

## Rule-Based Classification of MMPI Data of Patients with Mental Disorders: Experiments with Basic and Extended Profiles

Jerzy Gomuła<sup>2,3</sup>, Krzysztof Pancerz<sup>1</sup>, Jarosław Szkoła<sup>1</sup>

<sup>1</sup> *Institute of Biomedical Informatics, University of Information Technology and Management,  
Sucharskiego Str. 2,  
Rzeszów, 35-225, Poland*

*E-mails: kpancerz@wsiz.rzeszow.pl, jszkola@wsiz.rzeszow.pl*

<sup>2</sup> *The Andropause Institute, Medan Foundation,  
Warsaw, Poland*

*E-mail: jerzy.gomula@wp.pl*

<sup>3</sup> *Cardinal Stefan Wyszyński University,  
Warsaw, Poland*

### Abstract

In the research presented in the paper, we test the classification potential for different configurations of the MMPI data used for nosological diagnosis of patients with mental disorders. Originally, each patient is described by an MMPI profile consisting of a set of values of thirteen attributes (the so-called scales). A profile can be extended by different specialized indexes defined in the professional domain literature. Each created index has been proposed by practitioners and has a proper meaning from the clinical point of view. Adding such indexes leads to the extension of the attribute space for cases to be classified. In the paper, we present results of experiments with basic (original) and extended profiles.

*Keywords:* classification, rules, decision support, MMPI data.

### 1. Introduction

We can distinguish two main tasks in building decision support systems in medicine. The first one is to search for efficient and accurate methods of diagnostic classification of new cases (patients). The second one is to extract the useful knowledge from data which has the clinical meaning and interpretation, and that is important for diagnosticians. Our research concerns psychometric data coming from the Minnesota Multiphasic Personality Inventory (MMPI) test<sup>16</sup>. MMPI is one of the most frequently

used personality tests in clinical mental health as well as psychopathology (mental and behavioral disorders). The test builds upon multidimensional and empirical presumptions. It was designed and published first in 1943 in a questionnaire form by a psychologist S.R. McKinley and neuropsychiatrist J.Ch. Hathaway from the University of Minnesota. Later the inventory was adapted in above fifty countries.

The MMPI-WISKAD personality inventory is a Polish adaptation of the American inventory. The test originally was translated by M. Choynowski (as WIO)<sup>5</sup> and elaborated by Z. Płużek (as WISKAD)

in 1950<sup>18</sup>. The American norms were accepted there. The clinical scales in the MMPI-WISKAD are not the scales of clear symptoms, since the questions (items) simultaneously draw, in most cases, upon several scales.

On the basis of the received responses (*Yes, Cannot Say, No*) to selected questions we may make up a total count for the subject being examined, measuring obtained raw results against the reference and clinical scales, as being directly related to specific questions (items). The profile (graph) that is built for such a case always has a fixed and invariable order of its constituents as distributed on the scales:

- the validity part consists of three scales: *L* - the scale of lying which is sensitive to all false statements aiming at representing ourselves in a better light, *F* - the scale which detects atypical and deviatory answers to all items in the test, *K* - the scale which examines self defensive mechanisms and detects subtler attempts of the subject being screened at falsifying and aggravation.
- the clinical part: basic clinical scales have numbers attributed so that a profile can be encoded to avoid negative connotations connected with the names of scales: 1. (*Hp* - Hypochondriasis), 2. (*D* - Depression), 3. (*Hy* - Hysteria), 4. (*Ps* - Psychopathic Deviate), 5. (*Mf* Masculinity/Femininity), 6. (*Pa* - Paranoia), 7. (*Pt* - Psychastenia), 8. (*Sc* - Schizophrenia), 9. (*Ma* - Hypomania), 0. (*It* Social introversion).

In years 1998-1999, a team of researchers consisting of W. Duch, T. Kucharski, J. Gomuła, R. Adamczak created two independent rule systems devised for the nosological diagnosis of patients that may be screened with the MMPI-WISKAD test<sup>9</sup>. In this paper, we continue research carried out by that group. Our research is focused on creating a new computer tool called Copernicus. It is a tool for computer-aided diagnosis of mental disorders based on personality inventories. The first version of this tool has been presented in<sup>11</sup>. The presented results shall be helpful for selection of suitable approaches and algorithms for this tool.

The rest of the paper is organized as follows. Section 2 gives a brief description of data used in experiments (profiles and added specialized indexes).

In section 3, computer tools used in experiments are mentioned. We especially describe our tool - the Copernicus system - created for effective and specialized analysis of psychometric data to support psychologists in diagnostic decision making. Experiments performed by us are depicted in Section 4. Finally, section 5 consists of some conclusions and directions for the further work.

## 2. MMPI Data for Experiments

In case of the MMPI test, each case (patient)  $x$  is described by a data vector  $a(x)$  consisting of thirteen descriptive attributes:  $a(x) = [a_1(x), a_2(x), \dots, a_{13}(x)]$ . If we have training data, then to each case  $x$  we additionally add one decision attribute  $d$  - a class to which a patient is classified.

In our research, we have obtained input data which have classes (nosological types) assigned to patients by specialists. Our data are categorized according to 19 nosological classes plus the reference (norm) class. Each class corresponds to one of psychiatric nosological types: norm (*norm*), neurosis (*neur*), psychopathy (*psych*), organic (*org*), schizophrenia (*schiz*), delusion syndrome (*del.s*), reactive psychosis (*re.psy*), paranoia (*paran*), (sub)manic state (*man.st*), criminality (*crim*), alcoholism (*alcoh*), drug addiction (*drug*), simulation (*simu*), dissimulation (*dissimu*), and six deviational answering styles (*dev1*, *dev2*, *dev3*, *dev4*, *dev5*, *dev6*).

For the training data (which are used to learn or extract relationships between data), we have a tabular form (see example in Table 1) which is formally called a decision system (decision table)  $S = (U, A, d)$  in the Pawlak's form<sup>17</sup>.  $U$  is a set of cases (patients),  $A$  is a set of descriptive attributes corresponding to scales, and  $d$  is a decision attribute determining a nosological type (class, category).

Attributes in  $A$  represent scales in the following way:  $a_1$  corresponds to the scale of lying  $L$ ,  $a_2$  corresponds to the scale of atypical and deviational answers  $F$ ,  $a_3$  corresponds to the scale of self defensive mechanisms  $K$ ,  $a_4$  corresponds to the scale of Hypochondriasis (1.*Hp*),  $a_5$  corresponds to the scale of Depression (2.*D*),  $a_6$  corresponds to the scale of

Table 1: An input data for Copernicus (fragment)

Patient ID	$a_1$	$a_2$	$a_3$	$a_4$	$a_5$	$a_6$	$a_7$	$a_8$	$a_9$	$a_{10}$	$a_{11}$	$a_{12}$	$a_{13}$	class
#1	55	65	50	52	65	57	63	56	61	61	60	51	59	norm
#2	50	73	53	56	73	63	53	61	53	60	69	45	61	org
#3	56	78	55	60	59	54	67	52	77	56	60	68	63	paran
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

Hysteria (3.Hy),  $a_7$  corresponds to the scale of Psychopathic Deviate (4.Ps),  $a_8$  corresponds to the scale of Masculinity/Femininity (5.Mc),  $a_9$  corresponds to the scale of Paranoia (6.Pa),  $a_{10}$  corresponds to the scale of Psychasthenia (7.Pt),  $a_{11}$  corresponds to the scale of Schizophrenia (8.Sc),  $a_{12}$  corresponds to the scale of Hypomania (9.Ma),  $a_{13}$  corresponds to the scale of Social introversion (0.It). Values of attributes are expressed by the so-called T-scores. The T-scores scale, which is traditionally attributed to MMPI, represents the following parameters: offset ranging from 0 to 100 T-scores, average equal to 50 T-scores, standard deviation equal to 10 T-scores.

Data can be represented in a graphical form as the so-called MMPI profiles. The profile always has a fixed and invariable order of its constituents (attributes, scales). Let a patient  $x$  be described by the data vector  $a(x) = [56, 78, 55, 60, 59, 54, 67, 52, 77, 56, 60, 68, 63]$ . Its profile is shown in Figure 1.

The data set was collected by T. Kucharski and J. Gomuła from the Psychological Outpatient Clinic. The data for the analysis (i.e., profiles of patients) were selected using the competent judge method (the majority of two-thirds of votes of three experts). The sample consists of over 1000 selected profiles of women examined by means of the MMPI-WISKAD test, split by this method into 19 nosological categories (and norm) mostly used by clinicians for the differential psychiatric diagnosis. A number of profiles in each category is not the same (it seems to be natural), but it satisfies condition for rule generation: at least 10 cases (profiles) for one attribute (scale or index) in the attribute vector (space). It enables us to extend the basic profile even to 100 attributes (13 scales plus 87 indexes). Such a basic profile extended by additional indexes or systems

of indexes is called multiprofile. In our research, a basic profile is extended by 33 indexes. Different combinations of scales constitute diagnostically important indexes (e.g., Gough's, Goldberg's, Watson-Thomas's, L'Abate's, Lovell's indexes - see <sup>7</sup>) and systems of indexes (e.g., Diamond's <sup>8</sup>, Leary's, Eichmann's, Petersen's, Taulbee-Sisson's, Butcher's <sup>4</sup>, Pancheri's). They have been determined on the basis of clinical and statistical analyses of many patients' profiles. All mentioned and some additional (e.g., the nosological difference-configuration Gough-Plużek's system) indexes have been implemented in the Copernicus system. All indexes are visualized in the form of multiprofile or other charts.

Differential indexes  $G_1$ ,  $G_2$ , and  $G_3$  are called the Goldberg's indexes. Difference-descriptive indexes  $D_1, \dots, D_{17}$  are based on combinations of the most important and strongest (most credible) clinical scales 1.Hp, 2.D, 3.Hy, 4.Ps, 6.Pa, 7.Pt, 8.Sc, and 9.Ma. These indexes constitute the Diamond's diagnosis system. Indexes  $L_{I.a}, L_{I.b}, L_{II.a}, L_{II.b}$  differentiating basic personality dimensions and indexes  $L_1, \dots, L_8$  differentiating and describing basic styles of interpersonal functioning constitute the Leary's system. All indexes have been browsed, some of them redefined and normalized to the range of 120 T-scores (i.e., the upper range of scales with the  $K$  suppressor for the normalizing group). It helps in visualizing the patient's profile. Some indexes ( $L_4, L_5$ ) seem to be overestimated and they give overstated results in comparison to basic personality dimensions and the remaining styles of interpersonal functioning. Below, we list all indexes used by us in research presented in this paper.

Goldberg's indexes:

- $G_1 = Hp + 2Ps - Ma$ ,
- $G_2 = 2Ps - Hy - Sc$ ,

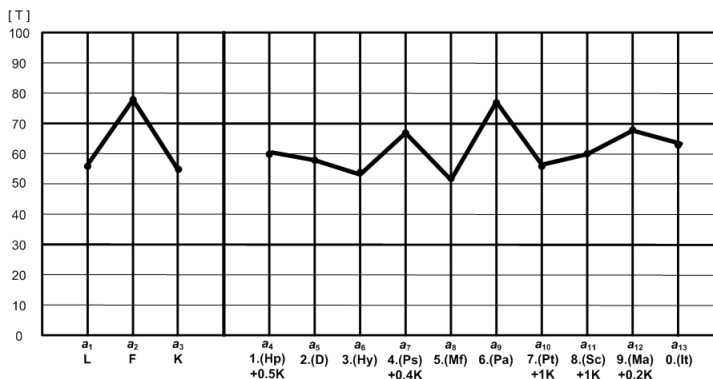


Figure 1: MMPI profile of a patient (example); suppressors +0.5K, +0.4K, +1K, +0.2K a correction value from raw results of scale K added to raw results of selected clinical scales

- $G_3 = (L + Pa + Sc) - (Hy + Pt)$ .

Diamond's indexes:

- $D_1 = (D + Ma)/2$ ,
- $D_2 = (Pt + Ps)/2$ ,
- $D_3 = (Hy + Pa)/2$ ,
- $D_4 = (Hp + Sc)/2$ ,
- $D_5 = (Hy + Ps)/2$ ,
- $D_6 = (Hy + Ps + Ma)/3$ ,
- $D_7 = (D + Pt)/2$ ,
- $D_8 = (Ma + Ps)/2$ ,
- $D_9 = (Hp + Hy)/2$ ,
- $D_{10} = (Pa + Sc)/2$ ,
- $D_{11} = (Hp + D + Hy + Pt)/4$ ,
- $D_{12} = (Ps + Pa + Sc + Ma)/4$ ,
- $D_{13} = (D + Hy + Hp)/3$ ,
- $D_{14} = (Pa + Pt + Sc)/3$ ,
- $D_{15} = (Ma + Ps + Pa + Sc)/4 - (D + Pt + Hy + Hp)/4$ ,
- $D_{16} = (D + Pt + Hy + Hp)/4$ ,
- $D_{17} = (Ma + Ps + Pa + Sc)/4$ .

A classical Diamond's system has been extended by the following indexes: impulsive behavior  $D_5$ , behavior disturbances  $D_6$ , supervision of impulses  $D_{11}$ , a lack of supervision of impulses (the so-called acting-out behavior)  $D_{12}$ , neurotism  $D_{13}$ , psychotism  $D_{14}$ , social aggression  $D_{15}$ .

Leary's indexes:

- $L_{I.a} = (K + Hy) - (F + Sc)$ ,
- $L_{I.b} = (K + Hy)/2 - (F + Sc)/2$ ,
- $L_{II.a} = (Ma + Hp) - (D + Pt)$ ,
- $L_{II.b} = (Ma + Hp)/2 - (D + Pt)/2$ ,
- $L_1 = (K + Ma) - (D + Pt)$ ,
- $L_2 = ((K + Hp + Hy)/3) - ((F + D + Pt + Sc)/4)$ ,
- $L_3 = (K + Hy) - (F + Sc)$ ,
- $L_4 = (D + Hy + Pt)/3$ ,
- $L_5 = (D + Pt)/2$ ,
- $L_6 = (F + D + Pt + Sc)/4$ ,
- $L_7 = (F + Ps + Ma)/3$ ,
- $L_8 = (Hp + Ma) - (D + Pt)$ .

### 3. Tools for Experiments

In our experiments, we used the following software tools:

- The Rough Set Exploration System (RSES) - a software tool featuring a library of methods and a graphical user interface supporting a variety of rough set based computations<sup>1, 20</sup>.
- COPERNICUS - a computer tool created by the authors for analysis of MMPI profiles of patients with mental disorders<sup>11</sup>.

The Copernicus system supporting clinical psychologists in differential and clinical diagnosis based on the overall analysis of profiles of patients examined by means of personality inventories is a

tool designed for the Java platform. The main features of the application are the following:

- **Multiplatforming.** Thanks to the Java technology, the application works on various software and hardware platforms. In the future, the tool can be adapted for platforms available in mobile devices (e.g. palmtops).
- **User-friendly interface** (see Figure 2). The interface is designed in order to make the use of the application in the medical environment possible. In the future, proper editors will be designed on the basis of defined domain ontologies for particular diseases supporting entering data. It will eliminate ambiguities in descriptions of cases (which is especially important for qualitative data).
- **The module of data visualization** allows presenting data in a clear and comprehensible way (for example, in a graphical way) for a person who must make a reasonable diagnostic decision.
- **Modularity.** The project of the application and its implementation takes into consideration modularity in order to make extending the application possible in the future and enlarging its usage on diagnosis based on different personal inventories. In the future, it will be obtained by using the architecture based on the so-called plug-ins.

The current version of this tool offers the following main functions:

- **Locating patients** in a profile space using a wide variety of measures and indexes (e.g., general distance measures, specialized measures, psychopathology indexes).
- **Matching patient profiles** to patterns of disorders using dendrograms generated by different clustering methods with a suitable visualization.
- **Creating diagrams** for psychopathological indexes defined in the professional literature (e.g., Goldberg's, Leary's indexes, Diamond's indexes).
- **Visualizing patient profiles** on the background of patterns of nosological classes as well as decision rules generated by popular data mining systems. An important thing is a unique visualization of decision rules (in the form of stripes put on profiles) supporting the nosological diagnosis.

- **Making diagnostic decisions** for patients described by MMPI profiles on the basis of classification functions obtained using the discriminant analysis module from the STATISTICA package.
- **Making diagnostic decisions** for patients described by MMPI profiles on the basis of rules embodied in the knowledge base of Copernicus. This knowledge base consists of several rule sets generated by different data mining and machine learning algorithms.

The Copernicus is released in two versions: Polish and English.

#### 4. Experiments

In our experiments, three rule generation algorithms available in the RSES system have been used, namely: genetic algorithm, covering algorithm, and LEM2 algorithm. The first algorithm originates in an order-based genetic algorithm coupled with heuristic (see <sup>2</sup>, <sup>21</sup>). It uses reducts for rule generation. Another two algorithms are based on a covering approach. A covering algorithm is described in <sup>2</sup> and <sup>22</sup>. The LEM2 algorithm was proposed by J. Grzymala-Busse in <sup>15</sup>. Covering-based algorithms produce less rules than algorithms based on an explicit reduct calculation. They are also (on average) slightly faster. It seems to be important if we extend a number of attributes in a decision table. Values of all scales (validity and clinical) can be treated as continuous quantitative data.

The total number of values covers a specific interval (from 0 to 100 T-scores). Building classification rules for such data can be difficult and/or highly inefficient. Therefore, for some rule generation algorithms, the so-called discretization is a necessary preprocessing step <sup>6</sup>. Its overall goal is to reduce the number of values by grouping them into a number of intervals. In many cases, discretization enables obtaining a higher quality of classification rules. In our research, for each algorithm, experiments have been performed for discretized data using the local method available in the RSES system <sup>2</sup>. This discretization technique is based on rough sets and Boolean reasoning.

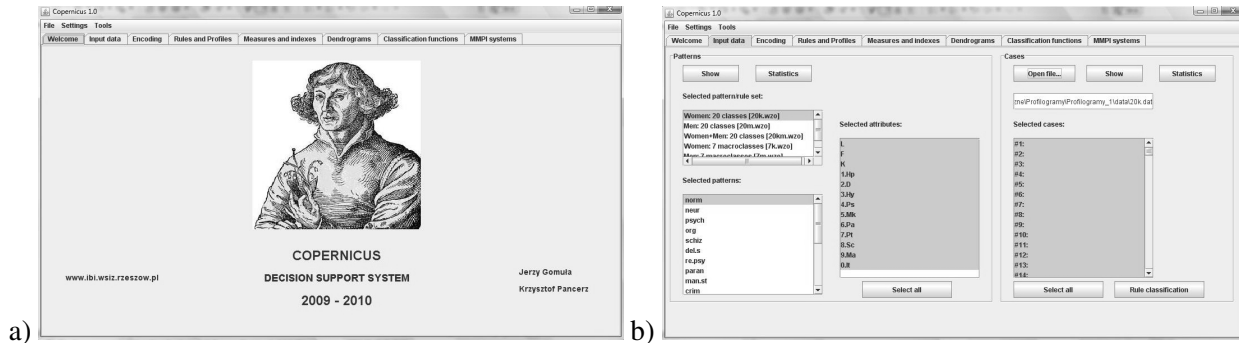


Figure 2: The Graphical User Interface of Copernicus: (a) welcome window, (b) input data window

To determine the accuracy of classification of new cases a cross-validation method has been used. The cross-validation is frequently used as a method for evaluating classification models. It comprises of several training and testing runs. First, the data set is split into several, possibly equal in size, disjoint parts. Then, one of the parts is taken as a training set for the rule generation and the remainder (a sum of all other parts) becomes the test set for rule validation. In our experiments, 5 cross-validation (CV-5) and 10 cross-validation (CV-10) tests were used.

We tested the following different situations according to the number of attributes:

1. Data table with basic (original) scales ( $L, F, K, 1.Hp, 2.D, 3.Hy, 4.Ps, 5.Mf, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It$ ).
2. Data table with basic scales and Goldberg's indexes ( $L, F, K, 1.Hp, 2.D, 3.Hy, 4.Ps, 5.Mf, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It, G_1, G_2, G_3$ ).
3. Data table only with Goldberg's indexes ( $G_1, G_2, G_3$ ).
4. Data table with basic scales and Leary's indexes ( $L, F, K, 1.Hp, 2.D, 3.Hy, 4.Ps, 5.Mf, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It, L_{I.a}, L_{I.b}, L_{II.a}, L_{II.b}, L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8$ ).
5. Data table only with Leary's indexes ( $L_{I.a}, L_{I.b}, L_{II.a}, L_{II.b}, L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8$ ).
6. Data table with basic scales and Diamond's indexes ( $L, F, K, 1.Hp, 2.D, 3.Hy, 4.Ps,$

$5.Mf, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It, D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, D_{14}, D_{15}, D_{16}, D_{17}$ ).

7. Data table with basic scales (without validity scales and the  $5.Mk$  scale) and Diamond's indexes ( $1.Hp, 2.D, 3.Hy, 4.Ps, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It, D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, D_{14}, D_{15}, D_{16}, D_{17}$ ).
8. Data table only with Diamond's indexes ( $D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, D_{14}, D_{15}, D_{16}, D_{17}$ ).
9. Data table with basic scales, Goldberg's indexes, Leary's indexes and Diamond's indexes ( $L, F, K, 1.Hp, 2.D, 3.Hy, 4.Ps, 5.Mf, 6.Pa, 7.Pt, 8.Sc, 9.Ma, 0.It, G_1, G_2, G_3, L_{I.a}, L_{I.b}, L_{II.a}, L_{II.b}, L_1, L_2, L_3, L_4, L_5, L_6, L_7, L_8, D_1, D_2, D_3, D_4, D_5, D_6, D_7, D_8, D_9, D_{10}, D_{11}, D_{12}, D_{13}, D_{14}, D_{15}, D_{16}, D_{17}$ ).

Experiments were carried out on a data set with over 1000 women.

Obviously, ranges of classification rule conditions can be restricted. We can replace  $\infty$  by:

- a maximal value of a given scale occurring for a given class in our sample,
- a maximal value of a given scale for all twenty classes,
- a maximal value of a given scale for a normalizing group (i.e., a group of women, for which norms of validity and clinical scales have been determined),

Table 2: The LEM2 rules for a data table with basic scales (fragment)

Rule ID	<i>L</i>	<i>F</i>	<i>K</i>	<i>1.Hp</i>	<i>2.D</i>	<i>3.Hy</i>	<i>4.Ps</i>
#1		$(-\infty, 68.5)$			$(73.5, \infty)$	$(69.5, \infty)$	$(57.5, \infty)$
#2		$(77.5, 85.5)$			$(73.5, \infty)$		$(57.5, \infty)$
#3		$(-\infty, 68.5)$		$(67.5, \infty)$	$(73.5, \infty)$	$(69.5, \infty)$	$(57.5, \infty)$

Rule ID	<i>5.Mf</i>	<i>6.Pa</i>	<i>7.Pt</i>	<i>8.Sc</i>	<i>9.Ma</i>	<i>0.It</i>	<i>class</i>
#1		$(-\infty, 72.5)$	$(68.5, \infty)$		$(-\infty, 56.5)$		<i>neur</i>
#2	$(-\infty, 52.5)$	$(72.5, 79.5)$	$(68.5, \infty)$	$(76.0, \infty)$		$(64.0, \infty)$	<i>schiz</i>
#3		$(-\infty, 72.5)$	$(68.5, \infty)$		$(-\infty, 56.5)$		<i>neur</i>

- a maximal value for all scales for a normalizing group, i.e., 120 T-scores.

A procedure of restricting ranges of classification rule conditions with the value  $-\infty$  is carried out similarly, but we take into consideration minimal values. A minimal value for all scales of normalizing group of women is 28 T-scores.

Indexes are linear combinations of scales. Therefore, restricting ranges of rule conditions for indexes is also possible and simple. Copernicus enables us to select the way of restricting ranges. Rule conditions are automatically restricted to the form readable for the diagnostician-clinician.

All experiment results have been collected in Tables 4, 5, 6 - for the LEM2 algorithm, the covering algorithm and the genetic algorithm, respectively.

Clinical research confirms high relevance of the Goldberg's indexes ( $G_1, G_2, G_3$ ) and their usefulness in differential diagnosis. However, from the analysis of results in Tables 4, 5, 6 we can conclude that classification accuracy and support for rules obtained in this way are significantly smaller than for rules obtained from basic scales and Diamond's and Leary's indexes (both taken separately and taken as an extended profile).

It is easy to see that the Goldberg's indexes do not take into consideration validity scales and they differentiate weakly non-validated profiles from validated ones (it is very important because the profile validation is the first step of differential analysis). Moreover, their border values do not take into consideration a gender of the examined person by means

of MMPI. The mentioned indexes differentiate well only four macroclasses and norm. Our previous research indicates that better results are obtained for linear classification functions (for 4 macroclasses and norm or for 19 nosological classes and norm - calculated separately for women and men). Significantly better results were obtained for a basic profile extended by classification functions (cf. <sup>12</sup>).

It is worth noting, on the basis of results from Table 3, that for rules extracted by the LEM2 algorithm, values of conditions in those rules for validity and clinical scales do not exceed the upper range 70 T-scores. 70 T-scores seems to be a border value. In the MMPI-WISKAD test, values above this threshold indicate an area of disturbances and pathologies of the examined person.

Amazingly, good differentiating properties are demonstrated by the Leary's indexes, which are rather regarded as auxiliary descriptive indexes. They are rather useful in the so-called interpretation quiet (i.e., for basic profiles in ranges 40 - 60 T-scores).

The classification accuracy and support of rules based on the Diamond's indexes or the basic profile and Diamond's indexes are always better than for the Goldberg's indexes (with or without the basic profile). For the LEM2 algorithm, the best results have been obtained for rules extracted from data for which a vector consisted of the 13 scale profile and 17 Diamond's indexes or the 13 scale profile and all considered indexes.

On the basis of rules, a proper diagnostic deci-

Table 3: The LEM2 rules for a data table with basic scales and Leary's indexes (fragment)

Rule ID	L	F	K	1.HP	2.D	3.Hy	4.Ps	5.MF	6.Pt	7.Pt	8.Sc	9.Ma	0.H
#0		(-∞;68.5)			(88.0,67.0)	(-∞;60.5)			(-∞;72.5)	(54.5,62.5)	(-∞;62.5)		
#1		(-∞;68.5)		(67.5,∞)	(73.5,∞)	(69.5,∞)	(57.5,∞)		(-∞;72.5)	(68.5,∞)		(-∞;56.5)	
#2		(-∞;68.5)		(51.0,56.5)		(-∞;60.5)	(57.5,∞)		(-∞;72.5)	(54.5,62.5)	(68.0,74.5)		
#3			(51.5,54.5)	(61.0,63.5)			(57.5,∞)		(-∞;72.5)	(54.5,62.5)			
#4		(77.5,85.5)			(73.5,∞)		(57.5,∞)		(-∞;52.5)	(72.5,79.5)	(76.0,∞)		(64.0,∞)
#5		(-∞;68.5)		(56.5,61.0)	(73.5,∞)	(63.5,69.5)	(57.5,∞)	(56.5,62.0)		(68.5,∞)		(-∞;56.5)	
#6		(69.5,74.0)			(73.5,∞)		(57.5,∞)	(52.5,56.5)		(68.5,∞)		(-∞;56.5)	
#7	(68.0,74.5)			(56.5,61.0)			(57.5,∞)				(76.0,∞)		
#8					(-∞;58.0)	(-∞;60.5)	(57.5,∞)		(-∞;72.5)			(64.5,∞)	
#9		(-∞;68.5)	(46.5,51.5)		(67.0,70.5)		(57.5,∞)		(-∞;72.5)	(62.5,68.5)	(68.0,74.5)		
#10	(51.5,55.5)					(-∞;60.5)	(57.5,∞)	(-∞;52.5)		(-∞;54.5)	(-∞;62.5)	(-∞;56.5)	
#11							(57.5,∞)	(56.5,62.0)				(64.5,∞)	
#12		(85.5,102.5)					(57.5,∞)		(79.5,∞)	(68.5,∞)	(76.0,∞)		(55.5,59.5)
#13		(-∞;68.5)	(54.5,65.5)	(51.0,56.5)		(-∞;60.5)	(57.5,∞)	(52.5,56.5)	(-∞;72.5)	(-∞;54.5)	(-∞;62.5)	(-∞;56.5)	
#14			(69.0,∞)	(67.5,∞)	(73.5,∞)	(69.5,∞)	(57.5,∞)		(-∞;72.5)			(-∞;56.5)	
#15		(102.5,∞)	(-∞;44.5)					(62.0,∞)					
#16		(102.5,∞)	(54.5,65.5)	(67.5,∞)	(73.5,∞)		(57.5,∞)		(79.5,∞)		(76.0,∞)	(64.5,∞)	
#17		(102.5,∞)		(67.5,∞)	(58.0,67.0)	(63.5,69.5)	(57.5,∞)	(62.0,∞)	(79.5,∞)	(68.5,∞)	(76.0,∞)		
#18	(74.5,∞)	(85.5,102.5)		(67.5,∞)		(69.5,∞)	(57.5,∞)	(62.0,∞)	(-∞;72.5)			(64.5,∞)	
#19		(102.5,∞)			(73.5,∞)	(63.5,69.5)	(57.5,∞)	(-∞;52.5)					

Rule ID	L <sub>1a</sub>	L <sub>1b</sub>	L <sub>1a</sub>	L <sub>1b</sub>	L <sub>1</sub>	L <sub>2</sub>	L <sub>3</sub>	L <sub>4</sub>	L <sub>5</sub>	L <sub>6</sub>	L <sub>7</sub>	L <sub>8</sub>	class
#0	(-23.5,∞)				(39.5,∞)					(58.5,∞)	(-∞;69.5)		nom
#1	(-23.5,∞)				(-∞;-39.5)					(58.5,∞)	(-∞;69.5)		neur
#2			(-17.5,∞)		(-39.5,∞)					(58.5,∞)	(-∞;69.5)		psych
#3										(58.5,∞)	(-∞;69.5)		org
#4						(-∞;-13.5)				(58.5,∞)			schiz
#5			(-∞;-34.5)		(-∞;-39.5)					(58.5,∞)	(-∞;69.5)		del.s
#6			(-∞;-34.5)		(-∞;-39.5)	(-∞;-13.5)				(58.5,∞)			rec-psys
#7										(58.5,∞)			paran
#8	(-51.5,-23.5)		(-17.5,∞)		(-39.5,∞)					(58.5,∞)			man.s
#9	(-51.5,-23.5)		(-17.5,∞)		(-39.5,∞)					(58.5,∞)	(69.5,75.5)		crim
#10	(-23.5,∞)		(-17.5,∞)		(-39.5,∞)	(-7.0,∞)				(58.5,∞)	(-∞;69.5)		alcoh
#11			(-17.5,∞)		(-39.5,∞)	(-13.5,-7.0)				(58.5,∞)	(75.5,∞)		drug
#12	(-∞;-51.5)				(-∞;-13.5)					(58.5,∞)	(-∞;69.5)		sinu
#13	(-23.5,∞)		(-17.5,∞)		(-39.5,∞)	(-7.0,∞)				(58.5,∞)			dissim
#14					(-∞;-39.5)					(58.5,∞)			den1
#15										(58.5,∞)			den2
#16	(-∞;-51.5)		(-17.5,∞)		(-39.5,∞)	(-∞;-13.5)				(58.5,∞)	(75.5,∞)		den3
#17	(-∞;-51.5)		(-17.5,∞)		(-39.5,∞)	(-∞;-13.5)				(58.5,∞)	(75.5,∞)		den4
#18													den5
#19													den6



Table 4: Results of experiments using the LEM2 algorithm

Attributes	CV-5		CV-10	
	Accuracy	Covering	Accuracy	Covering
Basic scales	0.950	0.694	0.957	0.706
Basic scales + Goldberg's indexes	0.956	0.682	0.935	0.702
Goldberg's indexes	0.760	0.494	0.793	0.509
Basic scales + Leary's indexes	0.959	0.732	0.970	0.729
Leary's indexes	0.896	0.630	0.897	0.640
Basic scales + Diamond's indexes	0.965	0.747	0.967	0.745
Diamond's indexes	0.949	0.737	0.961	0.730
Basic scales (without scales <i>L, F, K, 5.Mf</i> ) + Diamond's indexes	0.967	0.732	0.962	0.747
Basic scales + Goldberg's, Leary's and Diamond's indexes	0.962	0.727	0.962	0.729

Table 5: Results of experiments using the covering algorithm

Attributes	CV-5		CV-10	
	Accuracy	Covering	Accuracy	Covering
Basic scales	0.746	0.492	0.764	0.448
Basic scales + Goldberg's indexes	0.828	0.606	0.833	0.586
Goldberg's indexes	0.676	0.505	0.680	0.513
Basic scales + Leary's indexes	0.807	0.471	0.790	0.457
Leary's indexes	0.751	0.517	0.786	0.529
Basic scales + Diamond's indexes	0.786	0.450	0.786	0.413
Diamond's indexes	0.751	0.373	0.792	0.352
Basic scales (without scales <i>L, F, K, 5.Mf</i> ) + Diamond's indexes	0.823	0.432	0.820	0.411
Basic scales + Goldberg's, Leary's and Diamond's indexes	0.788	0.355	0.777	0.380

Table 6: Results of experiments using the genetic algorithm

Attributes	CV-5		CV-10	
	Accuracy	Covering	Accuracy	Covering
Basic scales	0.862	1.000	0.880	1.000
Basic scales + Goldberg's indexes	0.884	1.000	0.897	1.000
Goldberg's indexes	0.620	0.836	0.620	0.844
Basic scales + Leary's indexes	0.900	1.000	0.911	1.000
Leary's indexes	0.772	1.000	0.782	1.000
Basic scales + Diamond's indexes	0.894	1.000	0.907	1.000
Diamond's indexes	0.862	1.000	0.870	1.000
Basic scales (without scales <i>L, F, K, 5.Mf</i> ) + Diamond's indexes	0.890	1.000	0.895	1.000

sion can be made. In the Copernicus system, we have implemented, among others, several aggregation factors for the different sets of rules indicating the class to which a given profile is classified. Copernicus enables us to visualize such a meta-diagnosis in the form of the so-called classification star, bar charts, or stripes put on profiles. It is readable and user-friendly for diagnosticians-clinicians. This area is a subject of further research.

## 5. Conclusions

In this paper, we have examined several combinations of MMPI data for the classification purpose. Our main goal is to deliver to diagnosticians and clinicians an integrated tool supporting the comprehensive diagnosis of patients. The Copernicus system is flexible and it can also be diversified into supporting differential diagnosis of profiles of patients examined using another professional multidimen-

sional personality inventories, for example MMPI-2/MMPI-2-RF<sup>3</sup> (including additional control scales and dozens of clinical scales and subscales, RC scales and additional scales<sup>19</sup>), MMPI-A (for adolescents), CBA2.0, 16EPQ-R, NEO PI-R, MBTI.

## Acknowledgement

This paper has been partially supported by the grant from the University of Information Technology and Management in Rzeszów, Poland.

## References

1. J.G. Bazan and M.S. Szczuka, "The Rough Set Exploration System", *Transactions on Rough Sets III*, ser. LNCS, vol. 3400, Springer, Heidelberg, 2005, pp. 37-56.
2. J.G. Bazan, H.S. Nguyen, S.H. Nguyen, P. Synak and J. Wróblewski, "Rough set algorithms in classification problem", *Rough Set Methods and Applications*, ser. Studies in Fuzziness and Soft Computing, Physica-Verlag, Heidelberg, 2000, pp. 49-88.
3. Y.S. Ben-Porath and A. Tellegen, *MMPI-2-RF (Minnesota Multiphasic Personality Inventory-2 Restructured Form): Manual for administration, scoring, and interpretation*, University of Minnesota Press, Minneapolis; 2008.
4. J.N. Butcher (ed.), *MMPI: Research Developments and Clinical Application*, McGraw-Hill Book Company; 1969.
5. M. Choynowski, *Multiphasic Personality Inventory*, Psychometry Laboratory, Polish Academy of Sciences, Warsaw; 1964 (in Polish).
6. K. Cios, W. Pedrycz, R.W. Swiniarski and L.A. Kurgan *Data mining. A knowledge discovery approach*, Springer, New York; 2007.
7. W. Dahlstrom, G. Welsh and L. Dahlstrom, *An MMPI handbook*, vol. 1-2, University of Minnesota Press, Minneapolis; 1986.
8. W.G. Dahlstrom, G.S. Welsh G.S, *An MMPI. A Guide to use in clinical practice*, The University of Minnesota Press, Minneapolis; 1965.
9. W. Duch, T. Kucharski, J. Gomuła and R. Adamczak, *Machine learning methods in analysis of psychometric data. Application to Multiphasic Personality Inventory MMPI-WISKAD*, Toruń; 1999 (in Polish).
10. J. Gomuła, W. Paja, K. Pancierz and J. Szkoła, "A preliminary attempt to rules generation for mental disorders", in *Proceedings of the HSI 2010*, Rzeszow, Poland, 2010, pp. 763-765.
11. J. Gomuła, K. Pancierz and J. Szkoła, "Analysis of MMPI profiles of patients with mental disorders - the first unveil of a new computer tool", *Applications of Systems Science*, Academic Publishing House EXIT, Warsaw, 2010, pp. 297-306.
12. J. Gomuła, K. Pancierz and J. Szkoła, "Classification of MMPI profiles of patients with mental disorders - experiments with attribute reduction and extension". *Rough Set and Knowledge Technology*, ser. LNAI, vol. 6401, Springer, Heidelberg, 2010, pp. 411-418.
13. J. Gomuła, K. Pancierz and J. Szkoła, "Copernicus - an expert system supporting differential diagnosis of patients examined using the MMPI test - an index-rule approach", in *Proceedings of the HEALTHINF 2011*, Rome, Italy, 2011.
14. R.A. Greenes, *Clinical Decision Support: The Road Ahead*, Elsevier; 2007.
15. J. Grzymala-Busse, "A new version of the rule induction system LERS", *Fundamenta Informaticae*, 31, 27-39 (1997).
16. D. Lachar, *The MMPI: Clinical assessment and automated interpretations*, Western Psychological Services, Fate Angeles; 1974.
17. Z. Pawlak, *Rough Sets - Theoretical Aspects of Reasoning About Data*, Kluwer Academic Publishers, Dordrecht; 1991.
18. Z. Płużek, *Value of test WISKAD-MMPI for nosological differential diagnosis*, The Catholic University of Lublin; 1971 (in Polish).
19. PsychoCorp: <http://psychcorp.pearsonassessments.com>
20. The Rough Set Exploration System, <http://logic.mimuw.edu.pl/rses>
21. J. Wróblewski, "Genetic algorithms in decomposition and classification problem", *Rough Sets in Knowledge Discovery*, vol. 2, 1998, pp. 471-487
22. J. Wróblewski, "Covering with reducts - a fast algorithm for rule generation", *Rough Sets and Current Trends in Computing*, ser. LNAI, vol. 1424, Springer, Heidelberg, 1998, pp. 402-407.