pattern recognition, k-NN classifier,
amyotrophic lateral sclerosis,
matrix metalloproteinase, cerebrospinal fluid

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

ANALYSIS OF MATRIX METALLOPROTEINASES (MMPS) IN CEREBROSPINAL FLUID OF PATIENTS WITH AMYOTROPHIC LATERAL SCLEROSIS (ALS)

Matrix metalloproteinases (MMPs) are implicated in the pathogenesis of motor neuron degeneration in amyotrophic lateral sclerosis (ALS) and might be potential markers of diagnosis, prognosis and monitoring treatment effects. The aim of the present study was evaluation of the MMPs significance in cerebrospinal fluid (CSF MMPs) of patients with ALS in relation to severity of the disease. Metalloproteinases MT-MMP-1, MMP-2, MMP-9 and additionally age of subjects and disease duration were analyzed. The results demonstrate that the error of differentiation between healthy subjects and ALS patients (for MMP-2 feature) as well as between mild and severe ALS states (for CSF MMPs set) equalled to 0.033. In conclusion, the pattern recognition approach may be useful for differentation of ALS progressing on the basis of CSF MMPs features.

1. INTRODUCTION

Extracellular matrix (ECM) of the cells takes part in the development of normal function of plant, invertebrate and vertebrate organisms. Important regulators of cell functions are extracellular proteases. Various types of proteinases are implicated in the regulation of cell-cell and cell-matrix interactions. Among of them over 20 zinc-dependent endopeptidases (i.e. matrix metalloproteinases, MMPs) play an important role in the maintaiment of EMC homeostasis [9, 15]. MMPs are classified as the matrix in subfamily of zinc metalloprotease family M10 in the MEROPS database (http://www.metrops.sanger.ac.uk/) [10]. They are currently also implicated in the pathogenesis of several neurological diseases, as multiple sclerosis, Alzheimer disease and amyotrophic lateral sclerosis [12, 14, 15].

Amyotrophic lateral sclerosis (ALS) is a rare degenerative disorder of the large motor neurons of the cerebral cortex, brain stem, and spinal cord that results in progressive wasting and paralysis of voluntary muscles [13]. It is recommended to use some selected MMPs analysis for prognosis of the disease [1, 3], or as a marker for monitoring the treatment effects [12].

The aim of the present study was evaluation of the matrix metalloproteinase significance, such as membrane type matrix matalloproteinase-1 (MT-MMP-1), gelatinases A (MMP-2) and B (MMP-9) in cerebrospinal fluid of patients with ALS, in order to establish if they can be used a prognostic factor and potential marker for monitoring ALS treatment.

2. MATERIALS AND METHODS

2.1. SUBJECTS AND MEASUREMENTS

Thirty patients with amyotrophic lateral sclerosis (ALS) and fifteen control healthy subjects were studied (from 26 to 69 yrs). The clinical status including muscle strength, respiratory function, assessment of symptoms and neurological examination of the patients were evaluated. According to the clinical severity symptoms the ALS patients were divided into two subgroups: (i) with mild steady progressing over 2 yrs, and (ii) the severe ALS with rapidly developed impossent [2]. The lumbar cerebrospinal (CSF) samples were collected during diagnostic procedures. Membrane type matrix metaloproteinase (MT-MMP-1) and gelatinases A (MMP-2) and B (MMP-9) were determined by immunoassay procedures (biochemical methods are described in detail elsewhere [11]). The Local Ethical Committee of Warsaw Medical University approved the study protocol.

Following variables were used for the calculations: (i) MT-MMP-1^{CSF} (feature 1), (ii) MMP-2^{CSF} (feature 2), (iii) MMP-9^{CSF} (feature 3), (vi) AGE (the age of the subjects, feature 3), and (v) T_{dis} (the disease duration, feature 5). The studied groups (in each ones n=15) were: (i) control (healthy subjects, as class I), (ii) mild ALS patients (class II), and (iii) severe ALS patients (class III).

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

3. THE PATTERN RECOGNITION METHODS

A possibility of recognition of the clinical status of patients with amyotrophic lateral sclerosis (ALS) in relation to severity of the disease was studied. In our previous paper we analyzed concentration of erythropoietin (EPO) in patients with ALS by applying pattern recognition methods [8]. In present paper the another version of the distance based classifier was used for analyzing the relation between the severity of ALS disease and the concentration of some metalloproteinses in cerebrospinal fluid (CSF MMPs). The pattern recognition approach was in detail described elsewhere [4-7]. All error rates, in the present study, were calculated with using the leave-one-out method. The data were standardized on the basis of feature mean values and standard deviations. The city metric was used as the distance function.

It can be noticed that the decision is difficult only for the objects that lie in the class overlap areas. The class areas can be determined according to the formulas (1) and (2). Only such kind of objects that lie in the intersections of these areas ought to be classified by the *k*-NN rule. So, different decision rules should be used for different objects. Such an idea was introduced in [7]. To present this idea more strictly, some class areas and a class overlap area will be defined.

Let us assume that the reference set consists of *nc* subsets: X_1, X_2, ..., X_{nc} and each X_i contains samples from the class *i* only**.** We associate these sets with the positive real numbers $e_1,e_2,...,e_{nc}$ and the class areas A_i defined according to the following formulas:

$$ei = \max_{xj \in Xi} d(Xi-\{xj\},\{xj\}), \quad (1)$$

$$Ai=\{x: d(Xi,\{x\}) \leq ei\}, \quad (2)$$

where X_i is a subset of the reference set which corresponds to the class *i* and *d* denotes a distance function between two sets. As a distance between two sets we understand the distance between two nearest points each of them belonging to a different set. The symbol $\{\boldsymbol{x}_j\}$ is a set containing a single $\boldsymbol{x}_j$. The areas A_i for two classes and two features are illustrated on the Figure 1.

Now, we define the two-stage classification scheme. The first stage decides whether the classified point lies:

a) outside of each area A_i, i=1,2,...,*nc*,

b) exactly in one of the areas A_i, then the class *i* is assigned or

c) simultaneously in two or more areas A_i.

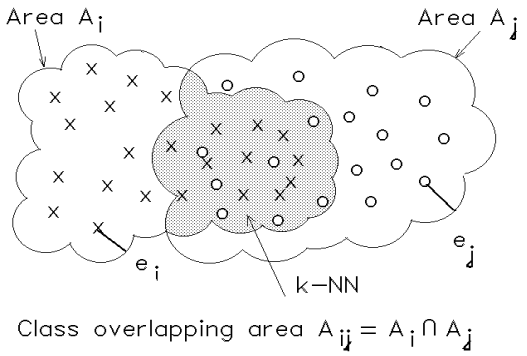

Fig. 1. An example of the areas A_1, A_2 and their intersection

In the situation (a) the decision is refused, if the case (b) takes place then the class *i* is assigned. For points that satisfy the condition (c), the *k*-NN rule is applied. The *k* nearest neighbours are found in the whole reference set. Some of these *k* nearest neighbours can lie outside of the intersection of all A_i containing the classified sample.

4. RESULTS AND DISCUSSION

It was estimated how well the considered of several metalloproteinases in cerebrospinal fluid (i.e. MT-MMP-1^{CSF}, MMP-2^{CSF} and MMP-9^{CSF}) and additional variables as subject age (AGE) and disease duration (T_{dis}), differentiate the examined groups (i.e. control, mild and severe ALS ones). Differentiation was investigated as between three classes and also 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. The four out of five features (MT-MMP-1^{CSF}, MMP-2^{CSF}, MMP-9^{CSF}, AGE) were available for the all three classes, while T_{dis} could be taken into account for comparison of the patient groups. Thus, the mild and severe ALS classes might be differentiated using four as well as five 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, mild and severe progressing ALS groups

FEATURES	3 CLASSES	2 CLASSES		
	I. CONTROL II. MILD ALS III. SEVERE ALS	I & II	I &III	II & III
MT-MMP-1^{CSF} (F1)	0.622	0.300	0.300	0.600
MMP-2^{CSF} (F2)	0.089	0.100	0.033	0.033
MMP-9^{CSF} (F3)	0.378	0.100	0.200	0.333
AGE (F4)	0.489	0.467	0.300	0.267
T_{DIS} (F5)	-	-	-	0.233
ALL FEATURES TOGETHER WITHOUT FEATURE SELECTION: {F1,F2,F3,F4} OR {F1,F2,F3,F4,5}	0.111 {F1,F2,F3,F4}	0.100 {F1,F2,F3,F4}	0.033 {F1,F2,F3,F4}	0.067 {F1,F2,F3,F4} 0.033 {F1,F2,F3,F4,F5}
AFTER FEATURE SELECTION	0.044 {F1,F2,F3}	0.033 {F1,F2,F3}	0.033 {F2}	0.033 {F2} 0.033 {F2}

Looking at the results (Tab. 1) concerning the pair of classes I & II (control *vs*. mild ALS), it can be seen that two features 2 and 3 (i.e. MMP2^{CSF} and MMP-9^{CSF}) offers low error rate (0.1), the feature 1 (i.e. MT-MMP-1^{CSF}) offers the error rate that equals 0.3 and the error rate for the feature 4 is nearly 0.5. The feature 2 (MMP-2^{CSF}) was also very good for the pair consisted of the classes I & III (i.e. control *vs*. severe ALS) as well as for the classes II & III (mild *vs*. severe ALS), and was three time better than for the classes I & II. In the remaining cases, the error rates varied between 0.2 and 0.6. The feature 5 (i.e. T_{dis}) as single feature better differentiate patients groups than features 1, 3, 4, and worse than feature 2.

If the three classes were analyzed jointly the lowest error rate offered the CSF MMPs set (E_r=0.044). All features together increased the value of E_r, although it decreased above twice after feature selection. It can be seen that differentiation of class pairs is yet easier. After feature selection error rates achieved the value of 0.033. Even the single feature MMP-2^{CSF} enabled very good recognition between two classes I & III as well as II & III (E_r=0.033). But the same recognition quality of the classes I & II was reached for the set of three CSF MMPs. Other two features, i.e. AGE, T_{dis}, were not necessary.

In general (see Fig. 2), *k*-NN classifier offers the best recognition for CSF MMPs set (E_r=0.017). Moreover, differentiation between mild and severe progressing cases of ALS is the best for MMP-2^{CSF} and it may be good support "diagnostic" variable in the ALS disease. Other features give worse recognition.

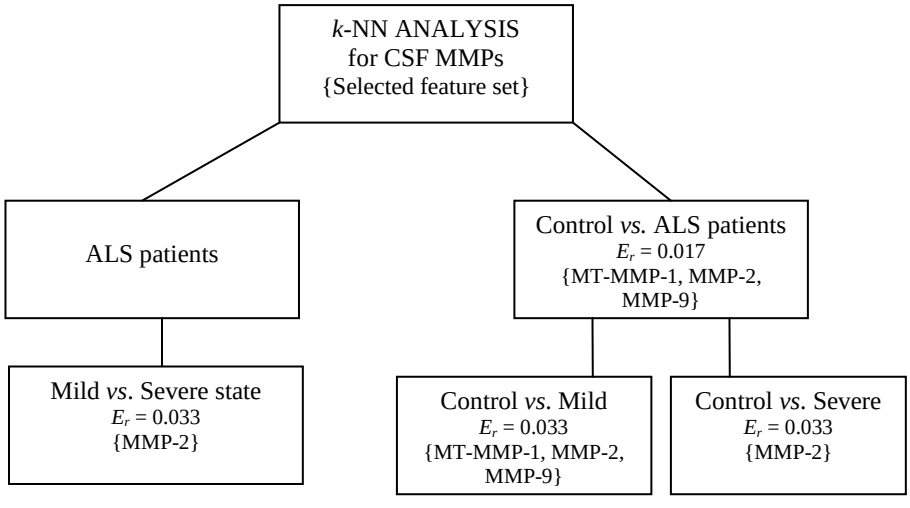

Fig. 2. Results of k-NN analysis for CSF MMPs of ALS disease

Let us now find the contents of the class areas defined in the section 2.2. and their intersections. Furthermore, let $O_1=A_1-(A_2\cup A_3)$, $O_2=A_2-(A_1\cup A_3)$, $O_3=A_3-(A_1\cup A_2)$, $O_{12}=A_1\cap A_2-A_3$, $O_{13}=A_1\cap A_3-A_2$, $O_{23}=A_2\cap A_3-A_1$ and $O_{123}=A_1\cap A_2\cap A_3$. Thus, all these defined areas are disjunctive. We can see (Tab. 2) that 5 cases from the class I lie only in the area of the class I (area O_1), next 6 cases in the area O_{13} and 4 cases in the intersection O_{123} of all the three classes. So, 5 out of 15 cases can be assigned to the class I in the first stage, because all cases from the area O_1 are assigned to the class I. All cases from the class II belong to the area O_{13} or to the intersection O_{123} of the all three classes. For this reason, the cases from this class will be classified by second classifier stage based on the *k*-NN rule. The most convenient situation appears for the class III because 14 out of 15 cases lie only in the area of the class III, i.e. in O_3. So, only one out of 15 cases from the area O_3 will be send to the second stage of the classifier and 14 will be classified in the first stage. Jointly 19 out of 45 cases (5 from the class I and 14 from the class III – the extreme classes) can be classified in the first stage. Thus, this two-stage classifier allows to classify 42.2% (19/45) cases with high certainty by the first stage of the two-stage classifier. The remaining 57.8% of cases must be classified by the standard k-NN classifier, i.e. by the second stage. The total error rate for the proposed two-stage classifier is exactly the same as for the single *k*-NN rule. However, it has significant advantage, because 42.2% of cases will be assigned with a high safety – the ones that fall in the area O_1 or in the area O_3.

Table 2. Contents of the class areas and their intersections

DISJUNCTIVE AREAS	O1	O2	O3	O12	O13	O23	O123
CLASS I (CONTROL)	5	0	0	0	6	0	4
CLASS II (MILD)	0	0	0	0	0	2	13
CLASS III (SEVERE)	0	0	14	1	0	0	0

In summary, concentrations of some matrix metalloproteineses (as MT-MMP-1, MMP-2, MMP-9) in cerebrospinal fluid of patients with amyotrophic lateral sclerosis were studied in order to recognize their relation to the severity of the disease course. Analysis based on using *k*-NN classifier enabled to detect the difference between the mild and severe ALS patients with the misclassification rate $E_r=0.033$. Global error rate of distinguish between healthy and all ALS patients is 0.017 according the feature set of CSF MMPs. In our opinion they can be used a prognostic factor and potential marker for monitoring ALS treatment.

BIBLIOGRAPHY

[1] BEUCHE W, YUSHCHENKO M, MADER M, MALISZEWSKA M, FELGENHAUER K, WEBER F. Matrix metalloproteinase-9 is elevated in serum of patients with amyotrophic lateral sclerosis. *NeuroReport* 2000; 11: 3419-3422.

[2] BROOKS BR, MILLER RG, SWASH M, MUNSAT TL. World Federation of Neurology Research Group on Motor Neuron Disease. El Escorial revisited: revised criteria for the diagnosis of amyotrophic lateral sclerosis, *Amyotrophic Lateral Sclerosis and other Motor Neuron Disorder*, 2000, 1: 293-299.

[3] CUDKOWICZ M, QURESHI M, SHEFNER J. Measures and markers in amyotrophic lateral sclerosis. *NeuroRx* 2004: 1: 273-283.

[4] DEVIJVER PA, KITTLER J. Pattern Recognition: A Statistical Approach. Prentice Hall, London, 1982.

[5] DUDA RO, HART PE, STOCK DG. Pattern Classification. John Wiley and Sons, New York, 2001.

[6] FIX E, HODGES JL. 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.

[7] JÓŹWIK A, SERPICO S and ROLI F. A parallel network of modified 1-NN and k-NN classifiers – application to remote-sensing image classification, Patter Recognition Letters 19, pp. 57-62, 1998

[8] JÓŹWIK A, SOKOŁOWSKA B, NIEBROJ-DOBOSZ I, JANIK P, KWIECIŃSKI H. Pattern recognition approach to differentiation of disease severity in patients with amyotrophic lateral sclerosis, *Journal of Medical Informatics & Technologies*, 2008; 12: 143-147.

[9] MURPHY G, NAGASE H. Progress in matrix metalloproteinase research. *Molecular Aspects of Medicine* 2008; 29: 290-308.

[10] NAGASE H, VISSE R, MURPHY G. Structure and function of matrix metalloproteinases and TIMPs. *Cardiovascular Research* 2006; 69: 562-573.

[11] NIEBROJ-DOBOSZ I, JANIK P, SOKOLOWSKA B, KWIECINSKI H. Matrix metalloproteinases and their inhibitors in serum and cerebrospinal fluid of patients with amyotrophic lateral sclerosis (ALS) (*in press*).

[12] ROSENBERG GA. Matrix metalloproteinases and their multiple roles in neurodegenerative diseases. *The Lancet Neurology* 2009; 8: 205-216.

[13] ROWLAND LP, SHNEIDER N. Amyotrophic lateral sclerosis. *New England Journal of Medicine* 2001; 344 (22): 1688-1700.

[14] YONG VW, FORSYTH PA, BELL R, KREKOSKI CA, EDWARDS DR. Matrix metalloproteinase and diseases of the CNS. *Trends in Neurosciences* 1998; 21: 75-80.

[15] YONG VW, POWER C, FORSYTH P, EDWARDS DR. Metalloproteinases in biology and pathology of the nervous system. *Nature Reviews Neuroscience* 2001; 2: 502-511.