Cardiologist-Level Arrhythmia Detection with Convolutional Neural Networks

arXiv, 6 July 2017

Pranav Rajpurkar, Awni Y. Hannun , Masoumeh Haghpanahi, Codie Bourn, Andrew Y. Ng
Stanford ML Group

Stanford Machine Learning Group

Our mission is to significantly improve people's lives through our work in Artificial Intelligence

PROJECTS PEOPLE PROGRAMS VOLUNTEERING CONTACT

Projects

We work on developing AI solutions for a variety of high-impact problems

MURA

Introducing a large dataset for abnormality detection from bone x-rays.

PROJECT WEBPAGE

Palliative Care

Using Electronic Health Record Data to direct palliative care resources.

PROJECT WEBPAGE

CheXNet

Radiologist-level pneumonia detection from chest X-rays.

PROJECT WEBPAGE

Arrhythmia

Cardiologist-level arrythmia detection from ECG signals.

PROJECT WEBPAGE

Education

Designing natural language models to detect writing errors and provide feedback.

PROJECT WEBPAGE

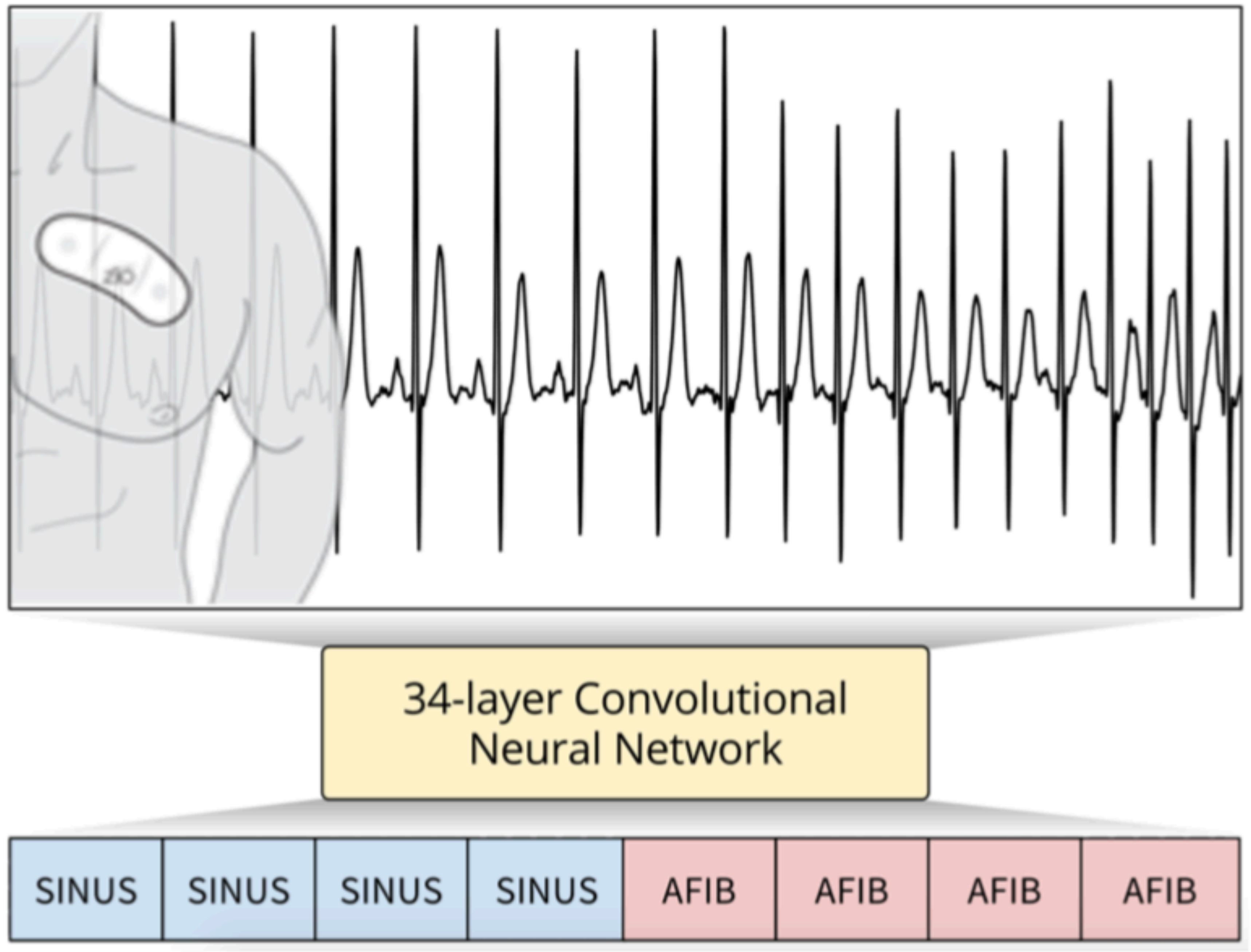
https://stanfordmlgroup.github.io

Cardiologist-Level Arrhythmia Detection with CNN

- Developed an algorithm which **exceeds the performance of board certified cardiologists** in detecting a wide range of **heart arrhythmias (irregular heart rhythms)** from ECG recorded with a **single-lead wearable monitor**

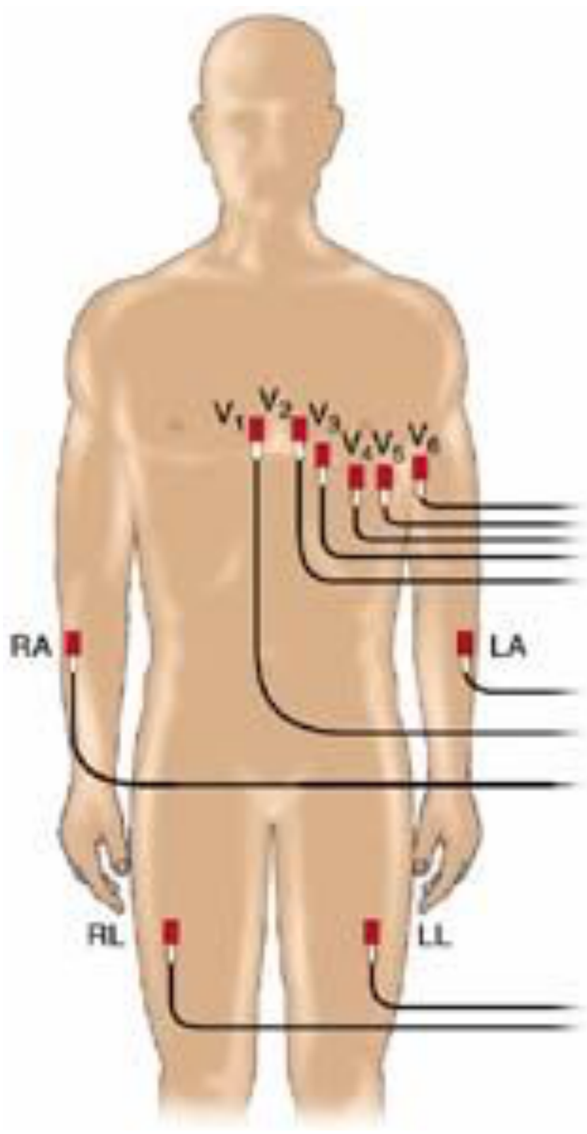
- Built a dataset with more than **500 times the number of unique patients** than previously studied corpora
  - Make class balance of the dataset more even

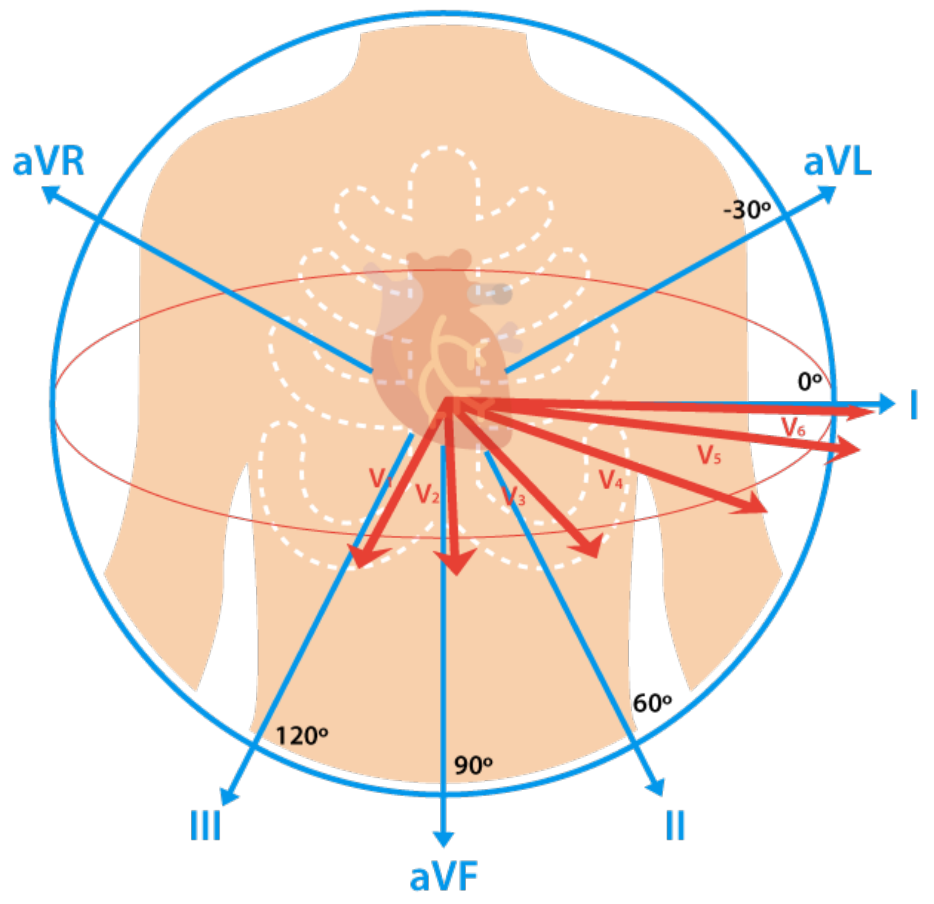
- Trained a **34-layer CNN** which maps a sequence of ECG samples to a sequence of rhythm classes



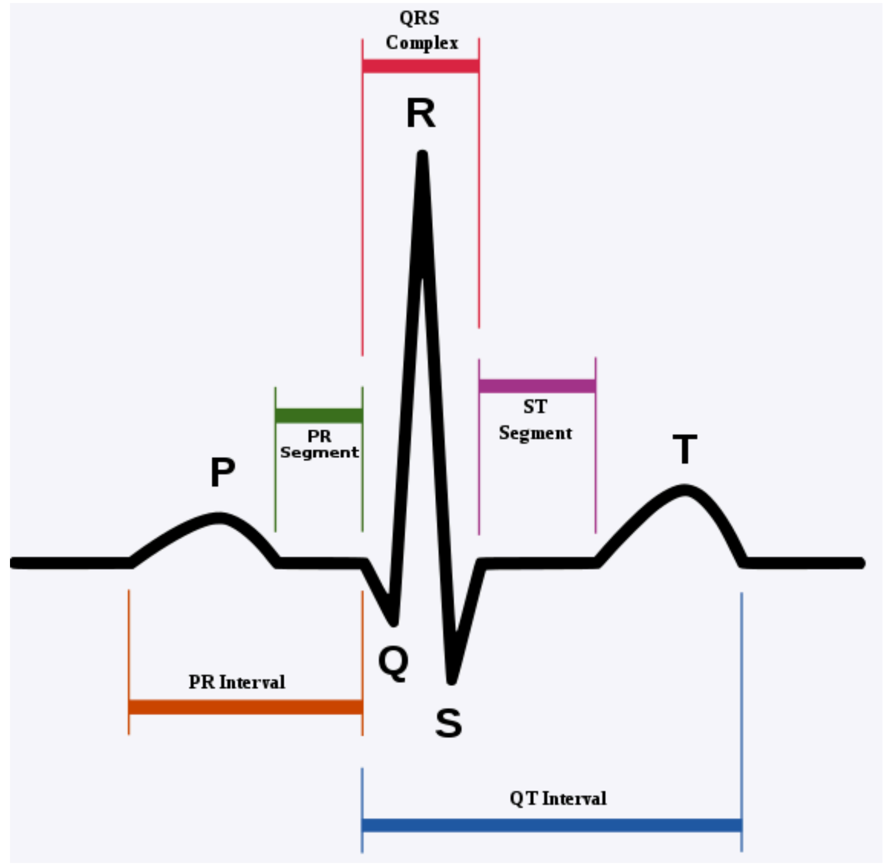
Electrocardiogram (ECG)

- ECG
  - The process of recording the **electrical activity of the heart** over a period of time using **electrodes** placed on the skin
  - Conventional **12-lead ECG**, 10 electrodes are placed and the overall magnitude of the heart's electrical potential is then measured from **12 different angles (leads)**





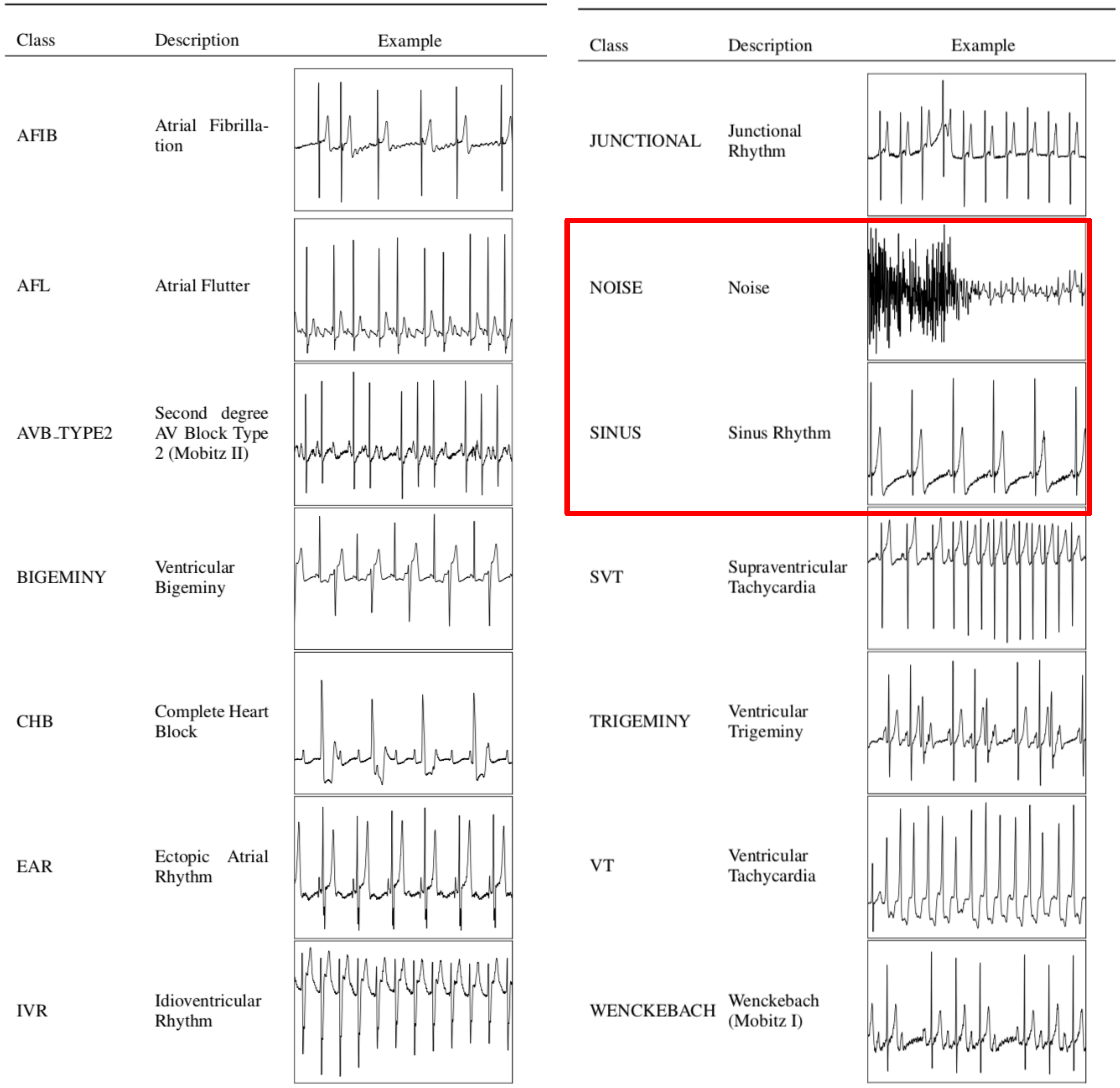
- ECG normal sinus rhythm
  - The **basic pattern of electrical activity** across the heart
  - It comprises three waves named **P, QRS, and T**



Heart arrhythmias

- Twelve Irregular heart rhythms

- Many heart diseases include Myocardial Infarction, AV Block, Ventricular Tachycardia and Atrial Fibrillation

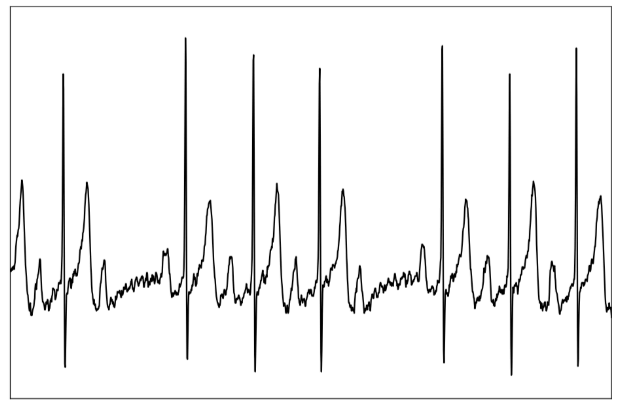
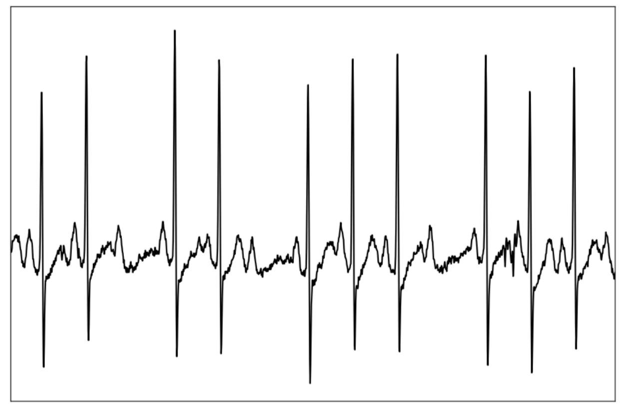


Class	Description	Example
AFIB	Atrial Fibrilla-tion	
AFL	Atrial Flutter	
AVB_TYPE2	Second degree AV Block Type 2 (Mobitz II)	
BIGEMINY	Ventricular Bigeminy	
CHB	Complete Heart Block	
EAR	Ectopic Atrial Rhythm	
IVR	Idioventricular Rhythm	
JUNCTIONAL	Junctional Rhythm	
NOISE	Noise	
SINUS	Sinus Rhythm	
SVT	Supraventricular Tachycardia	
TRIGEMINY	Ventricular Trigeminy	
VT	Ventricular Tachycardia	
WENCKEBACH	Wenckebach (Mobitz I)	

Motivation

- Arrhythmia detection is a challenging task
  - Must **implicitly recognize the distinct wave types** and discern the complex relationships between them over time
  - Difficult due to the **variability in wave morphology between patients as well as presence of noise**
  - Distinction between the rhythms can be subtle yet critical for treatment
    - Wenckebach is considered benign and Mobitz II is considered pathological requiring immediate attention

- Existing works performed poorly
  - One study, predictions for non-sinus rhythms, only about **50% were correct**
  - Another study, only **1 out of every 7 presentations** of second degree AV block were correctly recognized

WENCKEBACH	Wenckebach (Mobitz I)	
AVB_TYPE2	Second degree AV Block Type 2 (Mobitz II)	

Problem formulation

- ECG arrhythmia detection task is a **sequence-to-sequence task** which takes as input an ECG signals $X = [x_1, \dots x_n]$ and outputs a sequence of labels $r = [r_1, \dots r_n]$
  - Input: 30 second long ECG signal is sampled at 200Hz (30 x 200 length)
  - Output: a new prediction once every second (30 output)
    - Each labels take on one of the m different rhythm classes

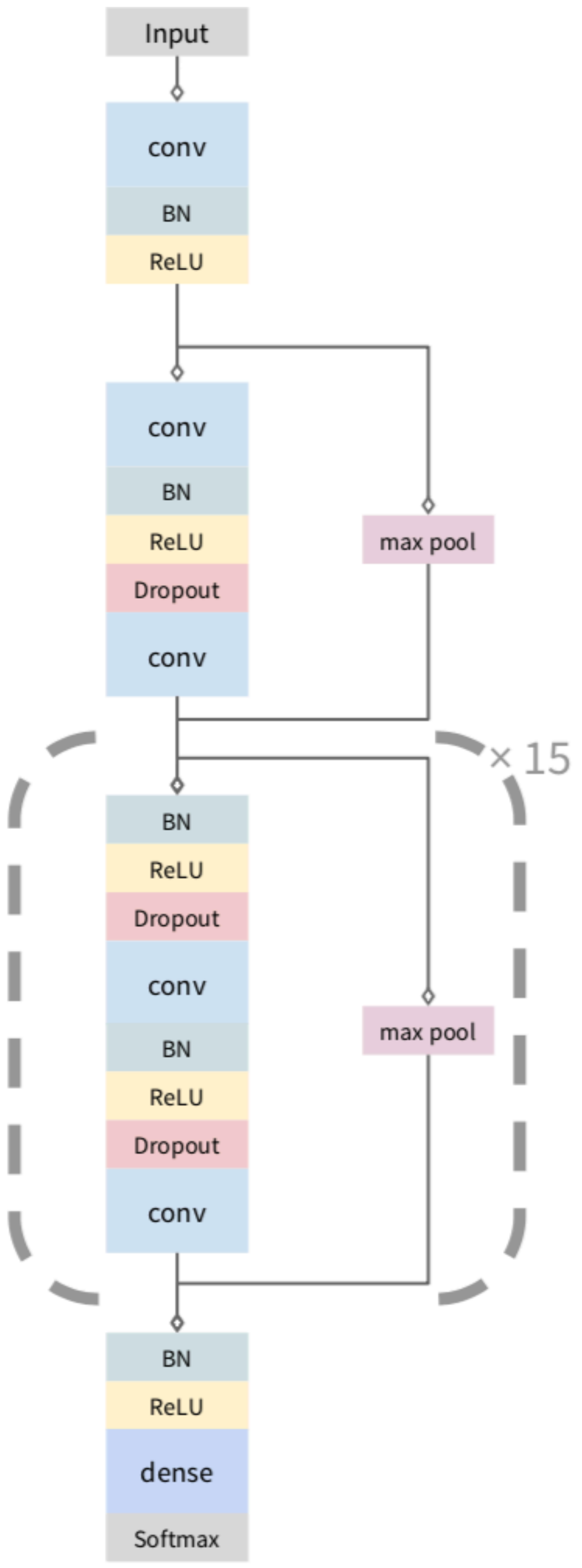
- For a single example in the training set, we optimize the cross-entropy objective function

$$\mathcal{L}(X, r) = \frac{1}{n} \sum_{i=1}^{n} \log p(R = r_i \mid X)$$

Model architecture & training

- Use CNN for the **sequence-to-sequence** learning task
- The network takes as input a time-series of raw ECG signal, and outputs a sequenc of label predictions

- Architecture
  - An architecture which is **33 layers of convolution** followed by a fully connected layer and a softmax
  - **Shortcut connections** (ResNet Architecture) employed
  - 16 residual blocks with 2 convolutional layers per block
  - Before input is fed into the network it is normalized using **robust normalization strategy**
  - Residual block: **Conv layer + Batch Norm + ReLU + Dropout + Conv layer**

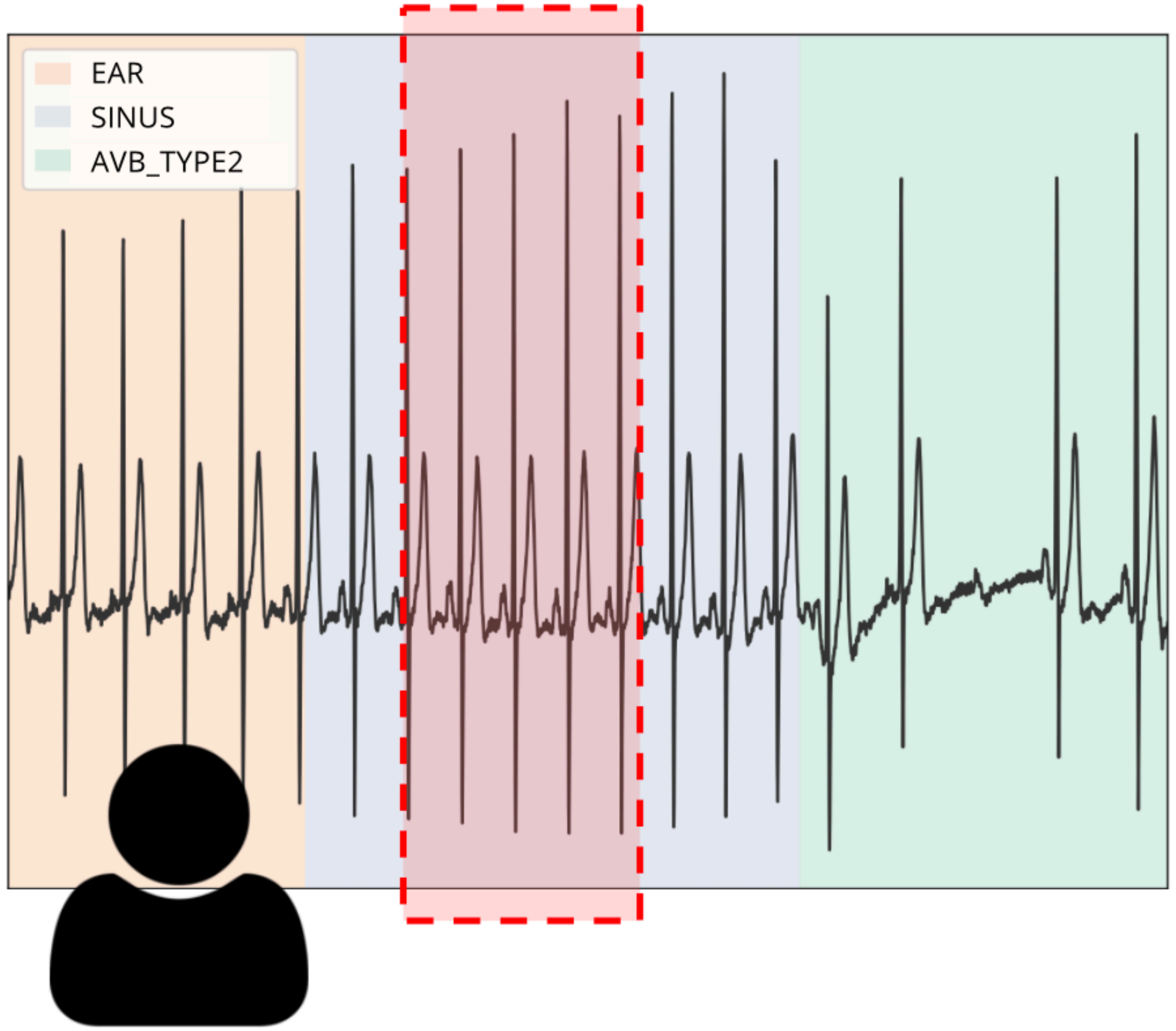


Data

- Annotation
  - The **three cardiologists discussed** each individual record **as a group and came to a consensus labeling**

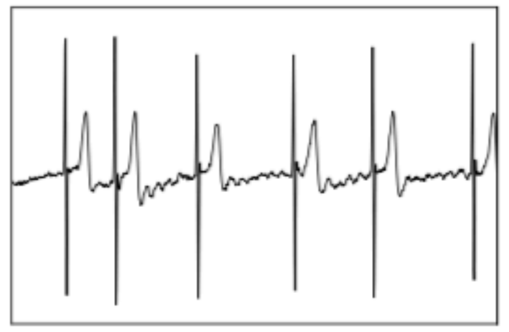
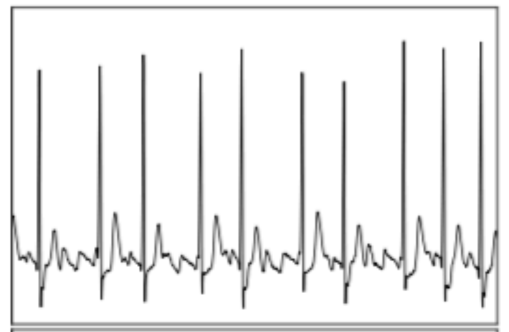
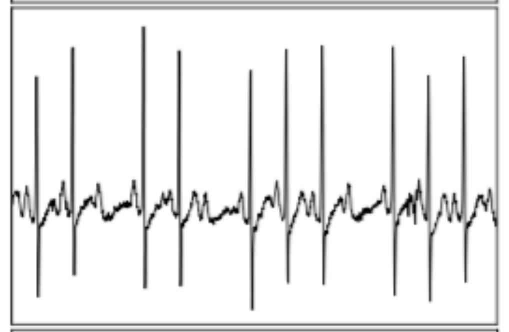
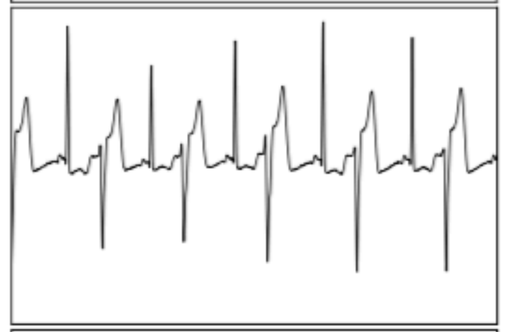
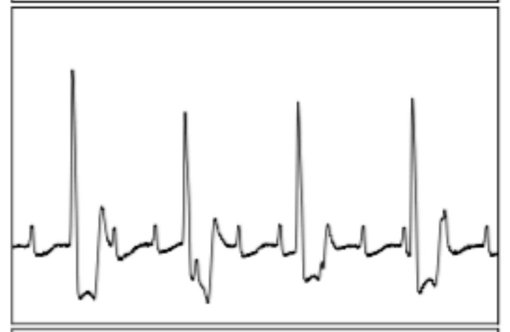
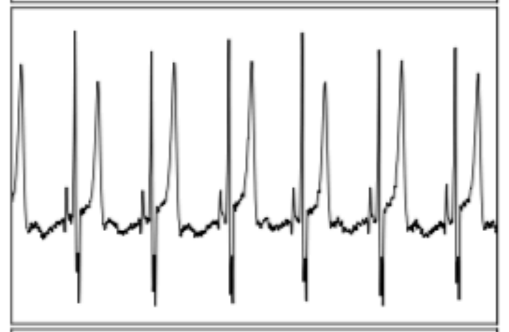
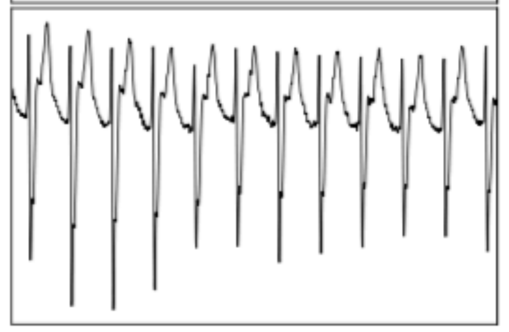
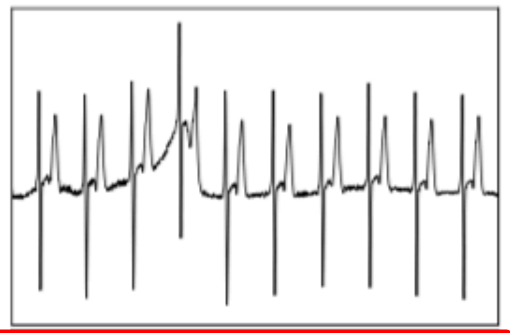
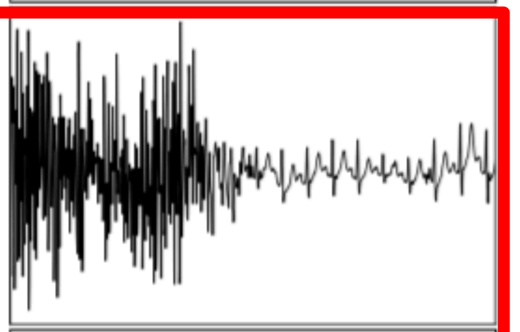
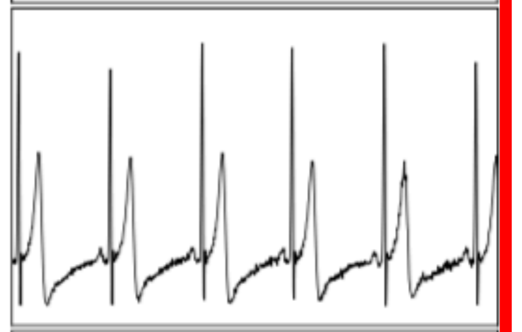
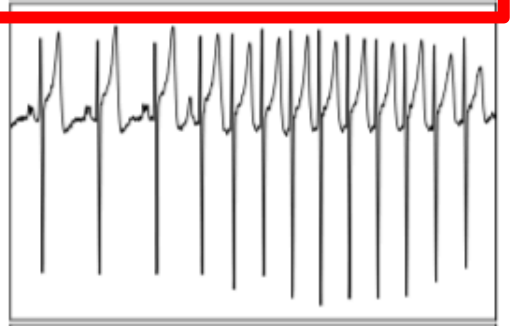
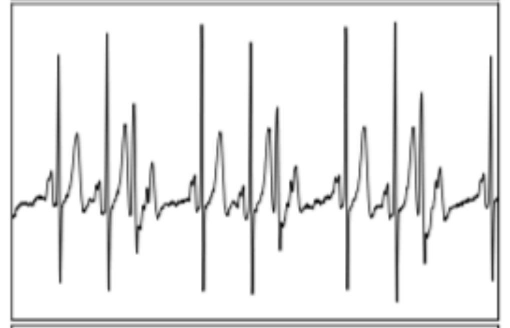
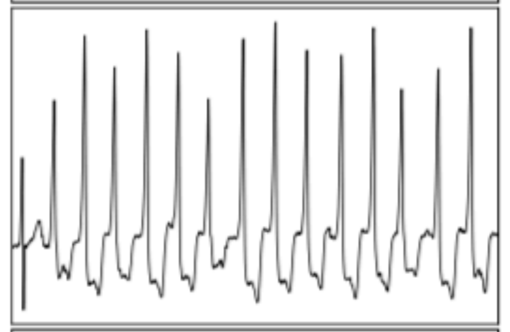
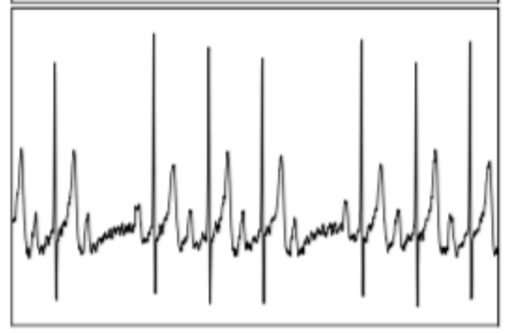
- Training
  - Collected a training dataset of **64,121 ECG records from 29,163 patients**
  - Sample rate: 200 Hz (collected from a single-lead monitor)
  - Each ECG record in the training set is 30 seconds long

- Testing
  - Collected a test set of **336 ECG records from 328 unique patients**
  - For each record in the test set **we also collect 6 individual annotations from cardiologists** not participating in the group (used for assessing performance of the model)



Rhythm classes

- The task is to identify 12 heart arrhythmias, sinus rhythm and noise for a total of **14 output classes**

Class	Description	Example
AFIB	Atrial Fibrillation	
AFL	Atrial Flutter	
AVB_TYPE2	Second degree AV Block Type 2 (Mobitz II)	
BIGEMINY	Ventricular Bigeminy	
CHB	Complete Heart Block	
EAR	Ectopic Atrial Rhythm	
IVR	Idioventricular Rhythm	
JUNCTIONAL	Junctional Rhythm	
NOISE	Noise	
SINUS	Sinus Rhythm	
SVT	Supraventricular Tachycardia	
TRIGEMINY	Ventricular Trigeminy	
VT	Ventricular Tachycardia	
WENCKEBACH	Wenckebach (Mobitz I)	

Evaluation Metrics

- Sequence Level Accuracy (F1)
  - **Average overlap** between the prediction and the ground truth sequence labels
- Set Level Accuracy (F1)
  - **The set of unique arrhythmias present** in each 30 second record is considered as the ground truth annotation
  - Set Level Accuracy does not penalize for time-misalignment within a record
- In both the sequence and set case, the F1 score for each class is computed separately

Model vs. Cardiologist Performance

- To assess cardiologist performance for each class, the average of all the individual cardiologist F1 scores are averaged using the group label as the ground truth annotation

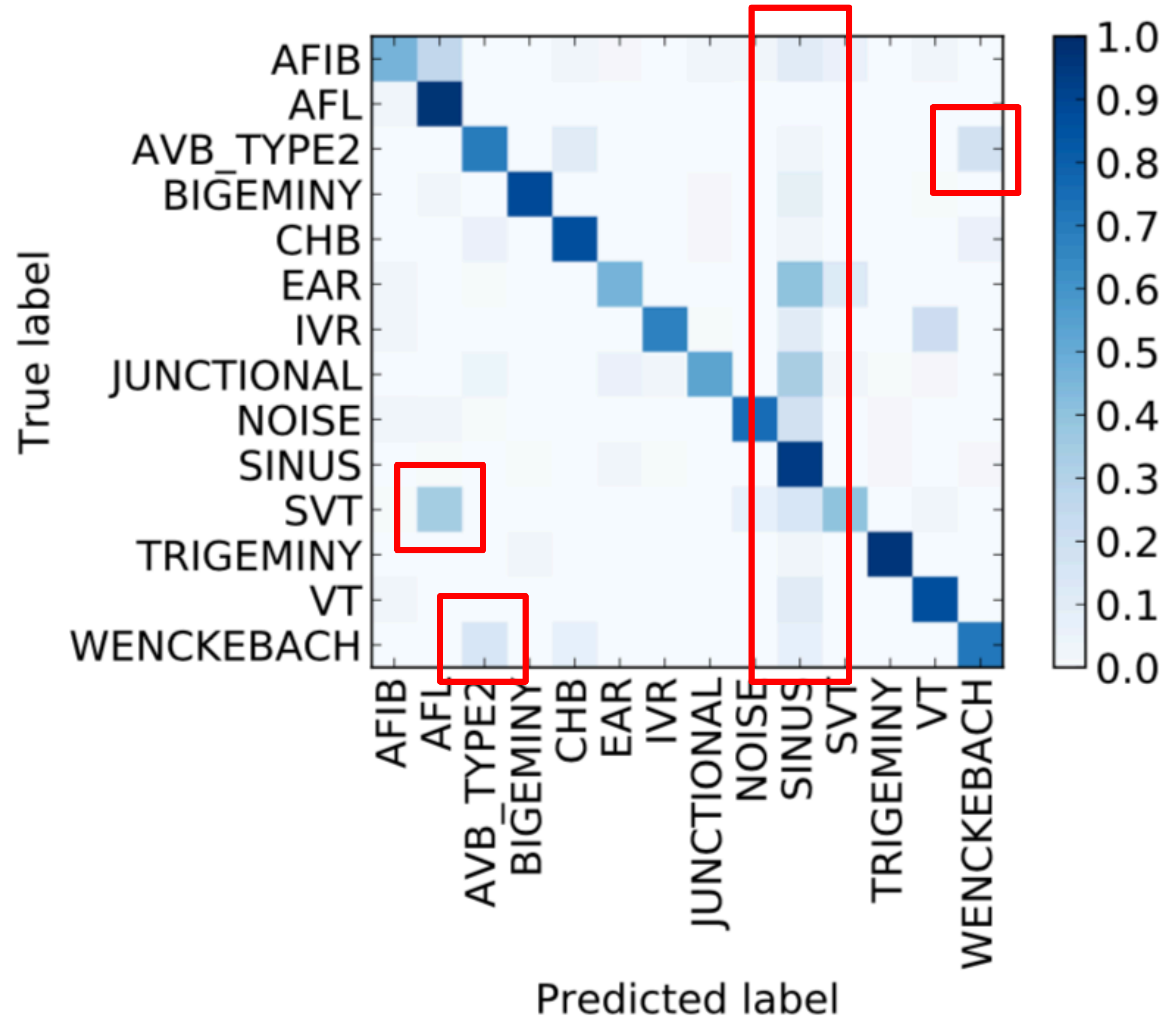
- **The model outperforms the average cardiologist performance on most rhythms**, noticeably outperforming the cardiologists in the AV Block set of arrhythmias which includes Mobitz I (Wenckebach), Mobitz II (AVB_type2) and complete heart block (CHB)

	Seq		Set	
	Model	Cardiol.	Model	Cardiol.
Class-level F1 Score				
AFIB	**0.604**	0.515	**0.667**	0.544
AFL	**0.687**	0.635	**0.679**	0.646
AVB_TYPE2	**0.689**	0.535	**0.656**	0.529
BIGEMINY	**0.897**	0.837	**0.870**	0.849
CHB	**0.843**	0.701	**0.852**	0.685
EAR	**0.519**	0.476	**0.571**	0.529
IVR	**0.761**	0.632	**0.774**	0.720
JUNCTIONAL	0.670	**0.684**	**0.783**	0.674
NOISE	**0.823**	0.768	**0.704**	0.689
SINUS	**0.879**	0.847	**0.939**	0.907
SVT	**0.477**	0.449	**0.658**	0.556
TRIGEMINY	**0.908**	0.843	**0.870**	0.816
VT	0.506	**0.566**	0.694	**0.769**
WENCKEBACH	**0.709**	0.593	**0.806**	0.736
Aggregate Results				
Precision (PPV)	**0.800**	0.723	**0.809**	0.763
Recall (Sensitivity)	**0.784**	0.724	**0.827**	0.744
F1	**0.776**	0.719	**0.809**	0.751

Mistakes made by the model are understandable

- Many arrhythmias are confused with the sinus rhythm, this is due to the sometimes **ambiguous location of the exact onset and offset of the arrhythmia in the ECG record**

- Often the **mistakes made by the model are understandable**
  - e.g., confusing Wenckebach and AVB_Type2 makes sense given that the two rhythms in general have similar ECG morphologies
  - Similarly, SVT and AFIB are often confused with AFL since they are all atrial arrhythmias
  - IVR is sometimes mistaken as VT which again makes sense given that two only differ in heart-rate and are difficult to distinguish close to the 100 beats per minute delineation



Conclusion

- Developed model **exceeds the cardiologist performance** in detecting a wide range of heart arrhythmias from single-lead EG records

- Key to the performance of the model is a **large annotated dataset and a very deep convolutional network** which can map a sequence of ECG samples to a sequence of arrhythmia annotations

- High-accuracy diagnosis from ECG can **save expert clinicians and cardiologists considerable time and decrease the number of misdiagnoses**

- Technology coupled with **low-cost ECG devices enables more widespread use of ECG as a diagnostic tool** in place where access to a cardiologist is difficult