1. Introduction

Background

Diabetes mellitus (DM) is a multifactorial disease, which is characterized by chronic hyperglycemia syndrome and disorders of carbohydrate, fat, and protein metabolism caused by insufficiency of insulin secretion or endogenous insulin activity, or both[1]. Chronic hyperglycemia is said to interfere with pancreatic beta cell function by triggering beta cell apoptosis, increasing intracellular calcium to cytotoxic concentrations, and resulting in increased synthesis of protein granules in beta cells including pro-insulin and pro-islet amyloid associated peptide – (ProlAPP) triggers endoplasmic reticulum stress[2]. Uncontrolled hyperglycemia can also cause many complications such as neuropathy, stroke, and peripheral vascular disease [1].

Diabetes mellitus (DM) ranks the fourth leading cause of death in developing countries[3]. In 2021 World Health Organization (WHO) reported an increase in the number of people with diabetes from 108 million in 1980 to 422 million in 2014. Between 2000 and 2016, there was a 5% increase in premature deaths from diabetes, and in 2019 it was estimated that 1,5 million deaths are caused directly by diabetes[4]. Indonesia is in seventh place for countries that have diabetes patients with as many as 10.7 million people[3].

Handling diabetes is done in two ways, i.e., the use of drugs and blood sugar-lowering therapy through the application of a diet. A person with diabetes will be given oral glycemic drugs (Oral Hypoglycemic Agents / OHA) to trigger insulin production. However, it has side effects with long-term use such as hypoglycemia and gastrointestinal disturbances[5]. Until now, no effective drug has been found to cure DM. Moreover, there are not many clues from preclinical and/or clinical trials because side effects are unexpected and only revealed on testing in intact biologic settings[6]. One solution that is being tried is to use an inhibitor of the Diacylglycerol Acyltransferase-1(DGAT-1) enzymes.

DGAT-1 is a microsomal enzyme that plays a central role in the metabolism of cellular glycerolipids. DGAT-1 catalyzes the final step in triacylglycerol (TAG) biosynthesis by converting diacylglycerol and fatty acyl-coenzyme A into triacylglycerol. DGAT-1 plays a fundamental role in the metabolism of cellular diacylglycerol and is important in higher eukaryotes for physiologic processes involving triacylglycerol metabolisms such as intestinal absorption, lipoprotein assembly, adipose tissue formation, and lactation [7].

Many drugs are usually developed using several trials which are costly, time-consuming, and fail to produce the proper results. In trait drug discovery theory, Quantitative Structure-Activity Relationship (QSAR) has been successfully developed to predict various important biopharmaceuticals, such as genotoxicity, toxicity, oral bioavailability, carcinogenicity, and mutagenicity [8]. The main objective of QSAR is to establish empirical rules to correlate the description of a chemical compound with its bioactivity. Multivariate modeling techniques have been widely used in QSAR studies such as Multiple Linear Regression, Partial Least Squares Regression, and Different Types of Artificial Neural Networks, Genetic Algorithms, and Support Vector Machines [8].

QSAR studies have been widely used to solve problems in DM. In 2019, Kumar et al. Conducted research on the design and development of novel focal adhesion kinase (FAK) inhibitors using Monte Carlo to validate QSAR. The value of r2 is 0.8398[9]. Then in 2020, Eduardo et al. Conducted research on drug reuse using QSAR, Docking, and Molecular Dynamics for possible SARS-CoV-2 Inhibitors. From 20 drug candidates, some of the best potential inhibitors with compound interaction rates were more than 50% [10]. Then in 2021, Vinicius et al. Conducting QSAR modeling of the Mpro SARS-CoV inhibitor identified other drugs as candidates for reuse against SARS-CoV-2. The results of the developed QSAR study obtained an accuracy of 98% [11].

In 2020, Kleandrova et. al. conducted cell-based multi-target QSAR model research for the design of virtual versatile inhibitors of liver cancer cell lines, where the use of QSAR opens a new horizon in the design of anti-cancer drugs compared to treating liver cancer using chemical therapy [12]. Then in 2021, Hammoudi et al. Conducted in-silico research on the discovery of Acetylcholinesterase and Butyrylcholinesterase enzyme inhibitor drugs based on QSAR and drug similarity. The resulting model has a high capacity, where 0.96 of the predicted compounds are outside the negative predictions for molecular weight and pH [13]. One of the challenges of QSAR is the determination of optimal features, which can be done using metaheuristic methods such as genetic algorithms and particle swarm optimization. based on the literature study, there are not many studies that implement the metaheuristic method in the QSAR Study in the case of DGAT-1.

In this study, we aim to develop a predictive model of the Diacylglycerol Acyltransferase-1 (DGAT1) inhibitor as anti-diabetes using a combination of Genetic Algorithms and Ensemble methods. A genetic algorithm is used to perform optimization in getting better features from a random initial sample[14]. The advantage of GA compared to optimization with the method differential is that GA can be used to determine the optimum conditions without Differentiating the data first. So for very complex data, optimization can be done easily[15]. The ensemble method is done by combining methods of machine learning for performance better than using one method. The purpose of the ensemble is to overcome the weaknesses contained in one classifier by using advantages over other classifiers. The ensemble method has proven to be able to improve classification performance[16].

Formulation of the Problem

The formulation of the problem from this research are:

1. What is the scheme and the effect of feature selection using a Genetic Algorithm?
2. What are the schemes and effects of building a predictive model using Ensemble Learning?
3. How is the performance of the Genetic Algorithm – Ensemble Learning in classifying?

Objectives

The objectives of this research are:

1. Implementing a Genetic Algorithm to perform Feature selection.
2. To find out the scheme and the effect of building a prediction model using the Ensemble Learning method to build a prediction model.
3. Knowing the performance of the Genetic Algorithm – Ensemble Learning method in classifying.