fibrillation/flutter (AFib), Supraventricular tachycardia (SVT), Junctional rhythm (Junc), Second-degree heart block type 1 (BII1), Second-degree heart block type 2 (BII2), Third-degree heart block (BIII), and Other. The consolidation algorithm generates Ventricular tachycardia (VT), Idioventricular rhythm (IVR), Intraventricular conduction delay (IVCD), Ventricular bigeminy (VBigem), Ventricular trigeminy (VTrigem), and Pause annotations using the BeatNet output and then splices together RhythmNet rhythms, ventricular rhythms, and artifact to create contiguous annotation files.

DL architecture

Both DL models rely on a similar architecture, which produces a sequence of classification results from a time series of single-channel ECG voltage values (Figure 2). The architecture is derived from preactivation ResNet,[13,14] a popular image classification architecture. Modifications to the architecture included replacing 2-dimensional (2D) convolutions with 1 dimension (1D) and removing the final pooling layer in order to repurpose the 2D image classification design to a 1D sequence-to-sequence classification design. As raw ECG flows through the network, it is compressed in the time dimension and extended depthwise. Compression occurs in the first convolution and at regular intervals throughout the remainder of the network. The input size and number of compression layers determine the model output resolution. The input to both DL models was 15,360 samples (60 seconds), which was compressed 5 times, resulting in 480 sequential outputs (every 0.125 second) for BeatNet, and compressed 8 times, resulting in 60 sequential outputs (every 1 second) for RhythmNet. Both architectures ended with a fully connected layer and softmax activation function, which produced classwise probabilities for each sequential output. The highest probability was selected as the label for each sequential output.

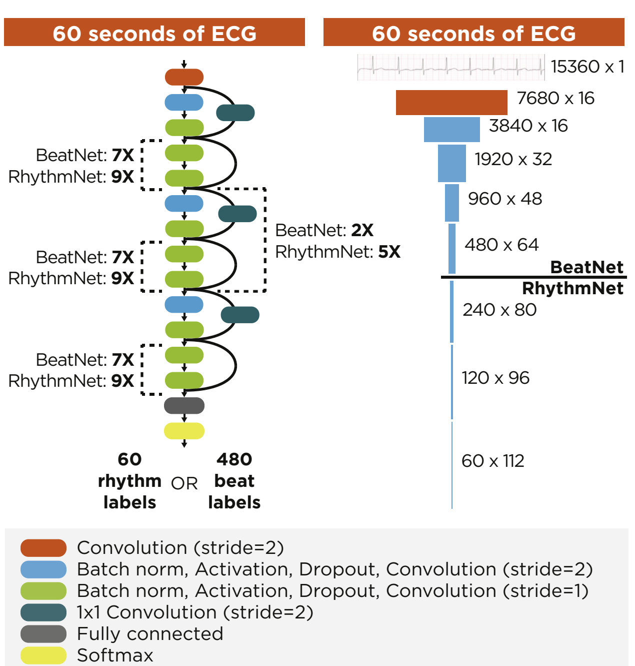

Figure 2 Deep learning model architecture. ECG = electrocardiogram.

ECG signal processing

ECG recordings were preprocessed using a wavelet high-pass (f_c = 0.5 Hz) filter[15] to remove baseline wander and 2 second-order Butterworth band-stop (f_c = 50 and 60 Hz) filters to remove powerline interference. After filtering, MIT-BIH and AFDB data were resampled to 256 Hz using linear interpolation.

Training record annotations

Training record annotations were generated for each model at the designed output resolution. BeatNet annotations were divided into 480 sequential classification labels consisting of Artifact, Not-a-beat, Ventricular ectopic, Bundle branch block, Normal, and Other. The Normal class included supraventricular ectopic beats, and the Other class included paced and unclassifiable beats. Sections with artifact onset/offset were labeled Artifact; sections with no beat and no artifact were labeled Not-a-beat; and sections in which a beat peak occurred anywhere within the 0.125-second window were labeled with the appropriate beat class label. Training records shorter than 60 seconds were padded using Other. RhythmNet annotations were divided into 60 sequential classification labels consisting of Sinus, AFib, BII1, BII2, BIII, SVT, Junctional, and Other. Rhythm transitions were labeled using the rhythm that spanned the majority of the 1-second region. Training records shorter than 60 seconds were padded using Other.

Model initialization and training

DL model weights were initialized in accordance with He et al[16] and trained using Adam[17] to optimize softmax cross-entropy. Padded and Other regions were masked in the training loss calculation. Mini-batch size and initial learning rate were optimized using the hyperparameter tuning process. A development dataset was partitioned from the training data, which was evaluated during training to implement early stopping and at the end of training to compare the performance of models with different hyperparameters. The development dataset contained at least 10 examples of each annotation, and no patient overlap was allowed between the development dataset and the remaining training dataset. After each training epoch (1 cycle through the full training dataset), micro-averaged training and development dataset F_1 scores were calculated, and the training dataset was randomly shuffled. During training, learning rate was reduced when the training dataset F_1 score did not improve for 5 consecutive epochs. Early stopping was invoked when the calculated PQ value[18] exceeded a threshold that was set using the hyperparameter tuning process.

Hyperparameter tuning

To fully define the model architecture and training procedure, model hyperparameters were optimized. Because preactivation ResNet was designed for image classification, this base

Table 2 Beat detection performance

Algorithm	Dataset	Se (%)	PPV (%)	F_1
Pan and Tompkins[24]	MIT-BIH	99.76	99.56	99.66
Christov[25]	MIT-BIH	99.74	99.65	99.69
Chiarugi et al[26]	MIT-BIH	99.76	99.81	99.78
Chouakri et al[27]	MIT-BIH	98.68	97.24	97.95
Elgendi[28]	MIT-BIH	99.78	99.87	99.82
State of the art	MIT-BIH	97.58	99.44	98.50
BeatLogic	MIT-BIH	99.60	99.78	99.69
Martinez et al[29]	MIT-BIH VFib excluded	99.80	99.86	99.83
Arzeno et al[30]	MIT-BIH VFib excluded	99.68	99.63	99.65
Zidelmal et al[31]	MIT-BIH VFib excluded	99.64	99.82	99.73
State of the art	MIT-BIH VFib excluded	97.58	99.57	98.56
BeatLogic	MIT-BIH VFib excluded	99.60	99.90	99.75
State of the art	Gold validation	95.79	96.32	96.05
BeatLogic	Gold validation	99.84	99.78	99.81

PPV = positive predictive value; Se = sensitivity; VFib = ventricular fibrillation.

architecture was reparameterized for sequence-to-sequence ECG classification in the context of BeatNet and RhythmNet. Hyperparameter optimization was performed using a combination of grid-search and tree-structured parzen estimator optimization[19] (for details see the Supplemental Material and Supplemental Table 4). The optimized BeatNet and RhythmNet models contained 81 and 113 convolutional layers.

State-of-the-art algorithm

The state-of-the-art algorithm was selected from several commercially available Food and Drug Administration (FDA)–cleared options capable of comprehensive beat and rhythm detection/classification. Candidate algorithms were evaluated using the EC57 standard, and the most accurate system was selected. Details of the selected algorithm are proprietary and were not disclosed to the authors for publication; however, the selected algorithm is known to leverage signal processing and classic machine learning techniques that are derived from the current ECG literature.

Validation procedure

Algorithm validation was performed in accordance with the EC57 guidelines.[10] EC57 is the FDA-recognized consensus standard and provides detailed instructions for measuring beat and rhythm detection/classification sensitivity (Se), and positive predictive value (PPV). Additionally, F_1 scores (0–100) were calculated for each validation metric per Equation 1.

$$F_1 = 2 \times \frac{Sensitivity \times PPV}{Sensitivity + PPV} \qquad \text{(Eq. 1)}$$

Results

Beat detection

On the MIT-BIH dataset, the BeatLogic platform performed equal to or better than 5 of the 8 previously published algorithms, whereas the state-of-the art algorithm outperformed only 1 published algorithm (Table 2). On the gold validation dataset, BeatLogic sensitivity was 99.84%, which exceeded the state-of-the-art algorithm by >4 percentage points. BeatLogic PPV was 99.78%, which exceeded the state-of-the-art algorithm by >3 percentage points (Table 2).

VEB classification performance

On the 11-record MIT-BIH data subset for measuring VEB performance, BeatLogic outperformed all other algorithms, achieving an F_1 score of 98.4, which is 0.8 points higher than the next highest performing algorithm (Table 3). On the gold validation dataset, BeatLogic outperformed the state-of-the-art algorithm, achieving sensitivity of 89.4% and PPV of 97.8% (Table 3).

Rhythm detection and classification

On the AFDB dataset, BeatLogic outperformed the previously published algorithms. The BeatLogic platform achieved episode Se/PPV of 97.7%/99.3% and duration sensitivity/PPV of 97.7%/99.7% (Table 4). On the gold validation dataset, BeatLogic outperformed the state-of-the-art algorithm for all 14 rhythms in measures of episode and duration sensitivity and PPV (Table 4). Three rhythm classes (junctional rhythm, second-degree heart block type

Table 3 Ventricular ectopic beat classification performance

Algorithm	Dataset	Se (%)	PPV (%)	F_1
de Chazal et al[22]	MIT-BIH 11	77.5	90.6	83.5
Jiang and Kong[3]	MIT-BIH 11	94.3	95.8	95.0
Ince et al[32]	MIT-BIH 11	90.3	92.2	91.2
Kiranyaz et al[20]	MIT-BIH 11	95.9	96.2	96.0
Zhang et al[8]	MIT-BIH 11	97.6	97.6	97.6
State of the art	MIT-BIH 11	73.2	96.3	83.2
BeatLogic	MIT-BIH 11	97.9	98.9	98.4
State of the art	Gold validation	36.0	51.2	42.2
BeatLogic	Gold validation	89.4	97.8	93.4

Abbreviations as in Table 2.

Table 4 Rhythm episode and duration performance

			Episode			Duration		
Rhythm	Dataset	Algorithm	Se (%)	PPV (%)	F_1	Se (%)	PPV (%)	F_1
AFib	AFDB	Petrucci et al[33] DRR	92.0	78.0	84.4	89.0	90.0	89.5
		Petrucci et al[33] RRP	91.0	92.0	91.5	93.0	97.0	95.0
		State of the art	63.3	100.0	77.5	65.3	99.3	78.8
		BeatLogic	97.7	99.3	98.5	97.7	99.7	98.7
AFib	Gold validation	State of the art	67.4	78.4	72.5	71.4	80.4	75.6
		BeatLogic	96.4	98.6	97.5	97.2	99.7	98.4
Sinus	Gold validation	State of the art	84.9	79.0	81.8	83.5	84.5	84.0
		BeatLogic	97.8	87.3	92.3	99.5	95.5	97.5
IVCD	Gold validation	State of the art	11.5	19.2	14.4	10.8	19.0	13.8
		BeatLogic	90.1	75.4	82.1	90.8	83.1	86.8
Artifact	Gold validation	State of the art	51.5	56.6	53.9	69.8	51.5	59.3
		BeatLogic	79.9	79.8	79.8	90.4	65.7	76.1
Pause	Gold validation	State of the art	69.8	100.0	82.2	67.6	99.9	80.6
		BeatLogic	97.7	93.2	95.4	92.0	93.7	92.8
SVT	Gold validation	State of the art	66.7	33.3	44.4	81.3	51.6	63.2
		BeatLogic	90.0	83.1	86.4	97.7	95.0	96.3
VT	Gold validation	State of the art	51.6	20.9	29.7	16.7	27.3	20.7
		BeatLogic	100.0	94.0	96.9	97.4	95.2	96.3
IVR	Gold validation	State of the art	61.7	33.8	43.7	60.5	28.5	38.7
		BeatLogic	83.0	98.0	89.8	63.8	96.4	76.8
Junctional	Gold validation	State of the art	—	—	—	—	—	—
		BeatLogic	91.3	73.9	81.7	84.9	77.5	81.0
VBigem	Gold validation	State of the art	62.3	75.3	68.2	29.1	77.5	42.3
		BeatLogic	100.0	98.6	99.3	99.2	98.7	99.0
VTrigem	Gold validation	State of the art	80.6	88.9	84.5	73.0	92.5	81.6
		BeatLogic	97.2	97.3	97.3	98.4	98.4	98.4
BII1	Gold validation	State of the art	—	—	—	—	—	—
		BeatLogic	56.9	93.2	70.7	72.6	97.7	83.3
BII2	Gold validation	State of the art	30.0	73.2	42.6	9.9	68.9	17.3
		BeatLogic	80.0	82.9	81.4	85.3	86.1	85.7
BIII	Gold validation	State of the art	—	—	—	—	—	—
		BeatLogic	98.7	95.8	97.2	93.2	97.2	95.1

AFib = atrial fibrillation/flutter; BII1 = second-degree heart block type 1; BII2 = second-degree heart block type 2; BIII = third-degree heart block; DRR = delta-RR; IVCD = intraventricular conduction delay; IVR = idioventricular rhythm; Junctional = junctional rhythm; RRP = RR prematurity; Sinus = sinus rhythm; SVT = supraventricular tachycardia; VBigem = ventricular bigeminy; VT = ventricular tachycardia; VTrigem = ventricular trigeminy; other abbreviations as in Table 2.

1, third-degree heart block) were not called at all by the state-of-the-art algorithm. State-of-the-art episode and duration F_1 scores exceeded 70 for 7 rhythms and exceeded 80 for episode detection of 3 rhythms. State-of-the-art episode and duration F_1 scores did not exceed 85 for any rhythm. BeatLogic episode and duration F_1 scores exceeded 70 for all 14 rhythms, exceeded 80 for 11 rhythms, exceeded 90 for 7 rhythms, and exceeded 95 for the following 5 rhythms: atrial fibrillation/flutter, ventricular tachycardia, ventricular bigeminy, ventricular trigeminy, and third-degree heart

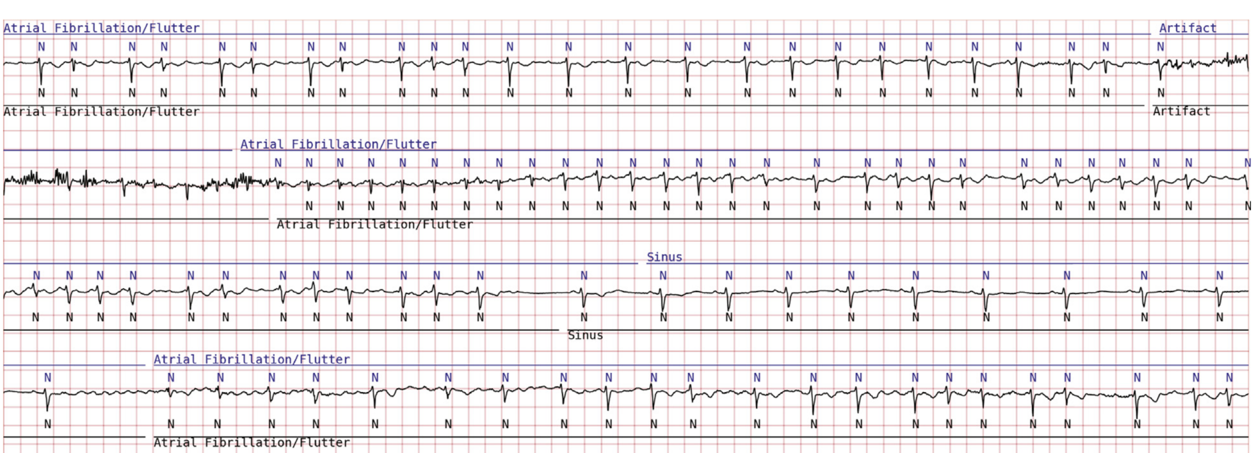

Figure 3 BeatLogic results *(blue)* compared with gold validation truth *(black)* demonstrating beat detection/classification, noise detection, and atrial fibrillation/flutter onset/offset.