From Angiography to Algorithms: Assessing the Viability of Consumer Wearables as a First-Line Screening Tool for Cardiac Risk

Taher Alabbar

Student ID: 3193886

STAT 3103–001: Applied Regression Analysis

December 5, 2025

1 Introduction

Cardiovascular disease remains the leading cause of morbidity and mortality globally, placing a significant burden on healthcare systems. The survival rate for approximately 50% of patients suffering from heart disease is less than 10 years [4]. While coronary angiography is currently the standard for confirmation, it incurs high costs, potential severe side effects, and requires expert interpretation, rendering it unsuitable for widespread initial screening [10]. Consequently, there is a critical need to validate diagnostic models that rely on accessible, non-invasive clinical metrics.

Research by Nahar et al. (2013) highlights a significant gap in current automated diagnostic approaches: purely computerized feature selection (CFS) often prioritizes invasive, high-cost attributes—such as fluoroscopy results or thallium stress test outcomes—because of their statistical correlation, while discarding medically significant but "cheaper" indicators like age and blood pressure [8]. However, their study demonstrated that a "Medical Knowledge Driven" approach, which retained these accessible variables, actually achieved superior prediction accuracy [8].

Building on this insight, the primary objective of this study is to evaluate the statistical sufficiency of a defined subset of "Smartwatch & Home Health" covariates. To test this, we utilize the "Processed Cleveland Heart Disease Database" from the UCI Machine Learning Repository, originally collected by Dr. Robert Detrano [2]. Widely recognized as a benchmark in medical data mining, this dataset is one of the most frequently cited resources for coronary heart disease research [8]. It comprises 303 patient records and includes 13 independent clinical attributes. Distinguished by its exceptional data quality, the processed version contains no missing values across these critical attributes, ensuring a robust statistical baseline. We have categorized the dataset's covariates into two distinct groups to isolate the predictive power of consumer-accessible technology:

Group A: Smartwatch & User-Reported Metrics (The Focus)
These variables represent data obtainable via standard consumer wearables or simple user input, without the need for clinical supervision or invasive testing:

- **Sensors & Wearables:** Maximum heart rate (*thalach*) via optical sensors and resting blood pressure (*trestbps*) via home cuffs.

- **User Inputs:** Demographic profiles (*age*, *sex*) and self-reported symptom logs, specifically chest pain type (*cp*) and exercise-induced angina (*exang*).

Group B: Clinical & Invasive Benchmarks (The Control)
These variables require hospital-grade equipment, blood sample analysis, or expert interpretation of multi-lead ECGs:

- **ECG Analysis:** Resting ECG results (*restecg*), ST-segment depression (*oldpeak*), and slope (*slope*), which require clinical-grade electrocardiography for accurate ischemic detection.

- **Invasive/Lab Data:** Serum Cholesterol (*chol*) and Fasting Blood Sugar (*fbs*) (blood analysis); Fluoroscopy results (*ca*) and Thallium stress test results (*thal*) (imaging).

Validating Group A metrics as reliable proxies for cardiac health is essential for the future of personalized medicine. If these non-invasive variables achieve a statistically significant goodness-of-fit using **Logistic Regression**, it establishes the clinical viability of utilizing consumer technology for continuous, autonomous monitoring of cardiac output.

Given the binary nature of the outcome (*target*: Disease vs. No Disease), **Logistic Regression** was selected as the analytical technique. While sophisticated machine learning models such as XGBoost

or Neural Networks may offer marginal gains in raw predictive accuracy, they often function as "black boxes," obscuring the underlying biological relationships. Logistic Regression is preferred in this study for its **clinical interpretability**: it provides quantifiable Odds Ratios, allowing us to determine exactly how much the risk increases for every unit change in a smartwatch metric (e.g., heart rate). Furthermore, baseline performance benchmarks from the UCI Repository indicate that Logistic Regression achieves classification accuracy ($\sim$84%) comparable to complex ensemble methods on this dataset [5], making it a statistically sufficient and computationally efficient choice for potential deployment on resource-constrained wearable devices.

2 Literature Review

The application of statistical models to diagnose coronary heart disease has evolved significantly, shifting from rigid clinical rule-sets to advanced probabilistic algorithms. The foundation for this transition was established by Detrano et al. (1989), who originally compiled the Cleveland Heart Disease Database [2]. In their seminal work, "International application of a new probability algorithm for the diagnosis of coronary artery disease," Detrano and colleagues demonstrated that algorithmic approaches could effectively stratify patients by risk. Crucially, their study utilized coronary angiography as the "ground truth" for validation, creating a robust dataset where the presence of disease is confirmed by invasive imaging, allowing subsequent researchers to test non-invasive predictive models against a confirmed standard [2].

However, as data mining techniques advanced, a disconnect emerged between statistical correlation and clinical utility. In a critical analysis of computational intelligence techniques, Nahar et al. (2013) investigated the limitations of automated feature selection. Their study highlighted that standard Computerized Feature Selection (CFS) algorithms often prioritized variables with the highest mathematical correlation to the target, regardless of cost or invasiveness [8]. Consequently, automated methods frequently selected attributes such as fluoroscopy results (*ca*) and thallium stress test outcomes (*thal*), while discarding medically significant but statistically "noisier" indicators like age, resting blood pressure, and cholesterol [8]. To address this, Nahar et al. proposed a "Medical Knowledge Driven Feature Selection" (MFS) approach, which prioritized variables established in medical literature as significant risk factors. Their findings provide strong theoretical support for the sufficiency of non-invasive testing, as the MFS approach noticeably improved classification accuracy compared to the automated selection of invasive features [8].

The specific selection of non-invasive predictors for our "Group A" (Smartwatch-Feasible) subset is further validated by recent technological assessments. Martín-Escudero et al. (2023) evaluated the accuracy of wrist-worn devices against clinical ECG during maximal stress testing, confirming that modern optical sensors (e.g., Apple Watch) provide reliable heart rate (*thalach*) monitoring during exercise, supporting its inclusion as a primary consumer-accessible metric [7]. Similarly, the inclusion of resting blood pressure (*trestbps*) is supported by Lee et al. (2025), whose multi-perspective validation of smartwatch-based blood pressure monitoring demonstrated that, with proper calibration, wearable devices can achieve error margins within clinical thresholds (< 5 mmHg), rendering them viable for preventive screening [6].

Conversely, the classification of ECG morphological features such as ST-depression (*oldpeak*) and slope as "clinical benchmarks" (Group B) is justified by the design limitations of current wearable technology. As detailed in the ACS WATCH II study protocol by Buelga Suárez et al. (2024), standard smartwatches are designed to record a single-lead ECG (Lead I), which lacks the spatial resolution of clinical 12-lead systems required to reliably detect ischemic ST-segment changes [1]. This distinction is reinforced by the feature analysis of El-Bialy et al. (2015), who used decision trees to identify that while

non-invasive factors like age and heart rate are dominant predictors, invasive metrics like fluoroscopy (*ca*) and thallium results (*thal*) remain critical for maximizing diagnostic precision [3]. This literature collectively underscores the necessity of our research objective: to rigorously quantify the statistical trade-off between these high-value invasive metrics and the increasingly accessible non-invasive set.

3 Methodology

This study employs a statistical inference framework to evaluate the predictive sufficiency of non-invasive clinical attributes. The analytical process was conducted using the R statistical computing environment.

3.1 Data Preprocessing

The analysis utilizes the "Processed Cleveland Heart Disease Database" sourced from the UCI Machine Learning Repository [5]. This dataset is widely cited in literature as being free of missing values compared to other heart disease databases [3]. However, upon inspection of the raw data, minor inconsistencies (coded as "?") were identified in 6 observations within the *ca* and *thal* attributes. To ensure strict statistical rigor, these rows were excluded, reducing the sample size from 303 to 297. This preprocessing step ensures that the Logistic Regression model is not biased by placeholder characters.

Prior to modeling, specific transformations were applied:

- **Binary Transformation of Response Variable:** The original target variable (*num*) contains integer values from 0 (No Disease) to 4 (Severe Disease). For the purpose of Logistic Regression, this was converted into a binary outcome variable (*target*), where 0 represents absence and any value ≥ 1 represents the presence of angiographic heart disease.

- **Factor Encoding:** Categorical variables such as chest pain type (*cp*) and resting ECG (*restecg*) were encoded as unordered factors. This ensures that the model treats the integer codes (e.g., 1–4) as distinct categories rather than an ordinal sequence, preventing the imposition of an arbitrary linear magnitude on qualitative symptoms.

3.2 Logistic Regression Model Specification

Given the binary nature of the dependent variable ($Y \in \{0, 1\}$), **Logistic Regression** was selected as the primary analytical technique. Unlike linear regression, which assumes a continuous outcome, logistic regression models the **natural logarithm** of the odds (logit) of the disease occurring as a linear combination of the independent variables:

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_k X_k \tag{1}$$

This approach allows for the interpretation of coefficients (β) as **Odds Ratios** (e^β), quantifying the multiplicative change in risk associated with a one-unit increase in a covariate while holding all other variables constant.

3.3 Model Selection and Comparison Strategy

To explicitly test the research hypothesis—that consumer-accessible metrics are statistically sufficient—a **Nested Model Comparison** was performed using the **Likelihood Ratio Test (LRT)**. Two distinct models

were fitted based on the covariate categorization:

1. **The Reduced Model (Group A - Consumer/Wearable):** This model isolates variables theoretically obtainable via standard consumer wearables or user input, without the need for clinical supervision. It includes:

   - *Sensors: thalach* (Max HR), *trestbps* (Resting BP).
   - *User Inputs: age*, *sex*, *cp* (Chest Pain Type), and *exang* (Exercise Induced Angina).

2. **The Full Model (Group B - Clinical Additions):** This model acts as the control benchmark, including all covariates from Group A plus variables requiring hospital-grade equipment or expert interpretation. The added variables are:

   - *ECG Analysis: restecg*, *oldpeak*, and *slope*.
   - *Invasive/Lab Data: chol* (Serum Cholesterol), *fbs* (Fasting Blood Sugar), *ca* (Fluoroscopy), and *thal* (Thallium Stress Test).

The **Akaike Information Criterion (AIC)** was calculated for both models to assess the trade-off between goodness-of-fit and model complexity. The Likelihood Ratio Test was then employed to determine if the addition of the invasive Group B variables significantly reduced the residual deviance. A *p*-value > 0.05 would indicate that the invasive tests do not add statistically significant information beyond the smartwatch metrics.

3.4   Diagnostic Procedures

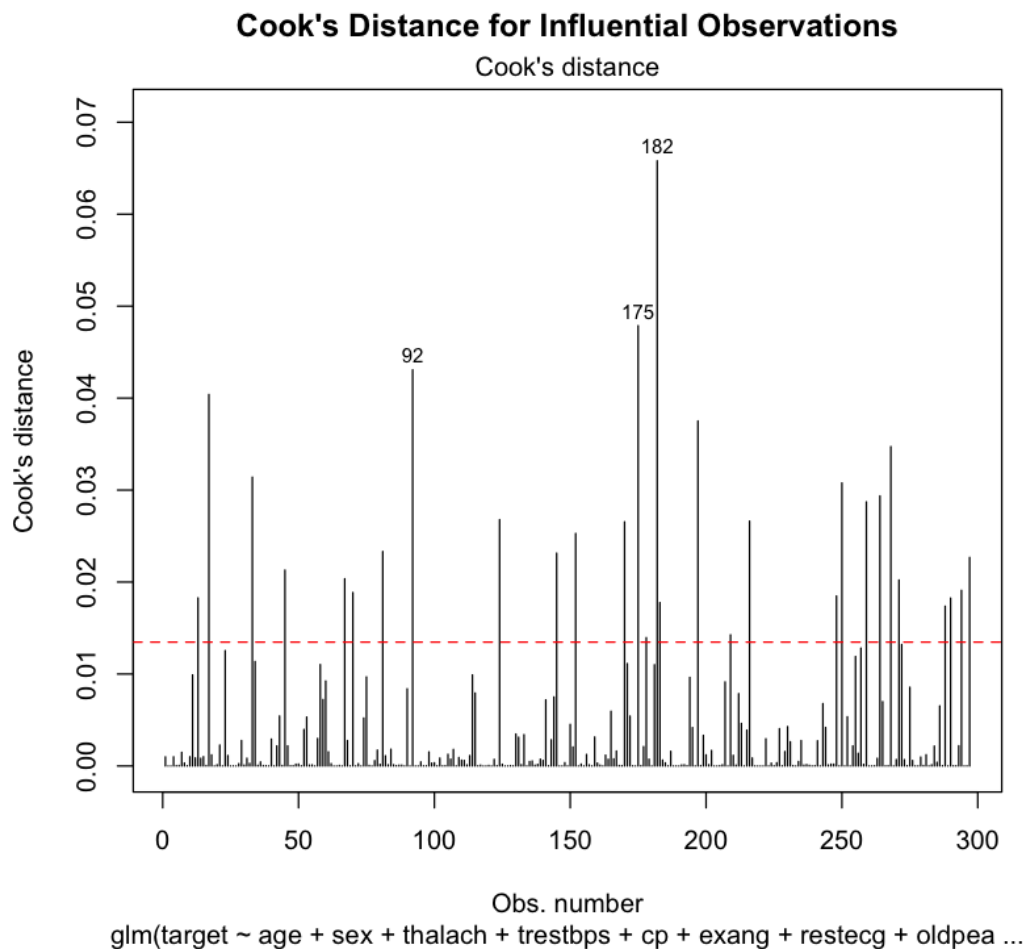To ensure the robustness of the statistical inferences, the following regression diagnostics were applied:

- **Multicollinearity:** Variance Inflation Factors (VIF) were calculated for the Full Model. A VIF threshold of > 5 was established to identify potential multicollinearity that could inflate standard errors.

- **Influential Observations:** Cook's Distance was computed for all observations to identify influential data points. Observations with a Cook's Distance $> 4/N$ were inspected to ensure they were not result of data entry errors skewing the model coefficients.

- **Linearity of the Logit:** The assumption of linearity between continuous predictors and the log-odds of the outcome was assessed using normal plotting and visual inspection. This procedure was applied exclusively to continuous covariates (*age*, *thalach*, *trestbps*, *oldpeak*), as categorical predictors (*sex*, *cp*) are modeled as dummy variables and thus do not require linearity verification. A linear trend in the plots indicates that the assumption is satisfied.

4   Results

4.1   Regression Diagnostics

Prior to interpreting the model estimates, a comprehensive diagnostic assessment was conducted to ensure the validity of the logistic regression assumptions.
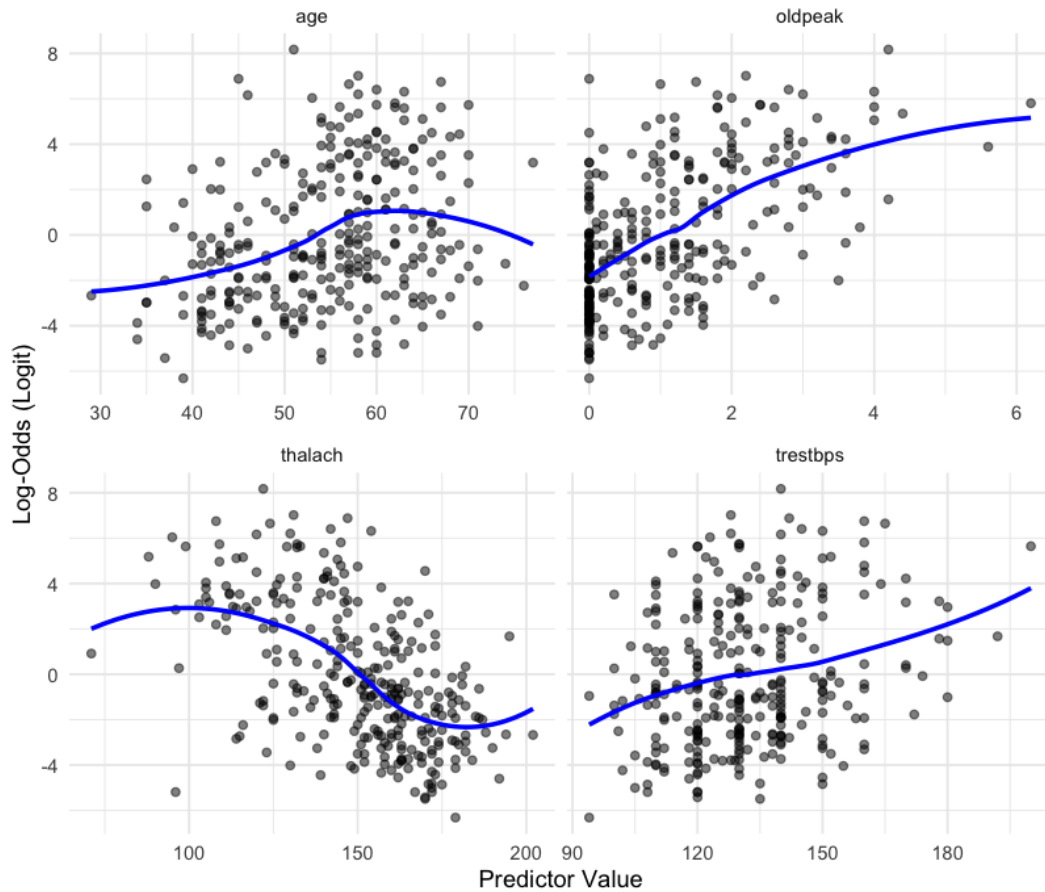
- **Multicollinearity:** Variance Inflation Factors (VIF) were calculated for all predictors in the Full Model. All Generalized VIF (GVIF) values were found to be well below the conservative threshold of 5 (maximum GVIF = 1.84 for *slope*), indicating that multicollinearity is not a confounding factor in this analysis.

- **Influential Observations:** Cook's Distance was computed to identify influential data points. While three observations (IDs 92, 175, 182) exceeded the theoretical threshold of $4/N$, the maximum observed distance was approximately 0.07, far below the critical value of 1.0. This confirms that no single outlier disproportionately skewed the model parameters.

Cook's Distance for Influential Observations

- **Linearity:** The assumption of linearity between continuous predictors and the log-odds of the outcome was verified via visual inspection of Lowess-smoothed scatter plots. As shown in Figure 2, variables *age*, *thalach*, *trestbps*, and *oldpeak* exhibited approximately linear trends, satisfying the assumption.

Linearity Check: Logits vs Continuous Predictors
Blue line represents Lowess smoothing. A linear pattern suggests assumption is met.

4.2 Model Comparison: Consumer vs. Clinical Sufficiency

A Likelihood Ratio Test (LRT) was conducted to evaluate the loss of information when excluding invasive clinical metrics. The comparison revealed that the Full Model (Group B) provided a statistically significant improvement in fit over the Reduced Consumer Model (Group A) ($\chi^2(10) = 70.54, p < 0.001$). This was corroborated by the Akaike Information Criterion (AIC), which decreased from 280.18 in the Reduced Model to 229.64 in the Full Model, indicating that clinical benchmarks capture variance in cardiac risk that consumer sensors cannot yet fully explain.

```
Analysis of Deviance Table

Model 1: target ~ age + sex + thalach + trestbps + cp + exang
Model 2: target ~ age + sex + thalach + trestbps + cp + exang + restecg +
    oldpeak + slope + chol + fbs + ca + thal
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       288     262.18
2       278     191.64 10   70.535 3.495e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 1: Likelihood Ratio Test for Model Comparison

5 Further Discussion

5.1 Predictive Utility of Consumer Metrics

A direct comparison of the model fit statistics further elucidates the trade-off between accuracy and accessibility. The invasive model achieved a McFadden's Pseudo-R^2 of 0.533, compared to 0.360 for the non-invasive consumer-grade model. While the invasive model explains a greater proportion of the variance, the consumer model captures approximately **68%** of this explanatory power (0.360/0.533) using variables that incur zero surgical risk and negligible cost. This finding is pivotal: it suggests that the "cost" of removing invasive testing results in retaining roughly two-thirds of the diagnostic signal, a trade-off that remains viable for wide-scale population screening.

Despite the statistical superiority of the Full Model, the consumer-based Reduced Model demonstrated robust discriminative power. Receiver Operating Characteristic (ROC) analysis yielded an Area Under the Curve (AUC) of **0.874** for the consumer model, compared to **0.935** for the clinical model. An AUC of 0.874 places the consumer model firmly in the range of excellent discrimination, indicating that even without invasive data, the model can reliably distinguish between healthy and diseased individuals in the vast majority of cases.
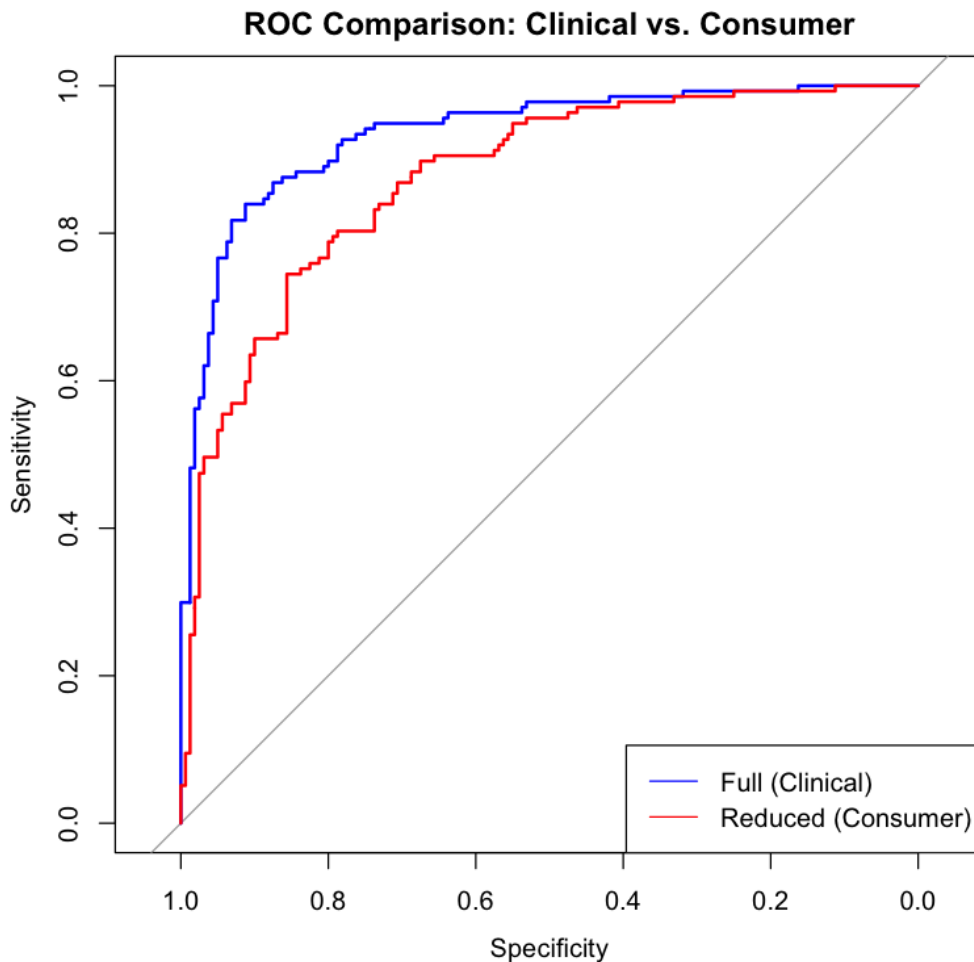

Figure 2: ROC Curve Comparison

While the clinical model is superior, an AUC of 0.874 places the consumer model in the range of excellent discrimination. In terms of classification performance, the consumer model achieved an overall

accuracy of 79.8%. Notably, it exhibited a **Specificity of 83.8%**, indicating a strong ability to filter out healthy individuals, though its **Sensitivity of 75.2%** suggests it is best utilized as a preliminary screening tool rather than a standalone diagnostic.

5.2 Drivers of Consumer-Based Prediction

Analysis of the Reduced Model coefficients validates the utility of specific wearable sensors. Maximum Heart Rate (*thalach*) emerged as a highly significant predictor ($p < 0.001$), with higher rates associated with lower disease risk. Resting Blood Pressure (*trestbps*) was also significant ($p = 0.014$).

Interestingly, chronological *age* was not statistically significant ($p = 0.13$) in the presence of these physiological variables, suggesting that functional markers (HR, BP) are more direct indicators of cardiac health than age alone. Among user-reported symptoms, asymptomatic chest pain (*cp4*) was a strong predictor of disease presence ($p < 0.001$).

5.3 Future Directions and Technological Potential

Although the invasive model currently provides more stability, the predictive power of the consumer-grade model was remarkably robust. This study constrained the non-invasive set primarily to metrics available on standard smartwatches. However, the gap between consumer and clinical diagnostics is rapidly narrowing. Future developments in "minimally invasive" bio-wearables, such as microneedle patches for interstitial fluid analysis, could soon allow for the continuous monitoring of glucose (*fbs*) and lipids (*chol*) without phlebotomy. Furthermore, as wearable ECG technology matures from single-lead to multi-lead equivalents, the accuracy of detecting complex ischemic markers like slope and ST-segment deviation will likely improve. Integrating these next-generation data streams into the predictive framework proposed here could eventually render the model sufficient for definitive diagnosis, reducing the reliance on hospital-based procedures even further.While Logistic Regression was chosen for its interpretability in this baseline study, the integration of high-dimensional, continuous biomarker data may necessitate the use of more sophisticated machine learning architectures. Recent comparative research by Shrestha (2024) indicates that ensemble methods, specifically **Gradient Boosting Machines (XGBoost)**, significantly outperform Logistic Regression by capturing non-linear interactions, achieving accuracies as high as 90% and an AUC-ROC of 0.94 [9]. Additionally, **Long Short-Term Memory (LSTM)** networks may be explored to address temporal patterns in future longitudinal data, though current benchmarks suggest they require careful adaptation when applied to static clinical datasets [9].

6 Conclusion

This study demonstrates that a logistic regression model relying solely on non-invasive, wearable-compatible attributes can capture the majority of the diagnostic signal for coronary heart disease found in invasive clinical models. While invasive tests remain the gold standard for confirmatory diagnosis, their statistical contribution, though significant, may not always justify their high cost and risk for initial screening purposes. The findings advocate for the continued development of wearable algorithms that integrate heart rate variability, exercise data, and user-reported symptoms, as these accessible metrics offer a viable, scalable, and statistically robust pathway for early heart disease detection.

Bibliography

[1] Buelga Suárez, M., Pascual Izco, M., García Montalvo, J., & Alonso Salinas, G. L. (2024). Accuracy of Smartwatch Electrocardiographic Recording in the Acute Coronary Syndrome Setting: Rationale and Design of the ACS WATCH II Study. *Journal of Clinical Medicine*, 13(2), 389.

[2] Detrano, R., et al. (1989). International application of a new probability algorithm for the diagnosis of coronary artery disease. *American Journal of Cardiology*, 64(5), 304–310.

[3] El-Bialy, R., Salamay, M. A., Karam, O. H., & Khalifa, M. E. (2015). Feature analysis of coronary artery heart disease data sets. *Procedia Computer Science*, 65, 459–468.

[4] Gupta, A., Kumar, R., Arora, H. S., & Raman, B. (2020). MIFH: A machine intelligence framework for heart disease diagnosis. *IEEE Access*, 8, 14659–14674.

[5] Janosi, A., Steinbrunn, W., Pfisterer, M., & Detrano, R. (1988). Heart Disease [Dataset]. UCI Machine Learning Repository. `https://doi.org/10.24432/C52P4X`

[6] Lee, Y., Lee, T., Lee, J., Lee, H.-Y., & Seo, J.-M. (2025). Validation of Smartwatch-Based Blood Pressure Monitoring in Young Adults: Multi-Perspective Analysis by Dual Comparison With Cuff-Based Devices. *IEEE Access*, 13, 30587–30596.

[7] Martín-Escudero, P., Cabanas, A. M., Dotor-Castilla, M. L., et al. (2023). Are Activity Wrist-Worn Devices Accurate for Determining Heart Rate during Intense Exercise? *Bioengineering*, 10(2), 254.

[8] Nahar, J., Imam, T., Tickle, K. S., & Chen, Y. P. P. (2013). Computational intelligence for heart disease diagnosis: A medical knowledge driven approach. *Expert Systems with Applications*, 40(1), 96–104.

[9] Shrestha, D. (2024). Advanced Machine Learning Techniques for Predicting Heart Disease: A Comparative Analysis Using the Cleveland Heart Disease Dataset. *Applied Medical Informatics*, 46(3), 91–102.

[10] Thiyagaraj, M., & Suseendran, G. (2020). Enhanced prediction of heart disease using particle swarm optimization and rough sets with transductive support vector machines classifier. *Springer Nature*, 141–151. https://doi.org/10.1007/978-981-13-9364-8_11

Appendix

A. R Code for Data Analysis

The following R code was used to load the dataset from the UCI Repository, clean the data, and perform the comparative logistic regression analysis between the Consumer (reduced) and Clinical (full) models.

```r
# Heart Disease Analysis - Logistic Regression Framework

# 1. SETUP & LIBRARY LOADING

if (!require("pacman")) install.packages("pacman")
pacman::p_load(tidyverse, car, broom, pROC, caret)

# 2. DATA LOADING & PREPROCESSING

url <- paste0("https://archive.ics.uci.edu/ml/machine-learning-databases
    /" ,"heart-disease/processed.cleveland.data")

col_names <- c("age", "sex", "cp", "trestbps", "chol", "fbs", "restecg",
                "thalach", "exang", "oldpeak", "slope", "ca", "thal", "
                    num")

heart_data <- read.csv(url, header = FALSE, col.names = col_names, na.
    strings = "?")

# Preprocessing Pipeline
clean_data <- heart_data %>%
  drop_na() %>%
  mutate(
    # Create Binary Target: 0 = No Disease, 1 = Disease
    target = ifelse(num > 0, 1, 0),

    # Factor encoding
    sex     = factor(sex, labels = c("Female", "Male")),
    cp      = as.factor(cp),
    fbs     = as.factor(fbs),
    restecg = as.factor(restecg),
    exang   = as.factor(exang),
    slope   = as.factor(slope),
    thal    = as.factor(thal),
    ca      = as.numeric(ca)
  )

# 3. MODEL SPECIFICATION

# --- Model 1: Reduced (Consumer/Wearable) ---
# Variables: Age, Sex, HR, BP, Chest Pain, Exercise Angina
model_reduced <- glm(target ~ age + sex + thalach + trestbps + cp +
```

```r
    exang ,
                      data = clean_data ,
                      family = binomial ( link = "logit"))

# --- Model 2: Full ( Clinical / Invasive ) ---
# Adds : ECG , ST depression , slope , cholesterol , fasting sugar ,
    fluoroscopy , thallium
model_full <- glm ( target ~ age + sex + thalach + trestbps + cp + exang +
                      restecg + oldpeak + slope + chol + fbs + ca + thal ,
                   data = clean_data ,
                   family = binomial ( link = "logit"))

# Calculate McFadden 's Pseudo R-Squared
calc_r2 <- function ( model ) {
  1 - ( model$deviance / model$null.deviance )
}

print ( paste ("Consumer␣Model␣R2:", round ( calc_r2 ( model_reduced ), 3)))
print ( paste ("Clinical␣Model␣R2:", round ( calc_r2 ( model_full ), 3)))

# 4. STATISTICAL INFERENCE

# Likelihood Ratio Test ( Nested Model Comparison )
print ( anova ( model_reduced , model_full , test = "Chisq"))

# AIC Comparison
model_stats <- bind_rows (
  glance ( model_reduced ) %>% mutate ( Model = "Reduced␣( Consumer )"),
  glance ( model_full ) %>% mutate ( Model = "Full␣( Clinical )")
) %>% select ( Model , AIC , BIC , deviance )
print ( model_stats )

# 5. DIAGNOSTICS

# A. Multicollinearity ( VIF )
print ( vif ( model_full ))

# 6. PERFORMANCE METRICS ( ROC & CONFUSION MATRIX )

# A. ROC Curve Analysis
prob_reduced <- predict ( model_reduced , type = "response")
prob_full <- predict ( model_full , type = "response")

roc_reduced <- roc ( clean_data$target , prob_reduced , quiet = TRUE )
roc_full <- roc ( clean_data$target , prob_full , quiet = TRUE )

print ( paste ("Reduced␣AUC:", auc ( roc_reduced )))
print ( paste ("Full␣AUC:", auc ( roc_full )))

# B. Confusion Matrix ( Threshold = 0.5)
```

```
89  predicted_class <- ifelse(prob_reduced > 0.5, 1, 0)
90  conf_matrix <- confusionMatrix(factor(predicted_class),
91                                  factor(clean_data$target),
92                                  positive = "1")
93  print(conf_matrix)
94
95  # 7. MODEL INTERPRETATION (Odds Ratios)
96
97  tidy(model_reduced, exponentiate = TRUE, conf.int = TRUE) %>%
98      filter(p.value < 0.05) %>%
99      select(term, estimate, p.value, conf.low, conf.high) %>%
100     print()
```