I. Introduction

Blood-brain barrier, commonly referred to as the BBB, is a membrane that separates the circulatory system from the extracellular fluid of the brain. [1]. The BBB is formed and protected by cells called pericytes and astrocytes [2]. The function of this membrane is to protect the central nervous system (CNS) from any potentially harmful foreign substances in the blood that are not needed by the brain while inhibiting drug transport to the brain [3]. This membrane has special features that make it very difficult for drug compounds to penetrate. The permeability of the blood-brain barrier has been a major consideration in drug design, particularly for Alzheimer's, Parkinson's and epileptic drugs. [4].

BBB penetration, also known as B3P, is currently measured using an *in vivo* approach. However, this approach requires large costs, complex operations, and time taken [5]. An alternative method that can be used is *in silico*. The *in silico* approach has received increasing attention and focus because it is considered to be able to quickly and reliably predict BBB permeability to molecules without requiring high costs [6]. One of the currently popular *in silico* methods is the implementation of machine learning. Research related to the implementation of machine learning on B3P has been carried out by several researchers. In 2012, Martins et al. implemented the Bayesian method in B3P prediction. Based on the results of their research, they found that the method gave results with the best model having an accuracy rate of 95% [1]. In 2019, Plisson et al. implemented 6 machine learning methods in B3P prediction. Based on the results of their research, they found that the Logistic Regression method gave the best results with an accuracy rate of 81% [7]. In 2020, Alsenan et al. implemented the Recurrent Neural Network method in B3P prediction. Based on the results of their research, they found that the method gave the best results with a 96.25% accuracy rate [5]. In 2021, Wu et al. implemented the Artificial Neural Network method in B3P prediction. Based on the results of their research, they found that the method gave the best results with an accuracy rate of 93% [6]. In 2023, Mazumdar et al. implemented the Deep Neural Network method in B3P prediction. Based on the results of their research, they found that the method gave results with an accuracy rate of 97.8% [8].

Based on the results of the literature survey, the results of the accuracy level using the *in silico* approach with the implementation of machine learning and deep learning produce very high accuracy values, even so these results can still be improved. Machine learning's

shortcomings include over-fitting, insufficient data, and difficulty in selecting control parameters appropriate to a given task. These weaknesses can be overcome by deep learning. Deep learning is one kind of machine learning that incorporates artificial intelligence in such a way that it mimics the way that human beings learn. This method is an essential part of the statistical and predictive analysis. The method in deep learning that is suitable for this research is Artificial Neural Network (ANN) which is the centre of deep learning [9]. ANN is a method that works similar to the human brain, having three main layers based on self-learning [10]. The artificial neural network (ANN) method has two main shortcomings. The first is the randomness of the initial weights and biases in the ANN training phase, which can result in suboptimal results. The second is the need to determine the optimal number of neurons in the hidden layer in order to adapt the neural network to each problem [9]. Both of these shortcomings are NP-hard problems, which are intractable and require significant computational resources to resolve. As a result, a non-traditional solution in the form of an optimizer is being sought to overcome them [9].

This study aims to utilize an ANN optimized by the Sine Cosine Algorithm to develop a B3P prediction model for drug candidates. Sine Cosine Algorithm (SCA) operates as an algorithmic optimizer, whereby a number of candidate solutions are randomly generated using a mathematical model based on sine and cosine functions. These candidate solutions are permitted to fluctuate in either an outward or inward direction, towards the optimal solution. This optimization serves to determine the number of neurons in the hidden layer and determine the initial weights and biases [9]. In addition, to emphasize the exploration and exploitation of the search space at the various optimization milestones, the algorithm incorporates a number of random and adaptive variables. [11].