

Supervised and Unsupervised Machine Learning for Improved Identification of Intrauterine Growth Restriction Types

Agnieszka Wosiak
Lodz University of Technology
Institute of Information Technology
ul. Wolczanska 215
90-924 Lodz, Poland
Email: agnieszka.wosiak@p.lodz.pl

Agata Zamecznik
Department of Pediatric
Cardiology and Rheumatology
2nd Chair of Pediatrics
Medical University of Lodz, Poland
Email: agazamek@gmail.com

Katarzyna Niewiadomska-Jarosik
Department of Pediatric
Cardiology and Rheumatology
2nd Chair of Pediatrics
Medical University of Lodz, Poland
Email: kasiajarosik@wp.pl

Abstract—This paper concerns automated identification of intrauterine growth restriction (IUGR) types by use of machine learning methods. The research presents a comparison of supervised and unsupervised learning covering single and hybrid classification, as well as clustering. Supervised learning techniques included bagging with Naïve Bayes, k-nearest neighbours (kNN), C4.5 and SMO as base classifiers, random forest as a variant of bagging with a decision tree as a base classifier, boosting with Naïve Bayes, SMO, kNN and C4.5 as base classifiers, and voting by all single classifiers using majority as a combination rule, as well as five single classification strategies: kNN, C4.5, Naïve Bayes, random tree and sequential minimal optimization algorithm for training support vector machines. Unsupervised learning encompassed k-means and expectation-maximization algorithms. The major conclusion drawn from the study was that hybrid classifiers have demonstrated their potential ability to identify more accurately symmetrical and asymmetrical types of IUGR, whereas the unsupervised learning techniques produced the worst results.

I. INTRODUCTION

IN MEDICINE there are many diseases and diagnoses where identification of their subtypes affects medical treatment. Many research papers concern cancer diagnosis, appropriate feature selection techniques [1]–[3], and its classification based on gene expression [4]. A big challenge is an accurate classification of medical imaging and sound recordings (see [5] and [6]). Moreover, in many cases classification process is performed on labelled and unlabelled data [7]. It may take place in situations where a medical expert diagnosis is imprecisely outlined, as described in this paper.

Intrauterine growth restriction (IUGR) is a fetal growth disorder which is associated with fetal hypoxia and increased perinatal mortality. IUGR may cause a significant risk factor for the development of many cardiovascular, metabolic, and pulmonologic diseases in adult life ([8]–[10]). It is a challenging problem for obstetrician, neonatologists and pediatricians, as the diagnosis is based on non-consistent definitions (see [11] and [12]). It occurs in about 3-10% of live-born newborns, and the most serious problem of IUGR exists in developing countries where it concerns up to 20-30% of newborn infants

[13]. The comparisons of absolute measurements of the fetuses with reference values, as well as birth weight percentiles, allow detection of deviations between expected and actual fetal growth and identification of newborns being possibly at risk for adverse health events [14].

Two types of IUGR can be distinguished: symmetrically impaired and asymmetrically impaired. Foetuses of the first type tend to have a decrease in all dimensions of the body and internal organs, and usually face a higher risk of reduction in growth potential. The problems occur in the first or second trimester of pregnancy and are often encountered in foetuses with infection or genetic and anatomic defects [15]. The second type - asymmetrical - constitutes 75-80% of all cases born as IUGR. It develops in the late second and third trimester of pregnancy and is a consequence of abnormal cell growth, rather than their quantity. In this type, infants have a low birth weight while body length and head circumference remain normal [16]. As asymmetric IUGR infants are more likely to have major anomalies than symmetric IUGR infants or non-IUGR infants [17], there is a need to distinguish between those two patterns of IUGR. Moreover symmetric and asymmetric growth restriction may have different influence on growth and development in preterms from birth to 4 years [18].

To discover risk factors and any parameters that impact IUGR, or to state the dependencies IUGR impacts on, it is necessary:

- to distinguish IUGR from normal fetuses,
- to identify the symmetrical or asymmetrical type of IUGR.

The problem of separating IUGR from normal fetuses has been the subject of analysis for researchers in the field of medicine, as well as computer science, including machine learning and artificial intelligence.

The authors of [19] used multiparametric classifier based on k-mean cluster analysis to separate pathological and normal fetuses. The identification of the intrauterine growth-restricted fetuses was performed on the basis of fetal heart rate variabil-

ity analysis in the antepartum period. The results attained up to 82.4% of accuracy.

In [20] an artificial neural network (ANN) classifier was developed to identify normal and abnormal fetuses based on features from ultrasound images. The accuracy of the classification equalled over 90%. Two ANN models, Multilayer Perceptron using Back propagation algorithm and Radial Basis Function, were also studied and used for IUGR identification in [12].

Lunghi et al. in [21] applied support vector machine algorithm for normal and pathological (IUGR) fetuses classification, based on the analysis of fetal heart rate recordings. The correct classification rate was high enough, above 84%. However, as a concluding remark for future work, the authors suggested using combined classifiers for better discrimination results.

In [22] statistical analysis (contingency tables, analyses of variance, and multiple regression) was applied to identify the problem of placental lesions associated with normal and abnormal fetal growth in infants delivered for obstetric indications at less than 32 weeks' gestation.

Although there are many studies that concern IUGR problem and its identification, few such studies explore different classification techniques. Therefore, automated or semi-automated identification of IUGR patterns is still an open topic.

The aim of this paper is to identify an appropriate classification technique as applied to the problem of intrauterine growth restriction types. Even though classification methods have been studied extensively over the past few years ([23]), no exact solution has been discovered. Moreover, the authors usually focus on one group of machine learning techniques: supervised or unsupervised, without comparisons between the groups. This research not only constitutes an independent contribution to the relevant literature, but also attempts to find a successful way to perform accurate classification of IUGR type.

The rest of the paper is organized as follows. Section II corresponds to the medical data used in this research and is followed by the description of methods used in the experimental part of the paper. Section III is dedicated to the experiments conducted on sample data and the results. Finally, in Section IV, the concluding remarks are discussed.

II. MATERIALS AND METHODS

The proposed methodology of indicating the best machine learning method to use in IUGR types identification consists of three steps:

- applying supervised learning by single classification methods,
- performing multiple classification,
- carrying out clustering as an example of unsupervised learning,
- comparing results of classification techniques by methods of statistical analysis.

TABLE I: Characteristics of the groups

Parameter	IUGR-1	IUGR-2	p-value
	Avg \pm SD (*)	Avg \pm SD (*)	
Birth weight(g)	2556.91 \pm 145.52	2516.77 \pm 301.68	<0.001
Birth length(cm)	52.68 \pm 1.51	50.06 \pm 2.40	<0.001
Head circ.(cm)	33.37 \pm 0.96	32.24 \pm 1.13	<0.001

(*) described as average values \pm standard deviations

A. Data Description

The research was based on a group of 68 children aged 5-10 years (average 7.4 ± 1.36) born on term with IUGR and birth weight below 10 percentile according to gestational age for the Polish population [24]. It consisted of 35 girls and 33 boys. All patients were selected during prospective studies at the Pediatric Cardiology and Rheumatology Department of Medical University of Lodz in 2010-2013. The study was approved by Medical Ethical Committee of the Health Sciences Faculty of Lodz University (No: RNN/760/10/KB).

Two subgroups were distinguished according to the type of hypotrophy:

- IUGR-1 – asymmetrically impaired based on birth weight and an appropriate remainder of the parameters (body length and head circumference above 10 percentile),
- IUGR-2 – symmetrically impaired, where all parameters to be considered (birth weight, body length and head circumference) were below 10 percentile.

Both subgroups were equinumerous - consisted of 34 cases. The IUGR-1 group was constituted by 15 boys and 19 girls, whereas IUGR-2 included 18 boys and 16 girls.

The characteristics of all parameters subjected to further analysis differed significantly between IUGR-1 and IUGR-2 (see Table I).

B. Supervised Learning by Single Classification Method

Classification is the form of supervised learning, which means assigning objects into pre-defined sets of categories or classes. The main purpose of classification is to identify which set of categories a new observation belongs to. This is performed on the basis of a training set consisting of instances that are already labelled the known classes.

A classifier is a mapping function that can be defined by (1):

$$A^i \rightarrow C \quad (1)$$

where:

- $a_1, \dots, a_i \in A$ – are i features that characterize a set of n input instances x_1, \dots, x_n
- $y_j \in C = c_1, \dots, c_m$ – are desired class labels.

C. Multiple Classification

Multiple classification combines individual classifiers in order to obtain a classifier that outperforms every single one.

There are two main questions that should be considered while performing multiple classification:

- the types of classifiers, that should be chosen,
- the way classifiers are combined to obtain a single classification result.

In the literature, there are two terms that refer to multiple classification: "ensemble methods" and "hybrid classifiers". The first one usually refers to collections of models that are minor variants of the same basic model, whereas hybridization allows combining classifiers from different families.

Regarding to combination rules for classifiers, in practice plurality voting is usually implemented (besides unanimity and simple majority) [25], [26]. It takes the result with the higher number of single classifiers' votes, which can be written as (2):

$$class(x) = \underset{c_i \in dom(y)}{arg\ max} \left(\sum_k g(y_k(x), c_i) \right) \quad (2)$$

where:

x – is an instance to be classified,
 $dom(y) = \{c_1, c_2, \dots, c_k\}$ – constitutes the set of labels,
 $y_k(x)$ – is the classification of the k^{th} classifier,
 $g(y, c)$ – is an indicator function defined as:

$$g(y, c) = \begin{cases} 1 & y = c \\ 0 & y \neq c \end{cases}$$

Bagging and boosting are techniques that improve the accuracy of a classifier by generating a composite model that combines multiple classifiers derived from the same inducer.

The term bagging was introduced by Breiman in [27] as an acronym for Bootstrap AGGREGatING. The idea of bagging is to create an ensemble classifiers based on bootstrap replicates of the training set. The classifier outputs are combined by the plurality vote [28].

A variant of bagging is a random forest [29]. It is a general class of ensemble building methods using a decision tree as the base classifier.

Boosting improves the performance of a weak learner as the method iteratively invokes a classifier on training data that is taken from various distributions. The classifiers are generated by resampling the training set and then combined into a single strong composite classifier. Boosting was based on an on-line learning algorithm called Hedge(β) [30]. This approach allocates weights to a set of strategies used to predict the outcome of a certain problem. The distribution is updated after each new outcome and strategies with the correct prediction receive higher weights while the impacts of the strategies with incorrect predictions are reduced.

One of the most popular ensemble algorithm that improves the simple boosting algorithm by an iterative process is AdaBoost (Adaptive Boosting). It was first introduced in [30]. The basic AdaBoost algorithm deals with binary classification. The classification of a new instance is performed according to (3):

$$class(x) = \underset{y \in dom(y)}{arg\ max} \left(\sum_{t: M_t(x)=y} \log \frac{1}{\beta_t} \right) \quad (3)$$

where:

x – is an instance to be classified,
 $dom(y) = \{c_1, c_2, \dots, c_k\}$ – constitutes the set of labels,
 M_t – is a base classifier,
 β_t – is defined as: $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$,
 ϵ_t – is defined as: $\epsilon_t = \sum_{i: M_t(x_i) \neq y_i} D_t(i)$,
 D_t – is a distribution defined as:
 $D_1(i) = 1/m; i = 1, \dots, m$
(m is a size of a training set)
 $D_{t+1}(i) = D_t(i) \cdot \begin{cases} \beta_t & M_t(x_i) = y_i \\ 1 & \text{Otherwise} \end{cases}$

Bagging and boosting use votes to combine the outputs of different classifiers. However in boosting, each classifier is influenced by the performance of predecessors, which means that the new classifier pays more attention to classification errors that were done by the previously built classifiers. Besides in boosting, instances are chosen with a probability that is proportional to their weight, whereas in bagging, each instance is chosen with equal probability.

Hybrid classifiers [25], [26], [31], [32] (also named multiple classifier systems) are designed to increase the accuracy of a single classifier by training several different classifiers and combining their decisions to output a single class label. The hybridization exploits the strength of each component [33] and it prevents the need to try each classifier and simplifies the entire process [31].

For hybrid approach, the diversity is supposed to provide improved accuracy and classifier performance [34]. Therefore most works try to obtain maximum diversity by different means: introducing classifier heterogeneity, bootstrapping the training data, randomizing feature selection, randomizing subspace projections or boosting the data weights. Nevertheless, the diversity hypothesis has not been fully proven [34].

D. Unsupervised Learning with Clustering

Cluster analysis groups objects taking into account a certain similarity metric. The algorithms divide all objects into a predetermined number of groups in a manner that maximizes a similarity function. There are two different approaches, that are commonly used in medical studies ([35] and [36]): the Expectation Maximization (EM) probabilistic method and deterministic k-means algorithm.

An expectation-maximization (EM) algorithm performs repeatedly 2 steps: an expectation (E) and a maximization (M). The first step (E) results in an expectation of the likelihood for observed variables, whereas the second step - maximization (M) computes the maximum expected likelihood found during the E step. EM generates a probability distribution to each instance which indicates the likelihood of its belonging to each cluster [37]. The number of clusters can be designated by cross validation. It is worth emphasizing, that the EM algorithm computes classification probabilities, not exact assignments of observations to clusters.

The k-means algorithm divides a data set into k clusters, where k is a user-defined value. The algorithm starts with k random clusters, and next moves objects between those groups to minimize variability within each of them and maximize variability between clusters. Usually, the means for each cluster on every dimension are calculated to assign objects into the closest group [39]. In most of the cases Euclidean metric is considered as the distance function for k-means algorithm [37], [40].

E. Statistical Analysis

Statistical analysis is a required part of any research investigations, including proposing new methods or comparing existing ones in any field of science. Many researchers in machine learning confirmed the need for statistical validation of results.

In cases where comparison of two classifiers is performed, the McNemar test and 5x2 cross validation were recommended [41]. The situation where many classifiers are verified, is more complex from statistical point of view. Although many research papers draw conclusions based on matrix of tests comparing all pairs of classifiers (e.g. a matrix of the McNemar tests), an appropriate test for multiple comparisons should be used. The Friedman test with the corresponding post-hoc analysis was proved to be suitable for comparison of many classifiers [38], [45].

The Friedman test was firstly introduced in [42], [43] for non-parametric measures. The goal of the test is to determine - basing on samples - that there is a difference among classification results. The original results are changed into ranks starting from the best one and the null hypothesis states that all algorithms give same results and their ranks are identical. The Friedman statistics is computed as follows (4):

$$\chi_F^2 = \frac{12n}{k(k+1)} \left[\sum_j R_j^2 - \frac{k(k+1)^2}{4} \right] \quad (4)$$

where:

- n – is the number of datasets being considered
- k – is the number of algorithms
- R_j – is the average rank of j th algorithm

If the null hypothesis is rejected, a post-hoc analysis should be performed to compare classifiers with each other and find statistically significant differences. The Nemenyi test can be applied. It states that two classifiers differ significantly if their ranks vary at least by the critical difference (5):

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6n}} \quad (5)$$

where:

- q_α – are critical values based on the Studentized range statistic divided by $\sqrt{2}$.

TABLE II: Single classification results

Method	ACC [%]	PREC	SENS	AUROC
kNN	75.00	0.751	0.750	0.754
C4.5	76.47	0.765	0.765	0.697
Logistic	80.88	0.809	0.809	0.874
DTable	76.47	0.782	0.765	0.779
NaiveBayes	77.94	0.785	0.779	0.846
RandomTree	73.53	0.742	0.735	0.732
SGD	79.41	0.794	0.794	0.794
SMO	73.91	0.775	0.739	0.746

III. RESULTS AND DISCUSSION

The purpose of experiments was to find the best method for IUGR type identification by examining the accuracy of different classification approaches, including single and hybrid classifiers, as well as clustering techniques.

The experiments were conducted on a real dataset consisted of 68 cases. Each case was described by 3 numerical attributes: birth weight, body length and head circumference according to the description presented in Section II-A.

The aim of the classification was to distinguish automatically symmetrical or asymmetrical type of IUGR. Therefore the set of labels consisted of two classes.

All experiments were based on WEKA Open Source Data Mining Tool [44].

In order to assess the performance of various classification methods, following comparison criteria have been used: accuracy (ACC), precision (PREC), sensitivity (SENS) and the area under ROC curve (AUROC). To verify experimental results, a detailed statistical evaluation was performed with use of Friedman test and post-hoc analysis.

A. Single Classification

In the first step of the experiments, single classification algorithms were applied. Eight approaches were considered: k-nearest neighbours (kNN), C4.5, logistic regression (Logistic), decision table, Naïve Bayes (NaiveBayes), random tree (a tree that considers K randomly chosen attributes at each node), stochastic gradient descent (SGD) and sequential minimal optimization algorithm for training support vector machines (SMO). The results of classification are presented in Table II.

The best single classification results attained 80% for logistic regression and 79% for SGD algorithm in terms of classification accuracy. Moreover, logistic regression and NaiveBayes gave the best results of AUROC (0.874 and 0.846 respectively) The average accuracy of single classification approach equalled 76.7%.

B. Multiple Classification

Next step of the experiments concerned performing classification using hybrid classifiers. Different combinations were applied:

- bagging with C4.5 and SMO as base classifiers,
- random forest as a variant of bagging with a decision tree (DTree) as a base classifier,

TABLE III: Hybrid classification results

Method	Base	ACC [%]	PREC	SENS	AUROC
Bagging	C4.5	79.41	0.799	0.794	0.852
Bagging	SMO	77.94	0.782	0.779	0.833
RandomForest	DTree	76.47	0.765	0.765	0.843
AdaBoost	DTable	80.88	0.824	0.809	0.835
AdaBoost	C4.5	80.88	0.809	0.809	0.826
AdaBoost	SGD	82.35	0.824	0.824	0.876
AdaBoost	SMO	77.94	0.780	0.799	0.806
Hybrid	all single	80.88	0.812	0.809	0.810

TABLE IV: Results of clustering

Method	No of cases in clusters	ACC [%]	PREC	SENS	AUROC
k-means	28 / 40	64.71	0.625	0.735	0.647
EM	63 / 5	57.35	1.000	0.147	0.574

- boosting with decision table (DTable), SMO, C4.5 and SGD as base classifiers, and
- hybridization by use of all single classifiers with majority voting as a combination rule.

The results of multiple classifications are shown in Table III. The best hybrid classification accuracy attained 82.35% for AdaBoost algorithm with SGD as a base classifier, whereas the worse one equalled 76.45% for RandomForest method. The average accuracy for all hybrid classification methods achieved 79.59%.

C. Clustering

The last step referred to clustering techniques. According to the methodology, two different approaches were considered: k-means algorithm and Expectation-Maximization method. Using EM algorithm we firstly used 10 fold cross-validation [37] to obtain clusters automatically, however it resulted in one cluster only. Therefore, for both techniques we defined 2 groups: for symmetrical and asymmetrical IUGR cases. The results of clustering are shown in Table IV.

One can notice that in the case of IUGR dataset, unsupervised techniques did not meet the expectations. Both algorithms resulted in accuracies below 65%, which is not satisfactory enough to implement this approach in practice.

D. Classification Comparison and Statistical Analysis

To compare the classifiers, the Friedman test and the corresponding post-hoc analysis were performed. The final results of absolute differences between average ranks for classifiers are presented in Table V where significant values are in bold, italic and underlined.

The results of the post-hoc tests can be clearly visualized with the diagram [38]. Figure 1 shows the results of the analysis of the data from Table V. The diagram compares all the algorithms against each other. The top line of the diagram is the axis on which we plot the average ranks of each method. Each number represents subsequent classification method sorted by the values of classification accuracy in the descending order, i.e. the lowest and best ranks are to the

TABLE V: Average ranks of post-hoc analysis

	kNN	C4.5	Log	DT	NB	RT	SGD	SMO	Bagg	C4.5
kNN	0	2	<u>10.5</u>	2	5	2	7.5	1	7.5	
C4.5	2	0	8.5	0	3	4	5.5	3	5.5	
Logistic	<u>10.5</u>	8.5	0	8.5	5.5	<u>12.5</u>	3	<u>11.5</u>	3	
DT	2	0	8.5	0	3	4	5.5	3	5.5	
NB	5	3	5.5	3	0	7	2.5	6	2.5	
RT	2	4	<u>12.5</u>	4	7	0	<u>9.5</u>	1	<u>9.5</u>	
SGD	7.5	5.5	3	5.5	2.5	<u>9.5</u>	0	8.5	0	
SMO	1	3	<u>11.5</u>	3	6	1	8.5	0	8.5	
Bag. C4.5	7.5	5.5	3	5.5	2.5	<u>9.5</u>	0	8.5	0	
Bag. SMO	5	3	5.5	3	0	7	2.5	6	2.5	
RF	2	0	8.5	0	3	4	5.5	3	5.5	
Boost DT	<u>10.5</u>	8.5	0	8.5	5.5	<u>12.5</u>	3	<u>11.5</u>	3	
Boost C4.5	<u>10.5</u>	8.5	0	8.5	5.5	<u>12.5</u>	3	<u>11.5</u>	3	
Boost SGD	<u>13</u>	<u>11</u>	2.5	<u>11</u>	8	<u>15</u>	5.5	<u>14</u>	5.5	
Boost SMO	5	3	5.5	3	0	7	2.5	6	2.5	
Hybrid	<u>10.5</u>	8.5	0	8.5	5.5	<u>12.5</u>	3	<u>11.5</u>	3	
kmeans	3	5	<u>13.5</u>	5	8	1	<u>10.5</u>	2	<u>10.5</u>	
EM	4	6	<u>14.5</u>	6	9	2	<u>11.5</u>	3	<u>11.5</u>	

	Bagg	RF	Boost	Boost	Boost	Boost	Hyb	km	EM
	SMO	DT	C4.5	SGD	SMO				
kNN	5	2	<u>10.5</u>	<u>10.5</u>	<u>13</u>	5	<u>10.5</u>	3	4
C4.5	3	0	8.5	8.5	<u>11</u>	3	8.5	5	6
Logistic	5.5	8.5	0	0	2.5	5.5	0	<u>13.5</u>	<u>14.5</u>
DT	3	0	8.5	8.5	<u>11</u>	3	8.5	5	6
NB	0	3	5.5	5.5	8	0	5.5	8	9
RT	7	4	<u>12.5</u>	<u>12.5</u>	<u>15</u>	7	<u>12.5</u>	1	2
SGD	2.5	5.5	3	3	5.5	2.5	3	<u>10.5</u>	<u>11.5</u>
SMO	6	3	<u>11.5</u>	<u>11.5</u>	<u>14</u>	6	<u>11.5</u>	2	3
Bag. C4.5	2.5	5.5	3	3	5.5	2.5	3	<u>10.5</u>	<u>11.5</u>
Bag. SMO	0	3	5.5	5.5	8	0	5.5	8	9
RF	3	0	8.5	8.5	<u>11</u>	3	8.5	5	6
Boost DT	5.5	8.5	0	0	2.5	5.5	0	<u>13.5</u>	<u>14.5</u>
Boost C4.5	5.5	8.5	0	0	2.5	5.5	0	<u>13.5</u>	<u>14.5</u>
Boost SGD	8	<u>11</u>	2.5	2.5	0	8	2.5	<u>16</u>	<u>17</u>
Boost SMO	0	3	5.5	5.5	8	0	5.5	8	9
Hybrid	5.5	8.5	0	0	2.5	5.5	0	<u>13.5</u>	<u>14.5</u>
kmeans	8	5	<u>13.5</u>	<u>13.5</u>	<u>16</u>	8	<u>13.5</u>	0	1
EM	9	6	<u>14.5</u>	<u>14.5</u>	<u>17</u>	9	<u>14.5</u>	1	0

right. As a result we start with number 1 for boosted SGD and end with number 18 for EM clustering. The positions of average ranks for each classifier are marked with vertical lines and captioned with their names. Moreover, the groups of algorithms that are not significantly different in terms of accuracy are connected with horizontal lines. Consequently, we can easily notice, that there is no significant difference between boosted SGD and hybrid approach, however both of them achieved statistically better accuracies when compared with, inter alia, NaiveBayes, SMO or kmeans.

To summarize the experimental studies, one can see, that none of the classification techniques significantly outper-

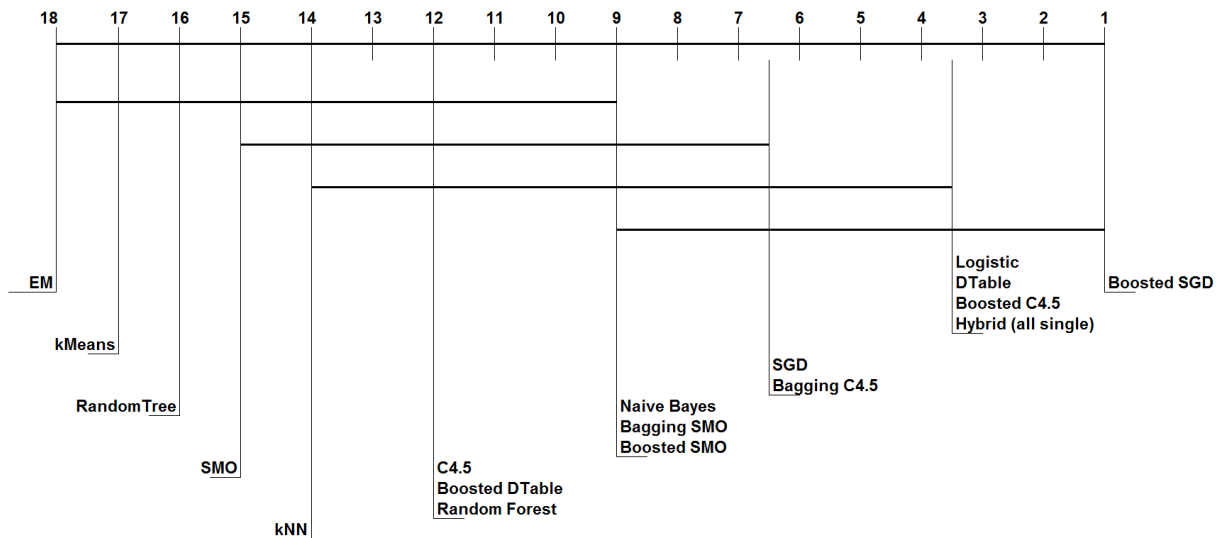


Fig. 1: Visualisation of post-hoc test for comparison of classifiers

formed the rest of them. However, it should be emphasized, that multiple classifications mostly exceeded single classifiers and grouping techniques in terms of classification accuracy. Even SGD - one of the best single classification method - when boosted, improved the accuracy to 82%.

IV. CONCLUSIONS

Classification of medical datasets is regarded as a challenging task, requiring extremely high accuracy. Therefore researches on finding the most appropriate methods for precise classification are conducted. Multiple classifiers constitute one of the most important advances in machine learning in recent years. In the absence of detailed a priori knowledge of the problem, they provide better performance.

The identification process of IUGR pattern (symmetrical or asymmetrical) is an important medical problem to solve, as symmetric and asymmetric growth restriction may have different influence on growth and development in childhood. Moreover asymmetric IUGR infants are more likely to have major anomalies than symmetric IUGR infants or infants appropriate for gestational age.

By comparing hybrid classifiers algorithms, single classification methods and clustering, it was demonstrated that the hybrid strategy resulted in the most satisfactory outcomes and confirmed other up-to-date researches on multiple classifier systems. Clustering, which is supposed to give good results in terms of unlabelled data and situations where label definitions are not precise, did not succeed in the case of IUGR classification.

In order to find the optimal solutions, future studies ought to involve other algorithms and strategies as well. Other combinations of various classifiers should be also investigated in depth. Furthermore, fuzzy logic can be applied to the

problem of IUGR classification, as its results on medical data proved their efficiency [47]–[49].

REFERENCES

- [1] Paja W.: "Medical diagnosis support and accuracy improvement by application of total scoring from feature selection approach", Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS 2015), Annals of Computer Science and Information Systems, eds. M. Ganzha and L. Maciaszek and M. Paprzycki, IEEE, 2015, pp. 281–286, DOI: 10.15439/2015F361
- [2] Gong, P. and Cheng, Y.-H. and Wang, X.-S.: "Benign or Malignant Classification of Lung Nodules Based on Semantic Attributes", Acta Electronica Sinica, 2015, vol. 43, no. 12, pp. 2476–2483
- [3] Pérez, N. and Guevara, M.A. and Silva, A. and Ramos, I. and Loureiro, J.: "Improving the performance of machine learning classifiers for Breast Cancer diagnosis based on feature selection", Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, IEEE, 2014, vol. 2, pp. 209–217, DOI: 10.15439/2014F249
- [4] Hijazi, H. and Chan, Ch.: "A Classification Framework Applied to Cancer Gene Expression Profiles" Journal of Healthcare Engineering, vol. 4, no. 2, pp. 255–283, 2013, DOI: 10.1260/2040-2295.4.2.255
- [5] Sun, S., Wang, H., Jiang, Z., Fang, Y., Tao, T.: "Segmentation-based heart sound feature extraction combined with classifier models for a VSD diagnosis system", Expert Systems with Applications, 41(4), 2014, pp. 1769–1780, DOI: 10.1016/j.eswa.2013.08.076
- [6] Montejo, L. D., Jia, J., Kim, H. K., Netz, U. J., Blaschke, S., Muller, G. A., Hielscher, A. H.: "Computer-aided diagnosis of rheumatoid arthritis with optical tomography, Part 2: image classification", Journal of biomedical optics, 2013, vol. 18(7), pp. 076002–076002, DOI: 10.1117/1.JBO.18.7.076002
- [7] Stamatis, K. and Nikos, F. and Sotiris, K. and Kyriakos S.: "A Semisupervised Cascade Classification Algorithm", Applied Computational Intelligence and Soft Computing, Article ID 5919717, 14 pages, 2016, DOI: 10.1155/2016/5919717
- [8] Baker, D.J.: "Maternal nutrition, fetal nutrition, and disease in later life", Nutrition, 1997, vol.13, pp. 807–813, DOI: 10.1016/S0899-9007(97)00193-7
- [9] Mahajan, S.D. and Singh, S. and Shah, P. and Gupta, N. and Kochupillai, N.: "Effect of maternal malnutrition and anemia on the endocrine regulation of fetal growth", Endocrine research, 2004, vol. 30(2), pp. 189–203, DOI: 10.1081/ERC-200027380

- [10] Mahajan, S.D. and Aalinkeel, R. and Singh, S. and Shah, P. and Gupta, N. and Kochupillai, N.: "Endocrine regulation in asymmetric intrauterine fetal growth retardation", *Journal of Maternal-Fetal and Neonatal Medicine*, 2006, vol. 19(10), pp. 615–623, DOI: 10.1080/14767050600799901
- [11] Gadagkar, A.V. and Shreedhara, K.S.: "Fetal Growth Diagnosis using Re-Initialization Free Level Set Method and Classification using Radial Basis Function Neural Network", *Proceedings of the International Conference on Multimedia Processing, Communication and Information Technology MPCIT 2013*, 2013, pp. 137–144, DOI: 03.AETS.2013.4.81
- [12] Bagi, K.S. and Shreedhara, K.S.: "Biometric measurement and classification of IUGR using neural networks", *Proceedings of the International Conference on Contemporary Computing and Informatics (IC3I 2014)*, 2014, pp. 157–161, DOI: 10.1109/IC3I.2014.7019613
- [13] Black, R.E. and Victora, C.G. and Walker, S.P. and Bhutta, Z.A. and Christian, P. and de Onis, M. and et al.: "Maternal and child undernutrition and overweight in low-income and middle-income countries", *Lancet*, 2013, vol. 382, pp. 427–451, DOI: 10.1016/S0140-6736(13)60937-X
- [14] Gürgen, F. and Zeynep, Z. and Füsün, V.: "Intrauterine growth restriction (IUGR) risk decision based on support vector machines", *Expert Systems with Applications*, 2012, vol.39(3), pp. 2872–2876, DOI: 10.1016/j.eswa.2011.08.147
- [15] Shreedhara, K.S. and Veena, A.: "Multiple sonographic features based IUGR diagnosis using artificial neural networks", *International Journal of Information Technology and Knowledge Management*, 2009, vol.2(1), pp. 73–78, DOI:10.1109/ICSIP.2014.54
- [16] Zamecznik, A. and Niewiadomska-Jarosik, K. and Wosiak, A. and Zamojaska, J. and Moll, J. and Stańczyk, J.: "Intra-uterine growth restriction as a risk factor for hypertension in children six to 10 years old", *Cardiovascular Journal of Africa*, 2014, pp.73–77, DOI: 10.5830/CVJA-2014-009
- [17] Dashe, J.S. and McIntire, D.D. and Lucas, M.J. and Leveno, K.J.: "Effects of symmetric and asymmetric fetal growth on pregnancy outcomes", *Obstetrics & Gynecology*, 2000, vol. 96(3), pp. 321–327
- [18] Bocca-Tjeertes, I. and Bos, A. and Kerstjens, J. and de Winter, A. and Reijnenveld, S.: "Symmetrical and Asymmetrical Growth Restriction in Preterm-Born Children", *Pediatrics*, 2013, vol. 133(3), pp. e650–e656
- [19] Ferrario, M. and Signorini, M.G. and Magenes, G.: "Complexity analysis of the fetal heart rate variability: early identification of severe intrauterine growth-restricted fetuses", *Medical & Biological Engineering & Computing*, 2009, vol.47(9), pp. 911–919, DOI: 10.1007/s11517-009-0502-8
- [20] Gadagkar, A.V. and Shreedhara, K.S.: "Features Based IUGR Diagnosis Using Variational Level Set Method and Classification Using Artificial Neural Networks", *Proceedings of the Fifth International Conference on Signal and Image Processing (ICSIP 2014)*, 2014, pp. 303–309, DOI: 10.1109/ICSIP.2014.54
- [21] Lungchi, F. and Magenes, G. and Pedrinazzi, L. and Signorini, M.G.: "Detection of fetal distress through a support vector machine based on fetal heart rate parameters", *Computers in Cardiology*, 2005, vol. 32, pp. 247–250, DOI: 10.1109/CIC.2005.1588083
- [22] Salafia, C.M. and Miniör, V.K. and Pezzullo, J.C. and Popek, E.J. and Rosenkrantz, T.S. and Vintzileos, A.M.: "Intrauterine growth restriction in infants of less than thirty-two weeks' gestation: associated placental pathologic features", *American Journal of Obstetrics and Gynecology*, 1995, vol. 173(4), pp. 1049–1057, DOI: 10.1016/0002-9378(95)91325-4
- [23] Jeetha Lakshmi, P. S. and Saravan Kumar S. and Suresh A.: "A Novel Hybrid Medical Diagnosis System Based on Genetic Data Adaptation Decision Tree and Clustering", *ARNP Journal of Engineering and Applied Sciences*, vol. 10, no. 16, 2015, pp. 7293–7299
- [24] Malinowski, A. and Chlebna-Sokół, D.: "Dziecko łódzkie-metody badań i normy rozwoju biologicznego", *Ankal*, 1998, (In Polish)
- [25] Kuncheva, L.I.: *Combining Pattern Classifiers. Methods and Algorithms.*, John Wiley & Sons, Inc., 2004, Hoboken, New Jersey, USA
- [26] Woźniak, M. and Graña, M. and Corchado, E.: "A survey of multiple classifier systems as hybrid systems", *Information Fusion*, 2014, pp. 3–17, DOI: 10.1016/j.inffus.2013.04.006
- [27] Breiman, L.: "Bagging predictors", Technical Report 421, Department of Statistics, University of California, Berkeley, 1994
- [28] Breiman, L.: "Bagging predictors", *Machine Learning*, 1996, vol. 26(2), pp. 123–140
- [29] Breiman, L.: "Random forests", *Machine Learning*, 2001vol. 45, pp. 5–32
- [30] Freund, Y. and Schapire, R.E.: "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of Computer and System Sciences*, 1997, vol. 55(1), pp. 119–139, DOI: 10.1006/jcss.1997.1504
- [31] Rokach, L.: *Pattern Classification Using Ensemble Methods*, World Scientific Publishing Co., Inc., 2010, River Edge, New York, USA
- [32] Seni, G. and Elder, J.F.: "Ensemble Methods in Data Mining: Improving Accuracy Through Combining Predictions", Morgan & Claypool, 2010
- [33] Michalski, R.S. and Tecuci, G.: "Machine Learning, A Multistrategy Approach", J. Morgan Kaufmann, 1994
- [34] Wang, S.L. and Li, X.L. and Fang, J.: "Finding minimum gene subsets with heuristic breadth-first search algorithm for robust tumour classification", *BMC Bioinformatics*, 2012, vol. 13(178), pp. 1–26, DOI: 10.1186/1471-2105-13-178
- [35] Pirooznia, M. and Yang, J. and Yang M.Q. and Deng, Y.: "A comparative study of different machine learning methods on microarray gene expression data", *BMC Genomics*, 2008 vol. 9, pp. 1471–2164, DOI: 10.1186/1471-2164-9-s1-s13
- [36] Wosiak, A. and Zakrzewska, D.: "Feature Selection for Classification Incorporating Less Meaningful Attributes in Medical Diagnostics", *Proceedings of the 2014 Federated Conference on Computer Science and Information Systems, Annals of Computer Science and Information Systems ACSIS*, 2014, pp. 235–240, DOI: 10.15439/2014F296
- [37] Witten, I.H. and Frank, E. and Hall, M.A.: "Data Mining: Practical Machine Learning Tools and Techniques" (3rd ed.), Morgan Kaufmann Publishers Inc., 2011
- [38] Demsar, J.: "Statistical Comparisons of Classifiers over Multiple Data Sets", *The Journal of Machine Learning Research*, 2006, vol. 7, pp. 1–30
- [39] MacQueen, J.B.: "Some Methods for Classification and Analysis of Multi-Variate Observations", *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, 1967, pp. 281–297
- [40] Ankita, V. and Satyanarayana, R.V. and Kamalakar, K.: "An Experiment with Distance Measures for Clustering", *Proceedings of the International Conference on Management of Data*, 2008
- [41] Dietterich, T. G.: "Approximate statistical tests for comparing supervised classification learning algorithms", *Neural Computation*, 2006, vol. 10(7), pp. 1895–1923, DOI: 10.1162/089976698300017197
- [42] Friedman, M.: "The use of ranks to avoid the assumption of normality implicit in the analysis of variance", *Journal of the American Statistical Association*, 1937, vol. 32, pp. 675–701, DOI: 10.1080/01621459.1937.10503522
- [43] Friedman, M.: "A comparison of alternative tests of significance for the problem of m rankings", *Annals of Mathematical Statistics*, 1940, vol. 11, pp. 86–92
- [44] Weka Data Mining Tool: <http://www.cs.waikato.ac.nz/ml/weka/index.html>
- [45] Garcia, S. and Fernandez, A. and Luengo, J. and Herrera, F.: "Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power", *Information Sciences*, 2010, vol. 180(10), pp. 2044–2064, DOI: 10.1016/j.ins.2009.12.010
- [46] Nemenyi, P.B.: "Distribution-free multiple comparisons", PhD thesis, Princeton University, 1963
- [47] Widjaja, M. and Darmawan, A. and Mulyono, S.: "Fuzzy classifier of paddy growth stages based on synthetic MODIS data", *Proceedings of the IEEE International Conference on Advanced Computer Science and Information Systems (ICACSIS 2012)*, 2012, pp. 239–244
- [48] Grochowina, M. and Leniowska, L.: "Comparison of SVM and k-NN classifiers in the estimation of the state of the arteriovenous fistula problem", *Proceedings of the 2015 Federated Conference on Computer Science and Information Systems (FEDCSIS 2015)*, *Annals of Computer Science and Information Systems*, eds. M. Ganzha and L. Maciaszek and M. Paprzycki, IEEE, 2015, pp. 249–254, DOI: 10.15439/2015F194
- [49] Lakshmi, PS Jeetha, S. Saravan Kumar, and A. Suresh. "Intelligent Medical Diagnosis System Using Weighted Genetic and New Weighted Fuzzy C-Means Clustering Algorithm." *Artificial Intelligence and Evolutionary Algorithms in Engineering Systems*. Springer India, 2015. pp. 213–220
- [50] Nawarycz, T. and Pytel, K. and Gazicki-Lipman, M. and Drygas, W. and Ostrowska-Nawarycz, L.: "A Fuzzy Logic Approach to The Evaluation of Health Risks Associated with Obesity", *Proceedings of the 2013 Federated Conference on Computer Science and Information Systems*, eds. M. Ganzha, L. Maciaszek, M. Paprzycki, IEEE, 2013, pp. 231–234