Classification of Participant Age from ECG Signals Using Various Machine Learning Methods

Trevor Hess

May 12, 2021

I. Problem Statement

Heart rate (HR) and heart rate variability (HRV) monitoring has gained increased popularity in recent years for its ability to provide valuable feedback on health status, such as through the use of fitness monitoring. In connection, one of the most popular methods for collecting HR and HRV data is by capturing electrocardiogram (ECG) signals. The benefit of doing so is the ability to provide a wearable, noninvasive nature for ECG-based HR and HRV monitoring.

Of particular interest for this project is to apply multiple machine learning methods to determine how well participant age can be classified from HR and HRV features collected from ECG signals. It's already established that in general, lower HR and higher HRV correlate to better health. However, what remains unclear is how drastically these measurements correlate with age. Thus, by understanding this correlation, individuals who own fitness wearables that measure HR and HRV may better understand not only their health status, but their biological age rather than chronological age. Without this knowledge, HR and HRV trends are difficult to observe over the long-term and are generally only useful for day-to-day insights. However, understanding one's biological age is much clearer and yields better long-term insights, so it's much more likely that users will take further steps towards improving their health, such as by exercising more frequently and making better nutritional choices. For future work, this project could be used in correlation with a graphical user interface (GUI) such as by being integrated into a fitness wearable. However, the aim of this project is to simply provide the machine learning foundation which can then ultimately be combined with a wearable in the future.

As a result, in this paper, Section II will first provide a general background on ECG signals as well as cover past literature that has explored machine learning methods and applications for ECG classification. Exploring various machine learning methods helps determine which methods may be most accurate for our specific purpose. Exploring various applications gives insight into how machine learning methods can be applied to maximize the potential use of ECG signals. Section III explains the dataset, methodology, and validation approach used for this project as well as the initial hypothesis. The public dataset, *Fantasia* [1], is used to evaluate the proposed machine learning methods in this project. In connection, this project explores the potential of Support Vector Machine (SVM) and Discriminant Analysis (DA) machine learning methods for translating ECG data to HR and HRV features, which can then ultimately be used to classify participant age. Section IV reveals the results of using both methods. Specifically, results of using RBF SVM, Polynomial SVM, Linear SVM, Sigmoid SVM, and Quadratic DA (QDA) models are revealed. Section V discusses differences between the initial hypothesis and the results, limitations of the project, as well as opportunities for future work. Finally, Section VI briefly summarizes the overall project, the results obtained, as well as the contributions of the project.

II. Literature Survey

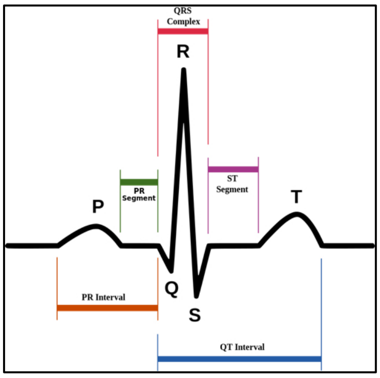

Fig. 1. Components of an ECG signal [2]

Electrocardiograms are electrical biopotentials of the heart, which are commonly recorded using 12 leads. The leads, or external electrodes, are used to measure the electrical conduction signals of the heart, which then can be graphed as ECGs. The ECG signals can be further broken up into individual parts, including the P, Q, R, S, and T waves, the PR and QT intervals, the PR and ST segments, as well as the QRS complex as shown in Figure 1. The ECG's parts are then generally used for computing HR and HRV as well as interpretating cardiac rhythms.

Furthermore, past work has explored various ECG signal classification methods to determine the accuracy of the methods. Multiple literature papers have explored ECG signal classification for the use of heartbeat identification, or distinguishing real and fake beats [3][4]. To accomplish this, Celin *et al.* [3] preprocessed input signals using filtered methods such as low pass, high pass, and butter worth to remove high frequency noise. After preprocessing, peak points were detected by using a peak detection algorithm and the features for the signals were extracted using statistical parameters. Finally, the extracted features were classified using SVM, Adaboost, ANN, and Naïve Bayes classifier methods to classify ECG signal data into normal and abnormal ECG signals. The paper found SVM and Naïve Bayes classifier methods to be the least and most effective methods, respectively. The paper provides valuable insights into the general framework for implementing classical machine learning methods like SVM. However, this project intends to go beyond the aims of Celin *et al.* [3] by using the framework and some of the methods to classify participant age from the ECG signals.

Furthermore, Zhang *et al.* [4] explored machine learning frameworks for instantaneous heart rate monitoring gathered from motion-artifact-corrupted ECG signals. To do this, SVM with a LASSO-based regularization ability for feature selection was implemented. Heartbeat identification then incorporated auto-segmentation, feature extraction, and classification to distinguish high confident heartbeats from large amounts of interferential spikes induced by artifacts, which was accomplished by using an SVM model. In addition, a refinement engine was used to remove outliers and interpolate missing heartbeats to yield better estimates. To do this, a two-step process was implemented. The first step was implementing a rule-based classifier, which incorporated two attributes, namely 'continuity check' and 'locality check,' for removing outliers. Then, the second step was a heartbeat interpolation strategy to address missing heartbeats. The paper found that the applied method was highly scalable, meaning new features could be easily added to the raw feature set. Thus, the paper provides valuable insights into useful features that can be used for this project like the mean and root mean square. However, this project intends to go beyond the aims of Zhang *et al.* [4] by implementing some of the features to then classify participant age from ECG signals.

Other papers were less applicable than the previously mentioned papers as they pertain to this project since they focused on exploring ECG signal classification for detecting arrhythmias, though still provided enough value to be mentioned [5][6][7]. Cattan *et al.* [5] implemented an unsupervised heart-rate estimation approach in wearables with liquid states and probabilistic readout. The paper's technique was first to encode spatio-temporal properties of ECG signals directly into spike train and use it to excite recurrently connected spiking neurons in a liquid state machine computational model. The paper then used a novel learning algorithm and designed an unsupervised readout based on fuzzy c-mean clustering of spike responses from a subset of liquid states using particle swarm optimization. Thus, the paper provides valuable insights into the difference between using supervised and unsupervised machine learning methods like Clustering, as well as evaluated their methods at the end using the training dataset, *Fantasia* [1], which is a useful dataset for this project. However, this project intends to go beyond the aims of Cattan *et al.* [5] by considering both supervised and unsupervised methods as well as using the *Fantasia* [1] dataset in order to classify participant age from ECG signals.

Mathews *et al.* [6] implemented a novel application of deep learning for single-lead ECG classification. To do so, Restricted Boltzmann Machine (RBM) and Deep Belief Network (DBN) methods were used for ECG classification following detection of ventricular and supraventricular heartbeats using single-lead ECG. Thus, the paper provides valuable insights into the difference between classical machine learning and deep learning methods like RBM and DBNs as well as illustrates that ECG classification can still be effective using lower sampling rates and simpler features like the mean of R-R intervals. However, this project intends to go beyond the aims of Mathews *et al.* [6] by considering the methods and features when selecting both for this project in order to classify participant age from ECG signals.

Finally, like Mathews *et al.* [6], Huanhuan *et al.* [7] explored classification of ECG signals using DBNs. To do this, original ECG signals were filtered. Then, DBNs were utilized to extract features with 3 R-R interval features composed into the feature vector, which are used for classification. Finally, four classifiers were selected to recognize ECG beats, with the nonlinear SVM with Gaussian kernel achieving the best results. Thus, the paper provides insights into the difference between classical machine learning methods and DBNs like SVM. However, this project intends to go beyond the aims of Huanhuan *et al.* [7] by considering the methods to then classify participant age from ECG signals.

As a summary, although the previously mentioned literature all aim at implementing machine learning to interpret ECG signals, this project aims to go beyond their work in several ways. First, this project intends to not only compute HR features from ECG signals, but compute HRV features as well. HR and HRV are considered to be equally important biometrics for quantifying age. Second, since HRV is being computed, this project uses more features for feature extraction. More specifically, features computed for this project include the mean, standard deviation, root mean square, and max-min slope for both HR and HRV. Lastly, previous literature aimed at measuring ECG signals and implementing machine learning methods to detect irregular or fake heartbeats. However, this project instead intends to measure ECG signals and implement Support Vector Machine and Discriminant Analysis machine learning methods to classify participant age from HR and HRV features.

III. METHOD

A. Dataset

The dataset, *Fantasia* [1], is used for this project, which is an ideal dataset considering the participants were rigorously screened to prevent health issues from being a concern within the data, as all participants were in a sinus rhythm. In addition, the participants were in a resting state when ECG signals were collected. This characteristic of the dataset is necessary, since in order to accurately measure a participant's health based on HR, the HR data must specifically be resting heart rate (RHR) data. If that's not the case, then the HR data won't be a good indicator of a participant's health. Additionally, it's difficult to record HRV signals when not in a resting state. Finally, the dataset is also ideal considering it consists of different ages of participants, which is necessary for being able to classify participant age from the ECG signals.

Regarding the specifications of the dataset, 20 young (21-34 years old) and 20 elderly (68-85 years old) participants, equal numbers of men of women, were subjected to 120 minutes of continuous supine resting while ECG, respiration, and continuous non-invasive blood pressure signals were collected. However, the ECG signals are only of interest for this project. There are 40 total participants, with half of the participants being those that included blood pressure signals and half that did not. In addition, 20 participants were of young and old age each. Participant 15 from the old participants was excluded for this project due to possessing significant outlier data. The filenames containing 'f2' are the participants datafiles that include blood pressure signals. In addition, the filenames containing 'y' and 'o' denote the young and old participants, respectively.

B. Proposed Methodology

In order to translate the ECG signals from the chosen dataset to HR and HRV features, which could then ultimately classify participant age, the approach was to implement (1) preprocessing, (2) peak detection and feature extraction, as well as (3) classification.

	p_signal	Age	Participant
0	7.952000	Young	P1
1	7.936000	Young	P1
2	8.004000	Young	P1
3	8.064000	Young	P1
4	7.980000	Young	P1
...	...	...	...
1693417	-0.041504	Old	P20
1693418	-0.039062	Old	P20
1693419	-0.043945	Old	P20
1693420	-0.029297	Old	P20
1693421	-0.012207	Old	P20

70036072 rows × 3 columns

Fig. 2. Dataframe of physical ECG signals for all participants

Preprocessing was implemented by extracting the physical ECG signals from the participant datafiles while ignoring the respiration and blood pressure signals. The physical ECG signals from all 39 participants were then combined to form a dataframe, with additional columns added to distinguish the age and participant number of each participant as shown in Figure 2. Again, the dataframe included 39 participants since participant 15 from the old participants was removed due to possessing significant outlier data.

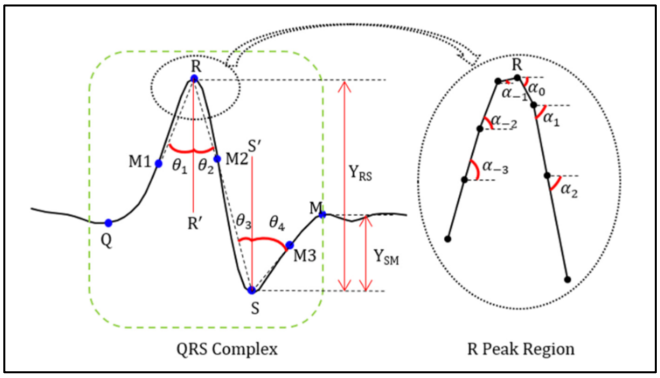

Fig. 3. QRS complex and R peak of ECG signals used for peak detection and feature extraction [4]

Next, peak detection is generally performed by detecting the QRS complex and R peak from ECG signals, as depicted in Figure 3. Thus, for this project, peak detection was accomplished using an 'xqrs' QRS detection algorithm. The algorithm first applies a bandpass filter between 5 and 20 Hz. Then, a moving wave integration (MWI) with a Ricker wavelet is applied onto the signal. Finally, the square of the filtered wave is saved. To be classified as a QRS complex, it must come after the refractory period, cross the QRS detection threshold, and not be classified as a T-wave if it comes close enough to the previous QRS complex. If it's not a QRS complex, it's classified as peak noise. If no QRS complex is detected within 1.66 times the recent R-R interval, backsearch QRS detection is performed [8]. This algorithm yields an array of QRS indices. However, peak correction was then performed by adjusting the detected peaks to coincide with local signal maxima. This was done using a search radius of 20 and smoothing window size of 100. This algorithm then yields an array of corrected QRS indices.

Using the array of corrected QRS indices, feature extraction was performed. First, HR and HRV, also known as the R-R interval, were computed from the QRS indices, yielding arrays of both instantaneous HRs and R-R intervals. To do so, a sampling frequency of 250 Hz was used for both. Then, features were extracted from the two arrays. The features extracted include the mean, standard deviation, root mean square, and max-min slope for both HRs and R-R intervals. The equations used to calculate these features are as follows:

$$\mu = \frac{\sum_{i=1}^{N} x_i}{N} \tag{1}$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{N}(x_i-\mu)^2}{N}} \tag{2}$$

$$x_{rms} = \sqrt{\frac{\sum_{i=1}^{N} x_i^2}{N}} \tag{3}$$

$$x_{mms} = \frac{x_{max}-x_{min}}{2} \tag{4}$$

μ	Mean
σ	Standard Deviation
x_{rms}	Root Mean Square
x_{mms}	Max-Min Slope
x	Sample

	hr_mean	hr_std	hr_rms	hr_mms	rr_mean	rr_std	rr_rms	rr_mms	Age	Participant
0	77.794495	6.381002	78.055748	23.448772	193.566228	16.879100	194.300768	71.5	Young	P1
1	77.173874	5.947497	77.402695	22.518177	194.946136	15.889678	195.592632	70.0	Young	P1
2	74.839676	6.701762	75.139145	22.391993	201.334677	19.048709	202.233790	68.0	Young	P1
3	72.336052	6.784934	72.653549	24.492018	208.451209	20.262598	209.433712	81.0	Young	P1
4	71.947021	5.957207	72.193230	22.427362	209.362416	18.013172	210.135898	77.5	Young	P1
...	...	...	...	...	...	...	...	...	...	...
503	72.760345	5.245121	72.949158	64.694771	206.466942	10.717915	206.744944	84.5	Old	P20
504	74.091591	5.945535	74.329765	46.784897	203.125203	11.955905	203.476760	86.0	Old	P20
505	75.702736	12.619240	76.747284	240.384613	199.468850	14.107862	199.967132	122.5	Old	P20
506	75.916023	12.065890	76.868904	239.340042	198.596979	12.627621	198.998033	117.5	Old	P20
507	74.604370	4.819539	74.759880	45.242004	201.547581	9.274888	201.760875	84.5	Old	P20

508 rows × 10 columns

Fig. 4. Windowed dataframe of extracted features for all participants

As shown in Figure 4, a windowed dataframe was then constructed, which included columns for all 8 features as well as additional columns for labeling the age and participant number for each participant. To construct the windowed dataframe, a sampling rate of 250 Hz, window size of 1000, and stride of 500 were used.

Finally, classification was performed using Support Vector Machine and Discriminant Analysis machine learning methods. Multiple kernels were implemented for the SVM to determine which would yield the best accuracy, which included RBF, Polynomial, Linear, and Sigmoid. However, only a Quadratic Discriminant Analysis approach was implemented for DA. The equation used for the QDA approach is as follows:

$$g_i(x) = -\frac{1}{2}|S_i| - \frac{1}{2}(x^T S_i^{-1} x - 2x^T S_i^{-1} m_i + m_i^T S_i^{-1} m_i) + \log \hat{P}(C_i) \quad (5)$$

$g(x)$	Discriminant of Sample x
x	Sample
m	Sample Mean
S	Covariance Matrix
$\hat{P}(C)$	Estimated Prior of Class C

C. Validation

After carrying out the proposed methodology, the goal of this project was ultimately to be able to determine whether SVM or DA is more accurate for classifying participant age from ECG signals. To do so, a 90/10 train/test split was implemented for both methods. As a result, 457 of the 508 total rows in the windowed dataframe were used for the training dataset for both methods. However, the training and testing datasets were manually constructed for DA, where 18 young and 17 old participants were used for the training dataset while 2 young and 2 old participants were used for the testing dataset. Priors of 0.5208 and 0.4792 were used for the young and old training sets, respectively, due to the chosen train/test split. Thus, both models are general rather than subject specific.

Based on the chosen dataset and methods for this project, it was expected that SVM would outperform Discriminant Analysis, with at least one SVM kernel exceeding 80% accuracy. The reasoning behind this hypothesis is that SVM is an optimization problem whereas DA is an analytical solution. As a result, DA makes use of the entire dataset to estimate covariance matrices, and so the method is prone to outliers. Conversely, SVM is optimized over a subset of the data, which is the data points, known as support vectors, that lie on the separating margin. Thus, the accuracy of SVM depends more heavily on the points that discriminate between the two participant age groups. Although one clear participant outlier was removed during preprocessing, it was expected that there would be less obvious outlier data in the dataset that would ultimately decrease the accuracy of DA more than SVM.

IV. Results

Table 1. Resulting accuracies of ML methods

ML Method	Kernel	Accuracy
Support Vector Machine	RBF	Testing: 60.7843%
		Training: 52.7843%
	Polynomial	Testing: 52.9412%
		Training: 51.8600%
	Linear	Testing: 90.1961%
		Training: 88.6214%
	Sigmoid	Testing: 52.9412%
		Training: 52.2976%
Discriminant Analysis	Quadratic	Testing: 49.0196%
		Training: 82.4945%

Table 1 depicts the resulting accuracies obtained from the methods performed. As can be observed, the Linear SVM performed the best with a testing accuracy of 90.1961% and training accuracy of 88.6214%. However, when comparing training accuracies between the Linear SVM and QDA, the training accuracy of the QDA was only slightly lower at 82.4945%. However, the QDA performed far worse than the Linear SVM on the testing dataset with an accuracy of 49.0196%. Thus, the results support the hypothesis that SVM would outperform DA.

Furthermore, the non-linear SVM models performed significantly worse than the linear SVM. The testing and training accuracies of the non-linear SVMs fell between 52-61% and 51-53%, respectively.

V. Discussion

First, it should be noted that a Linear Discriminant Analysis (LDA) approach was not appropriate to use for this project since the data was not normally or identically distributed. Thus, the covariance matrices were not identical. Regardless, QDA is more efficient to use considering a QDA would ultimately reduce to an LDA if the covariance matrices were identical, eliminating the need to compute both LDA and QDA.

Furthermore, considering that the Linear SVM performed the best suggests that the data was linearly separable. This makes sense considering all other kernels for the SVM, which were non-linear, performed significantly worse. A Linear SVM is also less prone to overfitting as compared to a non-linear SVM, which may have improved the accuracy of the linear SVM as well. In addition, the slightly lower training accuracy for the QDA as compared to the Linear SVM is assumed to be as a result of the fact that DA uses the entire dataset as previously explained. Thus, less obvious outlier data in the dataset likely reduced the accuracy of the QDA as compared to the linear SVM. Similarly, the significantly lower testing accuracy for the QDA suggests that the QDA overfit the data too much, which is also likely due to the fact that DA, again, uses the entire dataset as previously explained.

There are also multiple limitations of this project that should be noted. First, the dataset used for this project was relatively small compared to the number of people who use fitness wearables. Thus, it's not appropriate to assume SVM and DA can accurately classify participant age from ECG signals when the dataset only consisted of 39 participants. Future work should utilize a larger dataset to test the generalizability of both methods.

Similarly, the dataset only included normal ECG signals, meaning none of the participants had abnormal heart rhythms. Thus, the dataset is not realistic when compared to the normal population. As a result, future work should explore the efficacy of SVM and DA methods for classifying participant age from ECG signals that are both normal and abnormal.

Finally, advanced machine learning methods such as deep learning methods were not explored in this project due to the limited number of features used. However, there are many advantages to using deep learning such as much higher accuracy. Thus, future work should look to extract more features from the ECG signals in order to determine the potential for using deep learning methods as opposed to classical machine learning methods for classifying age from ECG signals.

VI. Conclusion

In this paper, SVM and DA methods were used to classify participant age from ECG signals. To accomplish this, preprocessing, peak detection and feature extraction, as well as classification were performed on the ECG signals collected from 39 participants. Preprocessing constructed a dataframe of the physical ECG signals of each participant from the participant datafiles. Peak detection filtered the ECG signals and constructed an array QRS indices, which were then corrected. Feature extraction constructed a windowed dataframe consisting of the mean, standard deviation, root mean square, and max-min slope of both HR and HRV. Finally, classification implemented RBF SVM, Polynomial SVM, Linear SVM, Sigmoid SVM, and QDA methods to classify participant age using the extracted features. From the classification, the Linear SVM performed the best with a testing accuracy of 90.1961% and training accuracy of 88.6214%. The non-linear SVM methods did not perform well, yielding testing and training accuracies between 52-61% and 51-53%, respectively. Finally, the Quadratic SVM yielded testing and training accuracies of 49.0196% and 82.4945%, respectively.

The results illustrate the potential for Linear SVM and QDA methods for classifying participant age from ECG signals. Thus, this project demonstrates that HR and HRV, computed from ECG signals, do in fact correlate with age. Furthermore, this may infer that improving HR and HRV through incorporating healthier lifestyle practices could extend one's lifespan. Considering that a biological age metric is more interpretable than HR and HRV, and provides better long-term insights, this project provides the opportunity for current fitness wearables to begin implementing biological age as a biometric for their users to track.

References

[1] Iyengar N, Peng C-K, Morin R, Goldberger AL, Lipsitz LA. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. Am J Physiol 1996;271:1078-1084.

[2] L. Potter, "Understanding an ECG: ECG Interpretation," *Geeky Medics*, 11-Apr-2021. [Online]. Available: https://geekymedics.com/understanding-an-ecg/. [Accessed: 24-Apr-2021].

[3] S. Celin and K. Vasanth, "ECG Signal Classification Using Various Machine Learning Techniques," *Journal of Medical Systems*, vol. 42, no. 12, Oct. 2018.

[4] Q. Zhang, D. Zhou, and X. Zeng, "A novel machine learning-enabled framework for instantaneous heart rate monitoring from motion-artifact-corrupted electrocardiogram signals," *Physiological Measurement*, vol. 37, no. 11, pp. 1945–1967, Sep. 2016.

[5] Grégoire Cattan, A. Das, P. Pradhapan, W. Groenendaal, P. Adiraju, R. T. Rajan, F. Catthoor, S. Schaafsma, J. L. Krichmar, N. Dutt, and C. Van Hoof, "Unsupervised heart-rate estimation in wearables with Liquid states and a probabilistic readout," *Neural Networks*, vol. 99, pp. 134–147, Jan. 2018.

[6] S. M. Mathews, C. Kambhamettu, and K. E. Barner, "A novel application of deep learning for single-lead ECG classification," *Computers in Biology and Medicine*, vol. 99, pp. 53–62, May 2018.

[7] M. Huanhuan and Z. Yue, "Classification of Electrocardiogram Signals with Deep Belief Networks," *2014 IEEE 17th International Conference on Computational Science and Engineering*, Dec. 2014.

[8] *wfdb*, 2018. [Online]. Available: https://wfdb.readthedocs.io/en/latest/index.html. [Accessed: 01-May-2021].