1. Introduction

Mosquitoes are the main vector of several diseases those attack humans and animals that cause thousands of deaths every year. Aedes aegypti is one of the most dangerous mosquitoes that can cause several diseases, such as dengue fever, Chikungunya, Zika, and jaundice [1],[2]. Dengue fever is considered one of the dangerous diseases caused by Aedes aegypti because the mortality rate is high and continues to increase every year [1]. The mortality rate of dengue fever has grown significantly worldwide at this time [1]. It is estimated that nearly thirty-nine million people worldwide are infected annually [1]. Symptoms of dengue fever are characterized by high fever accompanied by severe headache, muscle and joint pain, nausea or vomiting, and swollen glands [3].

Recently, no specific drug has been found that can cure dengue fever [3]. One possible solution for the treatment of dengue fever is to inhibit the growth and development of Aedes aegypti larvae. Several chemical products have been tested against larvae of Aedes aegypti such as phenolic acids, spinosyns, coumarins, et. Al [2]. Those compounds can help to inhibit the growth and development of Aedes aegypti larvae [2],[4],[5]. However, some chemical products are toxic and harmful to the environment [2]. Therefore, it is expected that the larvicides of plant products will be a source of raw materials and a safer alternative, which results in little waste, and is non-toxic to non-target organisms [6],[7], one of them is a group of larvicidal phytocompound [8]. It is known that the design of conventional drugs is not effective because the new compounds with certain biological activities need to be synthesized to determine their activity [9]. Hence, we need a model that can predict drug candidate activities, such as Quantitative Structure-Activity Relationship (QSAR).

QSAR is an alternative method developed for linking chemical molecules with activity biologically based on their chemical structure [10]. QSAR uses chemometric methods to describe the biological activity or nature of varying physicochemical properties as a function of a molecular descriptor that describes the structure of the chemical molecule [11]. Therefore, computed descriptors can be used to predict new compounds [11]. One of the challenges in QSAR study is to obtain optimal feature. One of the solutions to solve is use meta heuristic algorithm, such as Genetic Algorithm (GA) to select the optimal feature.

Several QSAR studies with and without meta heuristic have been done by several researchers. In 2012, A. Bahesthi performed QSAR modeling to analyze the activity of 68 urea derivatives as antimalarials using the GA-Multiple linear regression method [12]. The model validation was validated using external validation, namely leave-one-out (LOO) cross validation and y-randomization test, and obtaining the squares of the correlation coefficients R^2 0.801 and 0.803, respectively [12]. In 2017, Doucet, et. al. predicts the toxicity of a derivative of piperidine Aedes aegypti using QSAR models [13]. They predict the toxicity of 33 piperidine derivatives against Aedes aegypti [13]. The predicted toxicity was calculated using Ordinary Least Squares-Multi Linear Regression from QSARINS and Support Vector Machine (SVM), with the coefficient of determination (r^2) 0.85 and 0.80, respectively [13].

In 2020, Javidfar performed modeling larvicidal phytocompounds against Aedes aegypti using the index of ideality correlation [2]. They developed three QSAR models to predict pLC5062 plant-derived compounds to fight Aedes aegypti by method-based Monte Carlo on the IIC criteria, with the excellent predictive of the models ($r^2_{Val} = 0.856$ to 0.977) [2]. In 2020, Farisi Rahman, et. Al. carried out a QSAR model derived from Fusidic Acid as an Antimalarial Agent using the Simulated Annealing (SA) – SVM method [14]. The results showed that SA as a feature selection resulted in a satisfactory combination of features. Then, for the best validation results are generated by the RBF kernel [14].

In 2021, Fajar, et. al. predicts the activity of indenopyrazole derivatives as anti-cancer drugs using the QSAR model with the SA-SVM method, with three kernel models for SVM, namely the RBF kernel, linear kernel, and polynomial kernel [15]. Based on the three kernels, the RBF kernel produces an R^2 score train and the best test is 0.79 and 0.60, respectively [15]. Also, QSAR Model has been implemented to identify other disease [16],[17],[18],[19],[20]. However, to the best of our knowledge there is no report of the implementation of meta heuristic, such as GA, to select the features for the case of larvicidal phytocompounds.

In this study, we aim to build QSAR Model to predict larvicidal phytocompound activity as anti-Aedes aegypti with the Genetic Algorithm-Support Vector Machine methods. GA is generally a search-based algorithm built on the concept of natural selection and descendants [21]. GA is a subdivision of a much larger area of computing known as Evolutionary computing [21]. Meanwhile, SVM is a supervised learning technique that determines to classify different categories of data from different disciplines for classification problem solving and regression analysis [22].