

# Bone Texture Characterization for Osteoporosis Diagnosis using Digital Radiography

Keni Zheng and Sokratis Makrogiannis

**Abstract**—We introduce texture classification techniques to effectively diagnose osteoporosis in bone radiography data. Osteoporosis is an age-related systemic bone skeletal disorder characterized by low bone mass and bone structure deterioration that results in increased bone fragility and higher fracture risk. Therefore, early diagnosis can effectively predict fracture risk and prevent the disease. Automated diagnosis from digital radiographs is very challenging since the scans of healthy and osteoporotic subjects show little or no visual differences, and their density histograms mostly overlap. We designed a system to separate healthy from osteoporotic subjects using high-dimensional textural feature representations computed from radiographs. These features were then reduced using feature selection to obtain the more discriminant subset that was finally classified by our methods. The top performing approach yields 79.3% accuracy and 81% area under the ROC over 116 bone radiographs.

## I. INTRODUCTION

Osteoporosis is a skeletal disorder characterized by decreased bone strength that may lead to susceptibility of fracture [1]. Aerial Bone Mineral Density (BMD) is computed in dual-energy X-ray absorptiometry (DXA) scans to diagnose osteoporosis [2]. However, BMD can predict fracture with only 60% accuracy. Analysis of trabecular bone microarchitecture can significantly improve the prediction rates, but this information requires bone biopsy with histomorphometric analysis. The task of obtaining trabecular bone microarchitecture information by noninvasive methods is a nontrivial scientific problem [3]. Previous approaches to evaluating bone structure on radiographs by 2D texture analysis were reported in [2], [4], [5]. Moreover, in [6], [7] the authors propose to use 2D texture analysis to characterize 3D bone microarchitecture.

Diagnosis of osteoporosis using bone radiograph scans presents some challenges, mainly because images of osteoporotic and healthy subjects are visually very similar. Here we introduce a method for computer-aided diagnosis of osteoporosis in bone radiography data. Our premise is that the deterioration of trabecular bone microarchitecture can be captured by textural features. We first computed texture features based on wavelet decomposition, discrete Fourier and Cosine transforms, fractal dimension, statistical

co-occurrence indices, and structural texture descriptors. We employed feature selection techniques that consider the individual feature predictive ability and inter-feature redundancy to find the most discriminant feature set. In the classification stage we employed Naive Bayes, Multilayer Perceptron, Bayes Network, Random Forests and Bagging models for diagnosis. The system produced a classification accuracy of 79.3% and a receiver operating characteristic area of 81% over 116 images. These results are particularly promising when we consider the level of difficulty of the specific dataset.

## II. METHODS

### A. Texture Features

In this stage we compute texture descriptors that can be used to form morphometric signatures for separation between groups of healthy and osteoporotic subjects. This is usually performed in a high-dimensional feature space to reduce the Bayes error rate in principle. Fig. 1 shows an image of a healthy and an osteoporotic subject, and their corresponding intensity histograms. It is difficult to distinguish between the two classes since these two figures are very similar and there is significant overlap between the two distributions. We describe our feature set next.

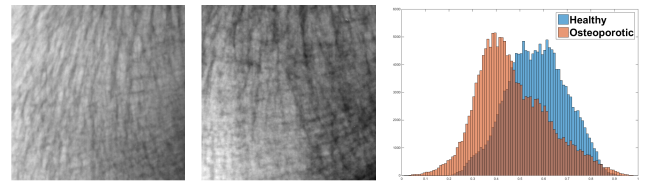


Fig. 1. Example of textures of a control subject (left) and an osteoporotic subject (middle). Histograms of two scans show significant overlap (right).

1) *Fractal Dimension*: We computed Fractal Dimension attributes that have shown promise in texture classification applications. A fractal is defined as a mathematical set whose Hausdorff dimension exceeds the fractal's topological dimension [8]. It has been shown that fractal dimension correlates well with a function's roughness. Therefore, we used fractal dimension to measure the roughness and granularity of the image intensity function. The topological dimension of this function is equal to 3, consisting of 2 spatial dimensions plus the intensity. We utilized the method of box counting to compute the fractal dimension explained as follows. Assuming a fractal structure with dimension  $D$ , we let  $N(\epsilon)$  be the number of non-empty boxes of size  $\epsilon$  required to

This research was supported by the National Institute of General Medical Sciences of the National Institutes of Health (NIH) under Award Number SC3GM113754 and by the Intramural Research Program of NIA, NIH. We also acknowledge the support of the Center for Research and Education in Optical Sciences and Applications (CREOSA) of Delaware State University funded by NSF CREST-8763.

Keni Zheng and Sokratis Makrogiannis are affiliated with the Department of Mathematical Sciences, Delaware State University, Dover, DE 19901-2277. Corresponding author email: smakrogiannis@desu.edu.

cover the fractal support. Using the relation  $N(\epsilon) \simeq \epsilon^{(-D)}$ , we can numerically estimate  $D$  from  $D = \lim_{\epsilon \rightarrow 0} \frac{\log N(\epsilon)}{-\log \epsilon}$  by least squares fitting. For the case of grayscale images or continuous functions, we generated 8 binary sets using multiple Otsu thresholding, then computed the fractal dimension, area, and mean intensity for each point set as in [9].

2) *Wavelet Texture Descriptors*: A multi-scale texture descriptor is usually very useful for classification. Gabor and wavelet transforms are both multi-scale spatial-spatial frequency filtering techniques. The discrete wavelet transform is frequently applied using tree or pyramid hierarchies for texture representation. Multi-band analysis offers advantages over the traditional discrete Fourier transform, but wavelet transform does not produce as exact a result as the Fourier transform.

a) *Discrete Wavelet Frames*: Discrete wavelet frames employ a filter bank for multi-scale decomposition. The Haar wavelet with a low-pass filter  $H(z) = (1+z)/2$  and a corresponding high-pass filter  $G(z) = (z-1)/2$  is frequently used because of its efficiency and computational simplicity. The largest filter kernels will have size  $2^{\maxlevel}$ , where the  $\maxlevel$  is the number of multiresolution levels. At each level, we filter the image by using the filter combinations:  $H_x H_y, H_x G_y, G_x H_y, G_x G_y$ , where  $H_x$  is the low-pass filter along the  $x$  direction, and  $G_y$  is the high-pass filter along the  $y$  direction. To produce the wavelet frame representation we compute the discrete wavelet transform for all possible signal shifts at multiple scales. The filters are used to decompose the image in subbands. We compute the orthogonal projections and residues for a full discrete wavelet expansion. We then compute energy, variance, entropy, and kurtosis signatures to form the texture descriptor.

b) *Wavelet Gabor Filter Bank*: The Gabor filter is a linear filter that can extract relevant characteristics for multiple frequencies and orientations (Fig. 2), similarly to the human visual system. Gabor functions form a complete but non-orthogonal basis. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave. Gabor filters are often used for texture identification, and good results have been achieved. The filter has a real and an imaginary component representing orthogonal directions, defined in complex form as follows:

$$g(x, y; \lambda, \theta, \psi, \sigma, \gamma) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \exp\left(i\left(2\pi \frac{x'}{\lambda} + \psi\right)\right) \quad (1)$$

and  $x' = x \cos \theta + y \sin \theta$ ,  $y' = -x \sin \theta + y \cos \theta$ ,  $\lambda$  is wavelength of the sinusoidal factor,  $\theta$  is orientation of the normal to the parallel stripes of a Gabor function,  $\psi$  is phase offset,  $\sigma$  is standard deviation of the Gaussian envelope and  $\gamma$  is spatial aspect ratio. The filter dictionary can be produced by dilations and rotations of the mother Gabor wavelet.

3) *Local Binary Patterns (LBP)*: For each pixel  $p$  in the image, we compare the intensity of  $p$  to the intensities of its eight neighbors. If the intensity of  $p$  is greater or equal to its  $i$ th (where  $i = 1, 2, \dots, 8$ ) neighbor, we set  $b_i = 0$ , otherwise  $b_i = 1$ . From these eight neighbors we construct an eight-digit

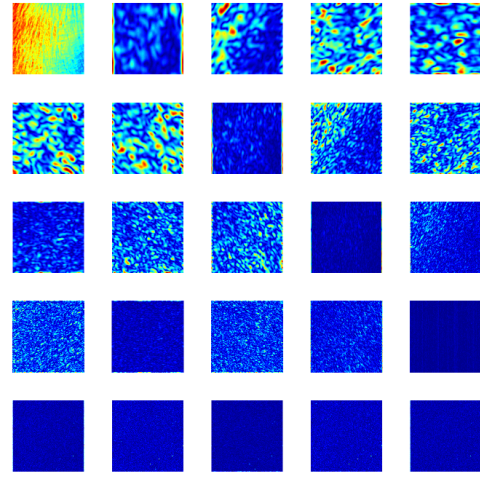


Fig. 2. The original scan and the Gabor texture components of a healthy subject using 4 scales and 6 orientations. While these maps pronounce the texture characteristics, visual interpretation is still particularly challenging. Therefore a machine learning technique is needed to distinguish healthy from osteoporotic subjects.

binary number  $b_1 b_2 b_3 b_4 b_5 b_6 b_7 b_8$ . We use the histogram of these numbers as a texture descriptor [10].

4) *Discrete Fourier and Cosine Transforms*: We utilize discrete Fourier transform and the discrete Cosine transform coefficients to capture spectral characteristics of texture. For example, fine texture has greater high frequency components, whereas coarse texture is represented by lower frequencies. We use the  $8 \times 8$  coefficients corresponding to lower frequencies for classification.

5) *Law's Texture Energy Masks*: The texture energy is computed by a set of  $5 \times 5$  convolution masks (level, edges, waves, spots, and ripples) to measure the amount of variation within a fixed-size window. We use the average level (intensity) feature to normalize intensity range and then we use the remaining 24 components to form the texture vector. Next, we calculate the mean, standard deviation, energy, skewness, kurtosis, and entropy for each component.

6) *Edge Histogram*: We compute the intensity gradient magnitude  $|\nabla f|$  and then calculate its histogram by  $p_{|\nabla f|}(|\nabla f| = r_k) = \frac{n_k}{N}$ ,  $k = 0, \dots, L-1$ .

7) *Gray Level Co-Occurrence Matrix (GLCM)*: The GLCM calculates how the frequency of occurrence of gray-level pairs  $(i, j)$  in horizontal, vertical, or diagonal pixel adjacencies on the image plane. Horizontal ( $0^\circ$ ), vertical ( $90^\circ$ ), and diagonal ( $-45^\circ, -135^\circ$ ) dimensions of analysis are denoted by  $P_0, P_{90}, P_{45}$ , and  $P_{135}$ , respectively. After we create the GLCMs, we compute contrast, correlation, energy and homogeneity measures.

## B. Feature Selection

Feature selection aims to select relevant and informative features for classification. It is applied to improve classification performance, to reduce computational complexity, and to interpret data.

1) *Correlation-based Feature Selection (CFS)*: This method selects features that are highly correlated with the

pattern classes, but have low correlation with the remaining features. The subset evaluation function is given by:

$$Merit_S = \frac{\bar{k}_{ref}}{\sqrt{k+k(k-1)r_{ff}}} \quad (2)$$

where  $Merit_S$  is the merit of the selected feature set  $S$ ,  $\bar{k}_{ref}$  is the mean correlation between the features and class with  $f \in S$ , and  $r_{ff}$  is the mean pairwise feature correlation. The numerator expresses predictive capacity, while the denominator expresses feature redundancy.

a) *Best First Search (BF)*: Searches the space of feature subsets by greedy hillclimbing that may include backtracking. Best first may search forward, or backward, or, consider all possible single feature additions and deletions at a given point using a bi-directional strategy.

b) *Genetic Algorithm-based Search (GA)*: Genetic search works by having a population of variables representing feature sets and performs the operations of reproduction, cross-over and mutation in each generation to get the offspring that optimizes a feature set-related objective function.

2) *Information Gain (IG)*: This function measures the information gain with respect to the class:

$$\text{InfoGain}(\text{Class}, \text{Attribute}) = H(\text{Class}) - H(\text{Class}|\text{Attribute}) \quad (3)$$

where  $H$  is the entropy of each class given by  $H(\text{Class}) = -p_{\text{Class}} \log p_{\text{Class}}$ . We select the attributes by individual ranking evaluation.

### C. Classifiers-Discriminant Functions

1) *Naive Bayes (NB)*: This model assumes conditional statistical independence  $p(\mathbf{x}|\omega_j) = \prod_{k=1}^D p(x_k|\omega_j)$  where  $\mathbf{x} = (x_1, x_2, \dots, x_D)^T$  and  $D$  is the dimensionality of the feature space. The posterior probability is based on Bayes' formula. The MAP decision rule is typically used for classification.

2) *Multilayer Perceptron (MLP)*: A multilayer perceptron (MLP) is a feedforward artificial neural network system that maps input patterns onto class labels. An MLP has multiple layers of nodes that are fully connected to the next layer. Each node is a neuron with a nonlinear activation function. MLP utilizes backpropagation for supervised learning [11], [12]. Because MLP has multiple layers of logistic regression models, it can distinguish data that are not linearly separable. In learning by backpropagation -that can be considered as an extension of the LMS algorithm- we adjust the connection weights, according to the amount of error in the output compared to the expected result.

3) *Bayes Network (BN)*: A Bayes network, is a probabilistic graphical model that uses a directed acyclic graph to represent a set of random variables and their conditional dependencies. In a Bayesian network the joint probability density function can be written as the product of univariate conditional density functions dependent on their parent variables:

$$p(x) = \prod_{v \in V} p(x_v | x_{pa(v)}) \quad (4)$$

where  $pa(v)$  denotes the parents of  $v$ . In the graph, the parents are vertices directly connected to  $v$  by a single edge.

4) *Bagging*: For a training set  $S$  with size  $k$ , Bagging generates  $j$  training subsets denoted as  $S_i$  with size  $k' < k$ , by sampling from  $S$  uniformly and with replacement. We denote the original set as  $A$ . In the training stage, we first have  $D = \emptyset$  and  $j$  is the number of classifiers to train. Then for  $p = 1, 2, \dots, j$ , we take a bootstrap sample  $S_p$  from  $A$  to train classifier  $D_p$ . Then we add the classifier  $D_p$  to the current ensemble,  $D = D \cup D_p$ . We obtain the class label prediction for the input  $x$  by majority voting on the individual classifier decisions produced by  $D_1, \dots, D_j$  [14].

5) *Random Forests (RF)*: Random forests are an ensemble learning method that constructs multiple decision trees from subsets of the training set and uses random feature selection for node splitting. RF decide the class after applying voting to the predicted classes by the individual trees for classification, or by calculating the mean prediction for regression. Random forests address the overfitting tendency of the decision trees and have shown robustness with respect to noise [13].

## III. EXPERIMENTS

In this section, we report experiments on the TCB challenge data, which include healthy and osteoporosis bone radiograph images. The TCB challenge data is publicly available online at <http://www.univ-orleans.fr/i3mto/data>. The experimental procedures involving human subjects were approved by the Institutional Review Board of the institution that provides the data. The dataset includes annotated images of 58 healthy and 58 osteoporotic subjects that display a region of interest in the trabecular bone.

In the feature computation stage we obtained a total of 723 texture-related features as described in Section IIA. We selected features using the approaches described in Section IIB that is correlation-based feature selection with best first search that we denote by CFS-BF, correlation-based feature selection with genetic algorithm search that we denote by CFS-GA, and information gain (IG).

We tested the following classifier models: Naive Bayes (NB), Multilayer Perceptron (MLP), Bayes Net (BN), Random Forests (RF). The performance validation measures that we utilized are True Positive Rate (TPR), True Negative Rate (TNR), Classification Accuracy (ACC), and Area Under the ROC Curve (AUC).

In Table I we observe that CFS-BF yields an overall better performance than no-feature selection, CFS-GA and IG. CFS-BF selected 6 features, generated by DCT (#29 feature), Law's texture energy masks (#129 and #166 features), LBP (#561 and #599 features) and Gabor filter (#636 feature). This means that CFS-BF effectively selects discriminant features from the complete set; because it selects only a few features, training time is reduced and simple classifiers can produce acceptable results without being affected by the curse of dimensionality.

In our experiments we first evaluated NB, BN, RF, MLP and Bagging techniques in a 10-fold cross-validation framework, to identify the classification technique of top performance. We note that Bagging produced the best AUC, which

TABLE I  
CLASSIFICATION PERFORMANCE (10-FOLD CROSS-VALIDATION)

| FSM    | CL        | TPR         | TNR         | ACC         | AUC         | Dimension                             |
|--------|-----------|-------------|-------------|-------------|-------------|---------------------------------------|
| No     | NB        | 50          | 63.8        | 56.9        | 54.1        | 723                                   |
|        | BN        | 65.5        | 56.9        | 61.2        | 59.2        |                                       |
|        | RF        | 62.1        | 58.6        | 60.3        | 61.3        |                                       |
|        | Bagging   | 51.7        | 58.6        | 55.2        | 57.8        |                                       |
| CFS-GA | NB        | 53.4        | 62.1        | 57.8        | 56.6        | 138                                   |
|        | MLP       | 55.2        | 58.6        | 56.9        | 61.7        |                                       |
|        | BN        | 70.7        | 70.7        | 70.7        | 72.3        |                                       |
|        | RF        | 62.1        | 69          | 65.5        | 67.3        |                                       |
| IG     | NB        | 50          | 63.8        | 56.9        | 54.1        | 723                                   |
|        | BN        | 65.5        | 56.9        | 61.2        | 59.2        |                                       |
|        | RF        | 55.2        | 60.3        | 57.8        | 59.5        |                                       |
|        | Bagging   | 56.9        | 63.8        | 60.3        | 60.9        |                                       |
| CFS-BF | NB        | 70.7        | 60.3        | 65.5        | 72.5        | 6<br>(29,129,<br>166,561,<br>599,636) |
|        | MLP       | 62.1        | 74.1        | 68.1        | 71.9        |                                       |
|        | BN        | 72.4        | 65.5        | 69          | 74.8        |                                       |
|        | <b>RF</b> | <b>74.1</b> | <b>74.1</b> | <b>74.1</b> | <b>76.7</b> |                                       |
|        | Bagging   | 70.7        | 65.5        | 68.1        | <b>77.8</b> |                                       |

TABLE II  
CLASSIFICATION PERFORMANCE (LEAVE-ONE-OUT CROSS-VALIDATION)

| FSM    | CL        | TPR         | TNR         | ACC         | AUC       | Dimension                            |
|--------|-----------|-------------|-------------|-------------|-----------|--------------------------------------|
| No     | NB        | 50          | 62.1        | 56          | 56.2      | 723                                  |
|        | BN        | 69          | 62.1        | 65.5        | 66.3      |                                      |
|        | RF        | 56.9        | 62.1        | 59.5        | 64.1      |                                      |
|        | Bagging   | 55.2        | 62.1        | 58.6        | 62.3      |                                      |
| CFS-GA | NB        | 50          | 62.1        | 56          | 57.6      | 137                                  |
|        | MLP       | 50          | 60.3        | 55.2        | 60.3      |                                      |
|        | BN        | 77.6        | 67.2        | 72.4        | 76.7      |                                      |
|        | RF        | 65.5        | 60.3        | 62.9        | 65        |                                      |
| IG     | NB        | 50          | 62.1        | 56          | 56.2      | 723                                  |
|        | BN        | 69          | 62.1        | 65.5        | 66.3      |                                      |
|        | RF        | 51.7        | 55.2        | 53.5        | 61        |                                      |
|        | Bagging   | 60.3        | 63.8        | 62.1        | 64.9      |                                      |
| CFS-BF | NB        | 72.4        | 60.3        | 66.4        | 71.6      | 6<br>(29,129)<br>166,561<br>599,636) |
|        | MLP       | 55.2        | 69          | 62.1        | 65        |                                      |
|        | <b>BN</b> | <b>82.8</b> | <b>75.9</b> | <b>79.3</b> | <b>81</b> |                                      |
|        | RF        | 74.1        | 69          | 71.6        | 76.2      |                                      |
|        | Bagging   | 69          | 72.4        | 70.7        | 76        |                                      |

is 77.8%, although the corresponding ACC was at the level of 68.1%, but still about the same with other comparable methods. The best results were obtained by RF for 74.1%, whose AUC at 76.7% was outperformed only by Bagging. RF effectively improved ACC and AUC when using CFS-BF. Furthermore, we conducted leave-one-out cross-validation experiments reported in Table II. These experiments also indicate that CFS-BF yields better class separation than the other feature selection techniques. Among the tested classifiers, Bayes Network achieves the top performance with an ACC of 79.3% and AUC of 81%. These are promising results especially considering the significant statistical overlap of pixel intensity distributions between healthy and osteoporotic subjects. In Fig. 3 we display the ROC graphs comparing all approaches listed in Tables I and II. We observe that BN produced the largest area under the curve for the leave-one-out experiment and RF performed better than the other classifiers in 10-fold cross-validation.

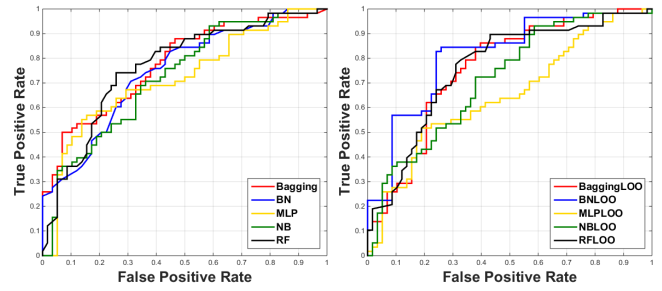


Fig. 3. ROC curves for 10-fold (left) and leave-one-out (right) cross-validation framework using CFS-BF, and NB, MLP, BN, RF and Bagging classifiers.

## IV. CONCLUSIONS

We introduced texture-based descriptors and classification techniques to separate healthy from osteoporotic groups using bone radiographs. Despite the fact that there were little to no visual differences between the two classes, the top performing techniques yielded 79.3% ACC and 81% area-under-the-curve of ROC. These results support the hypothesis that 2D texture analysis can contribute to identification of changes in trabecular bone microarchitecture. Hence the proposed system might be helpful for identification of individuals of high fracture risk and early intervention. Furthermore, we plan to test feature subdomain analysis techniques to improve the separability between the two classes.

## REFERENCES

- [1] R. Bartl and B. Frisch, *Osteoporosis: Diagnosis, Prevention, Therapy*, Second ed., Springer, 2009.
- [2] S. Hough, "Fast and slow bone losers: Relevance to the management of osteoporosis", *Drug Aging*, vol. 12, no. 1, 1998, pp. 1-7.
- [3] N.J. Macintyre and A.L. Lorbergs, "Imaging-Based Methods for Non-invasive Assessment of Bone Properties Influenced by Mechanical Loading", *Physiother Can.*, vol. 64, no. 2, 2012, pp. 202-215.
- [4] E. Martn-Badosa, A. Elmoutaouakkil, S. Nuzzo, D. Amblard, L. Vico and F. Peyrin, "A method for the automatic characterization of bone architecture in 3D mice microtomographic images", *Comput Med Imaging Graph*, vol. 27, no. 6, 2003, pp. 447-458.
- [5] F. Yger, "Challenge IEEE-ISBI/TCB: Application of Covariance matrices and wavelet marginals", *CoRR*, 2014.
- [6] R. Jennane, R. Harba, G. Lemineur, S. Bretteit, A. Estrade and C. L. Benhamou, "Estimation of the 3d self-similarity parameter of trabecular bone from its 2D projection", *Med Image Anal.*, vol. 11, no. 1, 2007, pp. 91-98.
- [7] R. Jennane, W. J. Ohley, S. Majumdar and G. Lemineur, "Fractal Analysis of Bone X-ray Tomographic Microscopy Projections", *IEEE Trans Medical Imaging*, vol. 20, no. 5, 2001, pp. 443-449.
- [8] A. P. Pentland, "Fractal-Based Description of Natural Scenes", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. PAMI-6, no. 6, 1984, pp. 661-674.
- [9] A.F. Costa, G. Humpire-Mamani, and A.J.M. Traina, *An Efficient Algorithm for Fractal Analysis of Textures*, SIBGRAPI 2012 ("XXV Conference on Graphics, Patterns and Images"), pp. 39-46.
- [10] L.G. Shapiro and G. Stockman, *Computer Vision*. Prentice Hall, 2001.
- [11] F. Rosenblatt, *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms*. Spartan Books, Washington DC, 1961.
- [12] D.E. Rumelhart, G.E. Hinton, and R.J. Williams, *Learning Internal Representations by Error Propagation, Parallel distributed processing: Explorations in the microstructure of cognition*, vol. 1, MIT Press, 1986, pp. 318-362.
- [13] L. Breiman (2001), "Random Forests", *Machine Learning*, vol. 45, no. 1, 2001, pp. 5-32.
- [14] R.O. Duda, P.E. Hart and D.G. Stork. *Pattern Classification*. Wiley-Interscience, 2000.