MaxYield™: Validation and Comparison with ECG Technicians

Michael D. Feist, Talwinder Punni, and Hunter Vander Linden

NeuralCloud Solutions Inc.

ask@theneuralcloud.com

1. Executive Summary

Cardiovascular diseases (CVDs) remain a major global health concern, and while current tools for arrhythmia detection and diagnosis are widely used, each carries important limitations. Standard 12-lead ECGs provide only a brief snapshot of cardiac activity, often missing paroxysmal or intermittent arrhythmias. Holter monitors extend recording over multiple days, but the resulting volume of data is time-consuming to review, and existing algorithms frequently underperform in the presence of noise and artifacts. As a result, manual correction by experts is often required, contributing to delays, variability, and the risk of missed diagnoses.

To address these challenges, NeuralCloud Solutions developed MaxYield™, a neural network algorithm designed for the analysis of long-duration ambulatory ECG recordings. MaxYield™ reduces noise and artifacts while accurately identifying the onsets and offsets of P-waves, QRS complexes, and T-waves, the foundational features of ECG interpretation. The combination of a cleaned ECG signal and precise beat-level labeling enables rapid post-processing to identify cardiac events and areas of concern, thereby streamlining diagnosis and reducing variability.

Validation studies demonstrated that MaxYield™ performs on par with expert human annotators in terms of accuracy, while providing superior consistency as reflected in lower median error and narrower interquartile ranges across P-wave, QRS complex, and T-wave detection. Importantly, this performance is maintained even when more noise is introduced, highlighting the robustness of its signal-cleaning capabilities. Furthermore, MaxYield™ supports the analysis of extended ECG recordings of up to 14.5 days, substantially increasing the likelihood of detecting intermittent arrhythmias that would be missed in shorter studies.

These findings demonstrate that MaxYield™ provides accurate and consistent beat-level labeling over long-duration ECG recordings, supporting its use as a reliable foundation for detecting clinically important cardiac events.

2. Introduction

Cardiovascular diseases (CVDs) continue to represent a significant global health burden, remaining a leading cause of morbidity and mortality worldwide. While standard 12-lead electrocardiograms (ECGs) provide a rapid "snapshot" of the heart's electrical performance, their utility is constrained by their brief recording duration. Many arrhythmias are paroxysmal or intermittent, appearing unpredictably and often absent during short clinical assessments. This limitation frequently leads to delayed or missed diagnoses of transient but clinically significant events, underscoring the urgent need for more comprehensive monitoring solutions.

Holter monitors, which enable continuous multi-day ECG recordings, were designed to address this limitation but introduce new complexities due to the sheer volume of data generated. Holter studies present technicians with millions of heartbeats, where pervasive noise and artifacts hinder reliable labeling and make the detection of clinically significant abnormalities especially challenging. Current Holter software attempts to identify P, QRS, and T waves and then flag potential abnormalities using pattern recognition, but performance varies widely between manufacturers and is often undermined by signal quality. The result is frequent false positives, missed events, and inconsistent labeling. This forces technicians into labor-intensive manual correction, a process that is inefficient, prone to "alarm fatigue," and introduces variability into final reports.

To address these limitations, we developed a solution focused on two foundations of ECG interpretation: robust noise reduction and accurate, consistent detection of P-waves, QRS complexes, and T-waves. These features underpin reliable arrhythmia detection and clinical decision-making.

NeuralCloud Solutions introduces MaxYield™, an advanced platform that analyzes raw ECG signals to precisely detect PQRST features and extract interval metrics. The core strength of MaxYield™ lies in its ability to combine a cleaned ECG signal with accurate beat-level outputs. This precision enables rapid post-processing to identify cardiac events with reliability, streamlining diagnostic workflows.

The results, detailed in the following sections, highlight MaxYield™'s performance in accurate beat-level labeling, supporting its potential to reduce diagnostic variability, improve efficiency, and ultimately enhance the detection of critical cardiac abnormalities.

3. Background

A validation study was conducted to assess the variance between fifteen Registered Cardiology Technologists and the AI model in MaxYield™. Each of the fifteen technologists independently performed manual annotations of ECG waveforms, focusing on the precise identification of P-wave, QRS complex, and T-wave onsets and offsets. By involving a cohort of fifteen experts rather than relying on a single reference annotator, we were able to capture the natural variability that exists in human interpretation. This is critical because conventional approaches to evaluating AI models typically compare predictions against a single human label, an inherently flawed method that overlooks inter-technologist variance. Using multiple annotators not only provided a more rigorous benchmark but also allowed us to measure where MaxYield™ aligns within the spectrum of expert performance.

4. Methodology

4.1 Selection of ECG Samples

Representative ECG data were selected from publicly available datasets made accessible through Zenodo and PhysioNet. The test dataset consisted of 400 ECG files, specifically curated to represent real-world Holter-style and standard 12-lead ECG conditions. Synthetic or simulated ECG data were used. Importantly, all test files were withheld from model training and development to maintain the independence of the model validation.

The Holter data included sinus rhythm, atrial fibrillation, atrial flutter, premature atrial contractions (PACs), and premature ventricular contractions (PVCs). From these categories, files were randomly selected in equal proportions.

As for the 12-lead data, random files and leads were selected. This was done to include ECG tracings with a higher quality signal and different polarity waves.

This test set was used to analyze the analysts' and the model's ability to label waveform annotations under real-world signal conditions. The data was selected to reflect a broad range of physiologic and arrhythmic conditions.

4.2 Analysts

Registered Cardiology Technologists were selected to represent a diverse range of professional experience (1–31 years; mean: ~10 years), academic training (including diplomas, bachelor's, master's, and doctoral degrees), and certification credentials (primarily through the Canadian Society of Cardiology Technologists and the Ontario Society of Cardiology Technologists). Several participants possessed additional specialization in domains such as electrophysiology, IV therapy, and cardiac rhythm device programming. This heterogeneity was intentional, reflecting the variation commonly observed in clinical environments. By incorporating a spectrum of interpretive approaches informed by varying levels of training and practical experience, the labeling process simulates real-world variability in ECG interpretation. The sample size was chosen to balance annotation diversity with operational feasibility and to approximate the staffing scale of a mid-sized clinical ECG department. This approach enhances the ecological validity of the resulting dataset and supports the development and evaluation of robust, generalizable signal processing algorithms.

4.3 Labeling Process

An in-house labeling tool was created for labeling the ECG files. Each analyst was given all 400 files and presented with 10-second one-lead ECG tracings. These were individual 10-second strips, one strip for each unique ECG file in the dataset. They were tasked to label all the P-waves, QRS complexes, and T-waves. This included the onset and offset of each wave, as well as the type of wave, specifically P, QRS, or T.

The labeling process involved analysts selecting a wave type and then dragging their mouse over the ECG tracing to highlight a wave. If there were errors, they could delete the labeled wave and try again, or they could manually edit the onset and offset using some buttons. Once the analysts completed labeling the ECG, they could submit the labels to a central database.

4.4 Statistical Analysis and AI Model Evaluation

The labels were downloaded from the central database, and then each labeled wave was grouped. Waves were considered to be the same wave and grouped if labels were for the same type of wave, and if they had any overlap with at least one of the other analysts' labels. So if label A

overlaps with label B, and label B overlaps with label C, then label A and label C are considered to be labeling the same wave even if label A and label C do not overlap. Each group of labels represented one wave. If fewer than three analysts labelled a specific wave, then that group was removed and considered mislabelled.

The mean of each group's onset and offsets was calculated using the analysts' labels. The mean was used as the 'true' value when measuring the error of the analysts and the model. The root mean square error (RMSE) was calculated for each analyst and the model.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \hat{x}_i)^2}{N}}$$

RMSE was chosen over other methods because it penalizes outliers. In practice, labeling the onsets and offsets perfectly is not essential; however, it is crucial not to misslabel a wave.

The mean of the RMSE was calculated for each analyst's wave onsets and offsets. This average error was used to rank the analysts and the model. This method provides a slight advantage to the analysts, as their labels were used to calculate the 'true' value.

The median error and interquartile range (IQR) were calculated for each analyst and the model. This was done to better understand the accuracy of each analyst in labeling and to compare it to the model.

Due to low-quality signals and noise, not every analyst agreed on whether a P-wave existed. An analysis was performed to estimate the accuracy with which an analyst and the model can identify P-waves. Since not all analysts agreed, the probabilities of each P-wave existing were first calculated and later used to weight the accuracy of the analysts.

The probability that a wave is a P-wave was calculated by dividing the number of analysts who claimed the wave is a P-wave by the total number of analysts who labeled the file. For example, if two analysts labeled a P-wave out of a total of ten analysts who labeled the file, then the probability of that P-wave existing would be 0.2. The probabilities were used to weight the percentage of time the analyst labeled a P-wave versus the total number of P-waves. This ensured more weight was placed on P-waves where the majority of analysts agreed.

The model P-wave accuracy was evaluated in the same manner as the analysts. The only

difference is that the model's P-wave predictions were not used to calculate any probabilities for a P-wave to exist. This gives the analysts a slight advantage.

The model was also tested with added noise. For each file, we randomly added noise, which included white noise, baseline wander, powerline, and artifacts. The model then labeled the noisy file. We used the results in the statistical analysis. Since the noise was random, this method was run five times, and the average was taken. This is an unfair advantage for the human labelers, as they were only given the clean version of the signal and did not experience the same level of noise. Therefore, we expect the model with added noise to struggle more.

5. Results

There were 400 unique files, comprising a total of 4,517 QRS complexes. Each analyst labelled all 400 files.

The root mean square error (RMSE) shows the error of each analyst in labelling the waves. This is important as a mislabelled wave could lead to misdiagnosing crucial cardiac events. Table 1 shows the result for each analyst, indicating that MaxYield™ performed similarly to the top-2 ranked analysts by having the 2nd lowest Mean RMSE. This shows the high level of accuracy that MaxYield™ has in removing noise artifacts and correctly labelling waves and sections of the ECG, which is essential for diagnosing cardiac events.

Table 1: RMSE Rank Table

Analyst	Mean RMSE	Mean RMSE Rank
Expert 1	11.852382	1
MaxYield™	**13.469676**	**2**
Expert 2	13.599225	3
Expert 3	14.038519	4
Expert 4	14.425222	5
Expert 5	15.612630	6
Expert 6	15.734446	7
Expert 7	15.798977	8
Expert 8	16.165228	9
Expert 9	16.226465	10
Expert 10	16.970178	11
MaxYield™ (With Noise)	**17.445167**	**12**
Expert 11	20.039835	13
Expert 12	20.385723	14
Expert 13	23.468058	15
Expert 14	27.027354	16
Expert 15	30.422535	17

The Median error and IQR were calculated to better understand the accuracy of each analyst in labeling the P-wave, QRS complex, and T-wave compared to MaxYield™; the results are detailed in Table 2, Table 3, and Table 4.

Table 2 shows that MaxYield™ has a very low Median error and IQR compared to the other experts. MaxYield™'s lower IQR indicates its ability to label P-wave onsets and offsets consistently. Only two experts had a lower IQR, indicating that MaxYield™ has high accuracy in

detecting P-wave onsets and offsets. It should be noted that though MaxYield™ (with noise) had a higher IQR, it was still lower than some of the experts. It also had a lower Median error, which further shows the accuracy of MaxYield™ in detecting P-waves even with noise treatment not active.

Table 3 shows that MaxYield has a low Median error, outperforming 11 experts, and a low IQR, outperforming all experts. This indicates that MaxYield™ has a high accuracy and consistency in detecting QRS complexes, which is essential for detecting heartbeats and intervals to help diagnose arrhythmias if present. MaxYield™'s (with noise) algorithm has a low IQR and outperforms all experts, second to MaxYield™, further showing the accuracy of MaxYield™ in identifying onsets and offsets of QRS complexes.

Table 4 shows that MaxYield™ has a very low Median error, outperforming all but one expert, and a low IQR outperforming all other experts. This shows that MaxYield™ has a high accuracy and consistency in detecting T-waves, which helps with cardiac event diagnoses. MaxYield™'s (with noise) algorithm has a low IQR that outperforms all but one expert, further showing the accuracy of MaxYield™ in identifying onsets and offsets of T-waves.

Table 2: Median Error and IQR Table for P-Wave

Analyst	P-Wave Onset		P-Wave Offset	
	Median Error	IQR	Median Error	IQR
Expert 1	-3.2	12.4	**0.3**	11.5
Expert 2	5.6	12.3	-3.2	12.9
Expert 3	6.6	12.8	-8.5	12.3
Expert 4	4.0	**9.6**	-4.5	11.3
Expert 5	10.4	14.0	-6.0	14.7
Expert 6	-4.8	14.6	-3.1	17.1
Expert 7	-4.3	15.5	6.9	15.2
Expert 8	-6.7	12.8	2.1	12.5
Expert 9	-0.3	11.7	-3.5	11.7
Expert 10	8.0	14.3	-2.0	19.6
Expert 11	9.1	11.0	-11.5	15.7
Expert 12	**0.0**	17.1	10.0	20.4
Expert 13	-4.7	12.3	1.9	11.2
Expert 14	-8.0	20.6	12.6	16.8
Expert 15	-5.9	16.3	-5.3	21.8
MaxYield™	-1.1	10.7	-1.7	**10.9**
MaxYield™ (With Noise)	-0.5	14.3	-2.4	13.0

Table 3: Median Error and IQR Table for QRS Complex

Analyst	QRS Complex Onset		QRS Complex Offset	
	Median Error	IQR	Median Error	IQR
Expert 1	-4.1	8.8	5.3	10.1
Expert 2	3.1	9.6	-5.0	10.1
Expert 3	2.5	9.1	-4.3	9.6
Expert 4	4.0	7.6	**-0.8**	9.1
Expert 5	4.9	9.3	-2.9	12.0
Expert 6	0.9	11.6	-2.9	12.5
Expert 7	-2.4	11.7	5.3	13.6
Expert 8	**0.3**	8.5	-5.3	8.4
Expert 9	2.9	7.4	-8.0	8.0
Expert 10	0.6	14.1	4.0	19.9
Expert 11	6.4	8.0	-10.9	8.9
Expert 12	5.6	13.9	-4.8	15.7
Expert 13	-0.8	7.9	**0.8**	8.3
Expert 14	-17.9	17.6	28.0	18.9
Expert 15	-3.7	12.3	-5.3	12.4
MaxYield™	-2.0	**6.8**	-5.3	**7.9**
MaxYield™ (With Noise)	**-1.7**	7.5	-5.6	8.7

Table 4: Median Error and IQR Table for T-Wave

Analyst	T-Wave Onset		T-Wave Offset	
	Median Error	IQR	Median Error	IQR
Expert 1	3.2	17.7	0.3	13.1
Expert 2	11.2	19.5	-7.5	15.4
Expert 3	-13.0	19.9	-0.5	15.2
Expert 4	**1.3**	21.3	**0.0**	14.8
Expert 5	15.1	18.1	-8.0	15.4
Expert 6	8.0	23.2	1.1	17.1
Expert 7	10.8	20.6	-2.3	18.7
Expert 8	-18.7	25.6	0.0	15.5
Expert 9	10.9	15.1	-5.7	12.3
Expert 10	8.3	22.9	0.2	22.9
Expert 11	21.6	17.6	-10.0	14.2
Expert 12	17.1	20.1	-6.1	18.7
Expert 13	-40.3	29.5	17.4	19.5
Expert 14	-5.3	28.3	17.9	25.6
Expert 15	-7.2	34.6	-7.2	20.3
MaxYield™	-1.6	**12.4**	-5.1	**10.5**
MaxYield™ (With Noise)	-0.5	15.6	-6.5	12.9

The analysis to estimate the accuracy with which the experts and MaxYield™ could identify P-waves in ECG files was calculated, and the results are shown in Table 6. MaxYield™ had a lower agreement percentage than most of the experts, specifically Expert 2, who was ranked 3rd in Table 1 and had low error scores depicted in Tables 2 to 5. This shows that MaxYield™'s noise cleaning algorithm accurately cleans data and labels true P-wave onsets and offsets when

they occur.

Table 6: P-waves labelled agreement

Analyst	Agreement
Expert 12	92.7%
Expert 5	91.9%
MaxYield™ with noise	91.4%
Expert 14	90.9%
Expert 13	88.6%
Expert 8	88.5%
Expert 9	86.7%
Expert 7	86.4%
Expert 1	85.2%
Expert 4	85.0%
MaxYield™	**84.9%**
Expert 15	83.1%
Expert 2	82.0%
Expert 11	78.1%
Expert 3	77.9%
Expert 10	76.2%
Expert 6	70.7%

6. Conclusion

The results demonstrate that MaxYield™ consistently achieved low Median and IQR errors across all tests compared to human experts, underscoring its high accuracy and consistency in identifying the onsets and offsets of P-waves, QRS complexes, and T-waves. These features are fundamental to ECG analysis and essential for interpreting cardiac events and diagnosing arrhythmias and other cardiovascular diseases.

This level of accuracy, combined with MaxYield™'s ability to process extended ECG recordings (up to 14.5 days), enables the reliable detection of numerous cardiac events, particularly paroxysmal or intermittent arrhythmias that are often missed in short-duration assessments. While this study focused on the precision of P-wave, QRS complex, and T-wave annotation, MaxYield™'s ability to correctly identify these complexes after noise and artifact removal is expected to substantially increase the probability of capturing critical abnormalities (e.g., prolonged pauses, various types of heart block, short atrial fibrillation episodes, sustained SVT, or ventricular tachycardia). This precision also reduces the false positives that contribute to technician "alarm fatigue."

Accurate detection of P-waves and QRS complex onsets and offsets further supports the diagnosis of not only pauses but also their underlying nature (e.g., second-degree AV block, third-degree AV block, sinus pauses). Although T-wave onset and offset are not always central to routine Holter reporting, their accurate identification by MaxYield™ can provide important clinical insights in specific cases requiring comprehensive cardiac assessment. Future studies focused on direct arrhythmia event detection will further validate these benefits.

In addition, MaxYield™ demonstrated superior robustness in the presence of noise, significantly outperforming traditional algorithms. Its noise-canceling capability effectively removes artifacts while preserving critical ECG morphology, ensuring data integrity for downstream analysis.

These capabilities position MaxYield™ as a next-generation platform for precise ECG analysis, delivering clean signals and accurately detected PQRST features over long-duration recordings. This foundation facilitates faster post-processing, streamlines technician workflows, and supports timely detection of a wide spectrum of cardiac events.

7. References

Zenodo. (2021). *CODE-15%: a large scale annotated dataset of 12-lead ECGs* (Version 1.0.0) [Database].

https://zenodo.org/records/4916206

PhysioNet. (2022). *Icentia11k Single Lead Continuous Raw Electrocardiogram Dataset* (Version 1.0) [Database].

https://physionet.org/content/icentia11k-continuous-ecg/1.0/

Carl Böck, Christoph Mörtl, Christoph Mahringer, Mario Huemer, Jens Meier. (2023). Variability of expert assessments of ECG time domain parameters. European Journal of Anaesthesiology and Intensive Care. 10.1097/EA9.0000000000000020

https://journals.lww.com/ejaintensivecare/fulltext/2023/04000/variability_of_expert_assessments_of_ecg_time.2.aspx

World Health Organization. (2025, August 12), *Cardiovascular diseases (CVDs).*

https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

National Library of Medicine. (2025, August 12), *Arrhythmias*.
https://www.ncbi.nlm.nih.gov/books/NBK558923/

David A. Cook, So-Young Oh, Martin V. Pusic. (2020). Accuracy of Physicians' Electrocardiogram Interpretations: A Systematic Review and Meta-analysis. JAMA Internal Medicine. 10.1001/jamainternmed.2020.3989

https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/2771093#google_vignette