



Ελληνικό Μεσογειακό Πανεπιστήμιο

Σχολή Τεχνολογικών Εφαρμογών

Τμήμα Μηχανικών Πληροφορικής

Πτυχιακή εργασία

Τίτλος:

Κατηγοριοποίηση ιατρικών εικόνων με σύγχρονες τεχνικές τεχνητής όρασης και επεξεργασίας εικόνας./ *Classification of medical images with advanced computer vision and image processing techniques.*

Βασίλειος Μελησιανός (4229)

Επιβλέπων εκπαιδευτικός :Κωνσταντίνος Μαριάς

Επιτροπή Αξιολόγησης :

- Γιώργος Παπαδουράκης
- Μανώλης Τσικνάκης,
- Κωνσταντίνος Μαριάς

Ημερομηνία παρουσίασης : 20/09/2019

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές ευχαριστίες μου στον επιβλέποντα καθηγητή μου κ. Κωνσταντίνο Μαριά για τη βοήθεια, τις συμβουλές και τη καθοδήγηση του, που συνέλαβαν στην βελτίωση της εργασίας και στην επιθυμία περαιτέρω ενασχόλησης με τον κλάδο της έρευνας σε ιατρικές εικόνες. Ευχαριστώ επίσης τους υποψήφιους διδακτορικούς κ. Γιώργο Ιωαννίδη και κ. Λευτέρη Τριβιζάκη του εργαστηρίου Υπολογιστικής Βιοιατρικής του Ινστιτούτου Πληροφορικής του Ι.Τ.Ε. για τις άμεσες συμβουλές και ιδέες που βοήθησαν στην απόδοση μου όχι μόνο στην πτυχιακή αλλά και στον τρόπο ανάλυσης, μελέτης και κατανόησης πολύπλοκων δυσκολιών κατά την εκπόνηση της πτυχιακής μου περιόδου. Τέλος, θέλω να ευχαριστήσω την ομάδα του εργαστηρίου Υπολογιστικής Βιοιατρικής του Ινστιτούτου Πληροφορικής του Ι.Τ.Ε. για την όμορφη ατμόσφαιρα και άμεση βοήθεια για οτιδήποτε χρειαζόμουν.

Abstract

Mammogram-based diagnosis of breast cancer is hard especially in high tissue density areas of the breast. In addition, Additionally, breast density type has been linked with increased risk of breast cancer calling for automated breast-density scoring tools. In this thesis, we aim to automatically classify breast density for the purpose of assisting the clinician to assess high density breasts that potentially have hidden neoplasms. The proposed methodology involves a variety of advanced techniques such as Otsu thresholding, K-Means clustering etc., for image processing, Gabor Filters and LBP for texture analysis, HOG, SURF, and SIFT for orientation/salient point extraction analysis and machine learning algorithms (for feature selection (NCA), model training (LDA) and prediction(LDA)). The results of our method using several configurations are presented for the MIAS public mammogram database.

Περίληψη

Η διάγνωση καρκίνου του μαστού με βάση μαστογραφία είναι δύσκολη, ειδικά σε περιοχές με υψηλή πυκνότητα ιστού του μαστού. Επιπλέον, η κατηγορία πυκνότητας του μαστού έχει συνδεθεί με αυξημένο κίνδυνο εμφάνισης καρκίνου του μαστού που απαιτεί αυτοματοποιημένα εργαλεία βαθμολόγησης της πυκνότητας του μαστού. Σε αυτή τη διατριβή, στοχεύουμε στην αυτόματη ταξινόμηση της πυκνότητας του μαστού με σκοπό να βοηθήσουμε τον κλινικό για να αξιολογήσει μαστογραφίες υψηλής πυκνότητας που πιθανώς έχουν κρυμμένα νεοπλάσματα. Οι προτεινόμενη μεθοδολογία περιλαμβάνει μια ποικιλία προηγμένων τεχνικών όπως η κατάτμηση Otsu, η ομαδοποίηση K-Means κλπ., για την επεξεργασία εικόνων, τα φίλτρα Gabor και LBP για ανάλυση υφής, οι τεχνικές HOG, SURF και SIFT, για ανάλυση και εξόρυξη χαρακτηριστικών καθώς και αλγορίθμων μηχανικής μάθησης (Επιλογή χαρακτηριστικών (NCA), κατάρτιση μοντέλων (LDA), πρόβλεψη (LDA)). Τα αποτελέσματα της μεθόδου μας χρησιμοποιώντας διάφορους συνδυασμούς αλγορίθμων, παρουσιάζονται για την δημόσια βάση δεδομένων μαστογραφιών MIAS.

Table of Contents

Abstract	i
Περίληψη.....	ii
Table of Contents	i
List of Figures	iii
List of tables	v
1. Introduction	1
1.1 Research questions	1
1.2 Thesis outline	1
2. Breast density classification	3
2.1 Breast density classification methodology	3
3. State of the art	7
3.1 Public Databases	7
3.1.1 Mini-MIAS database	7
3.1.2 DDSM	7
3.2 Related work	7
4. Methodology for extracting breast density features	10
4.1 Feature extraction	10
4.1.1 HOG	10
4.1.2 LBP	14
4.1.3 SURF	16
4.1.4 SIFT	20
4.1.5 GABOR_LBP	24
4.2 Preprocessing	26
Image dimension processing	26
4.2.1 Background and patient's label information removal	26
4.2.2Pectoral muscle removal	27
4.3 Feature selection	28
4.4 Model training	29
4.4.1 LDA	29
4.5 Class prediction	29
4.5.1 LDA Prediction	30
5. Experimental results	31
5.1 Overall performance	34
6. Discussion	35
6.1 Future work	36

7. Bibliography 37

List of Figures

Figure 1 The first step is to preprocess the image so that only the breast remains.	3
Figure 2 .The next step is to extract features from the entire image and convert these features into numeric vectors.	4
Figure 3 The method splits into two different operations	4
Figure 4 .In this figure, the classification model is built based on the known categories that a dataset provides.	5
Figure 5.In prediction stage the theoretically and desired prediction is shown on the left part and the real prediction is depicted on the right side of it. False feature classifications often occur in real-life applications.....	5
Figure 6.In this picture the whole process of this thesis is depicted.	6
Figure 7.In this figure the derivative mask are shown in order to filter the image both horizontally and vertically.....	10
Figure 8. This figure depicts the application of the aforementioned masks. Vertical mask (left), Horizontal mask (right).	11
Figure 9.In this figure, the gradient image of figure 2 is presented.	11
Figure 10.In this figure,the image is divided into 20x20 cells in order to find histogram values in each one of them.....	12
Figure 11.In this figure’s cell, the gradient orientation and magnitude have been calculated in order to create the 9-bin histogram that will follow in Figure (12).	12
Figure 12.In this figure the angle value that is part of only one bin interval, its magnitude value, summed on the histogram bin. However, if the value of the angle belongs to two bin intervals (brown arrows indicating a pixel with magnitude of 16 and orientation of 120), then the magnitude value of this angle is separated into two equal values (equal to 8 each), and taken into consideration in both these two bins	13
Figure 13. In this figure, the frequency of the angles are shown.	13
Figure 14.In this figure, the concatenation of the 4 histograms is done by moving the red area 2 blocks horizontally. When there no more horizontal area to go, the area returns to its first position, and moves 1 block vertically. The process ends when the red area reaches the last 4 blocks of the image (bottom right).....	14
Figure 15.In this figure,the image is divided into cells.This preprocess is required for the following steps of the descriptor.....	15
Figure 16. In this figure, in every image’s cell the following process is performed: Each cell’s pixel is considered as a center pixel of a 3 × 3 neighborhood. This neighborhood has gray scale values that are compared with the center pixel in order to create a binary code which is then converted to decimal and replaces the value of the central pixel. Importance should be given to the expression of coding sequence which is clockwise or anticlockwise.....	15
Figure 17. In this figure, is shown the histogram of the decimal values of the cell that occurred during the neighborhood comparison in the previous section. This computation is done in all the cells.....	16
Figure 18.In this figure, the computation method of integral imaging is illustrated. Inside the rectangular region, the bottom-right pixel is the sum of all the rest parts of this region.....	17
Figure 19.In this figure the two pictures from the left ($L_{xxx}, \sigma, L_{xy}(x, \sigma)$) are calculated with the second order partial derivative of Gaussian and the result is on the two pictures to the right. Pixels which are positive represent a local maximum at x , the negative ones represent the local minimum at x and the gray ones are not inconclusive (image from [31]).	17
Figure 20.In this figure, the filter and interest point locations process is depicted (image from [31]).	18
Figure 21.Here is an example of how the interest points decay from growth of filter size in image (figure 6) (image from [31]).	18
Figure 22.In this figure, the dominant orientation that generated by summing the responses within a sliding orientation window of size $\pi 3$ is depicted (image from [31]).	19

Figure 23. In this figure, a square region is constructed around the key point. The line from the center of each square, represents the dominant orientations of its point (image from [14]).19

Figure 24. In this figure, the Haar-wavelet responses are computed (sums of dx, dx, dy, dy) in order to find the dominant orientation of each sub-region (image from [31]).20

Figure 25. In this figure, each image with same size belongs to an octave which there is progressively blurred (from left to right).21

Figure 26. In this figure, the (DoG) images are generated, by finding the difference from 2 images at a time. This process is done in all the available octaves.21

Figure 27. In this figure, the marked pixel ‘X’ is compared with its neighbors from its image, the one above it and one below it.22

Figure 28. In this figure, using 3 by 3 images from the DoG set, the extreme points are detected. Then using the Taylor expansion, we find the Location and Scale of Key points in each image. After that, in order to find principal curvatures, edge rejection takes place with Hessian matrix. The result is shown on the set with name “Principal Curvatures”.23

Figure 29. This histogram represents the degrees in 36 subsets. Each subset is increased by the presence frequency of a degree that a subset has (image from ([1])).23

Figure 30. In this figure each 4x4 sub-block’s direction values are inserted into 8-bin orientation histogram.24

Figure 31. In this figure we see the whole migration of the degree values into 16 8bin orientation histograms (image from [1]).24

Figure 32. In this figure, the entire process of this combination is shown. The clinical Image(left) is filtered by the optimal parameters of Gabor’s filter (middle) and then processed by the optimal values of LBP’s function (right).25

Figure 33. The procedure of cropping and resizing, resulted on a much smaller image size, making the overall calculation time shorter.26

Figure 34. Otsu’s method is to find the optimal threshold for the grayscale frequencies of the cropped image.27

Figure 35. This image describes the conversion of a cropped image (left) to a preprocessed image (right).27

Figure 36. In this picture the failed attempt of pectoral muscle removal is shown and how breast image is affected by the different operations. The selected image from these 3, was the middle one.28

Figure 37. This figure show the comparison between the three pectoral muscle approaches to define the best solution.28

Figure 38. In this figure the classification method is depicted. As shown in picture (A), a line crosses the labelled data for the purpose of representation each data into the line. Then, the average (mean) of each class data is computed which creates a value for each class (picture B). This value is used to measure the distance between the three classes for computing the first step which is known as between-class variance (picture C). Next a distance is for each labelled data is measured between the data’s sample distance and class’s mean known as within-class variance (picture D). The final step, combines the previous two processes resulting both the maximum distance between each class and minimum distance from each data with its center (picture E).29

Figure 39. In this figure, the prediction method is shown. The test data that have been generated, are placed according to their values in the classification area (picture B). In this case of having 3 classes, the process begins with taking one class and combine the other classes into one in order to have 2 classes every time (picture C1, C2, C3). The computation of this prediction refers to the Bayes theorem that estimates the likelihood probability of each test data compared to the classification classes. After that computation, the resulting prediction comes from mixing the previous processes (picture C1, C2, C3) into a new classification model (picture D).30

Figure 40. In this picture the cross validation and combination method between k-fold and hold-out method is depicted. This figure repeats as many times as the k in the k-fold.31

List of tables

Table 1. A concise view of relevant prior work in mammogram density classification9
Table 2. In this table 2 or 3 class results of different feature extraction algorithm configurations depicted.34
Table 3. Performance metrics for the several algorithms/configuration of our predictive methodologies for breast density characterization as well as from relevant works.35

1. Introduction

Breast cancer is a devastating disease that affects millions of women worldwide [1]. There are many published works in the field aiming to automatically assess breast density [2]. Despite of the significant progress made, suspicious tumors can possibly be masked by large portions of tissue density. The consequence of this is an increased number of tumors that aren't diagnosed, leading to metastases and patient's death. Breast density in itself is a risk factor for breast cancer appearance [3]. Most of the recent works classify density according to the BI- RADS [4] breast's density classification scheme. These categories are (i) Almost entirely fatty which indicates that the breasts are almost entirely composed of fat, (ii) Scattered areas of fibro-glandular density that there are some scattered areas of density, but the majority of the breast tissue is non-dense, (iii) Heterogeneously dense shows some areas of non-dense tissue, but that the majority of the breast tissue is dense and (iv) Extremely dense illustrates that nearly all of the breast tissue is dense. The possibility of neoplasms appearance in a patient's breast, is related to the BIRAD's density category. According to Byrne C et al [5], heterogeneously dense and extremely dense breasts have the highest risk of breast cancer. As noted by W. Duncan, breast cancer is curable if it can be diagnosed in early stages [6]. These concepts are central for our motivation to work in this area in this thesis aiming to develop and evaluate machine learning density classification techniques. The main purpose of this thesis is to assess the value of feature-based machine learning classification using public mammographic datasets.

1.1 Research questions

As already stated, the present work is to develop an automated computational framework for breast density characterization based on mammogram image features. Our attention is focused on the optimal configuration of a set of parameters to maximize system performance using the optimal method combination. The following questions are posed and investigated in this work:

- Can we achieve good classification of breast density based on robust feature extractors using public mammography data?
- Is pre-processing and feature selection essential for boosting the performance?
- Which feature extractor extracts the most valuable features for the implemented machine learning classifier?
- Which of the implemented machine learning classifiers provides the best performance?
- Is traditional machine learning enough to solve the problem?

1.2 Research publications

Additionally, part of the thesis is included in journal publication [7].

1.3 Thesis outline

This thesis comprises six chapters, including this introduction. Chapter 2 provides the explanation of both tools and method steps that are included in Breast tissue density classification application. Literature review related to automatic breast density classification is presented in section 3. Also, the relevant implemented algorithms are presented in chapter 3. Chapter 4 discusses the selected dataset, parameters of each algorithm, the types of cross validation that has been selected and presents in detail each step of the methodology. The experimental results, the different scores between machine learning

algorithms are presented in section 5. The final chapter presents a summary of the contributions of this thesis and the related future research directions are discussed.

2. Breast density classification

As was pointed out in the introduction to this thesis, breast density classification is critical for the clinical assessment of women undergoing breast cancer screening not least due to the fact that it has been associated with the risk of developing breast cancer [3]. To this end, our objective is to develop an automated breast density classification system. Our approach is based on the extraction of salient features of the breast parenchyma coupled with machine learning through the use of MATLAB. In summary, the proposed technique is in essence an image classification application that takes as input several salient mammogram image features which are used to train a breast density classification model. Based on this model, image features are extracted from the test dataset and are classified in order to evaluate the performance of the model. The proposed methodology is presented in the next sections.

2.1 Breast density classification methodology

The first step of this method is to preprocess each image for the purpose of the extraction of the region of interest. In essence this means that we need to remove any labels, background and in the case of medio-lateral mammograms, the pectoral muscle which appears as a homogenous, bright area in the top right corner of medio-lateral mammograms. This background information must be therefore removed by an image processing/computer vision pipeline so that the classification/prediction scores are not affected by irrelevant image areas or features. The techniques used in this thesis are Otsu's method [8] for background thresholding, K-means [9] with a combination of connected components techniques in the pectoral muscle area [10] for muscle segmentation and breast boundary detection [11] for label removal Figure (1).

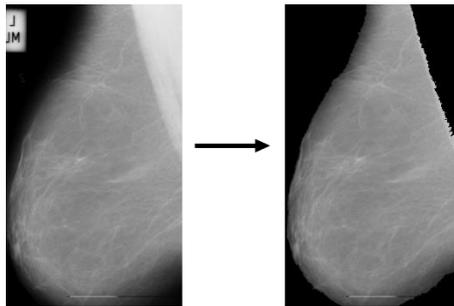


Figure 1 The first step is to preprocess the image so that only the breast remains.

Once this step is complete, feature extraction takes place. Feature extraction is a procedure based on the application of an algorithm on an image, aiming to detect/extract interesting points, lines or areas such as edges or salient/invariant points. This procedure can be done globally (whole image), or locally (region of interest). Since the size of the breast may differ significantly, we chose to extract features globally, from the entire image, in order to acquire same feature size. The algorithms used in our proposed methodology for feature extraction are: Histogram of Oriented Gradients (HOG) [12], Scale Invariant Feature transform (SIFT) [13], Speed Up Robust Features (SURF) [14], Local Binary Pattern (LBP) [15] and a combination of Gabor Filters and LBP [16] Figure (2).

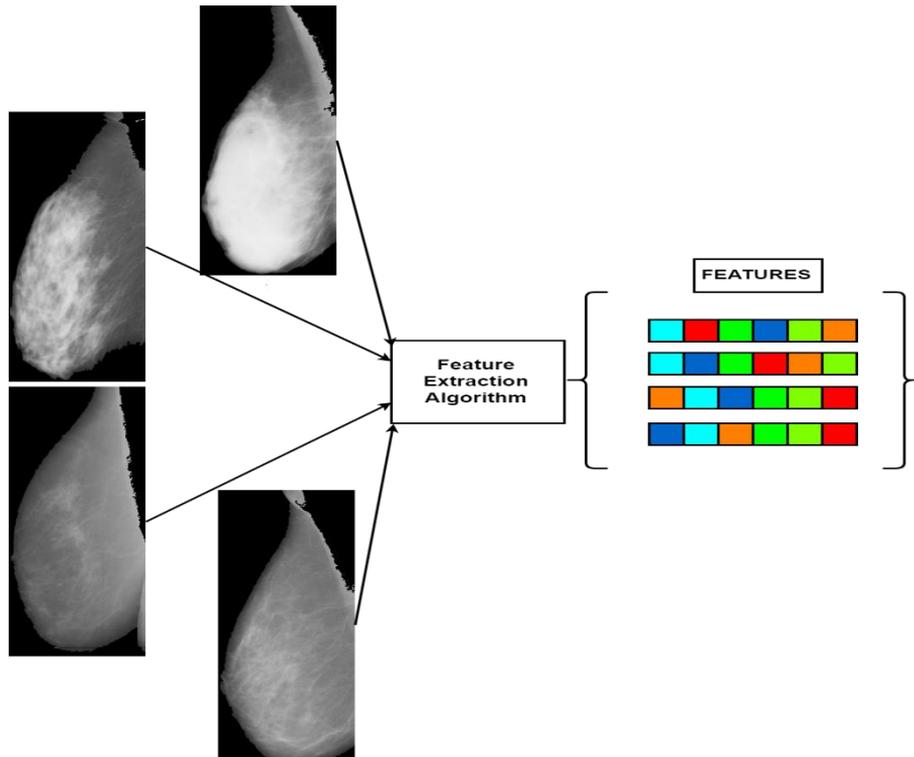


Figure 2 .The next step is to extract features from the entire image and convert these features into numeric vectors.

Before moving on building the classification model, the proposed method was tested with two feature extraction strategies. In the first one, the method uses the whole set of each image’s features, while in the second one, a feature reduction algorithm is applied (NCA) [17] resulting, theoretically, in a reduced but robust set of features Figure(3).

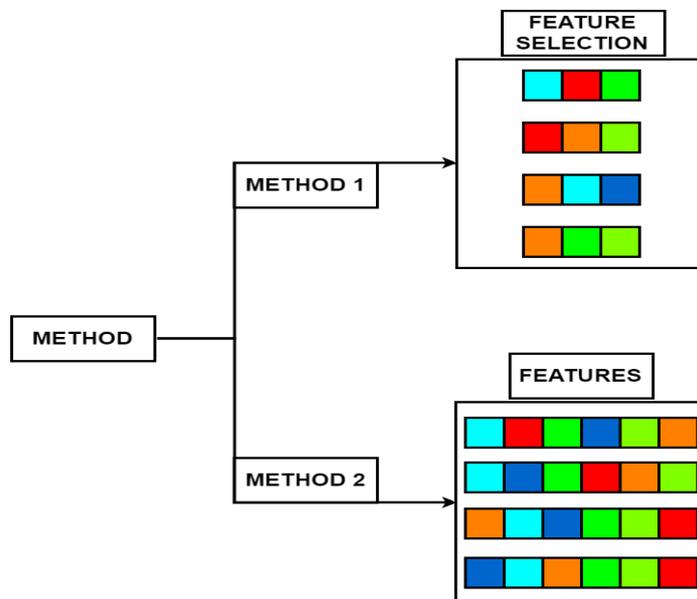


Figure 3 The method splits into two different operations

Afterwards, the calculated values that have been generated for each mammogram are divided into three parts: training set, validation set and test set. In our work the dataset is divided as follows: 64.4% training set, 15.4% validation set, and 20% test set.

Classification training can be either a supervised or unsupervised process. A supervised method, in contrast with unsupervised ones, works based on known categories/labels. So, according to the classes we have, the training features are distributed in these groups, generating a classification model. In this thesis, Linear Discriminant Analysis (LDA) [18] was used as training model Figure (4).



Figure 4 .In this figure, the classification model is built based on the known categories that a dataset provides.

The final phase of this method is to predict the breast density category of each image of the test data. This procedure requires the training model, that has been constructed previously, and the extracted image features of the test data. When trying to assign the unclassified numeric values into a category according to the established criteria, erroneous classifications are possible. Machine learning can improve the efficiency of such classification systems [19] Figure (5).

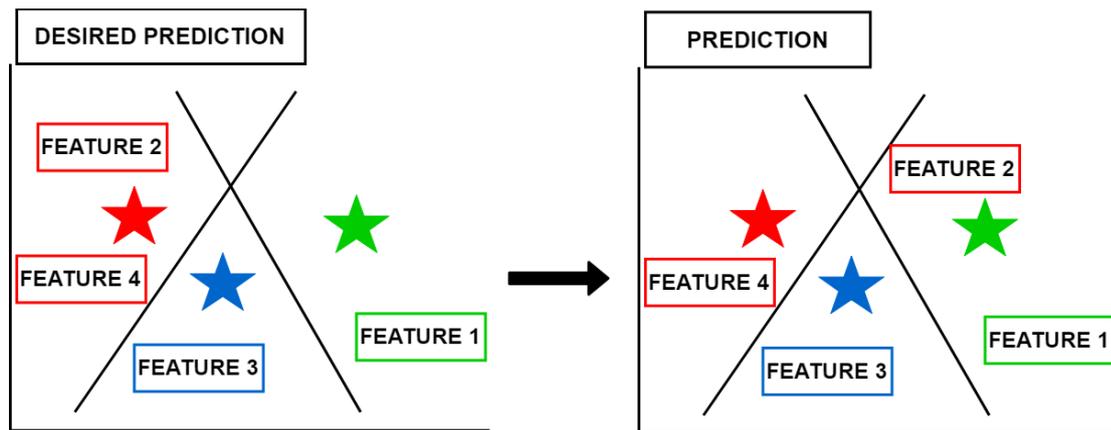


Figure 5.In prediction stage the theoretically and desired prediction is shown on the left part and the real prediction is depicted on the right side of it. False feature classifications often occur in real-life applications.

With the intention of evaluating the performance of our breast density classification method, we tested our method in two related public datasets. The difference between these two was that one of them had already removed the pectoral muscle area from the mammograms Figure (6).

Categorization of medical images with modern techniques of visual and image processing methods
Vasileios Melissianos

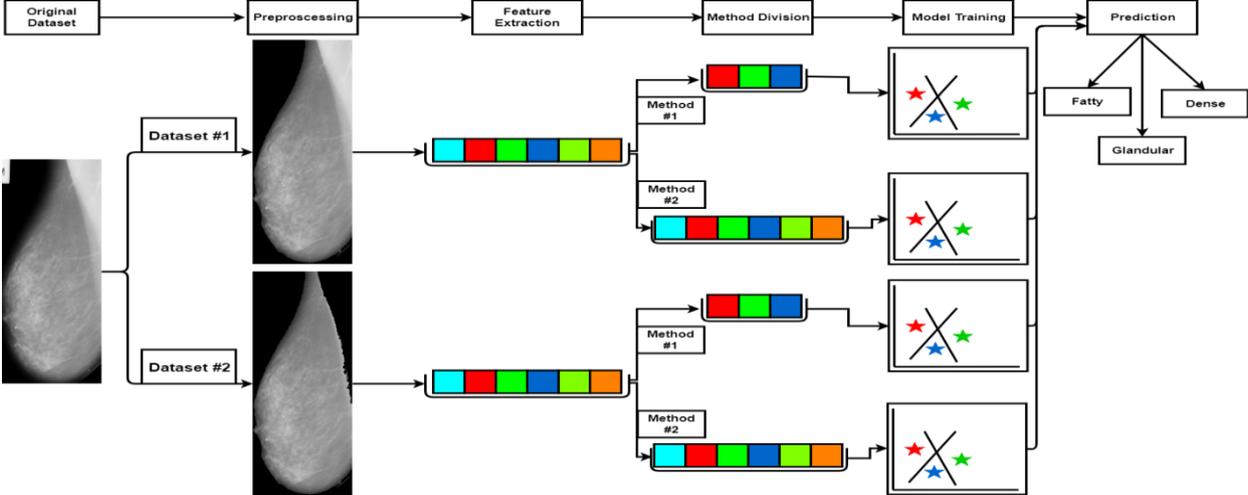


Figure 6. In this picture the whole process of this thesis is depicted.

3. State of the art

The link between increased breast cancer risk in women with dense tissue breasts, highlights the importance of automated systems for objective breast density classification. There are many previous efforts to develop and evaluate such systems in different databases.

3.1 Public Databases

3.1.1 Mini-MIAS database

The selected dataset of the thesis is the mini-MIAS (<http://peipa.essex.ac.uk/info/mias.html>) which contains a set of mammograms. This dataset is free for scientific research and consists of 161 pairs of medic-lateral oblique (MLO) view mammograms. The base of this set is the original MIAS Database but digitized at 50-micron pixel edge and reduced to 200-micron pixel edge and cropped so that each image becomes 1024×1024 pixels. The grayscale images of this database are from the United Kingdom National Breast screening program. They are annotated according to their breast density by expert radiologists, using three distinct classes: Fatty (F) (106 images), Fatty-glandular (G) (104 images) and Dense-glandular (D) (112 images).

3.1.2 DDSM

Digital Database Screening mammography (DDSM) (<http://www.eng.usf.edu/cvprg/Mammography/Database.html>) is a widely used open source database that comprises a set of large set of mammograms. In more details, this dataset includes approximately 2,500 studies. Each study includes a patient's two pairs of mammographic information (two mediolateral breast images and two craniocaudal ones). These pairs hold valuable information such as BI-RADS tissue category and clinical breast status (normal, benign, malignant). The most used images are the mediolateral images which give more mammography details. This grayscale set of images comes from the collaborative effort of Massachusetts General Hospital, the University of South Florida and Sandia National Laboratories.

3.2 Related work

For example, Keir Bovis and Sameer Singh [20] have used the technique of Chandrasekhar and Attikiouzel [21] for breast and background segmentation, Spatial Grey Level Dependency (SGLD) for feature extraction, PCA [22] for dimensionality reduction and Artificial Neural Networks (ANN) for prediction in 377 DDSM images.

Peter Miller and Sue Astley [20] applied Laws texture energy method [23] in order to find the texture energy of each image. Once the features were generated, a Bayesian classifier performs leave-one-out cross validation method to predict the remaining test data in a set of 40 images.

Stylianios Tzikopoulos et al.[24] [25], applied both image segmentation (breast boundary detection, pectoral muscle segmentation, nipple detection) and first order statistics as a preprocessing step. Then, Classification and Regression Trees (CARTs) algorithm took place for the leave one out classification using the mini-MIAS dataset.

Oliver et al used a mix of morphological and texture features for each one of 300 DDSM images. Then the relative area, the center of masses and the medium intensity are calculated from both clusters. After the calculation of contrast, energy, entropy, correlation and many other metrics, these values are fed into a histogram with the aim of finding related values from other images. The prediction was established with the application of three different algorithms: k-NN classifier, ID3 decision tree and a combination of these two.

Styliani Petroudi and Michael Brady segmented 32 mammogram images into texture regions with the utilization of textons in a Hidden Markov Random Field(HMRF).The extracted textures were separated/classified according to their textures.

Styliani Petroudi, Timor Kadir and Michael Brady [26] focused their attempt in 132 DDSM mammograms using texture classification, with the intention of determining a texture category as statistical allocation over texton type values. Classification was applied by calculating the measure between two histograms including texton values.

Bethany Percha et al. [27] developed a natural language mammogram processing method in 500 Stanford mammogram images and 100 Marshfield ones.

Boulenger et al. [28] edited and used the entire set of both mini-MIAS and DDSM, segmented the breast (label removal, pectoral muscle removal, subsampling), and classified the density of all the images through the use of tree classification.

The table below (Table I), summarizes the results about the state-of-the-art classification methods dealing with the problem of breast density classification with different databases. The results are expressed either in accuracy rate or in confusion matrices.

Author	Method	Results	Dataset																																													
K. Bovis et al.	SGLD matrices + PCA Artificial NN	71.4% 4 class problem 96.6% 2 class problem	DDSM 377 img																																													
Peter Miller et al.	Laws-Texture analysis FRR texture energy	80% 2 class problem(Fatty-Glandular)	40 img																																													
S. Tzidakopoulos et al.	Classification and Regression Trees (CARTs) Leave one out	<table border="1"> <thead> <tr> <th>Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th>Predicted class</th> <th>F</th> <th>G</th> <th>D</th> <th></th> </tr> </thead> <tbody> <tr> <td>F</td> <td>95</td> <td>5</td> <td>1</td> <td></td> </tr> <tr> <td>G</td> <td>11</td> <td>89</td> <td>19</td> <td></td> </tr> <tr> <td>D</td> <td>0</td> <td>10</td> <td>92</td> <td></td> </tr> </tbody> </table>	Breast density	Truth class				Predicted class	F	G	D		F	95	5	1		G	11	89	19		D	0	10	92		mini-MIAS 322 img																				
Breast density	Truth class																																															
Predicted class	F	G	D																																													
F	95	5	1																																													
G	11	89	19																																													
D	0	10	92																																													
S. Tzidakopoulos et al.	Regression Trees (CARTs)	<table border="1"> <thead> <tr> <th>Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th>Predicted class</th> <th>F</th> <th>G</th> <th>D</th> <th></th> </tr> </thead> <tbody> <tr> <td>F</td> <td>88</td> <td>21</td> <td>7</td> <td></td> </tr> <tr> <td>G</td> <td>10</td> <td>52</td> <td>28</td> <td></td> </tr> <tr> <td>D</td> <td>8</td> <td>31</td> <td>77</td> <td></td> </tr> </tbody> </table>	Breast density	Truth class				Predicted class	F	G	D		F	88	21	7		G	10	52	28		D	8	31	77		mini-MIAS 322 img																				
Breast density	Truth class																																															
Predicted class	F	G	D																																													
F	88	21	7																																													
G	10	52	28																																													
D	8	31	77																																													
A. Oliver et al.	k-NN classifier 40.3%	<table border="1"> <thead> <tr> <th>Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th>Predicted class</th> <th>I</th> <th>II</th> <th>III</th> <th>IV</th> </tr> </thead> <tbody> <tr> <td>I</td> <td>17</td> <td>24</td> <td>5</td> <td>4</td> </tr> <tr> <td>II</td> <td>27</td> <td>35</td> <td>30</td> <td>8</td> </tr> <tr> <td>III</td> <td>3</td> <td>25</td> <td>54</td> <td>18</td> </tr> <tr> <td>IV</td> <td>3</td> <td>9</td> <td>23</td> <td>15</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th>Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th>Predicted class</th> <th>I</th> <th>II</th> <th>III</th> <th>IV</th> </tr> </thead> <tbody> <tr> <td>I</td> <td>24</td> <td>20</td> <td>3</td> <td>3</td> </tr> </tbody> </table>	Breast density	Truth class				Predicted class	I	II	III	IV	I	17	24	5	4	II	27	35	30	8	III	3	25	54	18	IV	3	9	23	15	Breast density	Truth class				Predicted class	I	II	III	IV	I	24	20	3	3	DDSM 300 img
Breast density	Truth class																																															
Predicted class	I	II	III	IV																																												
I	17	24	5	4																																												
II	27	35	30	8																																												
III	3	25	54	18																																												
IV	3	9	23	15																																												
Breast density	Truth class																																															
Predicted class	I	II	III	IV																																												
I	24	20	3	3																																												

	ID3 decision tree 43.3%	<table border="1"> <tbody> <tr> <td>II</td> <td>34</td> <td>49</td> <td>10</td> <td>7</td> </tr> <tr> <td>III</td> <td>18</td> <td>17</td> <td>40</td> <td>25</td> </tr> <tr> <td>IV</td> <td>9</td> <td>5</td> <td>19</td> <td>17</td> </tr> </tbody> </table> <table border="1"> <thead> <tr> <th colspan="2">Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th rowspan="5">Predicted class</th> <th></th> <th>I</th> <th>II</th> <th>III</th> <th>IV</th> </tr> </thead> <tbody> <tr> <th>I</th> <td>23</td> <td>21</td> <td>3</td> <td>3</td> </tr> <tr> <th>II</th> <td>32</td> <td>48</td> <td>12</td> <td>8</td> </tr> <tr> <th>III</th> <td>3</td> <td>22</td> <td>52</td> <td>23</td> </tr> <tr> <th>IV</th> <td>5</td> <td>7</td> <td>20</td> <td>18</td> </tr> </tbody> </table>	II	34	49	10	7	III	18	17	40	25	IV	9	5	19	17	Breast density		Truth class				Predicted class		I	II	III	IV	I	23	21	3	3	II	32	48	12	8	III	3	22	52	23	IV	5	7	20	18	
II	34	49	10	7																																														
III	18	17	40	25																																														
IV	9	5	19	17																																														
Breast density		Truth class																																																
Predicted class		I	II	III	IV																																													
	I	23	21	3	3																																													
	II	32	48	12	8																																													
	III	3	22	52	23																																													
	IV	5	7	20	18																																													
	k-NN+ID3 combination 47%																																																	
S. Petroudi et al.	Markov Random Fields	Dence:96.7% Fatty:93.8% Breast edge:92.8%	32 img																																															
S. Petroudi et al.	Texton dictionaries	<table border="1"> <thead> <tr> <th colspan="2">Breast density</th> <th colspan="4">Truth class</th> </tr> <tr> <th rowspan="3">Predicted class</th> <th></th> <th>I</th> <th>II</th> <th>III</th> <th>IV</th> </tr> </thead> <tbody> <tr> <th>4 class</th> <td>91</td> <td>64</td> <td>70</td> <td>78</td> </tr> <tr> <th>2 class</th> <td>91</td> <td></td> <td>94</td> <td></td> </tr> </tbody> </table>	Breast density		Truth class				Predicted class		I	II	III	IV	4 class	91	64	70	78	2 class	91		94		DDSM 132 img																									
Breast density		Truth class																																																
Predicted class		I	II	III	IV																																													
	4 class	91	64	70	78																																													
	2 class	91		94																																														
Bethany Percha et al.	natural language processing.	Satnford-99.8% Marhfield-99.9%	Stanford 500 img Marshfield 100 img																																															
Boulenger et al.	Segmentation + tree building	<table border="1"> <thead> <tr> <th colspan="2">Breast density</th> <th colspan="3">Truth class</th> </tr> <tr> <th rowspan="4">Predicted class</th> <th></th> <th>F</th> <th>G</th> <th>D</th> </tr> </thead> <tbody> <tr> <th>F</th> <td>92</td> <td>15</td> <td>3</td> </tr> <tr> <th>G</th> <td>8</td> <td>60</td> <td>17</td> </tr> <tr> <th>D</th> <td>5</td> <td>28</td> <td>92</td> </tr> </tbody> </table>	Breast density		Truth class			Predicted class		F	G	D	F	92	15	3	G	8	60	17	D	5	28	92	mini-MIAS 320 img																									
Breast density		Truth class																																																
Predicted class		F	G	D																																														
	F	92	15	3																																														
	G	8	60	17																																														
	D	5	28	92																																														

Table 1. A concise view of relevant prior work in mammogram density classification

It is obvious from the above literature review summary table that performances in mammogram classification drop dramatically in the four (BIRADS) tissue class classification problem compared to the two class classification problem (dense vs. non-dense). Generally, in the bibliography there are many available datasets such as the mini-MIAS, DDSM, AMDI, MammoGrid etc. The oncoming chapter analyzes not only the optimal parameters and its algorithms but the whole methodology behind the breast density classification system.

4. Methodology for extracting breast density features

This chapter provides a descriptive workflow for each feature extractor including any other algorithm's pre-processing steps that were applied in breast density classification, as well as their optimal parameters that were selected via exhaustive parameter analysis in the training model stage between training set and validation set.

4.1 Feature extraction

One of the main parts of breast density classification, is feature extraction. Feature extraction in image processing, is a procedure that uses image's RGB or Grayscale values and converts them into numerical values.

4.1.1 HOG

The histogram of oriented gradients (HOG) [12] is a feature extraction methodology used in computer vision and image processing widely used for object detection. The method is based on computing the frequencies of gradient orientation in specific, local parts of the image and aims to improve accuracy by using overlapping local contrast normalization. In the next pages, the main parts of the algorithm are presented (Gradient computation, Orientation binning, Block normalization, Feature extraction).

Part1. Gradient computation

Algorithm's first phase corresponds to image's gradient values reckoning. The most frequent process is by filtering the image horizontally and vertically at a time, with a mask of one-dimension Figure (7).

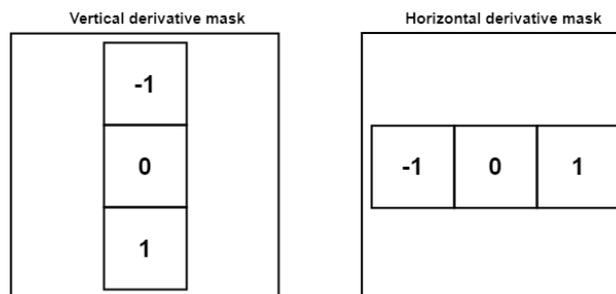


Figure 7. In this figure the derivative mask are shown in order to filter the image both horizontally and vertically.

With the use of vertical derivative mask, vertical textures of the image are shown and with horizontal mask, horizontal information appears. The result of applying these kernels in a mammogram is shown in Figure (8).



Figure 8. This figure depicts the application of the aforementioned masks. Vertical mask (left), Horizontal mask (right).

Assuming g_x and g_y to be the horizontal and vertical previously computed kernel images respectively, the gradient image is vector comprising g_x and g_y . The gradient magnitude (GM) [29] is the result of $GM = \sqrt{g_x^2 + g_y^2}$ and gradient's orientation [30] is calculated by $\theta = \arctan \frac{g_y}{g_x}$. Figure (9) shows the sum of g_x and g_y .

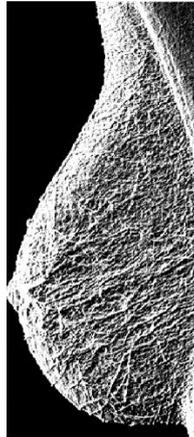


Figure 9. In this figure, the gradient image of figure 2 is presented.

Part2. Orientation binning

Once the gradient orientation and gradient magnitude are calculated, a generation of cell histograms follows. The image is divided into $n \times n$ cells. As an example, let us assume $n = 20$. These pixels within the cell, cast a weighted vote for an orientation-based histogram from 0 to 180 degrees, based on the values found in the gradient image Figure (10). Further on, each 20×20 patch of the image, generates two matrices, the gradient orientation and the gradient magnitude which are needed for the purpose of generating the histogram values. Figure (11).

Categorization of medical images with modern techniques of visual and image processing methods
 Vasileios Melissianos

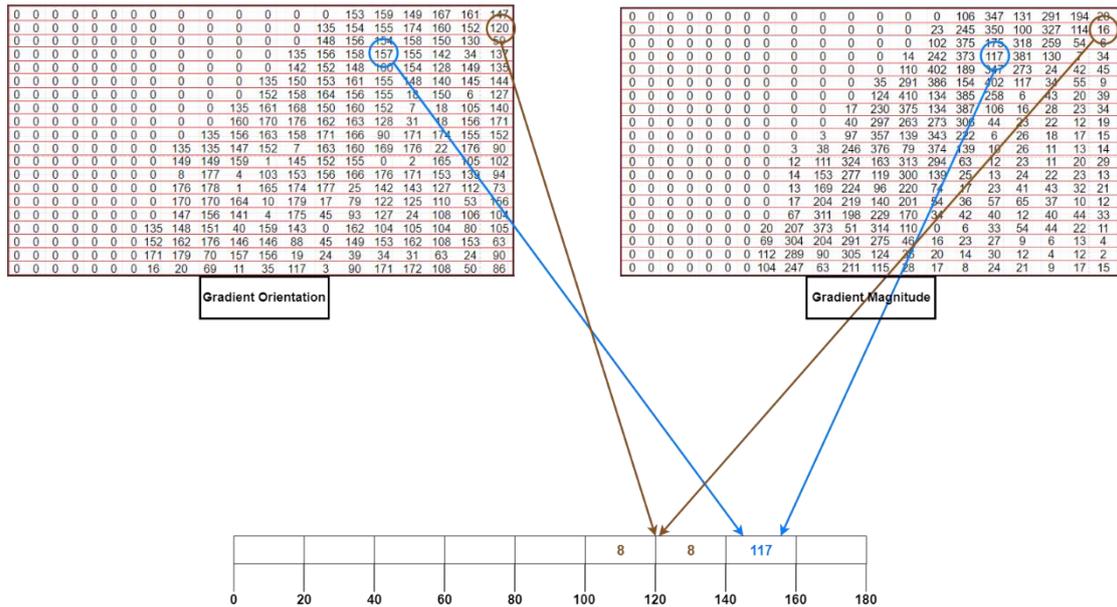


Figure 12. In this figure the angle value that is part of only one bin interval, its magnitude value, summed on the histogram bin. However, if the value of the angle belongs to two bin intervals (brown arrows indicating a pixel with magnitude of 16 and orientation of 120), then the magnitude value of this angle is separated into two equal values (equal to 8 each), and taken into consideration in both these two bins.

The sequence of this process generates a histogram which shows the frequency values of the direction in this 20 x 20 cell Figure (13).

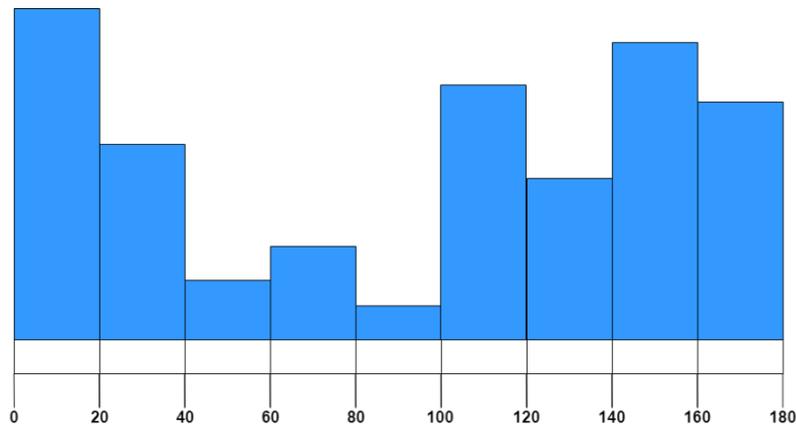


Figure 13. In this figure, the frequency of the angles are shown.

Part3. Block Normalization

In the previous step, a histogram based on the gradient of the image was created. Now, the next step is to normalize 4 histograms each time which means 4 blocks of the image. These 4 histograms are concatenated into a 36 x 1 vector and it can be normalized using the L2-norm [1]. This normalization refers to the first 4 block section of the image instead all of them. So, this calculation repeats itself every time when the calculation area moves 2 blocks at a time Figure (14).

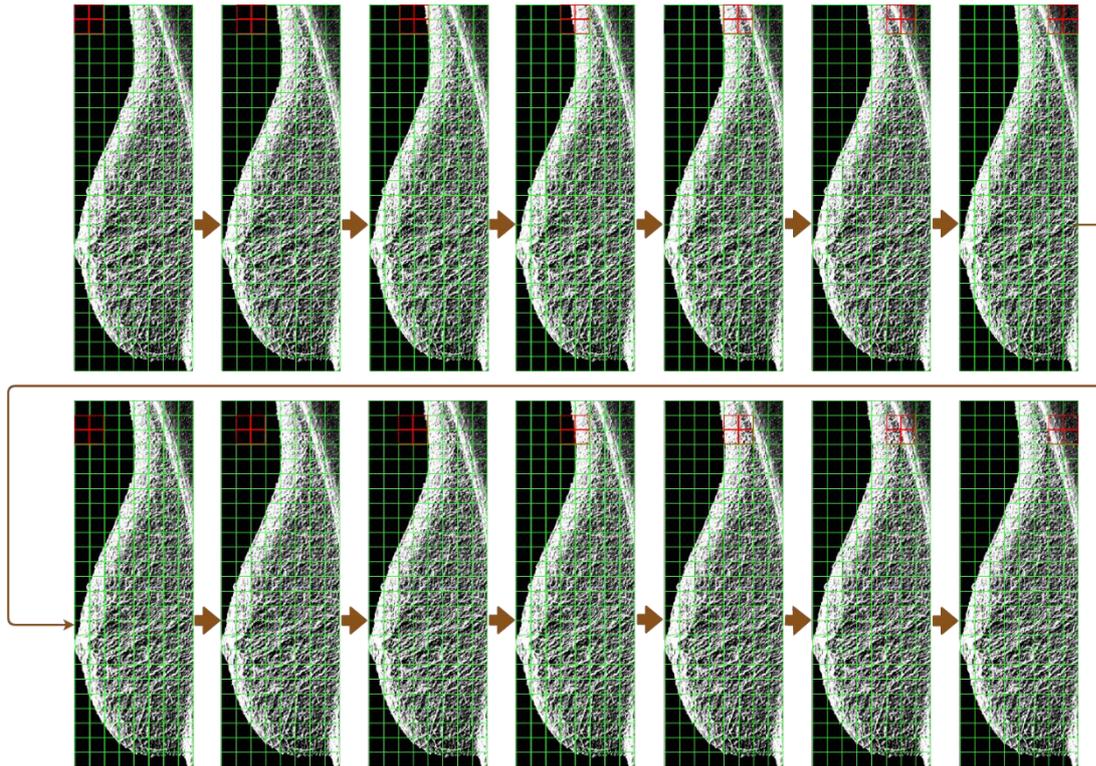


Figure 14. In this figure, the concatenation of the 4 histograms is done by moving the red area 2 blocks horizontally. When there no more horizontal area to go, the area returns to its first position, and moves 1 block vertically. The process ends when the red area reaches the last 4 blocks of the image (bottom right).

Part4. Feature Extraction

When all the blocks of the image are normalized, the 36 vectors of each 4x4 block are concatenated into one vector. The size of the vector is found by multiplying the horizontal and vertical 4x4 blocks of the image with the 36x1 normalized histogram of each area. The result is a $7 \times 25 \times 36 = 1 \times 6300$ dimensional vector.

After exhaustive analysis which was executed through the classification process with training and validation set, the optimal input parameters of this extractor algorithm were :

- Cell size=4
- Overlap division=2
- Block size=4
- Number of bins=15

4.1.2 LBP

The Local Binary Pattern (LBP) [15] is a type of visual descriptor used for classification in computer vision. It is an effective texture descriptor for images which computes a *local representation* of texture. This local representation is constructed by comparing each pixel with its surrounding neighborhood of pixels which is necessary, in order to provide a feature vector of the image based on concatenated histograms. This algorithm has three main parts: Neighborhood comparison, Histogram calculation, Feature extraction.

Part1. Neighborhood comparison

In the first section of the algorithm, the image needs to be converted from RGB into gray scale image (if necessary) for the purpose of applying this algorithm. Then, the image is divided into $n \times n$ cells. Let's say the cell division is 20×20 which is shown below Figure (15).

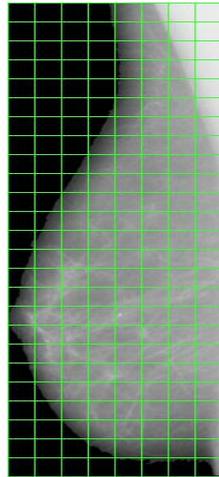


Figure 15. In this figure, the image is divided into cells. This preprocess is required for the following steps of the descriptor.

After applying the division, the center of each pixel's neighborhood of size (r) and points (p) is generated. Let's say the size of (r = 3) which means a neighborhood of 3 × 3 and points (p = 8). First, each neighborhood needs to be expressed as a binary sequence and then as a decimal value. To accomplish that each center pixel is compared with its neighbors, the comparison takes the intensity value of the pixels. So, if the intensity of the center pixel is greater-than (>) the intensity of its neighbor, then the value is set to 0 otherwise (<=), is set it to 1. The calculation begins with a start from any neighboring pixel and calculate clockwise or counterclockwise, but the ordering must be kept consistent for all pixels in the image Figure (2). In each instance, a binary code is created from the above values which is then converted to decimal and replaces the value of the central pixel, as illustrated in Fig. (16).

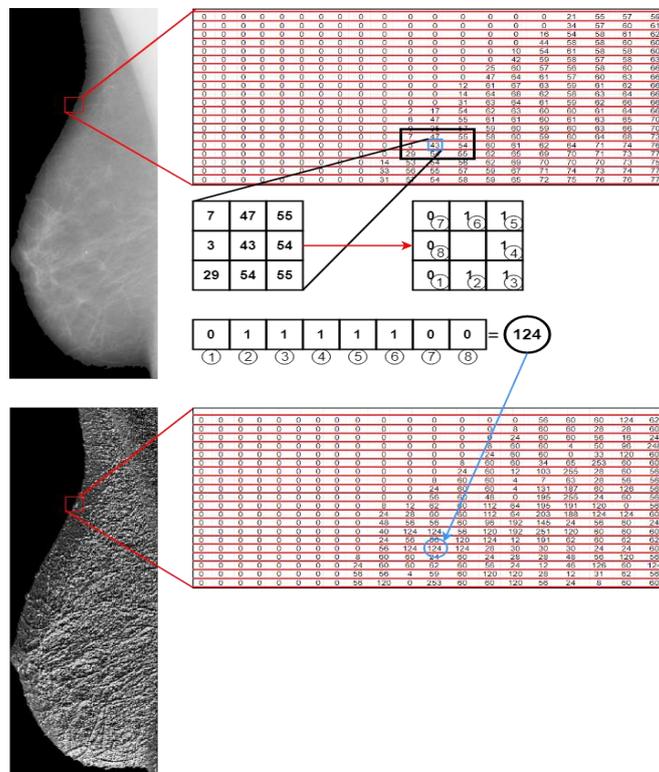


Figure 16. In this figure, in every image's cell the following process is performed: Each cell's pixel is considered as a center pixel of a 3 × 3 neighborhood. This neighborhood has gray scale values that are compared with the center pixel in order to create a binary code which is then converted to decimal and replaces the value of the central pixel. Importance should be given to the expression of coding sequence which is clockwise or anticlockwise.

Part2. Histogram calculation

Since the comparison is done, a computation of histogram takes place. According to the number of cells in the image, such histograms will be present Figure (17).

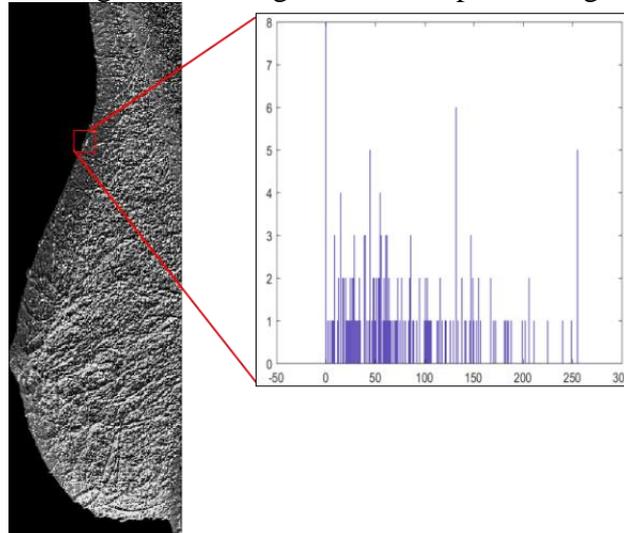


Figure 17. In this figure, is shown the histogram of the decimal values of the cell that occurred during the neighborhood comparison in the previous section. This computation is done in all the cells.

Part3. Feature extraction

Having computed the histograms for each cell, the final step of this descriptor is to concatenate all the cell histograms into 1×256 vectors. The size of the vector will be the multiplication of the number of cells times the 256 values of the gray scale image. In this situation, the result will be $25 \times 8 \times 256=51,200$ features.

As referred above, the optimal input parameters for this feature extractor are:

- Number of neighbors=24
- Radius=3
- Cell size=64

4.1.3 SURF

SURF (Speed Up Robust Features) [14] is comprised of a feature detector based on a Gaussian second derivative mask, and a feature descriptor that relies on local Haar wavelet responses. This framework shares many conceptual similarities with the most widely used feature detector in the computer vision community, called the Scale-Invariant Feature Transform (SIFT). The algorithm has three main parts: interest point detection, scale space representation, orientation assignment, Descriptor based on Haar Wavelet Responses.

Part1. Interest Point Detection

Firstly, for the interest point detection the algorithm uses the second derivative of the image (Hessian Matrix). This constitutes a great advantage as computational time is reduced significantly, because images are represented as integral images:

$$I_{\Sigma}(x) = \sum_{i=0}^{i \leq x} \sum_{j=0}^{j \leq y} I(i, j)$$

To be more comprehensive, the entry of an integral image $I_{\Sigma}(x)$ at a location $(x = (x, y)^T)$, represents the sum of pixel values in each image or a rectangular subset of the given image. A paradigm of such an image is given in Figure (18).

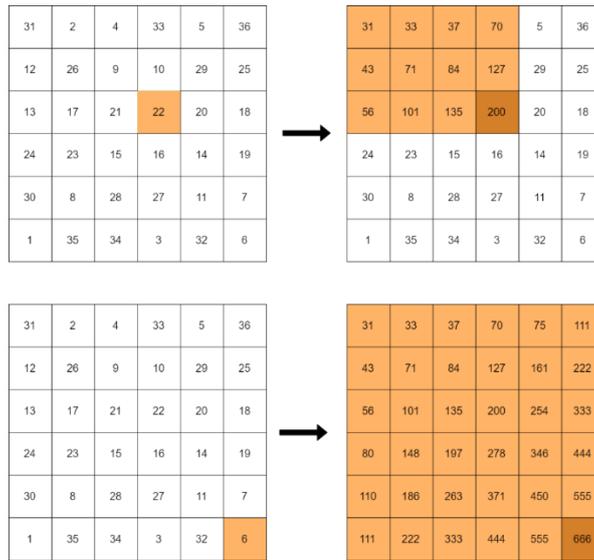


Figure 18. In this figure, the computation method of integral imaging is illustrated. Inside the rectangular region, the bottom-right pixel is the sum of all the rest parts of this region.

Hessian matrix $H(x, \sigma)$ is applied for measuring the point's neighborhood adjustments and thus, the selected points result from determinant's maximal locations:

$$H(p, \sigma) = \begin{bmatrix} L_{xx}(x, \sigma) & L_{xy}(x, \sigma) \\ L_{yx}(x, \sigma) & L_{yy}(x, \sigma) \end{bmatrix}$$

Where $L_{xx}(x, \sigma)$ is the convolution of the Gaussian second order derivative $\frac{\partial^2}{\partial x^2} g(\sigma)$ with the image I in point x at scale similarly for $L_{xy}(x, \sigma), L_{yy}(x, \sigma), L_{yx}(x, \sigma)$ Figure (19).

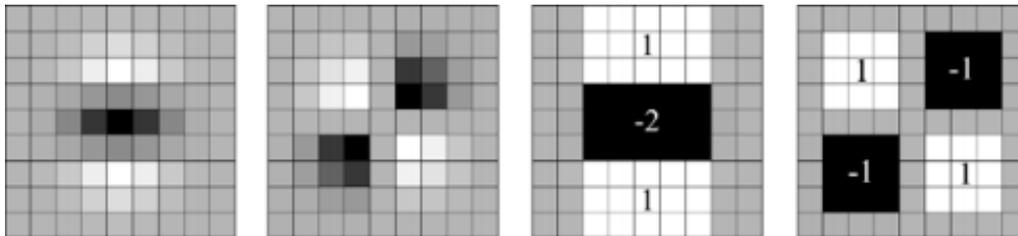


Figure 19. In this figure the two pictures from the left ($L_{xx}(x, \sigma), L_{xy}(x, \sigma)$) are calculated with the second order partial derivative of Gaussian and the result is on the two pictures to the right. Pixels which are positive represent a local maximum at x , the negative ones represent the local minimum at x and the gray ones are not inconclusive (image from [31]).

Part2. Scale-space representation and location of points of interest

In order to find points of interest, instead of resizing the image into different scales, the filter is being resized into different scales (see Figure 20).

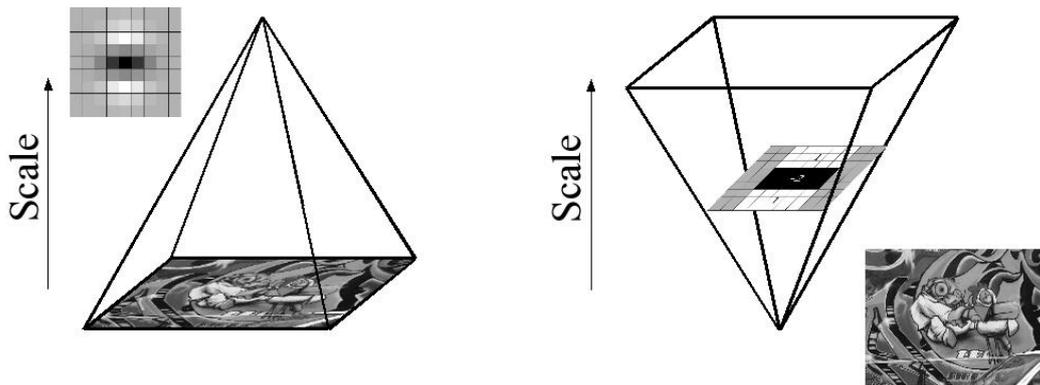


Figure 20. In this figure, the filter and interest point locations process is depicted (image from [31]).

Particularly, the filter sizes for the first octave are 9×9 , 15×15 , 21×21 , 27×27 , for the second octave are 15×15 , 27×27 , 39×39 , 51×51 . A third octave is computed with the filter sizes 27×27 , 51×51 , 75×75 , 99×99 and, if the original image size is still larger than the corresponding filter sizes, the scale space analysis is performed for a fourth octave, using the filter sizes 51, 99, 147, and 195. For each additional octave, the size of filter doubles (starting from 6 to 48 with step 6). But as the number of octaves is increased, the number of detected interest points decays Figure (21).

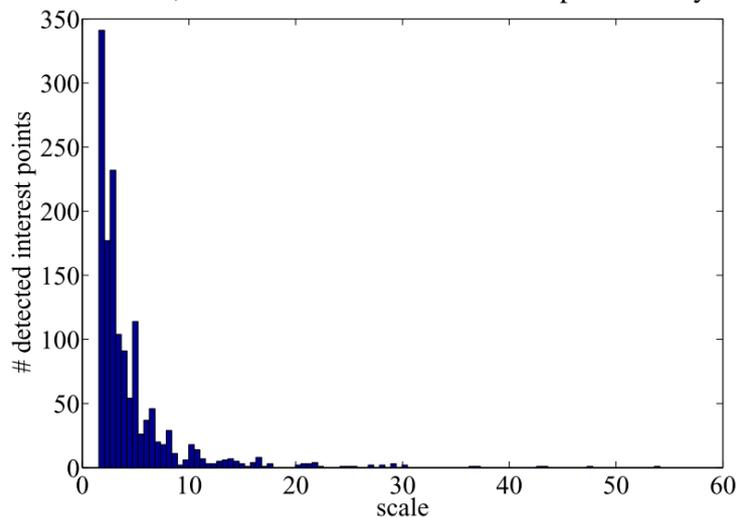


Figure 21. Here is an example of how the interest points decay from growth of filter size in an image (image from [31]).

Part3. Orientation Assignment

After having the image filtered and its interest points detected, the next step is to spot a recurring orientation of the points of interest. In order to achieve this the Haar-wavelet responses are calculated and weighted at x and y direction in a circular region of radius $6*s$ and centered at the key point. It is worth mentioning that the size of the wavelets is scale dependent and set to a side length of $4s$. 's' represents the scale at which the key point was determined.

Then, these responses are expressed as interest points in that circular region, and the dominant direction of this circular region is computed by counting the sum of all responses within a sliding orientation window of size $\frac{\pi}{3}$ Figure (22).

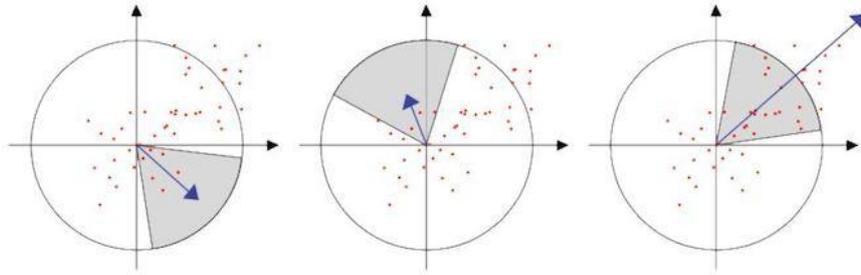


Figure 22. In this figure, the dominant orientation that generated by summing the responses within a sliding orientation window of size $\frac{\pi}{3}$ is depicted (image from [31]).

Part4. Descriptor based on Haar Wavelet Responses

Now that the orientation of each interest point is defined, the final step is to construct a square region around the key point with the orientation selected in the previous section as illustrated in Figure (23).



Figure 23. In this figure, a square region is constructed around the key point. The line from the center of each square, represents the dominant orientations of its point (image from [14]).

The region is split up regularly into smaller 4×4 square sub-regions. For each sub-region, the Haar-wavelet responses are computed in a 5×5 domain (sums of dx , $|dx|$, dy , $|dy|$) in order to find the dominant orientation of each region. Further on, the dx and dy responses are added together for each sub-area (Figure 24).

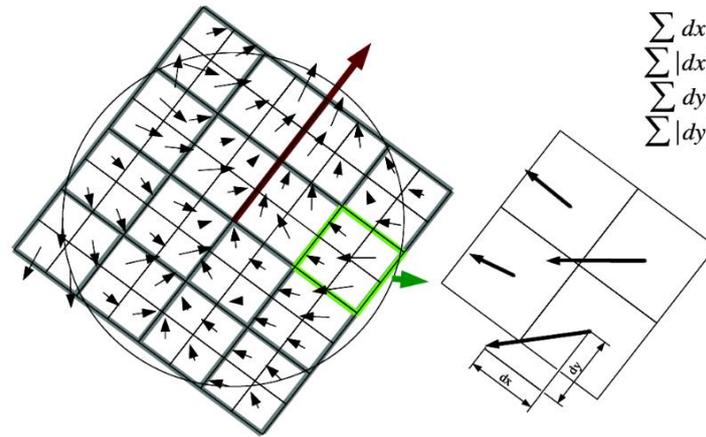


Figure 24. In this figure, the Haar-wavelet responses are computed (sums of \mathbf{dx} , $|\mathbf{dx}|$, \mathbf{dy} , $|\mathbf{dy}|$) in order to find the dominant orientation of each sub-region (image from [31]).

In order to acquire additional information for the opposition of intensity, the sum total of responses, $|\mathbf{dx}|$ and $|\mathbf{dy}|$ are extracted:

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y| \right)$$

Concatenating this for all 4×4 sub-regions, this results in a descriptor vector of length 64.

The optimal values for SURF were defined after extensive experimentation as :

- Sub-regions size= 8×8
- Filter size=64

4.1.4 SIFT

The scale-invariant feature transform (SIFT) constitutes a feature detection algorithm used in computer vision for detecting and describing local features in images [13]. Many algorithms have been presented to extract significant features from an image. Some of them use the Laplacian of Gaussian (LoG) to extract these features but they lack accuracy when the image is scaled. In this section we will briefly introduce the crucial steps of this algorithm and how the features are generated.

Step1. Scale space extrema detection

Unlike other rotation-invariant feature detection algorithms, such as Harris corner detector [32], SIFT uses difference-of-Gaussians (DoG) in different scales as an approximation of LoG (Laplacian of Gaussians) to locate interesting points which scale and orientations do not affect them. To begin with, progressively blurred images ($L(x, y, \sigma)$) are generated by convolving the original image ($I(x, y)$) with the Gaussian kernel ($G(x, y, \sigma)$). Mathematically, this procedure is described as, $L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$, where (σ) is the standard deviation or the “blur” value and $G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2}$. Then, the original image is resized to half size where a new **octave** (**octave** is the set of images generated by progressively blurring out an image by altering the blur value) is generated. And this process is repeated analogously with the original size of the image. This procedure is illustrated in Figure (25).

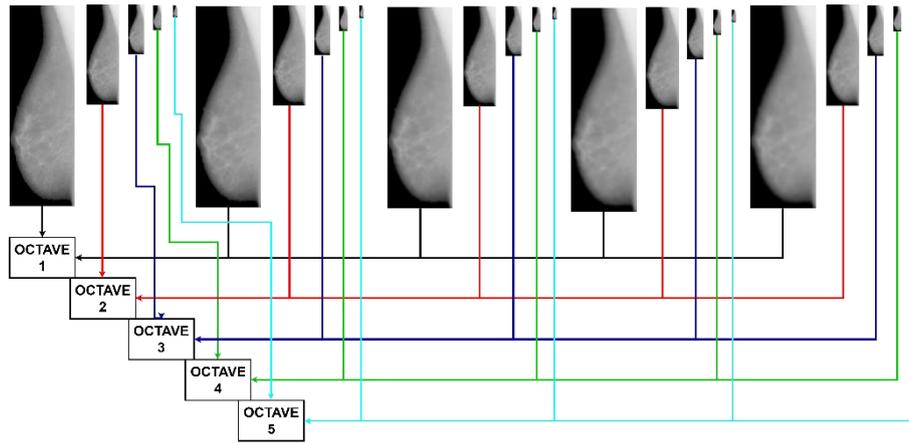


Figure 25. In this figure, each image with same size belongs to an octave which there is progressively blurred (from left to right).

After that, the blurred images are used to generate the DoG set of images for each scale. These set of images, are useful for finding interesting key points in the image. It is not worthy to mention, that DoG images approximate the LoG images and their great advantage is the lower computational cost. More analytically, DoG is computed by:

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$

An example of the DoG images, or equivalently, the possible interest points from the mias dataset is presented in the next figure (Figure 26)

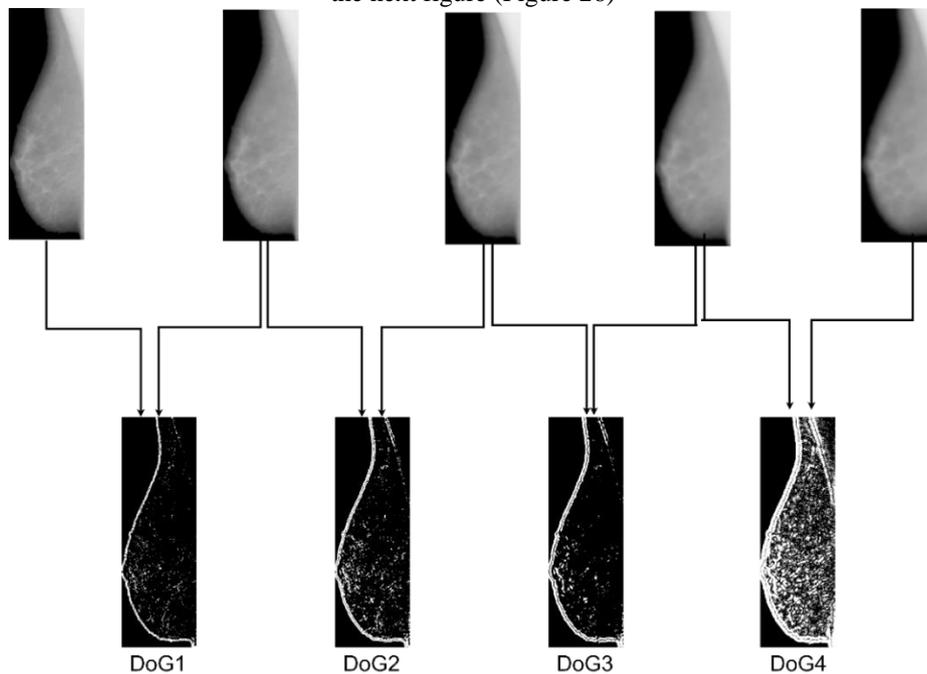


Figure 26. In this figure, the (DoG) images are generated, by finding the difference from 2 images at a time. This process is done in all the available octaves.

Now that the DoG images have been computed, images with less information are produced. The image regions with information contain the “extreme points” which are either the maximum or minimum value points in the image. In more detail, each pixel is being compared by its neighbors. The check is

done within the current image, and the one above and below it. One point is larger or smaller than all of its neighbors, then this point is selected as the extrema (candidate) (Figure 27).

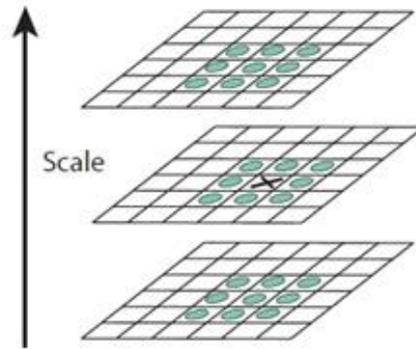


Figure 27. In this figure, the marked pixel 'X' is compared with its neighbors from its image, the one above it and one below it (image from [33]).

Step2.Keypoint localization

Once a candidate is found, the goal now is to find efficient key points. A selection is made by discarding those points that have low contrast (and are sensitive to light) and the ones that are poorly localized along an edge. The location and scale are calculated using the Taylor expansion of the image.

$$D(\mathbf{x}) = D + \frac{\partial D^T}{\partial \mathbf{x}} \mathbf{x} + \frac{1}{2} \mathbf{x}^T \frac{\partial^2 D}{\partial \mathbf{x}^2} \mathbf{x}$$

After that, the **edge** rejection follows, in order to find the principal curvatures. The principal curvatures can be computed from a 2x2 Hessian matrix, H, computed at the location and scale of the key point:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

The result is shown in Figure (28).

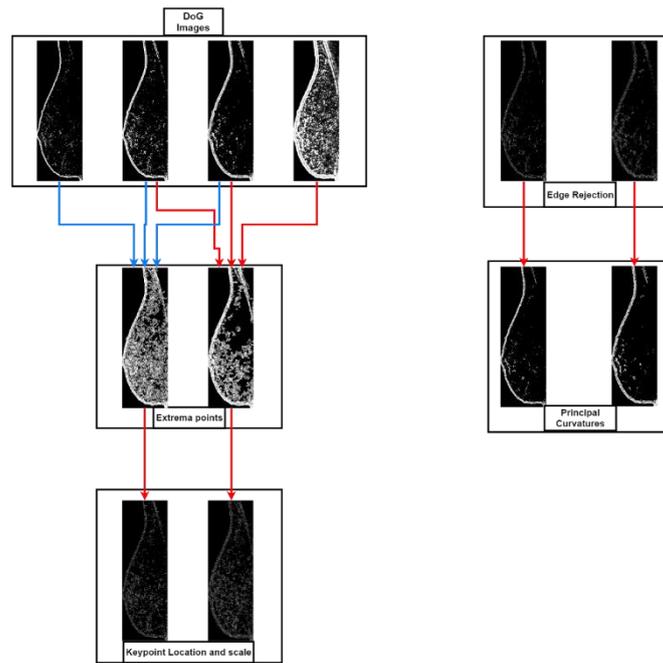


Figure 28. In this figure, using 3 by 3 images from the DoG set, the extreme points are detected. Then using the Taylor expansion, we find the Location and Scale of Key points in each image. After that, in order to find principal curvatures, edge rejection takes place with Hessian matrix. The result is shown on the set with name “Principal Curvatures”.

Step3.Orientation Assignment

The next step is to set orientation to the localized key points in order to maintain the image intact after any possible rotation. Around each key point a small region is defined according to scale, orientation and gradient magnitude are calculated separately for each key point region. A bar chart of orientations is created setting 36 equal subsets adding up to 360 degrees representing the, previously calculated, key point neighborhood direction. The key point orientation is estimated by taking the vertex of the histogram and any peak above 80% of it. New key points emerge with same location and scale but different directions Figure (29).

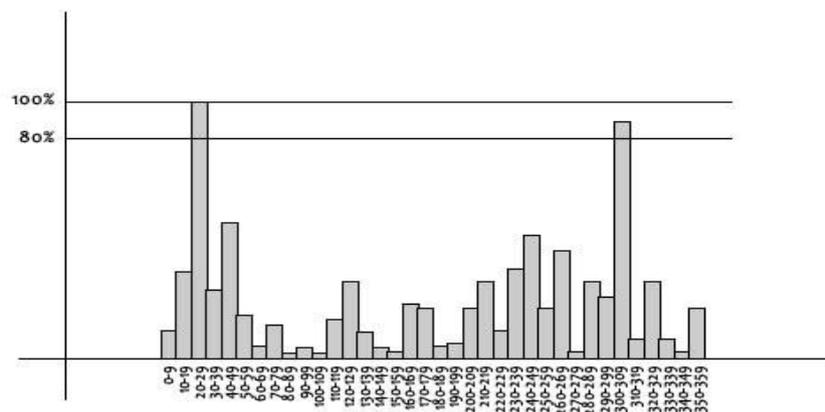


Figure 29. This histogram represents the degrees in 36 subsets. Each subset is increased by the presence frequency of a degree that a subset has (image from ([1])).

Step 4. Key point Descriptor

After the generation of key point descriptor, an area of 16x16 blocks around the key point is selected. The division of this area is resulted into 16 sub-blocks with size of 4x4. An 8-bin histogram of orientations is generated for each sub-block. To this end, a sum of 128 bin values are created.

Figures (30) and (31).

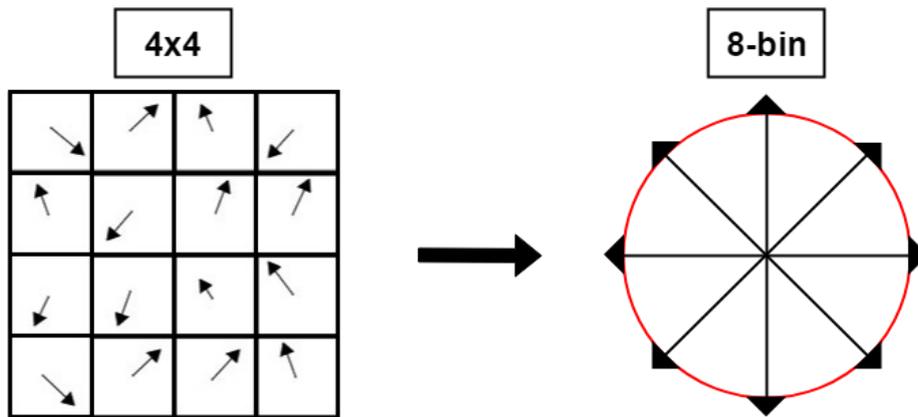


Figure 30. In this figure each 4x4 sub-block's direction values are inserted into 8-bin orientation histogram.

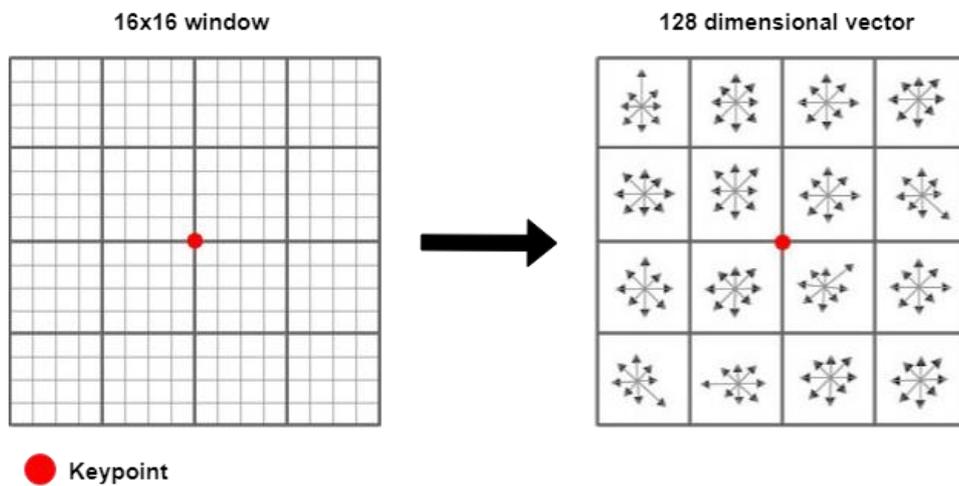


Figure 31. In this figure we see the whole migration of the degree values into 16 8bin orientation histograms (image from [1]).

SIFT optimal parameters were defined after extensive experimentation as:

- Edge rejection=3.5
- Ratio of principal curvatures =3
- Number of sub-block sizes=4,8,16

4.1.5 GABOR_LBP

Gabor filter part

Gabor filters were first introduced in 1946 by Dennis Gabor [34] and were extended, by J. Daugman [35] 42 years later to a 2-D Gabor filter which provides simultaneous optimal resolution in both spatial and frequency domains. The magnitude and orientation of the implemented image is

calculated using Gabor filter bank. The Gabor filter bank can be specified from many parameters that are shown below.

$$g_{x,y;\lambda,\theta,\phi,\sigma,\gamma} = e^{-\frac{(x'^2 + \gamma^2 y'^2)}{2\sigma^2}} \cos\left(i\left(2\pi\frac{x'}{\lambda} + \phi\right)\right)$$

Where: $x' = x\cos\theta + y\sin\theta$

$$y' = -x\sin\theta + y\cos\theta$$

The *standard deviation* σ of the Gaussian factor determines the size of the surrounding area of a pixel in which weighted summation takes place. The eccentricity of the Gaussian and the convolution kernel g is determined by the parameter γ , referred as *spatial aspect ratio*. It has a limited range of $0.23 < \gamma < 0.92$ [36].

The parameter λ is the *wavelength* of the sinusoidal factor and $\frac{1}{\lambda}$ the spatial frequency of the harmonic factor. The ratio $\frac{\sigma}{\lambda}$ determines spatial frequency bandwidth of the Gabor Filters. The half-response spatial frequency bandwidth b (in octaves) and the ratio $\frac{\sigma}{\lambda}$ are related as follows:

$$b = \log_2 \frac{\frac{\sigma}{\lambda}\pi + \sqrt{\frac{\ln 2}{2}}}{\frac{\sigma}{\lambda}\pi - \sqrt{\frac{\ln 2}{2}}}, \quad \frac{\sigma}{\lambda} = \frac{1}{\pi} \sqrt{\frac{\ln 2}{2}} \cdot \frac{2^b + 1}{2^b - 1}$$

The *orientation* parameter θ specifies the orientation of the Gabor function. Its value belongs in the interval $[0^\circ, 360^\circ]$. However, values in the interval $[0^\circ, 180^\circ]$ are taken, due to symmetry, the other directions become redundant [37]. Finally, parameter Φ , which is the *phase offset* of the harmonic factor, determines the symmetry of the function of the Gabor filter.

LBP part

After filtering the image with the Gabor bank, LBP takes place and enables its function upon the magnitude of the image with its optimal parameters Figure (32).

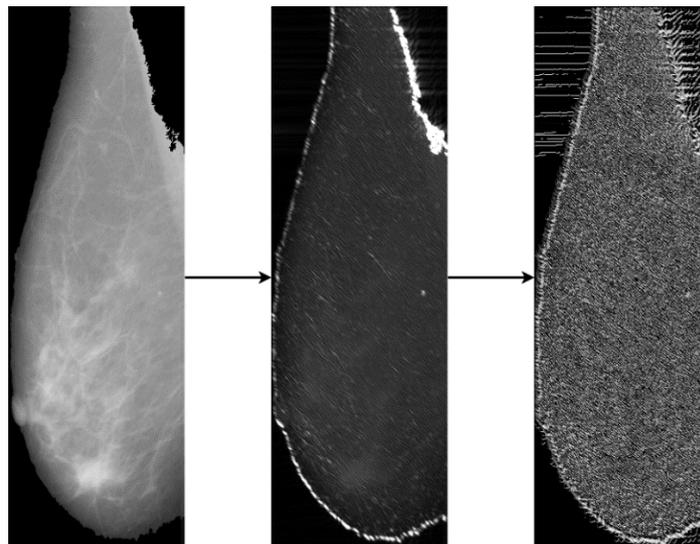


Figure 32. In this figure, the entire process of this combination is shown. The clinical Image(left) is filtered by the optimal parameters of Gabor's filter (middle) and then processed by the optimal values of LBP's function (right).

The optimal parameters for this combination were defined after extensive experimentation as:

Gabor:

- Wavelength=2
- Orientation=315
- $\Phi=1$

LBP:

- Number of neighbors=24
- Radius=3
- Cell size=64

4.2 Preprocessing

Image dimension processing

Although feature extraction has an important factor on the classification rate, preprocessing boosts this rate by removing the irrelevant image features. Mini-MIAS dataset's mammograms were cropped with the same motif and resized with bicubic interpolation to the point that only a minor area has background information Figure (33).

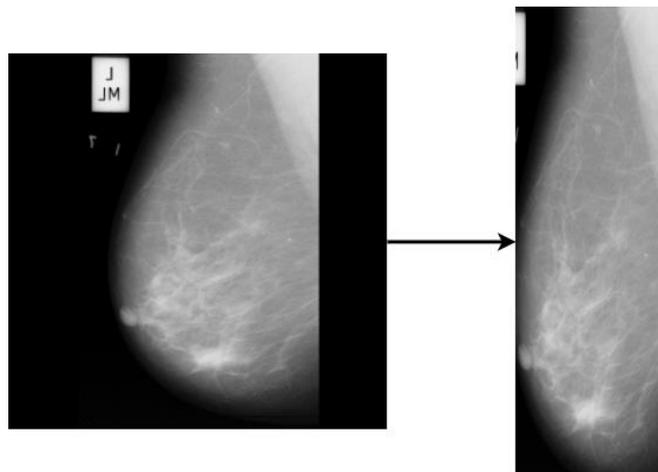


Figure 33, The procedure of cropping and resizing, resulted on a much smaller image size, making the overall calculation time shorter.

4.2.1 Background and patient's label information removal

After resizing and cropping, the next step is to remove the background using the Otsu's method [8] which corresponds to find the optimal greyscale value and minimize the interclass variance of breast and background through a histogram, Figure (34).

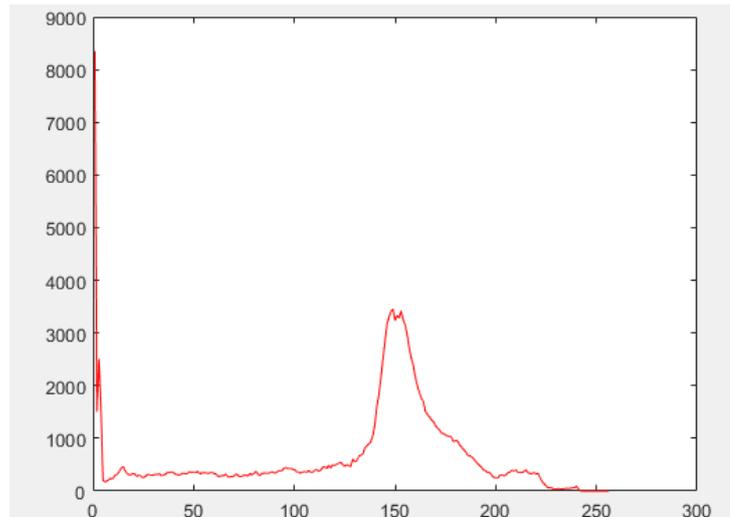


Figure 34. Otsu's method is to find the optimal threshold for the grayscale frequencies of the cropped image.

For the patient's label information removal, region connected analysis detection [11] was performed. Connected components analysis involves the detection of connected areas of binary images returns them as groups of coordinates. All the groups are deleted except the largest which represents the breast, as is lustrated in Figure (35).

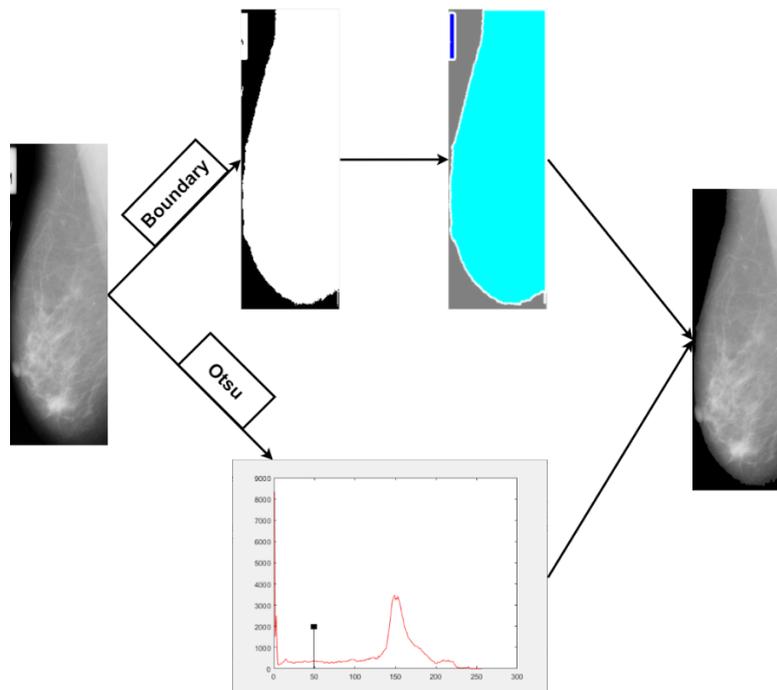


Figure 35. This image describes the conversion of a cropped image (left) to a preprocessed image (right).

4.2.2 Pectoral muscle removal

Last step of the preprocessing method is to remove the pectoral muscle next to breast tissue in order to preserve a set of only interesting mammogram areas. K-means [9] and gray connection [10] algorithm was combined for the completion of this task. K-means used as area clustering with the intention of segmenting the tissue from the muscle. The outcome was to use two different K values (3,4) because in some mammograms the pectoral muscle wasn't homogenous Figure (36). The solution was to compare mammograms the three approaches in order to define the optimal pectoral muscle segmentation as is shown in Figure (37).

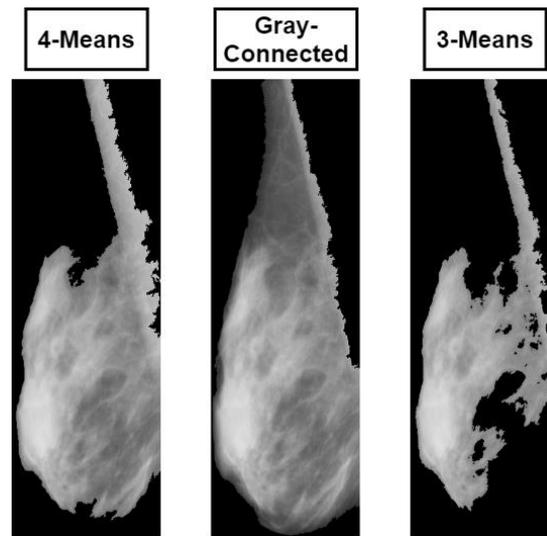


Figure 36. In this picture the failed attempt of pectoral muscle removal is shown and how breast image is affected by the different operations. The selected image from these 3, was the middle one.

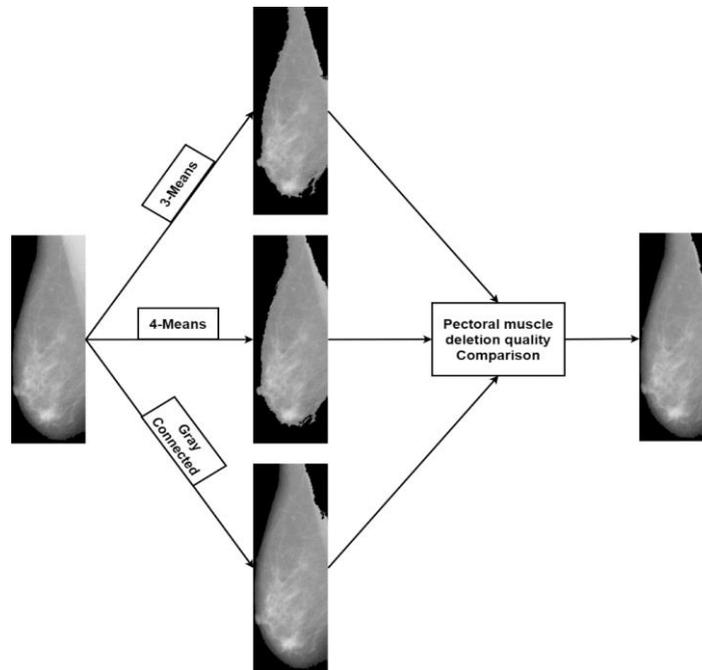


Figure 37. This figure shows the comparison between the three pectoral muscle approaches to define the best solution.

4.3 Feature selection

As mentioned in chapter 2, two types of classification methodology were established. One method refers to train the model with the extracted features, and another one in which a feature selection process is concluded before the model training phase. Feature selection/dimensionality reduction is an algorithm application that automatically selects the most significant features in order to achieve better performance score. Also, the total time of the entire system is dramatically reduced, offering a more compact system. Among the algorithms that Feature selection provides, Neighborhood Component Analysis (NCA) [17] was applied for this method's phase.

Feature Selection based on Neighborhood Component Analysis (NCA)

NCA is a dimensionality reduction methodology aiming to find a feature space that a stochastic linear neighbor algorithm will give the best accuracy also enforcing a visually meaningful clustering of

the data regardless the significant dimensionality reduction induced [17]. A detailed account of this method can be found in:[2]

4.4 Model training

After the selection or the extraction of features, model training takes place. Model training is a learning model that collects the appropriate information from the training set's labels and features for the purpose of predicting the labels of unknown features that belong to the test set. In more detail, model training is a fully trained multiclass that uses a classifier methods and training set's information to build a theoretically "trained classifier" capable of predicting the test's set labels properly. The applied classifier in this thesis was Linear Discriminant Analysis (LDA) [18] because this classifier is a robust, discriminative, and supervised learning algorithm that takes breasts class labels under consideration [38].

4.4.1 LDA

Linear Discriminant Analysis (LDA) is a supervised method used in machine learning to find a linear combination of features that separates two or more classes of objects. The resulting combination is used either as linear classifier or predictor for the test features.

Training the Classifier

Using LDA as a classifier, several computations are required for the purpose of creating this model (training the classifier with the training set). First thing to do is to calculate the separation between the 3 classes which is known as **between-class variance** and is defined as the distance between the mean of different classes. Afterwards, the **within-class variance** is computed within the class's mean and class's sample distance per time. The next and last stage of the process refers to **between-class variance** maximization and to **within-class variance** minimization in order to project the features in higher dimension space onto a lower dimensional space. With these steps, a classification model is created and ready to predict the upcoming unsupervised features (test set) Figure (38).

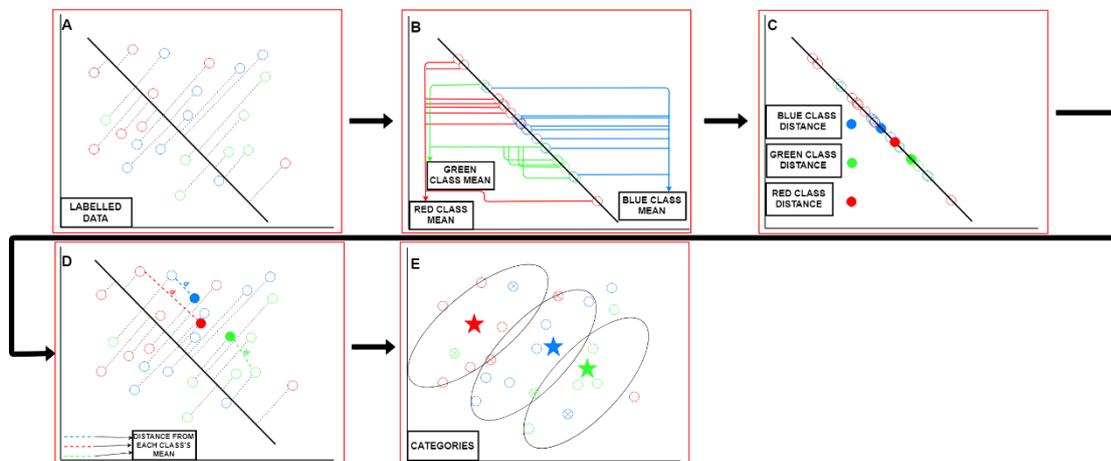


Figure 38. In this figure the classification method is depicted. As shown in picture (A), a line crosses the labelled data for the purpose of representation each data into the line. Then, the average (mean) of each class data is computed which creates a value for each class (picture B). This value is used to measure the distance between the three classes for computing the first step which is known as between-class variance (picture C). Next a distance is for each labelled data is measured between the data's sample distance and class's mean known as within-class variance (picture D). The final step, combines the previous two processes resulting both the maximum distance between each class and minimum distance from each data with its center (picture E).

4.5 Class prediction

4.5.1 LDA Prediction

As described previously, LDA can be used either as classifier or predictor. As a predictor, LDA cannot classify up to two classes instantly. For that reason, LDA either predicts 2 classes or predicts $N+2$ classes. In case of two classes, LDA estimates the probability that a new set of features (test set) belongs to every class. The class that gets the highest probability is the output class and a prediction is made. More precisely, this method uses Bayes Theorem [18] to estimate if the new feature's value provides a relative likelihood of being a part of a class.

In case of more than two ($N+2$) classes, multiclass LDA is needed. Multiclass LDA finds linear subspace which appears to contain all the class's variability. The approach of predicting more than two categories is to take one class and point the remaining classes as one. A common example of this is "one against the rest" where the points from one class are put in one group, and everything else in the other, and then LDA applied. This will result in $N+2$ classifiers, whose results are combined Figure (39).

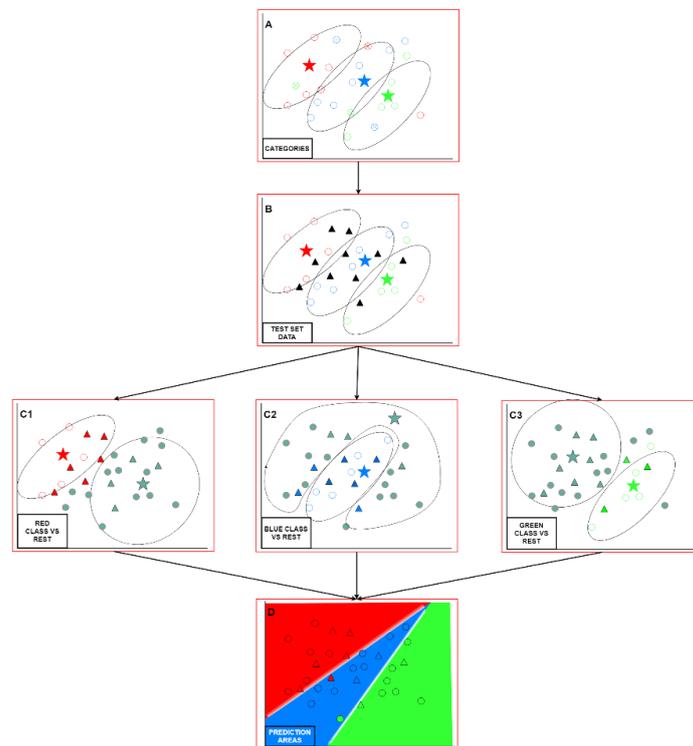


Figure 39. In this figure, the prediction method is shown. The test data that have been generated, are placed according to their values in the classification area (picture B). In this case of having 3 classes, the process begins with taking one class and combine the other classes into one in order to have 2 classes every time (picture C1, C2, C3). The computation of this prediction refers to the Bayes theorem that estimates the likelihood probability of each test data compared to the classification classes. After that computation, the resulting prediction comes from mixing the previous processes (picture C1, C2, C3) into a new classification model (picture D).

5. Experimental results

Machine learning offers a plethora of algorithms and methods for calculating the accuracy percentage of any given classification problem. For the purpose of evaluating our proposed breast-density classification methodology, cross validation was performed in order to increase the credibility of our results. Cross validation is a model validation process that evaluates the results (accuracy) of a classifier (LDA) upon a non-seen dataset. K-fold cross validation [39] technique was applied for evaluating the performance of our method in the mini-MIAS dataset separating it in equal k (5) sized sets(folds). Each fold at a time was treated as test set and the remaining 4 sets were split into validation set (for optimal parameter analysis) and training set with the application of holdout method [7]. The latter is a method that divides the method into 80% as training data (training set) and 20% as test set (validation set). After finding the optimal parameters from the 4 folds, the test set (i.e. the remaining fold) categories were predicted by the trained model. The generated score was then summed with the remaining ones and divided by the number of folds k(5). This process is depicted in Figure (40).

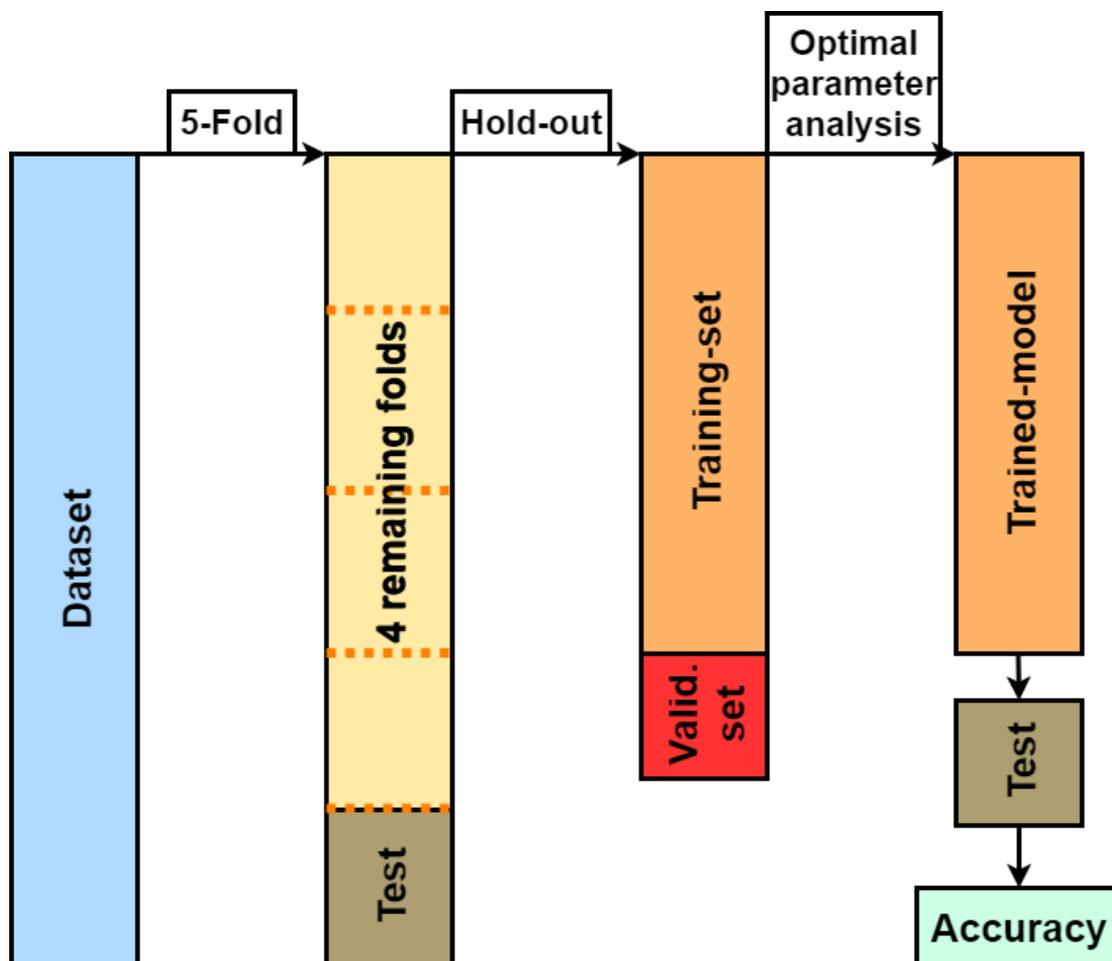


Figure 40. In this picture the cross validation and combination method between k-fold and hold-out method is depicted. This figure repeats as many times as the k in the k-fold.

The breast density classification system was tested in two related datasets: The mini-MIAS dataset including the information of breast tissue and pectoral muscle (Dataset #1), and the mini-MIAS dataset, with the pectoral muscle extracted and the only non-zero information being the breast tissue (Dataset #2). For each dataset, both the 2-class problem (Fatty + Glandular VS Dense) and the 3-class

problem (Dense VS Glandular VS Fatty) were assessed with our method. In the 2-class case, the AUC [40] was computed for accuracy evaluation, while in the 3-class case the Statistical 3-class confusion matrix were reported. Our methodology was tested in different configurations of the feature extractors to investigate which might be the best performer. In the following table (Table 2), we present the results of different algorithm configurations of feature extractors in the 2 or 3 class problems of both datasets used.

Breast density scoring												
2-CLASS ACC/AUC(%)			3-CLASS ACC/CONF.MATRIX(%)									
Method	Dataset #1	Dataset #2	Dataset #1				Dataset #2					
HOG	71.80/52.32	70.81/56.74	53.1 %		G	F	D	54.4 %		G	F	D
				G	39.0 8	24.88	36.04		G	41.28	34.00	24.71
				F	17.8 2	69.46	12.73		F	20.07	67.60	12.34
				D	27.5 6	20.14	52.30		D	20.98	24.81	54.21
LBP	83.25/78.01	82.03/75.23	74.2 %		G	F	D	71.5 %		G	F	D
				G	61.4 5	16.58	21.96		G	56.2 9	17.49	26.22
				F	9.89	88.13	1.98		F	9.06	88.96	1.98
				D	23.6 6	3.25	73.09		D	26.8 8	4.30	68.82
SURF	82.6/77.59	85.2/80.96	68.3 %		G	F	D	63.7 %		G	F	D
				G	59.7 7	20.96	19.27		G	51.4 0	20.48	28.12
				F	14.5 0	78.03	7.48		F	23.4 4	72.05	4.51
				D	28.2 4	3.82	67.95		D	30.2 4	3.36	66.40
GABO R + LBP	76.72/68.43	81.15/74.10	61.7 %		G	F	D	65.5 %		G	F	D
				G	42.7 1	30.66	26.63		G	47.6 9	22.96	29.35
				F	13.3 9	79.94	6.68		F	13.7 0	77.31	8.99
				D	28.0 9	9.23	62.69		D	19.4 6	10.46	70.08
SIFT- 4x4	57.17/44.90	63.50/53.38			G	F	D			G	F	D
				G	14.4 7	49.70	35.83		G	28.9 9	45.72	25.30

Categorization of medical images with modern techniques of visual and image processing methods
 Vasileios Melissianos

			28.2 %	F	34.7 5	39.90	25.35		35.1 %	F	20.4 2	48.40	31.17
				D	25.8 9	43.05	31.06			D	22.9 0	47.89	29.21
SIFT- 8×8	52.74/45.20	61.36/54.62	32%		G	F	D		35.5 %		G	F	D
				G	21.6 1	44.91	33.48			G	31.3 3	36.61	32.06
				F	25.4 9	43.42	31.09			F	29.6 4	43.34	27.02
				D	29.3 0	40.94	29.76			D	32.3 7	35.28	32.34
SIFT- 16×16	59.35/45.61	65.36/53.66	29.2 %		G	F	D		30.4 %		G	F	D
				G	22.3 2	39.49	38.19			G	10.6 1	46.96	42.43
				F	28.1 8	40.06	31.76			F	10.7 6	38.68	50.56
				D	23.9 6	48.94	27.10			D	14.1 6	44.23	41.62
Selected HOG	69.01/48.66	69.02/51.17	53.1 %		G	F	D		46.4 %		G	F	D
				G	37.66	24.87	37.47			G	28.6 5	41.78	29.58
				F	21.66	68.77	9.57			F	16.2 4	67.64	16.12
				D	25.65	21.25	53.10			D	24.3 9	30.48	45.13
Selected LBP	77.98/71.11	79.86/73.44	70.2 %		G	F	D		65.9 %		G	F	D
				G	42.12	15.27	42.61			G	39.7 9	18.62	41.58
				F	3.68	92.26	4.06			F	11.3 8	83.49	5.13
				D	17.69	7.21	75.10			D	22.2 6	4.44	73.30
Selected SURF	83.85/77.58	85.78/80.29	73.6 %		G	F	D		67.1 %		G	F	D
				G	66.5 1	16.09	17.40			G	54.7 0	20.20	25.10
				F	13.5 4	83.65	2.81			F	17.8 3	78.98	3.19
				D	22.9 3	6.28	70.79			D	28.9 3	4.99	66.08

Selected Gabor + LBP	64.93/60.99	66.60/62.01	50.9 %		G	F	D	42.9 %		G	F	D
				G	39.3 5	33.29	27.37		G	34.7 9	22.37	42.83
				F	27.7 3	54.93	17.34		F	35.0 3	45.06	19.91
				D	24.5 2	15.10	60.38		D	28.6 2	23.18	48.20
Selected SIFT 4×4	51.88/42.02	68.71/51.87	34.1 %		G	F	D	32.9 %		G	F	D
				G	31.7 1	38.01	30.28		G	20.2 8	35.23	44.50
				F	27.5 1	33.73	38.76		F	25.8 8	36.29	37.83
				D	19.7 6	42.98	37.26		D	19.0 4	34.21	46.75
Selected SIFT 8×8	54.37/46.07	65.70/56.60	28.9 %		G	F	D	35.4 %		G	F	D
				G	27.8 4	45.48	26.67		G	36.8 6	33.52	29.62
				F	43.1 7	27.79	29.04		F	38.8 0	35.97	25.23
				D	27.8 3	41.35	30.82		D	28.5 7	36.26	35.17
Selected SIFT 16×16	65.24/44.08	57.01/52.34	36.3 %		G	F	D	31.9 %		G	F	D
				G	29.4 6	43.18	27.37		G	27.7 8	38.32	33.91
				F	22.2 2	40.45	37.33		F	31.3 5	35.75	32.91
				D	28.7 5	30.44	40.80		D	31.8 6	36.22	31.92

Table 2. In this table 2 or 3 class results of different feature extraction algorithm configurations depicted.

5.1 Overall performance

LDA classifier was the main prediction tool for the entire table and LBP, SURF lead to the best results in the 2 and 3 class problem. In the 2-class problem, LBP scored 83.25% Accuracy/78.01% AUC (Dataset #1), 82.03% Accuracy/75.23% AUC (Dataset #2) while SURF scored 82.6% Accuracy/77.59% AUC (Dataset #1), 85.2% Accuracy/80.96% AUC (Dataset #2). These two robust algorithms show that Dataset #2 yields better results based on the AUC score. In the 3-class problem LBP scored 74.2% Accuracy (Dataset #1), 71.5% Accuracy (Dataset #2) and SURF scored 68.3% Accuracy (Dataset #1), 63.7% Accuracy (Dataset #2). The results indicate that our methods work better in the Dataset #1 for the 3-class problem.

6. Discussion

In the presented thesis, a machine learning methodology for breast density classification was developed and evaluated on mini-MIAS dataset. Several texture and orientation-based image features were examined in the context of density prediction rate. The rationale of this work lies in the clinical importance of the automated/objective characterization of mammographic breast density not least due to the fact that increased density is a risk factor for breast cancer. Most of the related published researches include binary tissue type prediction (non-dense versus dense). This was achieved in our case by fusing the fatty and glandular classes into the ‘non-dense’ class for binary prediction in the two datasets. According to our method, the results in Table 1, demonstrate that NCA feature selection did not increase the classification rate in most of the algorithms, indicating that NCA didn’t add value probably due to the relatively small number of features used.

In summary the tested configurations of our proposed methodology resulted in different scores, in the public mini-MIAS databases. Classification system scores in dataset#2 in comparison with dataset #1, reveal better prediction scores in dataset#2 based on AUC and confusion matrix results. HOG 2 and 3 class’s scores were for dataset #1 from 53.1% to 71.8% and for dataset #2 from 46.4% to 70.81% respectively. LBP results for dataset #1 ranged from 70.2% to 83.25% and for dataset #2 from 65.9% to 82.03%. SURF’s outcomes for dataset #1 ranged from 68.3% to 83.85% and for dataset#2 from 63.7% to 85.78%. Further on, Gabor and LBP combination scored for dataset #1 from 50.9% to 76.72% while on dataset #2, from 42.9% to 81.15%. Last but not least, SIFT had the worst performance in breast density classification, starting for dataset #1 from 28.2% to 65.24% and for dataset #2 30.4% to 68.71%.

Performance information is shown in table 3 which compares the prediction rates between the thesis algorithms with the already stated report rates.

Methodology or study, Authors (Refs.)	Mini-MIAS (2-class)	Mini-MIAS (3-class)
	Dataset #1/Dataset #2/Original Dataset %	Dataset #1/Dataset #2/ Original Dataset %
HOG	71.80/ 70.81/-	53.1 / 54.4 /-
LBP	83.25 / 82.03/-	74.2 / 71.5 /-
SURF	82.6/ 85.2/-	68.3 / 63.7 /-
Gabor + LBP	76.72/ 81.15/-	61.7 /65.5 /-
SIFT- 4×4	57.17/ 63.50/-	28.2/ 35.1/-
SIFT- 8×8	52.74/ 61.36/-	32 / 35 /-
SIFT- 16×16	59.35/ 65.36/-	29.2 / 30.4 /-
Selected HOG	69.01/ 69.02/-	53.1 / 46.4 /-
Selected LBP	77.98/ 79.86/-	70.2 / 65.9 /-
Selected SURF	83.85/ 85.78/-	73.6 / 67.1 /-
Selected Gabor + LBP	64.93/ 66.60/-	50.9 / 42.9 /-
Selected SIFT- 4×4	51.88/ 68.71/-	34.1 / 32.9 /-
Selected SIFT- 8×8	54.37/ 65.70/-	28.9 / 35.4 /-
Selected SIFT- 16×16	65.24/ 57.01/-	36.3 / 31.9 /-
S. Tzikopoulos et al.	-/-/-	-/-/ 85.25
S. Tzikopoulos et al	-/-/-	-/-/ 67.39
Boulenger et al	-/-/-	-/-/ 76.25

Table 3. Performce metrics for the several algorithms/configuration of our predictive methodologies for breast density characterization as well as from relevant works.

Although our method achieves very good classification results in the 2-class problem, Boulenger et al. and one study of Tzikopoulos et al. achieve better results in the 3-class classification problem comparing to our method as is reported in Table 3. It is obvious however that feature based

methods have a limit in mammogram density classification and only the use of more sophisticated techniques such as deep learning, can lead to much better results.

Breast cancer's radiographic appearance in high density breasts can be limiting factor for the early diagnosis of neoplasms in screening mammography. For this reason, computer-based density scoring systems can add value in the early detection of breast cancer. To this end, breast density classification is a challenging, yet not optimized method therefore further investigation and evaluation of novel methodologies are required.

6.1 Future work

New methodologies and further research will be applied with the aim to successfully generate or at least get closer as possible of creating a theoretically flawless classification system. For achieving that, more focus needs to be established in breast density classification methods (preprocessing, feature extraction, feature selection) and apart from machine learning, different Deep learning architectures [7] need to be implemented for improving performance.

7. Bibliography

- [1] <http://aishack.in/tutorials/sift-scale-invariant-feature-transform-keypoint-orientation/>
- [2] <https://www.mathworks.com/help/stats/neighborhood-component-analysis.html>
- [1] C. DeSantis, J. Ma, L. Bryan, and A. Jemal, "Breast cancer statistics, 2013," *CA. Cancer J. Clin.*, vol. 64, no. 1, pp. 52–62, 2014.
- [2] Y. P. Sun, S. Zhang, Z. Cui, and L. L. Qu, "CS based confocal microwave imaging algorithm for breast cancer detection," *Technol. Heal. Care*, vol. 24, no. 3, pp. S757–S765, 2016.
- [3] G. Pc and J. Kj, "Cochrane Database of Systematic Reviews Screening for breast cancer with mammography (Review) www.cochranelibrary.com."
- [4] E. Lazarus, M. B. Mainiero, B. Schepps, S. L. Koelliker, and L. S. Livingston, "BI-RADS lexicon for US and mammography: Interobserver variability and positive predictive value," *Radiology*, vol. 239, no. 2, pp. 385–391, 2006.
- [5] C. Byrne *et al.*, "Mammographic features and breast cancer risk: Effects with time, age, and menopause status," *J. Natl. Cancer Inst.*, vol. 87, no. 21, pp. 1622–1629, Nov. 1995.
- [6] F. B. QUEEN, "The curability of cancer," *Or. Health Bull.*, vol. 25, no. 53, p. 3, 1947.
- [7] K. M. ELEFThERIOS TRIVIZAKIS, GEORGIOS S.IOANNIDIS, VASILEIOS D.MELISSIANOS, GEROGIOS Z. PAPANAKIS, ARISTIDIS TSATSAKIS, DEMETRIOS A SPANDIDOS, "A novel deep learning architecture outperforming 'off-the-shelf' transfer learning and feature-based methods in the automated assessment of mammographic breast density," *Oncol. Rep.*, vol. 25, pp. 1–7, 2019.
- [8] P. Smith, D. B. Reid, C. Environment, L. Palo, P. Alto, and P. L. Smith, "A Threshold Selection Method from Gray-Level Histograms," *Pdfs.Semanticscholar.Org*, vol. 20, no. 1, pp. 62–66, 1979.
- [9] S. H. Ong *et al.*, "Medical Image Segmentation Using K-Means Clustering and Improved Watershed Algorithm View project Neuro vasculature modeling View project MEDICAL IMAGE SEGMENTATION USING K-MEANS CLUSTERING AND IMPROVED WATERSHED ALGORITHM," no. February 2001, 2001.
- [10] Y. Wang and P. Bhattacharya, "Image analysis and segmentation using gray connected components," *Proc. IEEE Int. Conf. Syst. Man Cybern.*, vol. 1, pp. 444–447, 1996.
- [11] W.-Y. Ma and B. S. Manjunath, "EdgeFlow : A Technique for Boundary Detection and Image Segmentation," *IEEE Trans. Image Process.*, vol. 9, no. 8, pp. 1375–1388, 2000.
- [12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005.
- [13] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.
- [14] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded Up Robust Features," Springer, Berlin, Heidelberg, 2006, pp. 404–417.
- [15] W. Zhou, A. Ahrary, and S. I. Kamata, "Image description with local patterns: An application to face recognition," *IEICE Trans. Inf. Syst.*, vol. E95-D, no. 5, pp. 1494–1505, 2012.

- [16] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol. Cybern.*, vol. 61, no. 2, pp. 103–113, Jun. 1989.
- [17] W. Yang, K. Wang, and W. Zuo, "Fast neighborhood component analysis," *Neurocomputing*, vol. 83, pp. 31–37, 2012.
- [18] P. J. Phillips, "Subspace Linear Discriminant Analysis for Face Recognition," no. April, 1999.
- [19] I. Domingues, "An automatic mammogram system: from screening to diagnosis," *Ph.D. Thesis*, no. May, 2015.
- [20] K. Bovis, "Classification of mammographic breast density using a combined classifier paradigm," *Med. Image Underst. Anal.*, no. c, pp. 1–4, 2002.
- [21] R. Chandrasekhar and Y. Attikiouzel, "A simple method for automatically locating the nipple on mammograms," *IEEE Trans. Med. Imaging*, vol. 16, no. 5, pp. 483–494, 1997.
- [22] D. J. Bartholomew, "Principal Components Analysis," *Int. Encycl. Educ.*, vol. 1, pp. 374–377, 2010.
- [23] H. Ellakany, K. Fábíán, I. Németh, and L. Stipkovits, "Antibody response detected by immunoblot in respiratory tract washings of chickens after infection with *Mycoplasma gallisepticum*," *Avian Pathol.*, vol. 27, no. 6, pp. 547–554, 1998.
- [24] S. D. Tzikopoulos, M. E. Mavroforakis, H. V. Georgiou, N. Dimitropoulos, and S. Theodoridis, "A fully automated scheme for mammographic segmentation and classification based on breast density and asymmetry," *Comput. Methods Programs Biomed.*, vol. 102, no. 1, pp. 47–63, 2011.
- [25] S. Tzikopoulos, H. Georgiou, M. Mavroforakis, and S. Theodoridis, "A fully automated scheme for breast density estimation and asymmetry detection of mammograms," *Eur. Signal Process. Conf.*, no. Eusipco, pp. 1869–1873, 2009.
- [26] S. Petroudi, T. Kadir, and M. Brady, "Automatic classification of mammographic parenchymal patterns: a statistical approach," pp. 798–801, 2004.
- [27] B. Percha, H. Nassif, J. Lipson, E. Burnside, and D. Rubin, "Automatic classification of mammography reports by BI-RADS breast tissue composition class," *J. Am. Med. Informatics Assoc.*, vol. 19, no. 5, pp. 913–916, 2012.
- [28] G. A. Boulenger, "Segmentation of the breast region with pectoral muscle suppression and automatic breast density classification.," *J. Nat. Hist. Ser. 7*, vol. 2, no. 8, pp. 133–134, 1898.
- [29] W. Xue, L. Zhang, X. Mou, and A. C. Bovik, "Gradient magnitude similarity deviation: A highly efficient perceptual image quality index," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 668–695, 2014.
- [30] Changming Sun and Deyi Si, "Skew and slant correction for document images using gradient direction," vol. 2, pp. 142–146, 2002.
- [31] H. Bay, T. Tuytelaars, and L. Van Gool, "Surf," *Eur. Urol. Suppl.*, vol. 13 (1), no. September, p. e227, 2014.
- [32] C. Harris and M. Stephens, "A Combined Corner and Edge Detector," pp. 23.1-23.6, 2013.
- [33] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, no. 3, pp. 346–359, 2008.
- [34] D. Gabor, "Theory of communication," *Journal of the Institution of Electrical Engineers - Part I:*

General, vol. 94, no. 73. pp. 58–58, 1947.

- [35] J. G. Daugman, “Complete Discrete 2-D Gabor Transforms by Neural Networks for Image Analysis and Compression,” *IEEE Trans. Acoust.*, vol. 36, no. 7, pp. 1169–1179, 1988.
- [36] N. Petkov and P. Kruizinga, “Computational models of visual neurons specialised in the detection of periodic and aperiodic oriented visual stimuli: Bar and grating cells,” *Biol. Cybern.*, vol. 76, no. 2, pp. 83–96, 1997.
- [37] K. Afrin, “Classification and feature extraction of binucleate cells using Mahalanobis distance and Gabor wavelet analysis,” *Int. J. Intell. Eng. Informatics*, vol. 2, no. 4, p. 304, 2014.
- [38] Q. Huo, Y. Ge, and Z. D. Feng, “High performance Chinese OCR based on Gabor features, discriminative feature extraction and model training,” *ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process. - Proc.*, vol. 3, pp. 1517–1520, 2001.
- [39] J. G. Moreno-Torres, J. A. Saez, and F. Herrera, “Study on the impact of partition-induced dataset shift on k-fold cross-validation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 23, no. 8, pp. 1304–1312, 2012.
- [40] C. Cortes and M. Mohri, “AUC optimization vs. Error rate minimization,” *Adv. Neural Inf. Process. Syst.*, 2004.