CHAPTER 1
INTRODUCTION

1.1. Background

Diabetes is one of the chronic diseases in the world. It causes the sufferer to have high blood sugar or hyperglycemia due to the body unable to respond or produce the required amount of insulin to regulate the glucose in their body. It may cause complications or it may increase the risk of developing into another disease such as heart disease, kidney failure, blindness, urinary organ diseases , etc. According to the World Health Organization (WHO) Report in 2016, everyone could be infected including children, women, men, young and old. Early fast diagnosis is one of the most important issues to fight this chronic disease . If there are a lot of patient records, the machine learning classification algorithms play a great role in predicting whether a person has diabetes or not [1].

The used dataset is Diabetes UCI Dataset from kaggle which has been collected using direct questionnaires from the patients of Sylhet Diabetes Hospital in Sylhet, Bangladesh, and approved by a doctor. The dataset has 520 data and 17 attributes. To make the performance efficient, the data was checked for missing values. After checked, the used dataset doesn't have any missing values, label encoding was applied or transforming the instance of diabetes into numerical values e.g. 1 or 0 [2]. After that the dataset was divided into training data and testing data for both algorithms [3].

Several studies have been made in the last few decades and some of them show that Artificial Neural Networks (ANN) are one of the best algorithms for diabetes predictions and. XGBoost is one of the popular machine learning algorithms used for classification, because of that reason the writer wants to find out whether XGBoost can be used on diabetes prediction and comparing it with ANN [3], [7], [8].

1.2. Problem Formulation

1. Can XGBoost be used in diabetes prediction?
2. How is the performance of XGBoost compared to the Artificial Neural Network on diabetes prediction?

3. How do we improve XGBoost's performance?

1.3. Scope

The study focuses on diabetes prediction using XGBoost and Artificial Neural Network algorithms and comparing both of them.

1.4. Objective

This main objective is to compare the Extreme Gradient Boosting(XGBoost) algorithm and the Artificial Neural Network(ANN) algorithm performance on diabetes prediction.