pattern recognition, k-NN classifier,
amyotropic lateral sclerosis, erythropoietin

Adam JÓŹWIK[*], Beata SOKOŁOWSKA[**], Irena NIEBROJ-DOBOSZ[***],
Piotr JANIK[****], Hubert KWIECIŃSKI[****]

PATTERN RECOGNITION APPROACH TO DIFFERENTIATION OF DISEASE SEVERITY IN PATIENTS WITH AMYOTROPHIC LATERAL SCLEROSIS

A possibility of recognition of the clinical status of patients with amyotrophic lateral sclerosis (ALS) in relation to severity of the disease was investigated. Three groups: (i) healthy controls (n=15) and two subgroups of ALS patients (ii) mild (n=15) and (iii) severe (n=15) were considered as classes. Four features of the subjects: (i) their age (AGE) (ii) erythropoietin concentration in serum (SERUM), (iii) in cerebrospinal fluid (CSF), and (iv) duration time of the disease (T_{dis}) were used for classifier construction based on the *k* Nearest Neighbours (*k*-NN) rule, known from pattern recognition theory. The presented results demonstrate that the pattern recognition approach may be useful for the evaluation of the severity of the ALS disease.

1. INTRODUCTION

Amyotrophic lateral sclerosis (ALS) is fatal motor neuron disease with progressive degeneration of cortical, bulbar and spinal cord motor neurons [4, 10]. The pathogenesis of this disease is still a mystery and no effective treatment is known yet. Important to recognize for particular patients is the possible course of the disease, either they would survive only up two years, or the disease would be longer lasting. Erythropoietin (EPO) in cerebrospinal fluid is suggested to differentiate the severe progressing ALS group from the more mild one [1]. However, it is also indicated, that EPO levels in CSF in the severe and mild ALS groups are comparable [9]. Solving this controversies in EPO concentration in CSF would be of interest, if we take into account the fact that EPO plays a neuroprotective and neurotrophic action [1, 2] and may be also a potential therapeutic factor for ALS [3, 8].

In the presented report the possibility of differentiating the severe and mild progressing ALS patients was evaluated by testing, except of the EPO concentration in serum and cerebrospinal fluid, also additional variables (the age of the patients and duration time of the disease). The pattern recognition methods were applied for this purpose.

2. MATERIALS AND METHODS

2.1. SUBJECTS AND BIOCHEMICAL MEASUREMENTS

Paired serum (SERUM) and cerebrospinal fluid (CSF) samples were collected and analyzed from 30 patients with amyotrophic lateral sclerosis (ALS group) and 15 healthy subjects (control group) as presented in Figure 1. The diagnosis of ALS was based on WFN El Eskorial criteria (World Federation of Neurology). Fifteen patients developed severe impairment in swallowing, speech, breathing, walking or motor function of upper limbs within two years (severe ALS group). In remaining 15 patients severe clinical symptoms appeared after 2 years, at least (mild ALS group). EPO in serum and CSF were determined using an enzyme-linked immunosorbent ELISA technique (Quantikine ® IVRD ® Epo kit R&D Systems Inc).

Following variables (features) were used for the calculations: (i) AGE (feature 1, age of the subjects), (ii) SERUM (feature 2, the EPO concentration in serum, (iii) CSF (feature 3, the EPO level in cerebrospinal fluid), and (iv) T_{dis} (feature 4, duration time of the disease). The classes were defined according to groups, which are presented in Figure 1: (i) control, (ii) mild ALS and (iii) severe ALS.

[*] Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland
Corresponding author, e-mail: adam.jozwik@ibib.waw.pl
[**] Department of Respiratory Research, Medical Research Center, Polish Academy of Sciences, Warsaw, Poland
[***] Neuromuscular Unit, Medical Research Center, Polish Academy of Sciences, Warsaw, Poland
[****] Department of Neurology, Medical University of Warsaw, Warsaw, Poland

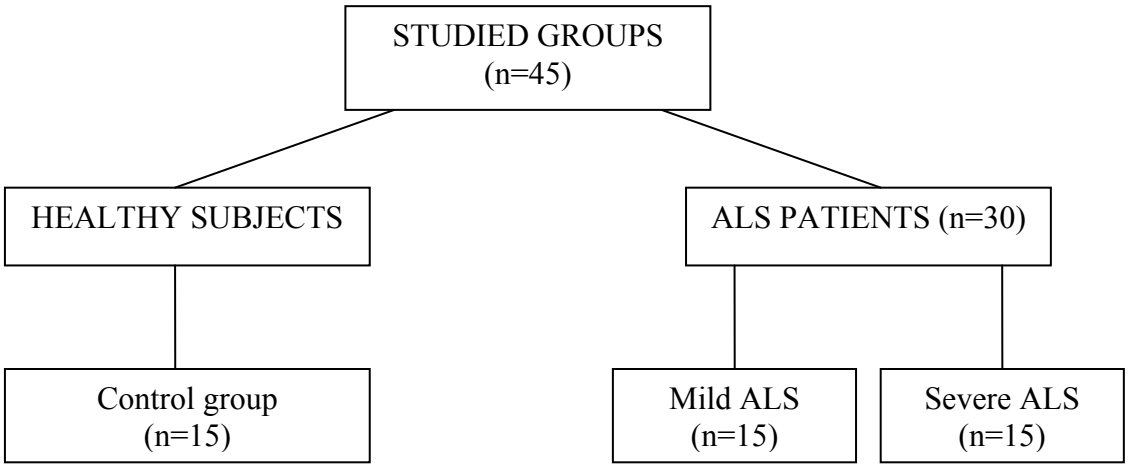

Fig.1 Scheme of studied groups

2.2. THE APPLIED PATTERN RECOGNITION METHODS

The statistical pattern recognition deals with methods of objects classification. Pattern recognition offers numerous algorithms for constructing the decision rule, i.e. the classifier. It is assumed that objects are represented by a set of parameters (variables) called features. Each of the objects can be treated as a point or a vector in a feature space. The set of all possible feature vectors (points) is called as a feature space. No class description is known. Instead of it, a set of objects with known class membership is given. It is called a reference set (or a training set). This set enables to divide the feature space into some regions, a separate one for each of the considered classes. Each determined region ought to contain the maximum number of objects from the class it is associated with, and minimum number of objects from the remaining classes. The classifier can be defined by the boundaries of these regions, described by the equations of the corresponding hypersurfaces. If the classified object falls in the decision region that corresponds to the class i then it is assigned to this class. Another approach requires defining the similarity measure between two objects, and next the similarity measure between the single object and the group of objects, i.e. the class of objects. The classification consists in indication the class most similar to the object being classified. Instead of similarity measure, a closeness measure, or distance measure, can be used. As a distance measure, between two objects, a city or Euclidean metric can be applied and the former was chosen in the present work. The distance measure must satisfy four conditions known as: non-negativity, identity, symmetry and triangle inequality. A closeness measure, between the single object and the class, can be defined on the basis of the distance measure. The mentioned above conditions do not need to be satisfied by the closeness measure.

One of the most popular decision rules of this type is the k Nearest Neighbour rule (k-NN) [5, 6, 7]. The closeness measure between the single object $\boldsymbol{x}$ and the class i is defined as a ratio k_i/k, where k denotes number of objects in the reference set X closest to $\boldsymbol{x}$ (nearest neighbours), and k_i is the number of object from the class i among these k nearest neighbours. Value of k is being established experimentally. The classified object is assigned to the class that corresponds to the greatest value of the ratio k_i/k. Usually the k is the same for each classified object, so, the value of the closeness measure can be identified with k_i. A simple example given on Figure 2 illustrates how the k-NN rule operates.

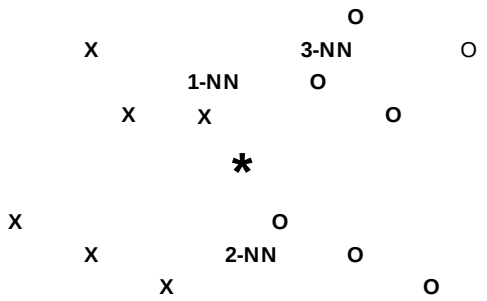

Fig. 2 Illustration of the k-NN rule. The symbols 1-NN, 2-NN and 3-NN denote the first, the second and the third nearest neighbour respectively. The point „*" is qualified to class 2, since two out of its three nearest neighbours come from this class ("X" - points from class 1, "O" – points belonging to class 2)

A value of k should be determined in such a way that offers the smallest probability of misclassification. To estimate the probability of misclassification, especially in the case of small data reference sets, the leave one out method is recommended. It consists in classification of each object $\boldsymbol{x}$ from the training set X by k-NN rule, with nearest neighbours searched for in the current set X-{$\boldsymbol{x}$}. A probability of misclassification is estimated by an error rate $E_r=r/m$, where r is a number of misclassified objects and m is a number of classified points, i.e. the number of points in the reference set. The Figure 3 explains how the leave-one-out method is used for an error rate calculation.

The training set may contain redundant features and features not related with the considered classes. Attendance of such kind of features can lead to error rate increasing and for this reason feature selection ought to be performed. In the case when the number of features is small, it is possible to review all possible feature combinations and to select the feature subset that offers the smallest misclassification rate.

1	2	3	4	5	6	7	8	9	10
x	**x**	**x**	**x**	**o**	**x**	**o**	**o**	**o**	**o**

Fig. 3 Illustration of the leave-one-out method. The 1-NN rule misclassifies three points, 4, 5 and 6 whereas the 3-NN rule misclassifies two points, 5 and 6 ("X" - points from class 1, "O" – points belonging to class 2)

Instead of single k-NN classifier, a parallel net of two-decision k-NN component classifiers can be used. The final decision is formed by voting these component classifiers. The value of k is being determined separately for each component classifier. Similarly, for each of the components classifiers, a separate feature selection is recommended.

3. RESULTS AND DISCUSSION

It was estimated how well the considered features (AGE, SERUM, CSF and T_{dis}) differentiate the examined classes, see Figure 1 and Table 1. Differentiation was investigated as between three classes as well as in the pairs of classes. The feature combinations consisted of (i) single features, (ii) the complete feature set and (iii) the selected features were evaluated (Table 1). The three out of four features (AGE, SERUM, CSF) were available for the all three classes, while T_{dis} could be taken into account for comparison the patient groups. Thus, the mild and severe ALS classes might be differentiated using three, as well as four features. The classifier recognizes the class of object on the basis of feature values. The lower is the error rate E_r the easier is the class differentiation.

Table 1 Misclassification rates (E_r) for healthy subjects and mild and severe progressing ALS

FEATURES	3 CLASSES	2 CLASSES		
	1. Control 2. Mild ALS 3. Severe ALS	1. Control 2. Mild ALS	1. Control 2. Severe ALS	1. Mild ALS 2. Severe ALS
AGE (f1)	0.489	0.467	0.300	0.267
SERUM (f2)	0.444	0.300	0.267	0.400
CSF (f3)	0.244	0.067	0.033	0.333
T_{dis} (f4)	-	-	-	0.233
All features together Without feature selection: {f1,f2,f3} or {f1,f2,f3,f4}	0.244 {f1,f2,f3}	0.133 {f1,f2,f3}	0.000 {f1,f2,f3}	0.233 {f1,f2,f3} 0.133 {f1,f2,f3,f4}
After feature selection	0.244 {f1,f2,f3} or {f3}	0.067 {f3}	0.000 {f1,f3}	0.233 {f1,f2,f3} 0.133 {f1,f3,f4}

If the three classes were analyzed jointly the lowest error rate offered the single feature CSF (E_r=0.244, Table 1). Additional features neither decreased nor increased the value of E_r. It can be seen that differentiation between the control and severe ALS class is the easiest. Even the single feature CSF enabled very good recognition between these two classes (E_r=0.033) and it was perfect for the three features {AGE, SERUM, CSF} or for CSF combined with the AGE (E_r=0.000). The error rates for differentiating the control class and mild were remarkably higher, but they were still low and reached the minimum value for the single feature CSF (E_r=0.067). Recognition between two patient classes was exactly twice worse (E_r=0.133 for set of {AGE, CSF, T_{dis}}). Furthermore, differentiation between the healthy controls and all ALS patients was

analyzed (without dividing them in mild and severe cases, the results are not showed in Table 1). It was found that the misclassification rate equalled 0.044, without feature selection. However, the single feature CSF indicated by feature selection, offered twice lower error rate value (E_r=0.022). In general, *k*-NN classifier offers the best recognition for feature CSF, and this variable may be good support "diagnostic" parameter in the ALS disease. The feature SERUM gives worse recognition. Differentiation between mild and severe progressing cases of ALS is the best for the set, which consists of features such as AGE, CSF and T_{dis} (i.e. the age of subjects, the EPO concentration in CSF and duration time of the disease).

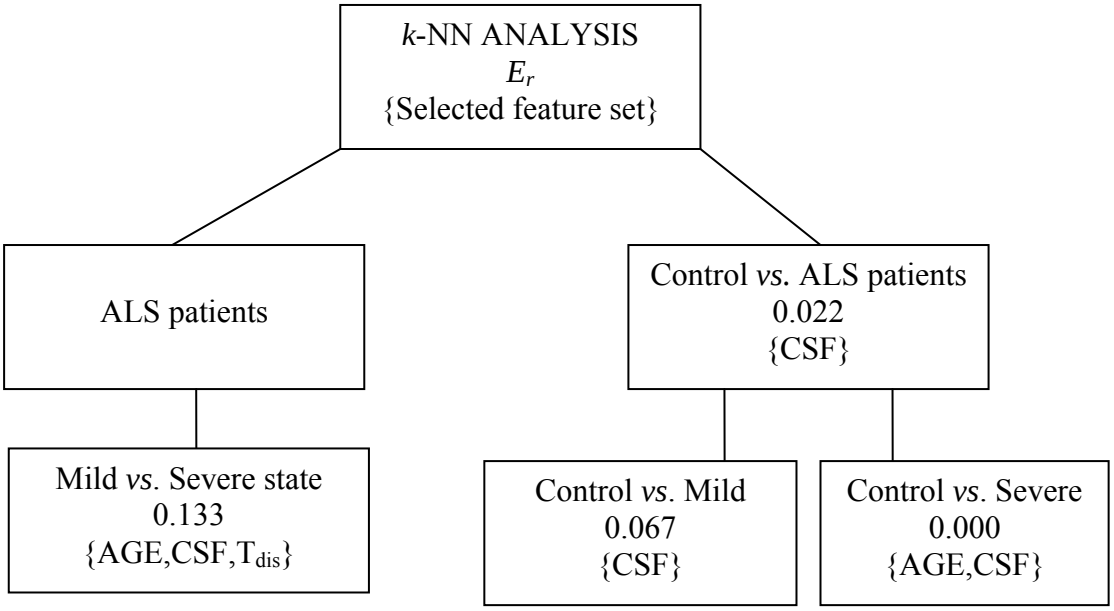

Fig. 4 Summary results of *k*-NN analysis

In summary, CSF and serum EPO were analyzed in order to recognize their relation to the severity of the disease course in the patients with ALS. The results of analysis received with the use of *k*-NN classifier are in agreement other studies on ALS [1, 9, 11]. It is reported that the EPO concentration is reduced in CSF of ALS patients, but in serum it is slightly decreased or normal. The variable of CSF is the best for recognition patients *versus* control subjects, among of all measured features. Controversies, however, are if the EPO level in CSF can predict the possible course of the disease, either it will be mild (that is over 2 years) or severe (up to two years). Contrary to the previous reports [1, 11], the differences in EPO concentration in CSF between the mild and severe ALS cases, using a classical statistic tests, were not significant [9]. However, analysis based on using *k*-NN classifier enabled to detect the difference between the mild and severe ALS patients.

BIBLIOGRAPHY

[1] BRETTSCHNEIDER J., WIDL K., SCHATTAUER D. et al., Cerebrospinal fluid erythropoietin (EPO) in amyotrophic lateral sclerosis, Neuroscience Letters, Vol. 416, pp.257-260, 2007.
[2] BUEMI M., CAVALLARO E., FLOCCARI F. et al., Erythropoietin and the brain: from neurodevelopment to neuroprotection, Clinical Science, Vol. 103, pp.275-282, 2002.
[3] CHUNG Y.H., JOO K.M., LEE W.B. et al., Enhanced expression of erythropoietin in the central nervous system of $SOD1^{G93A}$ transgenic mice, Brain Research, Vol. 1016, pp.272-280, 2004.
[4] COZZOLINO M., FERRI A., CARRI M.T., Amyotrophic Lateral Sclerosis: From current developments in the Laboratory to Clinical Implications, Antioxidants & Redox Signaling, Vol. 10, No. 3, pp.405-443, 2008.
[5] DEVIJVER P.A., KITTLER J., Pattern Recognition: A Statistical Approach. Prentice Hall, London, 1982.
[6] DUDA R.O., HART P.E., STOCK D.G., Pattern Classification. John Wiley and Sons, New York, 2001.
[7] FIX E., HODGES J.L., Discriminatory Analysis: Nonparametric Discrimination Small Sample Performance, project 21-49-004, Report Number 11, USAF School of Aviation Medicine, Randolph Field, Texas, pp. 280-322, 1952, reprinted in the book: Dasarathy BV "NN Pattern Classification Techniques" pp. 40-56, IEEE Computer Society Press, 1991.
[8] GRUNFELD J.F., BARHUM Y., BLODHEIM N. et al., Erythroietin delays disease onset in an amyothropic lateral sclerosis model, Experimental Neurology, Vol. 204, pp.260-263, 2007.

[9] JANIK P., KWIECIŃSKI H., NIEBROJ-DOBOSZ I., Erythropoietin concentration in serum and cerebrospinal fluid of patients with amyotropic lateral sclerosis (ALS), in press.
[10] ROWLAND L.P., SHNEIDER N.A., Amyotrophic Lateral Sclerosis, New England Journal of Medicine, Vol. 344, No. 22, pp.1688-1700, 2001.
[11] WIDL K., BRETTSCHNEIDER J., SCHATTAUER D. et al., Erythropoietin in cerebrospinal fluid: Age-related reference values and relevance in neurological disease, Neurochemical Research, Vol.32, pp.1163-1168, 2007.