Artificial neural networks help to identify disease subsets and to predict lymphoma in primary Sjögren's syndrome

C. Baldini[1], F. Ferro[1], N. Luciano[1], S. Bombardieri[1], E. Grossi[2]

[1]*Rheumatology Unit, University of Pisa;* [2]*Villa Santa Maria Foundation, Tavernerio, Italy.*

**The results of this study were partially presented at the 14th International Sjögren's Syndrome Symposium held in Washington DC, April 18-21, 2018.*

Chiara Baldini, MD, PhD
Francesco Ferro, MD
Nicoletta Luciano, MD
Stefano Bombardieri, MD
Enzo Grossi, MD

Please address correspondence to:
Dr Chiara Baldini,
Rheumatology Unit,
University of Pisa,
via Roma 67,
56126, Pisa, Italy.
E-mail: chiara.baldini74@gmail.com

Received on June 20, 2018; accepted in revised form on July 5, 2018.

Clin Exp Rheumatol 2018; 36 (Suppl. 112): S137-S144.

Key words: Sjögren's syndrome, artificial neural networks, lymphoma

Competing interests: none declared.

ABSTRACT

Objective. *Primary Sjögren's syndrome (pSS) is a complex chronic systemic disorder, for which specific and effective therapeutic interventions are still lacking. In this era of precision medicine, there is a clear need for a better definition of disease phenotypes to foster the research of novel specific biomarkers and new therapeutic targets.*

The main objectives of this work are: 1) to compare Auto Contractive Map (AutoCM), a data mining tool based on an artificial neural network (ANN) versus conventional Principal Component Analysis (PCA) in discriminating different pSS subsets and 2) to specifically focus on variables predictive of MALT-NHL development, assessing the previsional gain of the predictive models developed.

Methods. *Out of a historic cohort of 850 patients, we selected 542 cases of pSS fulfilling the AECG criteria 2002. Thirty-seven variables were analysed including: patient demographics, glandular symptoms, systemic features, biological abnormalities and MALT-NHLs. AutoCM was used to compute the association of strength of each variable with all other variables in the dataset. PCA was applied to the same data set.*

Results. *Both PCA and AutoCM confirmed the associations between autoantibody positivity and several pSS clinical manifestations, highlighting the importance of serological biomarkers in pSS phenotyping. However, AutoCM allowed us to clearly distinguish pSS patients presenting with predominant glandular manifestations and no or mild extra-glandular features from those with a more severe clinical presentation. Out of 542 patients, we had 27 cases of MALT-NHLs. The AutoCM highlighted that, besides other traditional lymphoproliferative risk factors (i.e. salivary gland enlargement, low C4, leukocytopenia, cryoglobulins, monoclonal gammopathy, disease duration), rheumatoid factor was strongly associated to MALT-NHLs development. By applying data mining analysis, we obtained a predictive model characterised by a sensitivity of 92.5% and a specificity of 98%. If we restricted the analysis to the seven most significant variables, the sensitivity of the model was 96.2% and its specificity 96%.*

Conclusion. *Our study has shed new light on the possibility of using novel tools to extract hidden, previously unknown and potentially useful information in complex diseases like pSS, facing the challenge of disease phenotyping as a prerequisite for discovering novel specific biomarkers and new therapeutic targets.*

Introduction

Primary Sjögren's syndrome (pSS) is a complex and chronic disorder that typically involves salivary and lachrymal glands leading to their chronic inflammation and dysfunction (1, 2). The disease may also affect any other organ and system, ultimately causing disability and impairment of quality of life (3). In a minority of cases lymphoproliferative complications may also occur with mucosa-associated lymphoid tissue (MALT) lymphoma of salivary glands being the most frequent haematological type of non-Hodgkin's lymphoma (NHL) detected (4). The heterogeneous clinical presentation of the disease is paralleled by a significant diversity in pSS serological manifestations and histopathology patterns (5, 6).

In this scenario, the current "one-size fits-all" medical approach based on a combination of symptomatic drugs and, eventually, steroid or other conventional immunosuppressants, leaves open several unmet needs. Unsatisfactory control of dryness and progression of glandular dysfunction, systemic damage accrual and prevention of NHL development

are still some of the most relevant key issues in pSS treatment (7, 8).

To overcome these limitations, there is a clear need of a better definition of disease phenotypes in order to move from phenotypes to endotypes and foster the research of novel specific biomarkers and new therapeutic targets.

Traditional statistical approaches have suggested some important associations among clinical and serological variables in pSS (9, 10); however, common algorithms of linear projections of variables, including principal component analysis (PCA), require a Gaussian distribution of data and seem to have limited power when the relationships between variables are non-linear.

In this era of precision medicine, based on patient individual clinical and serological features, behaviour, habits, comorbidities and genetic background and a better stratification of disease subsets is mandatory as prerequisite for opening new avenues towards individualised treatment in heterogeneous diseases such as pSS.

In this paper, therefore, to approach this complex situation we compared conventional statistical analysis (Principal Component Analysis) with a new methodology based on an Artificial Neural Network (ANN) architecture, the Auto Contractive Map (AutoCM) (11-13). This method of data mining is a new analytical process able to create a semantic connectivity map in which non-linear associations are preserved, connections schemes are explicated and the complex dynamics of adaptive interactions is captured. The AutoCM approach has been applied in recent years to the analysis of a growing number of different clinical diseases, demonstrating its value in identifying significant associations between clinical, serological and novel "omics" biomarkers (14-19).

In this study, by using AutoCM analysis we specifically aimed at concomitantly exploring simultaneous pathways, hidden trends or non-linear associations among clinical and serological pSS disease features, in order to generate prototypical variable profiles able to discriminate between different disease phenotypes. Moreover, we specifically focused on variables predictive of the risk of salivary MALT-NHL, assessing the previsional gain of the predictive models developed.

Patients and methods

This study included patients with pSS from a large historic single centre cohort followed at the Rheumatology Unit of the University of Pisa, Italy. To be included, patients should have a diagnosis of pSS made according to the AECG criteria 2002 (20).

Thirty-seven variables were analysed including: patients demographics, glandular symptoms, systemic features, MALT-NHLs and biological abnormalities; the latter included: haematological manifestations, C3/C4 levels, hypergammaglobulinaemia and disease-related autoantibodies (*i.e.* antinuclear antibodies, anti-Ro/SSA, anti-La/SSB, rheumatoid factor and cryoglobulins). In the analysis fibromyalgia and autoimmune thyroiditis were considered as comorbidities.

The study received the local ethics committee approval and was conducted according to the Declaration of Helsinki.

Statistical analysis

AutoCM, a data mining tool based on an artificial neural network (ANN) was used to compute the association of strength of each variable with all other variables in the dataset (*i.e.* in terms of many-to-many rather than dyadic associations). The architecture and mathematics of AutoCM are described elsewhere (11). Association strength across all variables were visualised by the concept of "closeness". In other words, variables whose connection weights were higher get relatively nearer and *vice versa*. By applying a minimum spanning tree to the matrix of distances, a graph, named "semantic connectivity map (SCM)," was generated, allowing a visual mapping of the complex web of connection schemes among variables.

A principal component analysis (PCA) of 1st-2nd component was applied to the same data set to compare AutoCM with a conventional statistical approach. Results of PCA represent a benchmark for the adaptive data mining based on ANN. After analysis of the variance explained by all the components, the first two components which explain maximum amount of variance (20%) were selected.

MALT-NHL prediction model

We used a pre-processing step based on traditional descriptive statistics to select variables differentiating the 27 patients who developed MALT-NHL from the 515 patients who did not. Two subsets were created, the first with 15 variables with p-value ranging from 0.1 to <0.01 and the second with 7 variables, all with p-value <0.01. Being the two subset highly imbalanced in numerical terms, in order to avoid a polarisation of the model toward non MALT-NHL patients, we extract randomly 100 non MALT – NHL to be used with 27 MALT-NHL in the modelling. AutoCM system was applied of the transposed matrix having records treated as variables. The ratio of cumulative number of correct links over incorrect links among records belonging to the MALT-NHL sub set was taken as sensitivity, and the ratio of cumulative number of correct links over incorrect links among records belonging to the Non MALT-NHL sub set was taken as specificity.

Results

The historic pSS cohort of the University of Pisa includes 850 patients followed since 1980s. Out of them we selected 542 cases to avoid missing data. Table I summarises the demographic and serological data of the patients included in the study.

Nonlinear and simultaneous pathways between clinical manifestations, serological abnormalities and comorbidities in pSS

Figure 1 shows a "semantic connectivity map" (SCM) performed by AutoCM analysis for all the variables. The SCM describes some significant correlations between pSS clinical manifestations, serological abnormalities, patients' demographics, and comorbidities providing insights into disease different phenotypes.

First of all, the SCM confirmed the associations between auto-antibody positivity and several pSS clinical manifes-

Table I. Patients' demographics, clinical and serological features.

Variables	
Sex	524:18
Age (mean±SD), yrs	59 ± 14
Disease duration (mean±SD), yrs	7 ± 8
Xerostomia	517/542 (95.4%)
Xerophtalmia	522/542 (96.3%)
Salivary gland enlargement (SGE)	154/542 (28.4%)
Ocular tests positive	503/542 (92.8%)
Arthralgias	271/542 (50%)
Purpura	42/542 (7.7%)
Raynaud's phenomenon	161/542 (29.7%)
Lymph nodes	125/542 (23.1%)
Glomerulonephritis	14/542 (2.6%)
Interstitial lung disease	64/542 (11.8%)
Peripheral nervous system	22/542 (4.1%)
Central nervous system	15/542 (2.8%)
MALT NHLs	27/542 (5%)
Fibromyalgia	271/542 (50%)
Fatigue	383/542 (70.7%)
Thyroiditis	163/542 (30.1%)
Low C3	99/542 (18.3%)
Low C4	65/542 (12%)
Leukocytopenia	121/542 (22.3%)
Hypergammaglobulinaemia	254/452 (46.9%)
Clonal gammopathy	44/542 (8.1%)
Antinuclear antibodies	497/542 (92%)
Anti-Ro/SSA	372/542 (68.6%)
Anti-La/SSB	158/542 (29.2%)
Rheumatoid factor	261/542 (48.2%)
Cryoglobulins	15/542 (2.8%)

tations highlighting the importance of serological biomarkers in pSS subtyping.

From this perspective cryoglobulins were strongly associated with pSS vasculitic manifestations (*i.e.* purpura) whereas anti-Ro/SSA positivity was more strongly connected to pSS non-vasculitic extraglandular manifestations.

The two hubs of the SCM were represented by the anti-Ro/SSA antibodies and rheumatoid factor (RF), respectively. In fact, anti-Ro/SSA, RF, antinuclear antibodies (ANA), anti-Ro/SSB and hypergammaglobulinaemia were clearly significantly inter-connected among them.

Anti-Ro/SSA antibodies acted as hub on two main branches. The first one connected anti-Ro/SSA to ANA, xerophtalmia and xerostomia. This branch well distinguished pSS patients presenting with predominant glandular manifestations and no or mild extraglandular features from those with a more "systemic" clinical presentation. Xerophtalmia was in turn connected with the ocular tests whereas xerostomia was linked to the variable "patients' age older than 60 years", highlighting the relationship between the dryness of the mouth and patients' aging. Moreover, ANA positivity was also associated with the axis fatigue-fibromyalgia with the latter strongly inter-connected one to each other, suggesting a pathogenetic link between them.

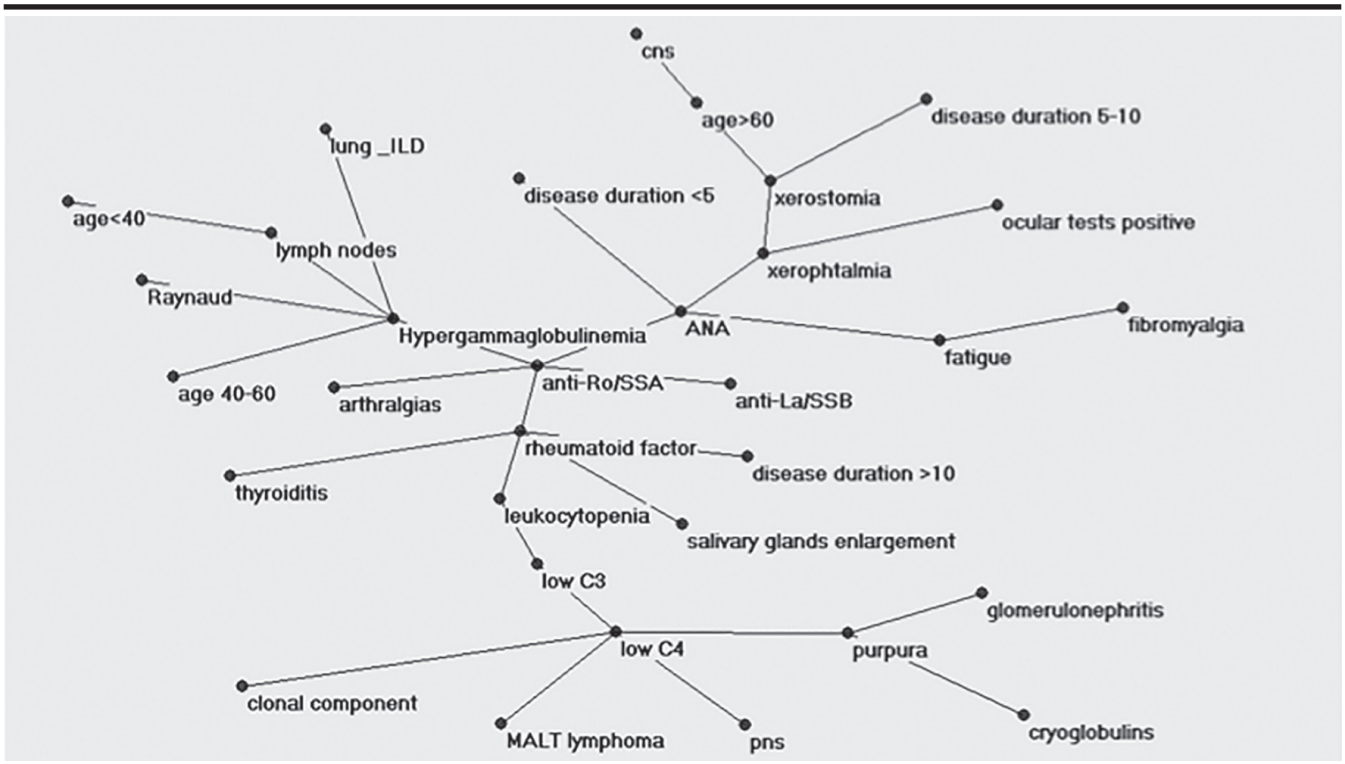

Fig. 1. Semantic Connectivity Map of variables on study with Auto-CM neural network.

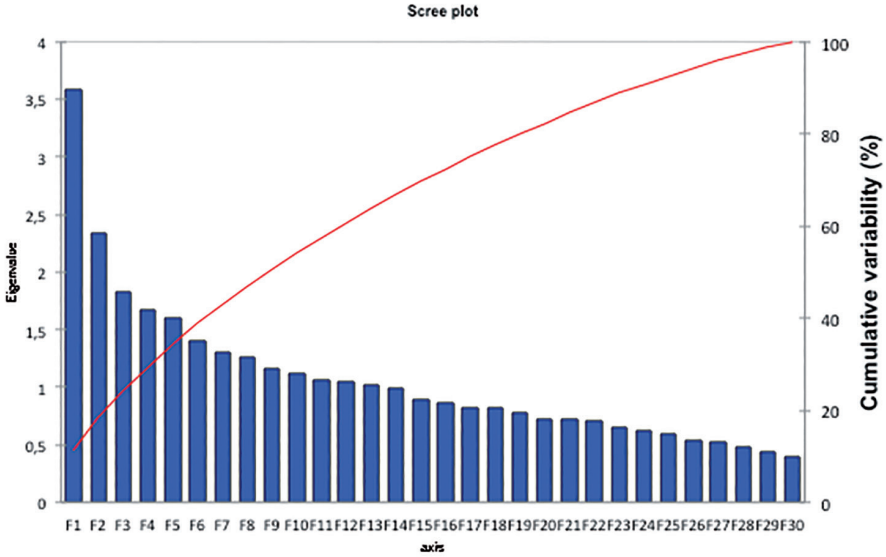

Fig. 2. Variance explained by different factors of Principal component analysis (PCA).

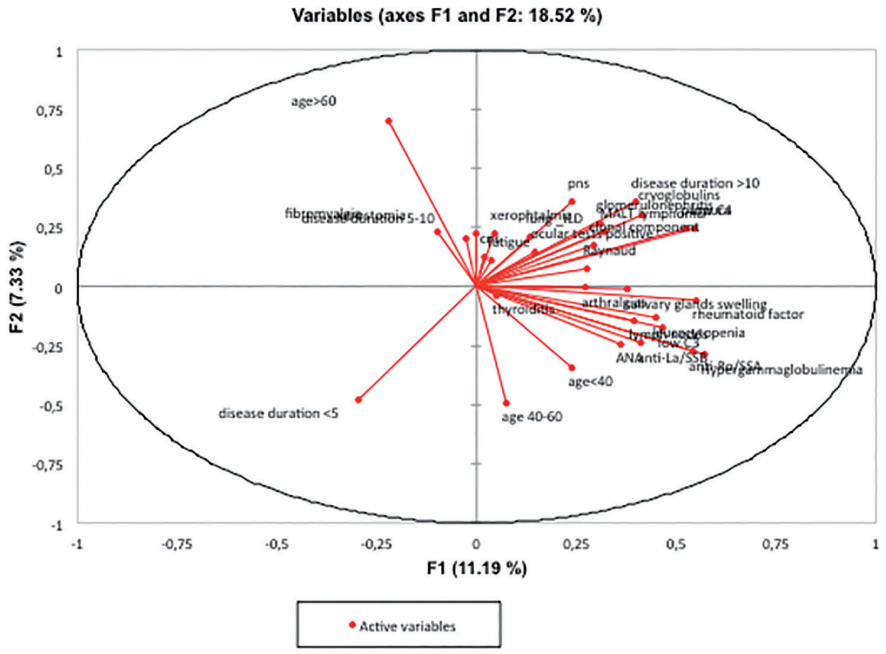

Fig. 3. Principal component analysis (PCA): first and second component.

By contrast, the second branch connected anti-Ro/SSA to hypergammaglobulinaemia and to a diverse spectrum of extra-glandular manifestations that included: lymphadenopathy, Raynaud phenomenon and lung interstitial disease. Noteworthy, patients included in this subgroup were apparently younger: in particular, the variable "patients' age younger than 40 years" was strongly associated to lymphadenopathy .

When compared to anti-Ro/SSA, RF acted as a hub for a diverse subset of pSS characterised by salivary gland enlargement, leukocytopenia, low C3/ C4, purpura, cryoglobulins vascultis manifestations (*i.e.* peripheral neuropathy, glomerulonephritis) and MALT-NHL. This branch, therefore, clearly distinguished pSS patients with vasculitic manifestations at higher risk for lymphoproliferative complications from those with non-vasculitic extra-glandular features.

MALT-NHL was placed at the end of a direct path starting with RF and passing through leukocytopenia low C3 and low C4. Other traditional lymphoproliferative risk factors like salivary gland enlargement, cryoglobulins, purpura and monoclonal gammopathy were close but remained in the background. Finally, lymphoproliferative complications and vasculitic manifestations appeared also close to the variable "disease duration longer than ten years" reinforcing the relationship between MALT-NHL and "time factor".

PCA analysis

The results of PCA represent a benchmark for the adaptive data mining based on ANN (Fig. 2, 3, 4). Figure 3 displays the first two components which explain only the 20% of the variance. When compared to ANN, the relationship between serological and clinical manifestations of pSS appears more difficult to understand in PCA. In addition, the vast majority of pSS manifestations are grouped together making it difficult to read the labels of single variables. Serological features and pSS biohumoral abnormalities still correlate with each other; however, PCA do not distinguish different disease subgroups particularly, patients with more prominent glandular manifestations from those presenting systemic features. In fact, MALT-NHL correlates with pSS-vasculitic manifestations, cryoglobulins low C4 and a longer disease duration. However, the spatial closeness between salivary MALT-NHL, RF and salivary gland swelling cannot be drawn by PCA. Finally, the closeness between fibromyalgia and fatigue highlighted by the AutoCM cannot be inferred from the PCA as well.

MALT-NHL prediction model

Focusing on MALT-NHL, out of 542 patients, we had 27 MALT-NHLs. Parotid MALT-NHLs were detected in 14 out of 27 patients. Seven other patients presented non-salivary MALT NHLs, mainly gastric lymphomas, but also splenic, brest, lung and hard palate NHLs. Six patients had a diffuse large B cell lymphoma. Table II summarises the more important 15 clinico-serological variables that differentiate patients who developed MALT-NHLs and patients who did not according to conventional statistics. Among them,

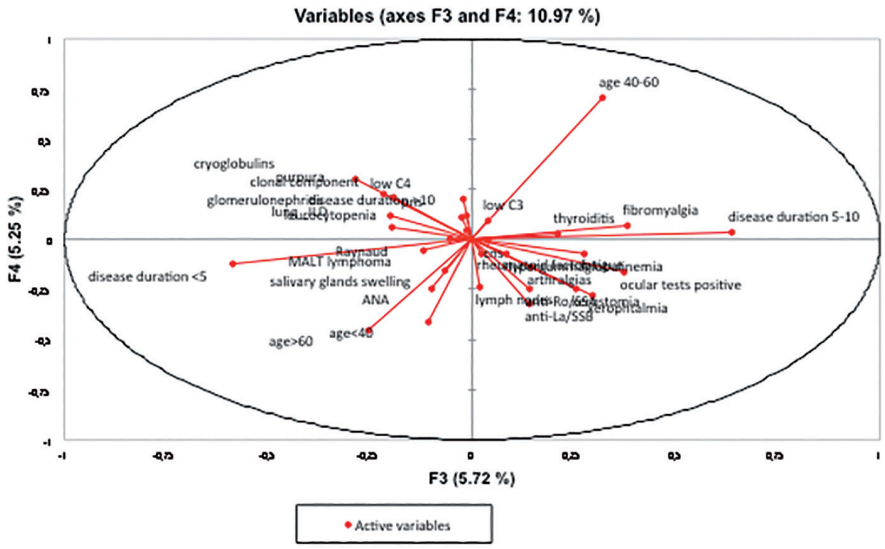

Fig. 4. Principal component analysis: third and fourth component.

Table II. Prevalence of demographic and clinic-serological features in patients with and without NHLs.

Variable	pSS without NHL (515)			pSS with NHL (27)			*p*-value
	prevalence	lower CI 95%	upper CI 95%	prevalence	lower CI 95%	upper CI 95%	
Age >60	44.9	40.6	49.1	22.2	6.5	37.9	<0.01
Disease duration <5	48.3	44	52.7	25.9	9.4	42.5	<0.01
SGE	26.0	22.2	29.8	74.1	57.5	90.6	<0.01
Purpura	7.0	4.8	9.2	22.2	6.5	37.9	ns
Lymphadenopathy	21.6	18.0	25.1	51.9	33.0	70.7	<0.01
Kidney	2.1	0.9	3.4	11.1	0	23.0	ns
Lung	11.1	8.4	13.8	25.9	9.4	42.5	ns
pns	3.5	1.9	5.1	14.8	1.4	28.2	ns
low C4	10.5	7.8	13.1	40.7	22.2	59.3	<0.01
Leukocytopenia	21.4	17.8	24.9	40.7	22.2	59.3	ns
Hypergamma	46	41.7	50.3	63	44.7	81.2	ns
Clonal component	7.4	5.1	9.6	22.2	6.5	37.9	<0.01
RF	47	42.7	51.3	70.4	53.1	87.6	<0.01
Cryoglobulins	2.3	1.0	3.6	11.1	0	23	ns
Thyroiditis	30.9	26.9	34.9	14.8	1.4	28.2	ns

7 were statistically significant whereas 8 showed a trend but were not statistically different between the two groups. When we applied the data mining analysis by using all the 15 variables we obtained a predictive model characterised by a sensitivity of 92.5% and a specificity of 98% (Fig. 5). If we restricted the analysis to the seven most significant variables (*i.e.* disease duration, age, RF, salivary gland enlargement, low C4, clonal gammopathy, and lymphadenopathy) the sensitivity of the model was 96.2% and its specificity 96% (Fig. 6).

Discussion

In this study we compared conventional PCA to AutoCM, a novel 4th generation unsupervised ANN approach to define the association strength between demographic and clinico-serological disease manifestations in a large cohort of pSS patients.

We found that both PCA analysis and AutoCM agreed on indicating the presence of a significant association between some of the pSS autoantibodies specificities and the type of organ involvement observed. Cryoglobulins were for example, not surprisingly, strongly associated to pSS vasculitic manifestations (*i.e.* purpura) (21). Similarly, anti-Ro-SSA antibodies were associated with hypergammaglobulinaemia, lymphadenopathy and lung involvement (22). Moreover, we also confirmed some of the previous acquisition regarding the temporal appearance of pSS clinical manifestations during the natural history of the disease: we observed for example that "vasculitic" features tended to appear later whereas peri-epithelitic organ involvement (*i.e.* lung interstitial disease) might be detected earlier during the disease course (23).

However, PCA has limits in the graphical representations as shown by the different mapping obtained in the first and second versus the third and fourth component analysis. By contrast, AutoCM overcomes some of the PCA well-known limitations (*i.e.* dealing with non linear data, not immediately evident graphical representations, etc.), allowing us to easily distinguish at the first glance three different phenotypes in our cohort, that resulted from the simultaneous combination of all the clinical, demographic and biohumoral variables, centred around two main hubs: the positivity for anti-Ro/SSA antibodies and for RF.

The first subset was represented by patient presenting with predominant glandular manifestations and none/mild extra-glandular features. These patients tended to be older and to present comorbidities including fibromyalgia. The direct axis fibromyalgia/fatigue *per se* might encourage future research lines aimed at clarifying the complex relationship between these two entities (24, 25).

The second subset was represented by anti-Ro/SSA positive patients presenting a spectrum of classical symptoms that reminded to the "area" of connective tissue diseases (*i.e.* systemic lupus erythematosus) including: Raynaud phenomenon, lymphadenopathy, hypergammaglobulinaemia and interstitial lung disease. These patients were intriguingly younger with the variable "age <40 years" strictly connected to the variable "lymphadenopathy".

The third subset recognised as hub the positivity of RF and was represented by patients with "vasculitic manifestations" and MALT-NHL lymphoma. The analysis of this part of the SCM was particularly of interest. Although the variable "MALT-NHL" appeared

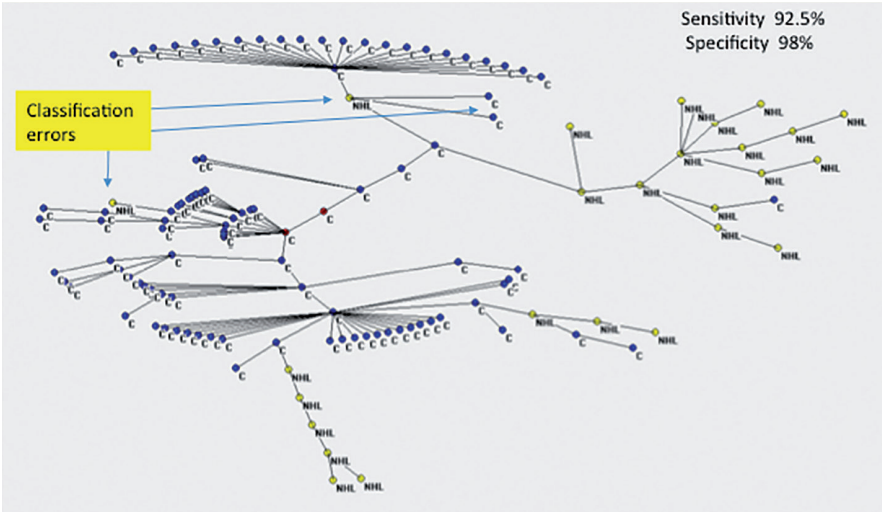

Fig. 5. Mapping of the records with Auto-CM systems using 15 discriminant variables.

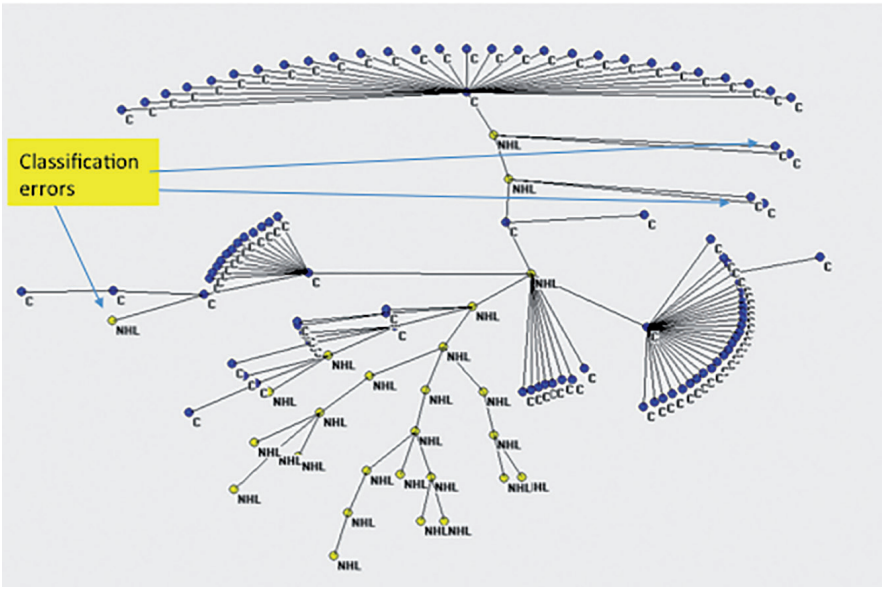

Fig. 6. Mapping of the records with Auto-CM systems using 7 discriminant variables.

spatially close to traditional risk factors such as salivary gland enlargement, cryoglobulins, clonal component and pSS-"vasculitic" manifestations (4, 26-28), nonetheless, the SCM showed a stronger association between MALT-NHLs, RF and other bioumoral abnormalities including leukocytopenia and low C4 levels. These findings did not emerge from PCA. By contrast, AutoCM placed the accent on the role of RF as a risk factor for lymphoma development, pushing into the background other traditional MALT-NHL predictors and mainly vasculitic manifestations.

Regarding the relationship between RF and parotid MALT-NHLs, our results are in line with those from Nocturne *et al*. (29) and Fragkioudaki *et al*. (30). Nocturne *et al*. (29), in their paper, highlighted the independent role of RF as a new predictive factor of lymphoma occurrence in 101 pSS patients. Noteworthy, in their series 43 out of 101 cases were represented by salivary MALT-NHLs. We confirmed the association between RF and salivary MALT-NHLs and demonstrated that RF when combined with other previously known risk factors (*i.e*. salivary gland enlargement, low C4, clonal gammopathy, lymphadenopathy) might be able to predict MALT-NHLs development with a high sensitivity and specificity. Similarly, Fragkioudaki *et al*. (30) in a case-control study of 381 pSS patients and 92 pSS patients with concomitant NHL, identified RF and salivary gland enlargement as independent predictors for NHL development together with lymphadenopathy, Raynaud phenomenon, anti-Ro/SSA or/and anti-La/SSB autoantibodies, monoclonal gammopathy, and low C4 levels.

Noteworthy, five out of the seven items proposed by Fragkioudaki *et al*. entered in our prediction model including not only RF but also salivary gland enlargement, lymphadenopathy, monoclonal gammopathy and low C4 levels. In addition, in our prediction model, we also confirmed the role of disease duration in lymphoproliferative complications, as previously shown by us and by other authors (23, 26, 28, 31). Overall, this model allowed us to predict lymphoma nearly in all cases.

To our knowledge, our study is the first one to have adopted AutoCM to face the challenge of disease phenotyping in pSS. From this perspective, pSS represents the ideal model to test the potentialities of ANNs in distinguishing different patients' subsets, given the clinical, serological and histological complexity of the disease as well as its diversity in the long-term outcomes (32). Indeed, ANN have been criticised as being "black box with limited ability to explicitly identify possible causal relationship" (33). However, they have been as well recognised to offer a number of advantages with respect to logistic regression in performing non-linear statistical modelling for systemic disorders, in general, and particularly, in autoimmunity (34, 35). Moreover, auto-CM connections matrix, filtered by a minimum spanning tree algorithm, is a 4th generation ANN that has already proven its usefulness in patients' phenotyping in several chronic heterogeneous conditions (15, 16, 18, 19)

This study was limited by several factors, including the fact that we restricted the analysis only to demographic, clinical and laboratory variables. We could not include information on the infiltrate composition of the minor salivary glands because some of the samples had been read using the Chisholm and Mason grading and not according to the

focus score (36). Moreover, none of the novel molecular or "omic biomarkers" or salivary gland ultrasonography findings entered in the analysis.

However, our results showed undoubtedly the potential of AutoCM for precision medicine when compared to conventional statistical analysis. AutoCM analysing the non-linear relationship between the variables was able in our study to put emphasis on some associations with clear clinical implications. The possibility of distinguishing different disease phenotypes with diverse outcomes might open the avenue to basic research projects as well as to novel individualised therapeutic approaches. From the perspective of clinical trials, moreover AutoCM may allow to design novel study for homogenous subgroups with better defined clinical endpoints fostering the development of novel effective drugs. Last but not least a reliable risk estimation for MALT-NHLs could offer an important tool in primary and secondary prevention as it can be used to expedite appropriate therapeutic intervention improving patients' prognosis and quality of life.

In conclusion, our study by comparing ANN analysis to PCA has shed new lights on the possibility of using novel tools to extract hidden, previously unknown and potentially useful information in a complex disease like pSS. The connections scheme evidenced by the Auto-CM system analysis suggest a novel comprehensive interpretation of pSS clinical complexity. The disappointing results obtained with alternative data mining statistical methods such as principal component analysis strengthen the idea that Auto-CM, thanks to its new sophisticated mathematics, could become, in the future, a reference approach to better understand the secrets of this disease. Proper further studies will reveal if this approach may be relevant also in actual clinical practice to improve the pertinence of medical interventions for management.

References

1. FERRO F, MARCUCCI E, ORLANDI M, BALDINI C, BARTOLONI-BOCCI E: One year in review 2017: primary Sjögren's syndrome. *Clin Exp Rheumatol* 2017; 35: 179-91.
2. BRITO-ZERON P, BALDINI C, BOOTSMA H *et al*.: Sjögren's syndrome. *Nat Rev Dis Primers* 2016; 2: 16047.
3. CORNEC D, DEVAUCHELLE-PENSEC V, MARIETTE X *et al*.: Severe health-related quality-of-life impairment in active primary Sjögren's syndrome is driven by patient-reported outcomes: data from a large therapeutic trial. *Arthritis Care Res* (Hoboken) 2017; 69: 528-35.
4. TZIOUFAS AG, VOULGARELIS M: Update on Sjögren's syndrome autoimmune epithelitis: from classification to increased neoplasias. *Best Pract Res Clin Rheumatol* 2007; 21: 989-1010.
5. GOULES AV, TZIOUFAS AG: Primary Sjögren's syndrome: clinical phenotypes, outcome and the development of biomarkers. *Immunol Res* 2017; 65: 331-44.
6. CHRISTODOULOU MI, KAPSOGEORGOU EK, MOUTSOPOULOS HM: Characteristics of the minor salivary gland infiltrates in Sjögren's syndrome. *J Autoimmun* 2010; 34: 400-7.
7. RAMOS-CASALS M, BRITO-ZERON P: New approaches in Sjögren's syndrome therapy. *Expert Rev Clin Immunol* 2007; 3: 195-204.
8. BARONE F, COLAFRANCESCO S: Sjögren's syndrome: from pathogenesis to novel therapeutic targets. *Clin Exp Rheumatol* 2016; 34 (Suppl. 98): S58-62.
9. BALDINI C, PEPE P, QUARTUCCIO L *et al*.: Primary Sjögren's syndrome as a multi-organ disease: impact of the serological profile on the clinical presentation of the disease in a large cohort of Italian patients. *Rheumatology* (Oxford) 2014; 53: 839-44.
10. MANOUSSAKIS MN, TZIOUFAS AG, PANGE PJ, MOUTSOPOULOS HM: Serological profiles in subgroups of patients with Sjögren's syndrome. *Scand J Rheumatol Suppl* 1986; 61: 89-92.
11. BUSCEMA M, GROSSI E, SNOWDON D, ANTUONO P: Auto-Contractive Maps: an artificial adaptive system for data mining. An application to Alzheimer disease. *Curr Alzheimer Res* 2008; 5: 481-98.
12. BUSCEMA M, GROSSI E: The semantic connectivity map: an adapting self-organising knowledge discovery method in data bases. Experience in gastro-oesophageal reflux disease. *Int J Data Min Bioinform* 2008; 2: 362-404.
13. GROSSI E, BUSCEMA M: Introduction to artificial neural networks. *Eur J Gastroenterol Hepatol* 2007; 19: 1046-54.
14. LICASTRO F, PORCELLINI E, CHIAPPELLI M *et al*.: Multivariable network associated with cognitive decline and dementia. *Neurobiol Aging* 2010; 31: 257-69.
15. COPPEDE F, GROSSI E, LOPOMO A, SPISNI R, BUSCEMA M, MIGLIORE L: Application of artificial neural networks to link genetic and environmental factors to DNA methylation in colorectal cancer. *Epigenomics* 2015; 7: 175-86.
16. GROSSI E, STOCCORO A, TANNORELLA P, MIGLIORE L, COPPEDE F: Artificial neural networks link one-carbon metabolism to gene-promoter methylation in Alzheimer's disease. *J Alzheimers Dis* 2016; 53: 1517-22.
17. DRAGO L, TOSCANO M, DE GRANDI R, GROSSI E, PADOVANI EM, PERONI DG: Microbiota network and mathematic microbe mutualism in colostrum and mature milk collected in two different geographic areas: Italy versus Burundi. *ISME J* 2017; 11: 875-84.
18. TOSCANO M, DE GRANDI R, PERONI DG *et al*.: Impact of delivery mode on the colostrum microbiota composition. *BMC Microbiol* 2017; 17: 205.
19. ULIVIERI FM, PIODI LP, GROSSI E *et al*.: The role of carboxy-terminal cross-linking telopeptide of type I collagen, dual x-ray absorptiometry bone strain and Romberg test in a new osteoporotic fracture risk evaluation: A proposal from an observational study. *PLoS One* 2018; 13: e0190477.
20. VITALI C, BOMBARDIERI S, JONSSON R *et al*.: Classification criteria for Sjögren's syndrome: a revised version of the European criteria proposed by the American-European Consensus Group. *Ann Rheum Dis* 2002; 61: 554-8.
21. QUARTUCCIO L, ISOLA M, BALDINI C *et al*.: Clinical and biological differences between cryoglobulinaemic and hypergammaglobulinaemic purpura in primary Sjögren's syndrome: results of a large multicentre study. *Scand J Rheumatol* 2015; 44: 36-41.
22. TZIOUFAS AG, KAPSOGEORGOU EK, MOUTSOPOULOS HM: Pathogenesis of Sjögren's syndrome: what we know and what we should learn. *J Autoimmun* 2012; 39: 4-8.
23. SKOPOULI FN, DAFNI U, IOANNIDIS JP, MOUTSOPOULOS HM: Clinical evolution, and morbidity and mortality of primary Sjögren's syndrome. *Semin Arthritis Rheum* 2000; 29: 296-304.
24. CHOI BY, OH HJ, LEE YJ, SONG YW: Prevalence and clinical impact of fibromyalgia in patients with primary Sjögren's syndrome. *Clin Exp Rheumatol* 2016; 34 (Suppl. 96): S9-13.
25. ARENDS S, MEINERS PM, MOERMAN RV *et al*.: Physical fatigue characterises patient experience of primary Sjögren's syndrome. *Clin Exp Rheumatol* 2017; 35: 255-61.
26. GIANNOULI S, VOULGARELIS M: Predicting progression to lymphoma in Sjögren's syndrome patients. *Expert Rev Clin Immunol* 2014; 10: 501-12.
27. PAPAGEORGIOU A, VOULGARELIS M, TZIOUFAS AG: Clinical picture, outcome and predictive factors of lymphoma in Sjögren's syndrome. *Autoimmun Rev* 2015; 14: 641-9.
28. BALDINI C, PEPE P, LUCIANO N *et al*.: A clinical prediction rule for lymphoma development in primary Sjögren's syndrome. *J Rheumatol* 2012; 39: 804-8.
29. NOCTURNE G, VIRONE A, NG WF *et al*.: Rheumatoid factor and disease activity are independent predictors of lymphoma in primary Sjögren's syndrome. *Arthritis Rheumatol* 2016; 68: 977-85.
30. FRAGKIOUDAKI S, MAVRAGANI CP, MOUTSOPOULOS HM: Predicting the risk for lymphoma development in Sjögren's syndrome: An easy tool for clinical use. *Medicine* (Baltimore) 2016; 95: e3766.
31. VOULGARELIS M, SKOPOULI FN: Clinical, immunologic, and molecular factors predicting lymphoma development in Sjögren's syndrome patients. *Clin Rev Allergy Immunol* 2007; 32: 265-74.
32. MANUEL RC, PILAR BZ, RAPHAELE S *et al*.:

Characterization of systemic disease in primary Sjögren's syndrome: EULAR-SS Task Force recommendations for articular, cutaneous, pulmonary and renal involvements. *Rheumatology* (Oxford) 2017; 56: 1245.

33. TU JV: Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996; 49: 1225-31.
34. OBERMEYER Z, EMANUEL EJ: Predicting the future - big data, machine learning, and clinical medicine. *N Engl J Med* 2016; 375: 1216-9.
35. ORANGE DE, AGIUS P, DICARLO EF *et al*.: Identification of three rheumatoid arthritis disease subtypes by machine learning integration of synovial histologic features and RNA sequencing data. *Arthritis Rheumatol* 2018; 70: 690-701.
36. FISHER BA, JONSSON R, DANIELS T *et al*.: Standardisation of labial salivary gland histopathology in clinical trials in primary Sjögren's syndrome. *Ann Rheum Dis* 2017; 76: 1161-8.