Studie/Poster

# «REAL-WORLD PERFORMANCE OF ATRIAL FIBRILLATION DETECTION FROM WEARABLE PATCH ECG MONITORING USING DEEP LEARNING»

2019

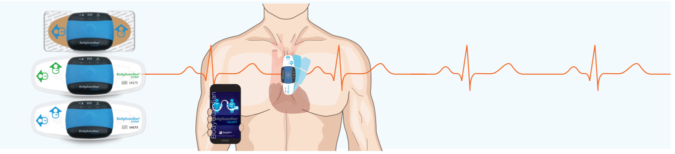
# REAL-WORLD PERFORMANCE OF ATRIAL FIBRILLATION DETECTION FROM WEARABLE PATCH ECG MONITORING USING DEEP LEARNING

MICHIGAN MEDICINE UNIVERSITY OF MICHIGAN | Preventice SOLUTIONS

Ben Teplitzky, PhD, Mike McRoberts, Pooja Mehta, Preventice Solutions, Rochester, MN and Hamid Ghanbari, MD, MPH, FACC, University of Michigan, Ann Arbor, MI

Disclosure: B. Teplitzky: Salary; Preventice. M. McRoberts: Salary; Preventice. P. Mehta: Salary; Preventice. H. Ghanbari: Consultant; Preventice.

## BACKGROUND



Wearable patch ECG monitoring enables continuous long-term monitoring outside of the clinic. During a monitoring study, service providers leverage human technicians and algorithms to analyze raw data and distill clinically relevant metrics into daily and end-of-study reports for the prescribing clinician. Atrial fibrillation (AFib) onset/offset detection and burden quantification of atrial fibrillation burden are important aspects of this reporting and must be performed with high sensitivity and precision in order to support clinical decision making.

New deep learning algorithms have demonstrated impressive performance when applied to AFib detection. However, many of these algorithms do not denote AFib onset/offset and very few have been validated on large, diverse, real-world datasets. Rigorous validation is particularly important for deep learning algorithms because of their capacity to "memorize" training data. This results in algorithms that do not perform well when presented with ECG that is significantly different than the training data. For example, algorithms suffering from an inability to generalize may perform well on ECG that contains 100% AFib or 100% normal sinus rhythm but may perform poorly when presented with rhythm transitions.

The following describes the Preventice BeatLogic™ deep learning platform for detecting and classifying cardiac arrhythmias. Real-world validation demonstrates performance of the algorithm for sinus rhythm and AFib using data that contains transitions into and out of AFib with varying durations.

## DEEP LEARNING MODELS

Two deep learning models form the foundation of the Preventice BeatLogic™ platform. The first detects and classifies beats and noise, the second detects and classifies rhythms. Both models leverage a convolutional neural network with residual connections and incorporate a repeating series of layers: 1-D convolution, batch norm, nonlinear activation, and dropout (below). Dropout is excluded following the first and last convolutions and classification is performed by a fully-connected and subsequent softmax layer. The rhythm classification model produces a prediction every one second and beat detection/classification produces a prediction every 0.125 seconds. Predictions from each model are generated within the context of a one-minute slice of ECG enabling each 1 or 0.125 second classification to gather context from the entire ECG. Results from the two models are merged in post-processing to create contiguous rhythm and artifact annotations.

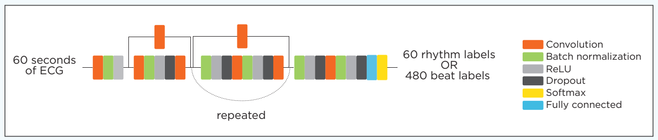



**Figure 1** Deep Neural Network Architecture

## TRAINING DATA

ECG training data was gathered from more than 10,000 patients who were monitored using the BodyGuardian® Heart (BGH) device. Rhythm labels were annotated and adjudicated by 3 certified ECG technicians, each with more than 5 years of experience. ECG records were captured from the mobile cardiac telemetry platform, which receives ECG segments that are generally 1 to 4 minutes in duration. Continuous and discontinuous atrial fibrillation were represented approximately equally in the dataset.

16,008 RECORDS
10,946 PATIENTS
395 HOURS OF ECG

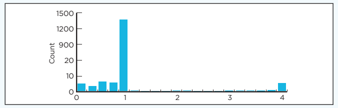


**Figure 2** Duration of AFib Episodes (min)

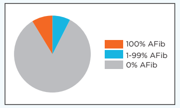



**Figure 3** AFib Burden

## VALIDATION DATA

ECG validation data was gathered from 512 patients who were monitored using the BodyGuardian® Heart (BGH) device. From the mobile cardiac telemetry platform, more than 2,500 ECG records were pseudo randomly captured to ensure rhythm and patient diversity within the validation data set. The large pool of candidate ECG records were filtered down so that at least 20 examples of each rhythm called by the Preventice BeatLogic™ platform were contained within the final validation data set. Rhythm labels were adjudicated by 3 board certified electrophysiologists. No patient crossover was allowed between the training and validation data sets.

522 RECORDS
512 PATIENTS
13.5 HOURS OF ECG

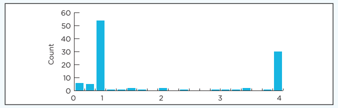


**Figure 4** Duration of AFib Episodes (min)

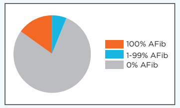



**Figure 5** AFib Burden

## CLASSIFICATION EXAMPLES

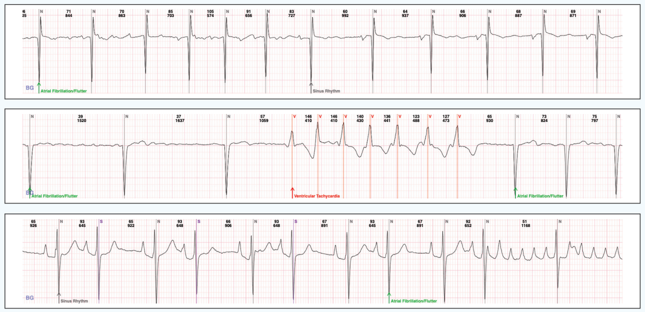


**Figure 6** Annotations Created by the Preventice BeatLogic™ Platform

## VALIDATION RESULTS

Algorithm validation was performed in accordance with the EC57 guidelines for assessing for cardiac rhythm measurement algorithms. This includes quantification of sensitivity and precision for episode (event detection) and duration (detected event overlap). The Preventice BeatLogic™ platform achieved state-of-the-art AFib and Sinus Rhythm classification performance.

Sensitivity was evaluated for AFib truth durations longer than 0, 30, and 60 seconds and precision was evaluated for AFib algorithm detections with durations longer than 0, 30, and 60 seconds. Episode sensitivity and precision improved to perfect for AFib events with duration longer than 1 minute.

**Table 1** Duration Performance for AFib and Sinus Rhythm

| | AFib | Sinus Rhythm |
|---|---|---|
| Duration Sensitivity (%) | 95.9 | 99.0 |
| Duration Precision (%) | 99.2 | 94.9 |

**Table 2** AFib Sensitivity by Duration Threshold

| | > 0 seconds | > 30 seconds | > 60 seconds |
|---|---|---|---|
| Episode Sensitivity (%) | 96.7 | 98.3 | 100 |
| Sensitivity Episode Count | 153 | 118 | 51 |
| Episode Precision (%) | 96.7 | 98.5 | 100 |
| Precision Episode Count | 153 | 116 | 51 |

## CONCLUSIONS

- The Preventice BeatLogic™ platform achieved state-of-the-art performance for detection and classification of both sinus rhythm and AFib, denoting onset and offset that is accurate to within a few heart beats.
- AFib sensitivity increased as the minimum duration for true AFib was increased to 30 and 60 seconds indicating that while the platform performs well on short bouts of AFib, much like humans, the system is better at correctly classifying longer durations.
- Within the academic literature most deep learning algorithms for AFib detection (1) fail to validate using real-world data, (2) do not demonstrate robust generalization by testing on a large unique patient data set, and/or (3) do not follow the standard EC57 guidelines for validation. These failures undermine published measures of performance and are addressed in this work.

## ACKNOWLEDGMENTS

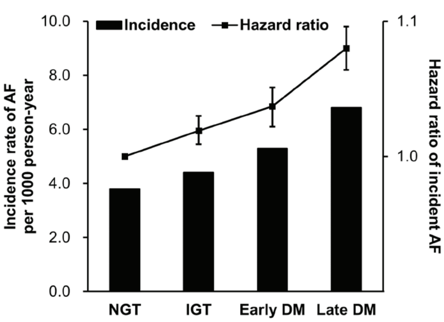
The authors would like to thank TM, ML, TA, and KP whose detailed annotations made this work possible as well as KW for lending their design expertise to this work.

## S-PO02-196

### INCREASED RISK OF ATRIAL FIBRILLATION WITH THE TIME COURSE OF DIABETES MELLITUS: A NATIONWIDE ANALYSIS

*Seokhun Yang, Eue-Keun Choi, MD, PhD, So-Ryoung Lee, Kyung-Do Han, Myung-jin Cha, MD and Seil Oh, MD, PhD, FHRS. Seoul National University Hospital, Seoul, Republic of Korea, Seoul National University Hospital Department of Internal Medicine, Seoul, Republic of Korea, Soon Chun Hyang University Hospital, Seoul, Republic of Korea, The Catholic University, Seoul, Republic of Korea, Department of Internal Medicine, Seoul National University Hospital, Seoul, Republic of Korea, Seoul National Univ Hospital, Seoul, Republic of Korea*

**Background:** Although diabetes mellitus (DM) is known for a risk factor of atrial fibrillation (AF), the relation of progression of DM and risk of AF is not well known.
**Objective:** We sought to find the association between the time course of DM and incidence of AF.
**Methods:** A total of 8,244,688 participants (aged ≥ 40 years, no history of AF) were included from Korean National Health Insurance Service database between January 2009 and December 2012. We divided the study population into 4 groups: subjects with normal glucose tolerance (NGT, n=4,922,678), subjects with impaired glucose tolerance (IGT, n=2,059,680), patients with DM duration < 5 years (early DM, n=730,411), and patients with DM duration ≥ 5 years (late DM, n=531,919). We compared the incidence and the risk of AF among groups.
**Results:** During a mean follow-up of 6.7±1.1 years, incidence rate of AF significantly increased across the groups (3.8, 4.4, 5.3, and 6.8 per 1000 person-years in NGT, IGT, early DM, and late DM, respectively, p-for-trend <.001). With NGT as reference group, the risk for incident AF increased significantly across the time course of DM (adjusted hazard ratio (aHR) 1.019, 95% confidence interval (CI) 1.009-1.030 for IGT; aHR 1.037, 95% CI 1.022-1.051 for early DM; aHR 1.080, 95% CI 1.064-1.096 for late DM). In subgroup analysis, increasing tendency of risk for incident AF according to duration of DM was more prominent in group of age <65 years.
**Conclusion:** Duration of DM showed significant association with the risk of AF, especially in those with younger (<65 years) patients.




AF: Atrial fibrillation
IGT: Impaired glucose tolerance
NGT: Normal glucose tolerance
DM: Diabetes mellitus
Early DM: DM duration < 5 years
Late DM: DM duration ≥ 5 years

## S-PO02-197

### REAL-WORLD PERFORMANCE OF ATRIAL FIBRILLATION DETECTION FROM WEARABLE PATCH ECG MONITORING USING DEEP LEARNING

*Benjamin Adam. Teplitzky, Michael McRoberts, Pooja Mehta and Hamid Ghanbari. Preventice Solutions, Eagan, MN, Preventice Solutions, Rochester, MN, Univesity of Michigan, Ann Arbor, MI*

**Background:** Wearable patch ECG monitoring quantifies atrial fibrillation (AF) burden using a combination of algorithms and trained technicians. New deep learning algorithms have improved the performance of atrial fibrillation detection in simple classification tasks where the goal is to detect the presence of AF without denoting the beginning and end of each AF event. However, the beginning and end of AF events must be known for accurate burdens to be calculated.
**Objective:** To demonstrate real-world performance of the Preventice BeatLogic deep learning platform for detecting the beginning and end of AF events.
**Methods:** We trained two deep residual convolutional neural networks (DNN), a beat detector/classifier (34 layers) and a rhythm classifier (42 layers) using 260 hours of ECG from 4,670 BodyGuardian® Heart (BGH) patients. Beats and artifact were detected/classified with 0.125 sec resolution while sinus rhythm and AF were detected/classified with 1 sec resolution. Results from the two DNNs were merged in postprocessing to produce rhythm beginning and ends for a variety of cardiac rhythms. Performance was measured using real-world data consisting of 400 hours of BGH data from 3,600 patients not included in the training dataset.
**Results:** AF duration sensitivity (Se) and a positive predictive value (PPV) were 93.8% and 94%, respectively. Episode detection Se and a PPV were 95.8% and 89.3% respectively. Episode detection Se increased to 97.6% and 99.2% for AF episodes with duration >30 sec and >60 sec, respectively.
**Conclusion:** ECG algorithms must be able to accurately detect the beginning and end of arrhythmias in order to ensure accurate burden calculation and maximize clinical value. Here we describe and validate a technique for achieving this goal by integrating the results from multiple DNNs. The proposed system detects AF episodes at rates comparable to the best reported values within the literature and detection performance improves to near perfect for episodes lasting one minute or more. The suboptimal episode PPV indicates the continued need for trained technicians to monitor low confidence AF classifications within a production environment in order to ensure that false AF detections are correctly excluded from burden calculations.

## S-PO02-198

### DIFFERENTIAL MANIFESTATION OF LATE-ONSET ATRIAL TACHYARRHYTHMIA IN PATIENTS WHO UNDERWENT ISOLATED VALVE SURGERY OR CORONARY BYPASS GRAFTING

*Min Soo Cho, MD, Jun Kim, Ungjeong Do, Gi-Byoung Nam, MD, Kee-Joon Choi, MD and You-Ho Kim, MD, FHRS. Asan Medical Center, Seoul, Republic of Korea, Asan Medical Center, Dept Of Internal Medicine and Cardiology, Seoul, Republic of Korea*

**Background:** Incidence and predictors of atrial tachyarrhythmias (ATA) after cardiac surgery are yet to be fully evaluated.
**Objective:** We hypothesized that the type of surgical procedure would have differential implications on the incidence and initial presentation of late-onset ATAs.
**Methods:** Data on 5,411 patients who underwent single cardiac procedure (801 isolated mitral valve surgery [MVS]; 1,066 isolated aortic valve surgery [AVS]; 3,544 bypass grafting [CABG]) at our center from 2006 to 2016 were analyzed. Those who had combined surgery, prior ATAs, or ATAs during hospitalization were excluded from the analysis. The major study outcome was the incidence of ATAs that occurred after discharge from index hospitalization for surgery.
**Results:** Patients who underwent MVS were younger, had lesser comorbidities, and had larger left atrial (LA) and left ventricular (LV) dimensions compared to other groups. During 5 years of follow-up, the primary outcome rate was higher in those