International Journal of Computer
Science and Engineering (IJCSE)
ISSN (P): 2278–9960; ISSN (E): 2278–9979
Vol. 12, Issue 2, Jul–Dec 2023; 11–20

International Academy of Science,
Engineering and Technology
Connecting Researchers; Nurturing Innovations

AN INSIGHT REVIEW ON BIG DATA ANALYTICAL MODELS OF COMPUTATIONAL PHARMACOLOGY IN ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD)

Ramakrishnan Varadharajan [1], Dr. P. Anbalagan[2] & Dr. Saravanan M.S[3]

[1]Research Scholar, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamil Nadu, India

[2]Assistant Professor, Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar, Tamil Nadu, India

[3]Professor, Department of Computer Science and Engineering, Saveetha School of Engineering, Chennai, Tamil Nadu, India

ABSTRACT

Attention-deficit/hyperactivity disorder (ADHD) is the foremost neurodevelopmental disorder in childhood. Trouble in paying attention and controlling impulsive behaviour. Predominant hyperactive and Impulsive behaviour are the various symptoms of ADHD. Drug Analysis is the most cogent process in the pharmaceutical industry. The main aspect of this Review is to scrutinize vvarious computational methods and models that curtail the cost and time of drug analysis. Multiscale computational methods used in identifying the drug efficacy of an individual drug with combinations& dosage levels of drugs, half-life period, and level of drug ligand interactions to identify the immunological reactivity of drugs in the human body were conferred in this Review. Various algorithms, namely Rank Correlations and scoring Functions, Association Rules,Artificial Neural Networks (ANN), Quantitative Association Rule,Decision trees (DT), convolutional neural networks (CNN), K-nearest neighbours (KNN), random forests (RF), Support vector machines (SVM), Advances in Deep Learning, principal component analysis (PCA) and independent component analysis (ICA), the Apriori algorithm, hierarchical clustering, and Anomaly detection was launched for drug analysis. These were compared with the conventional techniques of online screening systems (e.g., Pharmacophore Modelling, QSAR and molecular docking) forligand-based, novo drug design. This proves that the machine learning algorithms accord more accuracy and desired results in time with rapid drug analysis. Future trends in drug design and screening will involve combining various methodologies to handle complex problems of many sizes and dimensions.

KEYWORDS: *Virtual Screening, Multiscale Model, Machine Learning, De Novo Design, ADHD, ANN, K-Mean, KNN, PCA, QSAR, Molecular Docking, Pharmacophore Modelling*

Article History

Received: *02 Dec 2023* | ***Revised:*** *20 Dec 2023* | ***Accepted:*** *31 Dec 2023*

1. INTRODUCTION

Drug design and screening have considerably profited from numerous computation approaches along with the fast-developing computer hardware, software, and algorithms. These technologies significantly cleavage the time of drug analysis. Most reliable Factors (Ouyang and Smith, 2015), such as identifying the drug efficacy of an individual drug,

combinations of drugs, the Dosagelevel of drugs, half-lifeperiod, and level of drug ligand interactions, identifying the immunological reactivity of drugs in the human body are scrutinized. ADHD is the foremost neurodevelopmental disorder in childhood. Trouble in paying attention and controlling impulsive behaviour. Predominant hyperactive and Impulsive behaviours are the various symptoms of ADHD. Children affected by this ADHD disease can behave abnormally in various situations. This will conquer difficulties in concentrating on particular work or never allowing a person to sit in one position, which means a state of wandering mind or an unconditional state of the brain and ADHD-related aberrant brain development. (Latifoğlu et al. 2022).8% of young adults between the age of 18-35 years of age have concentration difficulties focusing on particular tasks. The importance of this research study is to scrutinize the uses of machine learning or deep learning algorithms to predict better accuracy at the first stage of drug analysis. That recommends the patient consume the best medicines among all the available medicines to recover soon from illness (Rydzewski and Nowak, 2016). However, it was essential to analyze and compare numerous classification algorithms for better accuracy (Gautam et al.2023). Hence, this work aims to compare the accuracy of predicting the efficacy of the medicine viloxazine used in adults, which is a norepinephrine inhibitor that is utilized in the treatment of depression in adults. Various research has shown its virtue in treating ADHD. (Faraone et al. 2022). AI tools are used in the Drug Analysis Process (Brown, 2020), and Bioinformatics tools are also used in the phase of Drug product development (Chavda et al.2023)**.** Drug design and analysis processes can be applied in machine learning models (Malviya and Sharma, 2021) for analysing drug accuracy. The utilization of computerized modeling and algorithms in biology has improved drug discovery, leading to the creation of "computational drug analysis." This phrase refers to all fields and resources that can be used to create and find new medications for intricate disease systems, how medications influence targets in one or more diseases, and how drug-disease networks can be built in such systems. Computational methodologies' high throughput capabilities can enhance the results of conventional experimental methods. Drug target discovery computational methods range from statistical to machine-learning strategies. (Jamali et al. 2016). The goal of the algorithm, a subfield of Artificial Intelligence (AI), is to create and use computer algorithms that can learn from unprocessed, raw data in order to carry out specific tasks in the future. Within a huge data collection, classification, regression, grouping, or pattern recognition are the primary tasks carried out by AI systems (Carracedo-Reboredo et al. 2021). In the pharmaceutical industry, Numerous ML techniques have been applied for the prediction of biological activities, novel molecular properties, interactions, and side effects of medications. , Support Vector Machines Deep Neural Networks, Naive Bayes, Random Forests, and, more recently, are some examples of these techniques (Kong et al. 2022). Drug analysis has considerably profited through numerous computational methods with the quick development of computer algorithms. These technologies significantly cut the expense of drug production. These methodologies and drug production are becoming more and more dependent on artificial intelligence, such as machine learning and deep learning. (Gautam et al. 2023). The 1st International Consensus Statement on ADHD was produced by an international group of researchers about two decades ago. They sought to dispel the fallacy of the disease that affected and diminished the trustworthiness of healthcare professionals.

The majority of computational pharmacology uses conventional models for virtual screening and docking. However, Machine learning models of k-means, Rank Correlations and scoring Functions, Association Rules, ANN, Quantitative Association Rule and KNN, PCA and Apriori, Independent Component Analysis, Hierarchical Clustering, and Anomaly detections for the aforementioned findings. Hence, it is believed that the effectiveness of the drug analysis procedure can be hiked by wielding machine learning. This Review possesses current analytical developments from the last 25 years in order to update the International Consensus Declaration. The main grail of this Review is not to afford the

exhaustive manual for identifying and treating ADHD or an encyclopaedia of the condition. The only objective is to provide accurate, thorough drug responses through machine learning algorithms that will be used for treating ADHD.

2. MATERIALS AND METHODS

The Department of Artificial Intelligence Laboratory of Computer Science and Engineering, Annamalai University, carried out this research review study. The Machine Learning algorithms used here for this Review study may use two groups for every individual finding. The power analysis test is to be done using the g*power analysis tool for each group to get the required Sample size. Machine learning algorithms might be restricted by changing the data size 4-5 times instead of 104 times.

The dataset for the above individual findings in the drug analysis is extracted from the Kaggle or Reddit website and comprises sufficient rows and attributes. Some of the attributes are the type of disease, nature of the disease (Chronic/Acute), type of medicines used to treat, gender, age, weight of the patient, dosage level, response time, WBC Count, lipid profile, and so on. Two sets of algorithms were implemented using the downloaded dataset in each discovery to find greater accuracy than the traditional approaches. Typically, split training is done in an 80:20 ratio. The data model could be created using the algorithms utilized for data training, and the accurate result would be predicted grossly. The proposed model would be tested in anaconda. Spyder©. The minimum hardware requirement of a computer with Windows 10 or higher version is a processor, 256 HDD, and 8GB RAM.

2.1 Possible Findings in Drug Analysis

There are various findings could be identified for the drug to treat the disease where a minimum basic level given below.

Table 1: Different Possible Findings in Drug Analytics

Author Name	Findings	Descriptions
Al-Barhantoshy and Motaweh (2017)	Efficacy of the drug	Its percentage or the ratio of Ability to attain the desired results – can be obtained from libraries of patients' responses.
Bosch-Bayard et al. (2020)	Effects of the combination of drugs	how quickly they heal due to the consumption of combinations of drugs.
Brown (2020)	The appropriate dosage level of a drug	Appropriate dosage levels are to be given to patients ranging in different age group between child and adult.
Carracedo-reboreedoetal.(2021)	The half-life period of a drug	How long (in terms of time durations) the drug can be stable in the body.
Chavda et al. (2023)	level of Drug ligand interactions	How close it's binding with the targeted cells which can find using a scoring function.
Faraone et al. (2021)	Immunological reactivity or response of a drug	how well the body responds to the medicine.

2.2 Machine Learning Algorithms used for the above findings in Drug Analysis

2.2.1.Rank Correlations & Scoring Function

To find the efficacy of drugs from patients' response data. It assesses the degree of similarity between the ranks of different ordinal variables. In this direction, the strength of the variables can be known(Silverstein et al., 2020).

2.2.2. Association Rule

Apply to find the efficacy of the combination of drugs (i.e.) Butter/Jam -i.e., pan40 (used to control the effects of the medicine in the stomach) +painkiller to cure). It is an unsupervised technique used to test the associations among the data set variables. It can be divided into Eclat, apriori, and F-P Growth Algorithm. It works on support, lifts, and confidence metrics(Leejin et al., 2018).

2.2.3. Artificial Neural Networks & Quantitative Association Rule –

To find the Appropriate Dosage level of the drug. Artificial neural network solves the issues by the echoic function of the brain. The brain uses the previous matter to sort out a problem. A network of neurons builds a pile of neurons that sway classifications, new decisions, and predictions based on old events. This element mimics the neurons. It is of three types: input, output, and hidden layers. Artificial neural networks sake top robustness self-organization, which is extensively utilized in the evaluation, prevention, and detection of diseases.ANN was utilized to assess the physical properties of antibodies, including melting temperature and aggregation temperature(Li et al., 2020).

2.2.4. KNN —- (k-Nearest Neighbors)-used to Find the Drug Ligand Interactions) with the target protein cells of the human body affected. (how close the ligand is getting reached and cures.)k close neighbor is a simple, foremost technique that is combined and used with other selected machine learning. It was utilized for Further regression analysis. Usually, they are divided by their neighbor's moleculesfor the distribution of molecules to their closest neighbors. So, the value of k is the number of closest neighbors(Sagala et al., 2019).

2.2.5. KNN & RF(random Forest)

It is an apparel approach that builds various training examples. It is also used for classification and regression. However, it is impossible that RF over-fits the data, and it has also been utilized for toxicity modeling, bioactivity data analysis, drug prediction, etc. (Kaur et al., 2022).

2.2.6.CNNAlgorithms

To find the immunological response (Body Response-how quick it's responding / any anomalies or side effects for the drug patients consume.) differs from patient - patient. Convolution Neural Network is a deep neural algorithm that enables the deep learning of neural images. Input layers, hidden layers, and Output layers are the three types of layers in this algorithm. Detection of images, audio signals, and videos can be done through the CNN Algorithm(Muñoz-Organero et al., 2018).

2.2.7. DT(Decision Trees)

The Decision Tree is an algorithm that is utilized in auxiliary diagnosis and disease prediction. The classified mode to create the disorder, management decision-making, etc. Chemical framework and utilizing a DT for obtaining ano. of options effectively assist the chemistsin performing prescreening of toxic compounds(Magyary et al., 2002).

2.2.8. Support Vector Algorithm Machine

Is the foremost machine algorithm method that utilizes descriptors of molecules to assess QSAR and also deals with the high-dimensional dataset. Regression analysis and ANN are utilized to build nonlinear and linear models that are compared with SVM. In Nonlinear, it utilizes mapping of the nucleus for transforming high-dimension space. In the linear approach, it utilizes the mapping space points. Currently, it is used in drug discovery (Anuradha et al., 2010).

2.2.9. Advances in Deep Learning

This is a concept interrelated to artificial neural networks. When a descriptors molecule was not targeted, this selects data from high-dimensional and original structures. Hence, it allows deep learning in drug discovery. The neural network was utilized in the drug analysis, which includes protein engineering and gene analysis. It is the toolbox that has been a potential game-changer method with a rapid rise of deep learning concepts. The successful research of current studies indicated that it would be helpful in toxicity prediction, pharmaceutical research, genome mining, and chemogenomic applications in the pharmaceutical industry (Chen et al., 2019).

2.2.10. Scoring Function Machine Learning

Is used to rank the responses given by the patients, and Image result for machine learning algorithms uses the scoring function. Classical scoring is the essential model that builds a linear map between the binding label and protein–ligand (Li et al., 2020).

2.2.11. Principal Component Analysis

It is one of the popular algorithms used for the dimensional reduction of large data sets by converting them into large variables. It enables Multidimensional data visualization. It was invented by Karl Pearson in 1901. It is mainly utilized for exploratory analysis. These findings are evaluated by the degree of variance. The No. Of Combinations increases with No, of Dimensions. Standardization, Covariance Computation, and Identifying the major Components are the various steps incorporated in PCA (Rasmussen et al., 2002).

2.2.12. Independent Component Analysis

It encodes the Multivariate signal to independent signals. In this source, signals are independent. Normality, Distribution, and Complexity were the effects of signaling. These contribute to the basic implementation of ICA. It is used for blind separation of the signal. The general discovery of the ICA was given by Bernard. It is used for Astronomy, Cosmetology, Face Recognition, and Optimal assessment of neurons. Separation of various signals, Unsupervised Learning, Non-Parametric, and screening are the various advantages of ICA (Hoekzema et al., 2014).

3. RESULTS AND DISCUSSIONS

In this study, utilizing the Python platform and the RAM of a laptop refers to 8.00 GB, access data quickly an Intel® Core i9 Processors and Windows 11 for data analysis tasks. In Table 2 and Figures (1-4), we have summarized the performance of machine learning algorithms such as k-means, Rank Correlations and scoring Functions, Association Rules, ANN, Quantitative Association Rule and KNN, PCA and Apriori, Independent Component Analysis, Hierarchical Clustering and Anomaly detections in terms of accuracy, sensitivity, AUC, and specificity. From Table 2, RF has obtained 96.39, while others have a lower accuracy rate. CNN has 98.21 AUC, which is higher than other algorithms, while RF obtained a higher Sensitivity of 98.82 than others, and KNN specificity is 98.26, which is higher than other machine learning algorithms.

Table 2: Performance Outcomes

Methods	Accuracy (%)	AUC (%)	Sensitivity (%)	Specificity (%)
Rank correlations & scoringnction (Silversteinet al. 2020)	69.70	72.61	75.64	74.31
Association Rule (Leejin and Myoung, 2018)	70.67	74.40	77.35	76.52
ANN & Quantitative Association rule (Li et al. 2020)	73.87	76.67	79.47	77.69
KNN(Sagala, 2019)	93.14	92.42	94.68	98.26
RF (Kaur andMahlon, 2022)	96.39	97.75	98.82	96.67
CNN Algorithms(MunozOrganero et al. 2018)	94.67	98.21	96.37	97.82
DT(Mayary and Brandt, 2002)	75.89	78.88	81.13	78.16
SVM (Anuradha et al. 2010)	76.90	79.54	83.62	80.22
Advances in Deep Learning (Chen et al. 2019)	79.30	81.34	85.37	81.76
Scoring function (Li et al. 2020)	82.66	83.77	87.26	82.63
PCA (Rasmessen et al. 2002)	84.48	85.38	89.39	84.71
ICA (Hoekzema et al. 2014)	87.92	86.58	91.60	86.17
Apriori algorithm (TaiandChiu,2009)	88.04	88.65	92.71	88.19
Hierarchical clustering (Thaler et al. 2013)	89.98	90.25	93.64	90.79

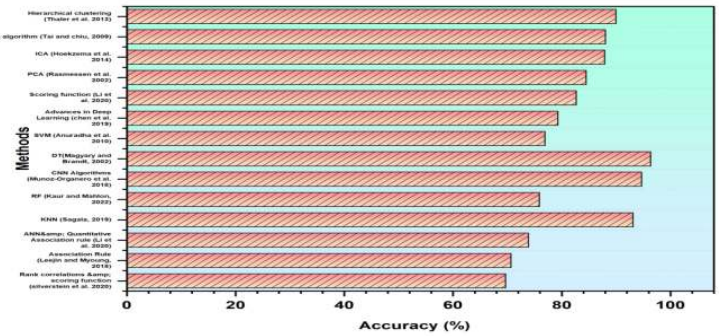

Figure 1: Accuracy.

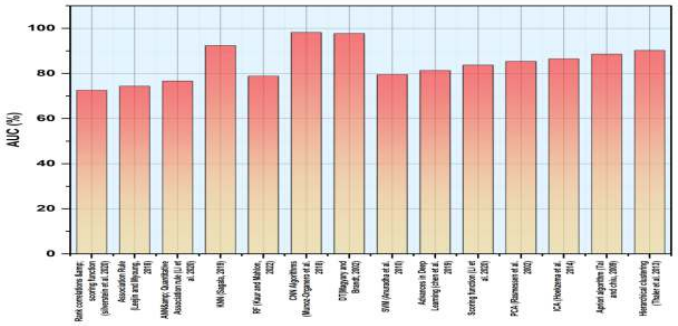

Figure 2: AUC

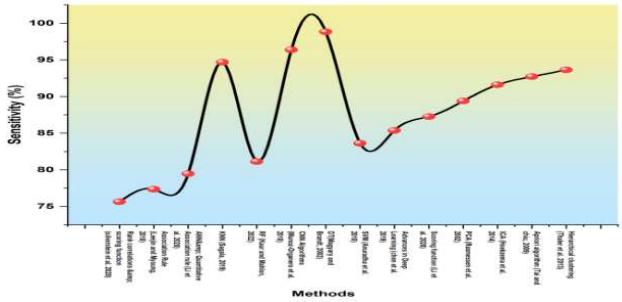

Figure 3: Sensitivity

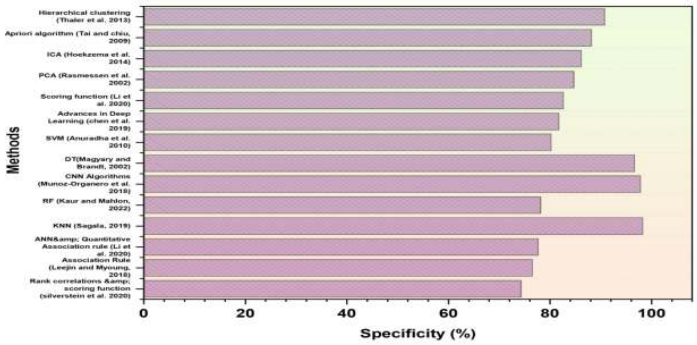

Figure 4: Specificity

3.1. DISCUSSION

Each of the mentioned machine learning techniques has its own limitations when applied to the field of Computational Pharmacology in ADHD. Rank correlations and scoring functions may struggle to capture complex nonlinear relationships in high-dimensional data. Association rule mining and Apriori algorithms are primarily designed for transactional databases and may not directly apply to the unique data structures found in ADHD research. ANN, while powerful, often lacks interpretability, making it challenging to understand the biological underpinnings of ADHD. KNN can be sensitive to distance metrics and struggle with high-dimensional data, while RF can be computationally expensive and prone to overfitting. CNNs are tailored for image data and may not effectively capture the spatial or temporal dependencies inherent in ADHD data. DT may fail to capture complex variable interactions or provide deep insights into ADHD's biological processes. SVMs may require additional techniques for multi-class or regression problems in ADHD research, and deep learning methods often demand large data volumes, which are often lacking in this domain. PCA and ICA may reduce

dimensionality but may not reveal specific biological markers associated with ADHD, and hierarchical clustering's effectiveness depends on data quality and characteristics. While each technique has its merits, KNN, CNN, and DT can be among the most suitable options when appropriately adapted to the specific challenges of ADHD data, but the choice depends on the nature of the data and the research objectives.

4. CONCLUSIONS

The contingent scrutiny of the literature exhibits that utilization of the Big Data Analytical algorithms in the Computational Pharmacology of ADHD disorder limits the time and cost of new insights. Various algorithms, namely KNN, CNN, and DT, Were found to be more effective algorithms in drug computation and compilation. SPSS software is found to be an effective tool for statistical analysis. In future endeavors, a novel drug analytical model can be used for various drug analyses and interpretations in treating more no. diseases.

5. REFERENCES

1. *Al-Barhamtoshy, H. M., & Motaweh, D. M. (2017). Diagnosis of Dyslexia using computation analysis. In the 2017 International Conference on Informatics, Health & Technology (ICIHT). https://doi.org/10.1109/iciht.2017.7899141*
2. *Bosch-Bayard, J., Girini, K., Biscay, R. J., Valdes-Sosa, P., Evans, A. C., & Chiarenza, G. A. (2020). Resting EEG effective connectivity at the sources in developmental dysphonetic dyslexia. Differences with non-specific reading delay. In International Journal of Psychophysiology (Vol. 153, pp. 135–147). https://doi.org/10.1016/j.ijpsycho.2020.04.021*
3. *Brown, N. (2020). Artificial Intelligence in Drug Discovery. Royal Society of Chemistry.*
4. *Carracedo-Reboredo, P., Liñares-Blanco, J., Rodríguez-Fernández, N., Cedrón, F., Novoa, F. J., Carballal, A., Maojo, V., Pazos, A., & Fernandez-Lozano, C. (2021). A review of machine learning approaches and trends in drug discovery. Computational and Structural Biotechnology Journal, 19, 4538–4558.*
5. *Chavda, V., Anand, K., & Apostolopoulos, V. (2023). Bioinformatics Tools for Pharmaceutical Drug Product Development. John Wiley & Sons.*
6. *Faraone, S. V., Banaschewski, T., Coghill, D., Zheng, Y., Biederman, J., Bellgrove, M. A., Newcorn, J. H., Gignac, M., Al Saud, N. M., Manor, I., Rohde, L. A., Yang, L., Cortese, S., Almagor, D., Stein, M. A., Albatti, T. H., Aljoudi, H. F., Alqahtani, M. M. J., Asherson, P., … Wang, Y. (2021). The World Federation of ADHD International Consensus Statement: 208 Evidence-based conclusions about the disorder. Neuroscience and Biobehavioral Reviews, 128, 789–818.*
7. *Faraone, S. V., Gomeni, R., Hull, J. T., Chaturvedi, S. A., Busse, G. D., Melyan, Z., O'Neal, W., Rubin, J., & Nasser, A. (2022). Predicting the efficacy of viloxazine extended-release treatment in adults with ADHD using an early change in ADHD symptoms: Machine learning Post Hoc analysis of phase 3 clinical trial. Psychiatry Research, 318, 114922.*
8. *Gautam, R. K., Kama, M. A., & Mittal, P. (2023). Computational Approaches in Drug Discovery, Development and Systems Pharmacology. Elsevier.*

9. *Jamali, A. A., Ferdousi, R., Razzaghi, S., Li, J., Safdari, R., & Ebrahimie, E. (2016). DrugMiner: a comparative analysis of machine learning algorithms for prediction of potential druggable proteins. Drug Discovery Today, 21(5), 718–724.*

10. *Kong, W., Midena, G., Chen, Y., Athanasiadis, P., Wang, T., Rousu, J., He, L., & Aittokallio, T. (2022). A systematic review of computational methods for drug combination prediction. Computational and Structural Biotechnology Journal, 20, 2807–2814.*

11. *Latifoğlu, F., Esas, M. Y., İleri, R., Özmen, S., & Demirci, E. (2022). Using electrooculography with visual stimulus tracking test in diagnosing of ADHD: findings from machine learning algorithms. Turkish Journal of Medical Sciences, 52(5), 1616–1626.*

12. *Malviya, R., & Sharma, A. (2021). Applications of Computational Methods and Modeling in Drug Delivery. In Machine Learning and Analytics in Healthcare Systems (pp. 163–190). https://doi.org/10.1201/9781003185246-9*

13. *Manilla, G. T., & de Braga, J. (2017). A New Dyslexia Reading Method and Visual Correction Position Method. In Global Pediatric Health (Vol. 4, p. 2333794X1773409). https://doi.org/10.1177/2333794x17734096*

14. *Ouyang, D., & Smith, S. C. (2015). Computational Pharmaceutics: Application of Molecular Modeling in Drug Delivery. John Wiley & Sons.*

15. *Pleša, I., & Mazure, L. (2017). THE CHILD'S AND ITS PARENTS' LEGAL STATUS AFTER CHILD'S PLACEMENT IN THE BABY BOX. In INDIVIDUAL. SOCIETY. STATE. Proceedings of the International Student and Teacher Scientific and Practical Conference (p. 222). https://doi.org/10.17770/iss2017.3011*

16. *Rello, L., Baeza-Yates, R., Ali, A., Bigham, J. P., & Serra, M. (2020). Predicting risk of dyslexia with an online gamified test. PloS One, 15(12), e0241687.*

17. *Rydzewski, J., & Nowak, W. (2016). Machine Learning Based Dimensionality Reduction Facilitates Ligand Diffusion Paths Assessment: A Case of Cytochrome P450cam. Journal of Chemical Theory and Computation, 12(4), 2110–2120.*

18. *Yang, Yang, Y., Tam, F., Graham, S. J., Sun, G., Li, J., Gu, C., Tao, R., Wang, N., Bi, H., & Zuo, Z. (2020). Men and women differ in the neural basis of handwriting. In Human Brain Mapping (Vol. 41, Issue 10, pp. 2642–2655). https://doi.org/10.1002/hbm.24968.*

19. *Silverstein, M.J., Faraone, S.V., Leon, T.L., Biederman, J., Spencer, T.J. and Adler, L.A., 2020. The relationship between executive function deficits and DSM-5-defined ADHD symptoms. Journal of attention disorders, 24(1), pp.41-51.*

20. *Leejin, K.I.M. and Myoung, S., 2018. Comorbidity study of attention-deficit hyperactivity disorder (ADHD) in children: applying association rule mining (ARM) to Korean National Health Insurance Data. Iranian Journal of public health, 47(4), p.481.*

21. *Li, D., Yang, D., Zhang, J., and Zhang, X., 2020. Ar-ann: incorporating association rule mining in an artificial neural network for thyroid disease knowledge discovery and diagnosis. IAENG International Journal of Computer Science, 47(1), pp.25-36.*

22. *Sagala, M., 2019. Algorithm Modified K-Nearest Neighbor (M-KNN) for Classification of Attention Deficit Hyperactive Disorder (ADHD) in Children. Login: Jurnal Teknologi Komputer, 13(1), pp.11-18.*

23. *Kaur, A. and Kahlon, K.S., 2022. Accurate identification of ADHD among adults using real-time activity data. Brain Sciences, 12(7), p.831.*