1. Introduction

Cancer is a medical condition characterized by the uncon- trolled proliferation of abnormal cells, which progressively damage healthy tissues in the body [1]. As these malignant cells continue to grow unchecked, they may invade normal cells, spread through the circulatory system, and affect various vital organs, leading to the emergence of multiple symptoms [2]. In 2014, approximately 1,665,540 individuals were diag- nosed with cancer, with 585,720 recorded deaths. Recent statistics indicate that nearly 1.96 million individuals are newly diagnosed with cancer, with more than 609,000 deaths linked to the disease. Between 2014 and 2019, following two decades of decline, cancer incidence has increased by approximately 3% annually [3]. Numerous approaches and pharmaceutical treatments have been developed to combat cancer. One of the most commonly employed methods is chemotherapy, which involves the use of agents that specifically target and eliminate cancer cells. However, chemotherapy may also damage healthy cells, leading to adverse side effects such as oral ulcers, nausea, and hair loss [2] [4]. Consequently, there is a critical need for the development of more targeted and less harmful anticancer therapies.

At present, a variety of drugs have been developed to target cancer cells, one of which involves matrix metalloproteinase-9 (MMP-9). MMP-9 is among the most structurally complex enzymes within the matrix metalloproteinase family and is classified under the gelatinase subgroup [5]. Structurally, MMP-9 is composed of three main domains: a propeptide domain, a catalytic core, and a fibronectin-like domain. The catalytic domain plays a crucial role, as it contains essential metal ions such as zinc and calcium. Due to its significant function in cancer progression, MMP-9 has become a primary target in the development of more potent and specific inhibitors [6].

Numerous strategies have been employed in the develop- ment of novel pharmaceutical agents, including traditional approaches. However, these methods are often considered time-consuming, costly, and relatively inefficient. Traditional drug development methods often demand substantial time and financial investment, and are frequently associated with a high likelihood of failure [7]. As a result, alternative strategies-particularly the use of machine learning methods-are gaining traction to speed up the drug discovery process and enhance its overall efficiency.

Several researchers have investigated the use of machine learning for anticancer drug development. In 2023, Banerjee et al. developed a predictive model for MMP-9 inhibitors using three classification methods: the Bayesian model, SARpy, and the RP model, which achieved accuracy scores of 87%, 78%, and 85%, respectively [6]. In 2015, Li et al. compared the performance of Support Vector Machine (SVM) and Random Forest (RF) for predicting MMP-3 and MMP-9 inhibitors, with RF and SVM achieving accuracy scores of 98.48% and 93.94%, respectively [8]. In 2022, Ikhsanurahman et al. used Simulated Annealing with SVM to model CDK2 inhibitors, reaching 0.986 accuracy and 0.987 F1-score [9].

Numerous machine learning methods have been utilized to forecast the bioactivity of MMP-9 inhibitors, and many investigations have shown encouraging levels of accuracy [6]. Nonetheless, further enhancement is possible- especially by optimizing feature selection techniques. Despite its potential, feature selection remains underexplored in this context [8]. Metaheuristic methods like Particle Swarm Optimization (PSO) offer an effective solution due to their fast convergence, flexibility, and ease of integration [10]. For predictive modeling, XGBoost is a suitable ensemble method, offering high speed, handling of missing data, regularization to reduce overfitting, and built-in feature importance and cross-validation capabilities [11].

This research seeks to construct a computational model for identifying potential MMP-9 inhibitors with anticancer properties, utilizing a hybrid PSO-XGBoost algorithm. The primary objective is to enhance the model's predictive accuracy by applying metaheuristic-based feature selection in combination with the XGBoost algorithm for robust prediction.