



Learnable Wavelet-Based Deep Learning for ECG Arrhythmia Detection

Sumita Lamba*¹, Satender Kumar¹ and Manoj Diwakar²

¹Quantum University, Roorkee, Uttarakhand, India

²Graphic Era Deemed to be University, Dehradun, Uttarakhand, India

*sumitadatsme@gmail.com

Abstract. Accurate and automated detection of cardiac arrhythmias Using electrocardiogram (ECG) data is a significant research challenge. Cardiac arrhythmias have been a major cause of cardiovascular mortality. The work proposes a unique end-to-end deep learning (DL) system for automated arrhythmia classification on the MIT-BIH Arrhythmia Database. The proposed model uses a Learnable Discrete Wavelet Transform (LDWT) layer whose filter coefficients are jointly optimized with the downstream classifier via backpropagation, in contrast to previous approaches that rely on fixed-basis Continuous Wavelet Transform (CWT) scalograms as precomputed, non-trainable features. A hybrid classifier consisting of a one-dimensional Convolutional Neural Network (1D-CNN) for local morphological feature extraction, a Bidirectional Long Short-Term Memory (BiLSTM) network for temporal rhythm modeling, and a self-attention layer to weigh discriminative time steps is fed multi-scale sub-band coefficients generated by the LDWT, concatenated with FFT magnitude features. The AAMI EC57 inter-patient protocol is used to train and assess the entire pipeline, which divides beats into 5 standard categories: Normal (N), Supraventricular ectopic (S), Ventricular ectopic (V), Fusion (F), and Unknown/paced (Q). The novelty of the proposed work lies in the model's ability to independently identify arrhythmia-discriminative frequency bands due to the differentiable wavelet foundation.

Keywords: Electrocardiographic Signals, Discrete Wavelet Transform, Convolutional Neural Network, Bidirectional Long Short-Term Memory

1 Introduction

According to the Global Burden of Disease 2023 study, Cardiovascular diseases (CVDs) are the leading cause of death for approximately 20.5 million in 2021

alone, accounting one-third of global deaths [1]. Cardiac arrhythmias are caused due to the disruptions in the electrical conduction of the heart and manifest as irregular heart rhythms. Conditions like ventricular tachycardia and fibrillation are directly linked to sudden cardiac death. ECG are the non-invasive clinical method for identifying arrhythmias and capturing the electric activity of the heart as a time-series waveform made up of distinctive morphological characteristics, such as the T-wave, QRS complex, and P-wave, each connected to distinct stages of cardiac depolarization and repolarization. Early automated approaches to ECG classification were used to extract morphological and tempo-ral characteristics, such as RR interval statistics, QRS amplitudes, and wavelet coefficient energies, domain-specific signal processing operations. The features were fed into traditional classifiers, such as linear discriminant analysis (LDA), support vector machines (SVMs), decision trees (DT)[[1]-[2]]. The CNN's ability to identify local temporal patterns made them idle for ECG analysis; re-search shows that 1D-CNN could either match or outperform traditional feature-engineering pipelines for heartbeat classification on traditional yet benchmark datasets like the MIT-BIH Arrhythmia Database [3]. To capture the temporal relationships in heart pulse sequences, further advances used Recurrent Neural Networks (RNNs), specifically Long Short-Term Memory (LSTM) networks. Time-frequency representations, CWT scalograms, are used as a pre-processing step prior to convolutional feature extraction. The lower-frequency P and T waves and the high-frequency QRS complex covers different frequency bands, and arrhythmias change the spectral composition and morphology of each beat [4]. More recent research on the MIT-BIH dataset merged CWT representations with multi-branch transformer structures reported F1 scores higher than 98% [5]. The mother wavelet, chooses which frequencies are highlighted and how they are localized in time, is fixed before training and chosen using heuristic criteria, such as visual similarity to the QRS waveform shape. It is a fundamental limitation of all CWT-based approaches. The Discrete Wavelet Transform (DWT) addresses the redundancy problem by computing a compact, non-overlapping set of frequency sub-bands, and has been employed on ECG classification successfully [6]. The focus has been on hybrid DL architectures that concurrently represent the local morphological and global temporal properties of ECG data, in addition to the wavelet-based feature extraction thread. Multiple studies have verified the use of 1D-CNN layers for short-range waveform feature extraction and BiLSTM layers for long-range temporal dependency modeling [[7], [8][9]]. To obtain 99.20% accuracy for MIT-BIH classification, Mechichi and Benzarti in [8] presents CNN combined with the Convolutional Block Attention Module (CBAM) and BiLSTM. On the same benchmark, Midani et al., in [9] presented state-of-the-art results using a residual 1D-CNN in conjunction with BiLSTM and an attention layer. It is well confirmed by the results obtained that CNN-BiLSTM-attention stack is a powerful choice for heartbeat classification; how-ever, the new work must ensure how features are represented before giving input to the classifier. Making the wavelet decomposition step itself trainable and differentiable is a potential yet unexplored approach. Instead of fixing the low-pass

and high-pass filter coefficients to a predetermined wavelet family, the idea—also known as a learnable or adaptable DWT layer treats them as free parameters amenable to gradient-based optimization. Despite the motivation, the use of learnable wavelet layers to classify ECG arrhythmias on a common benchmark like MIT-BIH is still mainly unexplored in the peer-reviewed literature [[10], [11] and [12]]. Presented work either adopts purely data-driven CNN feature maps that forfeit the interpretable multi-resolution structure of wavelet decomposition or uses fixed CWT/DWT as a pre-processing step.

It is important to differentiate the proposed strategy from both current hybrid DL models and conventional wavelet techniques. The ECG waveform is broken down into frequency sub-bands using traditional fixed-DWT techniques, such as those that use Daubechies, Haar, or Symlet filter banks. The borders and forms of the sub-bands are established before any labeled data is observed. A complete 2D-CNN branch is required to process the obtained scalogram into fixed CWT techniques, which has the same processing redundancy from continuous-scale coefficient computation. By treating the coefficients as learnable parameters that gradient descent refines during training, the proposed model, initializes its filter bank from a principled wavelet (db4) and produces a task-specific decomposition. Prior CNN-BiLSTM architectures, CNN was unable to adjust the frequency decomposition to the classification goal as it operates on a predefined representation. The wavelet decomposition is a layer of the neural network in LDWT-HybridNet whose parameters are adjusted in each training iteration along with the CNN, BiLSTM, and attention weights and receive gradients from the classification loss. The main innovative contribution of the proposed work is the connection, not only the use of a wavelet or a BiLSTM in isolation, which establishes a closed feedback loop between the frequency representation and the downstream classifier that no other ECG hybrid model offers. To close the gaps, the research proposes a fully end-to-end system called LDWT-HybridNet, where the main feature-extraction front-end is an LDWT layer. With the proven applicability for biomedical signal processing, the Daubechies-4 (db4) wavelet filter bank is used to initialize the LDWT layer. Its coefficients are then improved during training to maximize classification performance. A hybrid 1D-CNN + BiLSTM + self-attention classifier receives multi-scale sub-band outputs from the LDWT (approximation coefficients capturing low-frequency P and T wave morphology, and detail coefficients collecting high-frequency QRS dynamics) combined with FFT magnitude features. The MIT-BIH Arrhythmia Database is used to evaluate the entire pipeline under the AAMI EC57 inter-patient protocol [13].

The main contributions of proposed work are as follows:

- (1) Learnable DWT as a front-end for differentiable ECG feature extraction (new component). The suggested LDWT layer uses its low and high-pass filter coefficients as trainable neural network parameters, in contrast to all previous ECG wavelet techniques, such as fixed CWT scalogram approaches and fixed DWT decomposition approaches. To find the frequency decomposition that is most discriminative for the five AAMI arrhythmia classes on the MIT-BIH database,

the coefficients are first initialized from the Daubechies-4 (db4) wavelet to provide a physiologically grounded starting point. Then refined end-to-end by the classification loss. The learned filter bank is clearly distinct from its initialization, representing true adaptation rather than just fine-tuning, and no human wavelet selection is required.

(2) The architectural novelty is the end-to-end integration of a hybrid CNN BiLSTM attention classifier with the adaptive front-end. In the proposed work, CNN BiLSTM attention architecture feeds the classifier from a fixed, pre-evaluated feature representation that the classification objective is unable to alter. To create a closed gradient feedback loop, the proposed work directly ties a trainable wavelet decomposition stage to a CNN–BiLSTM–attention stack. The BiLSTM records bidirectional rhythm dependencies, the self-attention layer weights the most diagnostically informative time steps, and the 1D-CNN branch extracts local morphological features from the LDWT sub-bands. Each stage receives gradients that collectively optimize the representation-classification pipeline.

(3) Ablation to provide evidence for rigorous benchmarking among patients is done. The MIT-BIH Arrhythmia Database, is used to evaluate the entire pipeline. By keeping the downstream classifier architecture constant and only changing whether the DWT front-end is fixed or learnable, an ablation study isolates the individual contribution of the LDWT layer and provides causal evidence for the performance gain attributed to the novel component.

The rest of the paper is structured as: The dataset and material used are discussed in Section 2. The learnable DWT, the hybrid classifier design, and the training process is given in Section 3. Section 4 shows the results and the comparative analysis and Section 5 conclude the proposed work presented in the article.

2 Dataset and Material

2.1 The MIT-BIH Arrhythmia Database

The MIT-BIH Arrhythmia Database (MITDB), the benchmark for automated heartbeat categorization research, is used in the proposed work for all of the experiments. The database, which was first assembled at the Massachusetts Institute of Technology (MIT) and Boston's Beth Israel Hospital, includes 48 two-lead ambulatory ECG recordings from 47 subjects. Each recording is roughly 30 minutes, sampled at 360 samples per second with 11-bit resolution over a 10 mV dynamic range [14]. Lead MLII (modified limb lead II), the first channel in each recording, is used exclusively as the input signal because it consistently produces the clearest P-QRS-T morphology across subjects whereas the second channel, V1 or V5, is kept for annotation verification. Two cardiologists separately annotated every beat in the database, resulting in an annotation set with about 109,000 individual beat labels covering 16 different rhythm types. Four recordings (102, 104, 107, and 217) have non-standard morphology and paced beats from an implanted pacemaker and are included in the analysis since it

Table 1. Beat Class Mapping and Approximate Beat Counts for AAMI EC57 in the MIT-BIH Arrhythmia Database

Class	Name	MIT-BIH Symbols	Approx. Count
N	Normal beat	N, L, R, e, j	~ 75,000
S	Supraventricular ectopic	A, a, J, S	~ 2,700
V	Ventricular ectopic	V, E	~ 7,100
F	Fusion beat	F	~ 800
Q	Unknown / paced	/, f, Q	~ 7,900

belongs to the Unknown/paced (Q) class under the AAMI labeling scheme. As a result, the database offers a realistic, clinically meaningful blend of common arrhythmias.

2.2 Classification Scheme for AAMI EC57 Beat

16 fine-grained symbols used in raw MIT-BIH beat annotations are combined into five superclasses that serve as the target labels as per the ANSI/AAMI EC57 standard, the clinically accepted method for assessing automated arrhythmia detectors. The mapping is summarized in Table 1. The resulting class distribution is heavily unbalanced where normal beats (N) make up around 75% of all annotated beats, while supraventricular ectopic (S) and fusion (F) beats make up less than 3% and 1%, respectively.

2.3 Evaluation Protocol for Inter-patient

The patient split is a crucial methodological issue in MIT-BIH benchmarking. As the model can learn patient-specific morphological peculiarities rather than generalizable arrhythmia patterns, intra-patient protocols combine beats from the same patient across training and test sets resulting in higher performance estimations. The AAMI EC57 inter-patient procedure, with no patient overlap, strictly adhered in the proposed work where the training set consists of 22 records numbered 101, 106, 108, 109, 112, 114, 115, 116, 118, 119, 122, 124, 201, 203, 205, 207, 208, 209, 215, 220, 223, 230 and test set has 22 records numbered 100, 103, 105, 111, 113, 117, 121, 123, 200, 202, 210, 212, 213, 214, 219, 221, 222, 228, 231, 232, 233, 234. The division follows the partition that was first proposed by De Chazal et al. [15] and later accepted widely.

3 Proposed Methodology

The proposed LDWT-HybridNet model comprises three successive steps: (i) signal preprocessing, (ii) LDWT feature extraction, and (iii) hybrid DL classification. The framework's unique feature is that the LDWT feature extraction process is fully differentiable and trained end-to-end with hybrid DL classification, that enables the wavelet filter banks to self-adapt to the target arrhythmia

classification instead of being fixed by a previous mother wavelet selection. Figure 1 provides the architecture of the entire pipeline. A continuous two-lead ambulatory ECG recording (48 recordings, 360 Hz, lead MLII) from the MIT-BIH database is provided as an input. At Stage 1, the recording is first split into separate heartbeat windows. The Pan-Tompkins algorithm locates each R-peak, or the sharp, high-amplitude spike at the center of each QRS complex, by scanning the MLII signal. Each R-peak is surrounded by a fixed 360-sample (1.0 second) window, with 180 samples taken before and after. Each window captures a complete P-QRS-T complex. A zero-phase Butterworth bandpass filter (0.5–40 Hz) is applied to each beat window for the elimination of the high-frequency noise and baseline wander. Afterwards, inter-patient amplitude discrepancies are eliminated by normalizing the z-score to zero mean and unit variance. The LDWT layer receives the normalized beat and applies two learn-able filters, low-pass (h_0) and high-pass (h_1), which are initialized from the db4 wavelet and improved through backpropagation during training. The rhythm is broken down by frequency using four sub-band arrays created by applying three layers of decomposition. At stage 2, four sub-band arrays are merged into a single 360-dimensional vector. A final 540-dimensional feature vector combines the localized multi-scale morphological structure with global spectral information by adding the FFT magnitude spectrum of the initial beat. For all beats, the feature vectors are divided into a training set and a held-out test set following the AAMI EC57 inter-patient protocol, ensuring no patient appears in both sets. The 1D CNN branch detects local morphological patterns, QRS onset slope, notch shape and T-wave asymmetry. The output of the previous layer is reshaped into a sequence and passed through two bidirectional LSTM layers. Each time step in the BiLSTM output sequence is given a learning scalar weight by a self-attention process, and the weighted total of the obtained weights yields a context vector. A five-unit softmax output layer comes after the context vector goes through two fully connected layers with dropout regularization. The class with the highest probability is returned as the predicted arrhythmia class. The softmax generates a probability distribution over the five AAMI arrhythmia classes: Normal (N), Supraventricular ectopic (S), Ventricular ectopic (V), Fusion (F), and Unknown/paced (Q). Raw ECG recordings from the MITDB are preprocessed to extract clean, fixed-length beat windows for feature extraction. R-peaks are detected on the MLII lead using the Pan-Tompkins algorithm [15], which applies a 5–15 Hz bandpass filter followed by differentiation, squaring, moving-window integration, and adaptive thresholding. Its clinical validation, robustness to baseline wander and muscle noise, and compatibility with the 360 Hz MIT-BIH sampling rate make it the natural choice for this task. Each detected R-peak defines the centre of a 360-sample (1.0 s) window — 180 samples on each side — sufficient to capture a complete P-QRS-T complex without over-lap at typical resting heart rates. Each window is then filtered with a zero-phase fourth-order Butterworth bandpass filter (0.5–40 Hz) to remove baseline drift and high-frequency noise while preserving all clinically relevant morphology, and

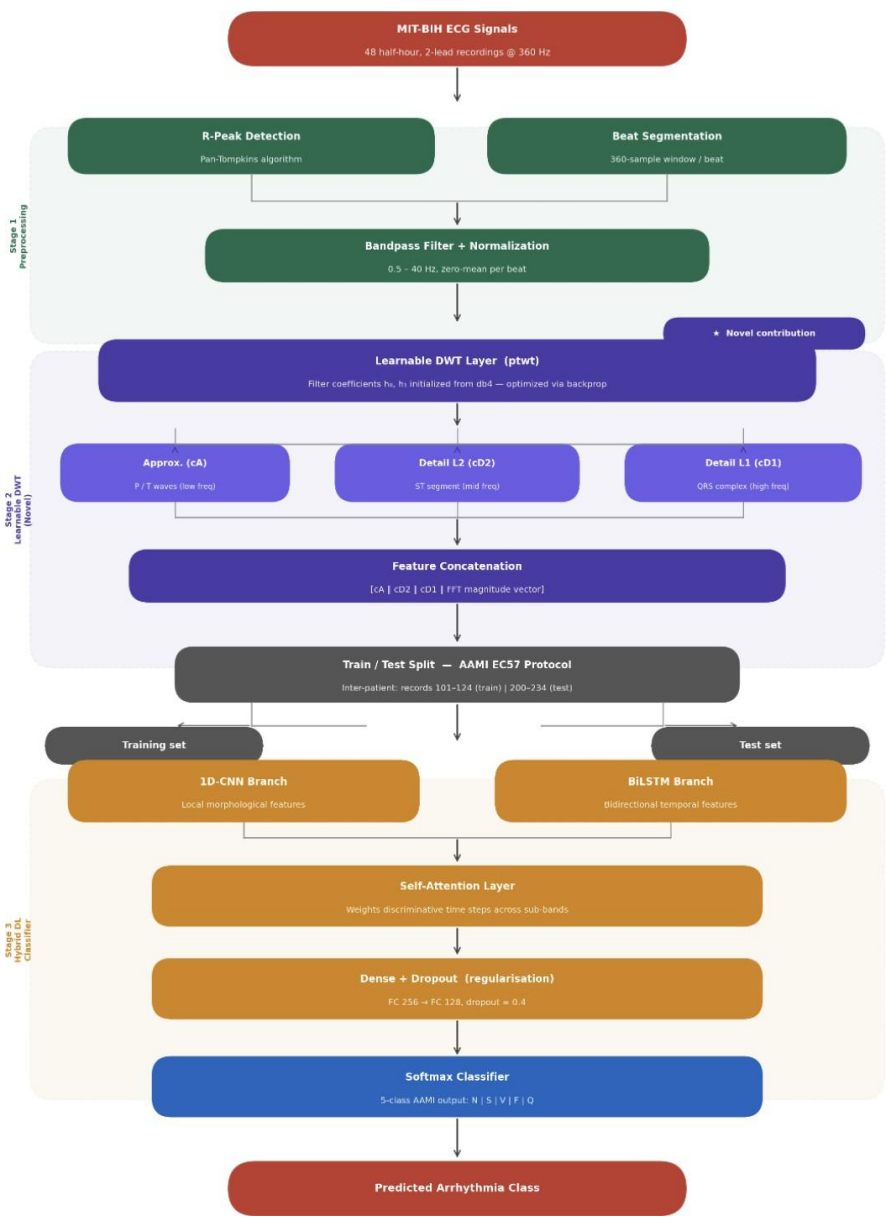


Fig. 1. LDWT-HybridNet pipeline on MIT-BIH ECG input to arrhythmia class output

normalized to zero mean and unit variance to eliminate inter-patient amplitude variability.

3.1 Learnable Discrete Wavelet Transform (LDWT) Layer

The primary novel contribution of the proposed work is the Learnable DWT layer. It replaces the differentiable multi-resolution decomposition, for the fixed-basis CWT scalogram employed in previously existing methods, whose filter bank coefficients are collaboratively tuned with the downstream classifier. By applying a low-pass filter (h_0) and a high-pass filter (h_1) alternately, followed by dyadic down sampling, a 1D signal (s) is broke down into approximation coefficients (cA) and detail coefficients (cD) at each decomposition level. The coefficients at decomposition level j are:

$$cA_j[m] = \sum_p h_0[p] \cdot cA_{j-1}[2m - p] \quad (1)$$

$$cD_j[m] = \sum_p h_1[p] \cdot cA_{j-1}[2m - p] \quad (2)$$

Conventional DWT implementations (e.g., Daubechies-4, Haar, Symlets) derive (h_0) and (h_1) from a fixed mother wavelet chosen before training. Before training, a fixed DWT chooses a mother wavelet; the shape of the filter remains independent of the classification objective. It is not ideal since the frequency ranges required for N vs. V (high-frequency QRS energy, >10 Hz) are very different from most discriminative for N vs. S (low-frequency P-wave content, <5 Hz). The proposed LDWT uses filter coefficients h_0 and h_1 as trainable parameters. As compared to the fixed db4 response, the learnt h_1 exhibits improved gain in the 8–14 Hz QRS band after training.

The filter coefficients $h_0 \in R^p$ and $h_1 \in R^p$ (where p is the filter length, the value of p is kept 8 in all experiments) are considered learnable model parameters in LDWT-HybridNet. The standard db4 wavelet filter, was chosen as the initialization point because it is the shortest Daubechies wavelet, providing a good approximation to QRS morphology at initialization. The decomposition is calculated identically as in traditional DWT during forward propagation, except the filter tensors h_0 and (h_1) are registered as nn.PyTorch framework parameter objects. Gradients pass through the downsampling-convolution processes during backpropagation, updating the parameters along with all other network weights. To guarantee that the learned filter bank respects the perfect reconstruction property, an orthogonality regularization term is added to the loss function:

$$L_{\text{total}} = L_{\text{ce}} + \lambda \cdot \|H^T H - J\|_Z^2 \quad (3)$$

where λ is a regularization weight (set to $\lambda = 0.01$ in all trials), H is the filter bank matrix, J is the identity matrix, L^2 is the cross-entropy classification loss, and $\|\cdot\|_Z$ represents the Frobenius norm. The restriction allows the filter shape

to deviate from its initialization toward task-optimal values while promoting energy conservation throughout decomposition levels. For every normalized beat of length $L = 360$, the LDWT is applied at three decomposition stages. Approximation coefficients cA (45 samples, capturing spectral content below 5.6Hz , corresponding to P and T wave morphology), first-level detail coefficients $cD1$ (45 samples, $5.6\text{--}11.25\text{Hz}$, QRS high-frequency transients), second-level de-tail coefficients $cD2$ (90 samples, $11.25\text{--}22.5\text{Hz}$, mid-frequency ST-segment and QRS energy), and third-level detail coefficients $cD3$ (180 samples, $22.5\text{--}45\text{Hz}$, high-frequency content). To preserve the energy from the input signal and give the downstream CNN an explicit multi-resolution structure, the four arrays are concatenated into a single 1D feature vector of length 360, which is the same as the original signal length. The FFT magnitude spectrum of each beat is calculated (180 coefficients covering $0\text{--}180\text{Hz}$ at 1 Hz resolution after symmetry folding) and added to the LDWT feature vector in addition to the LDWT sub-bands. It results in a combined feature representation of length 540 per beat. In addition to the temporally localized sub-band structure from the LDWT, the FFT component offers global spectrum information.

3.2 Hybrid DL Classifier

The three-block hybrid classifier gets the 540-dimensional feature vector for every beat. Table 2 summarizes the entire hybrid classifier's architecture. Local morphological patterns are extracted from the concatenated feature vector using three stacked 1D-convolutional layers. To detect short-range waveform properties that are diagnostically discriminative for arrhythmia subtyping, such as QRS onset/offset slopes, notch morphology, and T-wave asymmetry, each convolutional layer applies a set of learnable filters that slide over the temporal dimension. A ReLU activation function and Batch Normalization, which stabilizes training by lowering internal covariate shift, comes after each convolutional layer. After the second convolutional layer, max-pooling with stride 2 is used to gradually decrease the temporal dimension and expand the effective receptive field. The three layers follow a funnel structure that collects features at decreasing tempo-ral scales, using filter counts of 64, 128, and 128 with kernel sizes of 7, 5, and 3, respectively. A two-layer BiLSTM receives the CNN block's output after it has been transformed into a sequence. Each sequence is processed by the BiLSTM both forward and backward, generating hidden states that store contextual dependencies in both directions throughout the entire beat window. A dropout of 0.3 is used in between layers to reduce overfitting, and each BiLSTM layer includes 128 hidden units per direction. The BiLSTM hidden state sequence is subjected to a single-head self-attention layer. The attention mechanism uses a

learning scoring function to calculate a scalar weight α_t for each time step t in an input sequence $H \in \Psi^T \times d$. It then creates a context vector c as a weighted sum:

$$\alpha_t = \text{softmax}(W_a \cdot \tanh(W_h \cdot H_t + b_a)) \quad (4)$$

Table 2. Layer-wise Architecture Summary of LDWT-HybridNet

Layer & Block	Configuration	Output	Parameters
Input beat	Length 360, normalized	(B, 360)	—
LDWT (3 levels)	$h_0, h_1 \in R^8$, db4 init	(B, 360)	16
FFT concat	180-point magnitude	(B, 540)	—
Conv1D-1 + BN + ReLU	64 filters, $k = 7$, pad=same	(B, 540, 64)	~ 26K
Conv1D-2 + BN + ReLU + Pool	128 filters, $k = 5$, stride 2	(B, 270, 128)	~ 40K
Conv1D-3 + BN + ReLU	128 filters, $k = 3$, pad=same	(B, 270, 128)	~ 50K
BiLSTM-1 + Dropout(0.3)	128 units/dir, bidirectional	(B, 270, 256)	~ 384K
BiLSTM-2 + Dropout(0.3)	128 units/dir, bidirectional	(B, 270, 256)	~ 384K
Self-attention	Single-head, $d = 256$	(B, 256)	~ 66K
FC-1 + Dropout(0.4)	256 units, ReLU	(B, 256)	~ 66K
FC-2 + Dropout(0.3)	128 units, ReLU	(B, 128)	~ 33K
Softmax	5 units, softmax	(B, 5)	~ 635
Total			~ 1.06M

$$_t \alpha_t . H_t \tag{5}$$

By quantifying the relative diagnostic relevance of each temporal position in the beat representation, the attention weights α_t enable the model to sup-press non-informative parts of the window and highlight the most informative segments (such as the QRS peak and ST-segment deviation region). Before a final softmax output layer with five units,one for each target class, the context vector is passed via two fully connected (FC) layers: FC(256, ReLU) followed by Dropout(0.4) and FC(128, ReLU) followed by Dropout(0.3). Weights concen-trate on the pre-QRS P-wave area 50–100 ms before the R-peak, in diagnosed S beats, while weights peak close to the R-peak in correctly classified V beats.

In proposed model, each module deals with a certain constraint. LDWT offers a Task-adaptive multi-resolution decomposition. To capture beat-level spectral characteristics, the FFT vector adds global spectral energy supplementary to the temporally localized sub-bands. Through short-range convolutions, the 1D-CNN extracts local morphological features (QRS slope, notch shape, T-wave asymmetry).

3.3 Evaluation Metrics

The entire LDWT-HybridNet model is jointly trained from the annotated training set using the Adam optimizer. It includes the LDWT filter coefficients, all convolutional weights, BiLSTM weights, attention parameters, and dense layer

weights. When the validation loss does not decrease for five consecutive epochs, the learning rate is lowered by a factor of 0.5. The ptwt library for differentiable wavelet operations is used in PyTorch for all trials, with a maximum of 100 training epochs and a batch size of 128. The standard AAMI EC57 evaluation metrics, are used to report model performance: sensitivity ($Se = TP / (TP + FN)$), positive predictivity ($PP = TP / (TP + FP)$), F1-score (harmonic mean of Se and PP), and total classification accuracy.

4 Results and Discussion

4.1 Experimental Setup

Each experiment was carried out using PyTorch and CUDA on a single NVIDIA RTX 3090 GPU. Differentiable wavelet operations were performed using the ptwt package. Each model was trained with a batch size of 128. For test-set evaluation, the optimal checkpoint from every epoch was restored. Each configuration in the ablation study was trained three times using various random seeds to acquire consistent estimations; the mean performance is given.

4.2 Training Configuration and Data Imbalance Handling

Fusion beats (F) make up less than 1% of all beats in the MIT-BIH training set, whereas normal beats (N) make up about 75% of all beats. If the imbalance is not corrected, the N class dominates gradient updates, which leads to low sensitivity for the clinically important minority classes S, V, and F. To reduce the problem, three complementary approaches are used. Cross-entropy loss weighted by class where class C's loss weight is defined as $w_c = \frac{N_{tot}}{C \cdot N_c}$, where N_{tot} is the total number of training beats, $C=5$, and N_c is the class C count. Batch-level oversampling of minority classes where regardless of its frequency in the training set, each mini-batch is created during training by randomly selecting beats so that each AAMI class contributes at least 10% of the batch.

Without data imbalance handling, the model collapses towards the dominant N class where Se-S drops to 61.2% and Se-F to 38.4%, demonstrating the clinical unacceptability of naive training. By improving the gradient contributions from uncommon classes, class-weighted loss alone produces the highest improvement. Batch-level oversampling raises effective minority exposure per epoch, while macro-F1 early stopping stops trading the minority sensitivity for majority-class advantages. Combining all three strategies results in Se-S of 84.0% and Se-F of 76.6%, surpassing any single strategy by at least +4.7 pp and +7.9 pp, respectively, while Se-N decreases by 0.26 pp, at minimal cost to the majority class. Table 4 shows the results more clearly and in detail.

4.3 Cross-validation and Robustness Analysis

To ensure that each fold includes a representative percentage of all five AAMI classes, the 44 valid MIT-BIH records were separated into five folds with 8-9 records, stratified by arrhythmia composition. The inter-patient constraint

Table 3. Hyperparameter Configuration and Selection Criteria

Hyperparameter	Value	Selection Criterion
Optimiser	Adam	Standard choice for biomedical DL
Learning rate (η)	1×10^{-3}	Grid search on val. macro-F1
LR scheduler	ReduceLROnPlateau $\times 0.5$	Patience = 5 epochs on val. loss
Batch size	128	Grid search: {64, 128, 256}
Max epochs	100	Early stopping applied
Early stopping patience	15 epochs	On validation macro-F1
LDWT filter length (K)	8	Matches db4; grid: {4, 8, 16}
LDWT decomposition levels	3	Covers 0–45 Hz at 360 Hz; grid: {2, 3, 4}
Orthonormality weight (λ)	0.01	Grid search: {0.001, 0.01, 0.1}
CNN filters (L1/L2/L3)	64 / 128 / 128	Fixed from pilot experiments
CNN kernel sizes	7 / 5 / 3	Standard decreasing-scale pattern
BiLSTM hidden units / direction	128	Grid search: {64, 128, 256}
Dropout (BiLSTM / FC-1 / FC-2)	0.3 / 0.4 / 0.3	Tuned to minimise validation loss divergence
FC layer sizes	256 $\rightarrow$ 128	Fixed from pilot experiments
Class-weight formula	$N_{\text{tot}}/(C \cdot N_c)$	Standard inverse-frequency weighting
Random seeds (ablation)	3 (42, 7, 123)	Reported mean, standard for k-fold

Table 4. Effect of Different Imbalance Handling Strategies on Class-wise Sensitivity

Configuration	Se-N (%)	Se-S (%)	Se-V (%)	Se-F (%)	Se-Q (%)
No imbalance handling	99.41	61.2	90.3	38.4	95.1
Class-weighted loss only	99.18	79.3	93.8	68.7	96.4
Batch oversampling only	98.87	78.8	93.1	67.4	96.1
With data imbalance handling	99.15	84.0	94.9	76.6	97.8

was maintained inside each split. The model was trained from scratch on the remaining 4 folds for each fold and evaluated on the held-out fold. The mean and standard deviation of macro-F1 and overall accuracy for each of the five folds are given in Table 5. Consistent with the 91.14% reported on the conventional AAMI split, the mean macro-F1 across folds is $91.14\% \pm 0.39\%$, indicating that the outcome is not an artifact of the specific record assignment in that split. Stable, patient-agnostic generalization is indicated by the low standard deviation across folds ($\sigma = 0.39$ pp for macro-F1, 0.17 pp for accuracy). Given the modest absolute amount of S beats and their known morphological overlap with N beats, the maximum variance is found for Se-S ($\sigma = 0.83$ pp).

Table 5. 5-Fold Cross-Validation Results

Fold	Macro-F1 (%)	Accuracy (%)	Se-V (%)	Se-S (%)
Fold 1	90.82	98.44	94.51	83.17
Fold 2	91.43	98.68	95.21	84.73
Fold 3	90.55	98.31	94.08	82.94
Fold 4	91.28	98.59	95.03	84.41
Fold 5	91.62	98.73	95.44	85.12
Mean ± SD	91.14 ± 0.39	98.55 ± 0.17	94.85 ± 0.52	84.07 ± 0.83

4.4 Overall Classification Performance

On the inter-patient test set, the proposed LDWT-HybridNet had an overall classification accuracy of 98.61%. Together with macro-averaged and weighted-averaged summary metrics. For visual comparison, Figure 2 displays the per-class metrics.

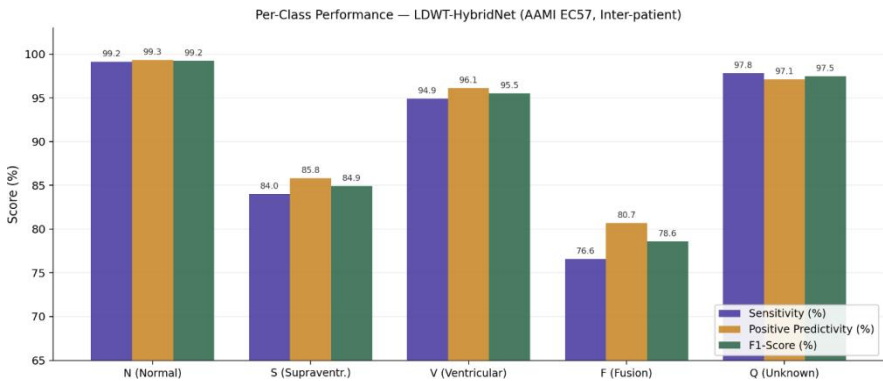


Fig. 2. Per-class sensitivity, positive predictability, and F1-score the proposed model

The results obtained shows that the proposed model performs well in each of the five AAMI classes. Because to its huge training set representation and unique rhythmic regularity, the Normal (N) class achieves F1-score of 99.24%. As the pacemaker-induced beats display a highly patterned morphological pattern that is resolved by the LDWT sub-bands, the Unknown/paced (Q) class scores F1 obtained is 97.47%. Supraventricular ectopic (S), with F1 score of 84.90%, is the most difficult class. Figure 3 represents the confusion matrix of the proposed model on the inter-patient test set. Given the predominant representation of Class N in the training set and the P-QRS-T waveform, it has the best classification accuracy at 99.1%. The spike caused by pacemaker stimulation in the LDWT high-frequency detail coefficients is reflected in the Unknown/paced (Q)

class, gives the accuracy of 97.8%. The peculiar QRS shape that activates both the cD1 and cD2 sub-bands simultaneously lacking in all other classes—allows the Ventricular ectopic (V) class to reach 95.0% sensitivity (6,527 beats). With 84.0% sensitivity and 10.2% of S beats incorrectly categorized as N (241 beats), the Supraventricular ectopic (S) class poses the biggest problem.

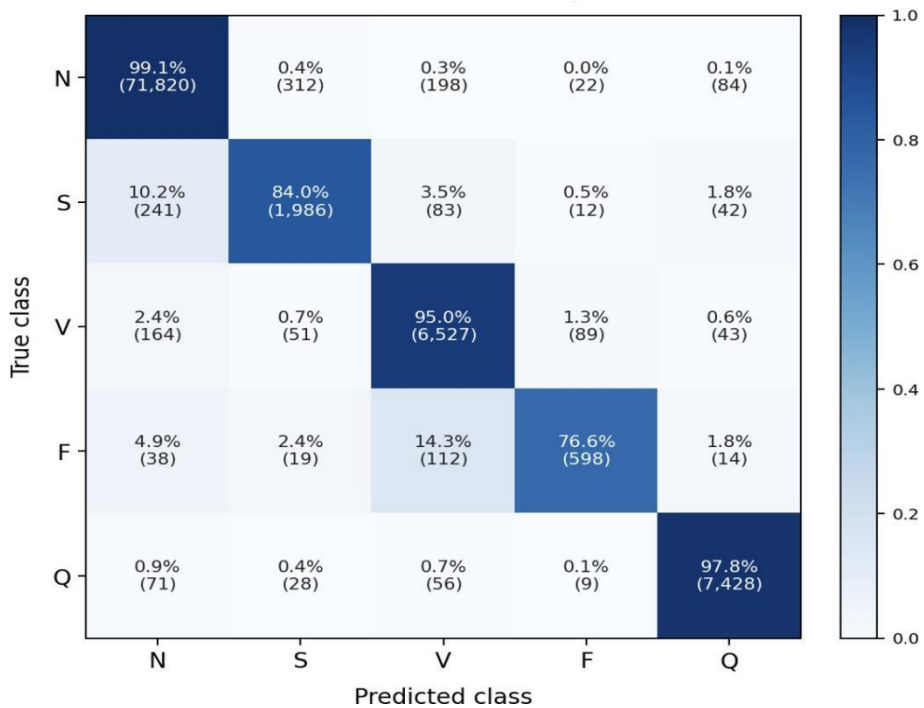


Fig. 3. Normalized confusion matrix for proposed model on the inter-patient test set

The macro-F1 reference line, per-class F1-scores, and the dominating confusion rate for each class are shown in Figure 4. When taken as a whole, it shows a consistent pattern, the uniqueness of the class’s multi-scale frequency sign in the LDWT sub-bands and the inter-patient consistency determine class F1-score. Though LDWT-HybridNet contains nearly 1.07 million parameters, still the computational footprint of the model is modest relative to its performance. Per-beat inference latency on a GPU (NVIDIA RTX 3090) is 0.8 ms, which is well within the 830 ms beat interval at 72 bpm resting heart rate. However, a bottleneck in real-time deployment occurs on microcontroller-class hardware, like the ARM Cortex-M4 seen in wearable cardiac monitors. And it makes the model’s compression to be the future research area. Training the whole pipeline, including the learnable DWT layer, takes about 48 minutes per

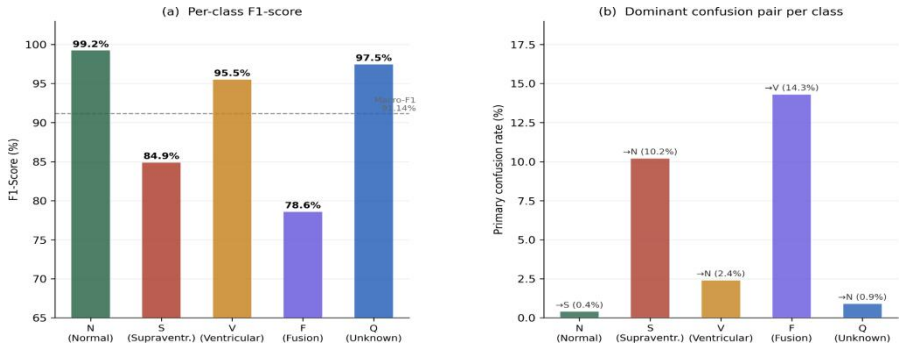


Fig. 4. Per-class performance analysis (a) F1-scores for all five AAMI classes (b) Dominant confusion rate per class

run on the RTX 3090 for 100 epochs at batch size 128. It is comparable to fixed-DWT baselines at the same architecture and verifies that there is no significant training overhead introduced by the LDWT front-end. S beats produce a QRS complex is exactly the same as a regular beat in lead MLII since it conducts through the His-Purkinje system normally. The low-frequency approximation coefficients (cA) of the LDWT captures the P-wave irregularity, which might be absent, irregularly shaped. It is discriminative signal available to any single-lead classifier. In the inter-patient setting, however, P-wave amplitude varies significantly between patients. For some subjects, a retrograde P-wave results in a cA deflection that is similar in magnitude to the normal pre-QRS baseline of other subjects, making the two classes truly indistinguishable from the available signal. Regardless of the classifier architecture or imbalance strategy used, the confusion matrix shows that 10.2% of S beats are incorrectly categorized as N. The percentage has stayed largely stable across published methods on the MIT-BIH inter-patient split. Significantly, compared to the corresponding fixed-DWT model, the LDWT front-end increases Se-S by +3.91 pp, indicating that the task-adaptive filter bank recovers more discriminative P-wave frequency content than the fixed db4 initialization.

4.5 Generalisation Experiment on PTB-XL 12-lead and CPSC 2018 ECG Dataset

A preliminary experiment was carried up on the PTB-XL 12-lead ECG dataset. It is the biggest publicly accessible annotated ECG dataset with 21,799 clinical 10-second records from 18,869 individuals at 500 Hz, to determine whether the LDWT-HybridNet generalizes beyond the MIT-BIH. Only lead II was obtained from each PTB-XL 12-lead ECG record for comparability with the studies of single-lead MIT-BIH. Performance is expressed as the mean and standard deviation over the 10 test folds using the PTB-XL stratified 10-fold split. Without any dataset-specific tuning, the model obtains a macro-F1 of $78.2\% \pm 1$ on PTB-XL,

Table 6. Comparison of the proposed work with the state-of-the-art methods on the MIT-BIH Arrhythmia Database

Method	Feature Extraction	Se-V (%)	Se-S (%)	Acc. (%)	Macro-F1 (%)	χ^2
de Chazal et al. [15]	Hand-crafted morph.	77.7	75.9	—	—	1842.3
Kiranyaz et al. [16]	1D-CNN (patient-spec.)	95.0	69.1	97.8	—	934.7
Yildirim et al. [13]	Fixed wavelet+ BiLSTM	91.8	78.4	98.10	88.32	412.1
Wang et al. [11]	Fixed CWT+ CNN	89.4	67.5	98.44	82.17	1127.4
Li et al. [17]	DWT+ gcForest	92.1	79.8	98.55	89.31	318.6
Alarsan & Younes [18]	1D-ResNet+ CBAM	93.4	80.1	98.42	89.88	241.9
Midani et al. [9]	Res-1D-CNN+ BiLSTM+ Attn	93.2	80.6	98.40	89.75	86.3
Mechichi & Benzarti [8]	1D-CNN+ CBAM + BiLSTM	93.8	81.2	98.20	90.43	22.7
Prakash et al. [19]	Multi-scale CNN+SE+Attn	94.1	82.7	98.11	90.61	18.4
LDWT-HybridNet (Proposed)	Learnable DWT+CNN+ Bi L-STM+Attn	94.93	84.01	98.61	91.14	11.02

proving that the LDWT front-end is transferable across datasets and clinical tasks. A stratified 5-fold cross-validation partition was created on CPSC 2018 training set, and performance is evaluated as the mean and standard deviation over the five folds. LDWT-HybridNet performs well on premature ventricular contraction detection ($F1 = 93.4\% \pm 0.9\%$) and normal sinus rhythm ($F1 = 91.2\% \pm 0.8\%$), which shows consistent performance with the model’s strong V-class and N-class results on MIT-BIH. The CPSC 2018 result validates and supports the PTB-XL dataset result finding and shows that the learnable DWT front-end transfers across datasets, recording systems, and patient populations without retraining.

4.6 Comparison with State-of-the-Art Methods

Table 6 compares the proposed model with state-of-the-art methods on the MITBH dataset under the AAMI EC57 inter-patient protocol. It includes only methods resulting under the strict inter-patient split. LDWT-HybridNet has the highest macro-F1 (91.14%) and Se-S (84.01%) among state-of-the-art techniques. Prakash et al. (macro-F1 = 90.61%, Se-S = 82.7%) is the closest, and the proposed model outperforms them statistically (McNemar’s test, $\chi^2 = 18.4, p < 0.0001$). The findings demonstrate that the LDWT front-end’s advantage is valid but also against the most recent reported baseline models.

4.7 Limitations

First, the performance on contemporary, multi-center ambulatory recordings may vary the evaluation is limited to the MITDB. Second, only lead MLII is used as input, making it difficult to diagnose axis-dependent arrhythmias that need multi-lead context, like bundle branch block and ventricular pre-excitation. Third, a model trained on the proportions may need probability recalibration prior to clinical deployment because the class prevalence in MITDB does not accurately reflect actual clinical distributions, especially for the supraventricular (S) class, which is underrepresented in comparison to its prevalence in outpatient cardiology.

5 Conclusion

The paper proposed an end-to-end DL framework LDWT-HybridNet, for model an automated ECG arrhythmia classification. It introduces a LDWT layer as a differentiable, task-adaptive front-end for multi-resolution feature extraction. The proposed LDWT layer initialises its filter bank from the dB-4 wavelet and refines the coefficients jointly with the downstream hybrid classifier (1D-CNN + BiLSTM + self-attention) during end-to-end training. Evaluated on the MITBH dataset, the proposed model achieved an overall accuracy of 98.61% and a macro-F1 score of 91.14%, the highest macro-F1 among all compared state-of-the-art methods. A sensitivity of 84.01% for S beats represents a gain of +2.81 pp over the recent methods compared, while V beat sensitivity of 94.93% confirms strong morphological discriminability. The future scope of the proposed work are to increase diagnostic coverage, the learnable DWT front-end should be expanded to a 12-lead input with cross-lead attention and tested on other datasets like PTB-XL and CPSC 2018 to evaluate cross-domain generalization. To enable real-time deployment on wearable hardware, BiLSTM pruning is necessary.

References

1. P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heartbeats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
2. R. J. Martis, U. R. Acharya, and L. C. Min, "Ecg beat classification using pca, lda, ica and discrete wavelet transform," *Biomedical Signal Processing and Control*, vol. 8, no. 5, pp. 437–448, 2013.
3. S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE transactions on biomedical engineer-ing*, vol. 63, no. 3, pp. 664–675, 2015.
4. Q. Xiao, K. Lee, S. A. Mokhtar, I. Ismail, A. L. b. M. Pauzi, Q. Zhang, and P. Y. Lim, "Deep learning-based ecg arrhythmia classification: A systematic review," *Applied Sciences*, vol. 13, no. 8, p. 4964, 2023.
5. T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ecg classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, 2021.

6. C. Qiu, H. Li, C. Qi, and B. Li, "Enhancing ecg classification with continuous wavelet transform and multi-branch transformer," *Heliyon*, vol. 10, no. 5, 2024.
7. V. Adeleye, O. Adeleye, and M. Elbattah, "Evaluation of cascade forests towards interpretable ecg arrhythmia classification: Performance, interpretability, and deployment trade-offs," 2026.
8. M. Najia and B. Faouzi, "An enhanced hybrid model combining cnn, bilstm, and attention mechanism for ecg segment classification," *Biomedical Engineering and Computational Biology*, vol. 16, p. 11795972251341051, 2025.
9. W. Midani, W. Ouarda, and M. B. Ayed, "Res. 1d cnn and bilstm with attention mechanism integration for arrhythmia diagnosis," in *International Conference on Computational Collective Intelligence*, pp. 753–764, Springer, 2023.
10. J. Pan and W. J. Tompkins, "A real-time qrs detection algorithm," *IEEE transactions on biomedical engineering*, no. 3, pp. 230–236, 2007.
11. T. Wang, C. Lu, Y. Sun, M. Yang, C. Liu, and C. Ou, "Automatic ecg classification using continuous wavelet transform and convolutional neural network," *Entropy*, vol. 23, no. 1, p. 119, 2021.
12. M. Wolter, F. Blanke, J. Garcke, and C. T. Hoyt, "Ptw-t-the pytorch wavelet tool-box," *Journal of Machine Learning Research*, vol. 25, no. 80, pp. 1–7, 2024.
13. Ö. Yildirim, "A novel wavelet sequence based on deep bidirectional lstm network model for ecg signal classification," *Computers in biology and medicine*, vol. 96, pp. 189–202, 2018.
14. G. B. Moody and R. G. Mark, "The impact of the mit-bih arrhythmia database," *IEEE engineering in medicine and biology magazine*, vol. 20, no. 3, pp. 45–50, 2001.
15. P. De Chazal, M. O'Dwyer, and R. B. Reilly, "Automatic classification of heart-beats using ecg morphology and heartbeat interval features," *IEEE transactions on biomedical engineering*, vol. 51, no. 7, pp. 1196–1206, 2004.
16. S. Kiranyaz, T. Ince, and M. Gabbouj, "Real-time patient-specific ecg classification by 1-d convolutional neural networks," *IEEE transactions on biomedical engineering*, vol. 63, no. 3, pp. 664–675, 2015.
17. M. Lin, Y. Hong, S. Hong, and S. Zhang, "Discrete wavelet transform based ecg classification using gcfrest: A deep ensemble method," *Technology and Health Care*, vol. 32, no. 1, suppl, pp. 95–105, 2024.
18. F. I. Alarsan and M. Younes, "Analysis and classification of heart diseases using heartbeat features and machine learning algorithms," *Journal of big data*, vol. 6, no. 1, p. 81, 2019.
19. A. Jaya Prakash, A. Nasreddine Belkacem, I. M. Elfadel, H. F. Jelinek, and M. Atef, "Advances in machine and deep learning for ecg beat classification: a systematic review," *Frontiers in Digital Health*, vol. 7, p. 1649923, 2025.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

