CardioPet ECG Algorithm validation study

By Allison Spake, DVM, DACVIM (Cardiology); Donald Szlosek, MPH; D.J. McCrann, PhD; and Sarah Shannon MBA

Abstract

The CardioPet® ECG Algorithm is a decision support technology that aides in the clinical interpretation of ECGs by cardiologists. The CardioPet ECG Algorithm technology is designed to accurately measure the waveforms of an ECG complex and detect potential ECG abnormalities.

A study was conducted to validate the performance of the CardioPet ECG Algorithm. The goal of the study was to compare the performance of the CardioPet ECG Algorithm to a panel of board-certified cardiologists and test for statistical equivalence.

When evaluating for arrhythmias alone, the algorithm identified arrhythmias in all but one case with a resulting sensitivity of 99.7% and a specificity of 100% compared to a panel of board-certified cardiologists. The CardioPet ECG Algorithm was found to be statistically equivalent to the panel of cardiologists for the detection arrhythmias on an ECG.

Introduction

ECGs in veterinary medicine are performed for a variety of indications: as a preoperative screening tool, to diagnose arrhythmias noted on auscultation, as part of a full cardiac workup for patients with known or suspected heart disease, as part of screening for arrhythmias during chemotherapy, or as a baseline for breeds at risk of developing clinically important arrhythmias (e.g., boxers and Doberman pinschers).

The CardioPet ECG Algorithm is now used in support of the IDEXX CardioPet ECG screening service. The algorithm grades each ECG as normal or flags certain abnormalities, such as abnormal ECG measurements or potential arrhythmias. By utilizing these flags as part of the decision support technology, IDEXX releases some ECG reports to the customer and elevates potentially abnormal ECGs to cardiologists for additional review.

An ECG can be graded "abnormal" due to the presence of an arrhythmia or based upon wave measurements detected outside of established intervals. However, these abnormalities do not carry equivalent medical relevance. Arrhythmias identify functional abnormalities and can represent potentially serious underlying cardiac disease; whereas ECG measurements alone are not considered reliable to detect underlying cardiac disease. For example, a study evaluating the P wave width to detect underlying left atrial enlargement found this measurement to be unreliable in the detection of underlying left atrial enlargement.[1] Therefore, the sensitivity for detection of abnormal measurements correctly identifying functional abnormalities is considered not as medically relevant when evaluating the performance of the ECG algorithm.

We evaluated the sensitivity of the algorithm in two different ways: 1) sensitivity for all abnormalities, including sensitivity for abnormal measurements, and 2) sensitivity for detection of arrhythmias alone. This algorithm was designed to be highly sensitive in the detection of arrhythmias and highly specific in the measurement of ECG parameters.

Validating the algorithm: A randomized and masked study

An expert panel result, defined as agreement between at least 2 of 3 reviewing cardiologists, was established to validate ECG findings. A group of seven cardiologists (C1–C7) were masked to the algorithm results and were asked to review 399 ECGs for the following parameters: mean electrical axis (MEA), R wave height, P wave height, P wave width, QRS duration, heart rate, PR interval, arrhythmia identification and final overall result (normal/abnormal).

Three cardiologists were randomly assigned to read each ECG. The results and measurements produced by the cardiologists and algorithm were then compared to one another. If there was disagreement between the consensus result of the cardiologists and the algorithm, then the findings of the cardiologists were compared to one another to define baseline agreement.

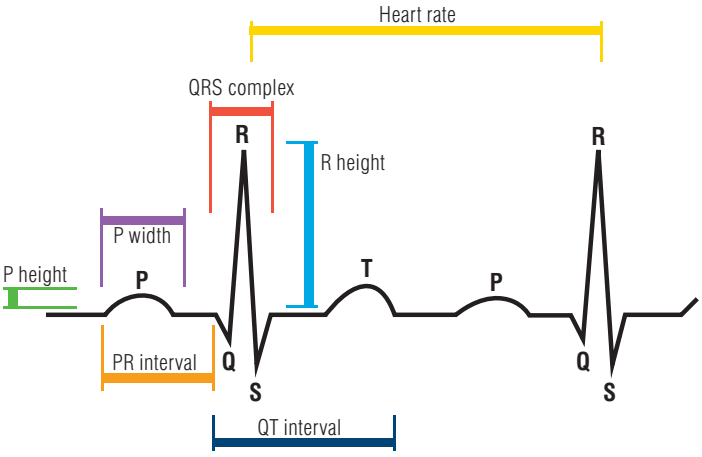

Figure 1. ECG deconstructed.

Once the cardiologists recorded their results, these ECGs were analyzed by the ECG algorithm. The algorithm recorded the following parameters: heart rate, P wave height, P wave width, PR interval, R wave height, QRS duration, MEA, and flag any potential arrhythmias. A final pass or fail value and a reason for the fail was also assigned by the algorithm.

After the initial review by the masked panel, an unmasked panel of 3 cardiologists reviewed all the false-positive and false-negative results to determine truth. From these evaluations, we arrived at our final sensitivity and specificity with regards to ECGs that were abnormal for arrhythmia only as well as a sensitivity and specificity for all potential ECG abnormalities.

Results

The sensitivity of the algorithm for detecting arrhythmias was 99.7% (95% CI: 98.5–100.0) and the specificity was 100.0% (95% CI: 88.8–100.0) compared to the panel of cardiologists. The sensitivity of the algorithm for picking up any anomaly, including abnormal measurements, was 82.7% (95% CI: 76.6–87.7) and specificity was 93.1% (95% CI: 89.0–96.02) compared to the panel of cardiologists (table 1).

ECG abnormality type	TP	TN	FN	FP	SUM	Sensitivity (95% CI)	Specificity (95% CI)
All ECG abnormalities	189	162	14	34	399	82.7 (95% CI: 76.6–87.7)	93.1 (95% CI: 89.0–96.2)
Arrhythmia only	31	367	1	0	399	99.7 (95% CI: 98.5–100.0)	100.0 (95% CI: 88.8–100.0)

Abbreviations key
TP: True Positive; TN: True Negative; FP: False Positive; FN: False Negative

Table 1. Sensitivity and specificity of the panel of cardiologists to the algorithm for the detection of arrhythmias and any anomaly

A grand kappa (κ) test was conducted to measure the overall agreement between the cardiologists and the algorithm for cases where 2 or more cardiologists agreed with each other and with the algorithm.[2] The detection of arrhythmias between the panel of cardiologists and the algorithm was observed to have almost perfect agreement 0.98 (95% CI: 0.89–1.00). Similarly, the panel of cardiologists' result compared to the algorithm showed substantial agreement for the detection of all anomalies: $\kappa = 0.76$ (95% CI: 0.70–0.82).

Almost perfect agreement was found across all 7 ECG parameters measured when the 7 cardiologists were compared against each other with an intraclass correlation coefficient (ICC) range of 0.96–0.99. High level of agreement was also found when comparing the algorithm to the cardiologists (table 2).

Following agreement testing a panel of 3 cardiologists, unmasked to the results of the algorithm and cardiologist panel, were presented with all discordant results. This group reviewed the discordant false-positive and false-negative results, and the algorithm was observed to be statistically equivalent (Wald-*P* value: > 0.995).

When comparing the panel of cardiologists to the algorithm for the detection of any ECG anomaly, the algorithm was not found to be statistically equivalent (Wald-*P* value: < 0.001). On all 7 ECG parameters and the final ECG result, statistical equivalence between the performance of the expert panel, cardiologists, and algorithm was found (all values, Yuen-TOST-*P* value: > 0.995).

ECG parameter	ICC (between cardiologist)	ICC (algorithm vs. cardiologist)	Interpretation (agreement)
MEA	0.96 (0.87–0.98)	0.96 (0.85–0.98)	Almost perfect
QRS duration	0.97 (0.93–0.98)	0.98 (0.95–0.99)	Almost perfect
PR interval	0.99 (0.95–1.00)	0.98 (0.98–0.99)	Almost perfect
P wave width	0.96 (0.95–0.96)	0.82 (0.79–0.84)	Almost perfect
Heart rate	0.99 (0.99–1.00)	0.97 (0.93–0.99)	Almost perfect
P wave height	0.98 (0.96–1.00)	0.97 (0.49–0.73)	Almost perfect
R wave height	0.99 (0.97–1.00)	0.95 (0.92 –0.97)	Almost perfect

Table 2. Agreement of ECG parameters measured by 7 cardiologists and the CardioPet® ECG Algorithm. Kappa and ICC agreement was interpreted from Landis and Koch[2] as the following: poor (< 0), slight (0.01-0.20), fair (0.21–0.41), moderate (0.41–0.60), substantial (0.61–0.80), and almost perfect (0.81–1.00).

Conclusion

The objective of this study was to compare the effectiveness of a computer-aided algorithm in evaluating an ECG to the performance of a team of board-certified cardiologists. The results indicate that the computer-aided algorithm performs in a manner statistically equivalent to the cardiologists for measuring ECGs.

One of the most important uses for the CardioPet® ECG Algorithm is the measurement and detection of arrhythmias. The algorithm's sensitivity for the detection of arrhythmias was 99.7%, and the specificity was 100%.

The CardioPet® ECG Algorithm has been validated to perform equivalently to a panel of 7 cardiologists in the measurement of 7 ECG parameters as well as to the combined detection of arrhythmias and abnormal waveforms by this same expert panel. The incorporation of the CardioPet® ECG Algorithm into medical practice will increase the efficiency and reliability of IDEXX cardiologists by automatically reviewing and flagging arrhythmias and normal ECGs with a high degree of sensitivity and specificity.

References

1. Savarino P, Borgarelli M, Tarducci A, Crosara S, Bello NM, Margiocco ML. Diagnostic performance of P wave duration in the identification of left atrial enlargement in dogs. *J Small Anim Pract*. 2012;53(5):267–272.
2. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159–174.