Journal of Theoretical and Applied Information Technology
15th September 2024. Vol.102. No. 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

# EXPLOITING THE EFFICACY OF THE DRUGS USED FOR ATTENTION-DEFICIT/HYPERACTIVITY DISORDER (ADHD) TREATMENT IN ADULTS AND CHILDREN USING A NOVEL BIG DATA-DRIVEN TIME-DEPENDENT FLEXIBLE DEEP RECURRENT NETWORK MODEL (TDF-DRN)

**MR. RAMAKRISHNAN VARADHARAJAN[1], DR. P. ANBALAGAN[2], DR. SARAVANAN M.S[3]**

[1]*Research Scholar, Department of Computer Science and Engineering, Annamalai University, India.*
[2]*Assistant Professor, Department of Computer Science and Engineering, Annamalai University,* India.
[3]*Professor, Department of Computer Science and Engineering, Saveetha School of Engineering,* Chennai, India.
[1]vramki2007@gmail.com, [2]anbalagansamy@gmail.com, [3]saravananms@saveetha.com

**ABSTRACT:**

Attention-deficit/hyperactivity disorder (ADHD) is a prevalent neurodevelopmental condition characterized by age-inappropriate symptoms of inattention and hyperactivity, often persisting into adulthood for a majority of diagnosed individuals. The primary objective of this study was to assess the effectiveness of combination medications in treating patients with ADHD, encompassing both children and adults. A comprehensive approach was taken, involving the collection of 1500 records from 740 trials, including patient responses. Rigorous data cleaning and partitioning were achieved using the Map Reduce method to eliminate duplicates and address missing values. Moreover, the Recursive Feature Elimination (RFE) technique was implemented to enhance machine learning model performance by removing less significant features from the dataset. To evaluate the efficacy of drug treatment for ADHD in both adults and children, a proposed time-dependent flexible deep recurrent network (TDF-DRN) technique was utilized. The preliminary findings suggested that combination medication therapy could be a successful strategy for managing ADHD symptoms across different age groups. Notably, the study revealed that optimal medication combinations and dosages varied depending on the age group, with children benefiting from a range of combination therapies and generally requiring lower dosages than adults. Additionally, the study conducted a thorough assessment of the safety considerations and potential adverse effects associated with different medication combinations.
***Keywords:*** *Attention Deficit/Hyperactivity Disorder (ADHD), Combinations of Drugs, Recursive Feature Elimination (RFE), Time-Dependent Flexible Deep Recurrent Network (TDF-DRN)*

## 1. INTRODUCTION

***1.1 Background of the study:*** The symptoms and signs of Attention Deficit Hyperactivity Disease (ADHD) are varied, including hyperactivity, daydreaming, distractibility, and impulsivity. Children with ADHD often exhibit these symptoms for longer periods than their peers, making it challenging for them to focus and control their impulses. Interestingly, children of Mexican descent have a lower prevalence of ADHD compared to other racial or ethnic groups [1]. Additionally, boys are diagnosed with ADHD three to five times more often than girls. It's worth noting that at least fifteen to twenty percent of children diagnosed with ADHD continue to exhibit symptoms into adulthood. ADHD is often accompanied by stress, sadness, resistance to change, and drug use disorders, affecting around 70% of cases. To diagnose ADHD, signs of inattention should begin before the age of

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

twelve and significantly impact psychological functioning, social interactions, academic performance, and vocational performance [2]. Figure 1 illustrates the symptoms of ADHD in children.

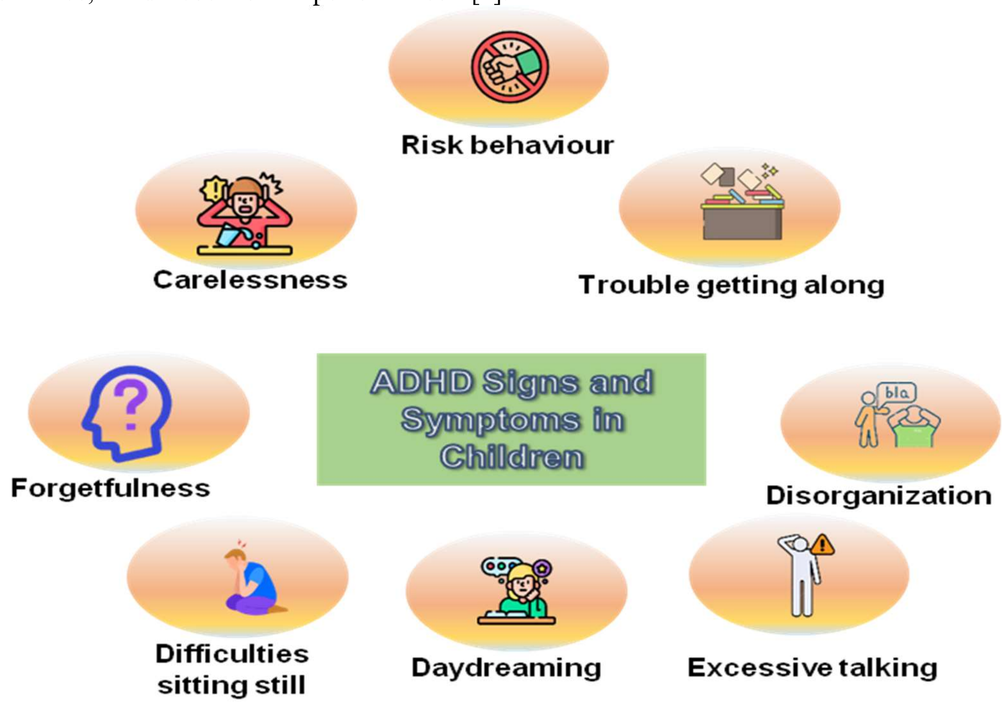


*Figure 1: ADHD signs in Children*

### *1.2 Identifying Young Children with ADHD*

Diagnosing ADHD in very young children presents a significant challenge, although symptoms of the disorder often appear in adolescents or older children. As a result, toddlers and preschoolers showing signs of ADHD will likely need to be evaluated by a specialized professional, such as a developmental pediatrician, psychiatrist, or psychologist [3].

### *1.3 Drug Efficacy for ADHD*

The eTNS device is intended for children aged 7 to 12 with ADHD who are not currently using prescription medication. The gadget administers minimal electrical stimulation to the child's forehead and stimulates microscopic areas. it's important to discuss expectations, safety measures, and potential negative effects. Medication effectiveness is a crucial part of the comprehensive treatment approach for individuals with this neurological condition[4]. Pharmaceutical treatments, including stimulant and non-stimulant medications, are primarily used to manage ADHD symptoms. Health professionals often prescribe stimulants such as amphetamine derivatives to enhance neurotransmitter activity, including norepinephrine and dopamine, improving cognitive function and impulse control [5]. It is essential to understand the dosage-dependent effects, recognizing that the optimal balance varies for adults and children. Variables such as developmental stage, metabolic differences, and potential long-term effects on cognitive function should be considered when developing effective treatment plans[6] .

### *1.4 Impulsivity and Hyperactivity*

Children under 17 must display at least 5 signs of hyperactivity-impulsivity, while individuals aged 17 and older should exhibit a minimum of six signs. These signs must have been present for at least a month and must disrupt the individual's normal development. They fidget and squirm when expected to be seated, and they often move around or interrupt when it's not appropriate [7]. As shown in Figure 2, inflammation can be observed in ADHD.

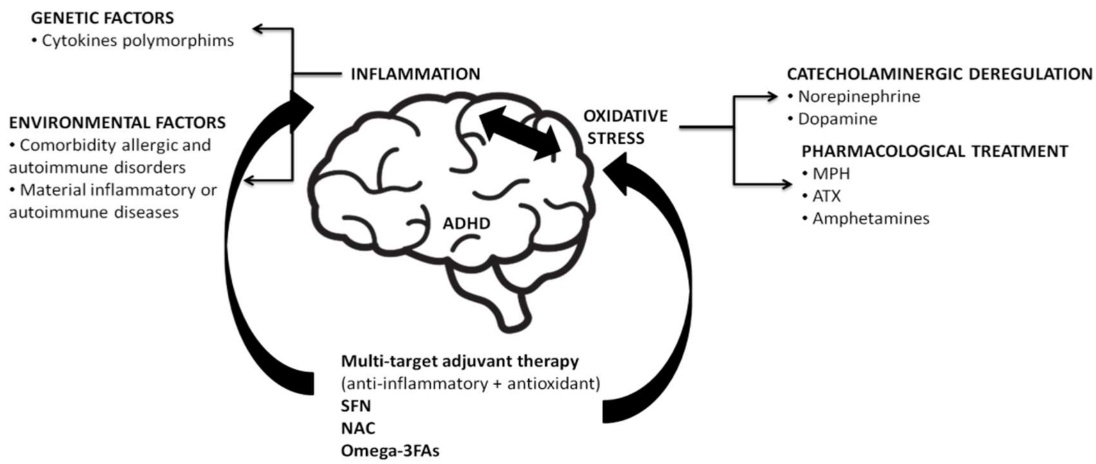


*Figure 2: Function of inflammation in ADHD*

The main symptoms of ADHD are correlated with problematic effects such as adolescent and adult drug use, risky sexual behavior, and criminal behavior. People exhibiting hyperactivity-impulsivity-inattention and conduct difficulties are at a higher risk of engaging in persistent criminal activity, particularly involving antisocial and transferring behaviors. However, a closer examination of individual symptoms and subtypes reveals distinctions, particularly concerning impulsivity and co-occurring psychopathology [8] .

The study aims to assist medical professionals in creating personalized, highly effective drug treatments for ADHD. Gaining a deep understanding of the effectiveness of combined medications in treating ADHD ultimately contributes to the broader goal of enhancing the quality of life for individuals affected by this complex neurodevelopmental disorder at any stage of development.

**Key contributions**

- An essential addition is the preparation of data using the Map Reduce methodology.
- This assists in providing a targeted and pertinent set of characteristics for estimating the effectiveness of ADHD treatment.
- According to the study, treating ADHD symptoms in both adults and children can be accomplished by combining drugs and other therapies.
- The study employs a comprehensive set of evaluation metrics to assess the model's efficacy, including Sensitivity, Specificity, Precision, Accuracy, AUC Curve, F1 Score, Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

In the paper's second section, we integrated a review of existing literature to provide context and understanding. Section 3 delves further into the methodology. Section 4 offers a thorough assessment and analysis of the findings. In section 5, the importance of the conclusion is extensively examined.

## 2. RELATED WORKS

The study assessed the tolerability, long-term safety, and effectiveness of cannabinoids in pediatric epileptic patients. Concomitant antiepileptic medication trough concentrations were evaluated at baseline, one, two, and three months into therapy, and as further clinically warranted [9]. The paper considered meta-analyses of biofeedback applied to children with ADHD. A new meta-analysis of randomized controlled trials (RCTs) was carried out, updating earlier findings and including methodological changes [10] . The study suggested a novel deep learning-based automated diagnosis methodology to distinguish between children who have two distinct types of ADHD, namely I EEG signals of healthy children [11]. A common neurnattentive ADHD and Combined ADHD, utilizing the development issue impacting many children is ADHD. Accordingly, computerized ADHD diagnosis provides a lot of potential benefits. The paper focused on identifying the most essential characteristics and automating the diagnosis process with existing classification approaches for better diagnosis. ASD is a neurodevelopment disease that is accompanied by sensory problems, such as excessive or insufficient Sensitivity to touch, noises, or odors [12]. The study involved the development of a virtual reality classroom that was integrated with persistent and focused attention activities. The attention tasks had to be carried out in the presence of visual-audio hybrid distractions [13]. The paper also focused on meta-analyses that have indicated that transcranial electrical stimulation can reduce clinical signs and enhance mental skills like memory retention and attention, which are compromised in ADHD. While pharmaceuticals account for the majority of treatments for ADHD, tES is becoming more and more popular as a substitute strategy [14] .

The study used the power spectrum, bicoherence, complexity, and biomarker candidates for detecting ADHD youngsters in a machine learning technique. Characterizing the EEG of children with ADHD is one of the critical objectives of this informative neuroimaging technology[14], [15]. One primary objective of the EEG, a proper neuroimaging technique for researching ADHD, is to describe the EEG characteristics of children with ADHD. The study proposed to perform an auditory oddball task, and Event-Related Potentials from EEG signals were analyzed over time and frequency. According to the information gathered, a machine-learning model was created to distinguish between patients with ADHD and healthy control subjects [16]. The paper discussed the neurodevelopment illness known as ADHD, which is characterized [16], [17]d by impulsivity, hyperactivity, and inattentiveness. The study examined whether a comparable tendency would hold in natural home and school environments by following study participants into the next academic year. According to a study done in an identical summer therapy environment, many children with ADHD who also received behavioral intervention did not require medication or very low doses of it to enhance Sensitivity [18]. The study assessed the interpolated dataset's capacity to discern between young people suffering from and without ADHD. It accomplished this by using a deep learning technique to fill in missing data in ADHD rating scales [18], [19]. The paper focused on the diagnosis of ADHD, a neurodevelopment illness with a wide range of manifestations, based on objective accounts of symptoms. The establishment of neuroimaging-based diagnostics for the diagnosis of ADHD has benefited from the use of machine learning classifications [20]. The paper presented a comprehensive approach to diagnosing the mixed form of ADHD. In youths, ADHD occurs more than any other neurobehavioral disorder. Unfortunately, there are no established diagnostic approaches that are dependable and affordable because their cause is unclear [21]. The study found that stimulant drugs are the gold standard for treating ADHD. However, these medications do not work for every patient, and they can have adverse effects and be misused for reasons other than medical. Adults with ADHD endure personal anguish, and the economy suffers as a whole [21], [22]. The study focused on an ANN-based Clinical decision support system (CDSS) to study the effectiveness of neurofeedback (NF) in the treatment of ADHD. To aid physicians in making more informed decisions regarding their patients, CDSS analyzes raw data and turns it into useful information [23].

## 3. MATERIALS AND METHODS

The data collection, data preprocessing, and feature extraction phases make up the first part of the method's multiple sections. We introduced a novel approach called the time-dependent flexible deep recurrent network (TDF-DRN) technique in this study, aiming to evaluate proposed model's effectiveness of the drugs classifications for treating ADHD in both adults and children. Figure 3 demonstrates the proposed method flow.

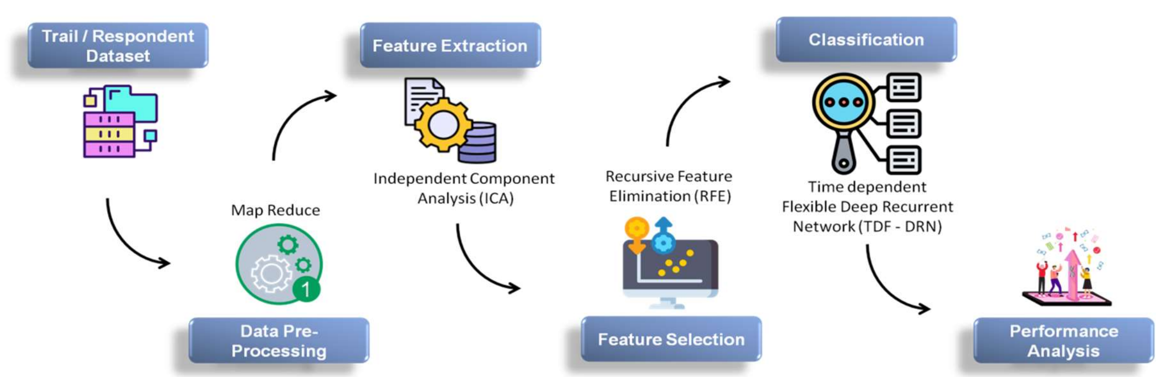


*Figure 3: Block diagram of the proposed method*

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

### *3.1 Data collection*

This data is organized into 1500 rows and 728 columns. An assessment status is usually considered high or excellent. The digit 839,999 seconds could indicate an examination or assessment that lasted around 839,999 seconds if they represent a period in seconds. Based on a 24-hour day, this time is translated into approximately 9.72 days. There will be a significant amount of time spent on the assessment. They could stand for a particular category or kind in a system of classifications. It denotes any topic that is up for review, such as a specific class, group, assessment type, technique, product, etc. Lower-level outcomes or measurements might be represented by numbers like 0, 1, 2, and 3. Results or measures with values such as 10, 11, 12, 13, 14, and 15 can be moderate to high. A common occurrence in datasets is the use of the placeholder value "-1" to denote unavailable or missing data or a specific condition code. It might stand for instances where the data was not collected, the response was irrelevant, or neither of those things.

In most cases, the positive numbers stand for recorded data, answers, or measurements. The study or experiment that yielded these values has influenced. These positive integers and the unit of measure would have different meanings depending on the domain of the study. An entire dataset called Last Trial has its values set to 360.

### *3.2 Data Preprocessing using Map Reduce*

Data preparation is a crucial step in the process chain allowing the distributed processing during massive databases using the Map-Reduce computing style. Data cleansing, feature extraction, duplicate item removal from datasets, format conversion, and many other procedures are examples of data preprocessing steps. When processing large datasets, Hadoop Map Reduce offers the appropriate framework for carrying out various operations in parallel. Processing several heterogeneous datasets is made possible by the Map-Reduce paradigm. Where $\alpha$ and $\beta$ stand for dataset lineages, $r$ for variables, and $w$ for variables entity as shown in Equation (1-2).

$$\text{Map: } (r_1, w_1)_\alpha \rightarrow [(r_2, w_2)]_\alpha \quad (1)$$

$$\text{Reduce: } (r_2, [w_2])_\alpha \rightarrow (r_2, [w_3])_\alpha \quad (2)$$

The map function transforms an input key and variables couple $(r1, w1)$ into a list of intermediate key/variable pairs $[(r2, w2)]$. The compilation of items is aggregated using the reduction method $[r2]$connected to $r2$ and generates an amount collection $[w_3]$, resulting in a connection to $r2$. It ought to remain mentioned that each of the operations' inputs and results belong to the equivalent category α. For mathematical connections, the merging process yields a self-merge if α = β, a procedure analogous to a self-adoption. Observe how nearly identical the Map and Reduce signatures in the new model are to those in the old Map Reduce. The dataset's ancestries and the fact that reduce produces a key/variables list rather than simple integers are significant variations. Since the decreased outcome in Google Map Reduce is final, consumers can include items they need in $[w_3]$, and sending $r_2$ for the next step is not necessary. This is to ensure that before the data can be correctly combined, partitioning it first and sorting it to implement duplicate keys are required. However, it provided the data that the altered variables correspond to remains structured identically, as the data mapping variables represented expressed in $r_2$, values continue to be converted across phases, and they even have been considered identical according to concept.

### *3.3 Feature extraction using Independent Component Analysis (ICA)*

One technique for extracting features from multivariate random signals that can be used to transform these features into greater significance features is Independent Component Analysis (ICA). Each observable signal is considered to be a linear combination of an equal number of independently floating indicators that have been measured or identified. Assuming $y = [y_1, y_2, \dots, y_i, \dots,]$ ym consists of m linear permutations, which are linear combinations. The combinations are the result of the m-way synthesis of separate linear components in Equation (3).

$$x_j = a_{j1}s_1 + a_{i2}s_2 + \cdots + a_{jn}s_n \quad (3)$$

Applying the symbol $a_k$, the n-way split could be broken down into its parts. Assuming a zero mean for the mixing variables and the independent components does not reduce generalizability. We can express the combinations using vector notation as below.

$y = [y_1, y_2, \dots, y_i, \dots, y_m]$ as y.

The independent variables are denoted by the $g$ in this context. The $g$ is used to represent the independent variables in this situation. This paradigm is known as ICA. It is possible to estimate the inverse of matrix $E$ and find its independent components. The ICA procedure is as follows: We can assume that the separate parts and mixing variables have mean values of 0 without limiting our generalizability, where ak is the total number of elements we desire. If we use vector notation to describe these combinations, we can refer to them as $x$. The following are the measures to take:

**Step 1:** Data normalization involves removing the Mean from the data to make it more consistent. So ensures that the sum of the parts is never less than zero.

**Step 2:** **S**treamlining Data handling the data preparation differently. To achieve whiteness, one must manipulate the mixture in that the components no longer interact with one another, and their disparities cancel each other out.

**Step 3:** A certain amount of independence is necessary for the analysis of the data. Many more instances exist, such as the Fast ICA method, ICA with non-Gaussianity maximization, ICA with variance maximization or reduction, ICA with negative entropy, and numerous additional instances. Here, the Negentropy approach was employed.

**Step 4:** Reconstruct the information. To get the result, we increase the whitening output by the result we got after applying the independence requirement. To get the output, multiply the supplied data by the transpose.

### *3.4 Feature selection using Recursive Feature Elimination (RFE)*

Our feature selection approach consists of two steps. RFE is a feature selection strategy that uses machine learning performance to eliminate features based on their relevance. To decrease the quantity of components utilized for model training, RFE is applied. We choose the number of features to use from the original parts which is why RFE is chosen. As a wrapper approach, RFE fits a learning model and eliminates less essential elements. This process is known as RFE. The model's coefficients, or feature importance properties, are used for evaluating the features. RFE attempts to remove potential correlations and collinearity in the framework by removing restricted amounts of characteristics for every repetition. RFE stipulates how many components must be retained; however, the initial number of acceptable characteristics is unspecified. To find the optimal number of features, cross-validation with RFE is employed to evaluate different feature subgroups and select the collection of characteristics with the greatest score.

### *3.5 Time-Dependent Flexible Deep Recurrent Network (TDF-DRN)*

The concept of time-dependent flexibility refers to the algorithm's ability to modify and react in an adaptable manner to alterations over time when used concerning the Time-Dependent Flexible Deep Recurrent Network (TDF-DRN). Time-dependent flex describes an analytical approach or modeling strategy that permits flexibility in the handling of time-dependent data. The following suggests that the framework takes the data's time component. When treating ADHD, a patient's symptoms and medication response change over time. As a result, a time-dependent model considers the evolution and change of these variables over various time intervals. The flexibility implies that the model can change or adapt to different circumstances. It suggests that the model's structure is relatively flexible, allowing it to capture various patterns and dynamics in the data instead of remaining imprecise. According to TDF-DRN, the model is designed to manage time-dependent data flexibly and adaptively. Such flexibility is essential for identifying the complex and evolving nature of symptoms and drug responses in the context of treating ADHD. According to the temporal patterns found in the data, the model might be able to modify its predictions, giving an improved and dynamic depiction of the effectiveness of the treatment.

The desired designation is the treatment result, while the ideal sets of attributes have been collected to constitute the analyzing input. This study used a deep learning method, the Deep Recurrent Network (DRN) to achieve this goal. The major goal of this approach is to develop the algorithm to solve the specified problem more quickly and accurately. Additionally, it instructs the system multiple times to get a high-performance frequency. A modified version of the used LSTM, RNN, is typically used to overcome problems with mood forecasting. The method for the construction of the DRN model of structure is shown in Figure 4. As a result, the suggested method predicts young individuals' anxiety levels using the DRN technique. Multiple layers of RNN variables make up the framework. Compared to traditional LSTM and RNN methods, DRN technology improves accuracy while reducing calculation time. Moreover, it

Journal of Theoretical and Applied Information Technology
15th September 2024. Vol.102. No. 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

gathers high-level data to forecast the label classification accurately. The optimal feature vectors are used to generate outcomes by fine-tuning the parameters of this approach. The total amount of filtering in this framework, which convolves the input using a 1D Convolutional layer, is used to construct the function of activation. Next, the dimensionality of the feature map is computed using the framework (Equation (4)) as follows:

$$FM_{Z2} = \frac{J_{Z1} - CK_Z + 2B_G}{G_{Gj}} + 1 \quad (4)$$

Where $B_g$.G. indicates the padding size, $G$ indicates the step size, $FM_{Z2}$ indicates the dimension of the attribute mapping, $IH_1$ shows the width of the input signal before convoluted and $CK_Z$ shows the combination kernel's gradient that will be obtained. Equations (5) through (6-7) are used to obtain the regional characteristics of the input employing the following convolution kernel:

$$IH_{1E}(j) = l(WM_E.Y(j{:}\,j+E-1)+P) \quad (5)$$

$$IH_{1E} = [IH_{1E}(1); IH_{1E}(2) \ldots . IH_{1E}(FM_{Z2})] \quad (6)$$

$$IHk'_{1E} = relu(IH_{1E}) \quad (7)$$

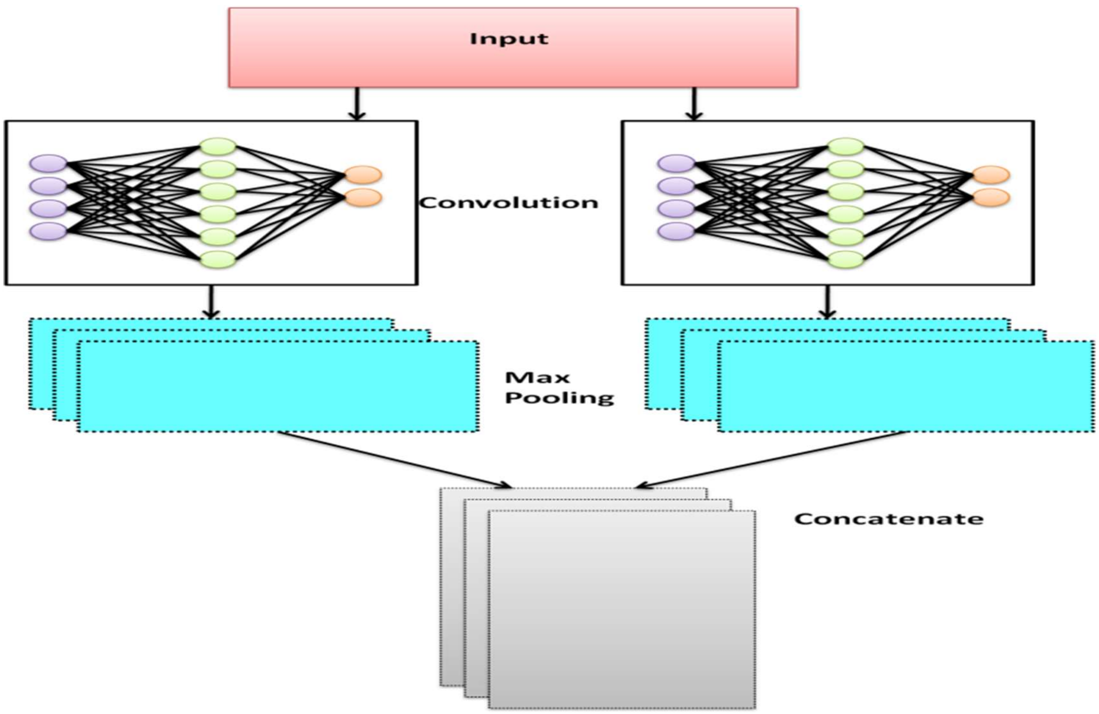


*Figure 4: Architecture of a DRN*

Where the result following convolution is denoted by $IH_1$, and $WM_E$ means the kernel of the convolution operation has height E. As a consequence, all organized findings of the language collection that have been obtained are pooled using the maximum pooling layer, as shown by Equations (8) and (9).

$$IHk'_{1E} = max\ (IHk'_{1E}) \quad (8)$$

$$IHk' = Concatenate\ (IHk'_{1E_1}, IHk'_{1E_2},) \quad (9)$$

## 4. RESULTS

A Windows 8 operating system, a 2.33 GHz CPU, and 6 GB of RAM were employed in the investigation. Python was utilized throughout the testing process. The accuracy, precession, f1-score, recall, specificity, Sensitivity, Area under the Average Accuracy Curve (AUC Curve), Root Mean Squared Error (RMSE), and Mean squared Error (MSE)of each categorization procedure classifier were used to gauge the performance. The outcomes show that our recommended model performance is much better than other traditional methods like Random Forest (RF), and Classification Tree (CT) [23, 24, 25], Support vector machines (SVM) [25].

### *4.1 Accuracy*

Accuracy is an outcome metric that measures a classifier's overall effectiveness. The overall number of reliable forecasts divided by the entire quantity of specimens is used to compute value. It demonstrates the accuracy with data samples are categorized using classification techniques. It is computed under Equation (10).

$$Accuracy = \frac{t_p+t_n}{t_p+t_n+f_p+f_n} \quad (10)$$

The relative effectiveness of Random Forest (RF) 83.3% and Classification Tree (CT) 88.9%) demonstrate how such techniques can predict how new medications will behave in clinical trials in addition to identifying important elements of the distinct cellular physiological processes that these drug classes produce. Figure 5 depicts the values that correspond to the accuracy measures. The suggested TDF-DRN outperforms other approaches with greater accuracy. The accuracy of the proposed system is shown in Table 1.

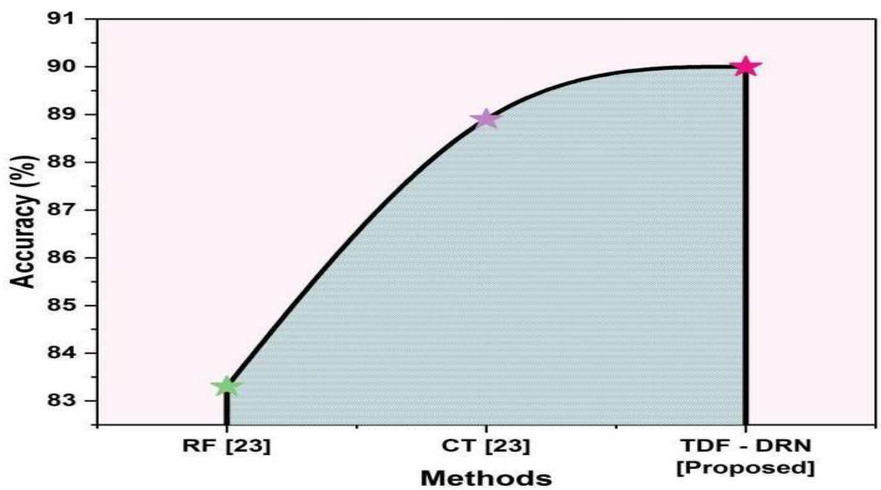


*Figure 5: Output for Accuracy*

*Table 1: Accuracy comparison*

| Methods | Accuracy (%) |
|---|---|
| **R.F. [23]** | 83.3 |
| **CT [23]** | 88.9 |
| **TDF - DRN [Proposed]** | 90 |

### *4.2 Precision*

Precision, or the proportion of truthfully correct positive estimates is calculated by dividing the amount of tp forecasts delivered by the simulation using the overall amount of positive forecasts estimated by the algorithm as shown in Equation (11).

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

The total accuracy that takes into consideration both features of the model is shown by the F1-score as shown in Equation (12), resulting in representing the harmonics means of the simulation's accuracy and recall.

$$F1 - score = \frac{2 \times precision \times recall}{precision + rec} \quad (12)$$

The table illustrates the performance metrics for a variety of classifiers, with the proposed TDF-DRN beating the others in terms of precision and F1-Score, obtaining percentages of 92.13%, 93.42%, 90.14%, and 95.95%, respectively. The table provides the performance metrics for the other classifiers, as shown in Table 2 and Figure 6.

Journal of Theoretical and Applied Information Technology
15th September 2024. Vol.102. No. 17

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

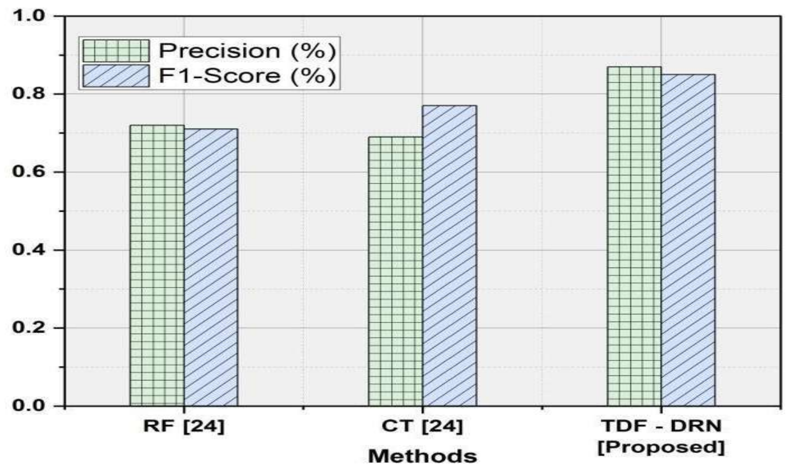


*Figure 6: Output for Precision and F1-score*

*Table 2: Precision and f1-score comparison with existing method*

| Methods | Precision (%) | F1-Score (%) |
|---|---|---|
| **RF [24]** | 0.72 | 0.71 |
| **CT [24]** | 0.69 | 0.77 |
| **TDF - DRN [Proposed]** | 0.87 | 0.85 |

### *4.3 Sensitivity*

The percentage of actual positive instances that are accurately classified as positive is known as Sensitivity. It suggests that a different percentage of real positive cases are mistakenly forecasted as negative. The definition of Sensitivity is in Equation (13):

$$Sensitivity = \frac{T_P}{T_P+F_N} \quad (13)$$

### *4.4 Specificity*

The percentage of actual negatives that were expected to be negatives is known as specificity. It suggests that there would be an additional percentage of accurate negative data that were misinterpreted as positive and referred to as false positives. It has the following definition in Equation (14):

$$Specificity = \frac{T_N}{T_N+F_P} \quad (14)$$

Sensitivity and specificity provide a thorough assessment of a model's performance by taking its accuracy in identifying both positive and negative examples, thereby offering valuable data about its overall categorization abilities. Figure 7 depicts the values that correspond to the Sensitivity and specificity measures. The suggested TDF-DRN outperforms existing approaches with greater output. The Sensitivity and specificity of the proposed system are shown in Table 3.

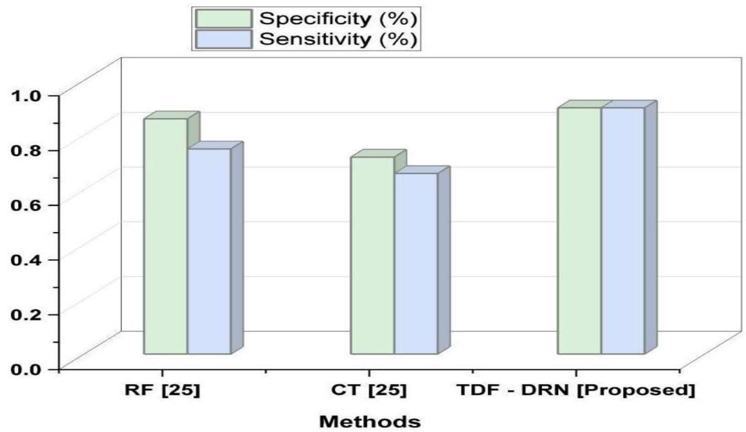


*Figure 7: Output for Sensitivity and specificity*

*Table 3: Sensitivity and specificity comparison between existing methods*

| Methods | Specificity (%) | Sensitivity (%) |
|---|---|---|
| **RF [25]** | 0.86 | 0.75 |
| **CT [25]** | 0.72 | 0.66 |
| **TDF - DRN [Proposed]** | 0.90 | 0.90 |

### *4.5 AUC Curve*

An indicator used to assess a binary classification model's performance is the AUC. The AUC curve, often referred to as the Receiver Operating Characteristic (ROC) curve, visually represents the percentage of true positives as well as the rate of false positives at various levels. Shown in Figure 8. True Positive Rate: The proportion of actual positive instances that the simulation accurately identifies in Equation (15).

$$TP\ Rate = \frac{T_P}{T_P + F_N} \tag{15}$$

The proportion of actual negative instances that the model incorrectly labels as positive is known as the false positive rate as shown in Equation (16).

$$FP\ Rate = \frac{F_P}{F_P + T_N} \tag{16}$$

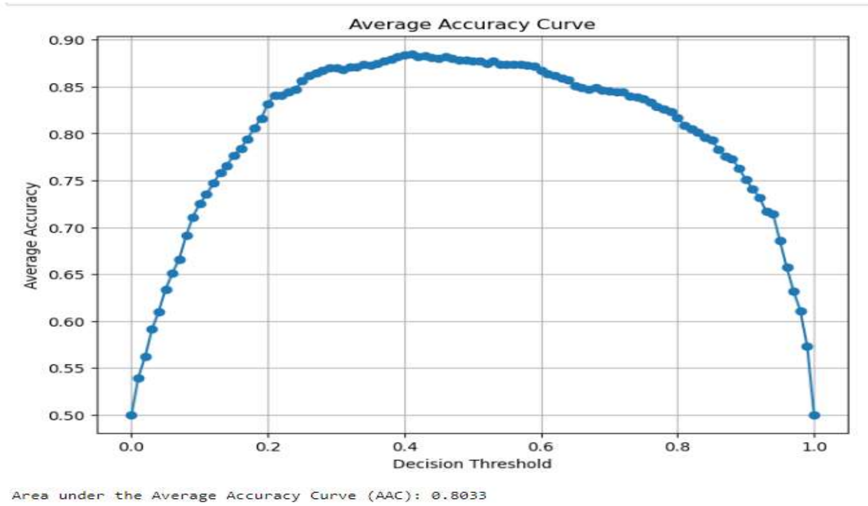


*Figure 8: Output for AUC curve*

The actual positive rate versus the false positive rate at different threshold values is plotted to form the ROC curve; we compute the Area under this ROC curve or AUC.

### *4.6 Mean squared Error (MSE)*

The MSE provides a statistic that sums together the ratio of the squared variance across the actual results and projected values, a popular statistic for assessing a regression algorithm's effectiveness. It provides a gauge of the actual values that match the model's predictions. The MSE efficiently computes the mean square variance between the actual and expected values. Considering a reduced MSE suggests predictions made by the model have been correlated compared to the actual information, it is indicative of higher accuracy. The MSE must be greatly affected by the presence of significant errors in a small number of data points, as the MSE is vulnerable to outliers in Equation (17).

$$MSE = \frac{1}{m} j = 1m\left(X_j - \hat{X}_j\right)^2 \qquad (17)$$

M - quantifies the amount of information, $X_j$ - is the value that has been observed for the j-th data point, and $\hat{X}_j$ - represents the anticipated value for the j-th component of data.

### *4.7 Root Mean Squared Error (RMSE)*

As an additional statistic, RMSE is utilized to assess the efficacy of a regression model. It takes the square root to make the metric more understandable and is derived from the MSE. It measures the average magnitude of the errors between expected and actual values in Equation (18).

$$RMSE = \sqrt{\frac{1}{m} j = 1m\left(X_j - \hat{X}_j\right)^2} \qquad (18)$$

m - Quantifies the amount of information, $X_j$ - is the value that has been observed for the i-th data point, $\hat{X}_j$ - represents the anticipated value for the ith component of data to find the most effective model for making accurate predictions, compare their RMSE values or use it as a benchmark. It is critical, though the details of the issue and the data. MSE and RMSE, two very comparable metrics, shift on considerations like punishing more significant mistakes and the statistic can be understood in the original data units. Both have their place in practice and picking one over the other usually comes down to the needs of the current issue. Figure 9 and Table 4 show the output of the existing method with the TDF-DRN method.

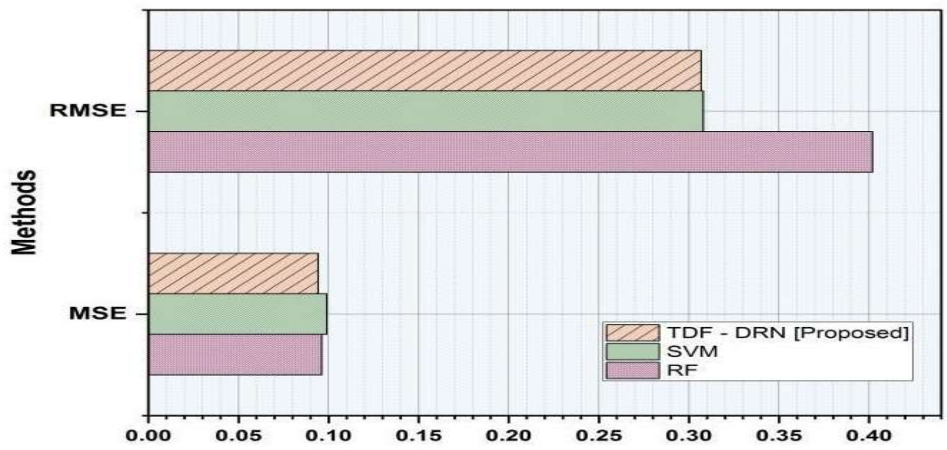


*Figure 9: Output for MSE and RMSE*

*Table 4: MSE and RMSE comparison between existing methods*

| Methods | RF | SVM | TDF - DRN [Proposed] |
|---|---|---|---|
| MSE | 0.0963 | 0.099 | 0.0942 |
| RMSE | 0.402 | 0.308 | 0.3069 |

## 5. DISCUSSIONS

Strong machine learning methods called Random Forests (RF) and classification trees (CT) are used to predict the effectiveness of medications for treating ADHD [24]. However, they have drawbacks, such as overfitting, which can result in highly variable forecasts and decreased generalization. Classification trees are prone to overfitting, which impairs predictive accuracy and leads to poor conception by collecting disturbance in training data instead of underlying patterns. Support vector machines (SVMs) are employed in investigations on the effectiveness of ADHD medications[22]; nevertheless, they have drawbacks, including interpretability problems, scalability problems, and Sensitivity to kernel function and parameter selection. Large datasets, heterogeneous data sources, and a variety of patient groups present challenges for them. Since it can learn hierarchical representations of characteristics, capture dynamic trends in patient data across time, and simulate temporal dependencies, a time-dependent flexible deep recurrent network is a potent tool for predicting drug efficacy in ADHD treatment. In the context of ADHD, this flexibility is critical for capturing complex interactions between factors influencing treatment success [25]. This research study does not cover the model analysis for the data collected of drugs half-life periods and immunological reactivity of the patients as well the toxicity of the drug. This analysis would be completed in the next subsequent research works.

## 6. CONCLUSIONS

The study was designed to discover if combined drugs work well for the treatment of ADHD in both adults and children. Data preprocessing operations, such as removing duplicate data and filling in missing values through efficient data partitioning, were handled using the Map Reduce technique. A TDF-DRN technique was devised to assess the effectiveness of the medication used to treat ADHD in both adults and children. According to preliminary research, combining drugs with other therapies can be an effective way to reduce ADHD symptoms in both age groups. The study also stresses that there could be differences in the best medication combination and dosage for adults and children. Compared to adults, children could need less medication and advantages of different combinations. When evaluating a model's efficacy in various categorization performance elements, it is frequently taken in conjunction with Sensitivity (90%), specificity (90%), precision (0.87%), accuracy (0.90), AUC curve (0.803) and the F1 score (0.85). The MSE (0.0942) and RMSE (0.3069) rates are lower than those of existing techniques. To improve the TDF-DRN's adaptability to the changing nature of ADHD symptoms, investigate integrating real-time monitoring data. Children require lower dosages than adults perform well and can gain advantages from different combinations.

## 7. ACKNOWLEDGMENT

This paper and the research behind it would not have been possible without the exceptional support of my Research supervisor, His enthusiasm, knowledge, and exacting attention to detail have been an inspiration. I am also grateful for the insightful comments offered by my co-supervisor. The generosity and expertise of one and all have improved this study in innumerable ways and saved me from many errors; those that inevitably remain are entirely my responsibility.

## REFERENCES

[1] M. Dobrosavljevic, H. Larsson, and S. Cortese, "The diagnosis and treatment of attention-deficit hyperactivity disorder (ADHD) in older adults," *Expert Rev. Neurother.*, vol. 23, no. 10, pp. 883–893, Sep. 2023.

[2] A. Koutsoklenis and J. Honkasilta, "ADHD in the DSM-5-TR: What has changed and what has not," *Front. Psychiatry*, vol. 13, p. 1064141, 2022.

[3] J. M. Halperin and D. J. Marks, "Practitioner Review: Assessment and treatment of preschool children with attention-deficit/hyperactivity disorder," *J. Child Psychol. Psychiatry*, vol. 60, no. 9, pp. 930–943, Sep. 2019.

[4] D. J. Heal, J. Gosden, and S. L. Smith, "New Drugs to Treat ADHD: Opportunities and Challenges in Research and Development," *Curr. Top. Behav. Neurosci.*, vol. 57, pp. 79–126, 2022.

[5] S. G. Reich and S. A. Factor, *Therapy of Movement Disorders: A Case-Based Approach*. Springer, 2019.

[6] G. A. Higgins and L. B. Silenieks, "The Effects of Drug Treatments for ADHD in Measures of Cognitive Performance," *Curr. Top. Behav. Neurosci.*, vol. 57, pp. 321–362, 2022.

[7] T. J. Dekkers, E. de Water, and A. Scheres, "Impulsive and risky decision-making in adolescents with attention-deficit/hyperactivity disorder (ADHD): The need for a developmental perspective," *Curr Opin Psychol*, vol. 44, pp. 330–336, Apr. 2022.

[8] K. Bock, *Brain Inflamed: Uncovering the hidden causes of anxiety, depression and other mood disorders in adolescents and teens*. Hachette UK, 2021.

[9] E. Rota *et al.*, "Direct Oral Anticoagulants and Concomitant Anti-seizure Medications: A Retrospective, Case-Control Study in a Real-World Setting," *Clin. Ther.*, vol. 46, no. 7, pp. e26–e30, Jul. 2024.

[10] R. D. Riley, J. F. Tierney, and L. A. Stewart, *Individual Participant Data Meta-Analysis: A Handbook for Healthcare Research*. John Wiley & Sons, 2021.

[11] S. Ayoubipour and N. Sho'ouri, "A Comparative Investigation of Wavelet Families for Classification of EOG Signals Related to Healthy and ADHD Children," *Clin. EEG Neurosci.*, vol. 55, no. 1, pp. 11–21, Jan. 2024.

[12] M. D. Hossain, M. A. Kabir, A. Anwar, and M. Z. Islam, "Detecting autism spectrum disorder using machine learning techniques: An experimental analysis on toddler, child, adolescent and adult datasets," *Health Inf Sci Syst*, vol. 9, no. 1, p. 17, Dec. 2021.

[13] S.-C. Yeh *et al.*, "A Virtual-Reality System Integrated With Neuro-Behavior Sensing for Attention-Deficit/Hyperactivity Disorder Intelligent Assessment," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 9, pp. 1899–1907, Sep. 2020.

[14] R. Lipka *et al.*, "Resolving heterogeneity in transcranial electrical stimulation efficacy for attention deficit hyperactivity disorder," *Exp. Neurol.*, vol. 337, p. 113586, Mar. 2021.

[15] C. Mulert and L. Lemieux, *EEG - fMRI: Physiological Basis, Technique, and Applications*. Springer Science & Business Media, 2009.

[16] I. Tachmazidis, T. Chen, M. Adamou, and G. Antoniou, "A hybrid AI approach for supporting clinical diagnosis of attention deficit hyperactivity disorder (ADHD) in adults," *Health Inf Sci Syst*, vol. 9, no. 1, p. 1, Dec. 2021.

[17] E. K. Coles *et al.*, "Randomized Trial of First-Line Behavioral Intervention to Reduce Need for Medication in Children with ADHD," *J. Clin. Child Adolesc. Psychol.*, vol. 49, no. 5, pp. 673–687, Sep-Oct 2020.

[18] C.-Y. Cheng, W.-L. Tseng, C.-F. Chang, C.-H. Chang, and S. S.-F. Gau, "A Deep Learning Approach for Missing Data Imputation of Rating Scales Assessing Attention-Deficit Hyperactivity Disorder," *Front. Psychiatry*,

vol. 11, p. 673, Jul. 2020.
[19] Y. Luo, T. L. Alvarez, J. M. Halperin, and X. Li, “Multimodal neuroimaging-based prediction of adult outcomes in childhood-onset ADHD using ensemble learning techniques,” *Neuroimage Clin*, vol. 26, p. 102238, Mar. 2020.
[20] P. Amado-Caballero *et al.*, “Objective ADHD Diagnosis Using Convolutional Neural Networks Over Daily-Life Activity Records,” *IEEE J Biomed Health Inform*, vol. 24, no. 9, pp. 2690–2700, Sep. 2020.
[21] K. M. Groeneveld *et al.*, “Z-Score Neurofeedback and Heart Rate Variability Training for Adults and Children with Symptoms of Attention-Deficit/Hyperactivity Disorder: A Retrospective Study,” *Appl. Psychophysiol. Biofeedback*, vol. 44, no. 4, pp. 291–308, Dec. 2019.
[22] A. Tajik, S. Nikfar, S. Elyasi, O. Rajabi, and M. Varmaghani, “Cost-effectiveness and budget impact analysis of lisdexamfetamine versus methylphenidate for patients under 18 with attention-deficit/hyperactivity disorder in Iran,” *Child Adolesc. Psychiatry Ment. Health*, vol. 17, no. 1, p. 115, Oct. 2023.
[23] L. Shahmoradi, Z. Liraki, M. Karami, B. A. Savareh, and M. Nosratabadi, “Development of Decision Support System to Predict Neurofeedback Response in ADHD: an Artificial Neural Network Approach,” *Acta Inform. Med.*, vol. 27, no. 3, pp. 186–191, Sep. 2019.
[24] E. C. Gunther, D. J. Stone, R. W. Gerwien, P. Bento, and M. P. Heyes, “Prediction of clinical drug efficacy by classification of drug-induced genomic expression profiles in vitro,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 16, pp. 9608–9613, Aug. 2003.
[25] T. Chen, I. Tachmazidis, S. Batsakis, M. Adamou, E. Papadakis, and G. Antoniou, “Diagnosing attention-deficit hyperactivity disorder (ADHD) using artificial intelligence: a clinical study in the UK,” *Front. Psychiatry*, vol. 14, p. 1164433, Jun. 2023.