Comparing and Normalizing the Measurement of Step Counts and Heart Rates of Selected Wristbands and Smartwatches

Siam Yamsaengsung
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
siam@sit.kmutt.ac.th

Thanapong Khajontantichaikun
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
thanapong.khajon@mail.kmutt.ac.th

Bunthit Watanapa
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
bunthit@sit.kmutt.ac.th

Saichon Jaiyen
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
saichon.jai@sit.kmutt.ac.th

Vajirasak Vanijja
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
vachee@sit.kmutt.ac.th

Pornchai Mongkolnam
School of IT
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
pornchai@sit.kmutt.ac.th

Chujit Treerattanaphan
School of Architecture and Design
King Mongkut's University of Technology Thonburi
Bangkok, Thailand
chujit.jea@kmutt.ac.th

Pairoj Likitthanasate
National Telecom Public Company Limited
Bangkok, Thailand
pairojl@ntplc.co.th

2023 27th International Computer Science and Engineering Conference (ICSEC) | 979-8-3503-4210-9/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSEC59635.2023.10329731

Abstract— **The usefulness and popularity of smartwatches in the marketplace has led to a large variety of brands and models, each utilizing different technologies and processing methods to detect physical activities and measure vital metrics. Consequently, the reported values of widely available metrics of heart rate and step count can exhibit significant variations across these devices compared to standard medical equipment. This research proposes a model to standardize these values using a medical-grade device as a reference for heart rate and a manual clicker for step count. Descriptive statistics reveal that while the detected median values of step count align closely with actual steps, there is a notable oscillation in the overall data distribution. Fitbit devices consistently report lower step counts with high variability, while Garmin devices demonstrate more accurate step counts. Fitbit devices provide more precise heart rate measurements, while Huawei's measurements are less so. A linear regression model is used to effectively refine smartwatch measurements across multiple models, achieving a high level of accuracy as compared to a medical-grade sensor. Gender does not significantly impact the modeling adjustments.**

Keywords- Smartwatch, Data Mining, Digital Health, Step Counts, Heart Rates.

I. Introduction

In today's fast-paced world, maintaining good health has become a top priority for many individuals. Health-conscious individuals are increasingly turning to smartwatches and wristbands as essential tools towards good health. These devices offer a range of features and functionalities, allowing users to conveniently track their exercise and health data. With capabilities such as workout reminders and measuring blood pressure, heart rate (HR) monitoring, measuring blood oxygen level, step counting, and sleep analysis, smartwatches have revolutionized the way we monitor our well-being [1, 2]. As a result, these digital devices have gained immense popularity, particularly among health enthusiasts who wear them consistently. In a survey conducted in 2019 [3], users of wearable devices were asked about their reasons for choosing to use this technology. The survey results revealed that the highest proportion, accounting for 60% of respondents, used wearable technology primarily for step tracking purposes. Additionally, 44% of respondents cited heart rate detection as their primary reason for using wearable technology, while 42% reported using it for tracking calories burned. Another reason mentioned by 40% of respondents was to track sleep patterns. These findings provide insights into the key motivations behind the adoption of wearable health devices among users.

With rising demand for smartwatches, a multitude of brands and models have flooded the market, each offering a diverse range of functions at a wide range of prices. However, studies have revealed inconsistencies in heart rate measurements and step counts across different smartwatch brands [4, 5]. These variations can be attributed to variations in sensing technologies, measurement methods, and the accuracy of sensors utilized, as well as differences in data processing algorithms. Consequently, these values may not be suitable for direct comparison to the high accuracy standards of medical instruments.

Thus, the primary objective of this research is to perform pilot experiments to identify a model or approach that can be used to optimize the heart rate and step count measurements obtained from smartwatches and wristbands. To achieve this goal, we conducted an experiment involving various brands and models of smartwatches and wristbands. Specifically, we selected six models for this study: Fitbit Charge 5, Fitbit Luxe, Garmin Forerunner 55, Garmin Vivoactive 4, Huawei Band 7, and Huawei Fit 2. By subjecting the data from these devices to rigorous analysis, we aim to enhance the accuracy and reliability of their health metrics, bridging the gap between consumer wearables and medical-grade instruments.

979–8–3503–4210–9/23/$31.00 ©2023 IEEE

979-8-3503-4210-9/23/$31.00 ©2023 IEEE

With a model for improving the precision of heart rate and step count measurements across multiple smartwatches and wristbands, individuals will have access to more reliable health data that can be more accurately compared with medical reference values, enabling them to make better informed decisions about their well-being. The findings of this research will contribute to the ongoing advancements in wearable technology and further empower users to proactively monitor and manage their health more effectively.

II. RELATED WORKS

C. Meza et al. [6] and I. Tomohiko et al. [7] conducted a study aimed to validate the use of consumer smartwatches, specifically the Fitbit Charge 5 (FC5) and Apple Watch, in screening for atrial fibrillation (AF) among stroke patients. The results of these two works demonstrate that smart watches can reliably detect AF under controlled conditions.

C. Dobbins et al. [8] aimed to address obesity prevention using lifelogging techniques. They utilized two datasets and employed a smartwatch as a display device to present the results. The datasets consisted of physical activities categorized into three levels: light, moderate, and highly energetic, based on their metabolic equivalent values. The study involved data collection, processing, classification, and visualization. The approach demonstrated higher classification accuracy compared to four previous studies, as evaluated by ten different classifiers.

G. M. Weiss et al. [9] explored the potential of smartphones and smartwatches, with their built-in sensors, for implementing mobile motion-based behavioral biometrics. This study considered both accelerometer and gyroscope sensors on smartphones and smartwatches, determining the optimal combination for performance. They also evaluated eighteen diverse activities of daily living for their efficacy in biometric identification and authentication. The results indicated that motion-based biometrics using smartphones and smartwatches yielded good outcomes across the tested activities.

P. Düking et al. [10] evaluated the accuracy of heart rate monitoring and energy expenditure measurement using four popular smartwatches (Apple Watch Series 4, Polar Vantage V, Garmin Fenix 5, and Fitbit Versa). The results showed that the Apple Watch Series 4 exhibits the highest validity, but HR data provided by the Garmin Fenix 5 and Fitbit Versa should be interpreted with caution due to higher error rates at certain intensities.

M. Nissen et al. [11] studied the accuracy of heart rate recording in two smartwatches, namely the Fitbit Charge 4 and Samsung Galaxy Watch Active 2. The findings revealed that the accuracy of these devices varied depending on the activity being performed. Overall, both smartwatches achieved a mean absolute percentage error of less than 10%. However, it was noted that neither device demonstrated sufficient accuracy during seated rest or keyboard typing activities.

The studies conducted by K. M. Tam et al. [12], D. Jones et al. [13], and L. Wang et al. [14] also had similar goals and tests, which involved assessing the accuracy and validity of step count measurements in commercial activity monitor devices. While the smartwatch models used in each study may differ, the overall findings were consistent. The results of all three articles indicate that the smartwatches tested in their experiments provided a high level of accuracy comparable to that of a clicker counter. This suggests that the commercially available electronic activity monitor devices, including the tested smartwatches, can reliably measure step counts with precision similar to traditional clicker counters.

E. A. Thomson et al. [15] studied the Fitbit Charge HR 2 and Apple Watch. The devices were compared for their accuracy in measuring HR against electrocardiogram (ECG) recordings. Thirty young adults participated in the Bruce Protocol exercise test, during which HR measurements were recorded from the ECG and both devices every minute. The findings suggest that as exercise intensity increases, the accuracy of real-time HR monitoring by both the Apple Watch and Fitbit Charge HR 2 decreases.

M. P. Støve et al. [16] conducted a study aimed to verify the accuracy of smartwatches in measuring HR during rehabilitation and sporting activities. The study involved twenty-nine participants who underwent various physical conditions, including rest and three submaximal exercise conditions: cycling, treadmill walking, running, and rapid arm movement. The researchers noted that the specific exercise conditions may influence the discrepancy in the obtained HR values.

From these previous works it can be seen that the health measurements among smartwatches varies, so we performed a pilot study to develop a model to optimize the HR measurements and step counts among devices.

III. METHODOLOGY

A. Data Collection

The data collection process involves recruiting a sample group of 6 individuals with no known health risk, consisting of 4 females and 2 males, within the adult age range of 18 to 50 years. Each participant was required to wear two smartwatches on their wrist during physical activity for a duration of 13 weeks. The order in which the brands/models of the smartwatches are worn was randomly assigned for each participant. During the data collection period, participants engaged in light exercise to collect HR and step count measurements. The participants were asked to speed walk or jog 1-2 times per day, for an average of 5 days each week. The participants were guided to aim for around 1,500 steps during each exercise session. This target is not strictly enforced, and the actual recorded steps range from the minimum of 430 to the maximum of 2,400 steps. Heart rate measurements were obtained both before and after exercise activities. These measurements were collected using two devices: wristband or smartwatch and a medical-grade HR monitor. And to accurately monitor their walking activity, each participant was equipped with a manual clicker counter for recording the actual number of steps taken.

The HR measurement and step count data obtained from the smartwatches are compared with values acquired from the high-precision medical instruments and the clicker counter. To establish the reference HR values, each participant was equipped with a heart rate monitor (Fingertip pulse oximeter LK89) calibrated to medical standards to ensure accurate heart rate measurements. The participants were asked to log all the measured data from each activity in an online form.

B. Data Analysis

The study leveraged exploratory data analysis techniques to understand the characteristics inherent in the collected data.

To this end, statistical tools such as measures of central tendency, measures of dispersion, and measures of position were utilized, tailoring the analysis to the unique attributes of each smartwatch brand. Additionally, data visualization techniques, including line charts, histograms, and box plots, were employed to present the data visually, helping to provide a more insightful comprehension and helping the analysis of the data.

For data analysis of the step count, all experimental data undergoes a validation process, identification of missing or erroneous data, and addressing any anomalies. Following validation, the recorded step counts are normalized into a relative standard value of 1,500 steps, serving as a dummy reference for measuring the discrepancy between the smartwatches and the manual step counter. This normalization allows for a comprehensive assessment of the accuracy and precision of the smartwatches in step tracking across multiple data records compared to the standardized reference value.

$$\text{SWCount\_at\_1500} = (\text{Count\_by\_SW} / \text{Count\_by\_Counter}) * 1{,}500 \tag{1}$$

As an example for equation (1), if the number of steps read from smartwatch A is 1,800 steps and the number of steps obtained from the clicker counter is 1,700 steps, the conversion of this data is (1,800/1,700) * 1,500 = 1,588 steps, which makes all data referenced to a single value of 1,500.

C. Model Evaluation

Evaluation of model performance is essential in the development and enhancement of models to achieve optimal operational efficiency. In this research, the effectiveness of the employed models was assessed using the following evaluation criteria:

Mean Squared Error (MSE) represents the average of the squared differences between the estimated values produced by the model and the corresponding actual values. In this context, the actual values are the standard values obtained from the medical instrument. This serves as the reference for assessing the model's deviation. The equation for MSE is as follows:

$$MSE = \frac{1}{n}\sum_i (y_i - \hat{y}_i)^2 \tag{2}$$

y_i is a standard value from the medical instrument, $\hat{y}_i$ is an estimate from the model, and n is the amount of data.

Root Mean Squared Error (RMSE) is a value used to measure the error of the model by taking the square root of MSE as shown in the following equation:

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2} \tag{3}$$

Residual Standard Error (RSE) is the error associated with data in a linear regression model. It is a measure of the model's predictive accuracy. RSE is computed by summing the squared differences between the actual values and the corresponding estimated values produced by the model, dividing that sum by the degrees of freedom, and then taking the square root of the result. A lower RSE value indicates a higher level of accuracy in predicting data values. In the field of data analysis, the RSE value serves as a widely employed metric for evaluating the accuracy of estimations. The equation for calculating RSE is as follows:

$$RSE = \sqrt{\frac{1}{df}\sum_i (y_i - \hat{y}_i)^2} \tag{4}$$

y_i is a standard value from medical instrument, $\hat{y}_i$ is an estimate from the model, and df is degree of freedom.

R-squared (Coefficient of Determination) or R^2 is a metric that evaluates the degree of fit between a model and a specific dataset of interest. It provides a measure of the model's suitability for the given data. Ranging between 0 and 1, a value nearing 1 indicates a strong correspondence between the model and the data, indicating a high level of appropriateness. Conversely, as the value approaches 0, it suggests that the model is poorly suited for the data. The calculation of R-squared can be achieved using the following equation:

$$R^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \tag{5}$$

y_i is a standard value from medical instrument, $\hat{y}_i$ is an estimate from the model

Adjusted R-squared is an improved method of the R-squared that considers the number of samples and independent variables used in the analysis. It provides a more accurate measure of the model's goodness-of-fit by adjusting for the complexity of the model and the sample size. The calculation for Adjusted R-squared is as follows:

$$Adjusted\ R^2 = 1 - [(1 - R^2) \times (n - 1)/(n - k - 1)] \tag{6}$$

where n is the number of samples, and k represents the number of independent variables included in the model. The Adjusted R-squared value ranges from 0 to 1, with a higher value indicating a better fit of the model to the data. It is important to note that Adjusted R-squared tends to be lower than the R-squared since it accounts for the number of independent variables used in the analysis. This adjustment helps mitigate the risk of overfitting in the final model.

Analysis of Variance (ANOVA) or One-Way ANOVA was employed to test the hypothesis that the mean of the model-adjusted values among all smartwatches does not exhibit a significant difference at the specified confidence level or similar at the given confidence level. For example, at the confidence level of 95%. The statistic used in the test was $F = \frac{MSB}{MSE}$ which was calculated from the One-Way ANOVA Table as shown in Table I.

TABLE I. ONE-WAY ANOVA TABLE

Source of Variance	Sum of Squares (SS)	Degrees of Freedom (df)	Mean Squares (MS)	F
Between Groups (Treatment)	$SSB = \sum_{j=1}^{k} n_j(\bar{X}_j - \bar{X})^2$	$k - 1$	$MSB = \frac{SSB}{k-1}$	$F = \frac{MSB}{MSW}$
Within Groups (Error)	$SSW = \sum_{j=1}^{k}\sum_{i=1}^{n}(X_{ij} - \bar{X}_j)^2$	$n - k$	$MSW = \frac{SSW}{n-k}$	

Total	$SST = SSB + SSW$	$n - 1$		

IV. EXPERIMENT

In this section, we describe the data analysis procedures for each smartwatch model and brand. Subsequently, we employed linear regression modeling techniques aimed at attaining a comparable level of accuracy to that of medical devices. The experiment is divided into four main areas: step count data analysis, step count modeling, heart rate data analysis, and heart rate modeling.

A. Step Count Data Analysis

First, the datasets obtained from participants have normalized the steps to 1,500. Then we used the data set to show a box and whisker plot at [Q1 - 1.5IQR, Q3 + 1.5IQR] as shown in Figure 1. The dataset exhibits a high level of dispersion, with a reduction in the number of data points by 173, accounting for approximately 27% of the initial dataset. Upon conducting experiments using a z-score range of (-3, +3), similar results were obtained, indicating that the data is distributed relatively evenly compared to a normal distribution. Additionally, we decided to adjust the threshold values using the Box and Whisker plot approach. After applying this method, the threshold values for a good distribution characteristic and data preservation were determined to be [Q1 - 3IQR, Q3 + 3IQR]. This adjustment led to the elimination of 80 records, representing approximately 12% of the data, resulting in a final dataset of 569 records. The comparative distribution of the remaining data points is depicted in Figure 1, illustrating the improved distribution characteristics.

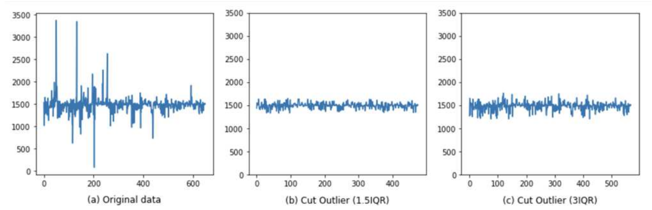

Figure 1. Comparison of the distribution of (a) initial data and (b), (c) data after cut-off outliers at 1.5 and 3 IQR.

We considered the datasets of all 3 brands of smartwatches (Fitbit, Garmin, and Huawei), the distribution of data is shown as box plots in Figure 2. The horizontal line in the Box Plot represents the median step count.

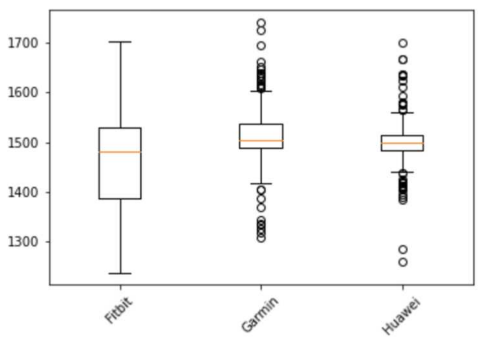

Figure 2. Box plot comparing the basic statistic characteristics of step count by smart watches from different brands.

Boxplot analysis reveals variations in step counts among different brands of devices compared to the reference value of 1500 steps. Fitbit shows a tendency towards lower step counts, with a median of 1,486 steps. Fitbit also exhibits a wider IQR compared to Garmin and Huawei. On the other hand, Garmin and Huawei demonstrate median values close to the reference value, with medians of 1,503 and 1,500 steps, respectively. Additionally, Garmin and Huawei show less deviation from the mean compared to Fitbit. Among the three brands, Huawei has the smallest interquartile range, indicating that 50% of the data deviates the least from the mean. It is important to note that the trial yielded 160, 153, and 163 records for Fitbit, Garmin, and Huawei, respectively.

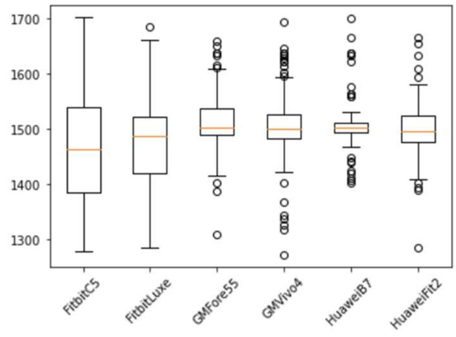

Figure 3. Box plot compares the basic statistics of smart watches of different brands.

Figure 3. presents the statistical data at the sub-model level for each smartwatch brand. Each brand comprises two sub-models, representing different price tiers: Fitbit Luxe vs. Fitbit Charge 5, Garmin Vivo 4 vs. Garmin Forerunner 5, and Huawei Band 7 vs. Huawei Fit 2. The Huawei Band 7, the most affordable option, demonstrates an accurate median step count of 1,503 steps with the least measurement deviation (IQR). Garmin Vivo 4 exhibits a median step count closest to the reference value of 1,499 steps. These findings highlight the variations in step count accuracy and deviation across different sub-models within each brand, offering insights into their performance at various price levels. Fitbit sub-models, the Fitbit Charge 5 and Fitbit Luxe, exhibit median step count values (1,465.5 and 1,494, respectively) that tend to be lower than the reference value. In contrast, Garmin and Huawei sub-models demonstrate step count results that are much closer to the reference value, indicating higher accuracy.

B. Step Count Modeling

Candidate models of Gaussian Normalization, Linear Regression, Logistic Regression, and Support Vector Machine, and Neural Network were compared in our preliminary test. Linear regression was found to be the best. We use the linear regression equation as follows (8).

$$Y = alpha + beta(X) \tag{8}$$

TABLE II. LINEAR REGRESSION MODELS BY BRANDS

Band	Model Alpha = *n*		Model Alpha = 0	
	Alpha	Beta	Alpha	Beta
Fitbit	414.03	0.761	0	1.027
Garmin	230.197	0.845	0	0.99
Huawei	104.375	0.931	0	0.996

Table II provides a comprehensive summary of the analysis outcomes, considering both the inclusion and exclusion of Alpha values. In the presence of Alpha, the linear step count adjustment model exhibits an initial value (the bias value) that ranges in the hundreds of steps or higher. Notably, Fitbit has the highest bias value. Conversely, when Alpha is set to zero, Fitbit demonstrates a step count adjustment factor of 1.027, providing the correlation of smartwatch step count values to clicker counter values. Hence, if a Fitbit device records a step count of 1,000 steps, the model will adjust it to 1,027 steps, while readings from Garmin and Huawei will be marginally adjusted to 990 and 996 steps, respectively.

Finally, to ascertain the necessity and effectiveness of the model, we employ a t-test with a required confidence level of 95% (or $p<0.05$). This statistical test allows us to assess three hypotheses.

1) Step counts from smartwatch is using gender as a parameter in the system.

The analysis results indicated from Figure 4 (a) and Figure 4 (b). The gender coefficient had a p-value of 0.538, which was greater than 0.05, which was the probability used to confirm the importance of the variable, while the smart watch coefficient had a p-value near zero. which supports that the model does not have to provide support for using gender as a variable in the system.

	coef	std err	t	P>|t|	[0.025	0.975]
StepsSmartWatch	0.9945	0.004	276.187	0.000	0.987	1.002
Gender	5.3024	8.603	0.616	0.538	-11.669	22.274

Figure 4. (a) Results of linear regression analysis with Gender as a covariate

	coef	std err	t	P>|t|	[0.025	0.975]
StepsSmartWatch	0.9960	0.003	377.735	0.000	0.991	1.001

Figure 4. (b) Results of linear regression analysis without Gender as a covariate

2) Step count from smartwatch is the same as the count from standard clicker counter.

The estimation of step count utilizing smartwatches and clicker counter involved conducting t-tests at a 5% confidence level, utilizing the complete dataset and adjusting it to the reference value of 1,500 steps according to Equation (7).

H_0: The mean step count derived from smart watches = the mean step count obtained from clicker.

H_1: The mean step count derived from smart watches $\neq$ the mean step count obtained from clicker.

The resulting t-value is -0.768805, with a corresponding p-value of 0.442185. These values indicate insignificance, leading to the acceptance of H_0. Therefore, based on the entirety of the experimental data, the average step count derived from smart watches does not differ significantly from the step count obtained from clicker counter.

3) Step count from the linear regression model is the same as the count from the manual clicker counter

The estimation of step count utilizing linear regression model and clicker counter involved conducting t-tests at a 5% confidence level, utilizing the complete dataset and adjusting it to the reference value of 1,500 steps according to Equation (7).

H_0: The average step count from the linear regression model = the average step count from the clicker.

H_1: The average step count from the linear regression models $\neq$ the average step count from the clicker.

The resulting t-value is -0.330918, with a corresponding p-value of 0.740773. These values indicate that there is no significant evidence to reject H_0. However, it is worth noting that the obtained p-value is considerably higher. This suggests that the model-adjusted values may provide an average step count that is similar or closer to the count obtained from the clicker compared to the direct readings from the smartwatch.

C. Heart Rate Data

An experiment was conducted to assess the heart rate data of six volunteers, comprising of both genders. The investigation involved measuring and comparing the pulse readings obtained from six different models of Smart Watch/Wristband with pulse acquired using medically calibrated heart rate monitor (an oximeter). In this study, a comprehensive total of 801 heart rate measurements were conducted utilizing an oximeter. These measurements were further categorized into 364 instances for males and 437 instances for females. For each data collection activity, the volunteers wore two smartwatches. As such, we have the heart rate data set of all smart watches at 1,602 data records. The data set was normalized and shown as a box plot to compare each smartwatch measured heart rate with the medical-standard heart rate monitor. The outliers were then removed with [Q1 - 1.5IQR, Q3 + 1.5IQR]. This result is shown in Figure 5.

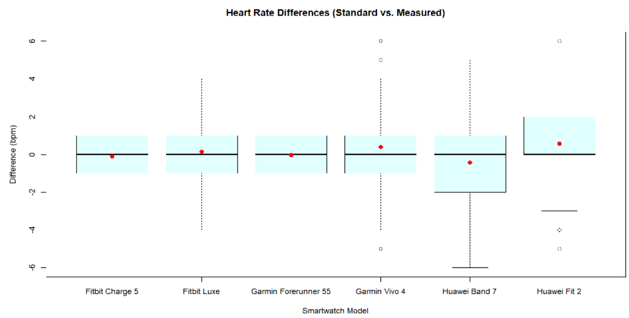

Figure 6. Heart rate differences (Standard vs Measured) with outliers remove.

After the outlier data were removed, we have 656 data for males from the original 728 and 768 data for females from the original 874. When taking the difference from the standard heart rate of each smartwatch to test the statistical hypothesis with the F-test using One-Way ANOVA, the results were $F = 3.683$ and $p = 0.00257 < 0.05$. Based on the test results, it can be concluded that there is a statistically significant difference in the mean variance between certain cohorts compared to others.

D. Heart Rate Modeling

We use a scatter plot (Figure 7) to visualize the values obtained for all smartwatch models. Including the values measured by the oximeter for all genders to examine the extent of gender differences in the results. Linear regression analysis was performed. The analysis found that, despite the observed gender differences (females in red data points, males in blue), the linear regression model did not detect significant differences between the genders.

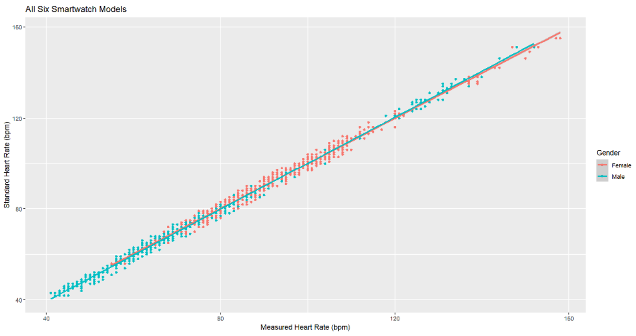

Figure 7. Scatter plot analyzed by linear regression.

We took the values from all smartwatch models and the measured values from the oximeter to create a linear regression model. The model's value of R-squared = 0.994 and the value of Adjusted R-squared = 0.994, which indicates that the model is very suitable for the data group. Additionally, separate linear regression analyses were conducted for each gender. The outcomes did not deviate significantly from those obtained when considering both genders collectively. Furthermore, multiple linear regression was also performed, incorporating multiple source variables. However, the results failed to yield a statistically significant difference compared to the previous analyses.

V. CONCLUSION

In our experiments we developed models to adjust the step count and heart rate to bring measured values from different smartwatches closer to the medical-standard device value and reduced the differences among brands. It was found that the linear regression model adjusted for step count at the brand level without needing to consider gender as a factor provide statistically satisfactory results. And the values obtained from different brands (Fitbit, Garmin, and Huawei) have similar count averages. However, this experiment is very limited, leading to an under-fitted model, but the experiment confirms the possibility of building a more rigorous model under specified requirements and clear terms of use of the model in the future.

The results were analyzed using ANOVA with all 6 Smartwatches/Wristbands of Fitbit Charge 5, Fitbit Luxe, Garmin Forerunner 55, Garmin Vivo 4, Huawei Fit 2, and Huawei Band 7. The results were not much different especially the brands Fitbit and Garmin, while the Huawei brand, both models, have different results from the two brands above.

For heart rate, experimentally and analytically using the Student's T-test and Z-test, the differences measured by smart watches and the reference device were different when considering gender. But the difference of Standard and Measured Heart Rate average is only about 0.5 bpm (beats per minute), so this does not need to be considered by gender.

This study utilized linear regression to examine the relationship between the dependent variable, Standard Heart Rate, and three primary variables, Measured Heart Rate, Gender, and Watch Models. The outcomes of R and R^2 did not exhibit significant disparities when comparing the use of Measured Heart Rate alone versus its combination with other variables. Consequently, it is viable to employ Measured Heart Rate as the default variable for approximating the standard heart rate variable.

REFERENCES

[1] J. Mander, C.Buckle, and, K. Gilsenan, "Digital healthcare understanding the evolution and digitization of healthcare," GlobalWebIndex Co., London, England, 2020.

[2] J. S. Sunny et al., "Anomaly Detection Framework for Wearables Data: A Perspective Review on Data Concepts, Data Analysis Algorithms and Prospects.," in Sensors (Basel, Switzerland) vol. 22, issue. 3, 756, Jan, 2022.

[3] J. W. Kim, J. H. Lim, S. M. Moon and B. Jang, "Collecting Health Lifelog Data From Smartwatch Users in a Privacy-Preserving Manner," in IEEE Transactions on Consumer Electronics, vol. 65, no. 3, pp. 369-378, Aug, 2019.

[4] A. O. Putri, M. A. M. Ali, and A. A. Almisreb, "Reliability and validity analysis of smartwatches use for healthcare.," in Periodicals of Engineering and Natural Sciences, vol. 9, no. 3, pp.82-89, 2021.

[5] Z., Zhang, and R.Khatami, "Can we trust the oxygen saturation measured by consumer smartwatches?.," in The Lancet Respiratory Medicine, vol. 10, March, 2022.

[6] C. Meza et al., "Accuracy of a Smartwatch to Assess Heart Rate Monitoring and Atrial Fibrillation in Stroke Patients., " in Sensors 2023, vol .23, issue. 10, 4632, May, 2023.

[7] I. Tomohiko et al., "Use of a Smart Watch for Early Detection of Paroxysmal Atrial Fibrillation: Validation Study.," in JMIR cardio, vol. 4,. 22 Jan. 2020.

[8] C. Dobbins et al., "Detecting Physical Activity within Lifelogs Towards Preventing Obesity and Aiding Ambient Assisted Living," in Neurocomputing, vol. 230, pp. 110-132, 2017.

[9] G. M. Weiss, K. Yoneda and T. Hayajneh, "Smartphone and Smartwatch-Based Biometrics Using Activities of Daily Living," in IEEE Access, vol. 7, pp. 133190-133202, 2019.

[10] P. Düking et al., "Wrist-Worn Wearables for Monitoring Heart Rate and Energy Expenditure While Sitting or Performing Light-to-Vigorous Physical Activity: Validation Study," in JMIR Mhealth Uhealth, vol. 8 , no. 5, May, 2020.

[11] M. Nissen et al., "Heart Rate Measurement Accuracy of Fitbit Charge 4 and Samsung Galaxy Watch Active2: Device Evaluation Study.," in JMIR formative research vol. 6, no. 3, Mar, 2022.

[12] K. M. Tam and S. Y. Cheung, "Validation of Electronic Activity Monitor Devices During Treadmill Walking," in Telemedicine and e-Health, vol. 24, no. 10, Oct, 2018.

[13] D. Jones et al., "Validity And Reliability Of The Fitbit Flex™ And Actigraph Gt3x+ At Jogging And Running Speeds," in International Journal of Sports Physical Therapy, vol. 13, no. 5, pp. 860-870, Aug, 2018.

[14] L. Wang, T. Liu, Y. Wang, Q. Li, J. Yi and Y. Inoue, "Evaluation on Step Counting Performance of Wristband Activity Monitors in Daily Living Environment," in IEEE Access, vol. 5, pp. 13020-13027, 2017.

[15] E. A. Thomson et al. "Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities," in Journal of Sports Sciences, vol. 37, issue. 12, pp. 1411-1419, Jan, 2019.

[16] M. P. Støve et al. " Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity," in Journal of Sports Sciences, vol. 37, issue. 8, pp. 895-901, Oct, 2018.