ABSTRACT

Predicting side effects of the drug is one of pivotal processes in drug development, as drug molecules can inadvertently interact with non-target proteins, potentially resulting in adverse effects. Traditional prediction methods such as in vivo and in vitro approaches, have weaknesses related to safety protocols, costs, and efficiency. This study employs an in-silico approach, utilizing numerical tools and the Simplified Molecular Input Line Entry System (SMILES) to represent chemical compounds, simplifying chemical compounds, and optimizing computations. This research, aims to enhance predictive models using SMILES2Vec, a process that converts SMILES into numerical vectors for deep learning efficiency. Additionally, Long Short-Term Memory (LSTM) architecture selected to process text data, optimized using Monarch Butterfly Optimization (MBO). By comparing models with and without tuning, the insights of the performance are gained, especially in terms of accuracy and F1-Score. The results show that the model without tuning, using a 3-unit LSTM layer, achieves an accuracy of 62.94% and an F1-Score of 75.54%, while the model with tuning achieves an accuracy of 65.34% and an F1-Score of 75.91%. Similar improvements are observed in other architectures. Thus, the best-performing model with tuning is the convolution layer with LSTM, achieving an F1-Score of 75.94%. This study concludes that the model with tuning, combining SMILES2Vec with LSTM optimized with MBO, has an advantage in improving predictive performance compared to the model without tuning.

Keywords: drug, side effect, blood and lymphatic disorder, smiles2vec, lstm, monarch butterfly optimization, optimization algorithm