CHAPTER I. INTRODUCTION

HD or cardiovascular disease (CVD) has consistently been the primary cause of death in numerous countries. Over 17.9 million deaths were caused by CVD in 2019 (World Health Organization, 2021). Coronary Heart Disease is the most common type of HD, killing 375,476 people in 2021 (Centers for Disease Control and Prevention, 2023). HD comprises a variety of factors that affect the structure or function of the heart (Shah, Patel and Bharti, 2020; World Heart Federation, 2020; Reshan *et al.*, 2023). Medical practitioners may have difficulties quickly and accurately diagnosing HD because of this factor (Reshan *et al.*, 2023).

The early detection of diseases to mitigate the risk of death can now be achieved by utilizing the accessibility of research data, hospital medical records, and computer technology. ML and AI are recognized as important contributors in the medical field. Models of ML and Deep Learning (DL) have the potential to predict or classify outcomes as well as diagnose disease. ML and AI are recognized as important contributors in the medical field. Models of ML and DL have the potential to predict or classify outcomes as well as diagnose disease (Bharti *et al.*, 2021). The utilization of ML algorithms in the system and methodology simplifies and makes easier data analysis (Grampurohit and Sagarnal, 2020). DL algorithms outperform traditional techniques for predicting disease status using genetic data (Wu *et al.*, 2018). Utilizing DL models such Artificial Neural Networks (ANN) and Convolutional Neural Networks (CNN) has been discovered to enhance the precision of multi-disease forecasting, encompassing conditions such as diabetes, breast cancer, and COVID-19 (Lalithadevi and Krishnaveni, 2022). As in the research studied by Riyadi, Rasyid, and Damarjati (2022), using deep learning models for recognizing COVID-19 in CT-Scan images demonstrated high accuracy, with each model achieving over 90% accuracy. Advancements in this field show that deep learning models are efficient in predicting diseases and offer potential for future exploration (Xie, Yu and Lv, 2021).

Currently, numerous studies utilize various ML and DL models to classify and predict disease diagnosis. In the study, Dayal, Shukla and Mahapatra (2023) who applied MLP for predicting diabetes risk. Other studies have also been conducted by Bhoyar *et al.* (2021) accurately predicted HD for in real-time using an MLP model.

Pal and Parija (2021) and Sumwiza *et al.* (2023) utilizeML to predict HD by utilizing the Random Forest (RF) algorithm. The study utilized a dataset with 303 samples and 14 features. The RF algorithm utilized in this study achieved an outstanding degree of accuracy, with scores of 93.3% and 96%.

Guo, Zhao and Yin (2017) carried out a study that combined ML and DL. They utilized a hybrid model that combined RF and MLP to predict the level of dielectric loss in nanocomposite films. This model demonstrates high efficiency, as demonstrated by its correlation coefficient (CC) score of 0.9447.

Feature selection can improve prediction accuracy and can overcome dimensional problems, so this is very important in training ML models. Selection of relevant features is important to improve model performance for predicting HD (Pathan *et al.*, 2022; Pudjihartono *et al.*, 2022). Remeseiro and Bolon-Canedo (2019) argue about the significance of feature selections in medicine for reducing dimensions as well as knowing causes of diseases. These works together emphasize an important role played by feature selection in enhancingML algorithms' performance primarily within medical diagnosis field.

As far as we know, there has been limited prior research examining the influence of feature selection based on correlation on HD. This study focuses on the feature selection process and the subsequent measurement of these features utilizing MLP and RF algorithms. Through this study, we aim to determine the impact of feature selection on the performance of MLP and RF algorithms utilizing predefined features.