

## Article

# Enhancing Self-Care Prediction in Children with Impairments: A Novel Framework for Addressing Imbalance and High Dimensionality

Eman Ibrahim Alyasin <sup>1,\*</sup>, Oguz Ata <sup>1</sup>, Hayder Mohammedqasim <sup>2</sup> and Roa'a Mohammedqasem <sup>2</sup>

<sup>1</sup> Department of Electrical and Computer Engineering, Institute of Science, Altinbas University, Istanbul 34218, Turkey; oguz.ata@altinbas.edu.tr or oguzata@gmail.com

<sup>2</sup> Faculty of Engineering, Istanbul Aydin University, Istanbul 34153, Turkey; hmohammedqasim@aydin.edu.tr or haydermohammedqasim@gmail.com (H.M.); rmohammedqasem@aydin.edu.tr (R.M.)

\* Correspondence: 203720791@ogr.altinbas.edu.tr or emanalyasin20@gmail.com

**Abstract:** Addressing the challenges in diagnosing and classifying self-care difficulties in exceptional children's healthcare systems is crucial. The conventional diagnostic process, reliant on professional healthcare personnel, is time-consuming and costly. This study introduces an intelligent approach employing expert systems built on artificial intelligence technologies, specifically random forest, decision tree, support vector machine, and bagging classifier. The focus is on binary and multi-label SCADI datasets. To enhance model performance, we implemented resampling and data shuffling methods to tackle data imbalance and generalization issues, respectively. Additionally, a hyper framework feature selection strategy was applied, using mutual-information statistics and random forest recursive feature elimination (RF-RFE) based on a forward elimination method. Prediction performance and feature significance experiments, employing Shapley value explanation (SHAP), demonstrated the effectiveness of the proposed model. The framework achieved a remarkable overall accuracy of 99% for both datasets used with the fewest number of unique features reported in contemporary literature. The use of hyperparameter tuning for RF modeling further contributed to this significant improvement, suggesting its potential utility in diagnosing self-care issues within the medical industry.

**Keywords:** ICF-CY; feature selection; oversampling; predictive models; expert system; Shapley value explanation; physical and motor disability; self-care; machine learning



**Citation:** Alyasin, E.I.; Ata, O.; Mohammedqasim, H.; Mohammedqasem, R. Enhancing Self-Care Prediction in Children with Impairments: A Novel Framework for Addressing Imbalance and High Dimensionality. *Appl. Sci.* **2024**, *14*, 356. <https://doi.org/10.3390/app14010356>

Academic Editor: Mustafa M. Matalgah

Received: 10 November 2023

Revised: 18 December 2023

Accepted: 24 December 2023

Published: 30 December 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Children having mobility or physical disabilities struggle with daily life and self-care because their usual schedule requires additional focus. Self-care evaluation has become a big and challenging issue since it requires a great deal of effort, particularly when there is a shortage of competent physical therapists [1]. The International Classification of Functioning, Disability, and Health for Children and Youth (ICF-CY) is a commonly used framework for diagnosing impairments. For identifying and categorizing self-service issues, this framework has been widely employed. In the medical literature, the problem of diagnosing disability is seen as a difficult challenge that requires the assistance of a qualified therapist, without which therapy becomes more difficult. Many studies use expert systems to diagnose and categorize disabilities [2,3]. In Section 1.2, we describe some of these major works of literature. Machine learning (ML) in computer science is a new direction that is used to develop different systems to support solutions in different fields. These expert systems assist clinicians in making more accurate diagnoses, resulting in more successful treatment [4]. An unbalanced dataset is a common challenge in ML, since the number of patients has a smaller distribution than the group of non-patients, and this is a major issue in ML computation. The self-care dataset used in this research has

a problem from an imbalance as patients with disabilities are less than patients without disabilities. To address this constraint, the resampling approach is a sophisticated ML methodology that creates a new sample at random from the minority sample and its neighbors, enhancing a class that has to be harmonized by raising the number of minority samples [5]. In recent years resampling method has greatly improved the overfitting condition due to the non-exploratory random sampling method, so it has been widely used in the class imbalance region in different areas such as network intrusion detection systems, and sentence boundary in speech [6].

Large amounts of data are another common issue when employing decision support systems in real-world applications. Large amounts of data can add complexity to a system and reduce its accuracy. Dimensional curse [7] is another name for this question. The selection of appropriate feature subsets in medical data mining systems is a hot research topic [8]. Feature selection appears as a preprocessing step in various ML tasks. Most medical datasets contain various features, and extraction of high-impact features for classification problems using the feature selection approach is necessary and difficult. The goal of feature selection is to find and remove duplicate and unnecessary features in the training step. As a result, a created model classifies with greater accuracy and efficiency, while taking less time to calculate. Several studies on the influence of feature selection on classification algorithms have demonstrated this feature extraction not only decreases data dimensionality but also enhances the quality of the model [9]. Within machine learning, feature extraction could be classed as a filtration system, a wrapper, or an embedding [10]. The proposed technique produces a subset of variables established on features also with the greatest ranking of the targeted goals. Rather than training models, they are frequently built on attribute datasets [11]. To select the best set of features, many analytical techniques, including relevance, gain ratio, and chi-squared, utilize scoring outcomes [12]. Because these outcomes are independent of one another, the analytical techniques are unable to delete extraneous features, but they may examine the data core behavior. The wrapper method iteratively searches the entire feature space for a subset of candidate features, using classifiers as a measure of importance. Considering a classifier as an indicator of importance, the wrapper method iteratively explores the whole feature space for a subset of potential features. Scientists compare prediction performance during training to different feature pairs to determine the optimum features using training methods. Forward, backward, and recursive deletion are illustrations of mathematical algorithms that can pick the best characteristics by expanding or changing the candidate solution, despite the high computational overhead [13]. The embedded method combines both filter and wrapper feature selection methods. It keeps a balance between computation speed and precision. To reduce complexity and delete features with fewer than minimal transactions, the penal technology is used in ElasticNet and other embedded methods [14].

### *1.1. Problem Statement*

In the medical industry, the use of machine learning (ML) algorithms is becoming more frequent. As the name implies, machine learning allows software to semi-automatically train data and build remarkable representations. ML techniques using a variety of data types have been used to diagnose children with impairments. ML also allows for the combination of data from multiple imaging systems to detect children with disabilities. To rely on these various measurements for the diagnosis of children with disabilities in early stages or with atypical structures, significant aspects that are not generally used in the diagnosis of disabilities are discovered using ML algorithms. An imbalance in the data might also pose issues with the ML model, perhaps biasing the classification process toward one class over another.

### *1.2. Related Works*

In this work, the ICF-CY self-care dataset was used to develop ML models for the detection of children with disabilities. Zarchi et al. [15] employed SCADI's original dataset

for self-care research and demonstrated the application of SCADI in expert systems. The author employed ANN for classification and a C4.5 decision tree (DT) to extract self-care problems; both models performed 10-fold cross-validation during training and testing. The experiment's findings demonstrated that ANN was highly accurate in anticipating self-care issues, with a prediction rate of 83%. In yet another study [16], several ML classification algorithms were employed to predict accuracy based on a 5-fold CV using principal component analysis (PCA) to minimize the dimensionality of features. With an accuracy of about 84%, KNN was chosen as the best classification approach for predicting self-care issues. Liu et al. [17] proposed a new method for feature selection based on information gain regression curve feature selection (IGTCFS) that uses a SCADI dataset to discover the most important properties for each class. In eight different datasets, the IGTCFS approach was tested against five different feature selection methods. With 10-fold cross-validation to assess Naive Bays (NB) after extracting the most relevant characteristics from the SCADI dataset using IGTCFS; the greatest accuracy obtained is 78%. In another work, Souza et al. [18] proposed a novel ML model on the SCADI dataset in which the dataset is turned into a binary class to identify whether children were identified with negative or positive self-care difficulties. Fuzzy neural networks were used to classify whether or not children had self-care problems; the model achieved a test accuracy of 85%. Akyol [19] used deep neural networks and extreme learning machine models to classify two datasets: Parkinson's disease and self-care problems. The split-train tool and test set were used to classify the models with 60% used for training and 40% for testing. Both models show the best accuracy of 97% and 88%, respectively. Moreover, Putatunda [20] also provides a hybrid auto-coding framework for classifying the self-care problem in multi-class and binary-class datasets. The system combines a method of automated encoding with deep neural networks. The classification model was tested using 10-fold cross-validation, and its accuracy was 84% for multi-class data and 91% for binary-class data. Prasetyowati et al. [21] suggested using the standard 0.05 threshold technique together with a correlation-based feature selection (CBF). Random forest was used to classify the ten original datasets that were processed using FFT and IFFT to evaluate the threshold value determination. The study demonstrated that using a standard deviation threshold when processing the initial dataset improved feature selection accuracy and sped up the random forest classification process on average. Sevinç [22] proposes a better learning model that makes use of a variety of machine learning strategies to predict the patients' severity. An adaptive boost method combined with a decision tree estimator and a new parameter tuning procedure are used in the suggested model.

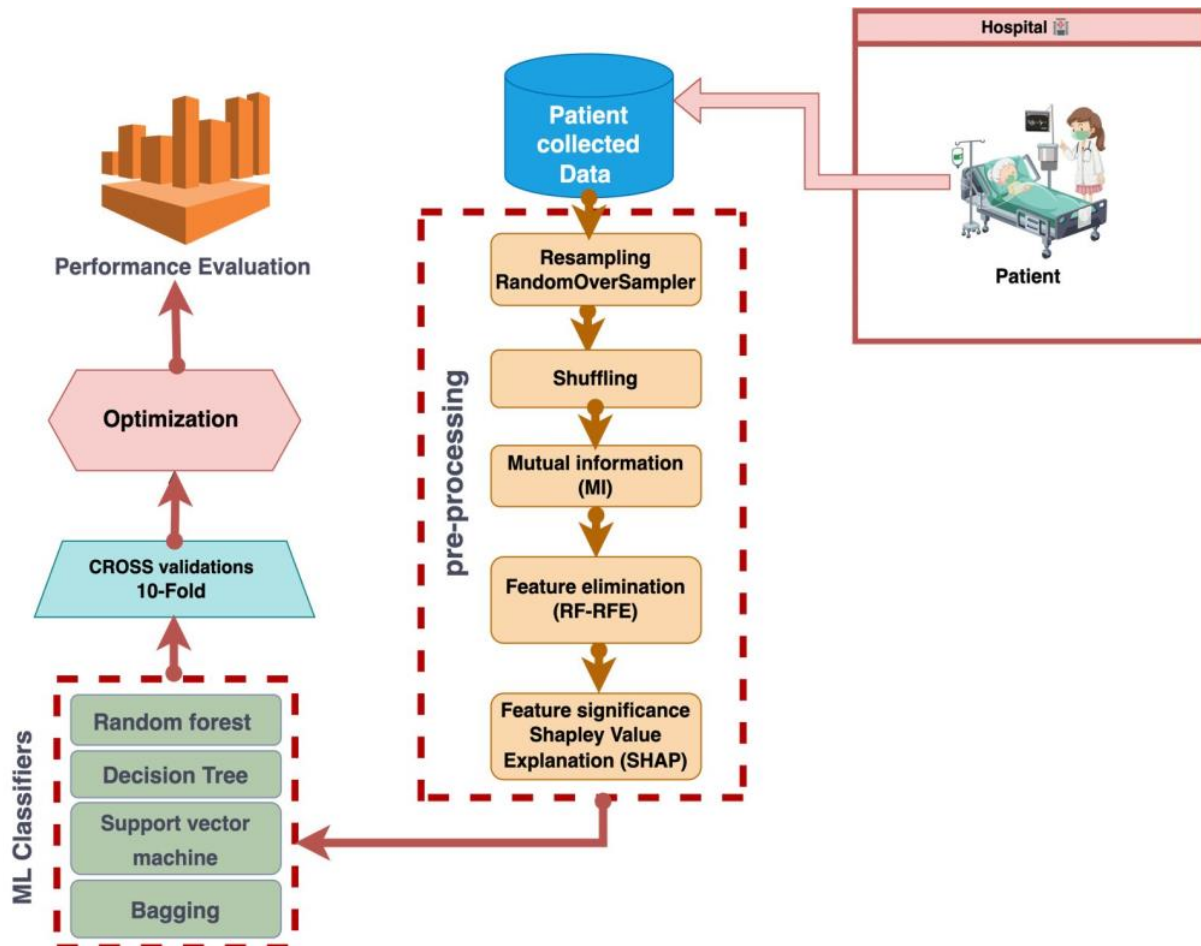
### 1.3. Objectives

The goal of this research is to address the aforementioned challenges in demand to evolve an additional effective predictive model by identifying the major influences on medical datasets. To achieve this, we employed a new hyperfeature selection process based on three main steps: rebalancing, feature selection, and optimization models to predict imbalance in medical datasets. The random over-sampler method was used to address imbalanced class selection in the data balancing step. The minority increase process in the dataset depends on contiguous and converging data values. Because of its computational efficiency, the MI filter method is employed to categorize the features found on their feature significance during the feature selection step. However, when defining the optimal subset of the characteristics, there are no decision limits for the selected parameters. Therefore, as in the second stage, a wrapper strategy known as recursive feature elimination (RFE) is used to remove unnecessary features.

## 2. Materials and Methods

The phases of the proposed methods for the SCADI dataset are detailed in this section. In the first phase, we collect the binary-class SCADI dataset and the multi-class SCADI dataset. In the second phase, the datasets are preprocessed. In the third phase,

we employ machine learning approaches to classify SCADI datasets. Finally, we analyze the performance of our technique using several evaluation measures. The process of the suggested technique for detecting SCADI datasets is displayed in Figure 1. Each phase of the suggested technique is discussed in the subsections that follow.



**Figure 1.** Machine learning diagram for self-care prediction.

### 2.1. Data

This study used the SCADI dataset, which is an openly available original dataset about self-care activity issues in children with impairments (self-care activities based on ICF-CY). The SCADI dataset was collected by experienced health practitioners at educational and health centers in Yazd, Iran, from 2016 to 2017. There are 70 subjects in the dataset, with 41% being female, and ages ranging between 6 and 18 years.

### 2.2. Resampling Method

ML algorithms are often restricted by an unequal dataset. There are two types of unbalanced data: unbalanced binary data, which consists of two classes, and unbalanced multi-class data, which consists of more than two classes [23]. The problem of unbalanced binary and multi-class datasets is addressed in this paper. The numbers of sampled groups have an uneven distribution, with certain groups having a much smaller prevalence than others, and are referred to as minority groups. As mentioned earlier, SCADI dataset have a binary and multi-class unbalanced data, and the random over-sampler technique (ROST) was employed to solve this issue. Unlike sampling methods, sampling algorithms process more minority samples than majority samples to maximize class efficiency. ROST [24] increases the prediction accuracy of a minority sample by overlaying a new minority

sample from a randomly selected identical adjacent sample. In addition, ROST is a robust ML technique whose primary goal is to randomly generate new data from minority group data and their neighbors, thus increasing the fraction of minority group data and the class's requirement for balancing.

### 2.3. Machine Learning (ML)

ML is a technique for giving a computer learning ability. It is frequently conducted using statistical techniques. Utilizing ML techniques, one of the many things that may be taught is categorization [25]. To improve learning results, numerous ML techniques have previously been created. Classification, clustering, and predicting problems may all be resolved with the use of ML techniques.

#### 2.3.1. Support Vector Machine (SVM)

SVM is the most frequently utilized supervised learning algorithm, and it is used to solve classification and regression issues [25]. However, it is employed in ML to solve classification issues. The goal of the SVM computation is to generate the optimal line or decision boundary that can categorize the new information point in n-dimensional space to assist in classification later. This best-case limit is known as a hyperplane. SVM selects strong focus points/vectors that support hyperplane [26].

#### 2.3.2. Random Forest (RF)

RF is an effective and simple strategy that uses supervised ML [25]. The RF technique outperforms a single decision tree by creating decision trees from randomly selected data samples, extracting prediction data from each tree, and selecting the best answer. Overfitting is also eliminated by result averaging. Because trees protect one another from particular flaws, the random forest method yields excellent results. Individual trees may give incorrect answers, but a group of them can obtain the right answer collectively.

#### 2.3.3. Decision Tree (DT)

DT is a member of the supervised learning family methods [25]. DT is a standard model in tree form that displays the relationships between features that lead to that classification, with leaves and branches of different classifications or outcomes. The root of the tree is the first component in the DT, which gains the ability to split according to feature values. An input variable is defined as a tree by passing it through a tree, starting at the root of the tree and ending at the leaf.

#### 2.3.4. Bagging

The bagging classifier is a powerful and effective learning ensemble technique. This approach employs several machine learning algorithms to improve algorithm stability and accuracy, and then the results of its work are evaluated.

### 2.4. Feature Selection

To determine the influence of choosing characteristics at this step, samples were created with and without them. The purpose of feature selection is to identify the most significant characteristics in the SCADI dataset. Furthermore, feature selection contributes to more accurate prediction by removing or underrepresenting less significant information, saving training time, and improving learning performance [27]. There are three methods for selecting characteristics (filter, wrapper, and embedded). Filtering and wrapping approaches are employed in this experiment to identify subsets of important features.

#### 2.4.1. Mutual Information (MI)

Filter-based selection: Statistical techniques are used in strategies to determine the link or dependence between independent features (input features) as well as the dependent feature (targeted feature). The filter methodology analyzes features based on common



features of the data, regardless of learning methods. In contrast, the outcomes obtained by the various statistical analyses used to evaluate them are used to determine which features to employ. MI is a metric evaluating the dependency or decrease in uncertainties between two variables [28]. Below is a list of the steps MI feature selection takes to select features.

Step 1: Input original Features as  $X_{\text{Original}}$ .

Step 2: Determine the connection among input (feature) and output by computing gain ratio scores (target).

Step 3: A strong correlation indicates a greater dependence on the target attribute. To develop the model, `SelectKBest()` was used to choose only the features with highest gain scores.

Step 4: Finally, all the features that will be used in the classification model will be transferred as  $X_{\text{MI}}$  based on the optimal gain scores.

#### 2.4.2. Feature Selection Using Wrapper Methods

This method mainly makes use of a research process to assess changing subsets of independent attributes by feeding them to the selected training algorithm and then assessing the effectiveness of the learning algorithm [29]. These techniques are performed iteratively until the necessary optimal groups are identified when the feature set in the dataset is  $N$ , at which point  $2^N$  subsets can be viewed. Recursive feature elimination (RFE) is a method for training a model iteratively by removing the least important feature as a criterion for each iteration employing method weights as the criterion. The objective of RFE is to choose features by iteratively examining smaller and smaller groupings of features.

#### 2.5. Hyperparameter

Due to architectural complexity, defining hyperparameters manually is a difficult and time-consuming operation. The manual approach frequently fails to produce results that are close to optimal [30]. Additionally, hyperparameters may affect how effectively the method performs for advanced machine learning models. There are some primary automated hyperparameter determination techniques (HPO) such as grid search [31], random search [31], gradient-based optimization [31], grey wolf optimizer [32], whale optimization algorithm [33], etc. A grid search (GS) approach was used, and our goal is to devise classifiers for every conceivable combination of selected hyperparameter values, evaluate these classifiers, and select the optimal model based on the performance of the validation data. Traditional GS entails an exhaustive search, but its complexity and the number of function evaluations grow exponentially with each added hyperparameter, rendering it impractical and intractable due to the curse of dimensionality. In this study, we employ a randomized grid search version, conducting evaluations in a random sequence through uniform sampling without replacement from the grid. This modification seeks to enhance efficiency and overcome the challenges associated with the exhaustive GS approach, offering a more pragmatic solution for hyperparameter tuning.

#### 2.6. Comparative methodology analysis

In Table 1, when comparing different machine learning algorithms applied to the self-care prediction dataset with our proposed model, the Empowered ADABOOST with Decision Tree (E-ADAD) model achieves the highest accuracy score, as documented in the literature [22]. Conversely, the deep neural network (DNN) models proposed by Zarchi et al. [15] and Putatunda [20] yield the lowest accuracy scores. Notably, the presented table indicates that the RF with our proposed model outperforms the other models listed in terms of performance.

**Table 1.** Comparative Analysis of Methodological Settings.

Method	Methodology	Key Techniques	Best Model	Accuracy
Zarchi et al. [15]	ANN, DT	205 neurons in the input layer, 40 neurons in the 235 hidden layer	ANN	83%
Islam et al. [16]	ELM, KNN, SVM, ANN, RF, GB	Principal Component Analysis (PCA)	KNN	84%
Souza et al. [18]	FNN, C.45,MLP,SVM	Artificial neural networks and fuzzy systems.	FNN	85%
Putatunda [20]	DT, DNN	Autoencoders and deep neural networks	DNN	81%
Prasetiyowati et al. [21]	RF	Correlation-Base Feature Selection, Fast fourier transform and inverse fast fourier transform	RF	84%
Sevinç [22]	Adaboost, DT, Bagging, Extra Tree, GB, RF, Hist.GB	Adaptive Boost Algorithm with Decision Tree	E-ADAD	85%
Proposed Model	MI-RFE	Resampling, Data Shuffling, Hyper-framework Feature Selection (MI, RF-RFE), SHAP	RF	99%

The methodological settings, innovations and advantages are presented below in a comparative manner with previous works.

Methodological Innovations:

- ensemble Approach: Unlike previous works, our study employs a combination of Random Forest, Decision Tree, SVM, and Bagging Classifier, providing a more robust and diverse approach.
- Resampling and Data Shuffling: In contrast to previous works, our approach addresses both data imbalance and generalization issues, ensuring a more comprehensive solution.

Feature Selection Strategy:

- Feature Selection Strategy: Hyper-framework Feature Selection (MI, RF-RFE): Differing from previous works feature selection, we utilize a hyper-framework incorporating mutual-information statistics and RF-RFE, resulting in improved model interpretability and efficiency.

Performance Evaluation:

- SHAP Analysis: In contrast to previous works reliance on traditional metrics, we employ Shapley Value Explanation (SHAP) for a more nuanced understanding of feature significance.

Outcome:

- Overall Accuracy (99%): Exceeding the accuracy reported in contemporary literature, our model achieves an outstanding accuracy of 99% for both binary and multi label SCADI datasets.

Efficiency:

- Fewest Unique Features: Distinctively, our model achieves superior accuracy while utilizing the fewest number of unique features, outperforming existing literature.

Generalizability:

- Applicability to Medical Industry: Through hyperparameter tuning, our model shows potential utility in diagnosing self-care issues within the medical industry, demonstrating broader applicability.

### 3. Results

SCADI datasets face two problems: unbalanced datasets, and high dimensional features that lead to inaccurate classification. We trained several ML classification algorithms

using a novel preprocessing approach. In the first step, the ROST was used to solve the imbalance issue in datasets I and II, increasing the minority class range by creating additional synthetic samples. In the second stage, a two-phase feature selection technique was used: filter (MI) method and wrapper (RFE) method, respectively. MI is an effective filtering method to remove irrelevant features. Also, RF was assigned as the RFE estimator because of its speed and efficiency. In this work, four ML approaches were used to evaluate whether patients had a disability: SVM, DT, bagging, and RF. The evaluation was performed using the 10-fold cross-validation method, which is used in ML models to avoid data overfitting. All features were normalized before being applied to classifiers.

Table 2 shows the performance of the binary class dataset's original and preprocessing results. We employed accuracy, precision, recall, f1-score, and kappa, with 10-fold cross-validation score as performance metrics. In this work, we employed oversampling to overcome the imbalance problem in the datasets and then used hyper MI-RFE feature selection due to the high feature space in the SCADI dataset may be decreased by utilizing filter-based (MI) guards to enhance the time complexity of the wrapping approach (RFE). In the evaluation metrics for the original binary class dataset, which comprises 206 features and 70 patient records, SVM was the second-best model, with an accuracy of 88.57%. After applying resampling and MI feature selection, the total patient records reached 108 patients and 80 features, and the accuracy of RF increased to 97.1% selected as the best model. Bagging reached 95.47% accuracy and was selected as the second best model. The reason for this increase is that feature selection can decrease data noise and choose the most important features to employ during the training process. RF-RFE feature selection was applied as the second feature selection, as the optimal feature obtained from MI was used and fed into the RFE to reach the best optimal features that affect the prediction of the ICF-CY dataset based on the best accuracy features obtained from the random forest classifier. The optimal features obtained from RFE are 14 features. DT model was selected as the best evolution model with 97.21% accuracy, and RF was selected as the second best model with 97.10% accuracy.

**Table 2.** Results of the experimental method for a binary SCADI dataset using 10-fold cross-validation.

Model	Processing	Accuracy	Precision	Recall	F1-Score	Kappa
RF	Original	91.42	81.25	81.26	81.37	75.69
	MI	97.10	94.70	99.10	97.29	94.34
	MI-RFE	97.10	94.77	99.12	97.29	94.44
SVC	Original	88.57	78.57	68.75	73.33	66.10
	MI	86.41	88.52	99.12	93.71	87.0
	MI-RFE	93.80	91.22	99.14	93.99	88.0
Bagging	Original	87.14	73.33	68.75	70.96	62.72
	MI	95.47	91.52	99.14	95.57	90.74
	MI-RFE	96.20	93.10	99.22	96.42	92.80
DT	Original	87.14	76.92	62.5	68.96	60.96
	MI	95.30	91.52	99.20	95.55	90.60
	MI-RFE	97.21	94.60	99.30	95.80	90.88

For the multi-class dataset, the same ML model of the binary class was used. Table 3 shows the results of the original and preprocessing of the multi-class dataset. The original dataset contains 70 total patients and 206 features. The bagging and DT models show the best accuracy, with 84.51% and 84.31%, respectively. When the resampling method was used to overcome the problem of multi-class unbalanced data, the data reached a total of 203 patients, and 70 features were obtained using MI feature selection. 'RF' was selected as the best model, with 97.10% accuracy, and bagging as the second model, with 96.37%. RFE was used as a second feature selection in the preprocessing stage after MI. All ML models show good classification improvement by reducing the number of features to 15.



RF and DT were selected as the best models for the multi-class dataset with 97.30% and 97.12%, respectively.

**Table 3.** Results of the experimental method for a multi-SCADI dataset using 10-fold cross-validation.

Model	Processing	Accuracy	Precision	Recall	F1-Score	Kappa
RF	Original	78.43	78.83	84.28	81.37	78.16
	MI	97.10	98.0	98.0	98.0	97.60
	MI-RFE	97.30	98.51	98.11	98.30	98.20
SVC	Original	78.35	77.41	84.28	80.50	77.90
	MI	89.33	91.70	90.60	90.20	89.0
	MI-RFE	96.50	97.20	97.0	97.0	96.50
Bagging	Original	84.51	81.99	88.57	84.91	84.16
	MI	96.37	97.12	97.0	97.0	96.50
	MI-RFE	97.10	97.40	97.50	97.52	97.20
DT	Original	84.31	83.90	88.57	85.84	84.0
	MI	96.0	96.62	96.50	96.49	95.97
	MI-RFE	97.12	97.50	97.40	97.30	97.18

All ML models for binary class I and multi-class II datasets were built using the default parameters as shown in Tables 1 and 2. Hyperparameters are used in ML algorithms to ensure the best results for the classification task. We employed grid-search to fine-tune the hyperparameters for every classification model for both datasets to obtain the best classification results (Table 4).

**Table 4.** Hyperparameter for SCADI datasets.

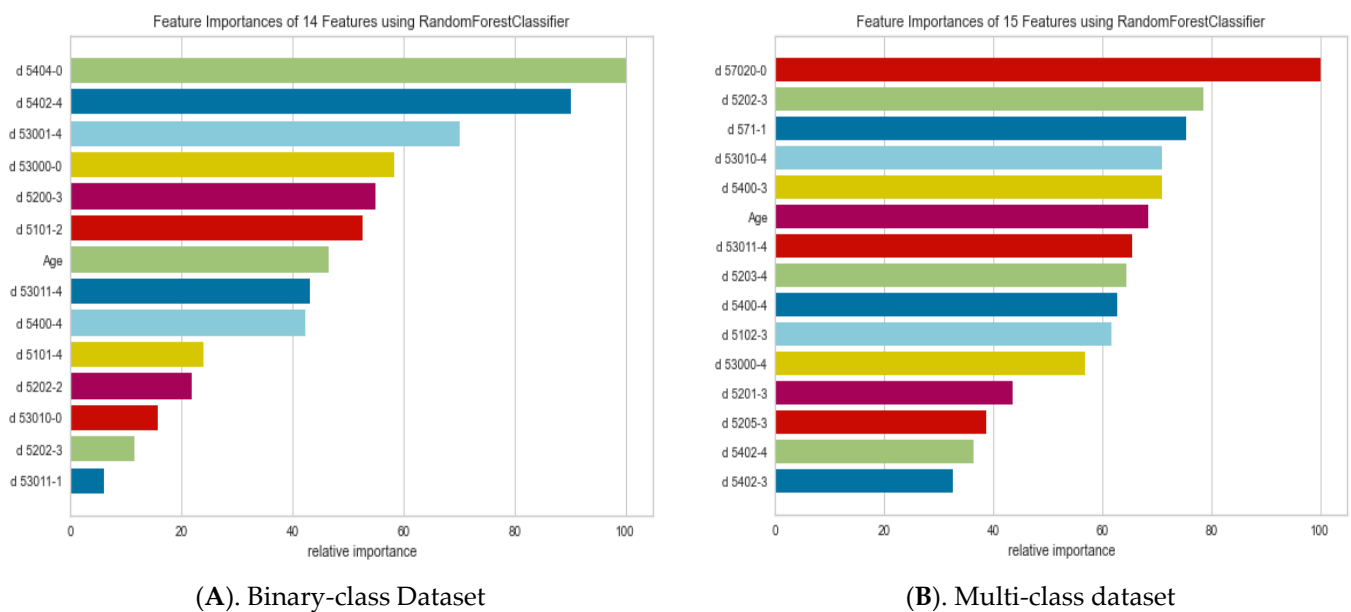
Class	Model	Hyperparameters
Binary	RF	Number of trees = 90, Criterion = gini, Max depth = 6
	SVC	C = 8, kernel = linear, gamma = 2
	DT	Criterion = gini, max depth = 12, max_features = 21
	Bagging	Estimator = RF, number estimators = 50, max samples = 0.9
Multi	RF	Number of trees = 150, Criterion = gini, Max depth = 10
	SVC	C = 11, kernel = linear, gamma = 22
	DT	Criterion = gini, max depth = 18, max features = 12
	Bagging	Estimator = RF, number estimators = 40, max samples = 0.7

As shown in Table 5, after the optimal hyperparameters were found, all the evaluation metrics of ML models show a good increase in performance where the RF model reached 99.10% classification accuracy in the binary dataset and 98.60% classification accuracy in the multi-class dataset.

**Table 5.** Experimental results for the SCADI datasets based on the optimal hyperparameters.

Class	Model	Accuracy	Precision	Recall	F1-Score	Kappa
Binary	RF	99.10	98.50	99.80	99.17	98.60
	SVC	95.55	91.52	99.90	95.57	90.77
	Bagging	98.0	96.43	99.20	98.10	96.30
	DT	98.20	96.45	99.81	98.20	98.30
Multi	RF	98.60	98.70	99.80	98.59	98.60
	SVC	97.55	98.21	99.90	95.57	90.77
	Bagging	98.30	98.55	98.20	98.60	98.27
	DT	98.12	98.05	98.01	98.10	97.70

Determining feature importance is an important strategy for understanding the ML model as well as the features of the underlying data. The Shapley value explanation (SHAP) is a method for properly evaluating the relevance of input features in a model output. Figure 2 depicts the most relevant features for both binary and multi-class dataset after employing MI-RFE feature selection methods based on the RF model.



**Figure 2.** Feature importance for SCADI datasets.

As shown in Table 6, we compared the effectiveness of our proposed approach to previous studies that employed the same SCADI dataset. Table 6 contains some abbreviations such as FS (feature selection), NF (number of features), and ACC (accuracy). A comparative evaluation of the proposed approach with relevant previous studies is summarized in Table 6. Overall, the proposed method improved all previous research results on the SCADI dataset, including: ANN [15], KNN [16], FNN [18], hybrid autoencoder [20], RF [21], and AdaBoost [22]. Comparing the results of other relevant work, our proposed model has the best performance by obtaining an accuracy of 98.60% and 99.10% for binary and multi-class datasets, respectively.

**Table 6.** Comparison results for SCADI datasets between this study and previous studies.

Method	FS	NF	Method Validation	Classes	ACC	Year
Zarchi et al. [15]	-	205	10-fold CV	Multi-class	83%	2018
Islam et al. [16]	PCA	53	5-fold CV	Multi-class	84%	2018
Souza et al. [18]	-	205	k-fold	Binary-class	85%	2019
Putatunda [20]	-	205	10-fold CV	Binary-class	84%	2020
	-	205	10-fold CV	Multi-class	81%	2020
Prasetiyowati et al. [21]	CBF	19	10-fold CV	Binary-class	84.14%	2021
Sevinç [22]	-	205	5-fold CV	Multi-class	85%	2022
Proposed Model	MI-RFE	14	10-fold CV	Binary-class	99.10%	-
	MI-RFE	15	10-fold CV	Multi-class	98.60%	-

Compared to those in the literature, the model used shows high prediction in accuracy, and the reason for this is selecting the most important features during the training process and reducing data noise by using hyper feature selection model based on optimal hyperparameters.

#### 4. Discussion

The ICF and ICF-CY are two common conceptual models used to assess and classify disabilities. Using survey questionnaires and manual methods to assess and classify disability is an expensive and time-consuming process [15]. Thus, a lack of experts can significantly increase costs and time. Specialized methods can help reduce the cost and time of the classification process. A standard dataset is essential for the learning phase when creating and training an expert system.

In this research, the aim is to contribute to previous research and achieve optimal performance in binary and multi classes imbalance SCADI dataset, so a new expert system based on the ML models has been proposed for SCADI datasets, which includes data pre-processing, re-sampling, feature selection model, evaluation, and hyper-tuning parameters. The random over-sampler technique is used to overcome the imbalance problem in both binary and multi-classes SCADI dataset.

Feature selection: It is a very useful preprocessing technique in medical applications because it not only reduces dimensions but also helps us identify the causes of disease [34]. The wrapping method is highly computational but has high classification performance, while filters are known to be fast but less accurate. In this study, filter and wrapper methods are combined to provide an advanced multi-objective model. As shown in Figure 2, the optimal features for binary and multi-class datasets were reduced to 14 and 15, respectively, using hyper filter and wrapper methods compared to the literature. The overall performance of four ML models was tested on SCADI datasets processed via feature processing and resampling methods to evaluate the approach to quickly and accurately diagnose binary and multi-class SCADI datasets. The effectiveness of each ML model was evaluated using 10 different cross-validation methods. The grid search method was used to find the best parameters for four ML models. The experimental results show that the RF classifier based on Gini impurity performs the best on both binary and multi-class SCADI datasets.

#### 5. Conclusions

This research aims to anticipate self-care issues in children with impairments, recognizing the intricate and time-consuming nature of identification by physical therapists. In response, our study focuses on the development of self-care prediction models, applying a unique and effective preprocessing framework to address two critical challenges: unbalanced multi-class and binary-class classification, and the curse of high dimensionality. The random over-sampler technique (ROST) was employed to overcome distribution bias in each self-care dataset, while mutual information (MI) and recursive feature elimination (RFE) were utilized as feature selection methods to identify crucial classification features. Evaluation metrics were applied to the original binary-class dataset, selecting the random

forest (RF) model as the best, achieving 91% accuracy and 81% precision. For the original multi-class data, the Bagging model emerged as the optimal classification model with 84% accuracy and 81% precision. Post-pre-processing, both self-care datasets achieved a balanced distribution and optimal feature numbers, with the RF model reaching 97% accuracy and 94% precision for the binary-class data and 98% accuracy and precision for the multi-class data. To attain optimum classification performance, the RF model achieved over 99% accuracy for the binary-class data and over 98.60% accuracy for the multi-class data using the Grid-search optimization method.

**Author Contributions:** Conceptualization, E.I.A.; Data curation, E.I.A. and O.A.; Formal analysis, E.I.A.; Investigation, H.M. and R.M.; Methodology, E.I.A. and O.A.; Project administration, H.M. and O.A.; Supervision, H.M., O.A. and R.M.; Writing—original draft, E.I.A.; Writing—review and editing, H.M. and R.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The test data used in the Results section are publicly available on UC Irvine Machine Learning Repository at: <https://archive.ics.uci.edu/dataset/446/scadi/> (accessed on 27 December 2023).

**Conflicts of Interest:** The authors declare that no conflict of interest.

## References

1. Le, T.; Son, L.H.; Vo, M.T.; Lee, M.Y.; Baik, S.W. A Cluster-Based Boosting Algorithm for Bankruptcy Prediction in a Highly Imbalanced Dataset. *Symmetry* **2018**, *10*, 250. [CrossRef]
2. Lan, K.; Wang, D.-T.; Fong, S.; Liu, L.-S.; Wong, K.K.L.; Dey, N. A Survey of Data Mining and Deep Learning in Bioinformatics. *J. Med. Syst.* **2018**, *42*, 139. [CrossRef] [PubMed]
3. Goshvarpour, A. A Novel Feature Level Fusion for Heart Rate Variability Classification Using Correntropy and Cauchy-Schwarz Divergence. *J. Med. Syst.* **2018**, *42*, 109. [CrossRef] [PubMed]
4. Rao, H.; Shi, X.; Rodrigue, A.K.; Feng, J.; Xia, Y.; Elhoseny, M.; Yuan, X.; Gu, L. Feature selection based on artificial bee colony and gradient boosting decision tree. *Appl. Soft Comput.* **2019**, *74*, 634–642. [CrossRef]
5. El Houby, E.M. A survey on applying machine learning techniques for management of diseases. *J. Appl. Biomed.* **2018**, *16*, 165–174. [CrossRef]
6. Chen, G.; Chen, J. A novel wrapper method for feature selection and its applications. *Neurocomputing* **2015**, *159*, 219–226. [CrossRef]
7. Sharifai, G.A.; Zainol, Z. Feature Selection for High-Dimensional and Imbalanced Biomedical Data Based on Robust Correlation Based Redundancy and Binary Grasshopper Optimization Algorithm. *Genes* **2020**, *11*, 717. [CrossRef]
8. de Sá, A.G.; Pereira, A.C.; Pappa, G.L. A customized classification algorithm for credit card fraud detection. *Eng. Appl. Artif. Intell.* **2018**, *72*, 21–29. [CrossRef]
9. Chen, H.; Li, T.; Fan, X.; Luo, C. Feature selection for imbalanced data based on neighborhood rough sets. *Inf. Sci.* **2019**, *483*, 1–20. [CrossRef]
10. Elhoseny, M.; Mohammed, M.A.; Mostafa, S.A.; Abdulkareem, K.H.; Maashi, M.S.; Garcia-Zapirain, B.; Mutlag, A.A.; Maashi, M.S. A new multi-agent feature wrapper machine learning approach for heart disease diagnosis. *Comput. Mater. Contin.* **2021**, *67*, 51–71. [CrossRef]
11. Albashish, D.; Hammouri, A.I.; Braik, M.; Atwan, J.; Sahran, S. Binary biogeography-based optimization based SVM-RFE for feature selection. *Appl. Soft Comput.* **2021**, *101*, 107026. [CrossRef]
12. Abdel-Basset, M.; El-Shahat, D.; El-Henawy, I.; de Albuquerque, V.H.C.; Mirjalili, S. A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection. *Expert Syst. Appl.* **2020**, *139*, 112824. [CrossRef]
13. Elavarasan, D.; Vincent P M, D.R.; Srinivasan, K.; Chang, C.-Y. A Hybrid CFS Filter and RF-RFE Wrapper-Based Feature Extraction for Enhanced Agricultural Crop Yield Prediction Modeling. *Agriculture* **2020**, *10*, 400. [CrossRef]
14. Amini, F.; Hu, G. A two-layer feature selection method using Genetic Algorithm and Elastic Net. *Expert Syst. Appl.* **2021**, *166*, 114072. [CrossRef]
15. Zarchi, M.; Bushehri, S.F.; Dehghanizadeh, M. SCADI: A standard dataset for self-care problems classification of children with physical and motor disability. *Int. J. Med. Inform.* **2018**, *114*, 81–87. [CrossRef] [PubMed]
16. Islam, B.; Ashafuddula, N.I.M.; Mahmud, F. A Machine Learning Approach to Detect Self-Care Problems of Children with Physical and Motor Disability. In Proceedings of the 2018 21st International Conference of Computer and Information Technology, ICCIT 2018, Dhaka, Bangladesh, 21–23 December 2018.

17. Liu, L.; Zhang, B.; Wang, S.; Li, S.; Zhang, K.; Wang, S. Feature selection based on feature curve of subclass problem. In Proceedings of the 2019 International Joint Conference on Neural Networks (IJCNN), Budapest, Hungary, 14–19 July 2019.
18. Souza, P.V.C.; dos Reis, A.G.; Marques, G.R.R.; Guimaraes, A.J.; Araujo, V.J.S.; Araujo, V.S.; Rezende, T.S.; Batista, L.O.; da Silva, G.A. Using hybrid systems in the construction of expert systems in the identification of cognitive and motor problems in children and young people. In Proceedings of the 2019 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), New Orleans, LA, USA, 23–26 June 2019.
19. Akyol, K. Comparing of deep neural networks and extreme learning machines based on growing and pruning approach. *Expert Syst. Appl.* **2020**, *140*, 112875. [[CrossRef](#)]
20. Putatunda, S. Care2Vec: A hybrid autoencoder-based approach for the classification of self-care problems in physically disabled children. *Neural Comput. Appl.* **2020**, *32*, 17669–17680. [[CrossRef](#)]
21. Prasetyowati, M.I.; Maulidevi, N.U.; Surendro, K. Determining threshold value on information gain feature selection to increase speed and prediction accuracy of random forest. *J. Big Data* **2021**, *8*, 84. [[CrossRef](#)]
22. Sevinç, E. An empowered AdaBoost algorithm implementation: A COVID-19 dataset study. *Comput. Ind. Eng.* **2022**, *165*, 107912. [[CrossRef](#)]
23. Qasim, H.M.; Ata, O.; Ansari, M.A.; Alomary, M.N.; Alghamdi, S.; Almeahmadi, M. Hybrid Feature Selection Framework for the Parkinson Imbalanced Dataset Prediction Problem. *Medicina* **2021**, *57*, 1217. [[CrossRef](#)]
24. Elyan, E.; Moreno-Garcia, C.F.; Jayne, C. CDSMOTE: Class decomposition and synthetic minority class oversampling technique for imbalanced-data classification. *Neural Comput. Appl.* **2021**, *33*, 2839–2851. [[CrossRef](#)]
25. Ayon, S.I.; Islam, M.; Hossain, R. Coronary Artery Heart Disease Prediction: A Comparative Study of Computational Intelligence Techniques. *IETE J. Res.* **2020**, *68*, 2488–2507. [[CrossRef](#)]
26. Senan, E.M.; Al-Adhaileh, M.H.; Alsaade, F.W.; Aldhyani, T.H.H.; Alqarni, A.A.; Alsharif, N.; Uddin, M.I.; Alahmadi, A.H.; E Jadhav, M.; Alzahrani, M.Y. Diagnosis of Chronic Kidney Disease Using Effective Classification Algorithms and Recursive Feature Elimination Techniques. *J. Healthc. Eng.* **2021**, *2021*, 1004767. [[CrossRef](#)] [[PubMed](#)]
27. Speiser, J.L. A random forest method with feature selection for developing medical prediction models with clustered and longitudinal data. *J. Biomed. Inform.* **2021**, *117*, 103763. [[CrossRef](#)]
28. Solorio-Fernández, S.; Carrasco-Ochoa, J.A.; Martínez-Trinidad, J.F. A review of unsupervised feature selection methods. *Artif. Intell. Rev.* **2020**, *53*, 907–948. [[CrossRef](#)]
29. Mohammedqasem, R.; Mohammedqasim, H.; Ata, O. Real-time data of COVID-19 detection with IoT sensor tracking using artificial neural network. *Comput. Electr. Eng.* **2022**, *100*, 107971. [[CrossRef](#)]
30. Mohammedqasim, H.; Mohammedqasem, R.; Ata, O.; Alyasin, E.I. Diagnosing Coronary Artery Disease on the Basis of Hard Ensemble Voting Optimization. *Medicina* **2022**, *58*, 1745. [[CrossRef](#)]
31. Kadam, V.J.; Jadhav, S.M. Performance analysis of hyperparameter optimization methods for ensemble learning with small and medium sized medical datasets. *J. Discret. Math. Sci. Cryptogr.* **2020**, *23*, 115–123. [[CrossRef](#)]
32. Zhang, R.; Wu, X.; Chen, Y.; Xiang, Y.; Liu, D.; Bian, X. Grey Wolf Optimizer for Variable Selection in Quantification of Quaternary Edible Blend Oil by Ultraviolet-Visible Spectroscopy. *Molecules* **2022**, *27*, 5141. [[CrossRef](#)]
33. Bian, X.; Zhao, Z.; Liu, J.; Liu, P.; Shi, H.; Tan, X. Discretized butterfly optimization algorithm for variable selection in the rapid determination of cholesterol by near-infrared spectroscopy. *Anal. Methods* **2023**, *15*, 5190–5198. [[CrossRef](#)]
34. Piri, J.; Mohapatra, P.; Singh, H.K.R.; Acharya, B.; Patra, T.K. An Enhanced Binary Multiobjective Hybrid Filter-Wrapper Chimp Optimization Based Feature Selection Method for COVID-19 Patient Health Prediction. *IEEE Access* **2022**, *10*, 100376–100396. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.