Neurological Disorder Detection Based on Optimized Machine Learning via Physiological Signals in Children

Heba M. Afify[1]*, Ahmed M. Salaheldin[1], Neven Saleh[2]

[1]Systems and Biomedical Engineering Department, Higher Institute of Engineering, EL Shorouk Academy, Cairo, Egypt.

[2] Biomedical Engineering Department, Future University in Egypt, Cairo, Egypt.

hebaaffify@yahoo.com(Corresponding author)

ABSTRACT

The purpose of the study is to categorize particular neurological ailments to promote rehabilitation therapy plans that are based on serious games, namely, CatchAPet and LeapBall. To address this challenge, an imbalanced dataset named AKTIVES, which includes 25 children with specific needs, was used. In addition to being typically developing children, impaired children have dyslexia, intellectual disabilities, and obstetric brachial plexus injuries. The main idea for classification was to use machine learning (ML) principles in conjunction with the Bayesian optimizer and through different combinations of data augmentation, oversampling, and feature extraction. The decision tree (DT) and ensemble learning (EL) classifiers performed best on the dataset. As a result, four approaches have been proposed for classifying these disorders based on blood volume pressure (BVP), skin temperature (ST), and electrodermal activity (EDA). Through five evaluation metrics, one scenario yielded promising results with an approximate accuracy of 97%. This study is the first to classify signals of neurological disorders that a child had previously experienced while engaging in serious video games as a kind of physical rehabilitation. This work can assist physiotherapists in assigning appropriate physical rehabilitation plans for patients depending on their neurological and functional disorders.

Keywords: AKTIVES database; machine learning (ML); blood volume pressure (BVP) signal; skin temperature (ST) signal; and electrodermal activity (EDA) signal; Bayesian optimizer

1. Introduction

Video games are usually developed for entertainment. Recently, these games have been used as promising tools in physiotherapy rehabilitation plans. Video games that record scores, praise, and feedback are known as serious video games (Vural et al., 2024). Many disorders can be alleviated through the use of serious video games, including stroke rehabilitation(De Carvalho et al., 2012), anxiety (Loftness et al.,2022), dyslexia (Puritat et al., 2023), intellectual disabilities (IDs) (Vihriälä et al. ,2023), and obstetric brachial plexus injuries (OBPIs) (Chan et al., 2019). Additionally, a typically developing (TD) population has been investigated (Dobri, 2021). In the context of functional and mental disorders, several facts are highlighted. The World Health Organization estimates that 58 million children suffered from anxiety disorders in 2019, with an average recorded age of 6-18 years (Raines et al., 2019). Different forms of childhood anxiety disorders can be observed, such as behavioral disturbances, excessive fear, excessive worry, and sometimes failure to respond to the family and the external world (Raines et al., 2019). Another common type of learning disorder among children is called dyslexia (Puritat et al., 2023). This disorder affects children's ability to learn in terms of writing and reading. Additionally, dyslexia manifests as difficulty in short-term memory recovery (Junttila et al., 2023). Worldwide, statistics estimate that 9% to 12% of the population is affected by dyslexia (Cheng et al., 2022). Another form of neurological disability is intellectual impairment, which impairs the intelligence quotient (IQ) and interferes with the skills needed to live a self-sufficient life. During childhood, symptoms of this chronic illness first emerge (Wijayanti et al., 2013). Globally, ID is estimated to affect 1% to 3% of the world population (D'Adamo et al., 2021). Many limitations arise due to this type of disability, including academic interaction, learning capabilities, and control of attention (Ramadhaniet al., 2017). OBPIs occur during infant birth (O'Berry et al., 2017). Among 1000 births, a rate of 1.6 to 2.6 incidents is predicted (Coroneos et al., 2017). The most common risk factor for OBPI is shoulder dystocia. When the head is subjected to lateral traction to allow shoulder clearance during delivery, unilateral brachial plexus complex damage occurs. Although recovery is the dominant process, approximately 35% of injured children face functional impairment for the remainder of their lives (O'Berry et al., 2017). In addition to functional disorders, OBPI can impact the growth of children (Coroneos et al., 2017). In another setting, a child's fundamental motor skills (FMS) influence long-term activities such as literacy in addition to overall physical and mental progress (Zhang et al., 2023). These skills clearly reflect individuals' cognitive functions and social coexistence. A lack of physical activity and obesity are linked to low FMS competency in childhood, which increases a child's risk for health issues (Bieber et al., 2023). Thus, several neurological disorders may lead to negative consequences for learning and daily life activities.

In the literature, many related works have been conducted on neurological and physical disorders, considering video games as a treatment protocol (Dutra et al., 2021). Additionally, some of them (Anand et al., 2023) have claimed that video games not only improve people's cognitive skills but also enhance the decision-making process. Language learning is another issue that has been addressed in impaired children (Ramadhani et al., 2017). Through specific games and training, language learning skills are enhanced. Other studies have focused on stress recognition in people from different backgrounds. For instance, stress recognition while playing serious video games based on facial images was discussed in (McGinnis et al., 2019).

In a related context, wearable devices are used to detect physiological signals that indicate specific neurological illnesses. Electromyography (EMG), electrocardiography (ECG), electroencephalography (EEG), blood volume pressure (BVP), and electrodermal activity (EDA) are various examples of biosignals. For instance, a multimodal wearable sensor is used to diagnose a child's anxiety (Loftness et al., 2024). An analysis of the kid's motion was obtained using a wearable belt inertial measurement unit (IMU) that recorded the angular velocity, tilt angle, and acceleration (McGinnis et al, 2019). Moreover, both heart rate and voice datasets were used to assess children's stress (Xu et al.,2022). In addition, hand movement can be an indicator of treatment effectiveness. In related works, hand motions were observed during a physiotherapy recovery plan to reflect the effect of the treatment (Arman et al., 2021).

According to the literature review, recovery outcomes may be impacted by physical therapy regimens involving serious video games for people with neurodevelopmental illnesses. To address this issue, the aim of the study was to investigate the effects of specific neurological and physical disorders while following a physiotherapy plan. This study highlights a variety of impaired characteristics in children, including dyslexia, ID, and OBPI, in addition to a TD population.

In 2023, Coşkun et al. released a new dataset for stress recognition known as AKTIVES. It comprises multimodal data that combines facial images, videos, and physiological signals. The study was conducted on 25 impaired children to categorize them as stressed or unstressed. The dataset was collected during physiotherapy rehabilitation sessions for the hand by playing two serious video games. The data included face images and videos to indicate children's interactions with the games. Moreover, a wearable wristband named Empatica E4 was used to record BVP, EDA, and ST biosignals (Coşkun et al, 2023).

Based on this dataset, a recent publication in which stress and non-stress were identified from acquired facial images was published (Vural et al., 2024). This previous study has two shortcomings. First, the authors employed facial images to recognize only stress in children. Although the AKTIVES data included BVP, EDA, and ST signals, they were not connected to the neurological disorders that were previously described in the study.

Second, due to the imbalanced facial dataset, the classification accuracy was relatively low (92%) for images extracted from the AKTIVES database.

Thus, the proposed study aimed to classify the neurological disorders that were described in (Coşkun et al., 2023) based on the BVP, EDA, and ST biosignals. The ML methodology based on DT and EL(Wu et al., 2008) is presented to solve the underlying problem. Moreover, the classification accuracy for those disorders should be improved. Therefore, the contributions of this study are summarized as follows:

(1) This is the first study in which the recent AKTIVES dataset was used to classify neurological disorders namely ID, OBPIs, dyslexia, and TD through BVP, EDA, and ST biosignals in children previously experienced while engaging in serious video games as kind of physical rehabilitation.

(2) Different scenarios are applied to the BVP, EDA, and ST biosignals using ML techniques to reduce the training time needed, signal filtration to remove noise, and augmentation and oversampling to balance the database.

(3) Compared with that in the most relevant study, the classification accuracy was improved by approximately 5%. This is due to the use of augmentation and oversampling techniques for imbalanced data, in addition to the use of optimized hyperparameters through a Bayesian optimizer.

2. Related Works

A few numbers of children-dataset emotion recognition databases have been released (Coşkun et al., 2023). Children's ability to identify emotions and stress has garnered a lot of attention lately, particularly in those with neurological disorders. The children's emotion recognition databases are divided into facial expression databases and physiological signals databases. The examples of children's facial expression databases are Child Affective Facial Expression (CAFÉ) (LoBue et al., 2014), NIMH-ChEFS (Egger et al., 2011), Radboud Faces (Langner et al., 2010), and Dartmouth (Dalrymple et al, 2013), while the examples of audio, video, and physiological signals databases are EmoReact (Nojavanasghari et al.,2019), and MMDB (Rehg et al., 2013) as shown in Table 1. Previously, it should be noted that there are no datasets that can be used to identify emotion in children with various special needs based on their physiological signals and facial expressions. Recently, the AKTIVES database (Coşkun et al, 2023) consists of 50 videos with two labels (stressed / not stressed) and four classes of physiological signals for 25 children. In this study, we focused on physiological signals extracted from the AKTIVES database.

Table 1: Comparative databases for children's emotion recognition

	Modalities	**No. of Samples**	**No. of Children**	**No. of Labels**	**Age Range**
CAFÉ (LoBue et al. , 2014)	Facial expression	1192 images	154	7	2-8
NIMH (Egger et al.,2011)	Facial expression	482 images	59	5	10-17
Radboud (Langner et al, 2010)	Facial expression	80 images	10	8	8-12
Dartmouth (Langner et al.,2013)	Facial expression	640 images	80	8	6-16
EmoReact (Nojavanasghari et al., 2016)	Audio and video	1102 videos	63	17	4-14
MMDB (Rehget al, 2013)	Audio, video, and physiological	160 videos	121	2	1-2

AKTIVES (Coşkun et al,2023)	Facial expression images-physiological signals	50 videos	25	2	10.2 ±1.27

Different studies focused on the investigation of mental health and emotion recognition through various aspects and different datasets. Bargal et al. (Bargal et al., 2016) focused on the impact of stress, emotion recognition, and maltreatment history on maternally sensitive behaviors. Dumontheil et al. (Dumontheil et al., 2023) demonstrated the positive effects of mindfulness meditation on attentional reorienting and amygdala reactivity in adolescents and adults. In another context, Jiang et al. (Jiang et al, 2024) explored digital biomarkers for psychiatric evaluation, showing promise for detecting psychiatric disorders. Garg et al. (Garg et al., 2024) delved into AI-driven natural language processing for emotion detection in social activities, emphasizing ethical considerations. Bargal et al. (Bargal et al., 2016) presented the implementation details of an emotion recognition system for video-based emotion recognition, achieving a test data recognition rate of 56.66%.

3. Materials and Methods

3.1. Schematic diagram of proposed model

The study objective was to classify children with ID, OBPI, dyslexia, and TD based on the AKTIVES database (Coşkun et al., 2023) using optimized ML principles. The disorders were detected according to BVP, EDA, and ST physiological signals. According to the imbalanced characteristics of the database, the authors proposed four approaches to address this problem. Fig.1 shows a schematic diagram of the four approaches applied to the AKTIVES database.

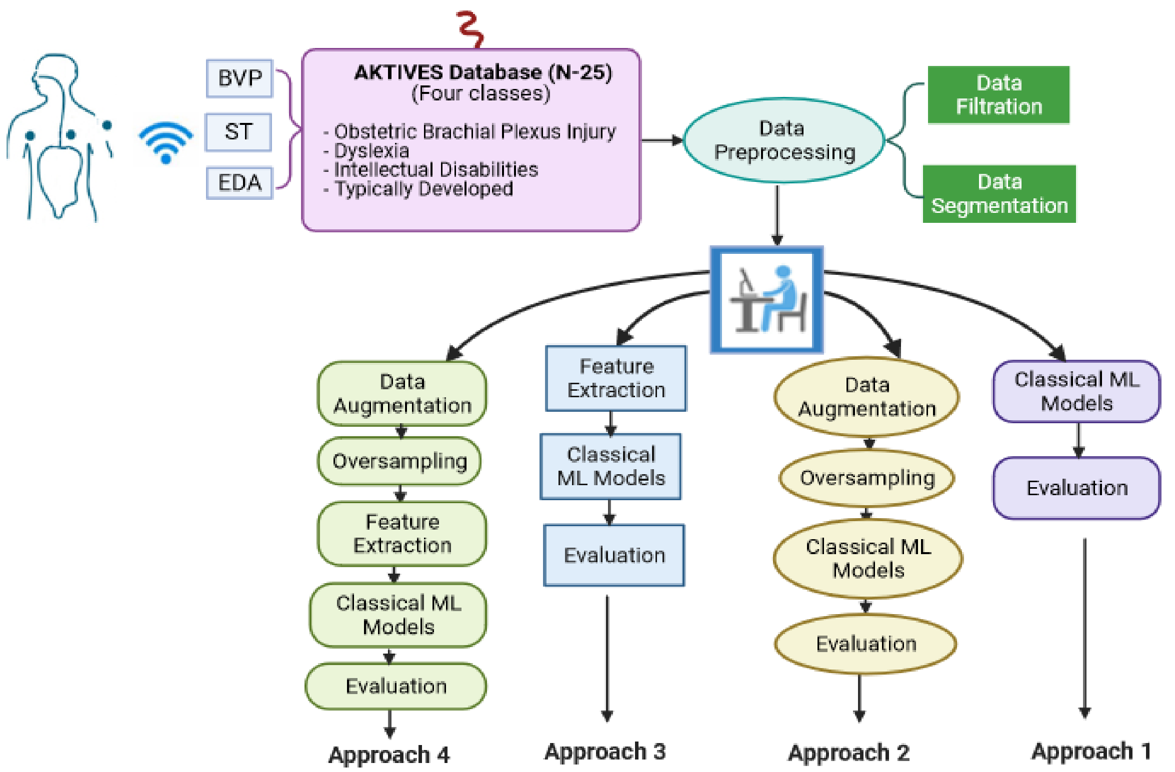


Figure 1: Schematic diagram of proposed model applied to the AKTIVES database for classification of four classes based on BVP, EDA, and ST signals

The proposed methodology is divided into three steps: (i) obtaining the raw physiological signals from the AKTIVES database, (ii) dividing the signal-preprocessing step into segmentation and filtration steps, and (iii) applying optimized ML methods to the biosignals for the classification of disorders. Signal segmentation involved dividing three signals into small signals, each lasting for ten seconds. Only BVP signals were filtered with a finite impulse response (FIR) lowpass filter with a cutoff frequency of 5 Hz and 32 orders. After that, four approaches were proposed to detect children's disorders automatically by selecting the best-fit approach. All the approaches used Bayesian optimization (Snoek et al., 2012; Saleh et al. 2022) to improve the classification results by using DT and EL (bagging/AdaBoost) classifiers. The following is a summary of the four approaches that were applied in this study:

1. The first approach consists of classical ML models and evaluation.
2. The second approach consists of four steps, namely, data augmentation (white Gaussian noise, random amplitude scaling, a moving average filter, the addition of a low-frequency sinusoid, and time warping), oversampling, classical ML, and evaluation.
3. The third approach consists of three steps: feature extraction (14 statistical and frequency features), classical ML model execution, and evaluation.
4. The fourth approach consists of five steps, namely, data augmentation (white Gaussian noise is added, random amplitude scaling occurs, a moving average filter is used, a low-frequency sinusoid is added, and time warping is applied), oversampling, feature extraction (14 statistical and frequency features), classical ML models, and evaluation.

3.2. Physiological signal database

Recently, the AKTIVES database (Vural et al., 2024) was collected at Istanbul Medipol University by occupational therapists. In this database, three physiological signals extracted from the AKTIVES database were recorded with Empatica E4 for 25 children with three disorders and TD children. The dataset was collected for 25 children, including 10 with TD, 8 with ID, 5 with dyslexia, and 2 with OBPI. There are two rehabilitative games by the Becure Company, namely, CatchAPet and LeapBall, used to detect disorders. All signals are stored in Comma-Separated Values (CSV) format. During data acquisition, the sampling frequencies are 64 Hz, 4 Hz, and 4 Hz for BVP, EDA, and ST, respectively. There are 720 total signal points in the AKTIVES database. This database is available on the GitHub repository by this link (https://github.com/hiddenslate/aktives-dataset-2022).

After filtration, the CatchAPet game achieved a BVP frequency range of 0.4-2.1 Hz, 0.1-2.5 Hz, 0.6-2.0 Hz, and 0.1-3.5 Hz for dyslexia, ID, OBPI, and TD, respectively. After filtration, the LeapBall game achieved a BVP frequency range of 0.2 -2.1 Hz, 0.2 -2.2 Hz, 0.3-2.1 Hz, and 0.1-2.2 Hz for dyslexia, ID, OBPI, and TD, respectively.

Table 2 displays the number of signals in the AKTIVES database for each game and each class. Notably, the AKTIVES database has imbalanced signals for each class. There are a total of 863 BVP signals, 867 EDA signals, and 877 ST signals for the CatchAPet game. There are a total of 941 BVP signals, 944 EDA signals, and 945 ST signals for the LeapBall game.

Table 2: Distribution of AKTIVES database according to four classes in terms of three signals and both games

Classes	CatchAPet			LeapBall		
	BVP	EDA	ST	BVP	EDA	ST
Dyslexia	97	98	100	187	189	189
ID	304	306	308	274	274	275
OBPI	68	68	69	69	69	69

TD	394	395	400	411	412	412
Total	863	867	877	941	944	945

By using the CatchAPet game**,** the original sample of three signals acquired from the AKTIVES database for dyslexia disorder is presented in Fig.2. On the other hand, Fig.3 shows the investigation of the BVP-filtered signals acquired from the AKTIVES database for dyslexia disorder with the CatchAPet game. The BVP-filtered signals were shown to have good interpretability without noise.

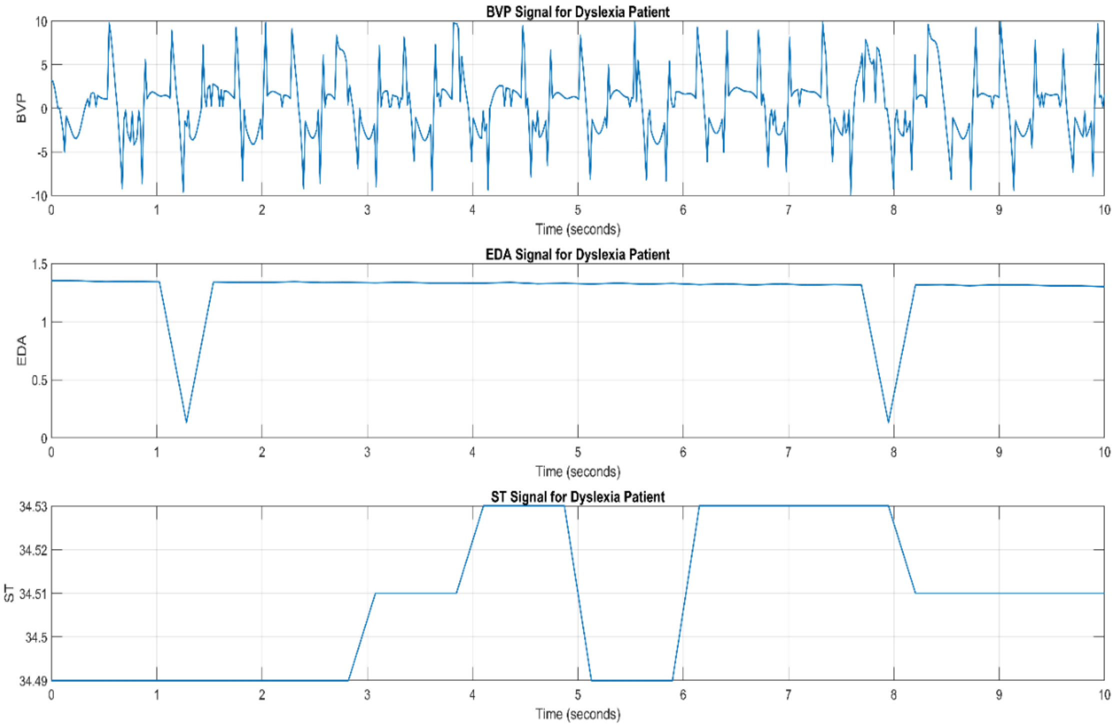


Figure 2: Original sample of three signals acquired from the AKTIVES database for dyslexia disorder via the CatchAPet game

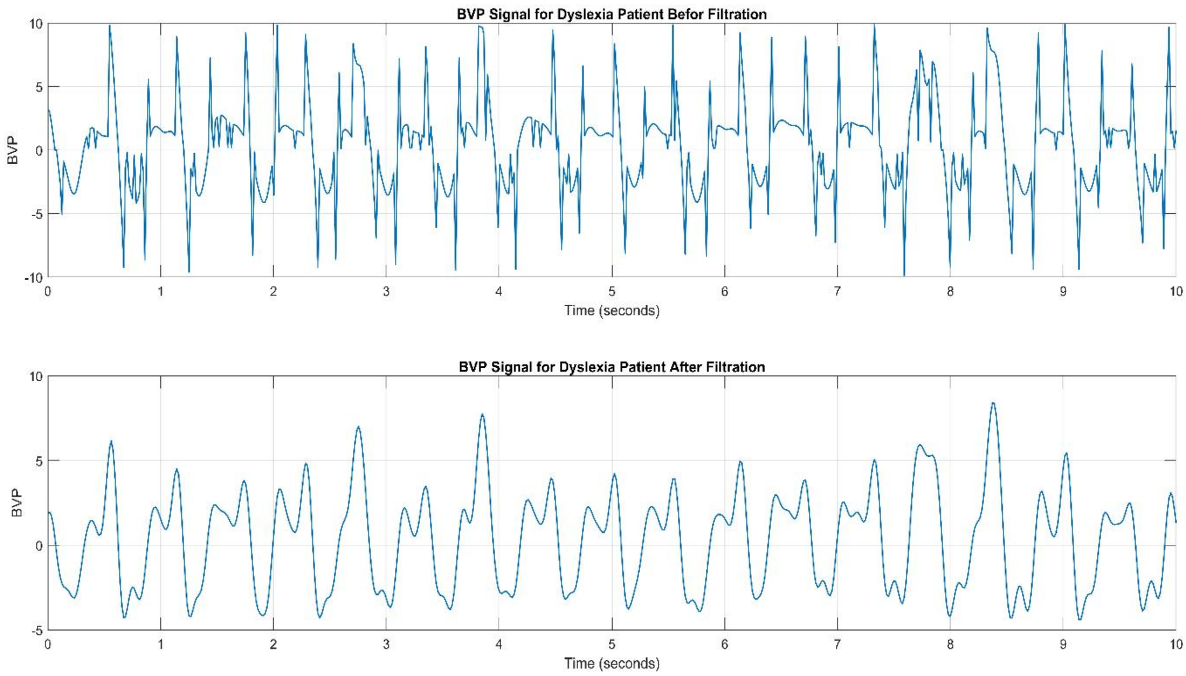


Figure 3: BVP-filtered signals acquired from the AKTIVES database for dyslexia disorder via the CatchAPet game

3.3. Data augmentation and oversampling

The problem in the AKTIVES database is the imbalanced distribution of disorders according to three physiological signals. In the proposed model, a data augmentation strategy and oversampling are used to overcome this problem during the training of the second and fourth approaches. After data augmentation and oversampling, the total number of BVP, ED, and ST signals became 394 for each disorder according to the CatchAPet game and 411 for each disorder according to the LeapBall game.

A data augmentation strategy is used to handle imbalanced data by reproducing actual data and altering it in various subsequent steps, including adding white Gaussian noise (Wang et al., 2018), random amplitude scaling (Zhang et al., 2022), and a moving average filter (Kawala et al., 2020); adding a low-frequency sinusoid (Scheeringa et al., 2011); and applying time warping (Jiang et al., 2023) as follows:

- Add white Gaussian noise as described in Eq. 1

$$X_j = X_i + X_n \qquad (1)$$

where X_j represents the augmented signal, X_i represents the original signal and X_n is the Gaussian distribution noise

- Random amplitude scaling was performed as described in Eq. 2.

$$X_j = X_i[0.8 + (1.2 - 0.8) \times rand(1)] \qquad (2)$$

where X_j represents the augmented signal and X_i represents the original signal.

- The average filter replaces each data point with the neighboring data point average (specified within its span) to function. It resembles low-pass filtering in certain ways. The moving average filter is described in Eq. 3.

$$X_j = \frac{1}{M} \sum_{k=-\frac{M-1}{2}}^{\frac{M-1}{2}} X_i[n+k] \tag{3}$$

where M represents the window size.

- Add a low-frequency sinusoid as described in Eq. 4.

$$X_j = X_i \times [0.1 \sin(0.1\pi n)] \tag{4}$$

where X_j represents the augmented signal and X_i represents the original signal.

- Apply time warping as described in Eq. 5.

$$W_f \in (0,1), W_i = round\left(\frac{n}{W_f}\right), X_j = X_i[W_i] \tag{5}$$

where n is the discrete time index, W represents the warping factor, X_j represents the augmented signal, and X_i represents the original signal.

On the other hand, oversampling (Duan et al., 2022) is used to assist in preventing phase distortion and aliasing while also increasing the resolution and signal-to-noise ratio.

3.4. Feature extraction

The primary application of ML models lies in feature extraction to efficiently handle data-reduced parameters. Feature extraction is crucial for converting original data points into features before utilizing ML classifier models (McGinnis et al., 2019). This paper focuses on detailing the feature extraction process in the third and fourth approaches, leveraging a set of 14 characteristics derived from statistical and frequency features. The selection of these features is underpinned by compelling reasons and advantages for enhancing the performance and interpretability of ML models.

Statistical features, as defined in the work (Qaisar et al., 2023), play a pivotal role in capturing essential characteristics of the data distribution. The mean provides a measure of central tendency, while the variance and standard deviation quantify data dispersion. Skewness and kurtosis offer insights into the asymmetry and shape of the distribution, respectively. The interquartile range and median range contribute robustness against outliers. Incorporating these statistical features enhances the model's ability to comprehend and generalize patterns within the data.

On the other hand, frequency features, as outlined in (Zhang et al., 2023), provide valuable information about the data's spectral characteristics. Frequency means and variances offer insights into the central tendency and spread of frequencies, while frequency max amplitude reflects the maximum intensity. Frequency entropy quantifies the distribution of frequencies, and skewness and kurtosis in the frequency domain provide additional details about the spectral shape. Leveraging these frequency features aids in capturing nuanced patterns related to the frequency components of the data, enriching the overall feature set. The adoption of statistical and frequency features in feature extraction brings about a comprehensive representation of the underlying data, contributing to the robustness, interpretability, and performance of ML models. The 14 characteristics derived from these features play a vital role in enhancing the model's ability to discern complex patterns and make informed predictions in various applications.

3.5. Machine learning classifiers

The automated diagnosis of neurological disorders depends heavily on the appropriate choice of classification technique. In this study, four approaches were tested with ML classifier models (Han et al., 2011) combined with Bayesian optimization (Snoek et al., 2021), which are used to select the best hyperparameters for classifying three disorders and TDs extracted from physiological signals. The extracted features are used to train a model that was created using ML models, including support vector machines (SVMs) (Lin et al., 2009) and K-nearest neighbors (K-NNs) (Duda et al., 2020); DT (Aydemir et al., 2014); and EL, including AdaBoost (Yaman et al., 2019) and bagging (Breiman et al., 1996).

AdaBoost (Qaisar et al, 2023) is a boosting algorithm related to the EL model. The main concept is to explicitly change the weights of each training sample that are provided to each classifier in the training data distribution. For all the training data, the distribution is initially uniform. Once each classifier has finished training, the distribution is modified during the boosting process. The weights are reduced for samples that are correctly classified and increased for those that are misclassified. Each classifier is combined into the final ensemble based on its accuracy. On the other hand, bagging incorporates the bootstrap sampling method to control the training data selection by (Breiman er al., 1996). To learn a unique classifier, T samples are randomly chosen each time from the initial training set of T samples. An ensemble predicts a test sample by having each classifier vote uniformly and with a majority. Theoretically, it is demonstrated that bagging accuracy increases significantly if bootstrapping causes notable variations in the individual classifiers built.

The DT (Fletcher et al., 2020) is based on the creation of a tree by repeating this procedure recursively. Every leaf node in a decision tree is assigned a class label; nonterminal nodes—the root node and other internal nodes—have attribute testing requirements for distinguishing records with various attributes. To assess the classifier's reliability and performance, data are initially supplied in the split strategy, which splits the data into 80% for training with 10-fold cross-validation and 20% for testing.

4. Results

The study objective was to classify neurological disorders, including ID, OBPI, and dyslexia, in addition to TD. The two serious video games, CatchAPet and LeapBall, during which the AKTIVES datasets (Coşkun et al., 2023) were gathered, were used in physical therapy. To categorize each illness, the authors used BVP, EDA, and ST biosignals. Fig.4 shows how the datasets were originally categorized. Furthermore, every disorder was classified according to BVP, EDA, and ST, resulting in a total of 2607 biosignals associated with the CatchAPet game and 2830 associated with the LeapBall game as shown in Table 2.

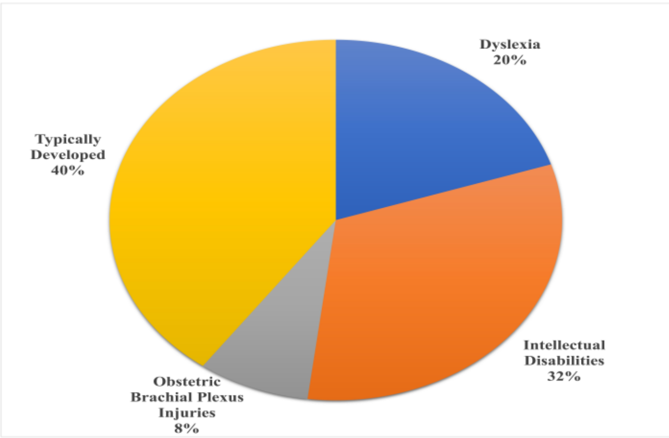


Figure 4: Data distribution of mental disorders in the investigated children

Owing to the imbalanced dataset of disorders, numerous approaches have been proposed to achieve signal balance. Four scenarios were established for the purpose of identifying the four groups based on the two video games that were used in the study (Coşkun et al., 2023). The basic steps that are applied to the four approaches include data segmentation and filtration. The first approach is very simple: after applying the basic steps, the ML classifiers were applied, and the evaluation metrics were measured. The extension of the first scenario is carried out in the second approach by adding data augmentation and oversampling before the ML classification. The third approach involves feature extraction before the ML classification step. The last scenario combines the second and third approaches together. This means that data augmentation, oversampling, and feature extraction are executed before ML classification. Therefore, the results are introduced for each video game according to the adopted approach. The computer specifications used to perform the calculations in this study were an Intel® Core i7-8565U 3.79 GHz processor, an NVIDIA GeForce MX 130, Windows 11, 64 bits, and 16 GB of RAM. Additionally, all the experiments were carried out using the MATLAB R2021b program.

4.1. Evaluation metrics

Five metrics—accuracy, sensitivity, specificity, precision, and F1-score—are proposed to assess each scenario. Each criterion is calculated in terms of true positives (TPs), true negatives (TNs), false negatives (FNs), and false positives (FPs). Eqs. 6-10 describe how each criterion is computed (Scheeringa et al., 2011). In addition, the confusion matrix and receiver operating characteristic (ROC) curve (Kuremoto et al., 2017) are displayed for each evaluated approach. Notably, after implementing four ML-based methods (K-NN, SVM, DT, and EL), only two classifiers were employed: DT and EL. The EL methods include either bagging or AdaBoost, depending on the optimization results.

$$\text{Accuracy} = (TP + TN)/(TP + FN + FP + TN) \quad (6)$$

$$\text{Sensitivity} = TP/(TP + FN) \quad (7)$$

$$\text{Specificity} = TN/(TN + FP) \quad (8)$$

$$\text{Precision} = TP/(TP + FP) \quad (9)$$

$$\text{F1-Score} = (2 \times (\text{Sensitivity} \times \text{Precision}))/(\text{Sensitivity} + \text{Precision}) \quad (10)$$

4.2. Performance of proposed model during the "CatchAPet" game

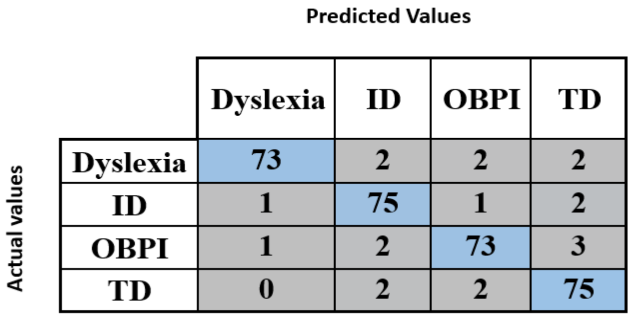
The AKTIVES data was collected while a child was playing two serious games. The first is the Becure CatchAPet, and the second is the Becure LeapBall (Coşkun et al., 2023). By considering the CatchAPet game, the results are presented in terms of accuracy, sensitivity, specificity, precision, and F1-score for the four proposed approaches. Table 3 shows the classification results for the specified disorders according to the DT and EL (bagging/AdaBoost) classifiers. The classification findings displayed in Table 3 indicate that, with an accuracy of 96.98%, the second approach for EL (AdaBoost) produced the best results. Typically, the other criteria reflect a similar conclusion.

Table 3: Results of evaluation metrics related to disorder classification based on the CatchAPet game

	Accuracy	**Sensitivity**	**Specificity**	**Precision**	**F1-Score**
Approach 1 - DT	90.70%	83.85%	93.21%	84.74%	83.75%
Approach 1 - Bagging	93.31%	88.02%	94.85%	87.44%	87.59%
Approach 2- DT	94.25%	88.44%	96.18%	88.18%	88.28%

Approach 2- AdaBoost	**96.98%**	**93.98%**	**97.99%**	**94.33%**	**94.12%**
Approach 3- DT	90.41%	73.58%	92.97%	79.50%	75.34%
Approach 3- AdaBoost	92.09%	77.18%	94.08%	84.47%	79.73%
Approach 4- DT	92.32%	84.66%	94.88%	84.38%	84.42%
Approach 4- Bagging	93.59%	87.25%	95.74%	90.18%	88.59%

The values of TP, TN, FP, and FN were calculated according to the resulting confusion matrix. Because we have eight confusion matrices based on the adopted approach and classifier, we present only the confusion matrix of the best scenario. Fig.5 depicts the confusion matrix of the second approach with the AdaBoost classifier. Additionally, the ROC curves of the best scenario are presented in Fig.6. These curves were used to classify children with underlying neurological disorders via the second approach based on the AdaBoost classifier.



Predicted Values / Actual values

	Dyslexia	ID	OBPI	TD
Dyslexia	73	2	2	2
ID	1	75	1	2
OBPI	1	2	73	3
TD	0	2	2	75

Figure 5: Confusion matrix of the best scenario for classifying the investigated disorders while playing the CatchAPet game

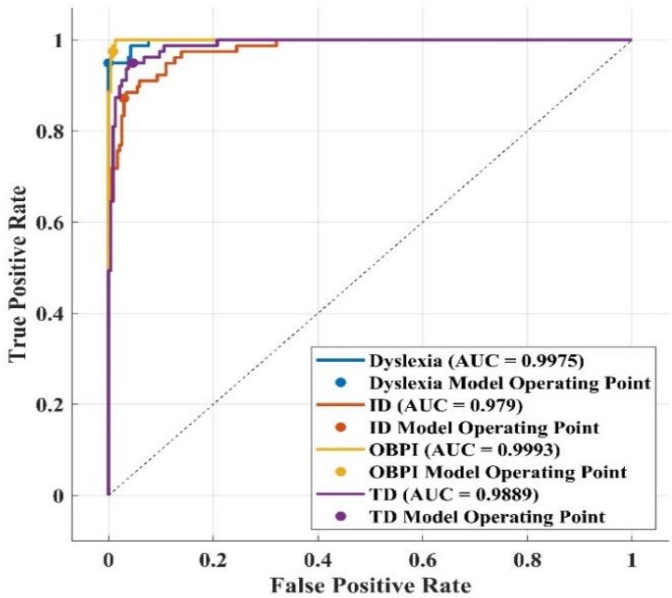


Figure 6: ROC curve for the best scenario for classifying investigated disorders while playing the CatchAPet game

In addition to classifying the underlying illnesses, the Bayesian optimization process for selecting the best combination of parameters plays a crucial role in the study's results. The optimization curve plots an estimated error for the classifier with the observed error to indicate the best combination of parameters.

As a result, points with minimum errors are presented as the best hyperparameters that should be selected. The optimization curves for each approach in terms of both games were run over 30 iterations according to the iteration on the X-axis and the minimum classification error on the Y-axis.

The best hyperparameter values achieved by the AdaBoost method with the maximum number of splits are 110, the number of learners is 430, and the learning rate is 0.908 for the CatchAPet game. For the first game, Fig.7 portrays the optimized parameters that were chosen for the AdaBoost classifier implementation.

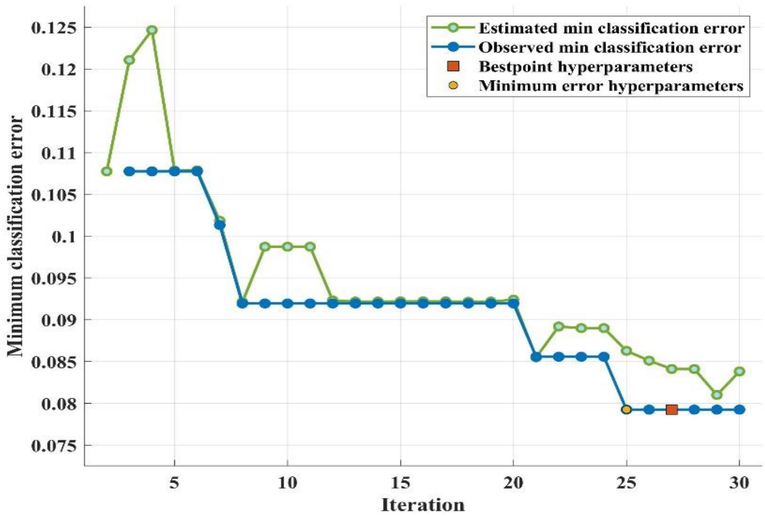


Figure 7: Optimized hyperparameters for the AdaBoost classifier according to the CatchAPet game

4.3. Performance of proposed model during the "LeapBall" game

While the children were playing the second game, Becure LeapBall, the BVP, EDA, and ST physical signals were obtained. The proposed scenarios were implemented to classify the intended disorders. As previously explained in the first game, the results were assessed for both classifiers against five metrics. The evaluation results of the second game are illustrated in Table 4. The second approach with EL using AdaBoost leads to the best results. An accuracy of 95.74% was achieved for classification among the other classifiers. Furthermore, the other evaluation criteria outperform the other related classifiers.

Table 4: Results of evaluation metrics related to disorder classification based on the LeapBall game

	Accuracy	Sensitivity	Specificity	Precision	F1-Score
Approach 1 - DT	93.39%	84.91%	95.45%	82.82%	83.73%
Approach 1 - Bagging	95.74%	84.87%	96.91%	93.33%	87.95%
Approach 2- DT	93.39%	84.91%	95.45%	82.82%	83.73%
Approach 2- AdaBoost	**95.88%**	**91.77%**	**97.26%**	**94.46%**	**93.06%**
Approach 3- DT	93.35%	84.73%	95.32%	79.40%	81.82%
Approach 3- AdaBoost	90.69%	79.45%	93.11%	80.34%	79.85%
Approach 4- DT	93.29%	86.59%	95.53%	88.01%	87.29%
Approach 4- Bagging	91.01%	82.01%	94.00%	81.16%	81.57%

Like in the first game, the confusion matrix and the ROC curve for the second game are presented in Figs.8 and 9, respectively. Moreover, the obtained optimization curve for choosing the best hyperparameters based on the AdaBoost classifier is depicted in Fig.10. The best hyperparameter values achieved by the AdaBoost classifier with the maximum number of splits are 44, the number of learners is 476, and the learning rate is 0.976 for the LeapBal game.

Predicted Values

Actual Values	Dyslexia	ID	OBPI	TD
Dyslexia	71	11	0	0
ID	3	77	0	2
OBPI	0	2	80	0
TD	8	1	0	73

Figure 8: Confusion matrix of the best scenario for classifying investigated disorders while playing the LeapBall game

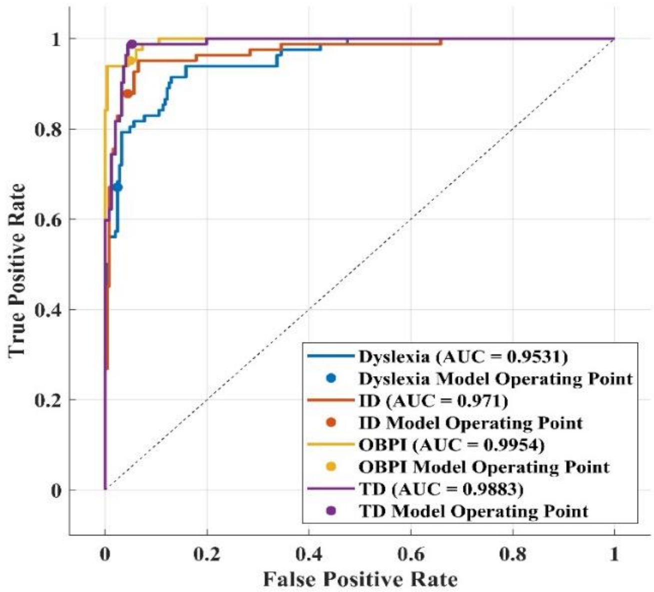


Figure 9: ROC curve for the best scenario for classifying investigated disorders while playing the LeapBall game

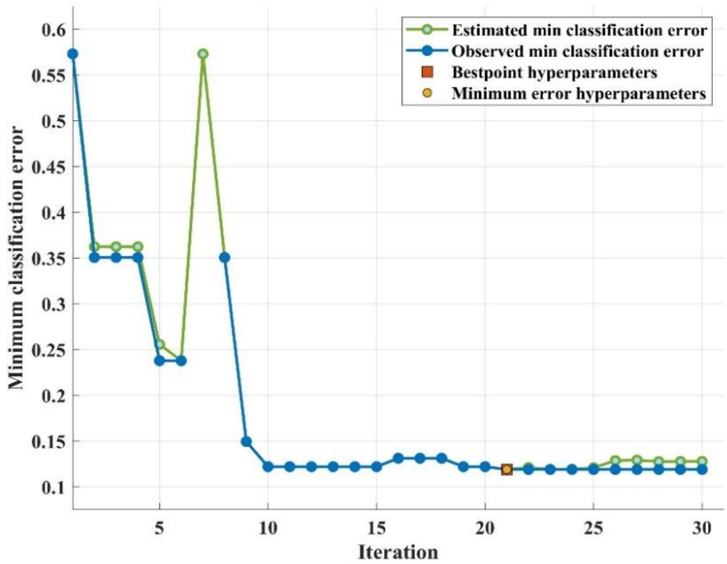


Figure 10: Optimized hyperparameters for the AdaBoost classifier according to the LeapBall game

5. Discussion

Traditionally, physical disorders are the main reason for performing physiotherapy rehabilitation. Recently, video games have proven effective at treating physical and neurodevelopmental disorders (López-Liria et al, 2022).

This study provides a framework for indicating the class of disorder in impaired children according to particular physiological signals. The AKTIVES database (Coşkun et al., 2023) is a recent collection of three physiological signals, namely, BVP, ST, and EDA, that are used to evaluate ID, OBPI, dyslexia, and TD in children by using two serious video games, CatchAPet and LeapBall, during physiotherapy. However, these databases are imbalanced, and the number of children varies widely for each disorder as shown in Fig. 4. Therefore, this study developed a schema for disorder classification in children based on three physiological signals as shown in Fig.1. The authors proposed a schematic flow comprising data preprocessing, including segmentation and filtration of BVP signals. Consequently, four approaches have been applied to select the best approach for accessing this imbalanced AKTIVES database. In this paper, four attempts were made to propose and explore the ML technique in addition to Bayesian optimization for disorder classification.

As shown in the results, the second approach is promising for augmenting and oversampling 720 data points from the three physiological signals to classify intended disorders using the AdaBoost classifier accompanied by Bayesian optimization. Bayesian optimization has exhibited a discriminating nature for adjusting the hyperparameters of classical ML models.

The first approach, based on the bagging classifier, achieved accuracies of 93.31% and 95.74% for the CatchAPet and LeapBall games, respectively. The second approach, using data augmentation based on the AdaBoost classifier, achieved maximum accuracies of 96.98% and 95.88% for the CatchAPet and LeapBall games, respectively as shown in Tables 3-4.

For the third approach, it was noted that the feature extraction step by 14 statistical and frequency features achieved an accuracy of 92.09% using the AdaBoost classifier for the CatchAPet game and 93.35% using the DT classifier for the LeapBall game. For the fourth approach, feature extraction and data augmentation achieved an accuracy of 93.59% using the bagging classifier for the CatchAPet game and 93.29% using the DT classifier for the LeapBal game.

Comparing the outcomes of the second and fourth approaches, the second approach, which used the original data points as features, produced a high rank of feature extraction, whereas the fourth approach used 14 different statistical features. Furthermore, signal augmentation and preprocessing have a beneficial impact on imbalanced AKTIVES database examination outcomes, demonstrating the resilience of the second approach, which makes use of the AdaBoost classifier. Initially, the second scenario approached the problem of database imbalance without conducting the conventional steps of feature extraction and biasing during the training process. Moreover, the AdaBoost (EL) classifier was combined with a Bayesian optimizer to automatically tune the hyperparameters. This reasoning explains why the second approach performed better than the others.

Benchmarking the study with relevant studies demonstrated the superiority of our findings. The most relevant work is a recent study in which children were categorized into stress and non-stress groups using the AKTIVES database (Coşkun et al., 2023). Only facial images were used without considering their acquired BVP, EDA, and ST signals. Additionally, due to imbalanced facial images among the different classes, the obtained data were relatively low, achieving 92% classification accuracy. In addition, in a previous study (Vural et al., 2024), a deep learning method for face image recognition was applied, which in turn consumed additional time. In this vein, there is no comparable study based on our point of view.

Several studies have compared the use of serious video games during physiotherapy treatment to our work. The work presented in (Dutra et al., 2021) only presented guidelines for improving computational thinking through video games for students who suffer from ID. Another study addressed the effects of two serious games in alleviating depression due to a cancer diagnosis (Khan et al., 2022). This study demonstrated the effectiveness of improving depression symptoms among young cancer patients. Obviously, the disorder and patients were different from those presented in our study. The ML fundamentals were applied to a database acquired from a wearable sensor to investigate internalizing impairment in children (McGinnis et al., 2019). Through logistic regression, 62 children were classified as having an internalizing disorder or a non-internalizing disorder, for an accuracy of 81%. The classification was performed based on the resulting angular velocity and motion acceleration of the participant. Although ML methods were employed, as in our study, a large difference was observed in classification accuracy. Furthermore, the mental disorders and the obtained data were not consistent with what was described in the proposed study. In another setting, heart rate and voice signals were obtained to classify a sample of children into stress and non-stress conditions. Using a pool of ML methods, the SVM was the best, resulting in an accuracy of 88.5% (McGinnis et al., 2019). Typically, the classification results were low compared to those of our study, in addition to the use of different physiological signals.

Various EL techniques have been applied to examine the degree of stress (Anand et al., 2023). Throughout regular data such as sleeping hours, weekly tasks, and productivity, stress was classified into three levels, high, medium, and low, with an accuracy of 93.48%. Despite several previous works using ML principles, our proposed method has been applied to different disorders and has achieved superior results in terms of accuracy and methodology.

The proposed study involved integrating three physiological signals generated by different sites in the body to significantly improve the discrimination effectiveness of neurological disorders in children instead of monitoring the psychological state of children by psychologists during some clinical tests (Hongn et al., 2025; Gandy, 2023).

The limitations of our study include the following: (i) a small number of children with special needs who participated in the AKTIVES database. (ii) Video recordings of facial expressions are available in this database, but video recordings of the disorders are unavailable. (iii) As far as the authors are aware, this is the first work in affective computing to use the AKTIVES database to classify neurological disorders in children. As such, no approach is directly comparable on the same database.

6. Conclusions and future work

The ability of researchers to recognize neurological disorders in children using physiological signals is encouraging because of the need to collect data and rehabilitate patients through video games. The development of optimized ML technology provides a more effective solution for neurological disorder classification based on wearable sensors. Improving the application of ML in such circumstances is feasible via an effective suggested strategy, which relies on BVP filtration to eliminate noise and signal augmentation to overcome the impact of the unbalanced AKTIVES database. The AdaBoost model and Bayesian optimization yielded an accuracy of 97% for the CatchAPet game. This tendency might be interpreted as a competitive tool in artificial intelligence guidelines, in which physiotherapists may ignore the effect of video games on rehabilitation. Physiological signals that are extracted from such wearable devices can guide the assessment of neurological disorders in children and adults. Additionally, applying diverse approaches to the AKTIVES database leads to improving the diagnosis of neurological disorders. The CatchAPet game is more suitable for identifying associated disorders in the AKTIVES database. Thus, we suggest that in subsequent performance comparisons, the consideration of a different game type be included as an extra tool. Finally, this study highlights the gap between the use of physiological signal databases during rehabilitation games and the recognition of neurological disorders in children. As a result, this gap should be addressed during rapid rehabilitation. Therefore, the demand for a large database of physiological signals during rehabilitation games is very important for identifying neurological disorders in children. This reflects how difficult it is to recognize these disorders through facial expressions only in videos or images. Future research will focus on the significant impacts of other video games on various neurological disorders in the rehabilitation domain. The proposed model can be extended to study the classification of children's emotions, such as happy, sad, angry, disgust, fear, and surprise. Additionally, other relevant databases might be compared with the existing study.

Data availability statement:
The datasets generated during and/or analyzed during the current study are available in this link: https://github.com/hiddenslate/aktives-dataset-2022

References

Anand, R.V., Md, A.Q., Urooj, S., Mohan, S. and Alawad, M.A., (2023), Enhancing Diagnostic Decision-Making: Ensemble Learning Techniques for Reliable Stress Level Classification. Diagnostics, 13(22), p.3455.

Arman, N., Oktay, A.B., Tarakci, D., Tarakci, E. and Akgul, Y.S., (2021), The validity of an objective measurement method using the Leap Motion Controller for fingers wrist, and forearm ranges of motion. Hand Surgery and Rehabilitation, 40(4), pp.394-399.

Aydemir O., Kayikcioglu T., (2014), Decision tree structure based classification of EEG signals recorded during two dimensional cursor movement imagery, Journal of Neuroscience Methods, Vol. 229, 68-75, ISSN 0165-0270.

Bargal, Sarah Adel, Emad Barsoum, Cristian Canton Ferrer, and Cha Zhang. (2016), Emotion Recognition in the Wild from Videos Using Images, In Proceedings of the 18th ACM International

Conference on Multimodal Interaction, ICMI ’16, New York, NY, USA: Association for Computing Machinery, 433–436.

Bavelier, D. and Green, C.S., (2019), Enhancing attentional control: lessons from action video games. Neuron, 104(1), pp.147-163.

Bieber, E., Smits-Engelsman, B.C., Sgandurra, G., Martini, G., Guzzetta, A., Cioni, G., Feys, H. and Klingels, K., (2023), Insights on action observation and imitation abilities in children with Developmental Coordination Disorder and typically developing children. Research in Developmental Disabilities, 139, p.104556.

Breiman, L.: Bagging predictors. (1996), Mach. Learn., Vol. 24, No.123–140 15.

Charles, D., Holmes, D., Charles, T. and McDonough, S., (2020), Virtual reality design for stroke rehabilitation. Biomedical Visualisation: Vol. 6, 53-87.

Chan, G.Y.Y., Nonato, L.G., Chu, A., Raghavan, P., Aluru, V. and Silva, C.T., (2019), Motion Browser: visualizing and understanding complex upper limb movement under obstetrical brachial plexus injuries. IEEE transactions on visualization and computer graphics, 26(1), pp.981-990.

Coşkun, B., Ay, S., Erol Barkana, D., Bostanci, H., Uzun, İ., Oktay, A.B., Tuncel, B. and Tarakci, D., (2023), A physiological signal database of children with different special needs for stress recognition. Scientific data, 10(1), p.382.

Coroneos, C.J., Voineskos, S.H., Christakis, M.K., Thoma, A., Bain, J.R. and Brouwers, M.C., 2017. Obstetrical brachial plexus injury (OBPI): Canada's national clinical practice guideline. BMJ open, 7(1), p.e014141.

Dalrymple, K., Gomez, J. & Duchaine, B. (2013), The dartmouth database of children’s faces: Acquisition and validation of a new face stimulus set. *PloS one* **8**, e79131.

De Carvalho Souza, A.M. and dos Santos, S.R., (2012), May. Handcopter game: a video-tracking based serious game for the treatment of patients suffering from body paralysis caused by a stroke. In 2012 14th Symposium on Virtual and Augmented Reality (pp. 201-209). IEEE.

D'Adamo, P., Horvat, A., Gurgone, A., Mignogna, M.L., Bianchi, V., Masetti, M., Ripamonti, M., Taverna, S., Velebit, J., Malnar, M. and Muhič, M., (2021), Inhibiting glycolysis rescues memory impairment in an intellectual disability Gdi1-null mouse. Metabolism, 116, p.154463.

Dobri, S.C., Samdup, D., Scott, S.H. and Davies, T.C., (2021), November. Differentiating Motor Coordination in Children with Cerebral Palsy and Typically Developing Populations Through Exploratory Factor Analysis of Robotic Assessments. In 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 5936-5939). IEEE.

Dumontheil, Iroise, Kristen E Lyons, Tamara A Russell, and Philip David Zelazo. (2023), A Preliminary Neuroimaging Investigation of the Effects of Mindfulness Training on Attention Reorienting and Amygdala Reactivity to Emotional Faces in Adolescent and Adult Females, Journal of Adolescence 95(1): 181–89.

Duan F, Zhang S, Yan Y, Cai Z. (2022), An Oversampling Method of Unbalanced Data for Mechanical Fault Diagnosis Based on MeanRadius-SMOTE, Sensors , 22(14), 5166.

Duda, R.O., Hart, P.E., Stork, D.G. 2000. Pattern Classification. John Wiley & Sons, New York.

Dutra, T.C., Felipe, D., Gasparini, I. and Maschio, E., (2021), Educational digital games and computational thinking for students with intellectual disabilities-guidelines for accessibility. In 2021 International Conference on Advanced Learning Technologies (ICALT) (pp. 314-316). IEEE.

Egger, H. *et al*. (2011), The nimh child emotional faces picture set (nimh-chefs): a new set of children's facial emotion stimuli. *International journal of methods in psychiatric research* 20, 145–56.

Fletcher S, Islam MZ (2020), Decision tree classification with differential privacy: a survey. ACM Comput Surv 52:83.

Garg, Muskan, and Chandni Saxena. (2024), Emotion Detection from Text Data Using Machine Learning for Human Behavior Analysis, In ed. D Jude B T - Computational Intelligence Methods for Sentiment Analysis in Natural Language Processing Applications Hemanth. Morgan Kaufmann, 129–44.

Gandy M., (2023), The role of psychologists in managing mental health comorbidities in adults with neurological disorders, Australian Psychologist, 58:3, 161-168.

Han J., Pei J., and KamberM., 2011. Data Mining: Concepts and Techniques, Elsevier, Amsterdam, Netherlands.

Hongn, A., Bosch, F., Prado, L.E. et al. (2025), Wearable Physiological Signals under Acute Stress and Exercise Conditions. Sci Data 12, 520.

Jiang S., Chen Z., (2023), Application of dynamic time warping optimization algorithm in speech recognition of machine translation, Heliyon,9, 11 ,e21625.

Jiang, Z et al. (2024), Multimodal Mental Health Digital Biomarker Analysis from Remote Interviews Using Facial, Vocal, Linguistic, and Cardiovascular Patterns, IEEE Journal of Biomedical and Health Informatics: 1–13.

Junttila, K., Smolander, A.R., Karhila, R., Kurimo, M. and Ylinen, S., (2023), Non-game like training benefits spoken foreign-language processing in children with dyslexia. Frontiers in Human Neuroscience, 17, p.1122886.

Kawala S. A, Podpora M, Pelc M, Blaszczyszyn M, Gorzelanczyk EJ, Martinek R, Ozana S. (2020), Comparison of Smoothing Filters in Analysis of EEG Data for the Medical Diagnostics Purposes, Sensors (Basel); 20(3):807.

Khan, S., Abbasi, A.Z., Kazmi, S.F., Hooi, T.D., Rehman, U., Hlavacs, H. and Arslan, F.S., (2022), Serious video games and psychological support: A depression intervention among young cancer patients. Entertainment Computing, 41, p.100479.

Kuremoto T, Baba Y, (2017), Obayashi M, Mabu S, Kobayashi K. A method of feature extraction for EEG signals recognition using ROC curve. In: Proceedings of 2017 International Conference on Artificial Life and Robotics, 654-657.

Langner, O. et al. (2010), Presentation and validation of the radboud faces database. Cognition and Emotion 24, 1377–1388.

López-Liria R, Checa-Mayordomo D, Vega-Ramírez FA, García-Luengo AV, Valverde-Martínez MÁ, Rocamora-Pérez P. (2022), Effectiveness of Video Games as Physical Treatment in Patients with Cystic Fibrosis: Systematic Review. Sensors (Basel), 22(5):1902.

LoBue, V. & Thrasher, C. (2014),The child affective facial expression (cafe) set: Validity and reliability from untrained adults. *Frontiers in psychology* **5**, 1532.

Lin, Y.-P., Wang, C.-H., Wu, T.-L., Jeng, S.-K., and Chen, J.-H. (2009), EEG-based emotion recognition in music listening: a comparison of schemes for multiclass support vector machine, in IEEE International Conference on Acoustics, Speech and Signal Processing, 2009. ICASSP 2009 (IEEE), 489–492.

Loftness, B.C., Halvorson-Phelan, J., O'Leary, A., Cheney, N., McGinnis, E.W. and McGinnis, R.S., (2022), UVM KID Study: identifying multimodal features and optimizing wearable instrumentation to detect child anxiety. In 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) (pp. 1141-1144). IEEE.

McGinnis, R.S., McGinnis, E.W., Hruschak, J., Lopez-Duran, N.L., Fitzgerald, K., Rosenblum, K.L. and Muzik, M., (2019), Rapid detection of internalizing diagnosis in young children enabled by wearable sensors and machine learning. PloS one, 14(1), p.e0210267.

Nojavanasghari, B., Baltrušaitis, T., Hughes, C. E. & Morency, L.-P. (2016), Emoreact: A multimodal approach and dataset for recognizing emotional responses in children. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI '16,

137–144 (Association for Computing Machinery, New York, NY, USA.

O'Berry, P., Brown, M., Phillips, L. and Evans, S.H., (2017), Obstetrical brachial plexus palsy. Current problems in pediatric and adolescent health care, 47(7), pp.151-155.

Puritat, K., Atsawakornkan, T. and Intawong, K., (2023), March. Using a Screening Risk of Dyslexia Game to Evaluate Students with Learning Disabilities Through the Use of Language-Independent Content and Machine Learning. In 2023 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunications Engineering (ECTI DAMT & NCON) (pp. 234-237). IEEE.

Qaisar S. M., Sibghatulla I. Khan, Kathiravan Srinivasan, Moez Krichen, Arrhythmia classification using multirate processing metaheuristic optimization and variational mode decomposition, (2023), Journal of King Saud University - Computer and Information Sciences, Vol. 35, No.1, 26-37.

Ramadhani, P.K. and Kustiawan, U., (2017), The effect of the big book media usage to simple sentences' reading ability for third grader with intellectual disability on elementary school for special needs. Journal of ICSAR, 1(1), pp.42-45.

Raines, E.M., Viana, A.G., Trent, E.S., Woodward, E.C., Candelari, A.E., Zvolensky, M.J. and Storch, E.A., (2019), Effortful control, interpretation biases, and child anxiety symptom severity in a sample of children with anxiety disorders. Journal of anxiety disorders, 67, p.102136.

Rehg, J. M. *et al.* (2013), Decoding children's social behavior. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3414–3421.

Scheeringa R., Pascal F., Karl-M., Robert Oostenveld, Iris G., David G. Norris, Peter Hagoort, Marcel C.M. Bastiaansen, (2011), Neuronal Dynamics Underlying High- and Low-Frequency EEG Oscillations Contribute Independently to the Human BOLD Signal, Neuron,69, 3, 572-583.

Saleh, N., Abdel Wahed, M. and Salaheldin, A.M., (2022), Computer-aided diagnosis system for retinal disorder classification using optical coherence tomography images. Biomedical Engineering/Biomedizinische Technik, 67(4), pp.283-294.

Snoek J., Larochelle H., Adams R.P., (2012), Practical Bayesian optimization of machine learning algorithms, Adv. Neural Inf. Process. Syst. 2951–2959.

Vihriälä, T., Raisamo, R., Merilampi, S., Leino, M. and Virkki, J., (2023), August. Textile-based AAC (TAAC) for Persons with Intellectual Disability: User Scenarios for Touch-based Activation. In 2023 IEEE 11th International Conference on Serious Games and Applications for Health (SeGAH) (pp. 1-5). IEEE.

Vural, Ş.F., Yurdusever, B., Oktay, A.B. and Uzun, I., (2024), Stress recognition from facial images in children during physiotherapy with serious games. Expert Systems with Applications, 238, p.121837.

Wang F., Zhong S.-h., Peng J., Jiang J., and Liu Y., (2018), Data augmentation for EEG-based emotion recognition with deep convolutional neural networks, MultiMedia Modeling, Part II, LNCS 10705, pp. 82–93, 2018.

Wijayanti, D.S., Hastuti, W.D. and Kristiawan, U., (2022), The Effect of the Numbers Lotto Media on the Ability of Simple Arithmetic in the Children with Intellectual Disability. In 2022 2nd International Conference on Information Technology and Education (ICIT&E) (pp. 217-220). IEEE.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., et al. (2008), Top 10 algorithms in data mining. Knowledge and Information Systems, 14(1), 1–37.

Wu, Y., Cheng, Y., Yang, X., Yu, W. and Wan, Y., (2022), Dyslexia: a bibliometric and visualization analysis. Frontiers in Public Health, 10, p.915053.

Xu, C., Yan, C., Jiang, M., Alenezi, F., Alhudhaif, A., Alnaim, N., Polat, K. and Wu, W., (2022), A novel facial emotion recognition method for stress inference of facial nerve paralysis patients. Expert Systems with Applications, 197, p.116705.

Yaman E, Subasi A. 2019. Comparison of Bagging and Boosting Ensemble Machine Learning Methods for Automated EMG Signal Classification. Biomed Res Int. 2019:9152506.

Zhang R., Ying Z., Li T., Jun S., Runnan L., Kai Y., Zhongrui Li, Bin Y., (2022), ERP-WGAN: A data augmentation method for EEG single-trial detection, Journal of Neuroscience Methods,376, 109621.

Zhang, D., Soh, K.G., Chan, Y.M. and Zaremohzzabieh, Z., (2023), Effect of intervention programs to promote fundamental motor skills among typically developing children: A systematic review and meta-analysis. Children and Youth Services Review, p.107320.