Assessment of Machine Learning Algorithms for the Purpose of Primary Sjögren's Syndrome Grade Classification from Segmented Ultrasonography Images

Arso Vukicevic[1,2(✉)], Alen Zabotti[3], Salvatore de Vita[3], and Nenad Filipovic[1,2]

[1] BioIRC, Bioengineering Research and Development Center, Prvoslava Stojanovica 6, 34000 Kragujevac, Serbia
arso_kg@yahoo.com, fica@kg.ac.rs
[2] Faculty of Engineering, University of Kragujevac, Sestre Janjica 6, 34000 Kragujevac, Serbia
[3] Azienda Ospedaliero Universitaria, Santa Maria Della Misericordia di Udine, Udine, Italy
zabottialen@gmail.com, devita.salvatore@aoud.sanita.fvg.it

Abstract. Primary Sjögren's syndrome (pSS) is a chronic autoimmune disease that affects primarily women (9 females/1 male). Recently, a great interest has arisen for salivary gland ultrasonography (SGUS) as a valuable tool for the assessment of major salivary gland involvement in primary Sjögren's syndrome. The aim of this study was to assess accuracy of state of the art machine learning algorithms for the purpose of classifying pSS from SGUS images. The five-step procedure was carried out, including: image pre- processing, feature extraction, data set balancing and feature extraction, classifiers (K-Nearest Neighbour, Decision trees, Naive bayes, Discriminant analysis classifier, Random forest, Multilayer perceptron, Linear logistic regression) learning and their corresponding assessment. The preliminary results on the growing HarmonicSS cohort showed that Naive bayes (72.8% accuracy on training set, and 73.3% accuracy on test set) and Multilayer perceptron (85.0% accuracy in training stage, and 70.1% accuracy at test stage) are the most suitable for the purpose of pSS grade classification.

Keywords: Sjögren's syndrome · Classification · Ultrasonography

1 Introduction

Primary Sjögren's syndrome (pSS) is a chronic autoimmune disease [1]. According to the clinical reports, the annual incidence of pSS among North and South European populations has been estimated at a range from 200 to 3000 per 100.000 individuals [2]. Moreover, among 394,827 affected individuals with systemic autoimmune diseases, pSS was found to be characterized by the most unbalanced gender ratio with almost 10 females affected per 1 male, followed by systemic lupus erythematosus,

O. Fratu et al. (Eds.): FABULOUS 2017, LNICST 241, pp. 239–245, 2018.
https://doi.org/10.1007/978-3-319-92213-3_35

systemic sclerosis and antiphospholipid syndrome (APS) (ratio of nearly 5:1) [3]. The pSS has a wide range of clinical presentations, from mild disease limited to exocrine glands to severe multi-systemic disorder, and increases the risk of developing a B-cell non-Hodgkin lymphoma, which occurs in about 5% of patients [1].

Currently, the involvement of salivary glands in pSS is assessed by means of complementary tests such as sialometry, sialoscintigraphy and sialography, in accordance with the American European Consensus Group (AECG) classification criteria. Such tests, added to biopsy of the minor salivary gland (MSGB), may provide valuable information on the anatomical and functional damage in these glands; however, their use in clinical practice is limited by its poor specificity for pSS diagnosis. Recently, a great interest has arisen for salivary gland ultrasonography (SGUS) as a valuable tool for the assessment of major salivary gland involvement in primary Sjögren's syndrome. Figure 1 shows an example of manual segmentation of SGUS, adapted from a literature [4]. The aim of this study was to develop tools for classification of pSS using segmented ultrasonography images of salivary gland. The overall workflow is given on Fig. 2 while each of the particular steps is explained in the remainder of this document.

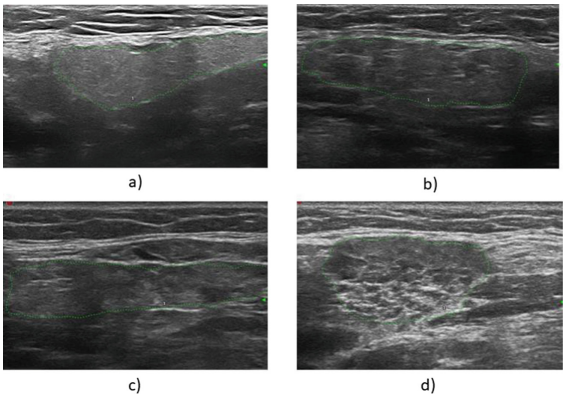

Fig. 1. Ultrasound images of pSS: (a) Grade 0, (b) Grade 1, (c) Grade 2, (d) Grade 3

2 Methods

2.1 Image Pre-processing

The purpose of image pre-processing was to reduce noises and artifact occurred during the image acquisition with a mobile ultrasonography device. Each DICOM image was pre-processed using the procedure sketched on Fig. 3. We first used Wiener filter [5]. Afterwards, we used Matlab's Image processing toolbox implementation of histogram equalizer. Finally, Salt and pepper noise on images war reduced using the two-dimensional median filter.

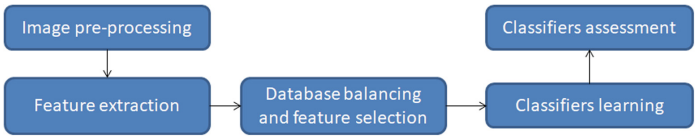

Fig. 2. Procedure workflow

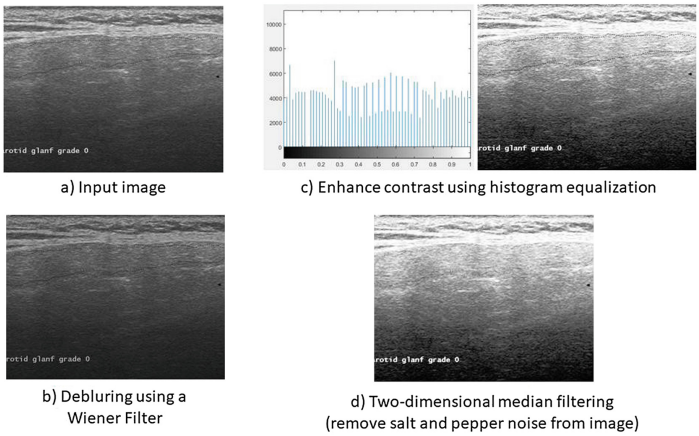

Fig. 3. Image pre-processing.

2.2 Feature Extraction

It is assumed that pSS region is manually segmented, so that only pixels inside the region of interest are further analyzed. The feature extraction represents the process of transformation of selected pixels into the descriptive values suitable for learning classifiers. Since the pSS region is of varying area and shape, we suggested using histogram-based descriptors: Local binary pattern (LBP) and Gray-level co-occurrence matrix (GLCM).

2.2.1 Local Binary Pattern

The example of processed image and the obtained LBP histogram is given on Fig. 4. The LBP feature vector was created in the following manner [6]: (1) Divide the examined window into cells (e.g. 16 x 16 pixels for each cell). (2) For each pixel in a cell, compare the pixel to each of its 8 (in general N) neighbors (on its left-top, left-middle, left-bottom, right-top, etc.). Follow the pixels along a circle (with diameter R), i.e. clockwise or counter-clockwise. (3) Where the center pixel's value is greater than the neighbour's value, write "0". Otherwise, write "1". This gives an 8-digit binary number (which is usually converted to decimal for convenience). (4) Compute the histogram, over the cell, of the frequency of each "number" occurring (i.e., each combination of which pixels are smaller and which are greater than the center). This histogram can be seen as a 256-dimensional feature vector. (5) Optionally normalize

the histogram. (6) Concatenate (normalized) histograms of all cells. This gives a feature vector for the entire window. In the present study, we used LPB of N = 8 and R = 4 presented below.

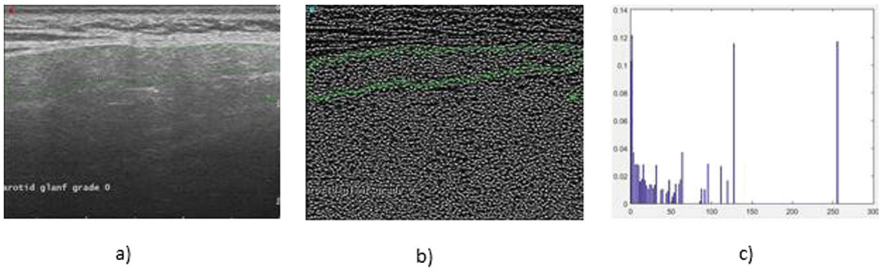

Fig. 4. Local binary pattern: (a) Original image; (b) LBP (N = 8, R = 4); (c) LBP histogram.

2.2.2 Gray-Level Co-occurrence Matrix

GLCM is a statistical method of examining texture that considers the spatial relationship of pixels. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix (Fig. 5). After creating the GLCMs, we derived several statistics from them. These statistics provide information about the texture of an image. The list of features extracted for the purpose of the present study was as it follows: Autocorrelation, Contrast, Correlation, Correlation, Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity, Homogeneity, Maximum probability, Sum of squares: Variance, Sum average, Sum variance, Sum entropy, Difference variance, Difference entropy, Information measure of correlation1, Information measure of correlation2, Inverse difference, Inverse difference normalized, Inverse difference moment normalized.

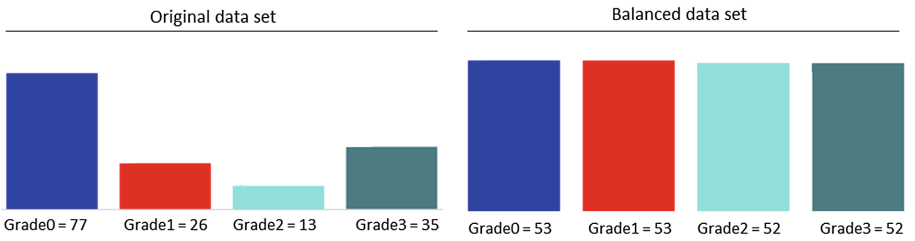

Fig. 5. Classes' distribution. Image left shows original-overall data set. Image left shows the learning data set balanced using the SMOTE algorithm.

2.3 Population Data

For the purpose of the present study we used a dataset of total 153 patients subjected for ultrasonography imaging of pSS. Distribution of each class (Grade 0–3) is shown on Fig. 5. The amount of samples for pSS Grade 4 and 5 was insufficient (2), so they were excluded. Moreover, 100 samples (∼66%) of each class were used for the independent training, while 51 samples (∼33% of each class) were used for independent subsequent testing. As it may be noted, the database had unbalanced distribution of classes – which was solved in training-stage by using the Synthetic Minority Over-sampling Technique (SMOTE) [7].

2.4 Features Selection

The feature selection if performed in order to minimize a number of features, by omitting these that are less correlated with the pSS. Using the wrapper for feature subset selection [8], out of 57 features extracted using LBP and GLCM, 21 features were elected as dominant for the learning process.

2.5 Learning Classifier

Seven different classifiers were considered: K-Nearest Neighbour (KNN), Decision Trees (DT), Naive Bayes (NB), Discriminant analysis classifier (DCA), Random Forest (RF), Multilayer Perceptron (MLP), Linear Logistic Regression Model (LR) implemented in Weka software [9]. Each of classifier was trained using 10-fold cross-validation training and data set described in Sect. 2.3.

3 Results and Discussion

The obtained results are given in Table 1. As it may be noted, the considered algorithms showed variations of the accuracy for learning-derivation data set and independent-test set. In terms of classification accuracy, three top-ranked algorithms on derivation set are: Multilayer perceptron, Random forest and K-nearest neighbour. Considering test set, three top-ranked algorithms on derivation set are: Naive Bayes, Multilayer Perceptron and Linear Logistic Regression. Assuming that, Naive bayes and Multilayer perceptron are suggested as most suitable for the purpose of pSS grade classification. It is worth to mention that the database used in this study represent result of on-going HarmonicSS H2020 EU project, which aims to integrate analysis of regional, national and international cohorts on primary Sjögren's Syndrome towards improved stratification, treatment and health policy making. Since it is well know that performances of machine learning algorithms are varying with database size scaling – presented results could be considered as preliminary indicators that could change as the cohort will growth over time.

Table 1. Classification accuracy of the considered learning algorithms.

Algorithm	Derivation set (Samples: 210)	Test set (Samples: 51)
Decision trees	75.4%	67.7%
Naive Bayes	72.8%	73.3%
Discriminant analysis	76.7%	66.8%
K-Nearest neighbour	83.2%	66.0%
Random forest	84.6%	68.2%
Multilayer perceptron	85.0%	70.1%
Linear logistic regression	82.9%	69.3%

4 Conclusion and Future Work

In the present study we performed assessment of state of the art classification algorithms for the purpose of pSS classification from ultrasound images. The assessment was performed using growing HarmonicSS (H2020 EU project which aims to integrate analysis of regional, national and international cohorts on primary Sjögren's Syndrome) data set. The preliminary results on the growing HarmonicSS cohort showed that Naive bayes (72.8% accuracy on training set, and 73.3% accuracy on test set) and Multilayer perceptron (85.0% accuracy in training stage, and 70.1% accuracy at test stage) are the most suitable for the purpose of pSS grade classification.

Acknowledgments. This study was funded by the grants from the Serbia III41007, ON174028 and EC HORIZON2020 HarmonicSS project.

References

1. Mavragani, C.P., Moutsopoulos, H.M.: Sjögren syndrome. CMAJ **186**(15), E579–E586 (2014). https://doi.org/10.1503/cmaj.122037
2. Shapira, Y., Agmon-Levin, N., Shoenfeld, Y.: Geoepidemiology of autoimmune rheumatic diseases. Nat. Rev. Rheumatol. **6**(8), 468–476 (2010). https://doi.org/10.1038/nrrheum.2010.86
3. Ramos-Casals, M., Brito-Zerón, P., Kostov, B., Sisó-Almirall, A., Bosch, X., Buss, D., Trilla, A., Stone, J.H., Khamashta, M.A., Shoenfeld, Y.: Google-driven search for big data in autoimmune geoepidemiology: analysis of 394,827 patients with systemic autoimmune diseases. Autoimmun. Rev. **14**(8), 670–679 (2015). https://doi.org/10.1016/j.autrev.2015.03.008
4. Baldini, C., Luciano, N., Tarantini, G., Pascale, R., Sernissi, F., Mosca, M., Caramella, D., Bombardieri, S.: Salivary gland ultrasonography: a highly specific tool for the early diagnosis of primary Sjögren's syndrome. Arthritis Res. Ther. **17**(1), 146 (2015). https://doi.org/10.1186/s13075-015-0657-7
5. Wiener, N.: Extrapolation, Interpolation, and Smoothing of Stationary Time Series. Wiley, New York (1949). ISBN 0-262-73005-7
6. Ojala, T., Pietikäinen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recogn. **29**, 51–59 (1996)

7. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: SMOTE: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
8. Kohavi, R., John, G.H.: Wrappers for feature subset selection. Artif. Intell. **97**(1–2), 273–324 (1997)
9. Frank, E., Hall, M.A, Witten, I.H.: The WEKA Workbench. Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", 4th edn. Morgan Kaufmann, Massachusetts (2016)