BME 772 Course Project Summary Sheet

Project Title: An Examination of the Impact of Cerebromicrovascular Disease on Elderly People with Diabetes

Lastnames (A, B, C): Ashemi, Harder, Van Greuning

Dataset Used: CVES database 24-hour ECG signals

Dataset Link:
https://www.physionet.org/content/cves/1.0.0/data/24h-electromyography/#files-panel

Number of Signals and Number of Classes	2 classes (Diabetic & Control) 60 control signals and 60 diabetic signals 28 signals were examined (13 control & 15 Diabetic)
Normal Signals	13 normal signals
Preprocessing Techniques Used	Notch Filtration Pan-Tompkins Algorithm
Feature Extraction Methods Used	Pan-Tompkins Algorithm to detect PQRST points
Machine Learning Methods Used	Supervised Logistic Regression, 5-fold cross-validation
Overall Accuracy Rate or other ML metrics	Testing: 54%, Validation: 62.25%, Sensitivity: 66.67%, Specificity: 53.85%, Classification Accuracy: 60.71%, AUC: 0.60
Best Result in the Literature (if known)	Test accuracy (80.86%), Validation accuracy (≈ 75-85%), Sensitivity (46.34%), Specificity (86.66%), Classification accuracy (≈ 66.5%), AUC (0.7932)

Toronto Metropolitan University

Department of Electrical, Computer, & Biomedical Engineering
Faculty of Engineering & Architectural Science

Course Title:	Biomedical Signal Analysis
Course Number:	BME772
Semester/Year (e.g.F2016)	Fall 2024

Instructor:	Dr. Sridhar Krishnan

Assignment/Lab Number:	5
Assignment/Lab Title:	BME772 Course Project

Submission Date:	November 25, 2024
Due Date:	November 25, 2024

Student LAST Name	Student FIRST Name	Student Number	Section	Signature*
Ashemi	Syed	xxxx51173	1	S.A
Harder	Robert	xxxx42149	4	R.H.
Van Greuning	Aiden	xxxx40587	2	A.V.

*By signing above you attest that you have contributed to this written lab report and confirm that all work you have contributed to this lab report is your own work. Any suspicion of copying or plagiarism in this work will result in an investigation of Academic Misconduct and may result in a “0” on the work, an “F” in the course, or possibly more severe penalties, as well as a Disciplinary Notice on your academic record under the Student Code of Academic Conduct, which can be found online at:
https://www.torontomu.ca/content/dam/senate/policies/pol60.pdf

Toronto Metropolitan University
Department of Electrical, Computer, & Biomedical Engineering
Faculty of Engineering & Architectural Science

An Examination of the Impact of Cerebromicrovascular Disease on Elderly People with Diabetes

By: Syed Ashemi, Robert Harder, Aiden Van Greuning

I. INTRODUCTION

In the realm of biomedical engineering, electro- cardiogram (ECG) signals serve several functions, chief among them being the monitoring of the human heart. Many diagnoses, such as diabetes, arrhythmia, heart disease, and cardiomyopathy, can result from heart monitoring [1]. Diabetes is one of the most prevalent illnesses in the world, affecting hundreds of millions of people globally. Type 1 and type 2 diabetes affect almost 5.7 million individuals in Canada [2]. In 2017, a study was conducted showing that of all the patients that were diagnosed with diabetes, 90% of them were listed as type 2 while 9% were listed as type 1 [3]. A patient's diabetes can be inferred indirectly from their ECG readings. Diabetes affects the human body in many ways, and living with the disease can be challenging. In some cases, the more serious the diabetes condition, the more potentially fatal it might become. Numerous symptoms, including increased frequency of urination, blurred vision, hunger after meals, weight loss even after eating more (type 1), and tingling, pain, or numbness in the hands or feet (type 2), are associated with diabetes; however, the heart is especially vulnerable to the potentially deadly effects of the disease [4].

Both the cardiac muscle and blood arteries may sustain damage from diabetes. Consequently, persistent damage to the heart muscle and blood arteries will lead to cardiac failure. The brain's tiny blood vessels sustain significant damage, leading to cerebral microvascular disease, also known as Cerebral Small Vessel Disease (CSVD) [5]. It is brought on by an accumulation of proteins called plaques, which block blood flow in the brain's tiny arteries [5]. An examination of the impact of Cerebral Small Vessel Disease on an ECG may provide crucial information about the relationship between diabetes and cardiovascular health.

Conversely, diabetic patients frequently exhibit anomalies in the produced ECG signal. This results from the link between diabetes and cardiac disease [5]. The body's cardiovascular system is then impacted, which results in the irregularities shown on an ECG signal.

II. BACKGROUND

Signal extraction and processing are the first procedures utilized to look at and notice these irregularities in the ECG. Through signal processing, extraneous artifacts and noise may be eliminated from the signal. Throughout these processes, the Pan Tompkins method, derivative, bandpass, notch, and other analytic tools are often used. These phases are referred to as preprocessing phases. To get the signal ready for further analysis, pretreatment steps are necessary. The heart rate (RR interval of the QRS complex), mean and standard deviation of the designated RR interval, mean and standard deviation of the ST interval, and mean and standard deviation of the QT interval are significant characteristics that can be extracted and analyzed for both normal ECG signals and signals from diabetic patients [6]. The RR interval is used to calculate a person's heart rate variability. Since early diagnosis may result in the assessment and prevention of cardiovascular illnesses, this knowledge is essential. While a long-term fluctuation suggests that there may be autonomic control or underlying cardiovascular diseases, a short-term variability indicates that there might be problems with respiratory effects [6].

The time difference between two R waves in the QRS complex is known as the RR interval [6]. As previously established, the patient's heart rate is determined by the RR interval [6]. The ST interval, which is utilized to diagnose myocardial ischemia and infarction, determines the measure of ventricular repolarization inside the QRS complex [7]. Furthermore, the temporal difference between ventricular depolarization and repolarization is measured by the QT interval inside the QRS complex [7]. Long QT waves can indicate problems like arrhythmia, which QT complexes can see. However, a shorter QT wave leads to the conclusion that atrial and ventricular fibrillation are more likely to occur [7]. Following the extraction of these characteristics, a machine learning algorithm—such as the logistic regression model—will be used to confirm the correctness and validity of the given dataset. Both the

control and diabetes patient groups utilize machine learning models. To forecast and determine the likelihood of a binary event happening, the logistic regression model would be utilized [8]. In this instance, the diabetic group would be identified by using this model to identify the binary events that are taking place within the signals as well as any potentially aberrant occurrences.

III. METHODS

A. Signal Pre-processing:

The data to be processed was the "Cerebro- microvascular disease in elderly with diabetes" dataset, found in the CVES database. Within this dataset were 120 total subjects with an age range of 55-75. 50% of the subjects were in the control group while the other 50% were diabetic. With a sampling rate of 1000 Hz, the signals were collected over a time interval of 24 hours [9]. The specific number of signals extracted was 15 diabetic signals and 13 control signals.

Method 1:

The first method for signal processing utilized was the notch filter to remove power line interference. Power line interference is when the power supply encounters electromagnetic interferences with the typical range between 50-60 Hz [10]. The notch filter was generated by placing zeroes on the unit circle at the spot that correlates to the point of noise that must be extracted. For the samples utilized, the zero was placed at the frequency of 60 Hz, resulting in noise removed at that frequency where the power line interference occurs. The following three equations were utilized for the notch filter with the third one being the final notch filter transfer function:

(1) $\Theta = 2\pi\frac{f_0}{f_s}$

(2) $H(z) = (1 - z^{-1})^2$

(3) $H(z) = 1 - 1.86z^{-1} + z^{-2}$

Method 2:

The second method for signal processing was the utilization of the Pan Tompkins (PT) algorithm. As one of the most important biomedical signals, it is paramount that ECG signals are devoid of noise and any external variables that can cause signal disturbance. Sources of signal disturbance for ECGs range from high and low-frequency noise, muscle contraction noise, power-line interference, and baseline wander. The PT algorithm first utilizes a digital bandpass filter that is a cascading of a low-pass and high-pass filter to attenuate the high and low-frequency noise respectively. With a passband frequency of 5-15 Hz; the low-pass part of the bandpass filter can also attenuate EMG noise and T wave noise while the high-pass part can attenuate baseline wander [11]. The following two equations were utilized for the low-pass and high filter respectively:

(4) $H(z)_{LP} = \left(\frac{1-z^{-6}}{1-z^{-1}}\right)^2$

(5) $H(z)_{HP} = z^{-16} - \frac{1}{32}\left(\frac{1-z^{-32}}{1-z^{-1}}\right)$

The next portion of this important algorithm is the high pass differentiator filter; utilized to detect the QRS cineplex specifically by isolating the high peaks of the QRS complex for analysis [11]. The following equation was utilized for the differentiator:

(6) $H(z) = \frac{1}{8}(2 + z^{-1} - z^{-3} - z^{-4})$

The PT algorithm then uses the squared operator to square the signal point by point to further emphasize the specific positive values and amplify the QRS peaks [11]. The equation for this is expressed below:

$$(7)\quad y(n) = [x(n)]^2$$

Finally, the signal is passed through a moving average integrator with a sampling frequency of 1000 Hz and a window length of 150 samples [12] [13]. Each individual sample is approximately 150 ms in width and this integrator is expressed in the following equation:

$$(8)\quad H(z) = \frac{1}{150}(1+ z^{-1} + z^{-2} + z^{-3} + z^{-4} + \ldots z^{-149})$$

After these two methods are utilized, peaks will be identified, and concerning the visible QRS peaks; their signal features will be analyzed.

B. Feature Analysis

The two categories of feature extraction are detection and extraction. Detection simply consists of finding the peaks while peak extraction can occur by finding either the QRS points or ST points. By implementing the following equations, the methods can be executed:

Heart rate equation:

$$(9)\quad HR(bpm) = \frac{60}{RR\ interval}(1000)$$

Mean equation:

$$(10)\quad \bar{x} = \frac{\Sigma x}{N}$$

Standard deviation equation:

$$(11)\quad SD = \sqrt{\frac{\Sigma|x-\bar{x}|^2}{n}}$$

Through what is known as feature extraction, parts of the data can be categorized and classified. There are unique features of the ECG in the time domain that can be visualized in the PQRST point. In this scenario, the ECG toolbox was used to identify all 5 locations of all 5 individual points of the ECG signal. The mean and standard deviation (SD) intervals are found in between these points, with the quantity a potential hindrance for further machine learning processes. To avert this, 6 required features were utilized in the machine learning classification. These were: heart rate, mean and SD intervals from ST, mean and SD intervals from QT, and the SD interval from RR. Due to the greatest effect on the diabetes ECG signals being from these features, they were chosen accordingly [14] [15].

Method 1: Finding the PQRST points

To apply peak detection for all points of R on the ECG signal, the moving window average plot was used. Directly following the R peaks was the S point, meaning that the ECG toolbox was searching for the minimum point in between the R point and the sampling frequency multiplied by 0.1 (0.1*fs) [16]. Due to the Q point being seen before the R peak, the same ECG toolbox was searching for the minimum point between the R point and the sampling frequency multiplied by 0.08 (0.08*fs). In turn, with the toolbox searching for half the RR distance between the R and S peaks, the P and T peaks were found. Using the aforementioned equations, the mean and SD for the RR interval and heart rate were calculated.

Method 2: Finding PQRST intervals

After detecting the PQRST peaks with Method 1, the intervals are located by utilizing the ECG toolbox. This toolbox identifies the closest R points and pairs them with the corresponding closest P, Q, S, and T points [16]. Due to period length, these points are required to be near the R peaks; otherwise, the period length would be voided. Interval calculations were conducted by finding the difference between the points. The mean and SD were both calculated using the same equations for Method 1.

C. Machine Learning

Machine learning algorithms are typically chosen for classification purposes. In this case, the algorithms used were chosen for the degree of diversity and variety provided. The key objective the model was utilized for was observing the typical behaviour of the data across a multitude of methods relating to classification. Furthermore, the model chosen was found to be the closest to reliable for the classification of diabetes from an ECG signal. Three models were considered: the Logistic Regression Model, the Decision Tree Model, and the Linear Discriminant Model. Ultimately, the Logistic Regression model proved to be the most accurate for implementation.

Logistic Regression Model:

The Logistic Regression model is the one that shall be tested. It is similar in functionality to the Linear Discriminant model; however the former has no division of the classifications alongside the linear discriminant boundary. Essentially, the model attempts to fit the most accurate and precise S-shaped predictor line to the data. On one axis are the class labels while the feature required to be set is on the opposite axis. In turn, a logit line can be set to get as many labels as possible on a variety of points that follow alongside the scale. To accurately map the highest amount of labels possible, a shift of the line is conducted. This line becomes the linear regression line and is used based on the feature. In turn, precise identification of the relative importance of the feature and its weight concerning the rest of the features within the broader dataset is enabled by the line [8]. If the specific feature has a greater weight, there will be a stronger influence on the overall prediction of the model.

IV. RESULTS

Part A: Pre-processing of Signal

A. Control Subject Signal:

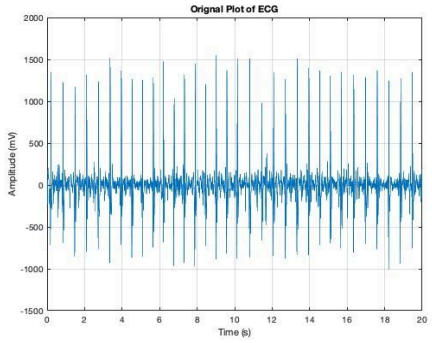

Figure 1: *The unfiltered ECG signal for Subject 6 is displayed here as the control signal.*

Method # 1 -

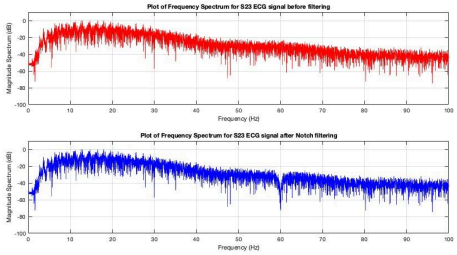

Figure 2: *The frequency spectrum of the ECG signal for Subject 6 control signal shows the before and after plots following notch filtering.*

Method # 2 -

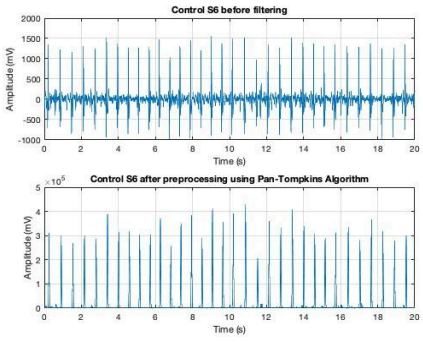

Figure 3: *After applying notch filtering and Pan-Tompkins filtering to the control signal, the ECG signal of Subject 6 is plotted before and after in the figure above.*

B. Diabetic Subject Signal:

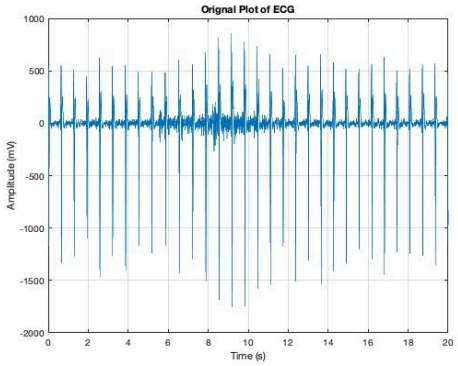

Figure 4: *The unfiltered ECG signal for Subject 23 is displayed here as the diabetic signal.*

Method # 1 -

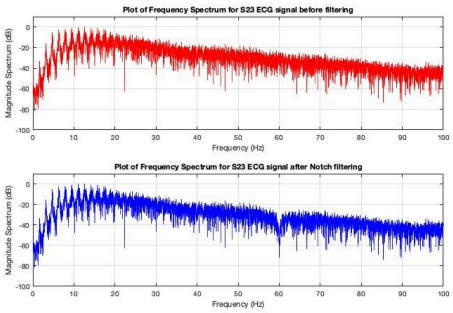

Figure 5: *The frequency spectrum of the ECG signal for Subject 23's diabetic signal shows the before-and-after plots following notch filtering.*

Method # 2 -

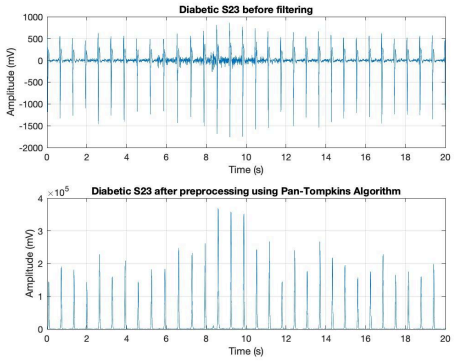

Figure 6: *After applying notch filtering and Pan-Tompkins filtering to the control signal, the ECG signal of Subject 23 is plotted before and after in the figure above.*

Part B: Feature Analysis of ECG Signal

A. Control Subject Signal:

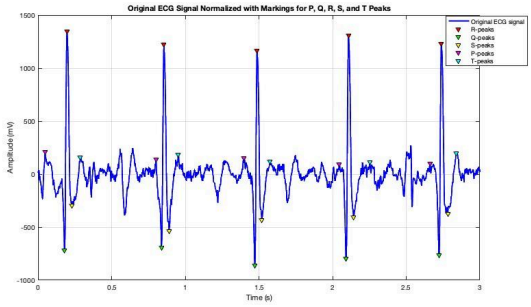

Figure 7: *When peak detection was applied, the graphic displays every PQRST point found in the ECG signal of Subject 6. Uses peak detection to display every PQRST point in the ECG signal.*

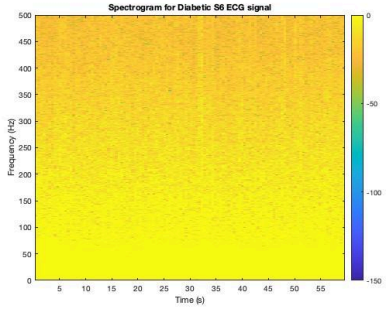

Figure 8: *A spectrogram for ECG signal Subject 6 is represented by the figure above.*

Signal	NumberofBeats	RRInterval	SDofRRInterval	AvgQRSLength	HR	MeanSTInterval	SDofSTInterval	MeanQTInterval	SDofQTInterval
ECG56	102	589.57	40.508	35.912	101.77	114.07	30.766	242.63	32.45

Figure 9: *Shows ECG characteristics extracted from Subject 6's control filtered signal.*

B. Diabetic Subject Signal:

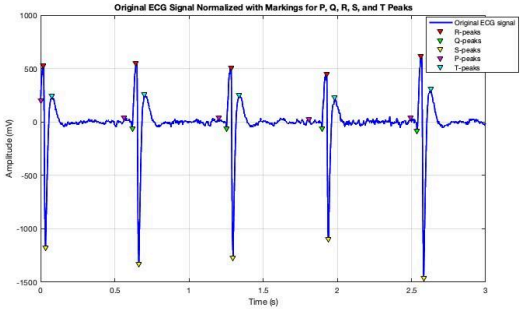

Figure 10: *When peak detection was applied, the graphic displays every PQRST point found in the ECG signal of Subject 23. Uses peak detection to display every PQRST point in the ECG signal.*

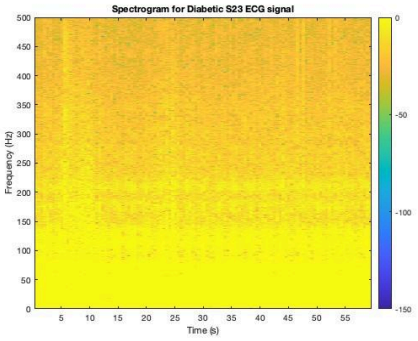

Figure 11: *A spectrogram for ECG signal Subject 23 is represented by the figure above.*

Signal	NumberofBeats	RRInterval	SDofRRInterval	AvgQRSLength	HR	MeanSTInterval	SDofSTInterval	MeanQTInterval	SDofQTInterval
ECGS23	95	633.53	14.84	33.493	94.707	162.52	49.694	336.01	54.311

Figure 12: *Shows ECG characteristics extracted from Subject 23's diabetic filtered signal.*

Part C: Machine Learning (Logistic Regression)

```
Overall Testing Accuracy (Testing Percentage): 54.00%
Overall Validation Accuracy (Validation Percentage): 62.25%
```

Figure 13: *Displays Testing and Validation Percentage using Logistic Regression.*

To verify the correctness of the retrieved characteristics, the Logistic Regression model—a sort of machine learning algorithm—was used overall. Because the data pool was small, a TA 5-fold cross-validation was employed rather than the 10-fold conventional cross-validation. The data was used to test the models after they had been validated, with the testing group accounting for approximately 23% of the total. The results of the test accuracy and validity are displayed in Table 1.

Model	Testing %	Validation %
Logistic Regression	54	62.25

Table 1: *The Decision Logistic Regression model's testing accuracy and validation percentages are shown in the table below.*

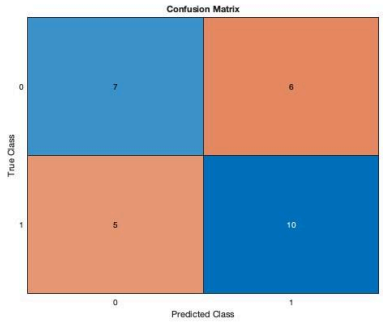

Figure 14: *The confusion matrix for logistic regression where 0 represents the control and 1 represents the diabetic is depicted in the figure above.*

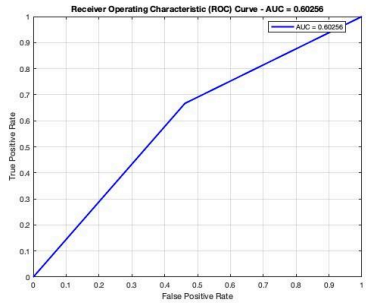

Figure 15: *Shows the plot of ROC with the area under curve (AUC).*

```
Confusion Matrix:
     7     6
     5    10

Sensitivity: 66.67%
Specificity: 53.85%
Classification Accuracy: 60.71%
AUC (Area Under Curve): 0.60
```

Figure 16: *Displays values of sensitivity, specificity, classification accuracy, and AUC.*

V. DISCUSSION

A. Signal Pre-processing

The first step in signal pre-processing was to apply a notch filter configured to specifically filter out 60 Hz noise. After referring to outside sources, it was determined that ECG signals most commonly experience a basic powerline interference at either the 50 Hz or 60 Hz locations [17]. The notch filter was specially designed for 60 Hz noise because the dataset used for this project came from North America, where electricity is delivered at 60 Hz. The frequency spectra for the control, subject 6, show a distinct drop in magnitude at the 60 Hz mark after the notch filter was applied that was not present beforehand (see Figure 2 above). This noticeable magnitude drop at the 60 Hz mark after notch filtering can also be seen for the diabetic case, subject 23 (see Figure 5 above). Overall, the notch filter was very effective at reducing noise in the 60 Hz range.

The final step in pre-processing was the implementation of the Pan-Tompkins algorithm, consisting of a bandpass filter, differentiator, squaring operator, and a moving-window integrator. The purpose of the algorithm was to isolate the QRS complexes of the ECG signal to perform feature analysis. The final results of the Pan-Tompkins algorithm contained very little noise and displayed

definitive peaks in the signal (see Figures 3 & 6). This algorithm proved to be very useful in reducing unnecessary noise and isolating the QRS complexes for each ECG signal.

A. Feature Analysis

The moving average operator used for peak detection made the peaks very easy to identify and was quite effective at determining all of the PQRST points of each given ECG signal. Figures 7 & 10 above show the results of pitch detection and how it was able to accurately identify all of the necessary points for control subject 6 and diabetic subject 23, respectively. What made this so effective was the use of the R & S points as a reference for determining the other points. These points were then used to calculate the various intervals and values to be used in training the machine learning algorithm.

Healthy adults generally have a resting heart rate in the range of 60-100 beats per minute (bpm) [18]. A higher resting heart rate is associated with an increased risk of type 2 diabetes [19]. These statistical ranges are consistent with the obtained data. Among the healthy population of subjects, the minimum heart rate detected was 59.4 bpm and the maximum heart rate detected was 79 bpm. In the set of diabetic subjects, the minimum detected heart rate was 60.8 bpm and the maximum was 88.7 bpm. Since the R-R mean was used to calculate the heart rate, these trends in the heart rate are also present in the R-R mean. Lower heart rate variability (HRV) is associated with cerebral small vessel diseases such as type 2 diabetes [20]. In the context of these results, it means that the R-R standard deviation would tend to be smaller for diabetic subjects than their healthy counterparts. This did not prove to be the case, however, as there was no significant difference between the R-R standard deviations for diabetic and healthy patients. This may have been the result of a limited dataset of only 28 signals. Perhaps a larger dataset would have produced more conclusive results.

A study published by the National Library of Medicine found that diabetic patients tend to have higher changes in their S-T intervals [21]. This study corresponds with our obtained results as the diabetic patients tended to have higher S-T standard deviations than the control group. In addition to this, the diabetic patients tended to have higher mean S-T intervals than those of the control patients.

A healthy adult's normal Q-T interval is approximately 420 milliseconds for a heart rate of 60 bpm, whereas diabetic patients tend to have Q-T intervals longer than this [22]. These values correspond with our obtained results, as the mean Q-T interval for the diabetic subjects tended to be higher than those of the control subjects. The diabetic subjects also tended to have higher Q-T standard deviations than the control subjects. These data trends were indicative of diabetic patients having longer and more variable Q-T intervals.

B. Machine Learning

The Logistic Regression model was used for the machine learning (ML) stage of this project and produced a 62.25% validation accuracy, a 54% testing accuracy, and an area under curve (AUC) of 0.60 (see Table 1 above). With such a limited dataset, the Logistic Regression model is susceptible to overfitting, and is unable to account for significant outliers among the testing and validation datasets. A 2019 study of various machine learning techniques for type 2 diabetes risk prediction analyzed data on over 138,000 participants, including over 20,000 with type 2 diabetes. The Logistic Regression model used in this study produced a testing accuracy of 80.86%, sensitivity of 46.34%, specificity of 86.66%, and an AUC of 0.7932 [23]. This shows that our model still has room for improvement and would likely benefit from a larger dataset.

VI. CONCLUSION

After conducting signal extraction and processing on ECG signals from patients with type 2 diabetes; the accuracy of identification of diabetic patients was 62.25%. A notch filter and the Pan-Tompkins algorithm were both implemented to conduct the preprocessing of the ECG signal. Both feature extraction and peak detection were then conducted to find the R peaks within the signal. Next, to find the PQRST points an ECG toolbox was implemented as an aid. This toolbox was also key to finding important features within the ECG signals such as the mean and standard deviation of the S-T interval, the mean and standard deviation of the Q-T interval, and the heart rate. Finally, using machine learning the extracted features were analyzed and the Logistic Regression model was implemented.

VII. REFERENCES

[1] M. Kelly and C. Semsarian, "Multiple mutations in genetic cardiovascular disease," *Circulation: Cardiovascular Genetics*, vol. 2, no. 2, pp. 182–190, Apr. 2009. doi:10.1161/circgenetics.108.836478

[2] "March 03, 2022 diabetes rates continue to climb in Canada," DiabetesCanadaWebsite, https://www.diabetes.ca/media-room/press-releases/diabetes-rates-continue-to-climb-in-canada (accessed Nov. 15, 2024).

[3] P. H. A. of Canada, "Government of Canada," ARCHIVED Diabetes in Canada - Canada.ca, https://www.canada.ca/en/public-health/services/publications/diseases-conditions/diabetes-canada-highlights-chronic-disease-surveillance-system.html (accessed Nov. 15, 2024).

[4] "Good to know: Diabetes symptoms and tests," Clinical diabetes : a publication of the American Diabetes Association, https://pmc.ncbi.nlm.nih.gov/articles/PMC6969655/ (accessed Nov. 15, 2024).

[5] A. Gardener, "What is cerebral small vessel disease (CSVD)?," American Brain Foundation, https://www.americanbrainfoundation.org/julies-story/ (accessed Nov. 15, 2024).

[6] "RR interval," RR Interval - an overview | ScienceDirect Topics, https://www.sciencedirect.com/topics/medicine-and-dentistry/rr-interval (accessed Nov. 15, 2024).

[7] E. Burns, R. Buttner, and E. B. and R. Buttner, "The St Segment," Life in the Fast Lane • LITFL, https://litfl.com/st-segment-ecg-library/ (accessed Nov. 15, 2024).

[8] J. Brownlee, "Logistic regression for machine learning," MachineLearningMastery.com, https://machinelearningmastery.com/logistic-regression-for-machine-learning/ (accessed Nov. 15, 2024).

[9] T. Hardigan, R. Ward, and A. Ergul, "Cerebrovascular complications of diabetes: Focus on cognitive dysfunction," Clinical science (London, England : 1979), https://www.ncbi.nlm.nih.gov/pmc/arti cles/PMC5599301/ (accessed Nov. 15, 2024)

[10] C. Levkov, G. Mihov, R. Ivanov, I. Daskalov, I. Christov, and I. Dotsinsky, "Removal of power-line interference from the ECG: a review of the subtraction procedure," *BioMedical Engineering OnLine*, vol. 4, no. 1, p. 50, 2005, doi: https://doi.org/10.1186/1475-925x-4-50. (accessed Nov. 15, 2024)

[11] J. Pan and W. J. Tompkins, "A Real-Time QRS Detection Algorithm," *IEEE Transactions on Biomedical Engineering*, vol. BME-32, no. 3, pp. 230–236, Mar. 1985, doi: https://doi.org/10.1109/tbme.1985.325532.

[12] F. Liu *et al.*, "Performance Analysis of Ten Common QRS Detectors on Different ECG Application Cases," *Journal of Healthcare Engineering*, vol. 2018, p. 9050812, May 2018, doi: https://doi.org/10.1155/2018/9050812. (accessed Nov. 15, 2024)

[13] M. Elgendi, "Fast QRS detection with an optimized knowledge-based method," National Library of Medicine, https://www.ncbi.nlm.nih.gov/pmc/arti cles/PMC3774726/ (accessed Nov. 15, 2024)

[14] "Prevalence of ECG abnormalities in people with type 2 diabetes: The Hoorn Diabetes Care System cohort," *Journal of Diabetes and its Complications*, vol. 35, no. 2, p. 107810, Feb. 2021, doi: https://doi.org/10.1016/j.jdiacomp.2020.107810. (accessed Nov. 15, 2024)

[15] S. Gupta, R. K. Gupta, M. Kulshrestha, and R. R. Chaudhary, "Evaluation of ECG abnormalities in patients with asymptomatic type 2 diabetes mellitus," National Library of Medicine, https://www.ncbi.nlm.nih.gov/pmc/arti cles/PMC5449835/ (accessed Nov. 15, 2024)

[16] "ECG Signal PQRST Peak Detection Toolbox," MathWorks, https://www.mathworks.com/matlabce ntral/fileexchange/73850-ecg-signal-p qrst-peak-detection-toolbox (accessed Nov. 15, 2024)

[17] L. Sörnmo, P. Laguna, *Bioelectrical Signal Processing in Cardiac and Neurological Applications*, Academic Press, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/B9780124375529500076 (accessed Nov. 24, 2024)

[18] "Normal heart rate: Ranges, danger, and more". MedicalNewsToday. https://www.medicalnewstoday.com/articles/235710 (accessed Nov. 24, 2024)

[19] D.H. Lee, L.F.M. de Rezende, F.B. Hu, J.Y. Jeon, E.L. Giovannucci, "Resting heart rate and risk of type 2 diabetes: A prospective cohort study and meta-analysis", National Library of Medicine, https://pmc.ncbi.nlm.nih.gov/articles/PMC6398339/ (accessed Nov. 24, 2024)

[20] Qiu Q., Song W., Zhou X., Yu Z., Wang M., Hao H., Pan D., Luo X., "Heart rate variability is associated with cerebral small vessel disease in patients with diabetes", National Library of Medicine, https://pmc.ncbi.nlm.nih.gov/articles/PMC9685533/ (accessed Nov. 24, 2024)

[21] Gupta S., Gupta R.K., Kulshrestha M., Chaudhary R.R., "Evaluation of ECG Abnormalities in Patients with Asymptomatic Type 2 Diabetes Mellitus", National Library of Medicine, https://pmc.ncbi.nlm.nih.gov/articles/PMC5449835/ (accessed Nov. 24, 2024)

[22] Rosenthal, L., "Normal Electrocardiography (ECG) Intervals", Medscape, https://emedicine.medscape.com/article/2172196-overview?form=fpf (accessed Nov. 24, 2024)

[23] Xie, Z., Nikolayeva, O., Luo, J., Li, D., "Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques", U.S. Centers for Disease Control and Prevention, https://www.cdc.gov/pcd/issues/2019/19_0109.htm# (accessed Nov. 24, 2024)

VIII. APPENDIX

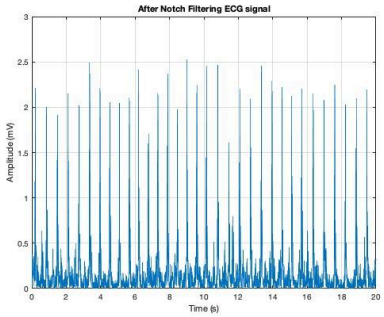

Figure 17: *Following notch filtering, the ECG signal of Subject 6 is shown as the control signal.*

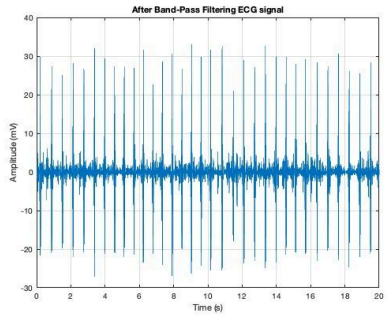

Figure 18: *Following bandpass filtering, the figure displays Subject 6's ECG signal for the control signal.*

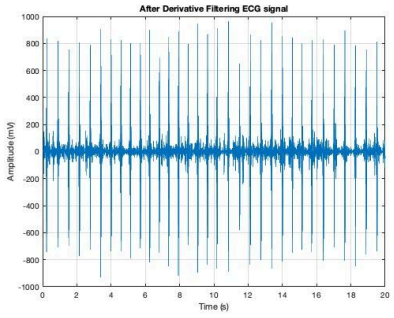

Figure 19: *Following derivative filtering, the figure displays Subject 6's ECG signal for the control signal.*

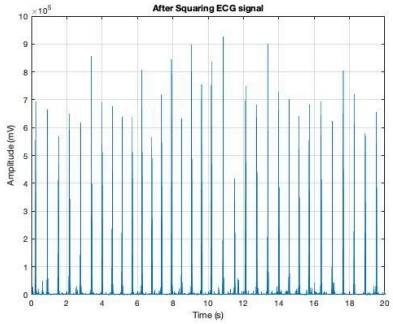

Figure 20: *Following the squaring operator, the figure displays Subject 6's ECG signal for the control signal.*

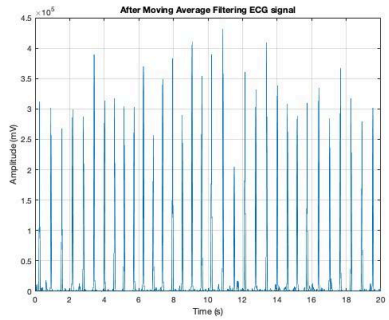

Figure 21: *Following moving average filtering, the figure displays Subject 6's ECG signal for the control signal.*

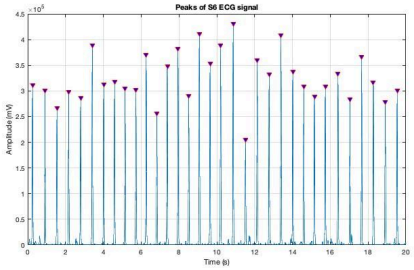

Figure 22: *Following moving average filtering and its peaks highlighted, the figure displays Subject 6's ECG signal for the control signal.*

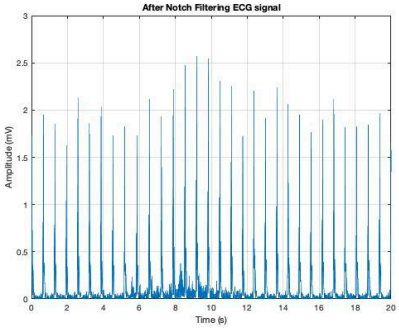

Figure 23: *Following notch filtering, the ECG signal of Subject 23 is shown as the diabetic signal.*

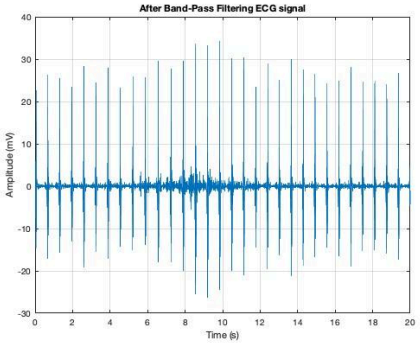

Figure 24: *Following bandpass filtering, the figure displays Subject 23's ECG signal for the diabetic signal.*

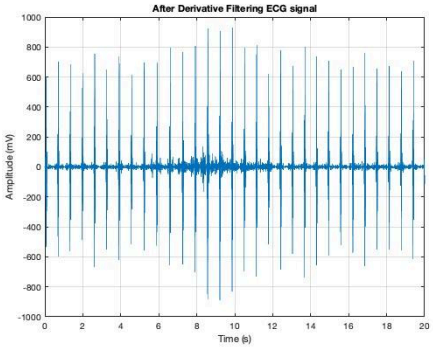

Figure 25: *Following derivative filtering, the figure displays Subject 23's ECG signal for the diabetic signal.*

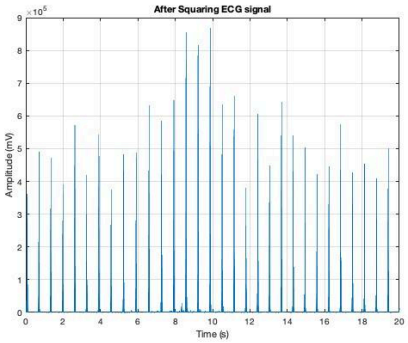

Figure 26: *Following the squaring operator, the figure displays Subject 23's ECG signal for the diabetic signal.*

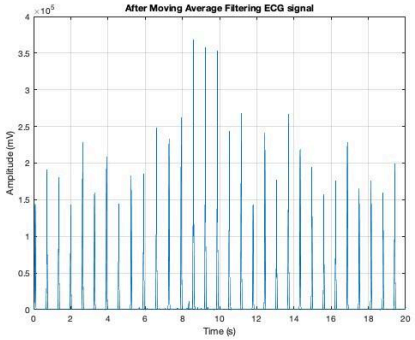

Figure 27: *Following moving average filtering, the figure displays Subject 23's ECG signal for the diabetic signal.*

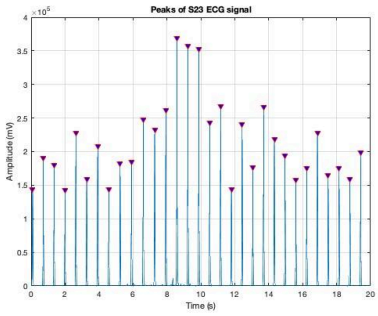

Figure 28: *Following moving average filtering and its peaks highlighted, the figure displays Subject 23's ECG signal for the diabetic signal.*

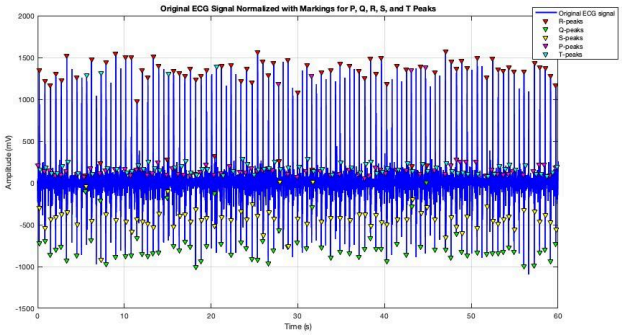

Figure 29: *The PQRST points found in the ECG signal of Subject 6 are zoomed out in the plot, which shows a minute-long time interval.*

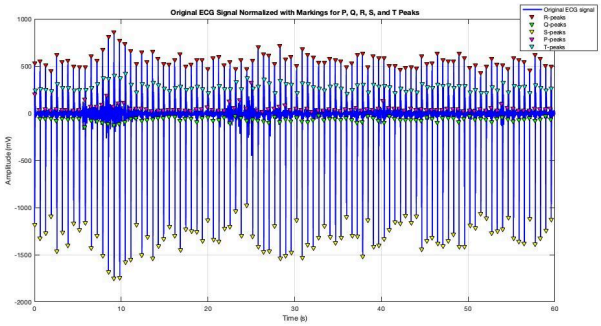

Figure 30: *The PQRST points found in the ECG signal of Subject 23 are zoomed out in the plot, which shows a minute-long time interval.*