I. Introduction

Cancer continues to be a major global health concern. According to the International Agency for Research on Cancer (IARC), around 19.3 million new cancer cases and 10 million deaths were recorded globally in 2020 [1]. Detecting cancer at an early stage with high accuracy plays a crucial role in enhancing patient survival. One promising method in cancer detection is DNA microarray analysis, which enables simultaneous measurement of thousands of gene expressions to identify cancer-related patterns [2].

Despite its potential, microarray data presents major challenges due to its high dimensionality, where the number of features significantly exceeds the number of available samples, a condition commonly referred to as the "curse of dimensionality". Moreover, medical datasets typically contain noise, have variations in feature values, and imbalanced class distributions, which can lead to overfitting and decreased classification accuracy [3]. To address these issues, dimensionality reduction methods are essential for eliminating non-informative features while preserving relevant information. Principal Component Analysis (PCA) is widely used for linear dimensionality reduction, while Autoencoders, as neural network-based models, are capable of capturing complex non-linear feature representations.

Several studies have demonstrated the benefits of these methods. Adiwijaya *et al.* [4] applied PCA on microarray data for cancer classification and achieved 96.07% accuracy using Levenberg-Marquardt Backpropagation (LMBP). Similarly, Kabir *et al.* [5] used Autoencoder for RNA-seq data and obtained a classification accuracy of 97.8%. However, PCA is inherently limited to capturing linear relationships between features, whereas Autoencoder may often require large datasets and careful parameter tuning to achieve optimal generalization. These limitations underline the need for a hybrid approach that combines the complementary strengths of PCA and Autoencoder.

To address these limitations, this study proposes a hybrid dimensionality reduction approach that integrates PCA and Autoencoder. PCA is first applied to reduce the dimensionality linearly, followed by Autoencoder to extract non-linear patterns from the reduced feature space. This combined method is evaluated using a Support Vector Machine (SVM) classifier and compared with standalone PCA and Autoencoder baselines. The goal of this study is to improve classification performance on microarray cancer datasets and contribute to the medical field through more accurate cancer detection.