Arch Health Sci Res. 2025;12:1-8
DOI: 10.5152/ArchHealthSciRes.2025.24136

Original Article

Comparison of Machine Learning Methods to Detect Stress and Participation for Children with Different Special Needs During Serious Game-Based Therapy: A Observational Study

Buket COŞKUN[1], Şevket AY[2], Duygun EROL BARKANA[1], İsmail UZUN[3], Hilal BOSTANCI SEÇKİN[4], Devrim TARAKCI[5]

[1]Department of Electrical and Electronics Engineering, Yeditepe University Faculty of Engineering, İstanbul, Türkiye
[2]Department of Computer Engineering, Sakarya University of Applied Sciences Faculty of Applied Sciences, Sakarya, Türkiye
[3]INOSENS Bilisim, Gebze Organize Sanayi Bölgesi Technopark, Kocaeli, Türkiye
[4]Department of Occupational Therapy, Ankara Medipol University Faculty of Health Sciences, Ankara, Türkiye
[5]Department of Ergotherapy, İstanbul Medipol University Faculty of Health Sciences, İstanbul, Türkiye

Cite this article as: Coskun B, Ay Ş, Erol Barkana D, Uzun İ, Bostancı Seçkin H, Tarakcı D. Comparison of machine learning methods to detect stress and participation for children with different special needs during serious game-based therapy: A observational study. *Arch Health Sci Res.* 2025, 12, 0136, doi: 10.5152/ArchHealthSciRes.2025.24136.

What is already known on this topic?

- *Serious games are commonly used in rehabilitation to support motivation and participation in children with special needs.*
- *Stress and participation levels can affect the effectiveness of rehabilitation.*
- *Machine learning methods have been applied to stress detection using physiological signals in various populations.*

What this study adds on this topic?

- *This study demonstrates how stress and participation levels in children with special needs can be detected using data from a wearable device during game-based rehabilitation.*
- *Machine learning methods have proven effective in classifying children's stress and participation levels based on physiological signals.*
- *The outcomes derived from processing physiological signals can guide therapists in selecting games and rehabilitation activities better suited to each child's needs.*

ABSTRACT

Objective: Serious games have shown promise as therapeutic tools for children with different special needs. However, understanding how children feel and participate in a game is also important. This study used machine learning (ML) methods to classify stress in children with different special needs and their participation in serious game-based therapies.

Methods: This cross-sectional observational study was conducted at the Pediatric Rehabilitation Laboratory of the Department of Occupational Therapy between March and May 2023. Physiological signals such as blood volume pulse, electrodermal activity, and skin temperature were collected from 25 children with obstetric brachial plexus injury, dyslexia, intellectual disabilities, and typically developing children during game therapy. Using these physiological signals, 12 ML models were applied to classify children's stress and participation. Descriptive statistics (mean, SD, frequencies) were used to summarize participant characteristics. Model performance was evaluated using metrics such as accuracy, precision, recall, and F1 score.

Results: The results demonstrate that the k-nearest neighbor (KNN) classifier after an autoencoder resulted in the highest F1 scores of 66% and 63% for stress and participation classification, respectively. Furthermore, the eXtreme Gradient Boosting (XGBoost) model achieved the highest F1 scores of 91% and 86% for the no-stress and no-participation classifications, respectively. When both minority and majority classes were taken into consideration, using KNN following an autoencoder yielded better results with average F1 scores of 68% and 65% for stress and no-stress and participation and no participation, respectively.

Conclusion: This study shows that ML methods are effective in classifying children's stress and engagement states using physiological signals.

Keywords: Machine learning, physiological, rehabilitation, stress

Introduction

The repetition and motivation of the patients during rehabilitation exercises are important issues during therapy sessions. Serious games, and in particular exergames in rehabilitation, are currently getting attention because they can motivate, engage, and increase patients' adherence to their treatment. Furthermore, serious games have been shown to enhance cognitive functions and motor skills, such as hand-eye coordination, attention, and visual perception.[1,2] Serious games have been used to improve sensorimotor function and motivation in people with cerebral palsy.[3] Additionally, serious game-based therapies have

Received: September 17, 2024
Revision Requested: November 5, 2024
Last Revision Received: June 25, 2025
Accepted: July 3, 2025
Publication Date: September 17, 2025

Corresponding author: Hilal BOSTANCI SEÇKİN, e-mail: hilalbstnc71@gmail.com

also been used for children with brachial plexus,[4] dyslexia,[5,6] and intellectual disability and enhance the motor and cognitive abilities of children with special needs.[7,8] Note that these children's motivation and participation are always important during these therapies.[3]

Emerging game-based rehabilitation technologies can potentially improve child participation in repetitive task practice and enhance function.[8] However, children may lose interest if they are stressed during game-based therapies, so detecting and managing stress during therapy exercises is crucial for better outcomes and monitoring the patient during the treatment.[9] Therefore, it is important to consider both the children's performance and their stress levels, and participation in the game to find suitable serious games for each child and improve the effectiveness of serious game-based therapies. It is good to understand if the children are stressed and if they participate or do not participate, and then modify the serious game difficulty to reduce their stress levels and increase their participation. Understanding stress levels and participation patterns can guide the development of personalized interventions for children with different special needs.[10]

Emotion or stress recognition methods use facial expressions, speech, and posture. Various emotion/stress recognition studies were developed for typically developing children and children with different special needs.[9,11–17] The emotions of children diagnosed with and without dyslexia have been recognized using physiological signals and statistical methods.[11] A stress detection method has been developed for children with autism spectrum disorder (ASD) using electrodermal activity (EDA) and blood volume pulse (BVP) signals.[12,16] The heart rate of children with and without neurodevelopmental disorders has also been examined under stress conditions.[13] The emotions of children with ASD have been classified using physiological signals.[14] The valence of children with ASD and typically developing children has been classified using an electrocardiogram (ECG) signal.[15]

Statistical, mathematical, and signal-specific features have been extracted from the BVP, EDA, and skin temperature (ST) physiological signals for use in machine learning (ML) models to classify emotions and stress.[18-23] Random forest (RF), neural network, support vector machines (SVM), and Naive Bayes (NB) classification methods have been used to classify 4 emotions, arousal, and valence.[18] Furthermore, the k-nearest neighbor (KNN), SVM, and RF have been used previously to classify arousal and valence.[20] Furthermore, KNN, SVM, decision tree (DT), support vector regression, ensemble learning, linear discrimination, and Gaussian process regression methods have been used to classify stress in.[21] Random forest, KNN, gradient boosting, and adaptive boosting (AdaBoost) methods have been utilized for emotion classification.[23] Emotions (negative, not negative, unknown) have been classified using RF[24] in children with profound intellectual and multiple disabilities. The KNN and ensemble classifier algorithms have been used for valence recognition in children with ASD. Support vector machines, RF, and artificial neural networks (ANN) classifiers have been utilized for emotion classification in children with hearing disabilities.[17]

This study uses the publicly available AKTIVES (Duygu Durum ve Vücut Hareketleri Tanıma Destekli Akıllı Aktivite ve Egzersiz Sistemi, Smart Activity and Exercise System Supported by Emotion and Body Movement Recognition) dataset to develop stress/no-stress and participation/no-participation models using ML methods.[25] The data includes the physiological signals (BVP, EDA, and ST) and facial expressions of 25 children with different special needs and typically developed children, which are collected during serious game-based therapy.[25] Children's stress has previously been identified through facial expressions using this dataset.[9] This study is the first to develop stress/participation detection models among children with diverse special needs during serious gaming sessions using physiological signals (BVP, EDA, and ST) within the AKTIVES dataset. In the dataset, the stress and participation classes were minority classes (13% stress and 20% participation). The synthetic minority oversampling technique (SMOTE) was applied to the dataset to address the imbalanced data issue. This study extracted 59 statistical and signal-specific features from the collected BVP, EDA, and ST signals. Twelve different models using eXtreme Gradient Boosting (XGBoost), ANN, SVM, logistic regression (LR), KNN, DT, NB, RF, autoencoder, and metric learning were used to develop stress/no-stress and participation/no-participation classification models. The accuracy, F1 score, precision, and recall were calculated to determine the best ML methods for the classification of stress/no-stress and participation/no-participation. This study aims to leverage ML methods to classify stress levels and participation in children with varying special needs during serious game-based therapy. The goal is to provide personalized therapy insights that address the unique challenges faced by these children, enhancing engagement and optimizing therapeutic outcomes. The research questions are:

- How effectively can ML algorithms classify stress levels in children with different special needs during serious game-based therapy sessions?
- To what extent can these algorithms assess and categorize children's levels of participation?

Materials and Methods

Twenty-five children with special needs who accepted the informed consent form were included in the study. Ethical approval was obtained from the Research Ethics Committee of Istanbul Medipol University on February 18, 2021 (Approval No: E-10840098-772.02-6580). The dataset collected and published by the authors of this study includes physiological signals, BVP, EDA, and ST of 25 children with different special needs and typically developed.[25] Three of them were children with obstetric brachial plexus injury, 7 of whom were children with intellectual disabilities, 4 of whom were children with dyslexia, and 11 of whom were typically developing children. The inclusion criteria required a diagnosis of obstetric brachial plexus injury for children classified under Narakas Classification Group I, intellectual disability for children identified with Mild Intellectual Disability, or dyslexia, and age between 5 and 14 years. Children with other chronic diseases were not included in this study. The mean age of the children (10 boys and 15 girls) was 10.2 ± 1.27 years.

Children played the Becure CatchAPet (becureglobal.com) (BC) and Becure LeapBall (becureglobal.com) (BL) games. The BC game aims for children to use wrist flexion/extension movements. The BL game aims for children to hold and release the ball using grasping movements. When the serious game ended, the children looked at a black screen for 30 seconds. In this order, each child was asked to play BC and BL games (Figure 1). Children and parents were informed about the procedure, and parents provided written consent. To familiarize the children with the games, they played for 2-3 minutes before the experiment. Instructions were given to avoid unnecessary movements and to cover their faces. The experiment began with a 30-second baseline on a black screen. Then, children played "Becure CatchAPet" or "Becure LeapBall" for 420 seconds. Afterward, the black screen was shown again for 30 seconds, and the games were repeated. All games were played on a fixed computer, with the E4 wristband securely placed on the child's arm to ensure accurate data collection.

The physiological signals (BVP, EDA, and ST) were collected with an Empatica E4 smart bracelet. E4 has previously been used to detect stress.[23,26] Furthermore, a video camera captured the facial expression data (Figure 2). Every 10 seconds, 3 experts observed the children and

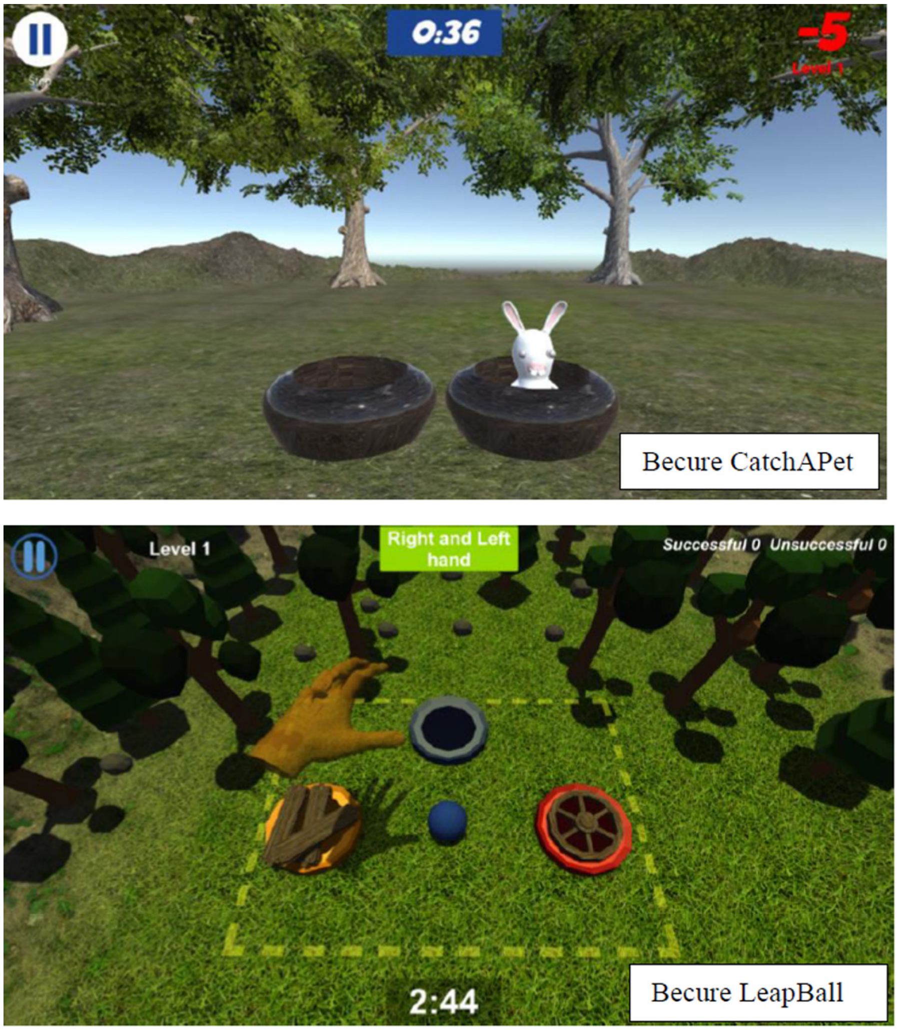

Figure 1. Becure CatchAPet and Becure LeapBal.

noted whether they thought they were stressed and whether they participated or did not participate in the game. Experts evaluated not only the physical presence of the children but also their active and voluntary engagement in the gameplay. Specifically, children were labeled as "participation" if they actively and willingly interacted with the game tasks, and as "no participation" if they appeared to engage passively or only due to external encouragement. Experts used the body language of children to decide between stress/no-stress and participation/no-participation. All 3 experts were occupational therapists with at least 2 years of experience. Each of the 3 experts was blinded to the annotation of the others. For each time slot, each expert indicated whether the children were stressed and whether they participated in the game. If at least 2 experts indicated that a child was stressed, that time period was labeled as "stress," otherwise, it was labeled as "no stress." Similarly, if at least 2 experts indicated that a child participated in the play, then the time period was labeled as "engagement", otherwise, it was labeled as "no engagement." In this way, binary categorical variables were obtained based on expert ratings rather than continuous values.

Figure 2. Experimental setup.

The dataset has 3 main parts: data processing, feature extraction, and classification for stress and participation detection (Figure 3).

Data Processing and Statistical Analysis

Physiological Data Collection

The EDA sensor was used to record skin conductivity data at a sampling rate of 4 Hz, whereas the photoplethysmograph (PPG) sensor was used to acquire BVP data at a sampling rate of 64 Hz. An infrared thermopile was used to collect the ST data at a sampling rate of 4 Hz. Finally,

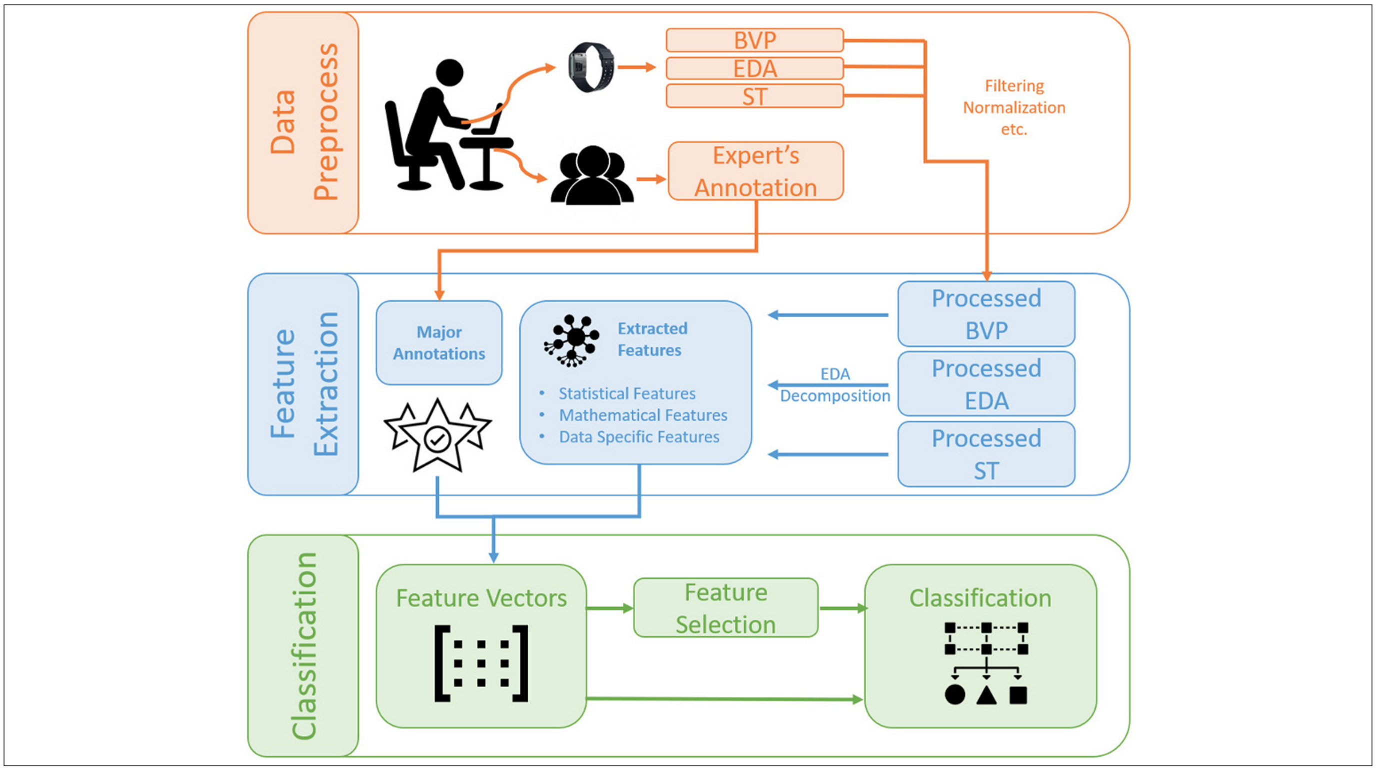

Figure 3. Flow chart for stress/participation detection.

the EDA and ST sensor data were upsampled to 64 samples per second to ensure that all the sensors captured equal samples.

Physiological Data Processing

All the experimental setups were combined into a single device for data synchronization. The collection time of the physiological signals was synchronized with the device clock using E4 Manager software. The E4 data were saved in Unix timestamp format. Synchronous time-stamps with the stimuli were gathered from each device during the experiment.

A sixth-order Chebyshev II filter (Stop Band Attenuation $R_s = 18$ dB and Normalized Stopband Edge Frequency $W_n = 0.1$ Hz), which was previously used for the BVP signal,[27] was applied to pre-process the BVP signal. A fifth-order Savitzky–Golay filter,[22] which was previously used to filter EDA signals, was selected for this study. The amplitude level of the physiological signal can differ in 2 individuals who express the same emotions.[18] Therefore, the amplitude of the BVP and EDA signals was normalized between 0 and 100 to overcome this difference. In this study, the raw data from the ST signal were used.

After filtering and normalization, the collected physiological signals were windowed as 10-second arrays with 5-second overlaps for feature extraction. First, the information from all sensors was fused at the data level using multisensor data fusion, and then data vectors (640 × 3) were produced for each 10-second slice.

Feature Extraction

The 59 features previously used for emotion recognition[18,19,21,23,28] were extracted from the BVP, EDA, and ST signals for each 10-second time-stamp. The EDA signal was separated into tonic EDA (SCL) and phasic EDA (SCR). Statistical and mathematical features were extracted from BVP, EDA, SCL, SCR, and ST signals (Table 1). The data-specific features, which were the number of positive peaks (BVP positive peak) related to systolic points, the number of negative peaks (BVP negative peak) related to diastolic points, the maximum distance between peaks (BVP peak maximum distance), and the minimum distance between peaks (BVP peak minimum distance) from the BVP signal, were also extracted.

Classification

Twelve different models were developed using ML methods for stress and participation detection. Twenty percent of the data was set aside for testing, and the data was selected equally for each child.

eXtreme Gradient Boosting (XGBoost) was selected as the first model in this study because it can increase accuracy through various arrangements. The XGBoost algorithm was configured with a learning rate of 0.0001, maximum depth of 6, minimum sample split of 2, 200 estimators, and a subsample ratio of 0.8. Subsequently, the SMOTE was applied to the dataset to address the imbalanced data issue before using the XGBoost classifier (model 2). In the dataset, the stress and participation classes were minority classes (13% stress and 20% participation). New samples were synthesized from existing data using SMOTE. The SMOTE uses minority class samples and applies the KNN

Table 1. Statistical and Mathematical Features

Feature	Explanation
Signal_zero_crossing	Zero crossing of signal
Signal_mean	Mean of signal amplitude
Signal_median	Median of signal amplitude
Signal_min	Minimum of signal amplitude
Signal_max	Maximum of signal amplitude
Signal_skew	Skewness of signal
Signal_kurtosis	Kurtosis of signal
Signal_025	0.25 quantile of amplitude
Signal_075	0.75 quantile of amplitude
Signal_050	0.50 quantile of amplitude
Signal_095	0.95 quantile of amplitude

approach to select the nearest neighbor randomly, and a random synthetic sample is generated in the feature space. After applying SMOTE to the data, the XGBoost classifier was used to balance the data. The SMOTE was applied to the data before training the ANN model to balance the imbalanced data (Model 3). Furthermore, the KNN algorithm was applied for classification after metric learning (Model 4). In this study, the value of k was empirically determined to be 5, as it exhibited optimal performance.

Subsequently, an autoencoder (Figure 4) was used to train the SVM (Model 5). The autoencoder was trained using a minority class. The primary objective of autoencoder training is to reduce the reconstruction error, thus justifying the choice to undertake training using a singular class. As a result, the autoencoder yielded a low reconstruction error for no-stress/no-participation data but a high reconstruction error for abnormal data (minority class). At this point, outliers became apparent in the hidden layer, and the hidden layers of the proposed autoencoder structure were provided as input to the SVM model. In the proposed autoencoder architecture, 4 fully connected layers were utilized for both the encoder and decoder. The hyperbolic tangent (tanh) was chosen as the activation function for these layers. Furthermore, the rectified linear unit activation function was employed for the latent layer, and a linear activation function was used for the output layer.

Autoencoder structures were also provided as input to the other models: LR (Model 6), extreme gradient boosting (XGBoost) (Model 7), DT (Model 8), KNN (Model 9), Naive Bayes (NB) (Model 10), RF (Model 11), and ANN (Model 12).

Results

The accuracy, F1 score, precision, and recall metrics are presented for the stress/participation and no-stress/no-participation target variables (Table 2). The F1 score was between 30% and 91% for the stress/no-stress model. The F1 score ranged between 18% and 86% when evaluating the participation/no-participation models. Higher F1 scores were observed in the no-stress and no-participation models because experts annotated more non-stress and non-participation than stress and participation.

The autoencoder + KNN model (Model 9) yielded the highest F1 scores for both stress (66%) and participation (63%) (Table 2). Model 9 selected only the data belonging to the no-stress and no-participation groups from the dataset and trained it with an autoencoder to determine stress and participation. Data points for stress/no-stress and participation/no-participation were generated by reconstructing the model using the hidden layers of the autoencoder. The generated model weights were then input into the ML models. This model was trained with KNN after the autoencoder. The autoencoder architecture used in this study was designed to augment the predictive accuracy of the minority classes of stress and participation. Observations revealed that models integrating this autoencoder architecture demonstrated enhanced F1 scores for both minority classes. The use of the autoencoder structure to improve predictions for minority classes was distinctly evident; models incorporating this structure exhibited notable improvements.

Furthermore, the SMOTE method did not provide a significant improvement. The SMOTE method generates synthetic samples through interpolation among existing instances of minority classes. Consequently, this model could lead to an overfitting issue, wherein it becomes excessively specialized in identifying synthetic samples solely, impairing its ability to generalize to unseen data effectively.

The study reveals that the autoencoder + KNN model (Model 9) outperforms others in predicting stress and participation by effectively addressing class imbalance, achieving the highest F1 scores for these minority classes. Although SMOTE was used to tackle class imbalance, it did not significantly enhance performance and posed overfitting risks. Conversely, XGBoost excelled in classifying the majority classes (no stress, no participation) without SMOTE. These results suggest that while SMOTE may not always be beneficial, the autoencoder + KNN approach provides a promising solution for improving predictions in imbalanced datasets. Future research should validate these findings with larger samples to better understand their generalizability and effectiveness.

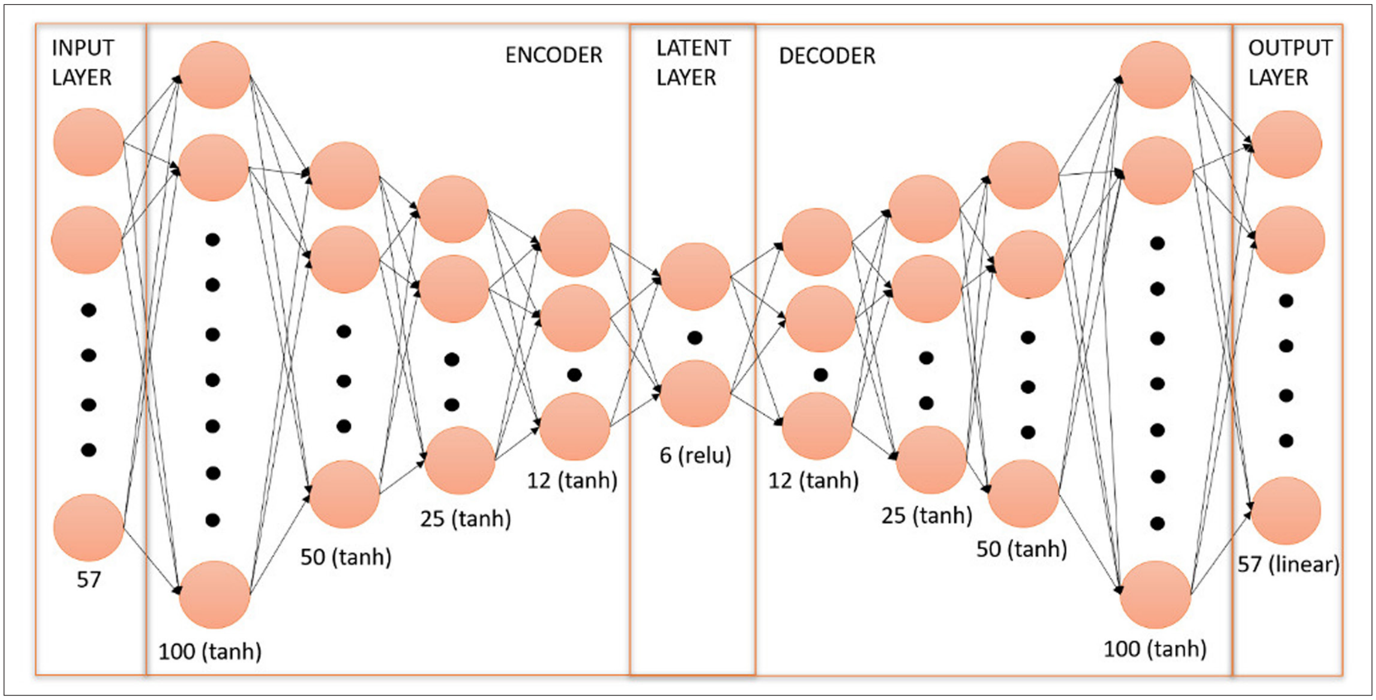

Figure 4. Autoencoder architecture.

Table 2. Evaluation Metrics: Stress and No-Stress

Model	Classifier	SMOTE	Accuracy	Average F1 Score	Stress			No Stress		
					F1 Score	Precision	Recall	F1 Score	Precision	Recall
Model 1	XGBoost	No	0.84	0.61	0.30	0.37	0.25	**0.91**	0.89	0.94
Model 2	XGBoost	Yes	0.81	0.64	0.39	0.34	0.45	0.89	0.91	0.87
Model 3	ANN	Yes	0.78	0.64	0.40	0.32	0.55	0.87	0.92	0.82
Model 4	Metric Learning + KNN	Yes	0.65	0.54	0.34	0.22	0.72	0.77	0.94	0.65
Model 5	Autoencoder + SVM	No	0.63	0.63	0.63	0.58	0.70	0.63	0.70	0.58
Model 6	Autoencoder + LR	No	0.58	0.56	0.54	0.52	0.56	0.58	0.61	0.56
Model 7	Autoencoder + XGBoost	No	0.63	0.63	0.63	0.58	0.68	0.63	0.68	0.59
Model 8	Autoencoder + DT	No	0.61	0.59	0.49	0.60	0.42	0.68	0.61	0.76
Model 9	Autoencoder + KNN	No	0.68	**0.68**	**0.66**	0.64	0.69	0.69	0.72	0.67
Model 10	Autoencoder + NB	No	0.54	0.54	0.52	0.49	0.56	0.55	0.59	0.52
Model 11	Autoencoder + RF	No	0.65	0.66	0.65	0.60	0.70	0.66	0.71	0.61
Model 12	Autoencoder + ANN	No	0.49	0.49	0.54	0.46	0.66	0.43	0.56	0.35

ANN, artificial neural networks; DT, decision tree; KNN, k-nearest neighbor; LR, logistic regression; NB, Naive Bayes; RF, random forest; SMOTE, synthetic minority oversampling technique; SVM, support vector machines; XGBoost, eXtreme gradient boosting.

The XGBoost model (Model 1) was trained with the entire dataset, resulting in high F-scores of 91% for the no-stress and 86% for the no-participation classification (Tables 2 and 3). This model detected no stress and participation without utilizing the SMOTE method. The observation indicated that XGBoost achieved higher accuracy in predicting the majority classes (no stress, no participation) without applying SMOTE. Given that there was no need for improvement in predicting the majority classes, it was evident that the simpler XGBoost model yielded better results. When the autoencoder was not used, a significant difference was observed between the minority and majority classes. The proposed autoencoder structure distinctly enhances the classification performance of the minority class, yielding optimal results for both classes and effectively addressing the imbalance issue within the dataset. Moreover, considering both minority (stress, participation) and majority classes, it was better to use Model 9 (autoencoder + KNN), which distinctly separated both classes with an average F1 score of 68% for stress and no-stress and 65% for participation and non-participation, respectively (Table 2).

Discussion

This study developed stress/no-stress and participation/no-participation detection models using ML methods for children with different special needs, which can be used during serious game-based rehabilitation. This study used publicly available labeled data (stress/no stress and participation/no participation).[25] The dataset included physiological signals collected from 25 children with special needs and typically developed children, where children played 2 serious games. Three experts observed the children and labeled the data to indicate whether they were stressed or not and whether they participated or not during the game. Various ML methods are used to develop stress/no-stress and participation/no-participation models.

Physiological responses to various stimuli, such as stress, physical activity, and pain, can negatively affect physical and psychological health. Stress can affect the patient's performance and reduce compliance with exercise programs.[29] Therefore, detecting and managing stress during therapy exercises is crucial for better patient outcomes. Such mechanisms also improve patient outcomes and contribute to the effective monitoring of patients during home treatment.[9] Emerging game-based rehabilitation technologies can potentially improve child participation in repetitive task practice and enhance function.[8] The findings suggest that a personalized understanding of stress and participation levels in children with different special needs can contribute to optimizing the frequency of therapy sessions. Practitioners can enhance the effectiveness of interventions by tailoring the therapeutic schedule to individual profiles. Evaluating children's motivation and stress levels during rehabilitation has been emphasized previously.[30] Understanding the stress-response patterns of each child may aid therapists in determining the optimal session length to maximize engagement and minimize potential stressors. Thus, a personalized understanding of stress and participation levels in children with

Table 3. Evaluation Metrics: Participation and No Participation

Model	Classifier	SMOTE	Accuracy	Average F1 Score	Participation			No Participation		
					F1 Score	Precision	Recall	F1 Score	Precision	Recall
Model 1	XGBoost	No	0.76	0.60	0.33	0.45	0.26	0.86	0.82	0.91
Model 2	XGBoost	Yes	0.71	0.57	0.32	0.34	0.31	0.82	0.81	0.83
Model 3	ANN	Yes	0.72	0.51	0.18	0.25	0.14	0.84	0.79	0.89
Model 4	Metric Learning + KNN	Yes	0.67	0.65	0.55	0.45	0.72	0.75	0.86	0.66
Model 5	Autoencoder + SVM	No	0.53	0.53	0.53	0.55	0.51	0.53	0.51	0.55
Model 6	Autoencoder + LR	No	0.55	0.56	0.58	0.57	0.59	0.53	0.54	0.52
Model 7	Autoencoder + XGBoost	No	0.60	0.60	0.61	0.62	0.60	0.59	0.58	0.61
Model 8	Autoencoder + DT	No	0.60	0.60	0.57	0.64	0.52	0.62	0.57	0.68
Model 9	Autoencoder + KNN	No	0.64	**0.65**	**0.63**	0.68	0.58	0.66	0.61	0.71
Model 10	Autoencoder + NB	No	0.54	0.54	0.54	0.56	0.52	0.53	0.52	0.55
Model 11	Autoencoder + RF	No	0.61	0.61	0.62	0.63	0.61	0.60	0.59	0.61
Model 12	Autoencoder + ANN	No	0.61	0.61	0.59	0.65	0.54	0.63	0.58	0.68

ANN, artificial neural networks; DT, decision tree; KNN, k-nearest neighbor; LR, logistic regression; NB, Naive Bayes; RF, random forest; SMOTE, synthetic minority oversampling technique; SVM, support vector machines; XGBoost, eXtreme gradient boosting.

different special needs can contribute to optimizing the frequency of therapy sessions.

The role of physiological signals, particularly EDA and ECG, in determining emotional states offers a significant advantage in rehabilitation processes. This study has developed models that utilize ML methods to detect stress and participation in children. The ability of the physiological signals to reflect emotional states has already been validated.[31] The results obtained in this study demonstrate the potential and contribution of the ability to accurately detect stress and participation in the therapeutic processes of children. Emotional data were anonymized, and visual data were collected in compliance with ethical and data protection regulations. Real-world applications of these systems are important for identifying the effects of therapy beyond motivation. The easy use of these systems in real-time can help assess the effectiveness of treatments and therapist-patient compatibility. Some of the data used in this study, such as those from smartwatches, can be easily integrated in the future, simplifying further use.

Delmastro et al[32] focused on stress detection in older adults with mild cognitive impairment using wearable sensors and ML. They analyzed physiological signals such as heart rate variability, EDA, and ST with algorithms like RFs and AdaBoost, achieving 89% accuracy in stress detection. In contrast, this study targeted stress and participation classification in children with special needs (e.g., dyslexia, intellectual disability) during serious game-based therapy. Signals like EDA, BVP, and ST were collected using the Empatica E4 wristband and analyzed through 12 ML models. Note that techniques such as SMOTE and class_weight are used to increase the representativeness of the minority class in imbalanced data sets in this study. However, oversampling methods such as SMOTE change the data distribution by creating new synthetic samples from existing minority class samples. This can lead to overfitting of the model to certain patterns. Especially in complex datasets, it can be observed that the synthetic data produced by SMOTE does not accurately reflect the real-world data. Similarly, the class_weight and scale_pos_weight parameters increase the error cost of the minority class, allowing the model to focus more on this class. However, such weighting methods can sometimes lead the model to overcorrect. When the model over-prioritizes recognizing the minority class, it can lead to increased false positive rates and lower overall accuracy. In this study, these methods have been tested, but the results showed a decrease in the model's generalization ability. Instead, an approach with an autoencoder that better discriminates minority class instances and is more representative of the data distribution was more successful. Therefore, different techniques were chosen to ensure that the model learns in a balanced way for both minority and majority classes.

Autoencoder combined with KNN achieved F1 scores of 66% for stress and 63% for participation detection, while eXtreme Gradient Boosting (XGBoost) reached 91% and 86% for no-stress and no-participation states. The KNN classifier provided a higher F1 score of 66% for stress and 63% for participation after the autoencoder method. In contrast, the XGBoost method achieved a maximum F1 score of 91% for the no-stress group and 86% for the no-participation group. When considering both the minority classes (stress and participation) and the majority classes (no-stress and no-participation), the autoencoder after the KNN model is more effective. This model distinguishes between the classes, achieving average F1 scores of minority and majority classes of 68% for stress and no-stress and 65% for participation and non-participation. Specifically, in this study, the XGBoost method distinguishes stress and engagement states with high accuracy, which has also been noticed in the multimodal emotional recognition models that use physiological signals recorded in therapeutic settings.[33] However, as emphasized, the limited information provided by a single physiological signal and the necessity to consider individual differences may impose some restrictions on the generalizability of the findings from this study.[34] Future research involving a more diverse sample with varied demographic characteristics and the combined use of multiple physiological signals can enhance the reliability and generalizability of the results.

Strengths and Limitations

The use of serious games demonstrated effectiveness in improving motor and cognitive functions in children with special needs, hereby enhancing therapeutic outcomes. The integration of physiological signals, such as electrodermal activity (EDA), blood volume pulse (BVP), and skin temperature (ST), builds on their established use in stress detection and emotional state classification in therapeutic contexts. Finally, the application of machine learning methods showed promise in recognizing stress and participation levels, suggesting potential for advancing individualized rehabilitation approaches. A key limitation is the small, homogeneous sample of 25 children within a narrow age range, which restricts the generalizability of the findings. Future studies should involve larger, more diverse samples to enhance external validity and assess the broader applicability of serious games and physiological signals in pediatric rehabilitation.

Conclusion

The proposed stress and participation model can be used to select suitable games for each child during therapy. Understanding stress and participation can help therapists choose games that align with each child's needs and preferences to enhance therapeutic outcomes. In future work, the selected ML models will be integrated into a real serious game-based rehabilitation environment to adjust game difficulty depending on children's stress and participation to increase their participation and decrease their stress.

A key limitation of this study is the small, homogeneous sample of 25 children with specific diagnoses and a narrow age range. Future research should involve larger, more diverse samples across regions, socioeconomic backgrounds, and age groups to improve generalizability and validate the use of serious games and physiological signals in pediatric rehabilitation.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author.

Ethics Committee Approval: Ethics committee approval was received for this study from the ethics committee of İstanbul Medipol University (Approval no: E-10840098-772.02-6580; Date: February 18, 2021).

Informed Consent: Written informed consent was obtained from parents or legal guardians of the children who participated in this study.

Peer-review: Externally peer-reviewed.

Author Contributions: Concept – D.E.B., D.T., I.U.; Design – D.E.B., D.T.; Supervision – D.E.B., D.T.; Resources – D.E.B., H.B.S., S.A.; Materials – D.E.B., I.U.; Data Collection and/or Processing – D.E.B., B.C., S.A.; Analysis and/or Interpretation – D.E.B., B.C., S.A.; Literature Search – B.C., S.A., D.E.B., H.B.S., D.T.; Writing Manuscript – D.E.B., D.T., H.B.S.; Critical Review – D.E.B., D.T.

Acknowledgments: The authors thank Tugsad Ocal and Emre Erken for their contributions to developing Becure CatchAPet and LeapBall games. They also thank Basak Tuncel in the Department of Ergotherapy, Istanbul Medipol University for their contributions to data acquisition.

Declaration of Interests: The authors declare that they have no competing interests.

Funding: The research work was conducted with the financial supports of The Scientific and Technological Research Council of Turkey Technology and Innovation Funding Programmes Directorate (TUBITAK-TEYDEB 1501) under grant number 3210036.

References

1. Rafiei Milajerdi H, Ordooiazar F, Dewey D. Is active video gaming associated with improvements in social behaviors in children with neurodevelopmental disorders: a systematic review. *Child Neuropsychol*. 2023;29(1):1-27. [CrossRef]
2. Tarakçı E, Arman N, Tarakçı D, Kasapcopur O. Leap Motion Controller–based training for upper extremity rehabilitation in children and adolescents with physical disabilities: a randomized controlled trial. *J Hand Ther*. 2019:33(2)1-9. [CrossRef]
3. Crebbin K, Grisbrook T, Elliott C, Thornton A. The use of serious gaming to improve sensorimotor function and motivation in people with cerebral palsy: a systematic review. *Games Health J*. 2023;12(3):169-197. [CrossRef]
4. Antunes A, Madeira RN, Tóth I, Bernardo C. Modeling serious games design towards engaging children with special needs in therapy. In: *Extended Abstracts of the 2021 Annual Symposium on Computer-Human Interaction in Play*. 2021:192-197. [CrossRef]
5. Jaramillo-Alcázar A, Venegas E, Criollo-C S, Luján-Mora S. An approach to accessible serious games for people with dyslexia. *Sustainability*. 2021;13(5):2507. [CrossRef]
6. Khaleghi A, Aghaei Z, Behnamghader M. Developing two game-based interventions for dyslexia therapeutic interventions using gamification and serious games approaches entertainment computing journal. *Entertain Comput*. 2022;42:100482. [CrossRef]
7. Kokol P, Vošner HB, Završnik J, Vermeulen J, Shohieb S, Peinemann F. Serious game-based intervention for children with developmental disabilities. *Curr Pediatr Rev*. 2020;16(1):26-32. [CrossRef]
8. Kanitkar A, Szturm T, Parmar S, et al. The effectiveness of a computer game-based rehabilitation platform for children with cerebral palsy: protocol for a randomized clinical trial. *JMIR Res Protoc*. 2017;6(5):e93. [CrossRef]
9. Vural ŞF, Yurdusever B, Oktay AB, Uzun I. Stress recognition from facial images in children during physiotherapy with serious games. *Expert Syst Appl*. 2024;238:121837. [CrossRef]
10. De Luca V, Schena A, Covino A, et al. Serious games for the treatment of children with ADHD: the BRAVO project. *Inf Syst Front*. Published online 2024. [CrossRef]
11. Sturm VE, Roy ARK, Datta S, et al. Enhanced visceromotor emotional reactivity in dyslexia and its relation to salience network connectivity. *Cortex*. 2021;134:278-295. [CrossRef]
12. Aktaş SNB, Uluer P, Coşkun B, et al. Stress detection of children with asd using physiological signals. In: *2022 30th Signal Processing and Communications Applications Conference (SIU)*. IEEE; 2022:1-4. [CrossRef]
13. D'Alvia L, Pittella E, Fioriello F, et al. Heart rate monitoring under stress condition during behavioral analysis in children with neurodevelopmental disorders. In: *2020 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. IEEE. 2020:1-6. [CrossRef]
14. Krupa N, Anantharam K, Sanker M, Datta S, Sagar JV. Recognition of emotions in autistic children using physiological signals. *Health Technol*. 2016;6(2):137-147. [CrossRef]
15. Anandhi B, Jerritta S, Anusuya IG, Das H. Time domain analysis of heart rate variability signals in valence recognition for children with autism spectrum disorder (asd). *IRBM*. 2022;43(5):380-390. [CrossRef]
16. Coşkun B, Uluer P, Toprak E, et al. Stress detection of children with autism using physiological signals in KASPar robot-based intervention studies. In: *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*. IEEE; 2022:1-7.
17. Uluer P, Kose H, Gumuslu E, Barkana DE. Experience with an affective robot assistant for children with hearing disabilities. *Int J Soc Robot*. 2023;15(4):643-660. [CrossRef]
18. Zhao B, Wang Z, Yu Z, Guo B. EmotionSense: emotion recognition based on wearable wristband. In: *2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI)*. IEEE; 2018:346-355. [CrossRef]
19. Cosoli G, Poli A, Scalise L, Spinsante S. Measurement of multimodal physiological signals for stimulation detection by wearable devices. *Measurement*. 2021;184:109966. [CrossRef]
20. Bulagang AF, Mountstephens J, Teo J. Multiclass emotion prediction using heart rate and virtual reality stimuli. *J Big Data*. 2021;8:1-12.
21. Sevil M, Rashid M, Askari MR, Maloney Z, Hajizadeh I, Cinar A. Detection and characterization of physical activity and psychological stress from wristband data. *Signals*. 2020;1(2):188-208. [CrossRef]
22. Sevil M, Rashid M, Hajizadeh I, et al. Discrimination of simultaneous psychological and physical stressors using wristband biosignals. *Comput Methods Programs Biomed*. 2021;199:105898. [CrossRef]
23. Chandra V, Priyarup A, Sethia D. Comparative study of physiological signals from empatica e4 wristband for stress classification. In: *Advances in Computing and Data Sciences: 5th International Conference, ICACDS 2021, Nashik, India, April 23-24, 2021*, Revised Selected Papers; part II 5. Springer; 2021:218-229.
24. Tanabe H, Shiraishi T, Sato H, Nihei M, Inoue T, Kuwabara C. A concept for emotion recognition systems for children with profound intellectual and multiple disabilities based on artificial intelligence using physiological and motion signals. *Disabil Rehabil Assist Technol*. Published online 2023:1-8.
25. Coşkun B, Ay S, Erol Barkana D, et al. A physiological signal database of children with different special needs for stress recognition. *Sci Data*. 2023;10(1):382. [CrossRef]
26. Campanella S, Altaleb A, Belli A, Pierleoni P, Palma L. A method for stress detection using empatica E4 bracelet and machine-learning techniques. *Sensors (Basel)*. 2023;23(7):3565. [CrossRef]
27. Elgendi M. *PPG Signal Analysis: an Introduction Using MATLAB®*. Boca Raton: CRC press; 2020. [CrossRef]
28. Golgouneh A, Tarvirdizadeh B. Fabrication of a portable device for stress monitoring using wearable sensors and soft computing algorithms. *Neural Comput Appl*. 2020;32(11):7515-7537. [CrossRef]
29. Jack K, McLean SM, Moffett JK, Gardiner E. Barriers to treatment adherence in physiotherapy outpatient clinics: a systematic review. *Man Ther*. 2010;15(3):220-228. [CrossRef]
30. Yıldirim Y, Budak M, Tarakçı D, Algun ZC. The effect of video-based games on hand functions and cognitive functions in cerebral palsy. *Games Health J*. 2021;10(3):180-189. [CrossRef]
31. Picard RW, Vyzas E, Healey J. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Trans Pattern Anal Mach Intell*. 2001;23(10):1175-1191. [CrossRef]
32. Delmastro F, Martino FD, Dolciotti C. Cognitive training and stress detection in MCI frail older people through wearable sensors and machine learning. *IEEE Access*. 2020;8:65573-65590. [CrossRef]
33. Zhang Y, Cheng C, Zhang YD. Multimodal emotion recognition based on manifold learning and convolution neural network. *Multimed Tools Appl*. 2022;81(23):33253-33268. [CrossRef]
34. Dissanayake V, Seneviratne S, Rana R, Wen E, Kaluarachchi T, Nanayakkara S. Sigrep: toward robust wearable emotion recognition with contrastive representation learning. *IEEE Access*. 2022;10:18105-18120. [CrossRef]