Classification and action rules in identification and self-care assessment problems

Małgorzata Zdrodowska[a,*] and Agnieszka Dardzińska-Głębocka[b]
[a]*Institute of Biomedical Engineering, Faculty of Mechanical Engineering, Bialystok Technical University, Bialystok, Poland*
[b]*Institute of Mechanical Engineering, Faculty of Mechanical Engineering, Bialystok Technical University, Bialystok, Poland*

Abstract.
BACKGROUND: Disability, especially in children, is a very important and current problem. Lack of proper diagnosis and care increases the difficulty for children to adapt to disabilities. Disabled children have many problems with basic activities of daily living. Therefore, it is very important to support diagnosticians and physiotherapists in recognizing self-care problems in children.
OBJECTIVE: The aim of this paper is to extract classification and action rules, useful for those who work with children with disabilities.
METHODS: First, features and their impact on the accuracy of classification are determined. Then, two models are built: one with all features and one with selected ones. For these models the classification rules are extracted. Finally, action rules are mined and the next step in treatment process is predicted.
RESULTS: Seventeen features with the greatest impact on classifying a child into a particular group of self-care problems were identified. Based on the implemented algorithms, decision and action rules were obtained.
CONCLUSIONS: The obtained model, selected attributes and extracted classification and action rules can support the work of therapists and direct their work to those areas of disability where even a minimal reduction of features would be of great benefit to the children.

Keywords: Disability, ICF-CY, self-care problem, classification rules, action rules, feature selection, classification, data mining

1. Introduction

A disability is a limitation or lack of ability to perform an activity in the manner or to the extent considered normal for a person, resulting from damage to and impairment of bodily functions. According to the World Health Organization (WHO), persons with disabilities are defined as those who are unable

---

*Corresponding author: Małgorzata Zdrodowska, Institute of Biomedical Engineering, Faculty of Mechanical Engineering, Bialystok Technical University, ul. Wiejska 45C, 15-351 Bialystok, Poland. Tel.: +48 797 990 774; Fax: +48 85 746 92 10; E-mail: m.zdrodowska@pb.edu.pl.

to provide for themselves, partially or wholly, the ability for normal personal and social life as a result of congenital or acquired physical or mental impairment [1]. The division of the effects of disease into damage, incapacity, and role limitation has been accepted. Damage refers to abnormalities in the structure and appearance of the body and the function of organs or systems, regardless of the cause. Thus, it means a disorder at the organ level. Incapacity reflects the effects of damage on a person's performance of activities and actions. Role limitations are the consequences of damage and incapacity that affect interpersonal interactions and adaptation to the environment. Persons under the age of 16 are counted as disabled if they have a physical or mental impairment with an expected duration of more than 12 months. This impairment may be due to a congenital defect, long-term illness or bodily injury. This makes it necessary to provide these children with total care or assistance in meeting the basic needs of life in a manner that exceeds the support needed by a person of their age. Persons under the age of 16 are counted as disabled if they have a physical or mental impairment with an expected duration of more than 12 months. This impairment may be due to a congenital defect, long-term illness or injury and in many cases has a major impact on the child's progress through the developmental stages. These limitations make it necessary to provide these children with total care or assistance in meeting the basic needs of life in a manner that exceeds the support needed by a person of their age. Therapy, a multi-pronged process that focuses on physical activity, intellectual development and proper social functioning, is therefore very important. Along with the support of loved ones, this is often the basis of life for children with disabilities. Children with disabilities very often experience problems in their lives with performing basic everyday activities. Therefore, self-care is a very important issue, which more or less ensures independence and comfort in everyday life. It is crucial to choose the right therapy, which can only be achieved with a proper diagnosis and classification of the disability [2–6].

Diagnosing disabilities in children is, in many cases, a very elaborate process that requires a great amount of knowledge, experience, and collaboration among physicians, physical therapists, neuropsychologists, therapists, special educators, and other professionals. A very important aspect here is the identification of individual characteristics of disabilities, as well as their appropriate classification. The International Classification of Functioning, Disability and Health of Children and Young People version (ICF-CY) may be helpful for this goal [5]. ICF-CY is based on the International Classification of Functioning, Disability and Health (ICF) and thus uses common language and terminology to identify problems with disability. However, the ICF-CY delineates the child in relationship to the environment, and specifically the child in the context of the family. It is a modern, family-centred approach that recognises that it is impossible to observe, diagnose, treat, educate and support a child in isolation from the family situation. This affords an opportunity to plan and take effective action for the child in conjunction with an action to strengthen the family. The ICF-CY can be used to describe functioning and disability in children and to classify health and related conditions. The ICF-CY does not classify children. It enables a classification of the characteristics of children's health status in the context of individual life situations and environmental influences. It provides an opportunity to plan and implement effective support for children with disabilities. The ICF-CY can be used by all health care providers and consumers, and by all those who are involved in health, education and care of children and young people and seek to create optimal conditions for their development [2–6].

The ICF-CY classification is divided into two parts (functioning and disability and contextual factors), each of which is divided into two classification components (body functions and structures, activity and participation, environmental factors and personal factors, respectively). Problems related to self-care will be found in the section of activity and participation. They mainly concern activities such as daily toileting, caring for body parts, grooming, etc., which are the main activities of daily life that should be

fully developed by the age of about 7 years. The development of these daily living skills is influenced by contextual factors, that is, environmental factors and personal factors [2–6].

The analysis of self-care meaning shows that this ability has an important role in the quality of life of the children with physical and motor disabilities. The obtained information becomes valuable. Thus, it is very important to study and classify children's self-care problems. Action rules can reclassify patients from one group to another; therefore can be helpful in treatment methodology.

Literature survey uncovers a number of different, but also related, methods for knowledge extraction from datasets of children with disabilities. The authors mainly study the efficiency of selected classification algorithms, and investigate the effect of PCA (principal component analysis) on classification accuracy, as well as find some decision rules [3,4,6–9]. For example, Sharma [3] used Partitioned Multifilter with Partial Swarm Optimization to feature selection and then applied Naïve Bayes, Multilayer Perception, C-4.5 and Random Tree classifier receiving 64–80% accuracy. Islam et al. [9] applied Principal Component Analysis and used Extreme Learning Machine, K Nearest Neighbour, Support Vector Machine, Artificial Neural Network, Random Forest, as well as Gradient Boosting. More advanced analysis can be found in the work of Fatemi Bushehri and Zarchi [4]. They proposed an expert model based on Probabilistic Neural Network and Genetic Algorithm, and also used CART algorithm to find classification rules. In our work, we go a few steps further. In addition to feature selection and classifier evaluation of the obtained models, we will extract not only classification rules, but also action rules which have not been implied before in any works related to self-care problem.

The purpose of this paper is to extract classification rules and extract action rules that can be helpful for diagnosticians and physiotherapists who work with children with disabilities. This study can put a focus on those areas of self-care problems where even a minimal reduction of features would be of great benefit to the child.

2. Materials and methods

For research we used dataset SCADI of 70 children with physical and motor disability [4,10]. These data were obtained from UCI Machine Learning Depository [11]. The SCADI database was compiled and published in collaboration with occupational therapists as the first standard set of data on self-care activities of children with physical and motor disabilities [10]. The dataset contains information on 70 children aged 6 to 22 years (29 female and 41 male) from 3 health and educational centres in Iran. The data were collected from 2016 to 2017. The following diseases were diagnosed in the studied children: [10]:

- Cerebral palsy (Diplegia – 26 children, Quadriplegia – 12 children, Hemiplegia – 4 children);
- Myelomeningocele – 8 children;
- Muscular dystrophy – 7 children;
- Clubfoot – 3 children;
- Nanism – 2 children;
- Pompe disease – 1 child;
- Head injury – 1 child;
- Mucopolysaccharides – 1 child;
- Maple syrup urine disease – 1 child;
- Osteogenesis imperfecta – 1 child;
- Congenital hand anomaly – 1 child;
- Medulla oblongata tumour – 1 child;
- Hydrocephalus – 1 child.

Table 1

Self-care activities codes based on International Classification of Functioning, Disability and Health of Children and Young People [5,10]

Characterization	Code	Classification
Washing body parts	d5100-x	Washing oneself
Washing whole body	d5101-x	Washing oneself
Drying oneself	d5102-x	Washing oneself
Caring for skin	d5200-x	Caring for body parts
Caring for teeth	d5201-x	Caring for body parts
Caring for hair	d5202-x	Caring for body parts
Caring for fingernails	d5203-x	Caring for body parts
Caring for toenails	d5204-x	Caring for body parts
Caring for nose	d5205-x	Caring for body parts
Indicating need for urination	d53000-x	Toileting
Carrying out urination appropriately	d53001-x	Toileting
Indicating need for defecation	d53010-x	Toileting
Carrying out defecation appropriately	d53011-x	Toileting
Menstrual care	d5302-x	Toileting
Putting on clothes	d5400-x	Dressing
Taking off clothes	d5401-x	Dressing
Putting on footwear	d5402-x	Dressing
Taking off footwear	d5403-x	Dressing
Choosing appropriate clothing	d5404-x	Dressing
Indicating need for eating	d5500-x	Eating
Carrying out eating appropriately	d5501-x	Eating
Indicating need for drinking	d5600-x	Drinking
Indicating need for drinking	d5600-x	Drinking
Ensuring one's physical comfort	d5700-x	Looking after one's health
Managing diet and fitness	d5701-x	Looking after one's health
Managing medications and following health advice	d57020-x	Looking after one's health
Seeking advice or assistance from caregivers or professionals	d57021-x	Looking after one's health
Avoiding risks of abuse of drugs or alcohol	d57022-x	Looking after one's health
Looking after one's safety	d571-x	Looking after one's safety

Impairment codes x: 0 – no impairment, 1 – mild impairment, 2 – moderate impairment, 3 – severe impairment, 4 – complete impairment, 8 – not specified, 9 – not applicable.

Each child was characterized using 206 features: gender (1 – male, 0 – female), age, self-care activities based on International Classification of Functioning, Disability and Health of Children and Young People (202 features shown in Table 1: 1 – the case has this attribute, 0 – otherwise) and an attribute that classifies the child into a specific group [10]:

- Group 1 – problems with caring for body parts – 2 children;
- Group 2 – problems with toileting – 7 children;
- Group 3 – problems with dressing – 1 children;
- Group 4 – problems with washing oneself and caring for body parts and dressing – 12 children;
- Group 5 – problems with washing oneself, caring for body parts, toileting and dressing – 3 children;
- Group 6 – problems with eating, drinking, washing oneself, caring for body parts, toileting, dressing, looking after one's health and looking after one's safety – 29 children;
- Group 7 – no problems – 16 children.

The first step of the study was feature selecting and checking its impact on the accuracy of classifying children to particular groups. The following methods were used for attribute reduction:

1. Correlation Based Feature Selection (CFS) – method originally developed by Hall [12]. The CFS method is classified as a subset selector because it evaluates the subsets rather than individual

attributes. For this reason, CFS has to perform a candidate scan for all possible subsets. Due to the exponential complexity of the problem, it is unnecessary to perform a full scan of all possible subsets, so heuristic must be used to guide a partial scan. We can say that this heuristic is the core concept of the CFS algorithm. It is a filtering method that applies a principle derived from Ghiselly test theory – good subsets of features contain features highly correlated with the class but uncorrelated with each other [12–15]. The CFS feature subset evaluation function is defined as [12,14]:

$$M_s = \frac{kr_{cf}}{\sqrt{k + k(k-1)r_{ff}}} \tag{1}$$

The number of features in S is k, i.e., $k = |S|$. Coefficient r_{ff} is the mean correlation between the features in S. Coefficient r_{cf} is the mean correlation between the features and unserved load S that indicates resilience. The numerator in Eq. (1) indicates the utility of a feature set in predicting the output, whereas the denominator gives a measure of the redundancy between the features of the set. A set containing redundant features has a high r_{ff}, since the features are highly correlated with each other, and therefore they have a low merit score. Therefore, those features that are highly predictive of the output data but statistically dependent on other features are discriminated against. The CFS algorithm computes the merits for all subsets and orders them accordingly. In general, the subset with the highest merit score is used to select the most appropriate attributes for the training algorithm, resulting in improved performance [14,15].

2. Information Gain Based Feature Selection (IGFS) – it takes the characteristics of the dataset and performs dimensionality reduction. This method calculates the information gain (also known as entropy) for each attribute for the output variable. The input values range from 0 (no information) to 1 (maximum information). Those attributes that contribute more information will have a higher information gain value and can be selected, while those that do not contribute much information will have a lower score and can be removed [16,17]. Equations (2) and (3) are used to calculate the entropy of X and Y [17]:

$$En(X) = -\sum p(x) \log p(x) \tag{2}$$

$$En(Y) = -\sum p(y) \log p(y) \tag{3}$$

X and Y is the entropy before the observation of the variable sequentially in class X or Y, $p(x)$ is the probability distribution of the observation x, $p(y)$ is the probability distribution of the observation y, and the variables x, y are members of the class X and Y, respectively [(5), (6)]. When the entropy is computed for two instances, the difference is computed to know the value of the profit. In fact, the gain from X to Y is the reduction of the entropy value and is calculated using Eq. (4) [17]:

$$IG(Y, X) = En(Y) - En(Y/X) \tag{4}$$

Features that provide a greater gain value are considered important features for predicting class labels [17].

3. Chi-Square Test for Feature Selection – is a numerical test that measures the deviation from the expected distribution given that the trait event is independent of the class value. Pearson's Chi-Square statistical test is a method to determine if there is a significant difference between the expected values and the values observed in the distribution between two variables [18,19]. To verify this correlation the following equation [18] is used:

$$\chi^2 = \sum_i^r \sum_j^c \frac{(n_{ij} - E_{ij})^2}{E_{ij}} \tag{5}$$

where r is the number of rows, c is the number of columns, n_{ij} is the observed value and E_{ij} is the expected value. $E = \frac{n_{j} \cdot n_{.j}}{n}$, where n is the total number of instances. The value of χ^2 is used to determine the adequacy of the value by a significant interval. Thus, χ^2 approaches zero when the observed value is similar to the expected value. Therefore, if the observed values are similar to the mean, χ^2 indicates that there is no relationship between the two analyzed variables [18].

The second step of the research was building models that included all attributes and attributes extracted using feature selection. Then we applied the classification to check if feature selection has brought the expected results. The methods chosen for classification are those which allow to build the decision rules in a simple way, i.e. [20–24]:

1. JRip – is an implementation of the RIPPER (Repeated Incremental Running to Produce Error Reduction) algorithm. This scheme allows an efficient induction of decision rules using error reduction mechanisms and heuristics-based rule set optimization. JRip works in two stages. It first induces an initial set of rules, and then refines it in a relatively complex optimization process, improving the individual rules so that when used together they give the best possible results.

2. PART – generates rules from partial decision trees and is a combination of divide-and-conquer and separate-and-conquer strategies. This approach bypasses data optimization throughout the process, making the algorithm simple and transparent. Moreover, it is efficient as it does not require post-processing of data. A decision subtree is a regular decision tree that is constructed and pruned until a stable subtree is found that cannot be simplified any further. As soon as a partial tree is found, a rule is created and the tree is discarded. Thus, there is no generalization of rules and no over-extension of subtrees, which is the case with naive rule building.

3. J48 (C4.5) – is a custom implementation of Quinlan's C4.5 algorithm for Weka. It is a successor to ID3, with added features such as attribute discretization, missing value handling, descriptor weighting, and decision tree pruning. For the purpose of homogeneity evaluation, this algorithm uses the entropy value. The entropy value of a homogeneous sample is zero, while the entropy of a sample containing an equal number of examples from different classes is one. The difference between the entropy value before and after partitioning the data set according to the selected attribute is called the information gain. The descriptors with the highest value of this coefficient and the greatest decrease in entropy are selected for the next nodes. A branch with entropy value equal to zero is considered a leaf.

4. Random Tree – is a set of classification trees with binary splits. For a given observation, expressed as an input vector, each tree returns a decision or a tuple of classification probabilities. The decisions from the trees comprising the forest are treated as votes, and the decision with the highest average probability is returned as the result. Votes may or may not be weighted depending on the voting method selected by the user.

5. CART – is an algorithm for constructing classification and regression trees used to build predictive and descriptive models. Classification trees are used when the dependent variable (decision class) is expressed on a nominal or ordinal scale. Regression trees, on the other hand, are used when there is (at least) an interval level of measurement of the dependent variable (continuous values for the decision class). The construction of a predictive model aims at qualitative or quantitative prediction, while the construction of a descriptor model seeks to describe and present patterns in the community under study. The trees constructed by the CART algorithm are binary decision trees built according to the Gini or binary partitioning criterion. The constructed decision trees are subject to pruning based on the cost of complexity and admit both attributes with continuous and discrete values. The target variable, or decision class-value in the leaf of the tree, can have continuous values, i.e. that

belong to the range of real numbers. The CART algorithm constructs a so-called regression tree in such a case. The data used for tree learning and classification may have missing attribute values.

We used a 10-fold cross-validation for testing, training and validation. Cross-validation is a statistical method that divides a statistical sample into k subsets, then performs an analysis on some of them and uses the others to confirm the reliability of the results. The k times this process is repeated, each time the test set is changed. In this way, within one repetition, $k - 1$ sets are analysed. The analysis is performed k times, and averaging k results allows one final result [25].

To evaluate the above classifiers, we have chosen ACC (Total Accuracy) measure, which is the accuracy that specifies the probability of correct classification, that is, the ratio of correct classifications to all classifications. [24,26,27]:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \tag{6}$$

where: TP (True Positive) is a correctly classified tuple belonging to the class of interest, TN (True Negative) is a correctly classified tuple belonging to another class, FP (False Positive) is a misclassified tuple actually not belonging to the class of interest and FN (False Negative) is a misclassified tuple that actually belongs to the class of interest.

We also collated such measures as kappa statistic, mean absolute error, root mean squared error, relative absolute error, as well as root relative squared error [24,26,27]:

- The Kappa statistic is a reliability coefficient used in statistics for measuring the same variable twice, which is the nominal and dependent variable. The Kappa statistic takes values from -1 to 1. The closer the value is to 1, the more the classifier agrees with the true class. Conversely, the closer the value is to 0, the more divergent the grades are. Values between -1 and 0 are not used in practice, since a negative value of the Kappa statistic would mean that the grades agree less than if they were completely random variables.
- The mean absolute error shows how much, on average, over the prediction period the actual realizations of the predicted variable will deviate in absolute value from the predictions.
- The root mean squared error measures how much, on average, the realizations of the predicted variable deviate from the calculated predictions. A significant difference in value between mean absolute error and root mean squared error indicates the occurrence in the period of the forecast errors with very large values.
- The relative absolute error is the relative absolute difference between the expected and actual values (relative because the mean difference is divided by the arithmetic mean).
- The root relative squared error is relative to what it would be if a simple predictor was used. More precisely, this simple predictor is simply the average of the actual values. So, the relative square error takes the total square error and normalizes it by dividing by the total square error of the simple predictor. Taking the square root of the relative square error reduces the error to the same dimensions as the predictor.

For all analyses and calculations was used WEKA software (version 3.8.4., Machine Learning Group, University of Waikato, New Zealand). WEKA is a machine learning and data mining software developed by researchers at the University of Waikato in Hamilton, New Zealand [27,28].

The final steps of this study were construction of classification rules and extraction of action rules from classification rules. Classification rules are of interest to most data explorers to summarize the discriminant power of classes found in data. A classification rule is a statement that distinguishes the concepts of one class from other classes. They describe the relationship between conditional attributes

(usually many) and the decision attribute by means of implications: on the left side there are conditions expressed by a certain logical formula, on the right – the value of the decision attribute. Rule-based classification systems typically contain many rules, each of which may be based on a different set of conditional items [29–35].

More advanced are action rules, which provide guidance on possible actions the user should take to achieve the desired result. Action rules were first proposed by Raś and Wieczorkowska [36] and they are constructed from classification rules which suggest ways to re-classify objects, for example reclassify patients to a desired group. An action rule can be presented in the following form [32,34,37]:

$$[(\omega) \wedge (\alpha \to \beta)] \Rightarrow (\Psi \to \Omega) \tag{7}$$

where ω indicates a fixed condition features conjunction, that is a part of both groups, $(\alpha \to \beta)$ is recommended changes in flexible features value and $(\Psi \to \Omega)$ means an effect of the action, which the user wants to achieve. All action rules can be evaluated because of its support and confidence [24,26].

3. Results and discussion

The results of feature selection are shown in Table 2. The top 20 attributes were selected for each feature selection method. Then the obtained attributes were compared and only those that were repeated in each method or in at least two methods were selected for further study. The remaining attributes were removed. It is worth noting that as many as 17 attributes out of 20 were extracted in each of the selected feature selection methods. The selected attributes were: d5400-4, d5402-4, d5102-4, d5101-4, d5401-4, d5403-4, d5203-4, d5204-4, d5201-4, d5202-4, d5501-4, d53001-4, d53011-4, d5205-4, d5602-4, d5100-4, d5403-2, d53011-2 and d53001-2.

Table 3 contains classification results for the model where all available attributes are considered (205 + 1 classifier). We can observe, that the best classifier for this model is the PART algorithm. Almost 81.5% of instances were correctly classified here. Equally strong classifiers (80% of correctly classified instances) are J48 and CART algorithms, as well as Random Tree (over 77% of correctly classified instances). The weakest classifier was the JRip algorithm, which achieved an ACC score of about 71.5%.

Table 4 contains the quality parameters for the model containing the attributes extracted during feature selection (19 attributes + 1 classifier). Each of the selected classifiers achieved over 70% accuracy. The best algorithm, with over 77% correctly classified instances, was CART. Almost 67% of correctly classified instances were obtained using the JRIP algorithm, while the PART and J48 algorithms allowed to correctly classify more than 74% of instances. The weakest classifier was Random Tree with 70% ACC.

When comparing the results of the two models (Table 5) we can observe, that for most of the selected classification algorithms, the results for the model after feature reduction are slightly worse than for the model containing all available attributes. Better results are obtained only for the JRip algorithm. However, we cannot say here that feature selection does not make sense. The percentage differences in ACC here range from about 3 to about 7 percentage points. Considering the fact that we reduced 205 attributes to 19, we can assume this was a promising result. Attribute reduction may be helpful for diagnosticians and physiotherapists for faster diagnosis and classification of self-care problems in children with disabilities.

As mentioned earlier, the selected classifier algorithms allowed us to extract classification rules. Classification rules were compared for the model containing all features and the model after feature extraction. In most cases, the rules overlapped, demonstrating the benefits of feature selection. Since about 100 classification rules have been extracted, in this paper we will only present some of the rules for

the model after feature reduction. These rules are written below, while the underlined rules were also present in the model before feature selection:

– IF d5101-4 = false AND d53011-2 = false AND d5205-4 = false AND d53001-4 = false AND d5204-4 = true THEN class = Group 1

– IF d5101-4 = false AND d53011-2 = true AND d5403-2 = true AND d5203-4 = true THEN class = Group 1

Table 2
Results of feature selection

Correlation Based Feature Selection	Information Gain Based Feature Selection	Chi-Square Test for Feature Selection
attribute	attribute	attribute
d5400-4 ♠	d5402-4 ♣	d5402-4 ♣
d5402-4 ♣	d5400-4 ♠	d5400-4 ♠
d5102-4 ♥	d5101-4 ♦	d5101-4 ♦
d5101-4 ♦	d5102-4 ♥	d5102-4 ♥
d5401-4 ♪	d5204-4 ▼	d5204-4 ▼
d5403-4 ▲	d5203-4 ►	d5203-4 ►
d5203-4 ►	d53001-4 ●	d53011-4 ▫
d5204-4 ▼	d53011-4 ▫	d53001-4 ●
d5201-4 ◄	d5401-4 ♪	d5403-4 ▲
d5202-4 ◊	d5403-4 ▲	d5401-4 ♪
d5501-4 ○	d5202-4 ◊	d5202-4 ◊
d53001-4 ●	d5201-4 ◄	d5201-4 ◄
d53011-4 ▫	d5501-4 ○	d5501-4 ○
d5205-4 ◙	d5205-4 ◙	d5205-4 ◙
d5602-4 ◦	d5602-4 ◦	~~d53010-4~~
d5100-4 □	d5100-4 □	~~d53000-4~~
~~d5200-4~~	d5403-2 △	d5602-4 ◦
d5403-2 △	d53011-2	d5100-4 □
d53011-2	d53001-2	d5403-2 △
d53001-2	~~d5602-2~~	~~d5400-3~~

The same attributes obtained in each method are marked with the same symbols. Attributes that repeat in only two methods are highlighted. Non-recurring attributes were crossed out. Attributes: d5400-4 – putting on clothes with complete impairment, d5402-4 – putting on footwear with complete impairment, d5102-4 – drying oneself with complete impairment, d5101-4 – washing whole body with complete impairment, d5401-4 – taking off clothes with complete impairment, d5403-4 – taking off footwear with complete impairment, d5203-4 – caring for fingernails with complete impairment, d5204-4 – caring for toenails with complete impairment, d5201-4 – caring for teeth with complete impairment, d5202-4 – caring for hair with complete impairment, d5501-4 – carrying out eating appropriately with complete impairment, d53001-4 – carrying out urination appropriately with complete impairment, d53011-4 – carrying out defecation appropriately with complete impairment, d5205-4 – caring for nose with complete impairment, d5602-4 – indicating need for drinking with complete impairment, d5100-4 – washing body parts with complete impairment, d5403-2 – taking off footwear with moderate impairment, d53011-2 – carrying out defecation appropriately with moderate impairment, d53001-2 – carrying out urination appropriately with moderate impairment, d5200-4 – caring for skin with complete impairment, d5602-2 – indicating need for drinking with moderate impairment, d53010-4 – indicating need for defecation with complete impairment, d53000-4 – indicating need for urination with complete impairment, d5400-3 – putting on clothes with severe impairment.

Table 3
Quality parameters for a model containing all features

Algorithm/ measure	ACC (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Root relative squared error (%)
JRip	71.4286	0.6059	0.1091	0.2712	51.0009	83.3938
PART	**81.4286**	**0.7404**	**0.0627**	**0.2257**	**29.3097**	**69.3990**
J48 (C4.5)	80.0000	0.7276	0.0687	0.2194	32.1269	67.4608
Random Tree	77.1429	0.6890	0.0653	0.2556	30.5344	78.5934
CART	80.0000	0.7224	0.0866	0.2317	40.4958	71.2542

Table 4
Quality parameters for a model containing the attributes extracted during feature selection

Algorithm/ measure	ACC (%)	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error (%)	Root relative squared error (%)
JRip	75.7143	0.6591	0.1039	0.2441	48.5893	75.0787
PART	74.2857	0.6472	0.0991	0.2411	46.3531	74.1546
J48 (C4.5)	74.2857	0.6480	0.1009	0.2459	47.1763	75.6254
Random Tree	70.0000	0.5998	0.0962	0.2603	44.9869	80.0390
CART	**77.1429**	**0.6850**	**0.0955**	**0.2338**	**44.6643**	**71.9099**

Table 5
Comparison of correctly classified instances in two models

Algorithm/ model	Model 1 (all 205 attributes)	Model 2 (19 attributes extracted during feature selection)	Difference in percentage points
JRip	71.4286 %	75.7143 %	+4.2857 %
PART	81.4286 %	74.2857 %	−7.1429 %
J48 (C4.5)	80.0000 %	74.2857 %	−5.7143 %
Random Tree	77.1429 %	70.0000 %	−7.1429 %
CART	80.0000 %	77.1429 %	−2.8571 %

– IF d5101-4 = false AND d53011-2 = false AND d5205-4 = false AND d53001-4 = true AND THEN class = Group 2
– IF d5400-4 = false AND d53001-4 = true THEN class = Group 2
– IF d5101-4 = false AND d53011-2 = true AND d5403-2 = true AND d5203-4 = false THEN class = Group 4
– IF d53001-4 = false AND d53001-2 = true THEN class = Group 4
– IF d5400-3 = true THEN class = Group 4
– IF d5400-4 = false AND d53001-4 = false AND d53001-2 = true THEN class = Group 4
– IF d5101-4 = true AND d5602-4 = false AND d5202-4 = false THEN class = Group 5
– IF d5101-4 = true AND d5602-4 = false AND d5202-4 = true AND d53001-4 = true AND d5100-4 = true AND d5501-4 = false AND d5205-4 = true THEN class = Group 5
– IF d5101-4 = false AND d53011-2 = false AND d5205-4 = true THEN class = Group 6
– IF d5101-4 = true AND d5602-4 = false AND d5202-4 = true AND d53001-4 = false THEN class = Group 6
– IF d5101-4 = true AND d5602-4 = false AND d5202-4 = true AND d53001-4 = true AND d5100-4 = false THEN class = Group 6
– IF d5101-4 = true AND d5602-4 = false AND d5202-4 = true AND d53001-4 = true AND d5100-4

= true AND d5501-4 = true THEN class = Group 6

- IF d5101-4 = true AND d5602-4 = false AND d5202-4 = true AND d53001-4 = true AND d5100-4 = true AND d5501-4 = false AND d5205-4 = false THEN class = Group 6
- IF d5101-4 = true AND d5602-4 = true THEN class = Group 6
- IF d5400-4 = true THEN class = Group 6
- IF d5101-4 = false AND d53011-2 = false AND d5205-4 = false AND d53001-4 = false AND d5204-4 = false THEN class = Group 7
- IF d5101-4 = false AND d53011-2 = true AND d5403-2 = false THEN class = Group 7
- IF d5203-4 = false THEN class = Group 7
- IF d53001-4 = false THEN class = Group 7
- IF d5400-4 = false AND d53001-4 = false AND d53001-2 = false THEN class = Group 7

We can see that no classification rule was found for Group 3 for the model after feature selection. For the model with all features, only one rule was found for Group 3: IF d5501-1 = false AND d5202-4 = false AND d53010-0 = true AND d5205-2 = true AND Gender = male THEN class = Group 3.

The rules for classification into Group 7 are very important for diagnosis and physical therapy. As we can see, for the model after feature selection, these rules apply only to the negation of features associated with total impairment. Usually such impairment cannot be removed completely, it can only be reduced to some extent. Therefore, it is useful here to show the rules obtained for all available features:

- IF d5501-1 = true AND d5202-3 = false AND d5200-2 = false AND d5602-0 = false AND5700-1 = true AND d5200-1 = true THEN class = Group 7
- IF d5501-1 = true AND d5202-3 = false AND d5200-2 = true THEN class = Group 7.

The final stage of the study was to extract action rules that would allow specific patients to be reclassified from one group to another. Below there are some selected action rules extracted from the classification rules – both for all attributes and for the model after feature selection:

- [d5101-4, false] ∧ [d53011-2, false] ∧ [d5205-4, false] ∧ [d53001-4, false] ∧ d [5204-4, true→false] ⇒ [class, Group 1 → Group 7]
- [d5101-4, false] ∧ [d53011-2, true] ∧ [d5403-2, true] ∧ [d5203-4, false→true] ⇒ [class, Group 4 → Group 1]
- [d5400-4, false] ∧ [d53001-4, false] ∧ [d53001-2, true→false] ⇒ [class, Group 4 → Group 7]
- [d5400-4, false] ∧ [d53000-4, false] ∧ [d5200-3, true] ∧ [d53001-2, true→false] ⇒ [class, Group 4 → Group 1]
- [d5400-4, true] ∧ [d571-1, true] ∧ [d5404-0, true→false] ⇒ [class, Group 6 → Group 5]
- [d5501-1, true] ∧ [d5202-3, true] ∧ [d5400-2, false→true] ⇒ [class, Group 5 → Group 1].

These rules are written in the form of a specific scheme. For example, the last one can be interpreted as follows: If carrying out eating appropriately is mild impairment, caring for hair is severe impairment and we change putting on clothes to moderate impairment, then we can reclassify patient from Group 5 with washing oneself, caring for body parts, toileting, and dressing problem to Group 1 with only caring for body parts problem.

4. Conclusion

Children with disabilities face many problems in their daily lives. Undoubtedly, a very important aspect here is the problem of self-care and its diagnosis. It is therefore necessary to support diagnosticians and

physiotherapists in recognizing self-care problems in children. Data mining, which is being used more and more in medicine and related fields, can be helpful here. Appropriate data analysis allows building models and algorithms that can support the recognition of self-care problems in children with disabilities. Decision and action rules presented in this paper may be equally useful. They can direct therapists to those areas of disability where even a minimal reduction would be of great benefit to the disabled child.

Acknowledgments

Conflict of interest

None to report.

References

[1] WHO. Summary World Report On Disability. World Health [Internet]. 2011; 1–24. Available from: https://apps.who.int/iris/rest/bitstreams/66941/retrieve.

[2] Boettcher L, Dammeyer J. Development and Learning of Young Children with Disabilities. Springer; 2016.

[3] Sharma M. Categorization of self care problem for children with disabilities using partial swarm optimization approach. Int J Inf Technol [Internet]. 2020. Available from: doi: 10.1007/s41870-020-00426-8.

[4] Fatemi Bushehri SMM, Zarchi MS. An expert model for self-care problems classification using probabilistic neural network and feature selection approach. Appl Soft Comput J [Internet]. 2019; 82: 105545. Available from: doi: 10.1016/j.asoc.2019.105545.

[5] World Health Organization. International classification of functioning, disability and health: children and youth version: ICF-CY. 2007.

[6] Petersson C, Simeonsson RJ, Enskar K, Huus K. Comparing children's self-report instruments for health-related quality of life using the International Classification of Functioning, Disability and Health for Children and Youth (ICF-CY). Health Qual Life Outcomes. 2013; 11(1).

[7] Zarchi MS, Fatemi Bushehri SMM, Dehghanizadeh M. SCADI: A standard dataset for self-care problems classification of children with physical and motor disability. Int J Med Inform [Internet]. 2018; 114(March): 81–7. Available from: doi: 10.1016/j.ijmedinf.2018.03.003.

[8] Le T, Baik SW. A robust framework for self-care problem identification for children with disability. Symmetry (Basel). 2019; 11(1).

[9] Islam B, Ashafuddula NIM, Mahmud F. A Machine Learning Approach to Detect Self-Care Problems of Children with Physical and Motor Disability. 2018 21st Int Conf Comput Inf Technol ICCIT 2018. 2019; pp. 21–3.

[10] Zarchi MS, Fatemi Bushehri SMM, Dehghanizadeh M. SCADI: A standard dataset for self-care problems classification of children with physical and motor disability. Int J Med Inform. 2018 Jun 1; 114: 81–7.

[11] UCI Machine Learning Depository [Internet]. Available from: https://archiveics.uci.edu/ml/index.php.

[12] Hall MA. Correlation-based Feature Selection for Machine Learning. 1999; (April).

[13] Palma-Mendoza RJ, de-Marcos L, Rodriguez D, Alonso-Betanzos A. Distributed correlation-based feature selection in spark. Inf Sci (Ny) [Internet]. 2019; 496: 287–99. Available from: doi: 10.1016/j.ins.2018.10.052.

[14] Ranjan A, Singh VP, Mishra RB, Thakur AK, Singh AK. Sentence polarity detection using stepwise greedy correlation based feature selection and random forests: An fMRI study. J Neurolinguistics [Internet]. 2021; 59(June 2020): 100985. Available from: doi: 10.1016/j.jneuroling.2021.100985.

[15] Billah Kushal TR, Illindala MS. Correlation-based feature selection for resilience analysis of MVDC shipboard power system. Int J Electr Power Energy Syst [Internet]. 2020; 117(September 2019): 105742. Available from: doi: 10.1016/j.ijepes.2019.105742.

[16] Odhiambo Omuya E, Onyango Okeyo G, Waema Kimwele M. Feature selection for classification using principal component analysis and information gain. Expert Syst Appl [Internet]. 2021; 174(January): 114765. Available from: doi: 10.1016/j.eswa.2021.114765.

[17] Ramesh G, Madhavi K, Dileep Kumar Reddy P, Somasekar J, Tan J. Improving the accuracy of heart attack risk prediction based on information gain feature selection technique. Mater Today Proc [Internet]. 2021. Available from: doi: 10.1016/j.matpr.2020.12.079.

[18] Luna-Romera JM, Martínez-Ballesteros M, García-Gutiérrez J, Riquelme JC. External clustering validity index based on chi-squared statistical test. Inf Sci (Ny). 2019; 487: 1–17.

[19] Sumaiya Thaseen I, Aswani Kumar C. Intrusion detection model using fusion of chi-square feature selection and multi class SVM. J King Saud Univ – Comput Inf Sci [Internet]. 2017; 29(4): 462–72. Available from: doi: 10.1016/j.jksuci.2015.12.004.

[20] Hussain J, Lalmuanawma S. Feature analysis, evaluation and comparisons of classification algorithms based on noisy intrusion dataset. Procedia Comput Sci [Internet]. 2016; 92: 188–98. Available from: doi: 10.1016/j.procs.2016.07.345.

[21] Barraclough PA, Fehringer G, Woodward J. Intelligent cyber-phishing detection for online. Comput Secur [Internet]. 2021; 104: 102123. Available from: doi: 10.1016/j.cose.2020.102123.

[22] Asha Kiranmai S, Jaya Laxmi A. Data mining for classification of power quality problems using WEKA and the effect of attributes on classification accuracy. Prot Control Mod Power Syst. 2018; 3(1).

[23] Kozjek D, Vrabič R, Kralj D, Butala P, Lavrač N. Data mining for fault diagnostics: A case for plastic injection molding. Procedia CIRP [Internet]. 2019; 81: 809–14. Available from: doi: 10.1016/j.procir.2019.03.204.

[24] Aggarwal CC. Data classification: algorithms and applications. Boca Raton: CRC Press, Taylor & Francis Group; 2015.

[25] Zacharski R. Programmers Guide to Data Mining: The Art of the Numerati. 2012; 360.

[26] Stehman SV. Selecting and interpreting measures of thematic classification accuracy. Remote Sens Environ. 1997; 62(1): 77–89.

[27] Witten IH, Frank E, Hall MA. Data Mining: Practical Machine Learning Tools and Techniques. Morgan Kaufmann; 2011.

[28] Frank E, Hall MA, Witten IH. The WEKA workbench. Data Min. 2017; 553–71.

[29] Rahman SMM, Kotwal MRA, Yu X. Mining classification rules via an apriori approach. Proc 2010 13th Int Conf Comput Inf Technol ICCIT 2010. 2010; (Iccit): 388–93.

[30] Dardzinska A. Action Rules Mining [Internet]. Springer-Verlag, Berlin; 2013. Available from: http://www.springer.com/series/7092.

[31] Dardzinska A, Ras ZW. Cooperative Discovery of Interesting Action Rules. FQAS'06: Proceedings of the 7th international conference on Flexible Query Answering Systems. June 2006; pp. 489–497. Available from: doi: 10.1007/11766254_41.

[32] Ras ZW, Dardzinska A. Action Rules Discovery Based on Tree Classifiers and Meta-actions. In: Rauch J, Raś ZW, Berka P, Elomaa T. (eds) Foundations of Intelligent Systems. ISMIS 2009. Lecture Notes in Computer Science, vol 5722. Springer, Berlin, Heidelberg.

[33] Raś ZW, Tsay LS, Dardzinska A. Tree-based algorithms for action rules discovery. Stud Comput Intell. 2009; 165: 153–63.

[34] Ras ZW, Dardzinska A. Action Rules Discovery without Pre-existing Classification Rules. In: Chan CC, Grzymala-Busse JW, Ziarko WP. (eds) Rough Sets and Current Trends in Computing. RSCTC 2008. Lecture Notes in Computer Science, vol 5306. Springer, Berlin, Heidelberg.

[35] Ras ZW, Dardzinska A. Action Rules Discovery, a New Simplified Strategy. In: Esposito F, Raś ZW, Malerba D, Semeraro G. (eds) Foundations of Intelligent Systems. ISMIS 2006. Lecture Notes in Computer Science, vol 4203. Springer, Berlin, Heidelberg.

[36] Ras ZW, Wieczorkowska A. Action-rules: How to increase profit of a company. Lect Notes Comput Sci (Including Subser Lect Notes Artif Intell Lect Notes Bioinformatics). 2000; 1910: 587–92.

[37] Zdrodowska M, Dardzinska A, Chorąży M, Kułakowska A. Data mining techniques as a tool in neurological disorders diagnosis. Acta Mech Autom. 2018; 12(3): 217–220.