1. Indroduction

Malaria is one of the world's deadliest endemic diseases. The majority of malaria cases were reported in African and tropical countries. There were an estimated 241 million infections and 627 thousand malaria deaths during 2020 in 85 different countries. These numbers increased from 227 million malaria cases and 558 thousand malaria deaths in 2019 [1]. Children under five years old accounted for up to 77% of the total malaria deaths in 2020 [1]. Plasmodium parasites are responsible for malaria infection in the human body. Infected female Anopheles mosquitoes transmitted Plasmodium parasite through bites. Several species of Plasmodium parasite are known to cause malaria, but the most prevalent species is *Plasmodium falciparum*, which dominates almost all malaria cases in every region [1]. *P. falciparum* is also the most dangerous Plasmodium species, which can give the patients the most severe and complex forms of malaria, such as cerebral malaria [1].

Currently, there is no effective vaccine available yet to prevent malaria (efficiency > 50%) [2]. On the other hand, artemisinin-based combination therapies (ACTs) are antimalarial drugs that have been successfully developed with the highest efficacy to date. Unfortunately, the parasite's resistance to most available antimalarial drugs has increased in recent years, particularly in *P. falciparum* species, making ACTs less effective [1], [2]. Between 2010 and 2019, malaria treatment failures due to drug resistance were detected in the Africa, America, Southeast Asia, and Western Pacific regions, with an average failure rate of 10% [1]. Thus, novel anti-malaria with better efficiency are urgently needed to encounter the resistance problem

Study shows that *P. falciparum* is a blood-feeding parasite capable of breaking down human hemoglobin into amino acids for its protein synthesis. The cysteine protease enzyme that plays a crucial role in hemoglobin degradation is falcipain (falcipain-2, falcipain-3) [3]–[5]. Furthermore, the inhibition of falcipain slowed the development of malaria by preventing the parasite from getting amino acids for its metabolism [5]. Hence, falcipain is a promising target for developing novel anti-malaria. In this regard, conventional laboratory research to study the activity of compounds to inhibit target protein usually requires a long time and is very expensive. Quantitative Structure-Activity Relationship (QSAR) can be an alternative solution to overcome conventional laboratory testing shortcomings. QSAR is a method commonly used to accelerate drug discovery by developing a predictive mathematical model to find the relation between biological activities and molecular properties of compounds [6].

QSAR studies to predict anti-malaria have been successfully conducted several times by other researchers. In 2018, Shehu and coworkers developed a QSAR model using a genetic algorithm (GA) to predict Pyrrolones bioactivity against malaria. This study achieved a QSAR model with R^2 and Q^2 values of 0.93 and 0.89, respectively [7]. In 2019, Hadni and Elhallaoui conducted a QSAR study on hybrids anilinoquinoline-triazines derivatives as antimalarial agents. They used multiple linear regression (MLR) and artificial neural networks (ANN) and performed good results with R^2 values for training and testing sets of 0.70 and 0.74, respectively [8]. Kurniawan and coworkers conducted a QSAR study in 2020 to predict the bioactivities of cycloguanil analogues as potent antimalarial agents. Partial least square regressions (PLSR) were used in this study, and the R^2 values for the training and test sets were 0.85 and 0.70, respectively, with a Q^2 score of 0.77 [9]. QSAR models have also been used to study other compounds and diseases [10]–[14].

The main challenge of implementing QSAR is finding the optimal combination of the molecular descriptors. In this regard, metaheuristic optimization, such as a genetic algorithm, can be used as the solution to this problem. However, according to previous studies, we found that only few researchers implement metaheuristic optimization as a feature selection method on falcipain inhibitors compounds. Therefore, this study aims to build a QSAR model to predict the activity of falcipain inhibitors as potent antimalarial drugs using a combination of genetic algorithm-support vector machine (GA-SVM) methods. We used GA as a feature selection algorithm because GA was proven to be a good optimization for machine learning models by selecting the best feature to achieve optimal results [15]. Finally, the QSAR model was constructed using SVM with optimized hyperparameters to predict the inhibitor activity values (pIC50). The SVM was utilized because it was proven to perform better for the QSAR model compared to other machine learning methods [16].