

*Pattern recognition, feature selection, pair-wise linear classifier,  
metalloproteinases and their tissue inhibitors,  
Emery-Dreifuss muscular dystrophy*

Beata SOKOŁOWSKA<sup>1</sup>, Adam JÓŹWIK<sup>2,3</sup>, Irena M. NIEBROJ-DOBOSZ<sup>1</sup>, Irena HAUSMANOWA-PETRUSEWICZ<sup>1</sup>

## **A PATTERN RECOGNITION APPROACH TO EMERY-DREIFUSS MUSCULAR DYSTROPHY (EDMD) STUDY**

The algorithms of pattern recognition were used for differentiation between two forms of Emery-Dreifuss muscular dystrophy (EDMD), i.e. autosomal-dominant laminopathy (AD-EDMD) and X-linked emerynopathy (X-EDMD). A set of some matrix metalloproteinases (MMPs) and their tissue inhibitors (TIMPs) in serum of EDMD patients and healthy subjects were treated as features. In concluding MMPs and TIMPs levels are helpful to identifying the EDMD patients and the disease progress.

### 1. INTRODUCTION

The deficits of lamins A and C or an emerin in skeletal muscle and heart contribute to the emergence of rare, genetically transmitted disease called Emery-Dreifuss muscular dystrophy (EDMD). The genes encode proteins associated with the nuclear envelope (i) of lamins A/C (*LMNA*) that cause an autosomal-dominant EDMD form, i.e. AD-EDMD, and (ii) of emerin transmitted as X-linked trait and causes the other EDMD form, i.e. X-EDMD [3,6]. The cell defect is generalized, but skeletal muscles, heart and joints are selectively affected. The clinical symptoms of EDMD are manifested as skeletal muscle atrophy and weakness, joint contractures and dilated cardiomyopathy (DCM) with conduction disturbances. The identification of cardiac involvement in EDMD is of special interest because of a considerable mortality and a risk of sudden death. The pathogenesis of dilated cardiomyopathy in EDMD is not recognized yet. It is supposed that changes in the myocardial extracellular matrix (ECM) are important for ventricular stability and alignment of cardiomyocytes. They are responsible also for induction of matrix metalloproteinases (MMPs) and a decrease in levels of tissue inhibitors (TIMPs), seeming to evoke dilated cardiomyopathy [1,8]. In EDMD changed values of MMPs are observed and are of value for detecting cardiac involvement [9].

The aim of the present study was to estimate the contribution of MMPs, such as membrane type matrix metalloproteinase-1 (MT-MMP-1), gelatinases A (MMP-2) and B (MMP-9), and their TIMPs (TIMP-1-3) to differentiate of both EDMD forms. The usefulness of them as

---

<sup>1</sup> Mossakowski Medical Research Center, Polish Academy of Sciences, Warsaw, Poland

<sup>2</sup> Institute of Biocybernetics and Biomedical Engineering, Polish Academy of Sciences, Warsaw, Poland

<sup>3</sup> Lodz University, Faculty of Physics and Applied Informatics, Lodz, Poland

regards the diagnosis of EDMD patients is to be established. The pattern recognition approach was not applied for estimation of interaction between MMPs and TIMPs in EDMD studies, yet.

## 2. MATERIALS AND METHODS

### 2.1. SUBJECTS AND LABORATORY MEASUREMENTS

A total of 25 patients with Emery-Dreifuss muscular dystrophy (EDMD) were included. Ten patients had autosomal-dominant EDMD associated with laminopathy A/C, i.e. AD-EDMD form. Fifteen patients had X-linked EDMD associated with emerinopathy, i.e. X-EDMD form. The diagnosis of AD- and X-EDMD was established clinically and confirmed genetically by reference to the genetic defect in *LMNA* and *EMD* gene, respectively. The control group included fifteen aged-matched healthy subjects. In serum metalloproteinases MMP-9, MMP-2, MT1-MMP and their tissue inhibitors TIMP-1, TIMP-2, TIMP-3 were quantified using ELISA sandwich immunoassay procedures [9,10]. The set of MMPs and TIMPs was used as a feature set, and examined groups were defined as classes (see next section, Table 1).

The clinical study protocol was approved by the Bioethics Committee of the Medical University of Warsaw. All patients received information about the study and gave written personal consent for participation.

### 2.2. PATTERN RECOGNITION APPROACH

The clinical data is analyzed by application of the pattern recognition approach. The analysis will consist in construction a classifier and evaluation of a misclassification rate. The mentioned above data are used as a training set. Three classes and 6 features are shown in Table 1.

Table 1. Description of classes and features.

CLASSES	FEATURES
I: AD-EDMD	Metalloproteinases: 1: MMP-9 2: MMP-2 3: MT1-MMP
II: X- EDMD	
III: Controls	Their tissue inhibitors: 4: TIMP-1 5: TIMP-2 6: TIMP-3

## 3. RESULTS AND DISCUSSION

Before applying more sophisticated pattern recognition method it is verified whether any single feature does not allow to separate all 3 considered classes or at least some class pairs. The minimum and maximum feature values inside each class are determined (Table 2). It is to be noticed that none out of 6 considered features is sufficient to differentiate all 3 classes.

Some of these features, used separately, allow to distinguish some class pairs. The classes I (AD-EDMD) and II (X-EDMD) can be perfectly distinguished as by the feature 3 (MT1-MMP) as well by the feature 4 (TIMP-1). The feature 3 enables the ideal separation of the classes I (AD-EDMD) and III (Controls). The classes I and III may be perfectly separated

Table 2. Ranges of the feature values separately for each class.

CLASSES	1. MMP-9 [ng/ml]	2. MMP-2 [ng/ml]	3. MT1-MMP [ng/ml]	4. TIMP-1 [ng/m]	5. TIMP-2 [ng/ml]	6. TIMP-3 [pg/ml]
I: AD-EDMD	[138÷260]	[118÷189]	[3.8÷6.8]	[28÷54]	[10÷28]	[9.6÷12.6]
II: X-EDMD	[150÷244]	[103÷180]	[0.1÷2.9]	[64÷136]	[8÷25]	[9.4÷13.9]
III: Controls	[143÷222]	[100÷114]	[1.3÷3.5]	[28÷67]	[15÷56]	[12.6÷31.0]

also using the single feature 2 (MMP-2). The gap between the classes I and II is larger than between the classes I and III if the single feature 3 is used. The comparison of the gap between classes I and II offered by the feature 3 versus the gap observed for the feature 4 cannot be done directly. Before any comparing both these features ought to be standardized. The similar situation appears also in the case of comparing the gaps between the classes I and III for the feature 3 versus the feature 2. The described above gap values after feature standardization are presented in Table 3.

Table 3. Gaps for class pairs after standardization of features.

CLASS PAIR	2. MMP-2	3. MT1-MMP	4. TIMP-1
I vs II	—	0.522	0.289
I vs III	0.161	0.174	—

In the case of differentiation between the classes I and II as well as between the classes I and III, the feature 3 offer larger gaps than the feature 4 and the feature 2 respectively. Therefore the feature 3 will be chosen for distinguishing the class I from the class II and also for differentiation of the classes I and III.

The main goal of the present approach, as it was already stated, consists in construction of the classifier operating for all 3 considered classes. Such a classifier can be composed with component classifiers, each of them would decide between two classes only. The structure of such classifier is presented in Fig. 1, where  $\underline{x}$  is a feature vector of the classified object.

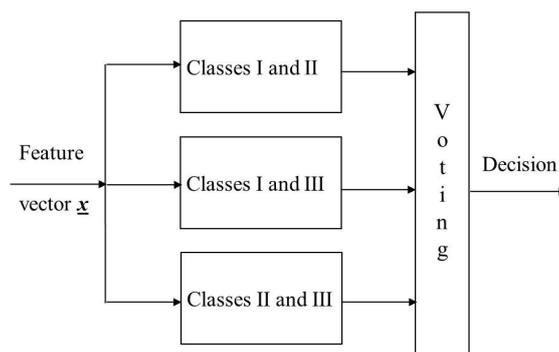


Fig. 1. The structure of the proposed classifier.

The component classifiers as for the class pair I and II as well as for the class pair I and III can be defined on basis of intervals given for the feature 3 in Table 2. The value  $3.35 = (2.9 + 3.8)/2$  of the feature 3 separates in the best way the classes I and II. The classes I and III separates in the best manner the value  $3.65 = (3.5 + 3.8)/2$  of this feature. Thus, the discriminant functions  $g_{12}(\underline{x}) = x_3 - 3.35$  and  $g_{13}(\underline{x}) = x_3 - 3.65$  define the component classifiers. The former for the classes I and II and the latter one for the classes I and III.

More difficult task concerns construction of the component classifier for the class pair II and III since none out of available features used alone differentiates these classes. It is worth to check whether the subsets  $X_2$  and  $X_3$ , representing in the training set the classes II and III respectively, are or not linearly separable. For this aim the iterative algorithm [5,7] can be used. In the case of linear separability it allows to construct the optimum separating hyperplane (the distance between this hyperplane and the nearest object from the set  $X_2 \cup X_3$  is maximum). The idea of this algorithm is shown in Fig. 2.

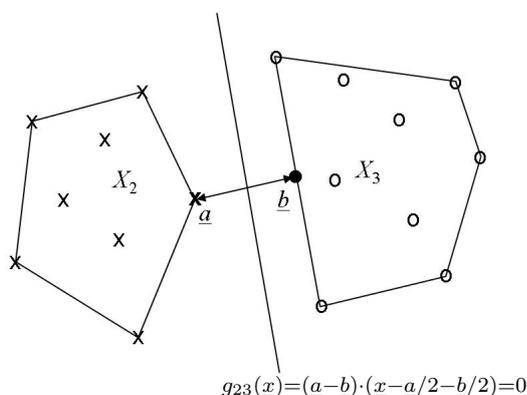


Fig. 2. The idea of the iterative algorithm for finding optimum separating hyperplane.

Two nearest points  $\underline{a}$  and  $\underline{b}$  ( $\underline{a}$  belongs to the convex hull of  $X_2$  and  $\underline{b}$  lies in the convex hull of  $X_3$ ) are searched for. If the distance between  $\underline{a}$  and  $\underline{b}$  will be greater than zero then the sets  $X_2$  and  $X_3$  can be treated as linearly separable and the relation  $g_{23}(\underline{x}) = (\underline{a}-\underline{b}) \cdot (\underline{x}-\underline{a}/2 - \underline{b}/2) = 0$  defines optimum separating hyperplane. In fact the sets  $X_2$  and  $X_3$  appeared to be linearly separable when all 6 features were used. However, one cannot exclude that some of these features are redundant. For this reason, it is reasonable to verify whether a feature combination contained smaller number of features allowing to construct the separating hyperplane can be found. Reviewing all possible combinations of two features, it appeared that there exist 3 combinations of 2 features which enable construction the separating hyperplane. They offer different distances (gaps) between the points  $\underline{a}$  and  $\underline{b}$  (Table 4).

Table 4. Gaps for class pairs after standardization of features.

FEATURE PAIR	DISTANCE (GAP)
2 and 4*	0.111
4 and 5	0.001
4 and 6*	0.122

Larger distances between  $\underline{a}$  and  $\underline{b}$  are more preferred. Therefore only 2 feature pairs, marked in Table 4 by the symbol \* will be taken into account in the further analysis. Running the iterative algorithm, for the classes II and III for both marked pairs of features, the optimum separating hyperplanes defined by the following relations can be found: the hyperplane  $g_{23}(\underline{x}) = 0.262 \cdot x_2 + 0.206 \cdot x_4 + 0.121 = 0$  if the features 2 and 4 are used and the hyperplane  $g_{23}(\underline{x}) = 0.228 \cdot x_4 - 0.265 \cdot x_6 - 0.103 = 0$  in the case of the features 4 and 6. These equations were found for the standardized features and for this reason also  $x_2$ ,  $x_4$  and  $x_6$  denote standardized feature values. To avoid feature standardization of the classified objects one can include it into weights of the functions  $g_{23}(\underline{x})$ . After this operation the separating hyperplanes will be described by  $g_{23}(\underline{x}) = 0.0099 \cdot x_2 + 0.0065 \cdot x_4 - 1.5690 = 0$  if the features 2 and 4 are used and by  $g_{23}(\underline{x}) = 0.0072 \cdot x_4 - 0.0625 \cdot x_6 + 0.2511 = 0$  if the classification is based on the features 4 and 6.

All discriminant functions, necessary to construct the pair-wise linear classifier of the type shown in Fig. 1, have been obtained. For the classes II and III two different versions of the discriminant function are available, i.e. finally two different classifiers for the three-class task were received. All obtained discriminant functions required for construction of the pair-wise linear classifiers are gathered (Table 5).

Table 5. Discriminant function for class pairs after standardization of features.

CLASS PAIR	CLASSIFIER A	CLASSIFIER B
I and II	$g_{12}(\underline{x}) = x_3 - 3.35$	$g_{12}(\underline{x}) = x_3 - 3.35$
I and III	$g_{13}(\underline{x}) = x_3 - 3.65$	$g_{13}(\underline{x}) = x_3 - 3.65$
II and III	$g_{23}(\underline{x}) = 0.009 \cdot x_2 + 0.0065 \cdot x_4 - 1.5690$	$g_{23}(\underline{x}) = 0.0072 \cdot x_4 - 0.0625 \cdot x_6 + 0.2511$

Each of the component classifier assigns the classified object  $\underline{x}$  to the class  $i$  if  $g_{ij}(\underline{x}) \geq 0$  else it is qualified to the class  $j$ . Both these constructed classifiers perfectly classify all objects from the training set. It means that hyperplanes  $g_{ij}(\underline{x}) = 0$  separate the sets  $X_i$  and  $X_j$ , represented the classes  $i$  and  $j$  in the training sets. It does not mean that the classifier offers ideal classification of the object from outside the training set.

To estimate the misclassification rate the leave one out method can be applied. It consists in classification of each object from the training set by the classifier obtained from the training set decreased by the object being currently classified. In another words, each object  $\underline{x}$  is classified by the classifier derived from the set  $X - \underline{x}$ , where  $X$  denotes the training set. While running the leave one out method it is enough to classify only those objects that are represented in the feature space by support vectors. There are two support vectors as in the case of the class pair I and II as well as in case of the classes I and III. In the case of classes II and III two features were used and generally there are at least 3 support vectors and this was the case in both classifier types determined in Table 5. Only one object from the class II was misclassified to the class III in the case of the classifier A. There was opposite situation in the case of the classifier B. This time one object from the class III was misclassified to the class II. So, the total error rates for both these classifiers were the same and equaled to 2.56%.

Although the misclassification rate obtained for the pair-wise linear classifier defined in Table 5 is small, it is worth to compare it with the pair-wise classifier based on the well known  $k$ -NN rule. The assumed numbers of nearest neighbors and the feature combinations for each class pair are presented in Table 6.

Table 6. Classifiers based on  $k$ -NN rule for class pairs after standardization of features.

CLASS PAIR	CLASSIFIER C	CLASSIFIER D
I and II	1-NN based on feature 3	1-NN based on feature 3
I and III	1-NN based on feature 3	1-NN based on feature 3
II and III	3-NN based on features 2 and 4	3-NN based on features 4 and 6

The leave one out method misclassified exactly the same objects as in the case of the pair-wise linear classifier, i.e. the classifier C misclassified the same one object from the class II to the class III as the classifier A and the classifier D misclassified exactly the same object from the class III to the class II as the classifier B. The component 1-NN classifiers for the class

pair II and III make exactly the same mistakes as the ones based on the 3-NN rule, no matter which of the classifiers C or D is used. However, using the 3-NN rule is more safely. There exists a risk that misclassified object can be an outlier and it could be chosen as the nearest neighbor whilst applying the 1-NN rule. So, this outlier could decide about the component classifier decision. The classifiers A and B seem to be more simple and can be easily realized in the spreadsheet.

Independently of the classifier type, the values of all features were standardized by subtracting their mean values and dividing the outcomes by standard deviations.

#### 4. FINAL REMARKS AND CONCLUSION

Matrix metalloproteinases (MMP), a family of proteases, are involved in the degradation of extracellular matrix proteins and determination of interstitial architecture. In the heart, MMPs play a important role in the development of myocardial remodeling and congestive heart failure. Tissue inhibitors of matrix metalloproteinases (TIMPs) represent a family of proteins which are known to regulate the expression and activity of MMPs. TIMPs are endogenous physiological inhibitors of MMPs and their concomitant down-regulation in heart failure suggests the existence of a critical balance between MMPs and TIMPs in the normal maintenance of myocardial homeostasis. Our previous study on MMPs in EDMD patients indicates that changed levels of matrix metalloproteinases may enable detection of cardiac involvement in EDMD [9]. Changes in TIMPs are associated with heart failure in EDMD [10]. In pathological condition there is an imbalance between MMPs and their TIMPs, leading in consequence to tissue destruction and degradation of EMC structural proteins. MMPs and their TIMPs play also a role in left ventricular remodeling, which is important in the progression of heart failure [2]. Deficiency of TIMP-3 is essential in promoting transition from compensated to end-stage cardiomyopathy and impaired left ventricular function, alters cardiac structure and leads to DCM [4]. Our results indicate on an altered MMP-TIMP system in Emery-Dreifuss muscular dystrophy, and giving the possibility to recognize the right diagnosis of EDMD patients.

##### In conclusion:

- 1) The performed analyses with proposed classifiers indicated that the pattern recognition approach may allow identification of patients with both form of EDMD patients and healthy subjects, using the selected feature sets of MMPs and TIMPs, especially: {MT1-MMP}, {MMP-2}, {TIMP-1}, {MMP-2; TIMP-1} or {TIMP-1; TIMP-3}.
- 2) Significant decreased of TIMP-3 in both forms of EDMD was observed in the study. The correct differentiation between AD- and X-EDMD was achieved for MT1-MMP. Additionally, X-EDMD form vs. healthy persons could be perfectly distinguished on the basis of TIMP-1.
- 3) The changes in the level of tissue inhibitors of metalloproteinases point to their significant participation in the evolution of dilated cardiomyopathy in Emery-Dreifuss muscular dystrophy.

#### ACKNOWLEDGEMENT

The authors thank to Agnieszka Madej-Pilarczyk and Michał Marchel for their consultation of EDMD patients. The study was partially supported by the Biocentrum-Ochota project [POIG.02.03.00-00-003/09].

Declaration of interest: The authors report no conflicts of interest.

BIBLIOGRAPHY

- [1] BOWERS S. L., BANERJEE I., BAUDINO T. A., The extracellular matrix: At the center of it all, *Journal of Molecular and Cellular Cardiology*, 2010, Vol. 48, pp. 474-482.
- [2] COHN N., FERRARI R., SHARPE N., Cardiac remodeling concepts and clinical implication: a consensus paper from international forum on cardiac remodeling, *Journal of the American College of Cardiology*, 2000, Vol. 35, pp. 569-582.
- [3] EMERY A. E., Emery-Dreifuss muscular dystrophy – A 40 year retrospective, *Neuromuscular Disorders*, 2000, Vol. 10 (4-5), pp. 228-232.
- [4] FEDAK P. W., SMOOKLER D. S., KASSIRI Z., OHNO N., LECO K. J., VERMA S., MICKLE D. A., WATSON K. L., HOJILLA C. V., CRUZ W., WEISEL R. D., LI R. K., Khokha, TIMP-3 deficiency leads to dilated cardiomyopathy, *Circulation*, 2004, Vol. 110, pp. 2401-2409.
- [5] FRANC V., HLAVAC V., An iterative algorithm learning the maximal margin classification, *Pattern Recognition*, 2003, Vol. 36, pp. 1985-1996.
- [6] HELBLING-LECLERC A., BONNE G., SCHWARTZ K., Emery-Dreifuss muscular dystrophy, *European Journal of Human Genetics*, 2002, Vol. 10, pp. 157-161.
- [7] KOZINIEC B. N., Recurrent algorithm separating convex hulls of two sets (V. N. Vapnik, Ed.), Soviet radio, Moscow, 1973, pp. 43-50 (in Russian).
- [8] NAGASE H., VISSE R., MURPHY G., Structure and function of matrix metalloproteinases and TIMPs, *Cardiovascular Research*, 2006, Vol. 69, pp. 562-573.
- [9] NIEBROJ-DOBOSZ I., MADEJ-PILARCZYK A., MARCHEL M., SOKOŁOWSKA B., HAUSMANOWA-PETRUSEWICZ I., *Acta Biochimica Polonica*, 2009, Vol. 56 (4), pp. 717-722.
- [10] NIEBROJ-DOBOSZ I., SOKOŁOWSKA B., MADEJ-PILARCZYK A. MARCHEL M., HAUSMANOWA-PETRUSEWICZ I., Tissue inhibitor of matrix metalloproteinases in serum as cardiac biomarkers in Emery-Dreifuss muscular dystrophy [Polish Heart Journal, in press].

