Application of Deep Learning and Fingerprint Modeling Methods to Predict Cannabinoid and Cathinone Derivatives

Widya Dwi Aryati[1], Gerry May Susanto[1], Muhammad Siddiq Winarko[1], Heru Suhartanto[2], Arry Yanuar[1*]

1. Faculty of Pharmacy, Universitas Indonesia, Depok, West Java, Indonesia
2. Faculty of Computer Sciences Universitas Indonesia, Depok, West Java, Indonesia

Presented on Pacific Rim Application and Grid Middleware Assembly 35 (PRAGMA 35) which be held on 3-6 October 2018 in Penang, Malaysia

ABSTRACT

Background: In recent years, new psychoactive substances (NPS) have rapidly emerged in market purportedly as "legal" alternatives to internationally controlled drugs, with the potential to pose serious health risks. 21 of 56 NPS were found as cannabinoid derivatives in 2016 in Indonesia. From 2013 to 2018 there was an increasing number of cathinone derivatives, 30 compounds in 2013 and 89 compounds in 2018. Artificial intelligence (AI) has become a tool for data processing and is applied for object recognition such as human pose and image classification.
Objective: The purpose of this study is to apply and gain the best AI method to classify new cannabinoid and cathinone derivatives by comparing deep learning method and fingerprint modeling method. The pharmacophore modeling was used as the reference method.
Methods: This study compared deep learning and fingerprint modeling methods. Both methods were compared with pharmacophore modeling as the reference method. Physicochemical property descriptor were used as learning parameters for the deep learning method. The two models produced by each method were used to classify new cannabinoid substances. As for the cathinone substances, the structure was transformed into a fingerprint form. This method was also compared with pharmacophore modeling as the reference method.
Results: Compared to the pharmacophore modeling method, the deep learning method for cannabinoids classification showed the higher accuracy and Cohen Kappa scores respectively (89.58% and 0,862) and (67,82% and 0,396) for pharmacophore modeling. Pharmacophore modeling in the classification of cathinone derivatives showed accuracy (91.11%) and Cohen Kappa scores (0.708). However, fingerprint modeling gave accuracy (71.8%) and Cohen Kappa (0.637).
Conclusions: These results conclude that the deep learning method with descriptor is a better instrument to be used for cannabinoid classification compared to pharmacophore modeling, but fingerprint modeling showed lower accuracy than pharmacophore modeling in cathinone classification.

Keywords: New Psychoactive Substance, Deep Learning, Fingerprint Modeling, Pharmacophore Modeling

INTRODUCTION

NPS was made by modificating chemical structure from substances that have already been banned by the government

↓

The consumption has been predicted to increase every year

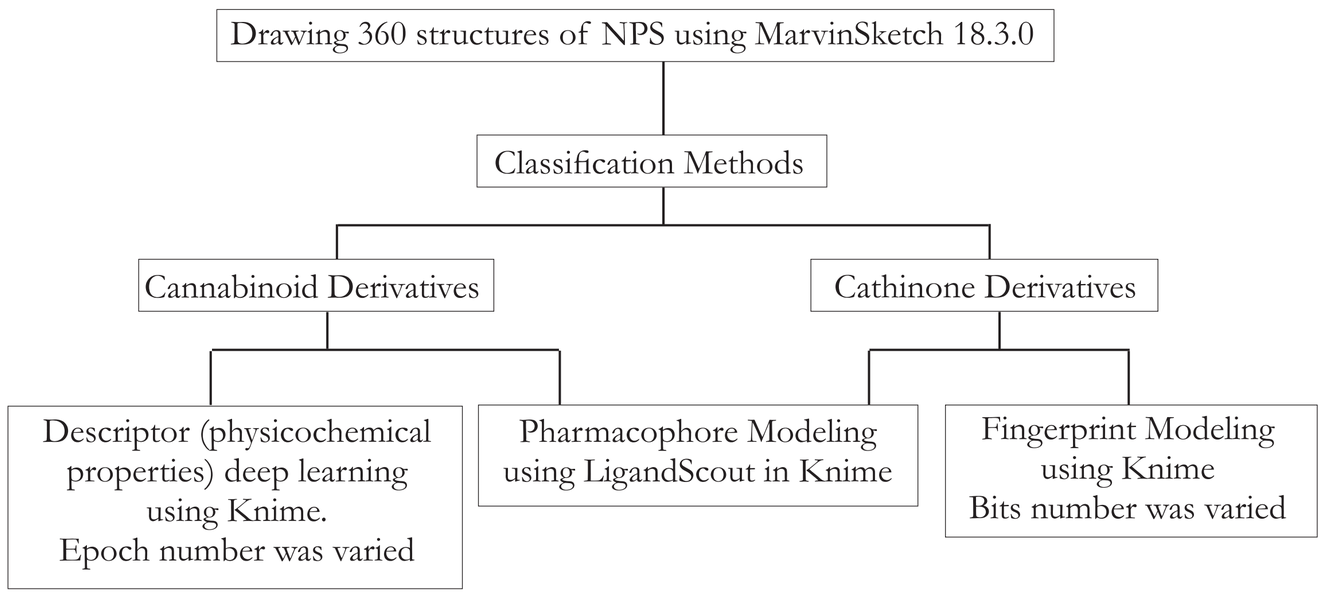
↓

An alternative method that can identify and classify the NPS is needed

OBJECTIVE

The purpose of this study is to apply and gain the best AI method to classify new cannabinoid and cathinone derivatives by comparing deep learning method and fingerprint modeling method. The pharmacophore modeling was used as the reference method.

METHOD



RESULTS AND DISCUSSION

Cannabinoid Derivatives

Pharmacophore Modeling

Table 1. Confusion Matrix of Pharmacophore Modeling of Cannabinoid Derivatives Classification

Database / Prediction	Active	Decoy	Total
Active	112	13	125
Decoy	99	124	223

The accuracy and Cohen Kappa score of classification with pharmacophore modeling method was 67.82% and 0.396, respectively. This is because the ligands used in the test set have diverse structure, so the data varied. The more identical the data in a class, it will have better accuracy. Therefore, the pharmacophore modeling couldn't differentiate the ligands and the decoys.[1]

Deep Learning

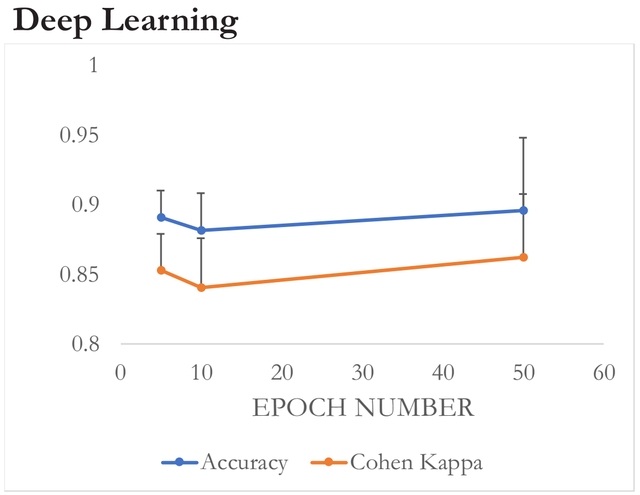


Figure 1. Comparison between Bits Number, Accuracy and Cohen Kappa Score

Figure 1 showed the relationship between epoch number, accuracy and Cohen Kappa score. It can be concluded that increasing the epoch number from 5 to 50 is affecting the accuracy and Cohen Kappa score. [2]

Canthinone Derivatives

Pharmacophore Modeling

Table 2. Confusion Matrix of Pharmacophore Modeling of Cathinone Derivatives Classification

Prediction / Database	Active	Decoy	Total
Active	44	0	44
Decoy	28	243	271

The accuracy and Cohen Kappa score of Classification with pharmacophore modeling method was 91,11% and 0,708, respectively. Table 2 showed the confusion matrix of this method. This method could perfectly predict 44 structures test set (cathinone) correctly, showed by the perfect score of sensitivity (1). However, this method predicted 28 non-cathinone compounds as cathinone, showed by 0.897 as the specificity score

Fingerprint Modeling

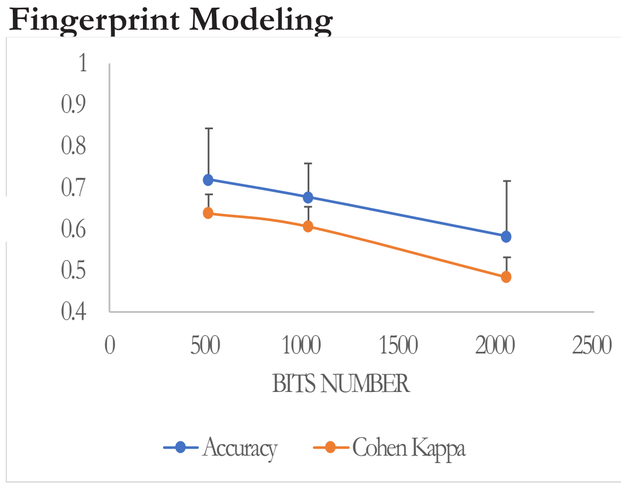


Figure 1. Comparison between Epoch Number, Accuracy and Cohen Kappa Score

Based on three configuration number of fingerprint bits in classification fingerprint modeling method, it could be concluded that higher fingerprint bits gave less accuracy score. This method shown lack of performance because The more data analyzed, it has the risk to decrease the prediction's accuracy.[3].

CONCLUSION

The deep learning method with descriptor is a better instrument to be used for cannabinoid classification compared to pharmacophore modeling, but fingerprint modeling showed lower accuracy than pharmacophore modeling in cathinone classification.

ACKNOWLEDGEMENT

Funded through a grant from the Universitas Indoensia for grant: Hibah PITTA 2018.

REFERENCES

1. Cho, J., Lee, K., Shin, E., Choy, G., and Do, S. (2015). How much data is needed to train a medical image Deep learning system to achieve necessary high accuracy
2. Batres-Estrada, G. (2015). Deep Learning for Multivariate Financial Time Series, 64-65
3. Freitas, A., & Lavington, S. (2000). Mining very large databases with parallel processing. Boston: Kluwer Academic Publishers.