BAB 1. INTRODUCTION

According to the World Health Organization, in 2019, it was estimated that there were 17.9 million deaths due to heart disease [1]. Coronary Heart Disease (CHD) is Indonesia's most significant cause of death. In 2016, the death rate from this disease reached 122 people per 100,000 population. Furthermore, CHD is a leading cause of mortality in the UK and across the world [2].

CHD is a condition in which the blood supply to the heart is obstructed or disrupted by fat in the coronary arteries [2]. Symptoms of this disease can be known if a person experiences chest pain, pain throughout the body, hard to breath, feeling weak and sick, and nausea. This disease can be caused by an unhealthy lifestyle, age, smoking, diabetes, obesity, or a family history of CHD. Heart Rate Variability (HRV) can be used as a valuable index to predict the occurrence of CHD and other heart diseases by accessing and analyzing the response of the Autonomic Nervous System (ANS) [3]. HRV is a computation of the time difference between each heartbeat. This variation is controlled by the ANS, a primitive part of the nervous system. It operates by automatically controlling heart rate, blood pressure, respiration, and digestion, among other important functions. [4].

Photophletysmograph (PPG) is a signal that can be used to determine the condition of the heart in a certain period. PPG is rapidly being used not only through medical equipment but also in wearable devices and smartphones to monitor oxygen saturation, cardiovascular disease symptoms, and HRV [5]. Currently, the HRV index can be known through the PPG signal by utilizing the camera from a smartphone so that CHD detection can be carried out. PPG uses a light source and a photodetector at the skin's surface to measure the volumetric variations of blood circulation [6]. The utilization of PPG signals to detect CHD is still little implemented. Electrocardiogram (ECG) signals are more often used to detect CHD [5]. Research [7] [8] [9] proposed a study to detect CHD using ECG signals. Therefore, developing a PPG signal for CHD detection is necessary as an alternative to the ECG signal.

In general, the detection of CHD using machine learning has four stages, namely denoising, feature extraction, feature selection, and classification. The denoising stage is the initial stage to remove noise in the signal but still retain important information in the signal. The feature extraction stage is the stage of grouping and distinguishing features based on the information on the signal. The feature extraction results are then evaluated to select the most influential features in the feature selection stage. The selected features are then used as input variables for CHD detection at the classification stage. The feature selection stage is essential because it can improve the detection performance of the built classification model. Moreover, using irrelevant features can reduce detection performance in the classification model. This can be proven in research [10] which tried to detect cardiovascular disease (CVD) using the Anova feature selection method with SVC classification. The datasets used are CVD and Framingham datasets. The highest accuracy achieved using the full feature set is 0.73 for the CVD dataset and 0.66 for the Framingham dataset. However, after using the features from the feature selection method, the accuracy increased to 0.75 and 0.71 for each dataset.

Research [7] conducted a study on feature selection methods and proposed a feature selection method, namely an ensemble algorithm based on multiple feature selection (EA-MFS). To find the best method for CHD detection, they compared feature selection methods. These methods are analysis of variance (Anova), mutual information (MutualInfo) and chi-square test (CHI). Classification is performed using a variety of classification algorithms using CHD datasets from demographics, symptoms and examination, ECG, and laboratory and echo features. As a result, the proposed method outperforms other methods in terms of classification performance and robustness on the CHD dataset.

Research [8] detect CHD using a feature selection Fisher-based Hybrid Extreme Learning Machine. The dataset used is from Z-Alizadeh sani, containing demographics, symptoms and examinations, ECG, laboratory and echo. Classification is done with selected features from the results of the Fisher method, which is the best ten features out of 15 features. The result achieved is accuracy of 97.6%.

Research [9] conducted a CHD prediction experiment using the Supervised algorithm with feature selection using an analytical approach. The data used comes from the FHS dataset in which there is a Heart Rate that comes from the ECG signal. The classification algorithm used is Random Forest, Decision Tree, and KNN. Classification using Random Forest has the highest accuracy of 96.80%.

Research [11] tried to detect Atrial Fibrillation (AF) using PPG signals with SVM classification with the Sequential Forward Selection (SFS) feature selection algorithm to find the best features heuristically. The experiment yielded an accuracy value of 95.7%, sensitivity of 94.2%, and specificity of 96.2%.

Research [12] identified fundamental features of HRV data which can be used for biometric recognition. HRV data is extracted from the PPG signal obtained from the IR-based pulse detection sensor. A Genetic Algorithm

(GA) was applied and used an adaptive search technique. The experiment resulted in 15 dominant features out of 101 features.

This research studied feature selection methods to detect CHD based on PPG signal. Three methods have been considered to this study, i.e., Analysis of Variance (Anova), Pearson Correlation, and Recursive Feature Elimination (RFE). The three methods will be compared to determine the best method in this study to detect CHD based on PPG signals. In this study, a comparison of the results of the accuracy, sensitivity, and specificity performance of the three proposed feature selection methods will be carried out.