Investigation of Dry Electrode ECG Signal Characteristics Towards Explainable AI

Abdelrahman Abdou[1], Niraj Mistry[2], Sridhar Krishnan[1]

[1]Toronto Metropolitan University, Toronto, Canada
[2]Scarborough Health Network, Toronto, Canada

Abstract

Single-lead electrocardiograms (ECG) are a reliable modality for long-term remote monitoring applications. The development of wearable ECG devices requires understanding of the inherent characteristics of the acquired signals to create ECG centric signal processing and machine learning (ML) algorithms. However, most single-lead ECG wearable signal analysis blindly follows algorithms that were built for clinical ECG. This approach yields inappropriate choice of filtering and ML models that create inaccuracies and increase unreliability. A framework to examine signal characteristics to guide development of explainable artificial intelligence (xAI) for newborn ECG monitoring applications is proposed. 3D printed dry electrode newborn ECG signals (n = 19) are examined under different testing paradigms. The signals undergo linearity, stationarity, and Gaussian testing. Also, the signals undergo cubic Hermite spline interpolation to show appropriateness of representing signals through spline polynomial functions. Across all newborn ECG signals non-linear, non-stationary, and hyper Gaussian behaviors are observed. By understanding the signal characteristics, adaptive algorithmic development is realizable especially for remote ECG monitoring applications.

1. Introduction

Single-lead ECGs are becoming increasingly prevalent in remote vital sign monitoring applications. These applications can include post-operative continuous observation of cardiac activity and telehealth monitoring for physician check-ups. This in turn requires the development of algorithmic approaches to appropriately perform application-specific ECG signal processing and analysis. The variability of ECG sensor used in signal acquisition including the use of wet, dry and/or fabric-based electrodes requires the use of specialized pre-processing and xAI algorithms which in turn utilizes the signal's characteristics to accurately analyze and perform correct diagnosis from the raw information being presented [1]. Furthermore, the acquired ECG can differ from patient to patient depending on their anatomy and age. For example, ECG for newborns contain shorter PQRST interval morphology when compared to adult ECGs [2]. This example shows how normal neonatal ECG if found in an adult is abnormal and vice versa. In turn certain algorithmic approaches and processing steps should be taken to account for the variable changes in subjects and their age. This process shows that not all algorithms are suitable for all ECGs. Certain assumptions are taken to ensure that the signal processing framework in question is suitable for the intended application. These assumptions include that ECGs are non-linear, nonstationary, and Gaussian probability density function in nature. However, these assumptions do not consider the acquisition sensor at hand, signal length, nor the presence of motion artifacts that may impact the quality and statistical nature of the ECG. This work is a systematic step in developing a framework to validate ECG acquisition assumptions and use an informed statistical approach to create the signal pre-processing algorithms used for dry electrode based neonatal ECGs.

Schwartz et Al. provided the guidelines for interpreting neonatal ECG, their work discusses the normal and abnormal behaviors of but not limited to the sinus rhythm, P-wave morphology, atrioventricular and intraventricular conduction mechanism, QRS and ST segment morphologies [3]. Their work offers physicians strategies to identify the abnormalities in newborn ECG which is different than adult ECG [3]. Their work is suitable clinically but do not consider the acquisition device in their analysis which in turn does not offer insights about the statistical nature of the newborn ECG signals.

While Hussain et Al. examined ECG features to detect asphyxia in neonates [4]. Their work examined different morphological shapes and sizes of the PQRST waveform including T-wave amplitude, QT interval, and T-wave slope. The authors found that asphyxiated neonates showed significant changes in ECG features when compared to non-asphyxiated groups [4]. Their work relied on established signal processing techniques to segment the ECG, and Pan-Tompkins algorithm to identify QRS waveforms. Their work used two out of the 12-lead standard ECG leads without identifying the type of recording device [4]. However, their results are

 ISSN: 2325-887X DOI: 10.22489/CinC.2024.151

promising and show that the current standard research accepted signal processing techniques are suitable for ECG feature extraction with wet electrode-based ECG acquisition in neonates.

It is evident there is a gap in the pre-analysis of ECG and its statistical features to guide signal processing and xAI development. This work is the first of its kind to propose a framework to guide dry electrode ECG feature extraction through an explainable statistics informed analysis. Also, signal characteristics are explored to direct development of future signal processing techniques and ML models for neonatal 3D printed dry electrode ECGs. While the signal reconstruction properties of dry electrode ECGs are explored to determine the appropriate polynomial order and function required to recreate the ECG signal effectively.

The methods section will discuss the dataset collected and tools used in the analysis. The results section will follow with signal characteristics and analysis. The discussion will deliberate on the outcomes of this study and provide interpretation towards the use of signal properties in the development of feature extraction mechanisms and areas of improvement for the examined framework. The conclusion section will provide a summary of this study.

2. Methods

2.1. Dataset

The dataset used to perform the investigation of signal characteristics under linearity, Gaussian, stationarity and modelling analysis contains 19 newborn ECG signals with 3-minute durations. The signals are collected using a novel single-lead ECG-based, Lead-I, heart monitor utilizing 3D printed dry electrodes and 500 Hz sampling rate. A hardware-based band-pass filter is present providing 3-48 signal filtering to eliminate electromagnetic interference (EMI) noise. This dataset collection is approved by Toronto Metropolitan University, REB 2023-070 and is part of a larger study of collecting ECG and HR information for improving neonatal resuscitation capabilities at point of care [5].

2.2. Statistical Analysis

This work examines dry electrode based neonatal ECG and its signal characteristics to guide future development of filtering techniques and xAI approaches. This informed approach would provide valuable information about the inherent attributes of the signals in question yielding a better design of signal processing approaches. In turn the signals underwent linearity, Gaussian, stationarity and modelling tests to inform the signal processing developer of the features that can be used in their algorithmic development. Linearity is defined as the capability of separating the desired signal into components that can be superimposed to recreate the signal. Superposition and homogeneity properties must be fulfilled to consider a signal linear. However, non-linear signals do not follow superposition and homogeneity traits. It is assumed that all physiological signals are inherently non-linear in nature. A tree portioned autoregressive non-linear model with model order one ($Isnlarx$), as shown in Eq. 1 is used to determine the linearity characteristics of the signals provided.

$$X(t) = L(t) + Fn(t) + E(t), \qquad (1)$$

where $X(t)$ is the data examined under the non-linear model, $L(t)$ is part of data explained by linear function, $Fn(t)$ is part explained by non-linear model and $E(t)$ is the remaining error.

While stationarity examines the statistical properties of a time series signal such as ECG. A signal is considered stationary when its statistical attributes do not change with time. Physiological signals are assumed to be nonstationary in nature. Kwiatkowski–Phillips–Schmidt–Shin (KPSS) and Augmented Dickey–Fuller (ADF) tests are used to determine stationarity depending on the type of stationary examined [6]. KPSS examines the trend stationary attribute which is defined as drift of the signal. While ADF is used to determine the stationarity depending on the presence of a unit root, α, as shown in Eq. 2.

$$Y(t) = \alpha Y(t-1) + \beta X_e + \epsilon, \qquad (2)$$

where, $Y(t)$ is value of amplitude at t and α is unit root. If α = 1, the time series signal is said to be non-stationary.

On the other hand, Gaussian characteristics of the signals can be performed by determining the skewness of the computed signal histograms and the Anderson-Darling (*AD*) test. The AD test is used to determine on whether a signal sample is present in a specified distribution. More specifically, the test is used to assess the sample's presence in a Gaussian distribution. The AD test and plotting signal histograms can give us an outline of the presence of different periodicities, which assists in identifying outliers that are representative of EMI noise and the period of the ECG sinus rhythm. While the use of cubic Hermite interpolation can assist us in recreating the desired ECG signals using N^{th} order polynomial functions. The signals undergo iterative cubic Hermite splicing and reconstruction to determine the minimum order of polynomial functions required to reconstruct the signals effectively. This approach will allow us to determine the right order number which can be used in the development of signal compression algorithms suitable for long-term embedded ECG monitoring applications. This framework of statistical analysis can act as a systematic framework that can be followed to inform the development of signal processing more specifically for 3D printed dry electrode neonatal ECG.

3. Results

All 19 subjects' dry electrode ECG underwent the statistical analysis. The signals showed non-linear behavior as evident by rejecting the null hypothesis (*Linearity* = 1), *p-value* < 0.05, as shown in Table 1. On the other hand, *NLValue* is identified as the standard deviation of the non-linear information, $Fn(t)$. Noise, σ, is considered the estimated SD of the unexplained errors, $E(t)$. The ECGs also showed low averaged *NLValue* of $8.6\text{x}10^{-3}$ and Noise, σ, estimations of $6.4\text{x}10^{-3}$. This information entails that low *NLValue* and σ represent low variability in the non-linear information present in the signals with low-noise characteristics, as shown in Table 1. However, *RatioDetect* (*RD*) provides the ratio of the *Isnlarx* test statistic and if it is above 2, the signals under test are considered non-linear, if below 0.5, they can be fitted using a linear model, and if the value is close to 1, the signals cannot be identified using the non-linear model in question requiring further investigation into another non-linear autoregressive model. All ECGs showed high *RD* values, above 2, with an average of 40.7 supporting the rejection of the linearity null hypothesis showing that all ECG signals show non-linear behaviors.

Table 1: Averaged Neonatal ECG signals' statistics

	Average of all 19 ECGs
Linearity	1
NLValue	$8.6\text{x}10^{-3}$
Noise σ	$6.4\text{x}10^{-3}$
Ratio Detect	40.7
KPSS	0.68
ADF	1
Polynomial Order	4

On the other hand, stationarity tests of KPSS and ADF showed different results. KPSS showed an average value below 1 which entails that some ECG signals failed to reject the null hypothesis (*KPSS* = 0), whereas ADF testing showed all signals rejecting the null hypothesis (*ADF* = 1), with *p-value* <0.05. This information entails that certain ECGs yielded an evident stationarity behavior under KPSS test conditions but not ADF. These ECGs are found in subjects; 29, 43, 57, 75, 58, 85, and 97. While the minimum polynomial order, 4, is required to reconstruct the ECGs efficiently as examined under the cubic Hermite polynomial test. This average value, as shown in Table 1, shows that all ECGs can be reconstructed using polynomial functions with a minimum order of 4 to rebuild the ECG. It is important to note that further investigation into polynomial functions, its orders and similarity metrics should be further explored to determine the appropriate function and order for single-lead ECG applications.

ECG signal histograms provide valuable information on the probabilistic distribution of the information. All 19 dry electrode ECGs showed similar slight left-tailed histograms as shown in Figure 1. The ECGs are all considered slightly skewed which is evident of the hyper gaussian behavior of ECGs in nature. This data can be used in identifying noise presence, and examining the different periodicities available in the signal.

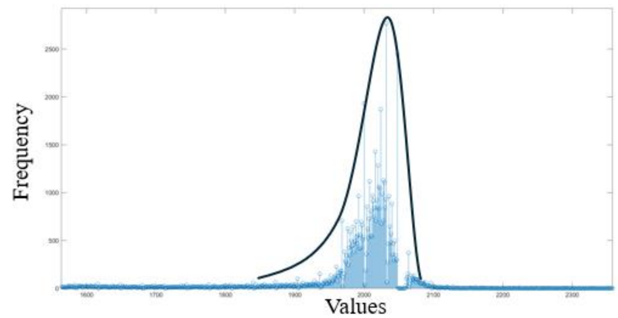

Figure 1: Histogram of Sub_97 showing left-tailed skew

4. Discussion

Dry electrode neonatal ECG signals possess non-linear, non-stationary, and Gaussian properties. The non-linearity tests included using an autoregressive non-linear model to determine the non-linear behavior of the signals. All signals showed non-linearity with high RD values indicating that strong non-linear behavior suitable for non-linear features for ML model or the use of non-linear ML such as artificial neural networks. Non-linearity is further confirmed by analyzing the input output relationship relative to the presence of linear data, represented by $\boldsymbol{L}(\boldsymbol{t})$, which showed insignificant values when compared to *RD* and $\boldsymbol{Fn}(\boldsymbol{t})$. Generally, ECG signals are considered non-linear which this study supports this conclusion with an informed statistical analysis thereby increasing the reliability of developers in creating their algorithms to include non-linear properties.

While non-stationary behavior of dry electrode ECG is evident from the ADF and KPSS tests. Most analyzed signals rejected the null hypothesis and showed non-stationarity. However, some signals showed conflicting ADF and KPSS results. It is important to note that both tests examine different aspects of non-stationarity. Certain signals are not able to reject the null hypothesis of KPSS and showed stationary behavior. KPSS is a test used to determine trend stationarity which cannot be relied upon alone in this type of analysis. ADF examines the unit root of stationarity. The signals that showed differences between both tests can be considered weakly nonstationary signals. Further testing paradigms should be explored other than ADF and KPSS to increase this framework's robustness. Some stationarity tests that can be further explored include non-parametric tests such as the Phillips-Peron and Zivot-Andrews tests for stationarity. Stationary behaviors of signals are important to identify because it guides the development of xAI. Stationarity can be used to build classification models, offering optimal results if a stationary behavior can be identified. This concept is highly relevant in stock market signal analysis which showed difficulty in predicting. Although ECG is considered nonstationary, adaptive segmentation techniques should be explored to offer stationarity in the signals that can be later used for prediction model creation for medical xAI applications [6]. Also, signal decomposition techniques such as empirical mode decomposition can be used for developing signal filtering and feature extraction paradigms for nonstationary signals such as ECG.

Also, the signals showed Gaussian distribution properties as evident in Figure 1. The histograms displayed hyper Gaussian characteristics with slight skewness represented as a left-tailed like wet electrode-based ECGs available in the literature. It is important to note that all examined ECGs provided similar histograms. Histogram analysis is valuable in identifying outliers which could be representative of noise. All analyses performed on the above-mentioned signals are computed on the whole signal of 3-minute duration. While fourth order piecewise polynomials can be used to in signal reconstruction of 3D printed dry electrode neonatal ECG. The order is determined using cubic Hermite polynomials which would offer the minimum required order number suitable for reconstruction. However, higher order polynomials and functions should be examined to provide a full picture of the ECGs in question. Overall, all the testing framework presented showed the different important statistical properties of the dry electrode-based neonatal ECGs that can be used in signal processing and ML development. Segmented portions of the ECGs should be examined as next steps. This approach will allow single-lead ECG developers to determine the changes in signal statistics and use that information to guide segmented based feature extraction and analysis.

5. Conclusion

Signal processing and feature extraction of neonatal ECGs could follow a systematic analysis framework for explainability and building trustworthy AI systems. Due to the changes in neonatal ECG compared to adult ECG, certain considerations should be examined to guide future signal analysis. This work utilized SAR-NE dataset [5], a dataset containing 19 newborn ECGs of 3-minute durations and developed a statistical informed framework to analyze the signals. The framework included examining linearity, Gaussian and stationarity of the ECGs. In addition, the polynomial order number is determined for signal reconstruction purposes using cubic Hermite polynomials. The analyzed neonatal ECGs showed non-linearity, nonstationary, and Gaussian behaviors which corresponds with the general nature of ECGs as a biological signal. This informed statistical approach allows us to add confidence when developing algorithms for filtering, prediction, and classification for single-lead dry electrode based neonatal ECGs. This work is a small step in the efforts to realize a holistic guidelines basis for the development of single-lead ECG signal processing and the future goals of building patient-centric technologies to improve care delivery and health outcomes.

References

[1] A. Abdou and S. Krishnan, "Horizons in Single-lead ECG Analysis from Devices to Data," *Front. Signal Process.*, Mar. 2022, doi: 10.3389/frsip.2022.866047.

[2] A. Bratincsák, C. Kimata, B. N. Limm-Chan, K. P. Vincent, M. R. Williams, and J. C. Perry, "Electrocardiogram Standards for Children and Young Adults Using *Z* -Scores," *Circ: Arrhythmia and Electrophysiology*, vol. 13, no. 8, p. e008253, Aug. 2020, doi: 10.1161/CIRCEP.119.008253.

[3] P. Schwartz, "Guidelines for the Interpretation of the Neonatal Electrocardiogram," *European Heart Journal*, vol. 23, no. 17, pp. 1329–1344, Sep. 2002, doi: 10.1053/euhj.2002.3274.

[4] N. M. Hussain, B. Amin, B. J. McDermott, E. Dunne, M. O'Halloran, and A. Elahi, "Feasibility Analysis of ECG-Based pH Estimation for Asphyxia Detection in Neonates," *Sensors*, vol. 24, no. 11, p. 3357, May 2024, doi: 10.3390/s24113357.

[5] A. Abdou, S. Krishnan, and N. Mistry, "Evaluating a Novel Infant Heart Rate Detector for Neonatal Resuscitation Efforts: Protocol for Proof of Concept Study (Preprint)," *JMIR Research Protocols*, Jul. 2023, doi: 10.2196/45512.

[6] S. Krishnan, "Biomedical Signal Analysis for Connected Healthcare", Academic Press, 2021.

Address for correspondence:
Abdelrahman Abdou
350 Victoria St. M5B 2K3, Ontario, Canada
abdelrahman.abdou@torontomu.ca