SCIENTIFIC REP⚙RTS

Texture analysis in gel electrophoresis images using an integrative kernel-based approach

Carlos Fernandez-Lozano[1], Jose A. Seoane[2,3], Marcos Gestal[1], Tom R. Gaunt[4], Julian Dorado[1], Alejandro Pazos[1,5] & Colin Campbell[6]

Texture information could be used in proteomics to improve the quality of the image analysis of proteins separated on a gel. In order to evaluate the best technique to identify relevant textures, we use several different kernel-based machine learning techniques to classify proteins in 2-DE images into spot and noise. We evaluate the classification accuracy of each of these techniques with proteins extracted from ten 2-DE images of different types of tissues and different experimental conditions. We found that the best classification model was FSMKL, a data integration method using multiple kernel learning, which achieved AUROC values above 95% while using a reduced number of features. This technique allows us to increment the interpretability of the complex combinations of textures and to weight the importance of each particular feature in the final model. In particular the *Inverse Difference Moment* exhibited the highest discriminating power. A higher value can be associated with an homogeneous structure as this feature describes the homogeneity; the larger the value, the more symmetric. The final model is performed by the combination of different groups of textural features. Here we demonstrated the feasibility of combining different groups of textures in 2-DE image analysis for spot detection.

Two-dimensional gel electrophoresis (2-DE) is the method of choice for analysing protein expression in proteomics due to its often-underestimated advantages: robustness, resolution and ability to separate entire proteins at high resolution[1].

A large number of proteins are separated in the same process according to their electrical charge and molecular mass in order to identify and characterize them. Thus, an array of dark spots on a gel (i.e. polyacrylamide, agarose) is generated (Fig. 1). The analysis of this gel is a non-trivial, tedious and time consuming task, resulting in a bottleneck in proteomics due to differences in protein expression and experimental conditions as well as appearance changes between proteins[2]. Advanced image analysis techniques could improve the performance and quality of this analysis, and one of the key stages of the analysis is the detection of protein spots and differentiation between noise and real protein (Fig. 1). Image classification usually involves computation of features (image attributes) and thus, every image is characterized by attribute vectors with thousands of dimensions. One of these image analysis techniques is texture analysis.

This technique has been widely used in many applications such as remote sensing or document segmentation, for example, but is of special relevance in biomedical imaging for characterization and quantification of different regions of interest (i.e control vs. lesions). There are multiple definition for texture[3], mainly due to the fact that it is used in a wide array of different applications. In this work textures are defined as the spatial distribution of the grey levels within an image and as a surface's property can be regarded as the almost regular spatial organization of patterns. Texture is always present even if it is a feature with a low discrimination power.

[1]Information and Communication Technologies Department, Faculty of Computer Science, University of A Coruna, A Coruna, 15071, Spain. [2]Bristol Genetic Epidemiology Laboratories, School of Social and Community Medicine, University of Bristol, Bristol BS82BN, UK. [3]Stanford Cancer Institute, Stanford School of Medicine, Stanford University, Stanford, 94305, USA. [4]MRC Integrative Epidemiology Unit, School of Social and Community Medicine, University of Bristol, Bristol BS82BN, UK. [5]Instituto de Investigacion Biomedica de A Coruña (INIBIC), Complexo Hospitalario Universitario de A Coruña (CHUAC), A Coruña, 15006, Spain. [6]Intelligent Systems Laboratory, University of Bristol, Bristol BS81UB, UK. Correspondence and requests for materials should be addressed to C.F.-L. (email: carlos.fernandez@udc.es)
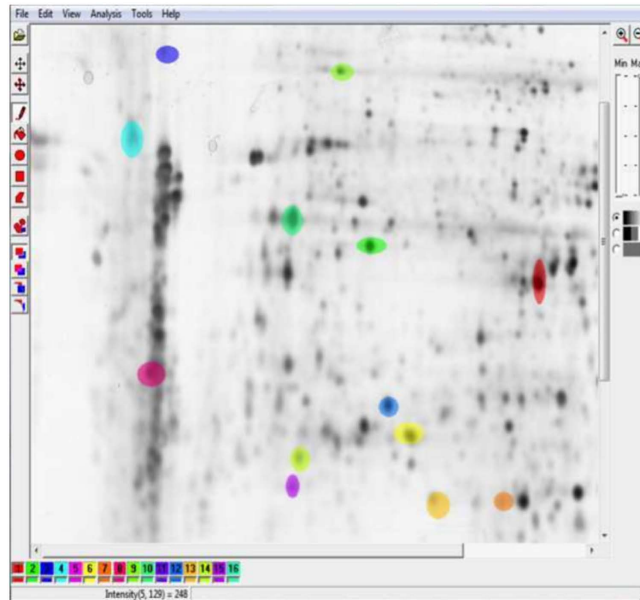
Figure 1. Two-dimensional gel electrophoresis image. Generated using human dermal fibroblasts in order to study the effect of a plant extract on the protein expression of IBR3. 1024 × 1024 8-bit image. 2D protein separation were visualized by silver staining using standard protocols. From the dataset of G.-Z. Yang and co-workers[44]. Spots manually segmented using Mazda software, first step in the image analysis pipeline (dataset generation).

Thus, image texture is also difficult to analyse. As with many other real-world problems, texture image analysis produces high-dimensional vectors (vectors made up of a large number of features) and it is of relevance to study the importance of each of these features.

Analyzing this collection of high-dimensional data is a challenge. There is also an obvious cost in computational time and memory consumption related with algorithms for analysing high-dimensional data. Furthermore, overfitting may also appear when the number of features exceeds the number of samples[4] leading to a poor predictive performance. This motivates the development of feature selection techniques to reduce dimensionality. These techniques are aimed at finding the subset of variables that describe in the best possible way the useful information contained in the data, allowing improved performance.

The aim of this work is to find complex combinations of textures that allow the use of the intrinsic information contained within the textures for classification purposes in the best possible way, as well as identifying the more relevant textures for protein classification in 2-DE electrophoresis images. Given there are several different approaches for feature Selection in Machine Learning (ML), in previous work[5] we chose to evaluate three different machine-learning feature selection approaches: subgroup-based Multiple Kernel Learning, Recursive Feature Elimination with different classifiers (Naïve Bayes, Support Vector Machines, Bagged Trees, Random Forest and Linear Discriminant Analysis) and a Genetic Algorithm based approach with a Support Vector Machines as decision function. Our study reflects that kernel-based approaches improve the interpretation of the results and further investigation should be done in order to find the best combination of variables from different groups of textures in order to measure the particular importance of each one of them to the final solution. Our study should evaluate the power of complex combinations of textures for classification purposes, therefore in this work we will focus in kernel-based methods.

Kernel-based techniques are widely used in bioinformatics and recognized as one of the state-of-the-art classifiers for supervised learning problems due to their ability to encode many different types of data[6–9], high test accuracy and ability to deal with high-dimensional datasets[10]. Different types of data can be encoded into kernels, quantifying the similarities of data objects[11]. In particular, for feature selection, these techniques have been widely applied in different areas such as ranking genes functional in cancer[12], predicting disease progression in breast cancer[13], microarray data classification[14] or autism detection[15]. With extensive applications to biomedical applications there are a number of reviews[8,10,16]. One necessary task in any investigation is a experimental evaluation of the performance of a proposed method against comparative state-of-the-art alternatives. The "no free lunch" theorem states that it is impossible to find one algorithm that is the best for solving every problem[17]. Thus, to validate our proposal, an experimental design has been concluded, consisting of four different phases: 1) Data extraction; 2) Data pre-processing; 3) Learning and 4) Selection of the best model to ensure the reproducibility of results.

In this work, a new approach for texture analysis in biomedical imaging is performed by means of integrating different types of features obtained from image textures. For this procedure, kernel-based techniques were used with different types of texture data for the selection of the most representative variables in order to improve the results achieved in classification and the interpretability of complex combinations of textures.
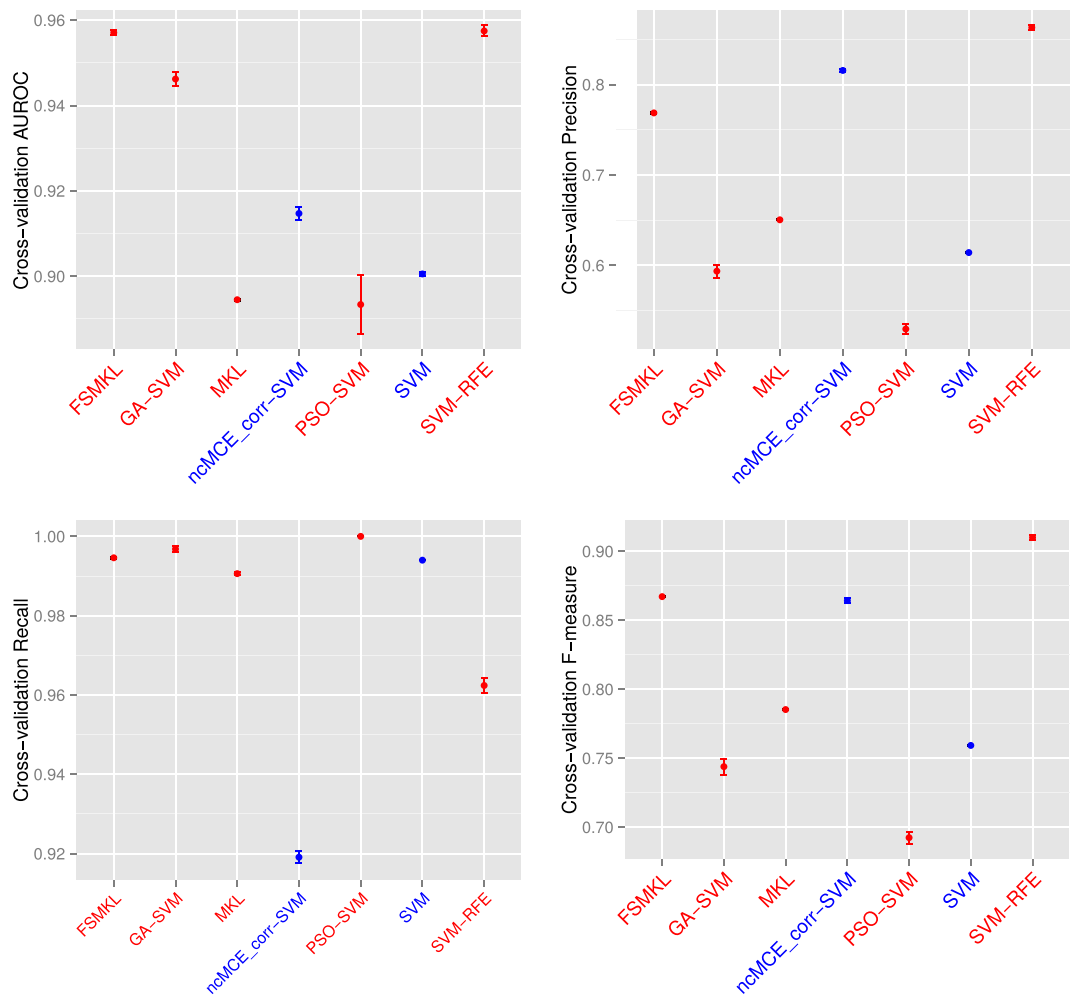
Figure 2. Classification models comparison during 10 experiments. Cross-validation AUROC, Precision, Recall and F-measure values error plot. Machine Learning techniques with Feature Selection are in red, without Feature Selection are in blue.

Furthermore, our novel approach allows scientists to investigate whether texture information within the 2-DE electrophoresis images, can be used in the proteomic field to better automatically distinguish spots from noise in the image analysis pipeline.

Results

Summary. We assembled six heterogeneous groups of textural features in order to generate the dataset from ten 2-DE images. Thus, we built a training set with 1000 samples and 274 textural variables. Eight different classification models have been used in order to generate the reference/baseline models. The values for each technique were expressed as the (mean $\pm$ standard deviation) of the Area Under the ROC curve (AUROC) measure as we performed 10 experiments for each model following a cross-validation scheme.

After the comparison study, we concluded that the integrative kernel-based approach called Feature Selection Multiple Kernel Learning[13] (FSMKL) for texture analysis is the best integrative technique for finding the best complex combination of textures required to solve the problem. This technique allows us to weight the importance of each particular textural feature in the final model.

Classification methods comparison. We determined the error plot in (Fig. 2) with four different performance measures (AUROC, Precision, Recall and F-measure. Supplementary Tables 1–4) and the mean ROC curves in (Fig. 3a) with the aforementioned classification models. We started performing experiments with a well-known Naïve Bayes (NB) method, achieving $89.84 \pm 0.11\%$, SVM $89.85 \pm 0.11\%$ and SVM applied in the 2D reduced space obtained by SVD-based MCE computed using the correlation norm (ncMCE_corr-SVM) $91.47 \pm 0.04\%$.

We continued performing feature selection (FS) with multiple kernel learning (MKL) (using only the most discriminant group of textures) we could not improve our results (achieving $89.25 \pm 0.07\%$) but this technique seemed more stable, using only fourteen features (wavelet textural features). The same occured if we considered PSO-SVM (achieving $89.13 \pm 2.15\%$), the algorithm cannot escape from local minima in most of the ten experiments performed. Finally we observed improved performance with three other FS approaches, with
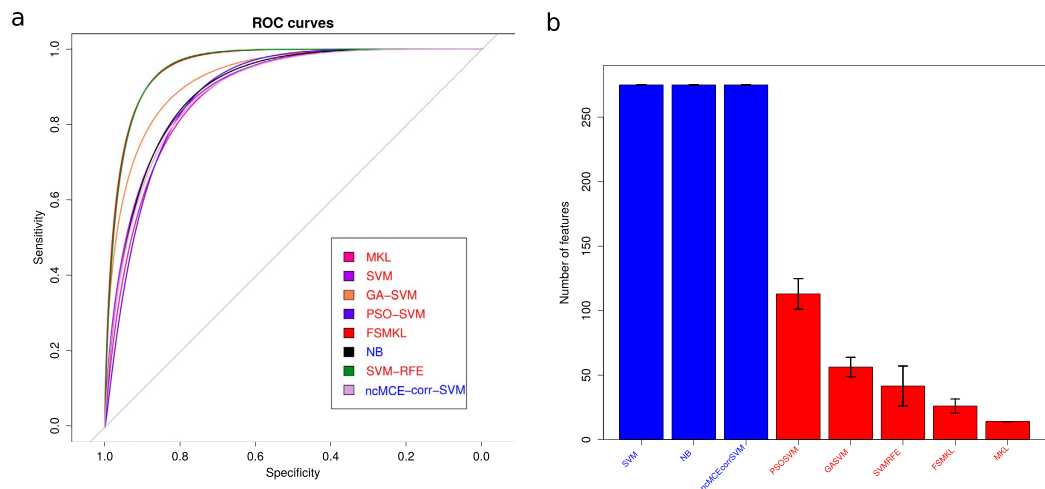
Figure 3. Classification models comparison during 10 experiments. (**a**) ROC curves. (**b**) Final number of features selected for each technique. Machine Learning techniques with Feature Selection are in red, without Feature Selection are in blue.

94.42 ± 0.48% with a GA-SVM, 95.50 ± 0.16% with FSMKL and 95.74 ± 0.40% with SVM-RFE (an SVM based on recursive feature elimination). The number of features selected for each technique is represented in (Fig. 3b).

SVM-RFE and ncMCE_corr-SVM show high precision at the expense of lower recall than the rest of the models, which means that both models penalize type II errors. This trade-off between precision and recall could be refined choosing different thresholds, and also can be controlled via the use of asymmetric soft margin parameters[18]. These two parameters can be adjusted via validation data and would directly affect the recall, precision and F-measures. Thus, for our stated results, ncMCE_corr-SVM provides a precision and F-measure, which is second best overall, but with worst recall performance (Fig. 2 and Supplementary Tables 2–4). Via use of L1 or L2-norm asymmetric soft margins[19], we can trade a loss of precision, for a gain in recall, if we wish, depending on our preferred data analysis outcome. The advantage of the AUROC as model selection score is that is independent of the threshold for binarization and includes both type I and type II errors.

Best model determination. We performed a normality analysis using the Shapiro-Wilk test[20] with the null hypothesis that the data follow a normal distribution. The null hypothesis was rejected with values $W = 0.8789$ and p-value $< 1.74 \times 10^{-06}$, therefore it could be considered that our results did not follow a normal distribution (Fig. 4a). We performed a Bartlett test[21] with the null hypothesis that our results were heteroscedastic. The null hypothesis was rejected with a value for Barlett's K squared measure of 150.110 with 7 degrees of freedom and p-value $< 2.2 \times 10^{-16}$. In this case, two of the three conditions required for a parametric test does not hold and thus, consistent with both tests, we performed a non-parametric Friedman test with the Iman-Davenport extension assuming the null hypothesis that all models have the same performance. The average rankings of the techniques compared are shown in Table 1 with Iman and Davenport statistic (distributed according to the F-distribution with 7 and 63 degrees of freedom: 78.09 and p-value $< 1.30 \times 10^{-28}$). Hence, at this point the null hypothesis is rejected for GA-SVM, SVM, NB, ncMCE_corr-SVM PSO-SVM and MKL with a high level of significance, showing that the winning models are SVM-RFE and FSMKL.

After the test for choose the significantly better models, a Finner[22] post hoc procedure must be used in order to correct and adjust the *p-values*. Results are shown in (Fig. 4b). Finner's procedure rejects hypothesis with a value ≤ 0.043, which means that MKL, PSO-SVM, SVM, NB, ncMCE_corr-SVM and GA-SVM are statistically significantly worse than the winning model. We present for each technique in the comparison in Table 2 the *p-value*, the adjusted *p-value* with Finner procedure and the final value achieved with Finner procedure. We failed to reject the null hypothesis with SVM-RFE and FSMKL. Thus, we performed a non-parametric Friedman test with Iman and Davenport statistic with FSMKL as the control model against NB, MKL, ncMCE_corr-SVM, PSO-SVM, GA-SVM and SVM and we reject the null hypothesis (distributed according to F-distribution with 6 and 54 degrees of freedom: 55.28 and p-value $< 2.67 \times 10^{-21}$). Finner's procedure rejects hypothesis with values value ≤ 0.050 as shown in (Fig. 4c). For the first image in the dataset we presented in the Supplementary Figures 1–7 the spots wrongly selected during the ten experiments by the best model.

We can conclude that both models behave similarly for this problem, being significantly better than the others, and thus there are no significant differences between FSMKL and SVM-RFE. Those techniques are not differentiable according to their AUROC values. In such case, we perform a pairwise Wilcoxon[23] test between them according to the final number of features, with the null hypothesis that both models have the same performance and we reject the null hypothesis with a p-value $< 7.74 \times 10^{-06}$. As shown in boxplot (Fig. 5), MKL and FSMKL have a lower number of features and are very stable in selecting the features whilst SVM-RFE has outliers, plotted as individual points, and GA-SVM and PSO-SVM selected a higher number of features and have long whiskers indicating variability outside the upper and lower quartiles. Considering this, we finally selected the FSMKL as model of reference.
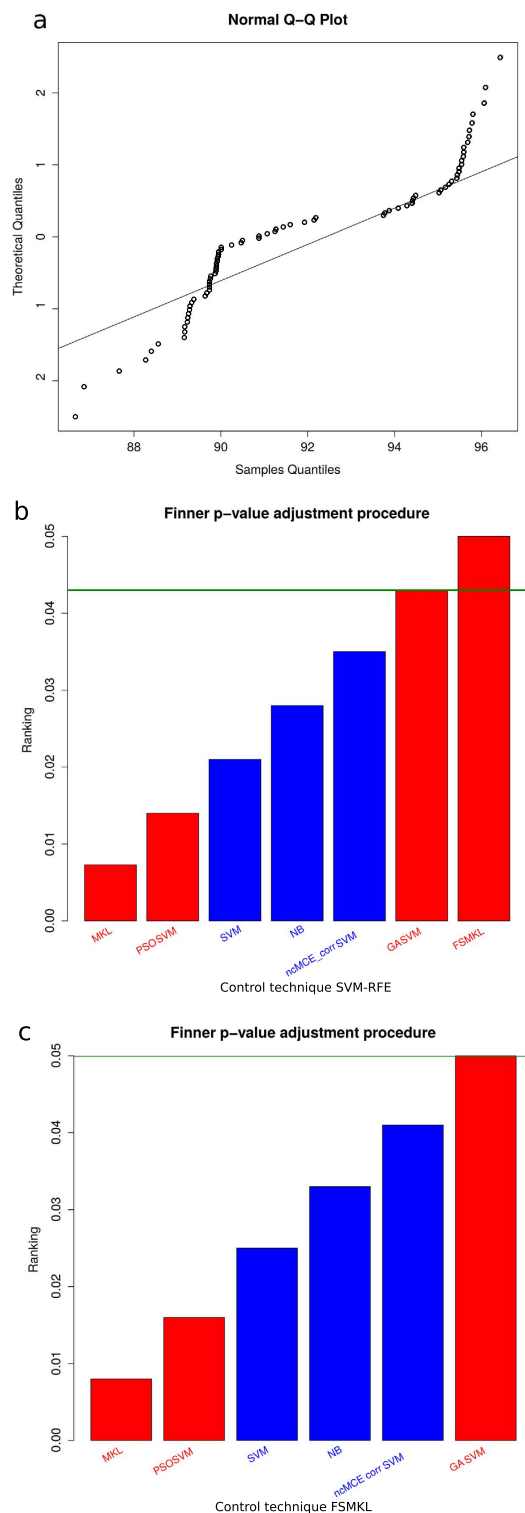
Figure 4. Best model determination. (**a**) Q-Q plot of observed versus expected values. (**b**) Finner p-value adjustment post hoc procedure. Control technique: SVM-RFE. (**c**) Finner p-value adjustment post hoc procedure. Control technique: FSMKL. Machine Learning techniques with Feature Selection are in red, without Feature Selection are in blue.

Determinant textures. Following an integrative kernel-based approach with FSMKL, we are able to weight the importance of each kernel in the final solution, measuring the relative importance of different groups of textural features in the final decision function. Thus FSMKL can indicate that the calculation of certain types of textural data may not be necessary. We used a large number of kernels, with a variable number of features per

Technique	Ranking
SVM-RFE[*]	1.29
FSMKL[*]	1.70
GA-SVM[*]	2.99
ncMCE_corr-SVM	4.1
SVM	5.89
NB	5.89
PSO-SVM[*]	6.7
MKL[*]	7.4

Table 1. Friedman's average ranking. [*]Average rankings of the different techniques used in this study using non-parametric Friedman test with Iman and Davenport extension. The lower the ranking, the better result: SVM-RFE is the control model. Feature Selection approach.

	Finner score for SVM-RFE as winner			Finner score for FSMKL as winner		
Technique	p-value	adjusted p-value	Finner score	p-value	adjusted p-value	Finner score
SVM-RFE[*]	–	–	–	–	–	–
FSMKL[*]	0.681	0.681	0.05	–	–	–
GA-SVM[*]	0.100	0.116	0.043	0.300	0.300	0.050
ncMCE_corr-SVM	3.00×10^{-3}	4.20×10^{-3}	0.035	0.029	0.035	0.041
NB	2.60×10^{-4}	6.08×10^{-4}	0.028	5.41×10^{-5}	1.08×10^{-4}	0.033
SVM	2.60×10^{-4}	6.08×10^{-4}	0.021	5.41×10^{-5}	1.08×10^{-4}	0.025
PSO-SVM[*]	5.01×10^{-6}	1.75×10^{-5}	0.014	1.14×10^{-6}	3.43×10^{-6}	0.016
MKL[*]	5.14×10^{-7}	3.60×10^{-6}	0.0073	2.27×10^{-8}	1.36×10^{-7}	0.085

Table 2. Adjusting p-values with Finner post hoc procedure. [*]In this table we present the p-value, adjusted p-value and the final value achieved with Finner procedure for the techniques evaluated in this study according with the winner model. Feature Selection approach.
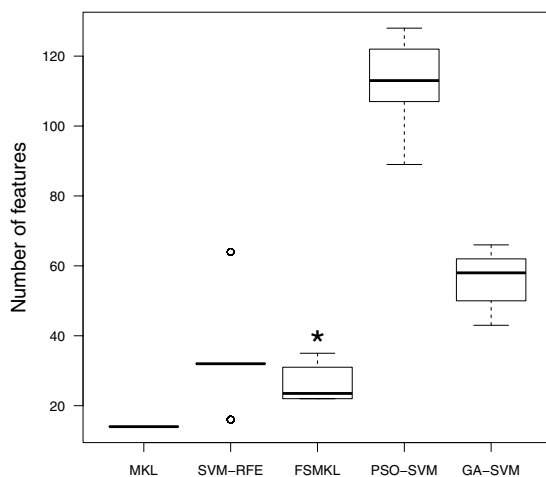

Figure 5. Number of features selected by each technique during the experimentation. This boxplot shows the stability and the number of features by each technique during the cross-validation process of the ten experiments. *Statistically significant difference with a p-value $< 7.74 \times 10^{-06}$ according with a pairwise Wilcoxon test: SVM-RFE and FSMKL.

kernel. Thus the algorithm finds which kernels, and hence which features per kernel are the most relevant for solving this classification problem[13].

We observed (Fig. 6) that the final solution is composed of 20 of the 546 kernels initially generated. With the first two kernels we explain 57% of the importance of the final solution. Both kernels only have two features S(0,5)InvDfMom and S(4,0)InvDfMom. With the inclusion of the third kernel in importance, we are able to explain 64% of the importance and we only add one new feature S(3,0)InvDfMom. Those three features are the same, but calculated with different values of distance d apart along a given direction with angle Θ. FSMKL selects 23 of
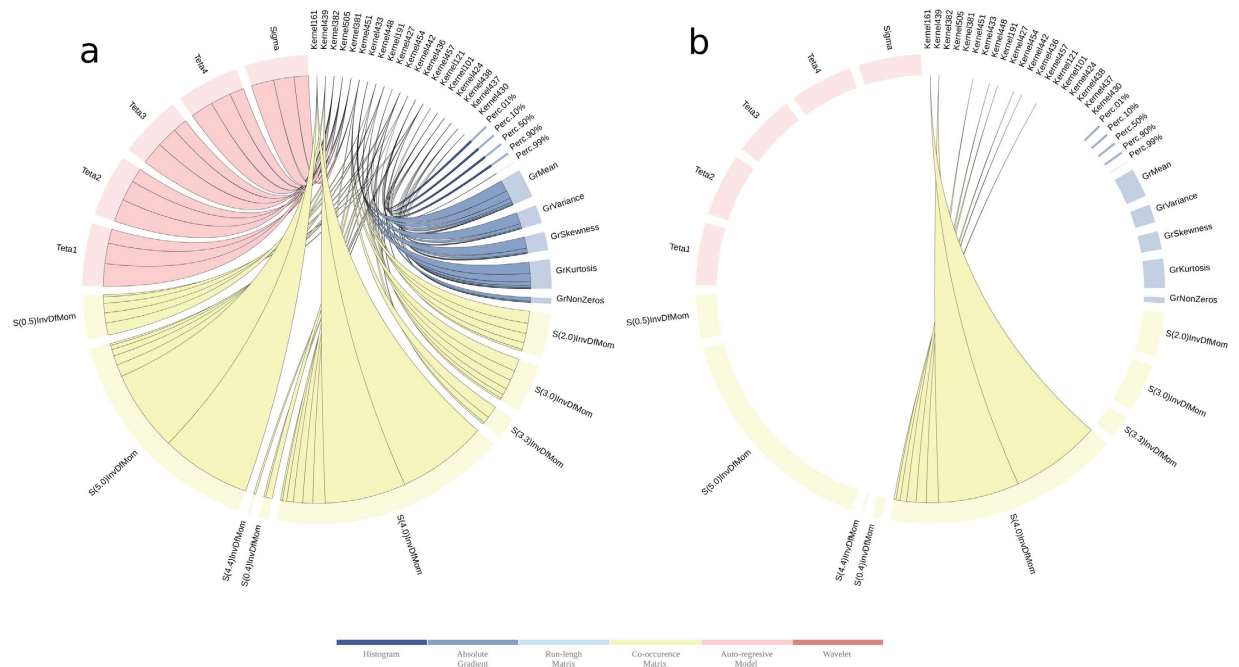
Figure 6. Importance of the kernels in the final solution in a FSMKL approach. This figure shows the importance of each kernel in the final solution. Non-zero values indicate the informative kernels. For each group of textural features we used several different types of kernels, the final decision function were only considering polynomial (with 1 and 2 degrees of freedom) and Gaussian kernels (with low values for the free parameter). The interactive version is available at http://sabia.tic.udc.es/resources/srep/.
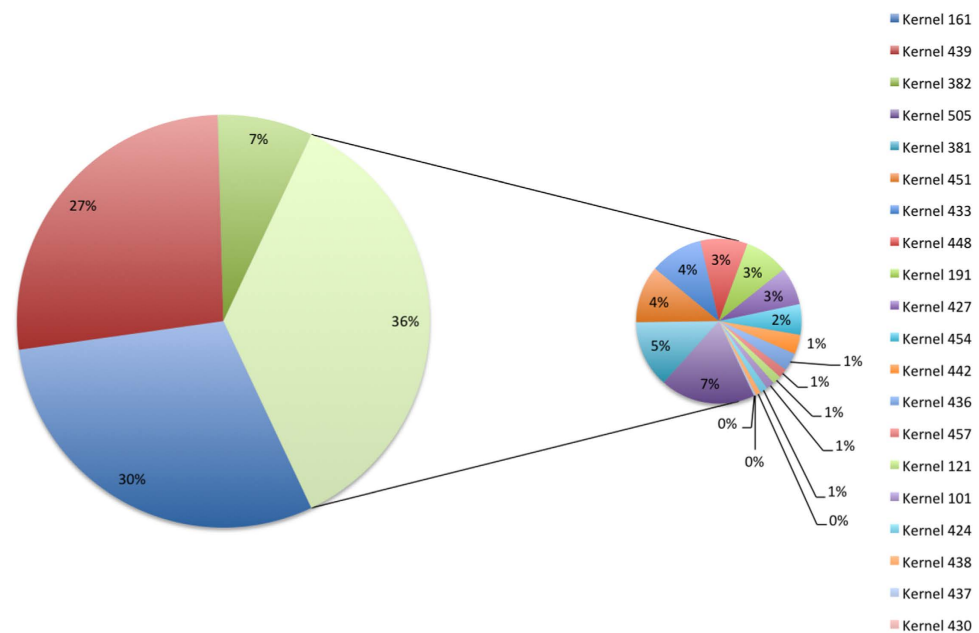

Figure 7. Importance in the final solution in a FSMKL approach. (**a**) This figure shows the importance of each feature in the final solution and the particular kernel in which the feature appears. Inverse difference moment (InvDfMom) is influenced by the homogeneity of the image. (**b**) This figure shows the importance of feature S(4,0)InvDfMom in the final solution and the particular kernels in which the feature appears. This feature appears in eight kernels with a combined weight importance of value 6.213.

the initial 274 textural features extracted as shown in Supplementary Table 5. Once FSMKL finds which kernels (and their particular importance in the final decision function), and which features per kernel, we are able to study the influence of each feature in the final solution as shown in (Fig. 7a), the interactive version is available at
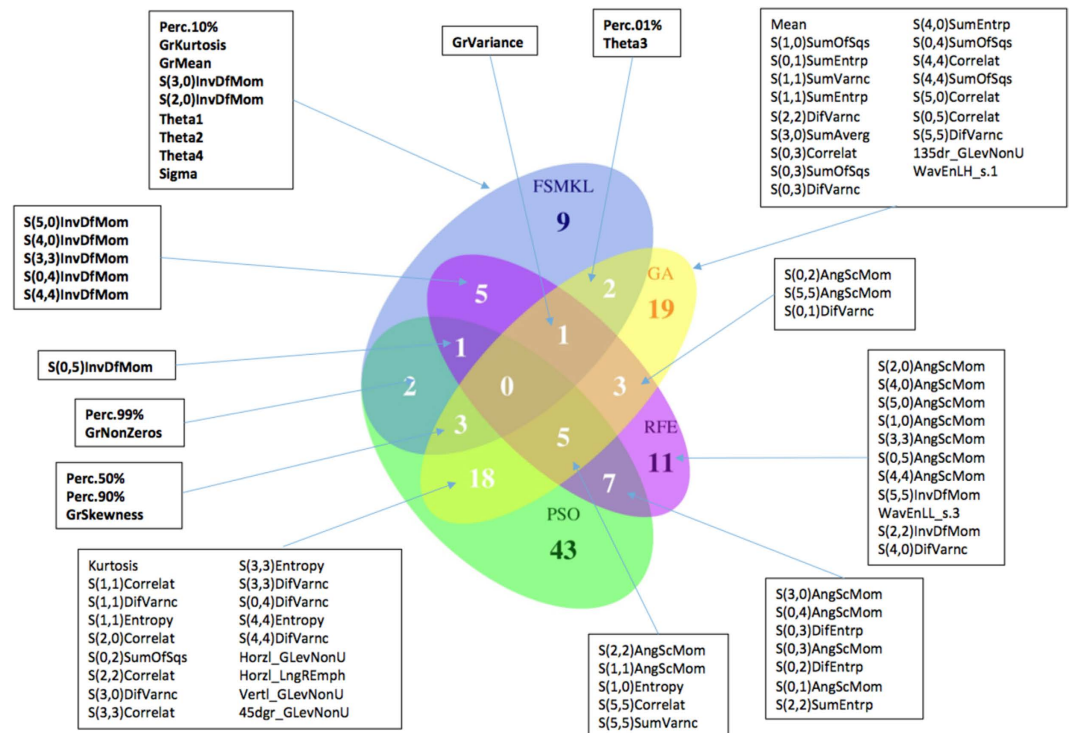
Figure 8. Venn diagram for the results obtained by different Feature Selection methods (from the R library VennDiagram). Comparing FSMKL, SVM-RFE, GA-SVM and PSO-SVM methods by checking the overlaps of their selected textural features (Top features are in bold character). Labels for the Histogram features: Perc. = percentile derived from the image histogram, Mean and Kurtosis. Labels for the absolute gradient features: Gr. = absolute gradient parameters (kurtosis, mean, skewness, variance and and percentage of pixels with nonzero gradient). Labels for the run-length matrix features: GLevNonU = grey-level non-uniformity, LongREmph = long-run emphasis. Calculated for vertical, horizontal, 45-degree and 135-degree directions. Labels for the co-occurrence matrix features: values in parenthesis represent coordinates, containing information about distance and direction between pixels (InvDfMom = inverse difference moment, AngScMom = angular second moment, DifEntrp = difference entropy, SumEntrp = sum entropy, Entropy, Correlat = correlation, SumVarnc = sum variance, DifVarnc = difference variance, SumOfSqs = sum of squares, SumAverg = sum average). Labels for the Autoregressive model features: Theta and Sigma. Labels for the Wavelet features: WavEnLH = energy of wavelet coefficients in subband LH, WavEnLL = energy of wavelet coefficients in subband LL.

http://sabia.tic.udc.es/resources/srep/. The most important feature, S(4,0)InvDfMom, appears in eight kernels with a combined importance weighting of 6.213 (Fig. 7b).

We compare the results of the four FS approaches with a Venn diagram (Fig. 8). This figure shows the overlap among the selected features. The top two features selected by FSMKL (please refer to interactive figure) are only present in the features selected by the SVM-RFE approach, both are statistically significantly better than the others. FSMKL has nine features that no other method uses and from the top ten of features from this method, there are six features that no other method use. Bio-inspired meta-heuristics (GA-SVM and PSO-SVM) are following the same paths in the feature space during its search process as they are sharing 26 features (close to 50% of the features selected by GA).

Furthermore, the most determinant textures are those derived from the Grey Level Coocurrence Matrix (GLCM), absolute gradient and autoregressive model. Second-order statistical features are computed from the intensity of pixels but taking into account spatial relationships of the two pixels in a pair. Extracting information from the GLCM provides information about the intrinsic structure of the texture[24]. The gradient of an image measures the spatial variations of grey levels across the image. Based on a first-order autoregressive model of the image, the model assumes that pixel intensity, in reference to the mean value of image intensity, may be predicted as a weighted sum of four neighboring pixel (left, top, top-left and top-right) intensities. The highest influence is achieved by the Inverse Difference Moment textural feature (IDM) which is a measure of local homogeneity.

Discussion

Discriminating proteins in 2-DE images is a difficult and challenging task. Our results showed that the combination of textural features using a data fusion approach, taking features from 4 different groups, provides effective classification with very high accuracy. One GLCM feature, IDM (with eight different values of distance d and angle σ) exhibited high discriminating power. For IDM, a higher value can be associated with a homogeneous or inhomogeneous structure as this feature describes the homogeneity of the ROI texture. It becomes reduced

in significance if local textures have maximal change: the larger the IDM, the more symmetric[25]. All the textural features from Autoregressive Model and Absolute Gradient are necessary together with five to nine Histogram features and the previously mentioned GLCM. Most previously published papers consider using only one group of textural features for feature selection strategy[26].

In biological research, GLCM parameters are generally seen as indicators of structural complexity, heterogeneity and homogeneity[27]. This study demonstrated the feasibility of using texture analysis to characterize the spatial distribution of proteins in 2-DE images in spots and noise.

IDM showed significant distinction between spots, demonstrating that these parameters may be able to differentiate them. The results of this study are consistent with those of previous studies in biomedical image texture analysis, which have reported the importance of second order textural features, in particular IDM, between normal and post-radiation therapy parotid glands in order to assess radiation-induced parotid injury[28], in the quantification of morphological changes in tissue collagen fibril organization cause by pathological conditions[29], exploring the cardioprotective effects of schisantherin A in myocardial ischemia-reperfusion (I/R) injury[30], in the quantification of the inherent heterogeneity of the relationship between biofilms images and the underling processes, such as mass-transport dynamics or substrate concentrations[31], in the quantification of chromatin structure in kidney in order to demonstrate a loss of complexity on kidney macula densa cells[32], in the characterization of the zonal dependence of biomechanical changes cause by compressive injury of immature articular cartilage[33], for plant cortical microtube quantitative analysis[34], in the quantification of lesion heterogeneity with respect to model-based enhancement kinetics parameters, directly related to tumour physiology[35] or in the analysis of multiple sclerosis[35].

High values of GLCM IDM indicate that textural values have minimal changes and are more homogeneous, these values are achieved in the noise. The IDM decreased in the spots (proteins in 2DE-images) indicating that are more inhomogeneous. The smaller the value of IDM, the more difficult the description of the texture because the texture is disorganized. Conversely, the higher the value, the easier the description of texture, because the texture is regular[34].

For example, the autoregressive model, absolute gradient and histogram-based textural parameters were also used previously in the literature for non-Hodgkin lymphoma response evaluation with MRI images during treatment with response controlled by quantitative volume analysis[36], in combination with other texture parameters such as first-, second- and high-order or wavelet for differentiation of adenocarcinoma and gastrointestinal stromal tumors, and between different grades of adenocarcinoma[37], on either T1- or T2-weighted images for the classification of focal liver lesions on MRI images[38] or for the classification of multiple sclerosis lesions[39,40], as a potential predictive tool for response evaluation in CT images after the neoadjuvant treatment in patients with lung adenocarcinoma[41] or aeophageal cancer[42] and were also used in the quantification of liver fibrosis[43].

Our results showed that FSMKL-based classification provided an AUROC score at 95.50%. This approach was stastistically significantly different from reference models. In this study, the use of kernel-based techniques was considered for the analysis of high-dimensional input spaces for scenarios such as texture analysis in biomedical imaging. Throughout different kernel-based techniques have been assessed to solve the classification task, both directly and combined with bio-inspired optimization techniques such as genetic algorithms and particle swarm optimization. Such techniques have been proven to solve these issues and, in addition, their use in combination with variable selection techniques renders them very powerful tools. In particular, in this study, kernel-based approaches have been evaluated to select variables via FSMKL (a filter approach), GA and PSO with SVM (wrapper approaches) and SVM-RFE (an embedded approach). The results obtained by FSMKL approach have proven to be significantly better than the others, taking into account the comparison made via different statistical tests. Moreover, as expected, the time required to conduct an experiment with this approach has proven to be the shortest. Furthermore, the winner model (FSMKL) shows that the combination of a kernel integration strategy (MKL) with the feature selection strategy improve the classification results in this texture analysis problem, by combining similar texture features in kernels and selecting the most important ones. This strategy improves also the interpretability of the results, showing the most informative subgroup of textures.

The proposed technique shows how texture analysis can be performed on 2-DE images to classify regions of interest corresponding to spots and noise. This is a very difficult task because of the high inter-and intra-variability (Supplementary Table 6) among different clinicians as they have to manually mark the areas to be studied. With this type of data it can be concluded that, from the entire space of input variables, the texture variable which most strongly enables the distinction between spots and noise, is the *inverse difference moment*, which is a measure of the homogeneity of the image. It is a second-order statistical operator which is calculated from the co-occurrence matrix of grey levels. This study supports the approach that second-order operators in combination with the Autoregressive Model and Absolute Gradient are better for distinguishing between spots and noise in a texture analysis: most of the previously published works considers only one group of textural features[26].

In summary, we demonstrated the feasibility of combining different groups of textures in 2-DE images images analysis for spot detection. Improved interpretability and the statistically significant difference measured against other state-of-the-art approaches indicate that our approach could be used as part of more complex 2-DE images images analysis pipelines and should be considered for the texture analysis of other types of biomedical images.

Methods

In this study we compared a set of kernel-based ML methods to see which can obtain better classification performance, and also analyse the nature of the textures selected by the methods.

The dataset.   In order to generate the dataset, ten 1024×1024 8-bit 2-DE images[1] were used, corresponding to an experiment where the effect of a plant extract on the protein expression of IBR3 human dermal fibroblasts was investigated. Spot separation patterns were visualized by silver staining using standard protocols. These

images are from the dataset owned by G.-Z. Yang[44] (Imperial College of Science, Technology and Medicine, London) and have been used in several publications[2,45,46].

For each image out of these ten, two different clinicians agreed (inter- and intra-variability in Supplementary Table 6) on 100 regions of interest (ROI), 50 spots representing proteins and 50 representing noise (noise, background, non-protein regions, cracks) manually segmented that were selected to build a training set with 1000 samples and 274 textural features using Mazda iteratively (Fig. 1). Each ROI were selected taking into consideration that there is an area of influence surrounding it, so it is slightly bigger than the visible black surface of the spot because texture information could exist also in the grey levels closest to white. Thus, one image is studied as shown in Supplementary Figure 1 where all the ROIs computed by Mazda are mixed in the same image.

We preprocess this dataset in order to have a standard normal distribution (a mean of zero and a standard deviation of one). The dataset is available for download at http://dx.doi.org/10.6084/m9.figshare.1368643.

Texture measures extraction. Texture can be determined in terms of patterns of homogeneity in appearance among spatially close pixels. Textural variables can be calculated by means of three different approaches: statistical, model-based and transform methods[3,47,48]. Statistical variables include those derived from grey-level co-occurence[49] and run-length matrix as well as mean, variance, skewness or percentiles derived from the image histogram. Model-based approaches interpret texture variables using different models such as the stochastic-like auto-regressive model. Finally, transform methods decompose an image in terms of frequency or wavelet: Gabor, Fourier or Wavelet transforms[50].

We considered six groups of textural features: Histogram-based (first-order statistical texture features), Absolute Gradient, Run-length Matrix (high-order statistical texture features), Co-occurrence Matrix (second-order statistical texture features), Autoregressive Model and Wavelet. These features are based on image histogram, co-occurrence matrix (information about the grey level value distribution of pairs of pixels), image gradients (spatial distribution of grey level values), auto-regressive models (description of texture based on statistical correlation between pairs of pixels) and wavelet analysis (information about image frequency at different scales). A more detailed description of those groups is available in the Supplementary Materials document.

We calculated those features using the specialized software *Mazda*[51]. Various approaches have demonstrated the effectiveness of this software, extracting textural features in different types of medical images[51,52]. A good review is provided in[53]. Final features used in this work are shown in Supplementary Table 7.

Machine learning techniques without feature selection. We started our experiments using two well-known machine learning methods for establishing the baseline AUROC performance for comparison. In this case we used NB[54] and SVM[55]. These techniques are suitable to deal with high-dimensional datasets. We performed our experiments with the Weka[56] implementation of the Naive Bayes algorithm. This algorithm is usually considered a naive approach because it assumes conditional independence and normal distribution in each class for the attributes. For more information about NB, please refer to[57]. The basic implementation of the SVM is for a binary (two well-separated classes) classification problem separating data via a hyperplane with maximal separation between the closest data points on each side (the support vectors). For more information about SVMs, we refer to[58,59]. For SVM and MKL we perform our experiments using MATLAB and SimpleMKL[60] code. We also included a novel SVM applied in the 2D reduced space obtained by SVD-based MCE computed using the correlation norm[61,62].

Feature selection. In high-dimensional spaces it is usual to perform a feature selection approach in order to reduce the number of features and to improve the performance of the algorithms. There are mainly three different approaches: Filter (assess the relevance of the features by looking at the intrinsic properties of the data ignoring the model), wrapper (embed the model and the feature subset search) and embedded (the feature subset search is built into the model construction) methods are the three main approaches[63]. In this study we perform experiments with two filter (MKL[64] filtering groups of features and FSMKL), two wrapper (PSO and GA with an SVM as decision function) and one embedded method (SVM-RFE) methods.

A kernel is a function that maps the input into a higher dimension in order to find a new space where the data are linearly separable. However, kernel functions and its parameters have to be determined. Thus, MKL provides a general framework for learning from multiple groups of data[65], encoding those groups in different kernels and combining kernels in a final decision function, each one with its particular value of importance[18] and automatically select these kernels and parameters. For more information about kernel-based learning machines please refer to[6,66].

PSO[67] and GA[68] are bio-inspired optimization meta-heuristics for finding the best subset[69] of input features that best reproduce the original structure of the data. PSO is based on the simulation of the social behavior of bird flocks, during its execution a set of particles moves within the function domain searching for the best fitness value whereas GA are inspired by Darwinian Evolution and an initial population of individuals (possible solutions within the function domain of a fitness function to be optimized) is evolved by means of genetic operators. Based on the latest Standard PSO implementation (SPSO-2011)[70,71] and GAlib[72], we modify them in order to add a binary representation for each particle/individual as a feature mask for the input feature space. SVM decision function is obtained by LIBSVM[73].

SVM-RFE[12] was originally developed for ranking genes in a cancer classification problem according to the hyperplane decision value of a SVM. RFE operates by removing genes (one or more at each iteration) according to the lowest score until the highest performance is achieved. The R Statistical Package[74] was used and it was necessary to enhance the Caret package[75] capacities in order to include a new ranking criterion $\|w\|$ for supporting SVM-RFE as initially proposed by Guyon[12], we used Kernlab[76] and pROC[77] for ROC curves and bar plot figures.

Non-parametric tests were performed using code provided by Garcia[17] *et al.* and are available in the Supplementary Materials of this work.

The FSMKL model presented in[13] used a sparse MKL minimization algorithm, which allowed selection of a low number of kernels and their ranking by importance for classification. For each of the datasets (in this particular case, for each group of textures), the FSMKL ranks the features of each group of textures in relation to the most statistically aligned with the class, and codes subsets of these ranked features as kernels. In this way, the FSMKL select, not only the most important texture subgroup for classification, but also a subset of the textures which compose that group. In this work, from the six textural groups presented, this method allows to find the most relevant texture features and groups for the given classification problem. Furthermore, FSMKL gives greater interpretability to the complex relationships among different groups of texture than simply performing the better feature selection combination as this technique can measure the final importance of each feature to the final solution.

Experimental analysis. In order to discover which of the proposed models are statistically significantly better, a set of tests was performed for the analysis of the behaviour of those techniques following the methodology proposed in[17,78,79]. In order to choose between a parametric or a non-parametric test to compare the models, we used three required conditions for using parametric tests: independence, normality and heteroscedasticity. The use of a parametric test is only appropriate when the results of those techniques fulfilled the three conditions aforementioned[17].

The independence condition is fulfilled because we perform different runs following a tenfold cross-validation approach for separating the data with a prior random reshuffling of the examples. A cross-validation approach splits the dataset into ten random equal-size subsets, nine of which are chosen ten times to train the model and the remaining set is used to test them (each iteration a random subset is chosen to be the test set).

For the normality condition we used the Shapiro-Wilk test[20] with the null hypothesis that the data follow a normal distribution and in order to evaluate the heteroscedasticity, we performed a Bartlett test[21] with the null hypothesis that the results were heteroscedastic. In order to compare the models, a non-parametric Friedman test with the Iman-Davenport extension was employed, where the null hypothesis is that all the models have the same performance. Once the test for check if a model is statistically better than the others, a post-hoc procedure had to be used in order to address the multiple hypothesis testing among the different models. A Finner[22] post-hoc procedure has to be used for detecting significance of the multiple comparisons[17,79,80] and the *p-values* should be corrected and adjusted. We perform our experimental analysis with a level of confidence $= 0.05$.

At this point, if we failed to reject the null hypothesis for two or more models, we can conclude that those models behave similarly for this problem and that there are no significant differences between them. The test reports in this case that those techniques are not differentiable with the particular performance measure used. In such case it is possible to keep the first one according to the ranking, considering that is the number one in the ranking but it is no significantly better that the others or it is possible to perform a new test taking into consideration other performance measure (number of features, time or simplicity depending on the particularities of the algorithms)[17,80].

References

1. Rabilloud, T., Chevallet, M., Luche, S. & Lelong, C. Two-dimensional gel electrophoresis in proteomics: Past, present and future. *J. Proteomics* **73,** 2064–2077 (2010).
2. Rodriguez, A., Fernandez-Lozano, C., Dorado, J. & Rabuñal, J. R. Two-dimensional gel electrophoresis image registration using block-matching techniques and deformation models. *Anal. Biochem.* **454,** 53–59 (2014).
3. Fernandez-Lozano, C., Gestal, M., Pedreira, N., Dorado, J. & Pazos, A. High order texture-based analysis in biomedical images. *Curr. Med. Imaging Rev.* **9,** 309–317 (2013).
4. Berthold, M. R. & Hand, D. J. *Intelligent Data Analysis: An Introduction* 1st edn (Springer-Verlag, Secaucus, 1999).
5. Fernandez-Lozano, C. *et al.* Texture classification using feature selection and kernel-based techniques. *Soft Comput.* **19,** 2469–2480 (2015).
6. Schölkopf, B. & Smola, A. J. *Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond* (MIT Press, Cambridge, 2001).
7. Möuller, K., Mika, S., Rätsch, G., Tsuda, K. & Schölkopf, B. An introduction to kernel-based learning algorithms. *IEEE Trans. Neural Netw. Learn. Syst.* **12,** 181–201 (2001).
8. Schölkopf, B., Tsuda, K. & Vert, J.-P. *Kernel Methods in Computational Biology*. Computational Molecular Biology (MIT Press, Cambridge, 2004).
9. Vert, J.-P. In *Kernel Methods in Bioengineering, Signal and Image Processing* (eds. Camps-Valls, G. *et al.*), Ch. 2, 42–63 (IGIGlobal, Hershey, 2007).
10. Ben-Hur, A., Ong, C. S., Sonnenburg, S., Schölkopf, B. & Rätsch, G. Support vector machines and kernels for computational biology. *PLoS Comput. Biol.* **4,** e1000173, doi: 10.1371/journal.pcbi.1000173 (2008).
11. Campbell, C. In *Springer Handbook of Bio-/Neuroinformatics* (ed. Kasabov, N.), Ch. 12, 185–206 (Springer, Berlin, 2014).
12. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Mach. Learn.* **46,** 389–422 (2002).
13. Seoane, J. A., Day, I. N. M., Gaunt, T. R. & Campbell, C. A pathway-based data integration framework for prediction of disease progression. *Bioinformatics* **30,** 838–845 (2014).
14. Sun, S., Peng, Q. & Shakoor, A. A kernel-based multivariate feature selection method for microarray data classification. *PLoS One* **9,** e102541, doi: 10.1371/journal.pone.0102541 (2014).
15. Kosmicki, J. A., Sochat, V., Duda, M. & Wall, D. P. Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Transl. Psychiatr.* **5,** e514, doi: 10.1038/tp.2015.7 (2015).
16. Borgwardt, K. In *Handbook of Statistical Bioinformatics* (eds. Lu, H. H.-S. *et al.*), Ch. 15, 317–334 (Springer, Berlin, 2011).
17. Garcia, S., Fernandez, A., Luengo, J. & Herrera, F. Advanced nonparametric tests for multiple comparisons in the design of experiments in computational intelligence and data mining: Experimental analysis of power. *Inf. Sci.* **180,** 2044–2064 (2010).
18. Campbell, C. & Ying, Y. Learning with Support Vector Machines. *Synthesis Lectures on Artificial Intelligence and Machine Learning* **5,** 1–95 (2011).
19. Veropoulos, K., Campbell, C. & Cristianini, N. Controlling the Sensitivity of Support Vector Machines. In *Proceedings. International Joint Conference on Artificial Intelligence*, Other: ML3, 55–60 (Stockholm, Sweden, 1999).

20. Shapiro, S. S. & Wilk, M. B. An analysis of variance test for normality (complete samples). *Biometrika* **52,** 591–611 (1965).
21. Bartlett, M. S. Properties of sufficiency and statistical tests. *Proc. R. Soc. Lond. A* **160,** 268–282 (1937).
22. Finner, H. On a monotonicity problem in step-down multiple test procedures. *J. Am. Stat. Assoc.* **88,** 920–923 (1993).
23. Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics* **1,** pp. 80–83 (1945).
24. Haralick, R. Statistical and structural approaches to texture. *Proc. IEEE* **67,** 786–804 (1979).
25. Yang, C., Zhu, H., Wu, S., Bai, Y. & Gao, H. Correlations between B-mode ultrasonic image texture features and tissue temperature in microwave ablation. *J. Ultrasound Med.* **29,** 1787–99 (2010).
26. Kassner, A. & Thornhill, R. Texture analysis: A review of neurologic MR imaging applications. *Am. J. Neuroradiol.* **31,** 809–816 (2010).
27. Pantic, I., Pantic, S., Paunovic, J. & Perovic, M. Nuclear entropy, angular second moment, variance and texture correlation of thymus cortical and medullar lymphocytes: grey level co-occurrence matrix analysis. *An. Acad. Bras. Cienc.* **85,** 1063–1072 (2013).
28. Yang, X. *et al.* Ultrasound GLCM texture analysis of radiation-induced parotid-gland injury in head-and-neck cancer radiotherapy: an *in vivo* study of late toxicity. *Med. Phys.* **39,** 5732–5739 (2012).
29. Mostaço-Guidolin, L. B. *et al.* Collagen morphology and texture analysis: from statistics to classification. *Sci. Rep.* **3,** 2190, doi: 10.1038/srep02190 (2013).
30. Chang, R. *et al.* Protective role of deoxyschizandrin and schisantherin A against myocardial ischemia-reperfusion injury in rats. *PloS One* **8,** e61590, doi: 10.1371/journal.pone.0061590 (2013).
31. Yang, X., Beyenal, H., Harkin, G. & Lewandowski, Z. Quantifying biofilm structure using image analysis. *J. Microbiol. Methods* **39,** 109–119 (2000).
32. Pantic, I. *et al.* Complexity reduction of chromatin architecture in macula densa cells during mouse postnatal development. *Nephrology* **18,** 117–124 (2013).
33. Rolauffs, B. *et al.* Vulnerability of the superficial zone of immature articular cartilage to compressive injury. *Arthritis Rheumatol.* **62,** 3016–3027 (2010).
34. Lu, Y., Huang, C., Wang, J. & Shang, P. An improved quantitative analysis method for plant cortical microtubules. *Sci. World J.* **2014,** 637183, doi: 10.1155/2014/637183 (2014).
35. Karahaliou, A. *et al.* Assessing heterogeneity of lesion enhancement kinetics in dynamic contrast-enhanced MRI for breast cancer diagnosis. *Br. J. Radiol.* **83,** 296–309 (2010).
36. Harrison, L. C. V. *et al.* Non-Hodgkin lymphoma response evaluation with MRI texture classification. *J. Exp. Clin. Cancer Res.* **28,** doi: 10.1186/1756-9966-28-87 (2009).
37. Ba-Ssalamah, A. *et al.* Texture-based classification of different gastric tumors at contrast-enhanced CT. *Eur. J. Radiol.* **82,** e537–e543, doi: 10.1016/j.ejrad.2013.06.024 (2013).
38. Mayerhoefer, M. E. *et al.* Texture-based classification of focal liver lesions on MRI at 3.0 Tesla: A feasibility study in cysts and hemangiomas. *J. Magn. Reson. Imaging* **32,** 352–359 (2010).
39. Harrison, L. C. V. *et al.* MRI texture analysis in multiple sclerosis: toward a clinical analysis protocol. *Acad. Radiol.* **17,** 696–707 (2010).
40. Zhang, J., Tong, L., Wang, L. & Li, N. Texture analysis of multiple sclerosis: a comparative study. *Magn. Reson. Imaging* **26,** 1160–1166 (2008).
41. Chong, Y. *et al.* Quantitative CT variables enabling response prediction in neoadjuvant therapy with EGFR-TKIs: are they different from those in neoadjuvant concurrent chemoradiotherapy? *PloS One* **9,** e88598, doi: 10.1371/journal.pone.0088598 (2014).
42. Yip, C. *et al.* Assessment of changes in tumor heterogeneity following neoadjuvant chemotherapy in primary esophageal cancer. *Dis. Esophagus* **28,** 172–179 (2015).
43. Barry, B. *et al.* Quantifying liver fibrosis through the application of texture analysis to diffusion weighted imaging. *Magn. Reson. Imaging* **32,** 84–90 (2014).
44. Veeser, S., Dunn, M. J. & Yang, G.-Z. Multiresolution image registration for two-dimensional gel electrophoresis. *Proteomics* **1,** 856–870 (2001).
45. Dowsey, A. W. *et al.* Image analysis tools and emerging algorithms for expression proteomics. *Proteomics* **10,** 4226–4257 (2010).
46. Fernandez-Lozano, C., Seoane, J., Gestal, M., Gaunt, T. & Campbell, C. In *Advances in Computational Intelligence* (eds. Rojas, I. *et al.*), Vol. 7902 of *Lecture Notes in Computer Science*, 427–434 (Springer, Berlin, 2013).
47. Tuceryan, M. & Jain, A. In *Handbook of pattern recognition and computer vision* 3rd edn, Vol. 2 (eds. Chen, C. H. *et al.*), Ch. 2, 235–276 (World Scientific, Singapore, 1999).
48. Henry, W. In *Biomedical Imaging* (ed. Mao, Y.), Ch. 4, 235–276, doi: 10.5772/8912 (InTech, 2010).
49. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst. Man Cybern.* **SMC3,** 610–621 (1973).
50. Szczypiński, P. M., Klepaczko, A. & Zapotoczny, P. Identifying barley varieties by computer vision. *Comput. Electron. Agric.* **110,** 1–8 (2015).
51. Szczypiński, P. M., Strzelecki, M., Materka, A. & Klepaczko, A. Mazda – A software package for image texture analysis. *Comput. Meth. Programs Biomed.* **94,** 66–76 (2009).
52. Mayerhoefer, M. E. *et al.* Texture analysis for tissue discrimination on T1-weighted MR images of the knee joint in a multicenter study: Transferability of texture features and comparison of feature selection methods and classifiers. *J. Magn. Reson. Imaging* **22,** 674–680 (2005).
53. Materka, A. & Strzelecki, M. Texture analysis methods – A review. *Technical University of Lodz, Institute of Electronics. COST B11 report* Technical Report. (1998). Available at: http://www.eletel.p.lodz.pl/programy/cost/pdf_1.pdf (Accessed: 30/09/2015).
54. John, G. H. & Langley, P. In *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, UAI'95, 338–345 (Morgan Kaufmann, San Francisco, 1995).
55. Vapnik, V. N. *Estimation of dependences based on empirical data* (Springer Verlang, New York, 2006).
56. Hall, M. *et al.* The weka data mining software: An update. *SIGKDD Explor.* **11,** 10–18 (2009).
57. Zhang, H. Exploring conditions for the optimality of nave bayes. *Int. J. Patt. Recogn. Artif. Intell.* **19,** 183–198 (2005).
58. Burges, C. J. C. A tutorial on support vector machines for pattern recognition. *Data Min. Knowl. Discov.* **2,** 121–167 (1998).
59. Cristianini, N. & Shawe-Taylor, J. *An Introduction to Support Vector Machines: And Other Kernel-based Learning Methods* (Cambridge University Press, New York, NY, USA, 2000).
60. Rakotomamonjy, A., Bach, F., Canu, S. & Grandvalet, Y. SimpleMKL. *J. Mach. Learn. Res.* **9,** 2491–2521 (2008).
61. Cannistraci, C. V., Ravasi, T., Montevecchi, F. M., Ideker, T. & Alessio, M. Nonlinear dimension reduction and clustering by minimum curvilinearity unfold neuropathic pain and tissue embryological classes. *Bioinformatics* **26,** i531–i539, doi: 10.1093/bioinformatics/btq376 (2010).
62. Cannistraci, C. V., Alanis-Lobato, G. & Ravasi, T. Minimum curvilinearity to enhance topological prediction of protein interactions by network embedding. *Bioinformatics* **29,** i199–i209, doi: 10.1093/bioinformatics/btt208 (2013).
63. Saeys, Y., Inza, I. N. & Larrañaga, P. A review of feature selection techniques in bioinformatics. *Bioinformatics* **23,** 2507–2517 (2007).
64. Lanckriet, G. R. G., De Bie, T., Cristianini, N., Jordan, M. I. & Noble, W. S. A statistical framework for genomic data fusion . *Bioinformatics* **20,** 2626–2635 (2004).
65. Gönen, M. & Alpaydin, E. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* **12,** 2211–2268 (2011).
66. Shawe-Taylor, J. & Cristianini, N. *Kernel Methods for Pattern Analysis* (Cambridge University Press, New York, 2004).

67. Kennedy, J. & Eberhart, R. Particle swarm optimization. In *Proceedings. IEEE International Conference on Neural Networks*, Vol. 4, 1942–1948 (IEEE, Perth, 1995).
68. Holland, J. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence*. (MIT Press, Cambridge, 1975).
69. Fernandez-Lozano, C. *et al.* Markov mean properties for cell death-related protein classification. *J. Theor. Biol.* **349,** 12–21 (2014).
70. Clerc, M. Beyond standard particle swarm optimisation. *Int. J. Swarm. Intell. Res.* **1,** 46–61 (2010).
71. Zambrano-Bigiarini, M., Clerc, M. & Rojas, R. Standard particle swarm optimisation 2011 at CEC-2013: A baseline for future PSO improvements. in *IEEE Congress on Evolutionary Computation*, 2337–2344 (2013).
72. Wall, M. GAlib: A C++ library of genetic algorithm components. (MIT Press, 1996).
73. Chang, C.-C. & Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.* **2,** 1–27 (2011). Software available at http://www.csie.ntu.edu.tw/cjlin/libsvm. Date of access: 30/09/2015.
74. R. Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria (2014). Software available at http://www.R-project.org. Date of access:30/09/2015.
75. Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat. Softw.* **28,** 1–26, doi: 10.18637/jss.v028.i05 (2008).
76. Karatzoglou, A., Smola, A., Hornik, K. & Zeileis, A. Kernlab–an S4 package for kernel methods in R. *J. Stat. Softw.* **11,** 1–20, doi: 10.18637/jss.v011.i09 (2004).
77. Robin, X. *et al.* pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* **12,** 77, doi: 10.1186/1471-2105-12-77 (2011).
78. Sheskin, D. *Handbook of Parametric and Nonparametric Statistical Procedures* 5th edn (CRC Press, Florida 2011).
79. Garcia, S., Fernandez, A., Luengo, J. & Herrera, F. A study of statistical techniques and performance measures for genetics-based machine learning: Accuracy and interpretability. *Soft Comput.* **13,** 959–977 (2009).
80. Demšar, J. Statistical comparisons of classifiers over multiple data sets. *J. Mach. Learn. Res.* **7,** 1–30 (2006).

Acknowledgements

Author Contributions

C.F.-L., J.A.S., M.G., T.R.G., J.D., A.P. and C.C. conceived the experiment(s), C.F.L. and J.A.S. conducted the experiment(s), C.F.L. and J.A.S. analysed the results. C.F.L. wrote de paper. C.F.L., J.A.S., M.G., T.R.G. and C.C. reviewed the manuscript. All authors read and approved the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Fernandez-Lozano, C. *et al.* Texture analysis in gel electrophoresis images using an integrative kernel-based approach. *Sci. Rep.* **6**, 19256; doi: 10.1038/srep19256 (2016).