Statistical Analysis of ECG Pattern Variations Based on Age and Gender

Mohammad Nazmul Alam[1], Dr. Vijay Laxmi[2], Dr. S. Quadir Patwari[3], Engr. Mohammad Shahriar Khan[4], Md Nannu Mian[5]

[1]Faculty of Engineering and Technology, Guru Kashi University, Talwandi Sabo, Bathinda, India
[2]Faculty of Computing, Guru Kashi University, Talwandi Sabo, Bathinda, India
[3]Public Health Specialist and Adjunct Faculty, Department of Pharmacy, Dhaka International University
[4]Founder, ICT Olympiad Bangladesh
[5]Aorora Legal Consulting Centre (ALCC), Dhaka, Bangladesh

***Abstract*—** This study analyzes how ECG features vary across different age groups and between genders using statistical techniques. ANOVA was used to find out whether differences in ECG features like P wave, QRS duration, and QT duration are statistically significant. While some features remained mostly stable, others like T wave repolarization (T_REPOL) and P wave polarity (P_POL) showed large variations, especially across age groups. These findings were confirmed using Tukey's HSD test, which identified exactly which age groups differed. Box plots clearly showed that age affects several important ECG parameters such as heart rate (HR), PR interval, QTcBz, and T_REPOL. Younger individuals showed more variation in ECG values, likely due to ongoing growth and development. This highlights the importance of considering age when building machine learning models for ECG analysis. Additionally, a two-sample t-test showed that some ECG features like QT, QRS, and P wave are significantly different between males and females, pointing to biological differences in heart activity. Overall, the results confirm that ECG features change with both age and gender, and these factors should be included in any accurate ECG classification system.

***Keywords*—ECG pattern, Statistical analysis, ANOVA, Tukey's HSD**

I. Introduction

Electrocardiography (ECG) is a fundamental diagnostic tool for assessing cardiac activity and identifying abnormalities. Understanding how ECG features vary across different demographics, such as age and gender, is crucial for improving diagnostic accuracy and developing age-specific or gender-specific medical interventions. This study aims to statistically analyze ECG features across age groups and genders to identify significant differences. By employing ANOVA, Tukey's HSD for post-hoc comparison in box plots after ANOVA test for age group, and two-sample t-tests for gender, this research provides insights into ECG variations that can enhance clinical decision-making and support AI-driven diagnostic models.

ECG is a widely used medical test that records the electrical activity of the heart [1]. During each heartbeat, the depolarization and repolarization of the atrial and ventricular chambers generate electrical impulses, which are detected by electrodes placed on the body's surface. These signals are sampled over time to produce a voltage-versus-time plot, representing the ECG waveform. Variations in ECG parameters, such as the PR interval and QRS complex, enable electrophysiologists to identify electrical abnormalities and diagnose cardiovascular diseases (CVDs). In contrast, imaging-based techniques like echocardiography (ultrasound), computed tomography (CT), and magnetic resonance imaging (MRI) provide a visual assessment of heart morphology and hemodynamics, aiding in CVD prediction. However, these methods require specialized healthcare facilities and trained professionals, making them less patient-friendly due to factors like high costs, radiation exposure, contrast agent injections, and lengthy acquisition times.

ECG remains a preferred heart monitoring tool due to its non-invasive nature, affordability, patient safety, and widespread availability [2]. It is commonly used in small clinics, outpatient departments, and even integrated into wearable devices, making it an accessible and efficient option for continuous heart monitoring.

A. ECG Overview

ECG records the electrical activity of the heart, which results from the depolarization and repolarization of the atrial and ventricular chambers.

These electrical signals fluctuate based on the contraction and relaxation of the heart muscles during each heartbeat and are represented as a time-voltage plot.

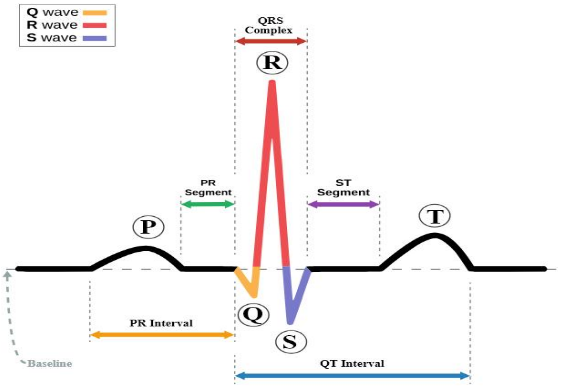

Fig. 1. Visual description of the different fundamental components of a single ECG heartbeat

Figure 1 illustrates an ECG signal corresponding to a single heartbeat. This signal consists of three key components that are analyzed by electrophysiologists to assess cardiac health. The first wave, known as the P-wave, represents atrial depolarization, which triggers the contraction of the atria and pushes blood into the ventricles. The most prominent feature of the ECG, the QRS complex, signifies the depolarization of the ventricles, leading to their contraction. The final elevation in the signal, the T-wave, corresponds to the repolarization of the ventricles, indicating their relaxation.

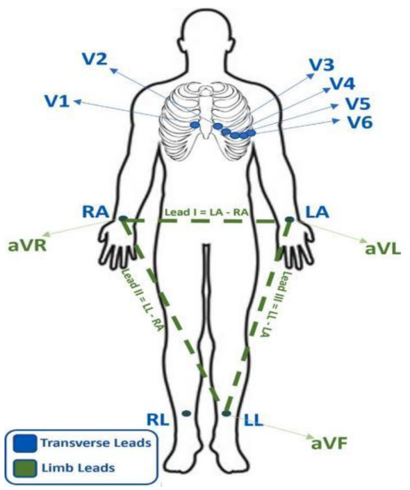

Fig. 2. Placement of ten different electrodes for measurement of 12-lead ECG signal [1].

The duration and amplitude of these waves, such as the peak of the R-wave, help cardiologists establish reference values for different age groups, distinguishing between healthy individuals and those with potential heart conditions. For example, the PR interval, which measures the time between atrial and ventricular depolarization, can indicate a pre-excitation syndrome if it is shorter than normal. Similarly, the QT interval represents the duration of ventricular electrical activity, and abnormalities in this interval may be linked to conditions such as myocardial infarction or ischemia. Additionally, ST-segment variations, including elevation or depression (such as an inverted T-wave), are often observed in patients with cardiac diseases.

A standard ECG recording consists of twelve leads, obtained using ten electrodes placed on specific locations of the body. Each lead captures the heart's electrical activity from a unique perspective, providing valuable insights for detecting abnormalities and diagnosing cardiac conditions. Among these, six leads (V1–V6) record electrical signals in the horizontal plane and are referred to as transverse leads. These leads utilize six unipolar chest electrodes. The remaining six leads capture electrical activity in the vertical plane and are classified as frontal or limb leads. This group includes three bipolar leads (I, II, and III), which measure signals concerning a negative reference electrode, and three unipolar leads (aVL, aVR, and aVF). Figure 2 visually represents the placement of transverse and frontal leads on the human body. Collectively, the twelve-lead ECG provides a comprehensive assessment of the heart's electrical function, making it a valuable tool for machine learning and deep learning models in detecting cardiac abnormalities [3-5].

B. Contributions

This research makes the following key contributions:

- Comprehensive Statistical Analysis: We conduct a detailed hypothesis testing framework using ANOVA and t-tests to determine significant differences in ECG features across age groups and genders.
- Age-Based Variability Identification: The study highlights significant variations in ECG parameters among different age groups, confirming age-related cardiovascular changes.
- Gender-Based Comparison: The analysis demonstrates that there are no statistically significant ECG differences between male and female participants.
- Data-Driven Insights for Clinical Applications: The results provide foundational knowledge for refining diagnostic models and developing personalized cardiovascular health interventions.

C. Organization

The remainder of the paper is structured as follows: The second section materials and methods present the data collection, preprocessing and statistical techniques. The third section analyze the results. Finally, the fourth section concludes with a summary.

II. Materials and Methods

A. Data Sources

For this study, after obtaining approval from the authority, we collected the ECG reports of Bangladeshi patients who were analyzed to determine the nature of their heart disease at the Khawaja Badrudduja Modern Hospital in Gazipur, Bangladesh. Labels (y) are assigned to instances (x) in the dataset (D) where D = $\{(x_i, y_i)\}^{n}_{i=1}$, where n is the number of samples. After gathering the raw ECG signals, it is processed into a structured format and we converted them into a CSV file.

B. ECG Signal

We used standard 12 -lead ECG traces sampled at 0.67-100 H_z, 25 mm/s, 10mm/mV, and 2*5.0s SE-1200 Express machine (Figure 3). The study involved 2840 subjects, 39,760 samples, and 67 classes, with the data stored in image format. The year of the study was 2024. A sample of the ECG report is shown in Figure 4. The features included in it are gender, age, hr, p, pr, qrs, qt, qtcbz, p_pol, qrs_depol, t_repol, rv5, and sv1. These features' information is displayed in table I.

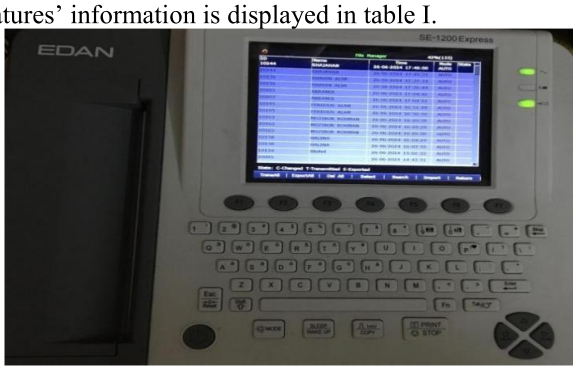

Fig. 3. SE-1200 Express ECG machine was used to produce the report.

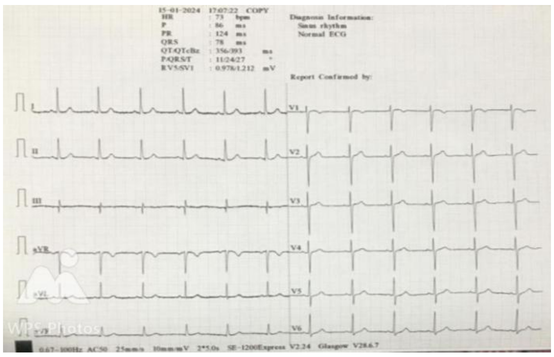

Fig. 4. Sample of 12-lead ECG signal collected from hospital.

C. Feature Extraction

Clinically significant features such as the P wave, QRS complex, T wave, heart rate, PR interval, and QT interval are extracted from ECG report. Figure 5 shows the extracted features and figure 6 shows the total no. of each features extracted from ECG datasets. Three types of features are used for this research which are described in the following sub sections.

1) Morphological Features: These features describe the shape and amplitude of ECG waveforms, providing insights into structural abnormalities [6-8]:

- P (Amplitude/Duration) **-** Related to atrial depolarization.
- PR (Interval) **-** Reflects the conduction time from atria to ventricles.
- QRS (Duration/Amplitude) **-** Represents ventricular depolarization.
- QT (Interval) **-** Total time of ventricular depolarization and repolarization.
- QTcbz - Corrected QT interval for heart rate.
- P_pol (Polarity) **-** Characteristics of the P wave.
- QRS_depol (Polarity/Amplitude) **-** Features of ventricular depolarization.
- T_repol (Amplitude) - Represents ventricular repolarization.
- RV5 (Amplitude) **-** Amplitude of the R-wave in lead V5.
- SV1 (Amplitude) **-** Amplitude of the S-wave in lead V1.

*2) Temporal Features***:** These features describe the timing aspects of the ECG signal and are essential for understanding the rhythm and conduction [9, 10]:

- PR (Time Interval) - Time from the onset of P wave to the start of QRS complex.
- QT (Time Interval) - Represents the total ventricular activity duration.
- QRS (Duration) - Time taken for ventricular depolarization.

3) Demographic Features:

- Gender
- Age

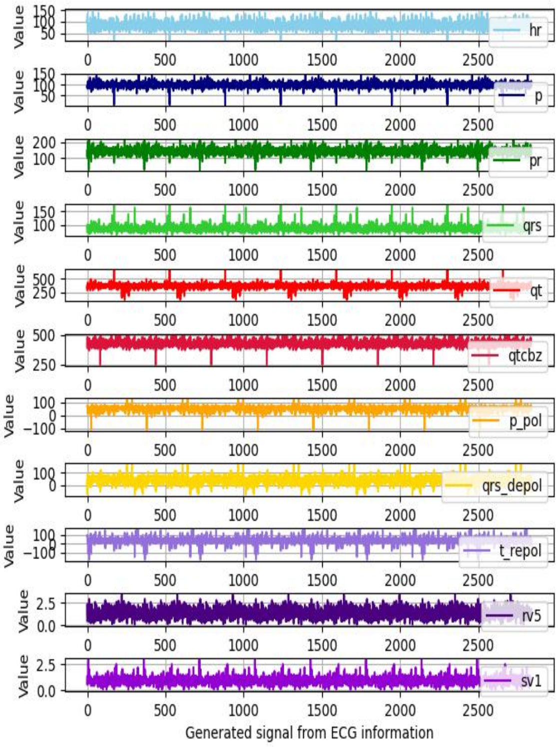

Fig. 5. Feature extraction after preprocessing.

TABLE I. ECG FEATURE'S DESCRIPTION

Feature	Mean	Std Dev	Distinct	Min	Max
Age	40.608	15.035	61	10	90
hr	81.893	16.334	74	21	150
p	91.887	12.059	36	8	144
pr	142.079	21.161	82	28	214
qrs	87.324	11.891	32	64	172
QT	366.016	45.613	84	144	664
QTCbz	425.552	25.83	102	246	500
p-pol	51.882	20.276	73	-110	117
qrs-depol	40.665	31.782	111	-77	160
t-repol	37.51	34.347	105	-178	154
TV5	1.333	0.583	311	0.089	3.414
SV1	0.919	0.374	277	0.051	2.927

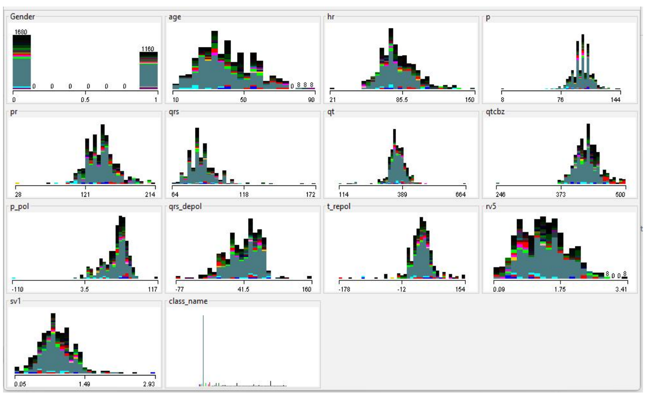

Fig. 6. Total features and classes extracted from ECG datasets.

TABLE II. DETAILS OF USED FEATURES

Features	Normal Range	Description
Age	0-120	Years since birth
Heart Rate (HR)	60-100 BPM	Beats per minute of the heart
P Wave Duration (P)	80-120 ms	Duration of atrial depolarization
PR Interval (PR)	120-200 ms	Time for electrical impulse to travel from atria to ventricles
QRS Duration (QRS)	70-100 ms	Duration of ventricular depolarization
QT Interval (QT)	Varies, typically <440 ms (men), <460 ms (women)	Total time for ventricular depolarization and repolarization
Corrected QT Interval (QTcBz)	<450 ms	Corrected QT interval supported on Bazett's formula
T Wave Repolarization (T_repol)	<440 ms	Duration of ventricular repolarization
RV5	0.5-1.5 mV	R wave in an ECG's lead V5
SV1	0.5-1.5 mV	S wave in an ECG's lead V1

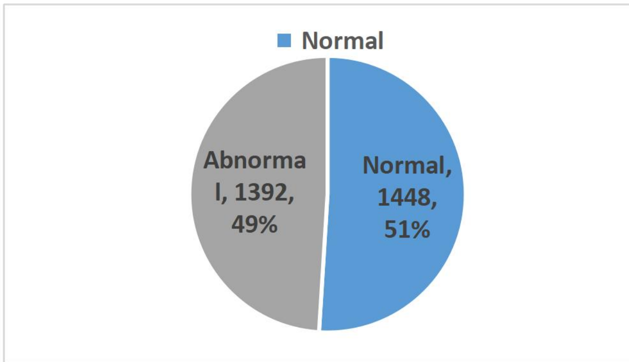

Fig. 7. ECG class distribution

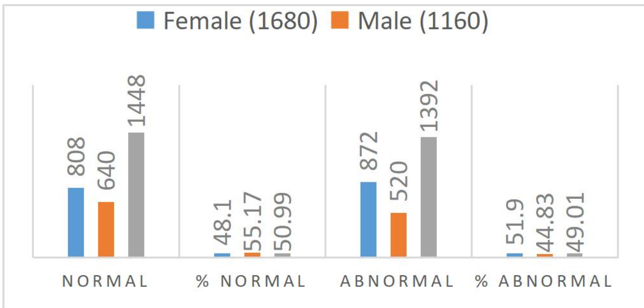

Fig. 8. ECG class distribution

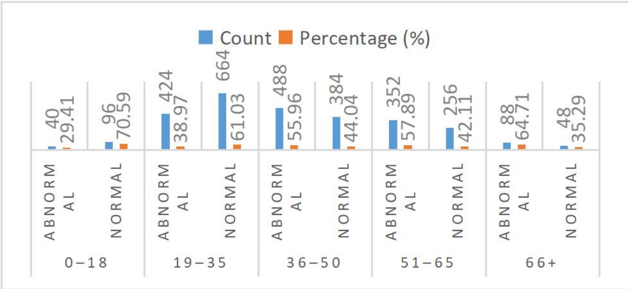

Fig. 9. ECG class distribution among age groups

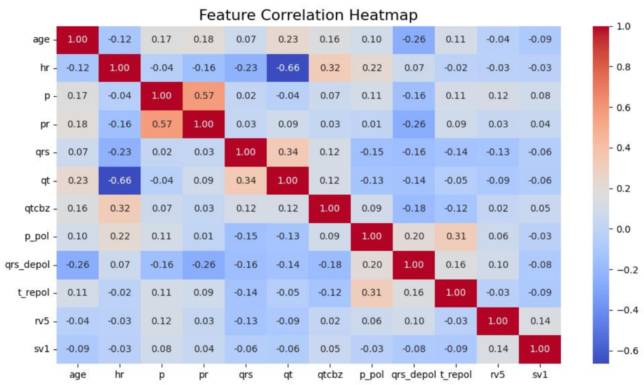

Fig. 10. ECG feature's correlation heatmap

D. Gender-Wise Features Correlation

Correlation is a statistical measure that expresses the degree to which two variables move in relation to each other. It ranges from -1 to 1:

- Positive Correlation (+1 to 0, in red): As one variable increases, the other also increases.
- Negative Correlation (0 to -1, in blue): As one variable increases, the other decreases.
- Zero Correlation (0, in neutral color): No relationship between the variables.

The heatmaps use a color gradient to represent correlation values between ECG features, ranging from deep blue (strong negative correlation, ≤ -0.6) to deep red (strong positive correlation, ≥ 0.6), with neutral correlations appearing in white or light shades [11]. In both male and female heatmaps, strong correlations (positive or negative) highlight significant relationships between ECG parameters, such as HR and QT interval shows negative correlation (Males: -0.78, Females: -0.62) and QRS duration with QT interval (positive correlation). The P wave (P) and PR interval (PR) show a moderate to strong positive correlation (Males: 0.56, Females: 0.58). QT and RV5 shows neutral correlation but slightly differ in male and female case. These insights help in analyzing gender-based variations in ECG characteristics.

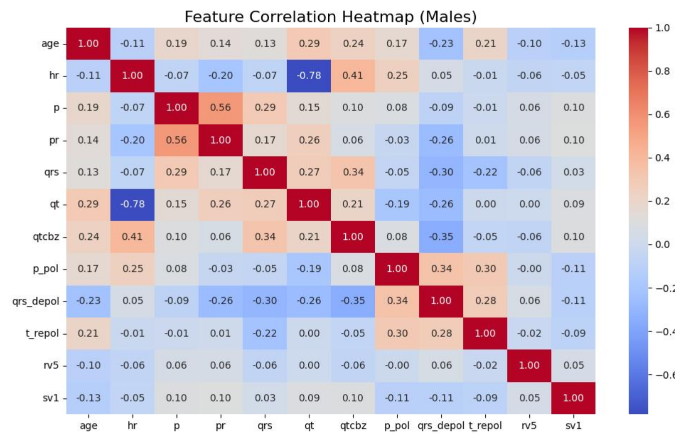

Fig. 11. ECG feature's correlation heatmap (males)

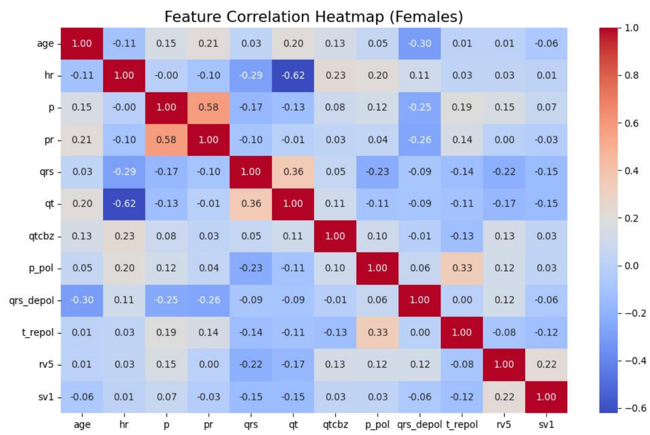

Fig. 12. ECG feature's correlation heatmap (female)

TABLE VII. DESCRIPTIVE STATISTICS BY AGE GROUP

Age Group	0-18	19-35	36-50	51-65	66+
P Mean	90.7058 82	98.4411 76	101.376 147	103.263 158	96.0000 00
P Std	12.7506 7	9.98897	10.7293 5	15.4701 1	9.51568
QRS Mean	84.2352 9	86.7058 8	87.5045 9	89.2105 3	85.7573 5
QRS Std	10.5468	9.2347	9.8325	11.5612	9.7655
QT Mean	330.529 41	333.411 76	336.155 96	338.473 68	331.084 91
QT Std	12.4567	14.3489	13.9786	15.2467	11.8942
QTCBZ Mean	340.764 71	342.411 76	344.725 69	347.052 63	341.330 19

QTCBZ Std	13.2948 2	15.0821 3	14.5673 2	16.1298 4	12.6724 9
P_Pol Mean	0.45176 5	0.48611 7	0.50423 5	0.52919 4	0.47561 0
P_Pol Std	0.05274 3	0.04875 6	0.05043 1	0.05386 1	0.04732 9
QRS_Depol Mean	0.34129 4	0.37094 1	0.39582 3	0.41876 5	0.35984 7
QRS_Depol Std	0.06583 2	0.05942 8	0.06218 3	0.06849 3	0.05481 2
T_Repol Mean	0.54397 1	0.56387 2	0.58126 4	0.59930 2	0.55294 7
T_Repol Std	0.06721 3	0.06102 9	0.06351 8	0.07029 4	0.05978 3

RV5 Mean	0.890353	0.962684	0.920541	0.867101	0.820706
RV5 Std	0.41662	0.51200	0.57844	0.68637	0.71261
SV1 Mean	0.361931	0.323883	0.360804	0.465568	0.340223
SV1 Std	0.25649	0.28647	0.29275	0.30546	0.25437

TABLE VIII. THE AGE GROUP: 0-18

Metric	Count	Mean	Std	Min	25%	50%	75%	Max
Age	136	14.88	2.69	10.00	14.00	15.00	17.00	18.00
HR	136	93.18	21.61	70.00	76.00	92.00	102.00	150.00
P	136	90.71	12.75	60.00	86.00	88.00	96.00	122.00
PR	136	129.06	14.75	104.00	124.00	128.00	140.00	152.00
QRS	136	84.24	10.21	68.00	78.00	84.00	88.00	108.00
QT	136	340.59	35.42	252.00	334.00	342.00	362.00	400.00
QTcbz	136	419.00	27.71	372.00	394.00	426.00	441.00	462.00
P_pol	136	57.59	17.25	21.00	53.00	60.00	72.00	79.00
QRS_depol	136	62.47	20.06	26.00	48.00	66.00	78.00	103.00
T_repol	136	37.53	16.75	12.00	25.00	43.00	46.00	81.00
RV5	136	1.53	0.42	0.69	1.32	1.42	1.60	2.51
SV1	136	0.89	0.36	0.40	0.64	0.80	1.17	1.46

TABLE IX. THE AGE GROUP: 19-35

Metric	Count	Mean	Std	Min	25%	50%	75%	Max
Age	1088	28.50	4.70	19.00	25.00	30.00	32.00	35.00
HR	1088	81.77	14.30	53.00	71.00	80.00	91.25	137.00
P	1088	98.44	9.99	66.00	94.00	98.00	104.00	132.00
PR	1088	138.55	17.27	80.00	126.00	140.00	150.00	184.00
QRS	1088	86.71	9.27	68.00	80.00	86.00	92.00	114.00
QT	1088	357.13	47.51	114.00	342.00	362.00	384.00	460.00
QTcbz	1088	421.65	20.13	378.00	407.75	420.50	437.00	477.00
P_pol	1088	48.99	22.78	-110.00	39.75	56.00	62.00	77.00
QRS_depol	1088	46.99	26.95	-56.00	26.00	54.00	69.00	87.00
T_repol	1088	34.85	21.94	-113.00	24.75	36.00	49.00	95.00
RV5	1088	1.31	0.51	0.37	0.89	1.28	1.75	2.27
SV1	1088	0.96	0.32	0.27	0.78	0.93	1.15	2.13

TABLE X. THE AGE GROUP: 36-50

Metric	Count	Mean	Std	Min	25%	50%	75%	Max
Age	872	42.57	4.70	36.00	40.00	41.00	45.00	50.00
HR	872	81.90	17.00	50.00	71.00	79.00	90.00	133.00
P	872	101.38	10.73	76.00	96.00	102.00	108.00	144.00
PR	872	144.13	18.78	106.00	132.00	144.00	154.00	208.00
QRS	872	87.50	11.83	64.00	82.00	86.00	90.00	138.00
QT	872	371.30	33.95	288.00	350.00	366.00	394.00	452.00
QTcbz	872	428.30	24.24	368.00	413.00	429.00	440.00	500.00
P_pol	872	51.39	16.46	-4.00	45.00	55.00	63.00	80.00
QRS_depol	872	38.03	27.11	-25.00	19.00	40.00	57.00	112.00
T_repol	872	35.61	32.06	-52.00	19.00	38.00	50.00	148.00
RV5	872	1.37	0.58	0.14	0.96	1.33	1.75	2.93
SV1	872	0.92	0.36	0.05	0.73	0.90	1.12	2.50

TABLE XI. THE AGE GROUP: 51-65

Metric	Count	Mean	Std	Min	25%	50%	75%	Max
Age	608	57.79	4.21	51.00	55.00	56.50	60.00	65.00
HR	608	79.14	15.41	21.00	70.75	77.00	87.50	116.00
P	608	103.26	15.47	8.00	97.50	104.00	110.50	140.00
PR	608	148.99	27.38	28.00	136.00	148.00	164.50	214.00
QRS	608	89.21	15.65	66.00	81.50	86.00	92.00	172.00
QT	608	376.29	52.79	154.00	352.00	374.00	392.00	664.00
QTcbz	608	426.00	33.31	246.00	410.75	424.00	438.50	496.00
P_pol	608	54.62	21.37	0.00	44.50	58.00	67.00	117.00
QRS_depol	608	29.51	41.26	-77.00	0.75	28.00	54.50	160.00
T_repol	608	40.58	53.20	-178.00	29.00	44.50	60.75	154.00
RV5	608	1.27	0.69	0.09	0.74	1.21	1.60	3.41
SV1	608	0.87	0.47	0.07	0.60	0.80	1.10	2.93

TABLE XII. THE AGE GROUP: 66+

Metric	Count	Mean	Std	Min	25%	50%	75%	Max
Age	136	73.82	5.70	69.00	70.00	72.00	75.00	90.00
HR	136	83.8	20.	50.0	69.0	76.0	99.0	123.

		2	00	0	0	0	0	00
P	136	96.00	9.52	80.00	90.00	94.00	106.00	118.00
PR	136	139.29	25.14	100.00	122.00	130.00	162.00	190.00
QRS	136	85.76	11.85	70.00	78.00	84.00	88.00	112.00
QT	136	382.71	46.20	292.00	348.00	390.00	412.00	492.00
QTcbz	136	443.65	26.00	394.00	426.00	442.00	459.00	494.00
P_pol	136	60.29	12.71	30.00	59.00	65.00	67.00	77.00
QRS_depol	136	35.00	33.98	-52.00	19.00	50.00	57.00	75.00
T_repol	136	57.29	27.90	1.00	45.00	62.00	70.00	112.00
RV5	136	1.42	0.71	0.23	0.91	1.59	2.03	2.60
SV1	136	0.82	0.34	0.06	0.55	0.81	1.03	1.42

E. Notable Observations from Different Age Group

- Age 0-18 Group: Has the highest heart rate (HR) mean (93.18 bpm) with greater variability. QTcbz values are also relatively high (419 ms), suggesting faster repolarization.
- Age 19-35 Group: Demonstrates relatively stable ECG characteristics with moderate HR (81.77 bpm) and depolarization values, indicating optimal heart function.
- Age 36-50 Group: Shows increased P and QRS durations with a more pronounced QT interval (371.3 ms), indicating prolonged ventricular repolarization.
- Age 51-65 Group: Features the highest P wave and QRS complex durations, with slightly prolonged QT intervals, reflecting age-related cardiac adaptations.
- Age 66+ Group: Depicts a decline in most parameters, including P, QRS, and polarization values, indicating age-related reductions in cardiac electrical activity.

F. ANOVA and Tukey's HSD Testing

For the statistical analysis of the datasets hypothesis testing was conducted based on age groups and gender. The analysis involved performing ANOVA to determine whether significant differences exist in ECG features across different age groups. Following the ANOVA, a Post-Hoc Analysis using Tukey's HSD (Honestly Significant Difference) test was applied to identify which specific age groups differed significantly from one another. Tukey's HSD compares all possible pairs of groups, providing details on the differences in their means, significance levels, and confidence intervals. Additionally, boxplots were created for each ECG feature, which allowed for visual observation of the distribution of the data across different age groups. The same steps were repeated for gender to investigate whether there were significant differences in ECG features between male and female participants and the two-sample t-test compares the means of a feature between two groups (e.g., males and females) to determine if there is a statistically significant difference. The following hypotheses were established to conduct hypothesis testing for the remainder of the analysis.

- For Age Groups:
- Null Hypothesis (H_0): There are no significant differences in the ECG features across different age groups.
- Alternative Hypothesis (H_1): There are significant differences in the ECG features across different age groups.
- For Gender:
- Null Hypothesis (H_0): There are no significant differences in the ECG features between male and female participants.
- Alternative Hypothesis (H_1): There are significant differences in the ECG features between male and female participants.

G. ANOVA Test for Age Groups

The table XIV (A-K) summarizes the results of ANOVA conducted to assess whether there are statistically significant differences in ECG feature values across multiple groups (age groups). The F-value measures the ratio of variance between groups to the variance within groups—higher values suggest more distinct group means. The p-value indicates the probability that the observed differences occurred by chance. Features are statistically highly significant ($p < 0.001$), meaning their variation across groups is unlikely to be due to random chance. A high F-value and a p-value below 0.05 indicate significant variation, leading to rejection of the null hypothesis. Therefore, there are significant differences in the ECG features across different age groups

TABLE XIII (A-K). ANOVA RESULTS FOR ALL FEATURES ON AGE GROUPS

A. HR (Heart Rate)

Source	**df**	**Sum of Squares**	**Mean Square**	**F-value**	**p-value**
Between Groups	4.0	23,705.36	5,926.34	22.8992	0.0000
Within Groups	2835.0	733,702.88	258.80	—	—
Total	2839.0	757,408.24	—	—	—
F-critical (0.05): 2.3751 → **Reject H_0**					

B. P Wave

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	4.99	1.25	5.0936	0.0004
Within Groups	2835.0	699.45	0.25	—	—
Total	2839.0	704.44	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

C. PR Interval

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	2,007.45	501.86	7.1523	0.0000
Within Groups	2835.0	198,874.64	70.15	—	—
Total	2839.0	200,882.10	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

D. QRS

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	1,000.90	250.23	12.2388	0.0000
Within Groups	2835.0	57,995.75	20.46	—	—
Total	2839.0	58,996.65	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

E. QT

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	15,563.74	3,890.94	13.4931	0.0000
Within Groups	2835.0	817,551.91	288.42	—	—
Total	2839.0	833,115.65	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

F. QTCBZ

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	29,064.01	7,266.00	23.2917	0.0000
Within Groups	2835.0	884,877.85	312.20	—	—
Total	2839.0	913,941.87	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

G. P Polarity

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	0.96	0.24	3.7824	0.0043
Within Groups	2835.0	181.88	0.06	—	—
Total	2839.0	182.83	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

H. QRS Depolarization

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	80.59	20.15	0.9559	0.4295
Within Groups	2835.0	59,709.35	21.06	—	—
Total	2839.0	59,789.94	—	—	—
F-critical (0.05): 2.3751 → **Fail to reject H₀**					

I. T Repolarization

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	2,489.77	622.44	3.7784	0.0043
Within Groups	2835.0	466,918.99	164.71	—	—
Total	2839.0	469,408.76	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

J. RV5

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	53,430.85	13,357.71	60.8783	0.0000
Within Groups	2835.0	622,179.02	219.52	—	—
Total	2839.0	675,609.87	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

K. SV1

Source	df	Sum of Squares	Mean Square	F-value	p-value
Between Groups	4.0	4,057.59	1,014.40	3.4002	0.0088
Within Groups	2835.0	845,225.38	298.01	—	—

Total	2839.0	849,282.97	—	—	—
F-critical (0.05): 2.3751 → **Reject H₀**					

H. Observations of Age Group by ANOVA Test

Null hypothesis is rejected of all the features except QRS Depolarization as the F-value is significant than F-critical value.

I. Tukey's HSD Test

Tukey's HSD test helps identify which specific age groups have statistically significant differences in ECG features, after ANOVA finds that there is at least some difference among groups. Tukey's HSD was conducted for further justification of ANOVA test. However, we have found that the test results were also rejected the null hypothesis. In the following figure (figure 13) the boxplots for all features across age groups have been displayed to visualize the distribution. These plots highlight the differences in the feature values among the age groups. Tukey HSD test is being used for multiple group comparisons. FWER (Family-Wise Error Rate) ensures that the probability of making one or more Type I errors (false positives) across the multiple comparisons is controlled at 0.05 (5%). Tukey's each box plot shows the median (middle line in the box), the interquartile range (IQR) (box = 25th to 75th percentile), whiskers (range within 1.5 × IQR), and Outliers (dots outside the whiskers). The HSD suggests that there is a difference in the ECG features across different age groups. As a result, the alternative hypothesis is established and the null hypothesis is rejected.

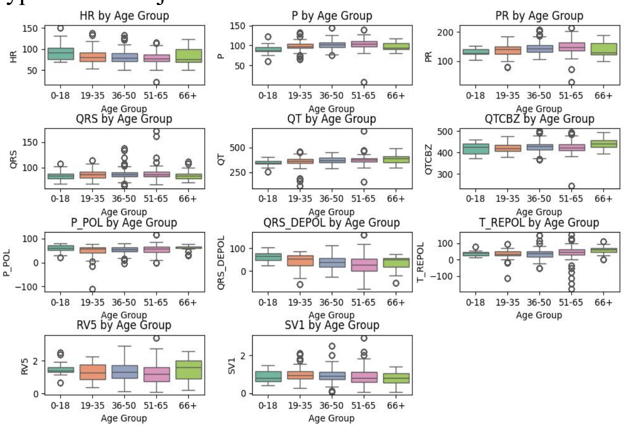

Fig. 13. Tukey's HSD test for feature values by age in box plot

Figure 13 presents box plots based on Tukey's HSD test, illustrating the distribution of various ECG features across five distinct age groups: 0–18, 19–35, 36–50, 51–65, and 66+. Each box plot corresponds to a specific ECG feature—namely P, QRS, QT, QTcBz, P_Pol, QRS_Depol, T_Repol, RV5, and SV1—with the y-axis representing the respective feature values. This visualization highlights age-related variations and helps identify statistically significant differences between groups.

1. HR by Age Group (Heart Rate)
   - Heart rate tends to slightly decline with increasing age.
   - The 0–18 group shows more variation (wider spread).
2. P by Age Group: P wave amplitudes remain relatively stable, but slightly higher variability is seen in younger groups.
3. PR by Age Group: PR interval increases with age, especially in the 66+ group, indicating possible age-related conduction slowing.
4. QRS by Age Group
   - QRS duration remains relatively stable across age groups.
   - Slightly higher values and variability in older adults.
5. QT by Age Group
   - QT interval increases gradually with age.
   - Older groups (51-65, 66+) show a higher median and more spread.
6. QTCBZ by Age Group (QT corrected using Bazett's formula)
   - QTc also increases with age, more noticeably in the elderly.
   - High variability in 66+ group suggests more cardiac electrical variability.
7. P_Pol by Age Group (P wave polarity)
   - Slight decrease in P wave polarity with age.
   - 66+ shows lower values with narrow spread.
8. QRS_Depol by Age Group (QRS Depolarization): Median remains stable, but younger age groups show more variation.
9. T_Repol by Age Group (T wave repolarization): Values are fairly stable but slightly decrease in older age groups.
10. RV5 by Age Group
    - RV5 amplitude (right ventricular voltage) drops with age.
    - A notable decline starts from the 36–50 age group onward.
11. SV1 by Age Group
    - SV1 (S wave in lead V1) shows a declining trend with age.
    - Highest in the youngest group, and lowers consistently with age.

- Older individuals (66+) show elevated QT/QTc and PR intervals, suggesting potential risk for age-related cardiac issues.
- This visualization can help clinicians or researchers understand how electrical heart activity evolves with age.

J. Two-Sample t-Test Results for Male and Female

The table XV presents the results of a two-sample t-test conducted to determine whether there are statistically significant differences in various ECG feature means between two independent groups (male vs. female). The t-statistic indicates the size and direction of the difference, while the p-value shows the likelihood that the observed difference occurred by chance.

TABLE XIV. SIGNIFICANCE OF ECG FEATURES BASED ON TWO-SAMPLE T-TEST BETWEEN GENDER GROUPS

Feature	T-Statistic	P-Value	Significance
HR	-9.3344	0.0000	***
P	5.0372	0.0000	***
PR	4.3033	0.000	***
QRS	11.1036	0.0000	***
QT	5.5220	0.0000	***
QTCBZ	-10.6258	0.0000	***
P_Pol	0.8792	0.3794	**ns**
QRS_Depol	3.0595	0.0022	**
T_Repol	7.6140	0.0000	***
RV5	4.5944	0.0000	***
SV1	3.6150	0.0003	**

Significance Codes:

- *** : $p < 0.001$
- ** : $0.001 \leq p < 0.01$
- * : $0.01 \leq p < 0.05$
- . : $0.05 \leq p < 0.1$
- ns : $p \geq 0.1$ (not significant)

K. Observations of Gender by Two Sample t-Test

Based on the positive and negative value of T-statistic (positive if mean (male) > mean(female), negative if mean (male) < mean(female), the two-sample t-test results reveal significant differences in various ECG features between male and female groups. Specifically, males exhibit higher mean values for the P wave, PR interval, QRS duration, QT interval, QRS depolarization, T repolarization, RV5, and SV1, as indicated by the positive t-statistics and highly significant p-values ($p < 0.01$ or $p < 0.001$). Conversely, the HR and corrected QT interval (QTcbz) has a negative t-statistic, suggesting that females have a significantly higher average value for this feature. The P_pol feature does not show a statistically significant difference between genders, as its p-value is well above the 0.05 threshold. Overall, these findings highlight that gender-specific variations are present in most ECG features, which may be important for personalized diagnosis or modeling in medical applications.

III. RESULTS AND DISCUSSION

ANOVA was conducted to determine whether the differences observed among age groups are statistically significant. In table XIII (A-K) shows that features such as P wave duration, QRS wave duration, and QT duration exhibit relatively stable distributions, with minor variations across different age groups. However, certain features, such as T-repol wave duration and P-pol wave duration, show substantial fluctuations, indicating possible significant differences that warrant further statistical validation.

The presence of outliers in specific age groups suggests that some individuals may have ECG values deviating from the general population trend, potentially due to physiological or pathological reasons. For features where significant differences are found, Tukey's HSD post-hoc test are employed to identify specific group differences.

Figure 13 provides key insights into how age influences various ECG features, as revealed through box plot visualizations supported by Tukey's HSD test. The analysis clearly shows that age significantly impacts several critical ECG parameters, particularly heart rate (HR), PR interval, QTcBz, and T wave repolarization (T_REPOL). These age-related variations are highly relevant for developing age-aware machine learning models aimed at classifying normal versus abnormal ECG signals. The visual patterns observed in the box plots support the need for incorporating age as an important factor in ECG anomaly detection, especially when using structured data and deep learning techniques. Additionally, the plots reveal that younger individuals in the 0–18 age group tend to exhibit greater variability in several ECG parameters, suggesting a broader range of normal values during early physiological development.

Table XIV shows a significant difference in features such as QT, QRS, and P-Wave between males and females, as determined by the two-sample t-test. In this test, the null hypothesis is rejected, indicating that there are significant differences in some ECG feature values between males and females. This suggests that these ECG parameters vary between genders, reflecting potential physiological differences in cardiac activity.

Finally, we can conclude that ECG parameters do vary significantly based on age and gender.

IV. CONCLUSION

This study clearly shows that ECG parameters vary significantly with both age and gender. Through ANOVA testing and Tukey's HSD post-hoc analysis, we found that certain ECG features—especially heart rate (HR), PR interval, QTcBz, and T wave repolarization (T_REPOL)—are strongly influenced by age. The visual box plots further support this, showing noticeable patterns and variations, particularly among younger individuals (0–18 years), who display a wider range of normal values due to developmental changes. Additionally, gender-based analysis using a two-sample t-test revealed significant differences in features like QT, QRS, and P wave durations between males and females. This suggests underlying physiological differences in cardiac function between genders. The presence of outliers in some age groups highlights individual variability, which may be due to health conditions or unique physiological traits. These findings confirm the importance of considering both age and gender when analyzing ECG data. Such insights are valuable for building accurate and personalized ECG classification models, especially in machine learning and deep learning applications for health monitoring.

REFERENCES

[1] M. Y. Ansari, M. Qaraqe, F. Charafeddine, E. Serpedin, R. Righetti, and K. Qaraqe, "Estimating age and gender from electrocardiogram signals: A comprehensive review of the past decade," Artificial Intelligence in Medicine, vol. 146. Elsevier B.V., Dec. 01, 2023. doi: 10.1016/j.artmed.2023.102690.

[2] T. Anbalagan, M. K. Nath, D. Vijayalakshmi, and A. Anbalagan, "Analysis of various techniques for ECG signal in healthcare, past, present, and future," Biomedical Engineering Advances, vol. 6, p. 100089, Nov. 2023, doi: 10.1016/j.bea.2023.100089.

[3] I. Tobore, J. Li, A. Kandwal, L. Yuhang, Z. Nie, and L. Wang, "Statistical and spectral analysis of ECG signal towards achieving non-invasive blood glucose monitoring," BMC Medical Informatics and Decision Making, vol. 19, Dec. 2019, doi: 10.1186/s12911-019-0959-9.

[4] C. Maji, P. Sengupta, A. Batabyal, H. Chaudhuri, and W. Bengal India, "Nonlinear and statistical analysis of ECG signals from Arrhythmia affected cardiac system through the EMD process."

[5] P. Ahmadi et al., "Age and gender differences of basic electrocardiographic values and abnormalities in the general adult population; Tehran Cohort Study," BMC Cardiovascular Disorders, vol. 23, no. 1, Dec. 2023, doi: 10.1186/s12872-023-03339-z.

[6] R. Rathi, N. Yagnik, S. Tiwari, and C. Sharma, "Analysis of statistical models for fast time series ECG classifications," Engineering Letters, vol. 30, no. 2, pp. [page numbers], 2022.

[7] Z. S. Al-Falahi et al., "Advanced electrocardiography heart age: a prognostic, explainable machine learning approach applicable to sinus and non-sinus rhythms," European Heart Journal - Digital Health, vol. 6, no. 1, pp. 45–54, Jan. 2025, doi: 10.1093/ehjdh/ztae075.

[8] B. Surawicz and S. R. Parikh, "Prevalence of male and female patterns of early ventricular repolarization in the normal ECG of males and females from childhood to old age," Journal of the American College of Cardiology, vol. 40, no. 10, pp. 1870–1876, Nov. 2002, doi: 10.1016/S0735-1097(02)02492-0.

[9] G. S. Nayak, C. Puttamadappa, and A. S. Kamatha, "Statistical analysis of ECG signals for arrhythmia detection," in 4th Kuala Lumpur International Conference on Biomedical Engineering 2008: BIOMED 2008, Kuala Lumpur, Malaysia, Jun. 25–28, 2008, pp. 251–253. Springer, Berlin Heidelberg.

[10] H. Mieszczanska, G. Pietrasik, K. Piotrowicz, S. McNitt, A. J. Moss, and W. Zareba, "Gender-related differences in electrocardiographic parameters and their association with cardiac events in patients after myocardial infarction," The American Journal of Cardiology, vol. 101, no. 1, pp. 20–24, 2008.

[11] A. Afzalian, P. Ahmadi, A. Jalali, F. Masoudkabir, A. Oraii, M. L. Tokaldani, A. Shafiee, M. Mohammadi, E. Sanei, and M. Tajdini, "Age and gender differences of electrocardiographic basic values and abnormalities in general adult population; Tehran Cohort Study," Authorea Preprints, 2022.

[12] S. Bansal, M. Agarwal, K. Yadav, Y. Yadav, and R. Choudhary, "Study of effect of age and gender on ECG pattern in normal paediatric population of Rajasthan," Journal of Medical Science and Clinical Research (JMSCR), vol. 6, no. 7, pp. [page numbers], July 2018.