International Journal of Multimedia and Ubiquitous Engineering
Vol.11, No.6 (2016), pp.241-254
http://dx.doi.org/10.14257/ijmue.2016.11.6.21

Performance Evaluation of Cancer Diagnostics Using Autoregressive Features with SVM Classifier: Applications to Brain Cancer Histopathology

D. Vaishali[1], R. Ramesh[2] and J. Anita Christaline[3]

[1, 3] *Department of Electronics & Communication Engineering, SRM University Chennai, India*
[2]*Department of Electronics & Communication Engineering, Saveetha Engineering College, Thandalam, TN, India*
saivaishali@gmail.com

Abstract

Until the recent past, cancer diagnosis was made using histopathology methods, where the pathologists study biopsy samples and make inferences. These inferences are based on cell morphology and tissue distribution which represent randomness in growth and/or in placement. These methods are highly subjective/arbitrary and can sometimes lead to incorrect diagnosis. Nowadays, computer-assisted diagnostic (CAD), based on very large database, can aid in objective judgment. This study emphasizes the contribution of a two-dimensional (2D) autoregressive (AR) model for analysis and classification of histopathological images. In AR model, the parameters consist of a feature set of histopathological images obtained from biopsy samples taken from patients. These features are further used for analysis, synthesis and classification of cancer cells. The Yule-Walker Least Square (LS) method has been used for parameter estimation. The test statistics for the choice of a model order has also been suggested in this paper. It has been inferred that for a given sample image, the neighborhood is unique and solely depends on the properties of samples under consideration. Based on the features of AR parameters, samples are classified into two – healthy tissue and malignant tissue. The feature data sets have been classified using the linear kernel Support Vector Machine (SVM) classifier. In this work, we focus on measuring the performance of cancer diagnostic tests in terms of their recall, specificity, precision and F score. We observe that the fourth-order AR model gives promising results in performance evaluation using SVM classifier.

Keywords*: Autoregressive model (AR), Least Square (LS), Computer-Assisted Diagnosis (CAD), Support Vector Machine (SVM)*

1. Introduction

1.1. Background

Histopathology refers to the study of cells and tissues of biopsy samples and the changes in their morphology and distribution of disease cells. [1] Human body consists of millions of cells, and mitosis is the most common form of cell division, leading to cell multiplication. Mitosis allows a body to grow. Due to mutation in the cell's DNA, the normal cells become cancerous. Cancer affects the entire life cycle of a cell and the cell division goes out of control, which is called uncontrolled mitosis, resulting in the production of more and more damaged cells. These damaged cells accumulate and form a tumour. As the entire process of cell division occurs at random, the detection of mitosis becomes very complex. [2] In cancer screening and analysis, the examination of mitotic

ISSN: 1975-0080 IJMUE

figures in histology sample sections is important. Earlier this examination was carried out manually by histologists with a high level of expertise. Automating this process reduces the time for analysis, cost for analysis, and errors in analysis. It also enhances the compatibility of the result with different laboratories. Using computer-assisted diagnosis (CAD), high-resolution histopathology images give accurate information about biopsies samples, which form the base for the diagnosis of diseases [3].

1.2. Recent Developments

In recent days, pathology labs are undergoing a complete transformation. With the digitization of histopathology slides, optical microscopes are being replaced with computer monitors. [4] The introduction of whole-slide imaging (WSI) techniques has enabled quantitative automatic image analysis methods. With these methods, one can avoid the subjectivity in the diagnosis made. [5] Using CAD, the high-resolution histopathology images give accurate information about tissue samples. Recent developments in medical instrumentation have made histopathology even simpler. Many researchers have proposed several algorithms for diagnosis and classification of diseases. Tissue preparation and staining is done before digitization of histopathology slides as a first step of analysis. To enable a better visibility of the structures on the slide, they are dyed with stains. Hematoxylin & Eosin (H & E) and immunohistochemicals (IHC) are the standard staining protocols used in histopathology. H & E stained slides enhance the spatial nuclei features like size, shape, texture, spatial arrangements, tubules, stroma, *etc.* [6] In IHC stained slide most of the information is given by changes in colour intensity.[7] CAD consists of three main computational steps – pre-processing and segmentation, feature extraction and classification. [8] The primary objective of pre-processing is to eliminate noise and improve the image quality. In feature extraction, the entire image is represented using the extracted features. They can be cell based or tissue based, showing spatial independency or spatial dependency, respectively. The image can be represented by morphological features, textural features, model-based statistical features, fractal features, *etc.* In this study, the entire image in autoregressive model parameters is used for classification of cancer cells. The objective of the third step is to distinguish healthy tissues from malignant tissues and to detect the different levels of malignancy as well. To classify the tissues/cells, various classifiers like KNN, SVM, MLP, Bayesian *etc.*, can be considered. For a given class of histopathology images, it is essential to find the right classifier that uniquely identifies the changes in the extracted features. [9]

1.3. Overview and Contribution

In this work, we have considered brain histopathological images. Texture is a prominent characteristic for histopathological images. When affected by a certain disease, the cells/tissues show randomness in their growth (cell-based features) and in their placement (tissue-based features). The textural properties computed are closely related to the class of histopathological images (cancer types). [10] We have considered histopathological sample as a texture image and processed it as a complete texture with random features. From the analysis and the synthesis point of view, textures are broadly classified into deterministic textures and stochastic textures. In deterministic texture, the primitives and placement rule describe a particular texture. Placement rules arrange the primitives to form the texture. A placement rule further considers the deterministic approach or stochastic approach. [11] In stochastic texture cases, textures do not possess easily identifiable primitives. With segmentation algorithms, we may extract such primitives, but a placement rule description for such textures might be extremely complicated [12].

This work also focuses on the analysis and classification of images based on textural attributes. Our aim is to find the right model for a given class of histopathological images and estimate the model parameters. Estimated parameters are feature set of an image which then used for classification and detection of cancer. [13]

The rest of the article is organized as follows. Section 2 gives related work done in the past. Section 3 describes the basics of stochastic models. Section 4 details the representation of AR model and parameter estimation procedure. Section 5 describes performance evaluation issues. Section 6 gives experimentation results and Section 7 is on conclusion.

2. Related Work Done

Many researchers have acknowledged the importance of histopathological image analysis. A histopathological image preserves the underlying tissue architecture and provides comprehensive data of the disease. For these reasons, histopathology remains a 'gold standard' in diagnosing several diseases including all types of cancers. [13] Research on optimizing the feature set for disease classification is highly essential for accurate diagnosis and prognosis. Most of the research work has been carried out on visual attributes showing nuclear features like boundary appearance of ductal, stromal, tubular and glandular structures. This forms the base for cytopathology. [14] There was always a need for more research work on histopathology imagery. Analyses of histopathological images are classified as object-level feature analysis and spatial-level feature analysis. Object-level features are further categorized based on the (i) size and shape, (ii) radiometry and densitometry, (iii) texture and (iv) chromatin specific. [15] Size and shape include area, length of major and minor axes, convex hull features, filled area features, centre of mass features, *etc.* Radiometric and densitometric features include image bands, intensity, optical density, hue, *etc.*, Textural features include co-occurring matrix features, fractal features, run length features, and wavelet features. Entropy and chromatin-specific features are area, integrated optical density, compactness, number of regions, distance, *etc.* [16-18] Texture-model-based intermediate representation has been discussed in the literature by O. Sertel *et al*. [19] in spatial feature analysis, the graph theory plays a vital role. The use of spatial arrangements for evaluating cellular arrangements was first proposed by R. Albert *et al*. [20] Different states of tissues are modelled through graph and then used for classification. While working with spatial arrangements most proposed algorithms use graph theory. These are like Voronol Tessellation, Delunae Triangulations, Minimum spanning tree, Connected Graphs, Relative Neighbour Graphs, K-NN Graphs *etc.* These graphs carry features like number of nodes, number of edges, edge length, number of neighbours, cyclometric number, number of triangles, number of k-walks, special radius, fractal index *etc.*, [22-23] Listed below are some of the cancer diagnostic–related research works carried out with histopathological imagery. Researchers in medical image processing referred some of the above mentioned methodologies for cancer diagnosis. In the image analysis of neuroblastoma, segmentation of cell nuclei approach has been proposed and studied. [23] Analysis of histopathological breast cancer images for lymphocytic infiltrations is given in the literature by Ajay Basavanhally. [24] K. L. Wiend *et al*. studied the spatial structures in histopathological images. [25] Their work shows analysis of carcinomas and fibroadenomas of breast tissues in the form of microvessel distribution. P. H. Bartels *et al*. described the Bayesian belief networks for histopathological images. [26] The choice of a right classifier is also important while handling large data set of histopathological images. After segmentation and feature extraction steps, the classifier is used to classify the images into abnormal and normal tissues. The classifiers are also used to determine the stage of cancer. Several machine learning algorithms have been proposed by the researchers. They are neural network multilayer perceptron (MLP), k-nearest neighborhood (KNN), logistic regression method, Fuzzy systems, Support vector machine

(SVM) *etc.* Detailed comparison of performance of standard machine learning classifiers for breast cancer detection is undertaken and described. [27] To classify oral sub-mucus fibrosis, use of SVM-based approach is discussed in the following works. [28] Scott Doyle *et al*. describes the Bayesian classifier for prostate cancer detection from digital needle biopsies. [29] Work on intra-ductal breast lesions and classification using expectation maximization and watershed transform has been carried out and given in the following literature. [30] Treatment of biopsy samples in random field of texture for unsupervised segmentation in cancer diagnosis is presented by Akif Burak *et al.* [31] In CAD, accurate grading of tissue is required and tissues are characterized by the organization of their own components. Tissue texture is a key factor in determining the changes in their organization and in diagnosis. Interpretation of these changes is very important. Several researchers have already worked on this concept. Research by K. P. Kennan *et al*. has described this organization through graph run length matrix (GRLM). This work shows an arrangement and structure of cells through GRLM. They followed grey-level co-occurrence matrix (GLCM) which has been used for segmentation of the abnormal tissues. [32] Many researchers have been using Gabor and wavelet filters for textural feature extraction. Histopathology images of meningioma have been already analysed and classified. [33] Follicular lymphoma grading can be performed with texture models for intermediate representation with colour texture images. Segmentation of histopathology images with follicular lymphomas can be performed with supervised colour texture analysis method. [34]

3. Stochastic Models

Several researchers have worked on image modelling, considering various approaches based on linearity and nonlinearity in data. Many statistical methods for the measurement, characterization and analysis of image data have been suggested in the form of statistical modelling and they are autocorrelation function, covariance function, co-occurrence matrix, grey-level probabilities, grey-level run lengths and regression analysis. Morkov random field models have been used to model nonlinearity in pixel grey levels. [35] In autoregressive model approach, Hawlick considers linear estimates of grey level by specific neighbourhood which describes texture found in histopathological image samples.

Basically in histopathological image modelling experiments, we have designed the model that finds the embedded properties of the image as a model parameter. [1] The two basic purposes of the model design are to find the hidden properties in image sample for analysis or classification, and the second is to generate the same image using model parameters in synthesis experiment. [36]

Stochastic model plays an important role in many image processing applications like image compression, analysis, segmentation, restoration and image retrieval. While pursuing this study in the stochastic-model-based approach for histopathology image analysis and classification, we met with certain challenges such as to (i) select the proper kind of model as AR, moving average (MA), ARMA, Markov models (MRF) and also model orders, (ii) select the proper method to estimate the model parameters (iii) select proper neighbourhood, checking complexity in estimation process and accuracy in classification and (v) select classifier, which is also very crucial as classifier output (accuracy of classification) is a feature-specific component. [1]

4. Autoregressive (AR) Models

In the spatial domain image processing, the grey-level dependency of one pixel on other pixels in the neighborhood is well described by the autocorrelation function. For the first time, this linear dependency was tested with an autoregression model by McCormick and Jayaramamurthy. [37] They have also used this dependency in synthesis experiments.

4.1. Representation of Two-Dimensional (2D) AR Model

We represent an image as 2D random field [1]

$$\{x[n, m], (n, m) \in N\}$$

Defining total lattice with

$$(i, j) \leq (s, t) \leftrightarrow i \leq s \text{ and } j \leq t$$

2D AR model defines N1 × N2 histopathological sample image **I,** where

$$I = \{x[n, m]: n \leq N1-1 \& m \leq N2-1\} \quad (1)$$

And x **[n, m]** represents the grey-scale value of the pixel located at **[n, m]** on the 2D space.

The image **I** can be represented by a difference equation [1].

$$x[n, m] + \sum_{i=0,j=0}^{i=p1,j=p2} a_{ij}\, x[n-i, m-j] = w[n, m] \quad (2)$$

Where p1 and p2 are row-wise and column-wise lag and decide the order of model or neighbourhood N, {**w [n, m]}** are white Gaussian noise with variance σ^2 and the coefficients $\{a_{ij}\}$ are parameters of the model. From equation (**2)** we considered an image **x[n, m]** as the output of linear time-invariant causal system with transfer function

H (z_1, z_2) and excited by white noise **w[n, m]** input. [1]

$$H(z_1, z_2) = \frac{1}{A(z1,z2)} = \frac{1}{\sum_{i=0,j=0}^{i=p1,j=p2} aij(z1-i, z2-j)} \quad (3)$$

4.2. Yule-Walker Least Square Parameter Estimation

If the noise sequence **w [n, m]** is known, the parameter estimation of AR model by equation **(2)** would be simple and it would be input–output system parameter estimation problem. In this problem, the input is noise and the output is the sample image under consideration, and we need to estimate AR parameters. It has been observed that the simplest method of parameter estimation is the least square method.

In the LS method, we express:

$$X[n, m] + \Phi^t[n, m]\, \theta = w[n, m] \quad (4)$$

$$\Phi^t[n, m] = [x[n, m-1], \ldots \quad x[n - p_1, m-p_2], -w[n, m-1], \ldots -w[n - p_1, m-p_2]].$$

$$\theta = [a_{01}, a_{02}, \ldots, a_{p1p2}]^t$$

Writing equation **(4)** in matrix form, [1]

$$x + \Phi\, \theta = w \quad (5)$$

Assume that we know Φ (Input white noise), then we can obtain a least square estimate ($\hat{\theta}$) of parameter vector θ by

$$\hat{\theta} = (\Phi^t \Phi)^{-1} \Phi^t x \quad (6)$$

For finite order AR model (p1, p2), the parameters in AR model can be estimated using a 2D extension of the Yule-Walker equation as follows

$$r[k, l]+\sum_{i=0}^{p1} 1 \sum_{j=0}^{p2} r[k-i, l-j]\, a_{ij}=\sigma^2\delta[k, l] \quad (7)$$

Where r[k, l] is the autocorrelation values of the {x[n, m]} calculated as

$$r[k, l] = \frac{1}{(N-k)(M-l)} \sum_{i=1}^{N-k} 1 \sum_{j=1}^{M-l} x[i, j]x[i+k, j+k] \quad (8)$$

Where **δ[k, l]** is the 2D Kronecker delta function. Form a matrix with linear equations like **(7)** and solved it for the coefficients a_{ij} (AR parameter estimates). [1]

4.3. Yule-Walker Least Square Algorithm

1. Read the sample image, perform pre-processing to remove noise and improve the quality of the sample.
2. Generate random image with Gaussian distribution and with variance σ^2.
3. Use Yule-Walker method given by equation **(7)** and estimate the parameters a_{ij} for the given neighbourhood **N**.
4. Find out an estimate of the noise field w [n, m].

$$w[n, m]\hat{}=x[n, m]+\sum_{i=0}^{p1} 1 \sum_{j=0}^{p2} x[n-i, m-j]\, a_{ij}\hat{} \quad (9)$$

for n = p1 + 1; to N1 and m = p2 + 1 to N2

Replace w[n, m] by w[n , m]^ computed in step 2 and obtain θ ^in equation (6).

4.4. Model Optimization: Choice of Neighbourhood (N)

Optimization process refers to a branch of applied mathematics concerned with the minimization or maximization of a certain function possibly under some constraints. We proposed a method to optimize model order, which is based on classification accuracy. Extracted AR parameter set with different neighbourhoods (model orders) is considered for classification. Parameters derived from histopathological image sample with different model order are taken for classification, and the set which gives highest classification accuracy is considered as an optimized model order. [1]

In a modelling experiment we encounter two major challenges related to the method of determining the best neighbour set ***N*** of the AR model and a method of estimating the parameters of the model. Least square (LS) estimates are consistent and efficient for parameter estimation. Figure 1 illustrates the causal (Quarter Plane) and semi-causal (NSHP) model neighbourhood. [38] In practice, the appropriate neighbourhood needs to be estimated from the given sample data. Asymptotically consistent, transitive and parsimonious decision rule is required to choose the appropriate N, which ultimately increases the classification accuracy.

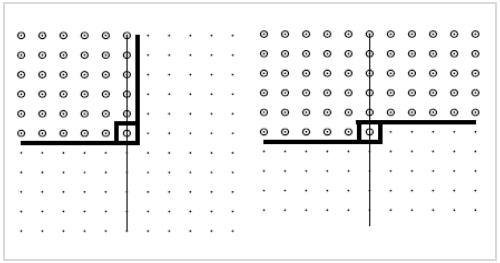


Figure 1. Quarter Plane Causal and Non-Symmetrical Half-Plane (NSHP) Semi-Causal Model

Several AR models, say $\mathbf{n_0}$ in number, are hypothesized for the given sets of sample. Each AR model has a different **N** (neighbourhood) with the corresponding unknown parameters. The unknown parameters are estimated for the class of histopathology image and considered for classification. The optimized order of model is the one which gives the maximum classification accuracy. If we deviate from this order, the classification accuracy will start decreasing.

Optimized model order = Model order @ Max (classification Accuracy)

5. Classification

5.1. Selection of Classifier

The next step after estimating right set of AR parameters is to determine the proper classifier to distinguish the malignant structures from others. In this step, we assign either the malignant or the healthy tissues as the sample tissues. Classifiers are also used to grade the malignancy in the tissues. There are two different ways to classify a particular sample. The first way is to execute some statistical tests on the features or to implement specific machine learning algorithms. While handling histopathological images, the statistical test results need to be interpreted cautiously. This is because statistical test assumes independent samples, but in practice all the tissue samples are collected from the same patient and are not independent. This may cause misleading conclusions. The different classes are distinguished using the machine learning algorithms. Many algorithms are proposed by researchers like neural networks (MLP), k-nearest neighborhood algorithm (KNN), and support vector machine (SVM), logistic regression, linear discriminate function, fuzzy systems, *etc.*

In our study the support vector machine algorithm has been considered to classify the malignant tissues from the healthy ones. The SVM algorithm is very sensitive to the elements of training data and testing data, and, therefore, a k-fold cross-validation approach is implemented. The average accuracy obtained through the k repetition of SVM classifier is considered as the final cancer detection accuracy. [39]

5.2. Performance Evaluation Issues

There are many ways to extract the information about a particular condition of a patient. Different diagnostic tests are used to investigate the accurate condition about a disease. In histopathology, many tests are carried out to extract the different features of biopsy samples. Thus, the treatment plan depends on the diagnostics test reports and the information obtained through these tests. Sensitivity (recall), specificity and accuracy, precision, F score, *etc.* are the preferred statistic measures to describe a diagnostic test. These statistic measures are used to measure the quality and reliability of the test. Sensitivity/recall detects the efficiency of the test in detecting a disease, and specificity estimates the likelihood of correctly ruling out the disease in a patient. Accuracy measures the efficiency of the test to identify and rule out the disease. [40] Although we accept the fact that higher the accuracy, the better the classifier, it is having limited use and it is always misleading. The disease prevalence, which is the statistical term used depending on population of disease, always affects the accuracy measure. Precision or confidence is the measure of predicted positive cases. [41] F score is another measure of testing accuracy based on precision and recall. Sometimes it is referred to as balanced mean or weighted average mean. Several terms as true positive (TP), true negative (TN), false positive (FP), false negative (FN) are associated with this study, and their definitions are as follows: [42]

True Positive (TP) - Sick person correctly diagnosed as sick.

False Positive (FP) - Healthy person incorrectly identified as sick.

True Negative (TN) - Healthy person correctly identified as healthy.

False Negative (FN) - Sick person incorrectly identified as healthy.

These terms are used to quantify the authenticity of the tests. In our study we have used the above-mentioned terms to describe sensitivity sometimes referred to as recall, specificity and accuracy, precision, F measures. They are described in TP, TN, FN and FP as follows:

Sensitivity / Recall = TP / (TP + FN)

Specificity = TN / (TN + FP)

Accuracy = (TN + TP) / (TN+TP+FN+FP)

Precision = TP / (TP+FP)

F score = 2 * (Precision * Recall) / (Precision + Recall)

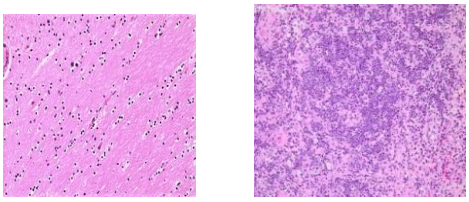
In this study of histopathology malignant images for brain cancer, we have used the above theory to find out classification accuracy using the SVM classifier with AR features and different neighbourhood size.

6. Experimentation

6.1. Experiment Setup

This work has been implemented in MATLAB environment proven for its performance in technical computing. This research work mainly consists of two parts. The first part is for image modelling (parameter estimation) and classification. The optimization of the model is also done using test statistics.

We test our model on histopathology images showing brain cancer tissue. Histopathological image database is taken from Tissue Microarray Database (TMAD), Dept. of Biochemistry, School of Medicine, Stanford University, which is an open source for research in pathology. [43] Sample histopathology images for healthy and malignant brain tissues are shown in Figure 2. In the first phase of this work, we cropped a portion of size 512 × 512 pixels from all the sample images before pre-processing. Pre-processing removes the noise and improves the quality of histopathology sample image, thereby making it suitable for key feature extraction. The filters in pre-processing smoothen or sharpen the image so that the disease-specific features are enhanced and correctly extracted. After pre-processing, we calculated the autocorrelation functions. Autoregressive parameters are then estimated using Yule-Walker least square method. These AR parameters form data sets for classification. We formed sets of AR parameters with different model orders. In the next phase, we performed classification to calculate classification accuracy. In this process, images are labelled either as normal or as cancerous. We tested our data sets using machine learning classifier, the linear kernel support vector machine (SVM). The classification procedure starts with training of the classifier. We used k-fold technique with ‘k’ value set at 5 and the number of iteration set at 100. To evaluate the result or to check the performance of the classifier on given data sets we used TP, TN, FP and FN measures and then calculated the sensitivity, specificity and accuracy, precision, and F score.



Healthy Brain Malignant Brain

Figure 2. Histopathology Image: Healthy Brain Tissue and Malignant Brain Tissue

5	4	3	2	2	2
5	4	3	2	1	1
5	4	3	2	1	X(i,j)

Figure 3. Quarter Plan Neighbourhood Up to Fifth Order

Table 1. Quarter Plane Causal Neighborhood Pixel under Consideration = X (i,j) = X (0, 0)

Model Order 1	Model Order 2	Model Order 3	Model Order 4
(i, j-1), (i-1, j-1), (i-1, j)	(i, j-1), (i-1, j-1), (i-1, j), (i, j-2), (i-1, j-2), (i-2, j-2), (i-2, j-1), (i-2, j-0)	(i, j-1), (i-1, j-1), (i-1, j), (i, j-2), (i-1, j-2), (i-2, j-2), (i-2, j-1), (i,j), (i,j-3), (i-1, j-3), (i-2, j-3), (i-3, j-3), (i-3, j-2), (i-3, j-1), (i-3, j)	(i, j-1), (i-1, j-1), (i-1,j), (i, j-2), (i-1, j-2), (i-2, j-2), (i-2, j-1), (i-2, j), (i, j-3), (i-1, j-3), (i-2, j-3), (i-3, j-3), (i-3, j-2), (i-3, j-1), (i-3,j), (i-4,j), (i-4, j-1), (i-4, j-2), (i-4, j-3), (i-4, j-4), (i-3, j-4), (i-2,j-4), (i-1, j-4), (i, j-4)

6.2. Results

In the proposed system for mitosis detection, the 2D autoregressive model with different neighbourhood has been considered to form the feature data sets. Figure 3 shows the 2D quarter plan neighbourhood and Table 1 shows the pixels under consideration with different neighbourhood or model orders. The estimated AR parameters for both healthy and malignant brain tissue samples are given in Table 2. As we increased the order of the model, the computation complexity also increased. As Table 3 shows, with AR features, the classifiers demonstrate good results in accurately detecting mitosis patients (TP = 87.50%) and eliminating non-mitotic patient as well (TN = 81.25). In a similar fashion, the system performance in false detection of mitotic patients (FP=18.75%) and losing the real mitotic patients (FN = (13.50%) shows greater precision and good sensitivity, respectively. Experimental results reveal that accuracy calculated with fourth-order (larger neighbourhood) AR model is 84.50% with SVM classifier which is found to be good. We also calculated the F score value which reveals the harmonic mean of precision and recall and gives better evaluation of performance test. The models with lower orders (small neighbourhood) are not able to detect the key features of cancerous tissue and hence not able to classify correctly.

Table 2. Estimated Model Parameters for Healthy Brain Tissue & Malignant Brain Tissue with Different Model Orders

HEALTHY BRAIN TISSUE						
Model Order 1	Model Order 2	Model Order 3		Model Order 4		
-0.9950	-1.1800	-1.1268	0.0393	1.1298	0.0204	0.1366
-0.9957	0.1860	0.0936	-1.1150	-0.0216	-1.0913	1.0522
0.9906	-1.1661	1.0730	0.1270	0.1178	-0.0653	0.0176
	1.1030	-0.0892	-0.0543	-0.0598	0.1719	-0.0720
	0.0591	0.1616	-0.0724	-0.0471	0.0064	0.0729

	0.1711	-0.0355	0.1753	-0.0139	-0.0405	-0.0312
	0.0733	-0.1118	-0.0950	0.0122	0.0847	-0.0847
	-0.2463	0.0302		-0.0271	0.0059	0.0210
MALIGNANT BRAIN TISSUE						
-0.7653	-1.0267	-1.0951	0.5910	-1.0892	0.5950	-0.3281
0.8474	0.2932	-0.3202	-1.4261	0.0065	-1.4446	1.4524
0.6147	-1.2675	1.4623	-0.7260	-0.6985	0.4264	0.0342
	1.2509	0.4518	0.9113	0.9692	-0.9024	0.3385
	-0.3247	-0.8943	0.3992	-0.2613	-0.0712	-0.4127
	0.4884	-0.3106	-0.3306	0.3845	-0.1303	0.1286
	-0.4750	0.3364	-0.1702	0.0344	0.0530	-0.0358
	0.0639	0.1222		0.0254	-0.0051	-0.0170

6.3. Discussion

Based on the results, it is understood that mitosis detection with autoregressive parameters is one way to extract the prominent features of histopathological image. In the autoregressive model, each pixel is represented by the estimated AR model parameters using an appropriate neighbourhood (order, window size). The neighbourhood should be selected in such a way that it captures sufficient information from the key region. For textural feature extraction, many researchers have attempted work on texture properties mentioned in Section 2. It shows work has been done with textural features, mainly considering the co-occurring matrix features, fractal features, run length features, wavelet features, entropy and chromatin-specific features, compactness, number of regions, distance, model-based intermediate representation are discussed by [19]. The work by Weyn *et al* represents the use of wavelet coefficients for textural analysis of histopathology image. [44] We believe that in the model-based study of histopathology image analysis, our work proposes an innovative method to extract image features using the AR coefficients, and the images are accurately classified using the SVM classifier. This work conceptualizes a new technique using AR model for CAD. The classification accuracy result shows the feature set formed by AR parameters are capable of detecting the hidden properties of histopathology images, which can be further classified. There are a few limitations to the proposed method. As we can see, classification accuracy is not only improving with model orders but also increasing the complexities in parameter estimation. Another limitation of this method is that it is not suitable for interacting with nonlinear stochastic fields (textures) in histopathology imaging as autoregressive models are linear stochastic models. Taking these limitations into consideration, we propose some future tasks, that is, the Markov random modelling (MRF) and radial basis function modelling can be considered while using nonlinear textures in histopathology imaging. Analysing the histopathology images in wavelet domain partially solves the problem related to higher model order as the correlation length in wavelet domain reduces to half and energy/data compression can also be achieved.

Table 3. Classification Accuracy with Different Model Orders

Test Parameters	Model Order 1	Model Order 2	Model Order 3	Model Order 4
TP	77.78	53.57	66.66	87.50
TN	70.45	57.14	62.50	81.25
FP	29.55	32.14	37.50	18.75
FN	22.32	43.86	33.34	13.50
Sensitivity/recall	77.78	53.57	66.66	87.50
Specificity	70.45	57.14	62.50	81.25
Accuracy	74.11	55.57	64.52	84.50
Precision	72.48	62.49	63.99	82.35
F score	75.74	57.68	65.29	84.85

7. Conclusion

Our work presents a statistical model that uses autoregressive features of histopathology images taken from brain tissues. A survey of histopathology image analyses and its applications in medical diagnostics are also carried out in this work. The estimated AR parameters are excellent discriminating features that are used as the basis for statistical analysis of histopathology images and proved to be useful for cancer diagnosis. Our work reveals that AR features give better results in classification experiments than the conventional approaches which use complex statistical techniques to quantify and classify tissue deformations. The work also shows the classifier performances. According to our results, the fourth-order AR model gives better results than the lower order models, as they are able to capture the deformations of cancerous tissues in detail. As a future work, one can extract more complex features from histopathology image using various statistical models like MRF, RBF, *etc.*, for different types of cancer tissues showing different textures and deformities in their key regions. To reduce the complexity in parameter estimation for higher order models, wavelet domain analysis may also be considered.

References

[1] D. Vaishali, R. Ramesh and A. Christaline, "International Review on Computers and Software", vol. 10, no. 2, **(2015)**, pp. 182-189.
[2] D. C. Ciresan, A. Giusti, L. M. Gambardella and J. Schmidhuber, "MICCAI 2013", 16th International Conference, Nagoya, Japan, **(2013)**, pp. 469-487.
[3] C. Demir and B. Yener, "Automated cancer diagnosis based on histopathological images: a systematic survey", Technical, Report, Rensselaer Polytechnic Institute, Dept. of Computer Science, **(2009)**, pp. 05-09.
[4] N. Stathonikos, M. Veta, A. Huisman and P. J. V, Diest, "Journal Pathology and Informatics", vol. 4, no. 1, **(2013)**, pp. 1-14.
[5] G. A. Meijer, J. A. Belien, P. J. V. Diest and J. P. Baak, "Journal of Clinical Pathology", vol. 50, no. 5, **(1997)**, pp. 365–370.
[6] M. Macenko, M. Niethammer, J. S. Marron, D. Borland, J. T. Woosley, X. Guan, C. Schmitt and N. E. Thomas, "IEEE International Symposium on", USA.
[7] R. A. Walker, "Histopathology", vol. 49, no. 4, **(2006)**, pp. 406-410.
[8] A. D. Belsare and M. M. Mushrif, "Signal & Image Processing", An International Journal (SIPIJ), vol. 3, no. 4, pp. 22-31, **(2012)**.

[9] M. N. Gurcan, T. Pan, H. Shimada and J. Saltz, "Proceedings of the 28th IEEE EMBS Annual International Conference New York City USA", **(2006)**.
[10] C. H. Chen, L. F. Pau and P. S. P. Wang, "The Handbook of Pattern Recognition and Computer Vision", World Scientific Publishing Co.2nd Edition, (Eds.), **(1998)**, pp. 207-248.
[11] R. Chellappa, "Stochastic models in image analysis and processing", Ph.D. Thesis, Purdue University, **(1981)**.
[12] R. M. Haralick, "Proceedings of the IEEE", vol. 67, no. 5, **(1979)**, pp. 786-804.
[13] A. Madabhushi and B. Yener, "Current Methods of Feature Extraction for Microscopic Images", Project Report, International Research Staff Exchange Scheme. (IRSES), **(2009)**.
[14] H. Irshad, A. Veillard, L. Roux and D. Racoceanu, "IEEE Reviews in Biomedical Engineering, vol. 7, **(2014)**.
[15] K. Rodenacker and E. Bengtsson, "Analytical Cellular Pathology", vol. 25, **(2003)**, pp. 1-36.
[16] A. J.Sims, M. K. Bennert and A. Murray, "Physics in Machine and Biology, vol. 48, **(2003)**, pp. 183-191.
[17] A. J. Gill, H. Wu and B. Y. Wang, "Microscopy Research and Technique", vol. 59, **(2003)**, pp. 109-118.
[18] L. E. Boucheron, "Object and Spatial level Quantitive Analysis of Multispectral Histopathology Images for Detection and Characterization of cancer", PhD Thesis, **(2008)**.
[19] O. Sertel, J. Kong, U. Catayurek, G. Lozansski, J. Saltz and M. Gurcan, "Journal of Signal Processing Systems", vol. 55, **(2009)**, pp. 169-183.
[20] R. Albert, T. Schinwolf, I. Baumann and H. Harms, "Cytometry", vol. 13, **(1992)**, pp. 759-765.
[21] C. Bilgin, P. Bullough, G. Plopper and B. Yener, "ECM Aware Cell Graph Mining for Bone Tissue Modelling and analysis", Renselear Polytechnic Institute, Computer Science Technical Report 08-07, **(2008)**.
[22] C. Gunduz, B. Yener and S. Gultekin, "Bioinformatics", vol. 20, **(2004)**, pp.145- 151.
[23] M. N. Gurcan, L. E. Boucheron, A. Madabhushi, N. M. Rajpoot and B. Yener, "IEEE REVIEWS on Biomedical Engineering", vol. 2, **(2009)**, pp. 147-171.
[24] A. N. Basavanthy, S. Ganesan, S. Agner, J. P. Monaco, M. D. Feldman, J. E. Tomaszewski, G. Bhanot and A. Madabhushi, "IEEE Transactions on Biomedical Engineering", vol. 57, no. 3, **(2010)**, pp. 642-53.
[25] K. L. Wiend, C. P. Maier, B. K. Rutt and M. Moussa, "Radiolology", vol. 208, **(1998)**, pp. 477-488.
[26] P. H. Bartels, D. Thompson, M. B. Bibbo and J. E. Weber, "Anal Quant Cyto. Histo. L", vol. 14, **(1992)**, pp. 459-73.
[27] R. Srivastava, S. Sharma and S. K. Singh, "Design, analysis and classifier evaluation for CAD tool for breast cancer detection from digital mammograms", International journal of Biomedical Engineering and technology, vol. 13, no. 1, **(2013)**, pp. 270-300.
[28] M. M. R. Krishnan, M. Pal, S. Bomminayuni, C. Chkraborty, R. R. Paul, J. Chatterjee and A. Roy, "Journal of Computers in Biology & Medicine", vol. 39, **(2009)**, pp. 1096-1104.
[29] S. Doyle, M. Feldman, J. Tomaszewski and A. Madabhushi, vol. 59, no. 5, **(2005)**, pp. 2005-2018.
[30] M. M. Dundar, S. Badve, G. Bilgin, V. Raykar, R. Jain, O. Sertel and M. N. Gurcan, "IEEE transaction on Biomedical Engineering", vol. 58, no. 7, **(2011)**, pp. 1977-84.
[31] A. Baurak, M. Kanemir, C. Sokmensure and C. G. Demir, "Pattern Recognition", vol. 42, **(2009)**, pp. 1104-1112.
[32] K. P. Kannan and A. Ananthakumarari, "International Journal of Management", IT and Engineering, vol. 2, no. 8, **(2012)**, pp. 1-10.
[33] O. A. Kadi, "Journal of Pattern recognition", vol. 43, **(2010)**, pp. 2043-2053.
[34] H. Kong, M. Gurcan and K. B. Boussaid, "IEEE Transactions on Medical Imaging", vol. 30, no. 9, **(2011)**, pp. 1661-77.
[35] R. L. Kashyap and R. Chellappa, "IEEE Transaction on Acoustics", Speech and Signal Processing, vol. 33, no. 1, **(1985)**.
[36] J. K. Hawlic, "Textural properties for pattern recognition", Picture Processing and Psychopictorics, Bernice Sacks Linkman Azriel Rosenfeld (EDS). New York: Academic Press, **(1969)**.
[37] H. McCormick and S. N. Jayaramamurthy, "International Journal Computer. Inform. S & IP, vol. 3, no. 4, **(1974)**, pp. 329-343.
[38] K. Deguchi, "Science on Form: of the First International Symposium on Form 441-449", KTK Science Publishers, and Tokyo, vol. 528, **(1986)**.
[39] R. Nateghi, H. Danyali, M. Sadegh and H. A. Tashk, "Computational Biology Journal", vol. 2014, **(2014)**, pp. 1-9.
[40] J. G. Nieto, A. L. Jourdad and E. Talbi, "Information Processing letters, Elsevier", **(2009)**, pp. 887-896.
[41] M. W. David, "Journal of Machine Learning Technologies", vol. 2, no. 1, **(2011)**, pp. 37-63.
[42] S. N. Japkowicz and S. Szpakowicz, "Proceedings of 19th Australian joint conference on Artificial Intelligence, **(2006)**, pp. 1015-1021.
[43] B. Weyn, G. V. D. Wouwer, A. V. Daele, P. Scheunders, D. V. Dyck, E. V. Marck and W. Jacob, "Cytometry", vol. 33, **(1998)**, pp. 32-40.
[44] "Stanford Tissue Microarray Database", https://tma.im/.

Authors

D. Vaishali, was born in Pune, India, in 1969.She received GRAD IETE in 1993. She received B.E Degree in Electronics and Telecommunication Engineering in 1994 and M. E. degree in Digital Systems in 2002, both from Pune University, India. She is a research scholar in SRM University and her research work is in progress in the field of Image Processing with wavelet Transforms. She is currently working as Asst. Professor in the Department of Electronics and Communication Engineering at SRM University, Chennai. Ind. She is a member of IET.

R. Ramesh, was born in Kanyakumari, India, 1976. He received the B.E. Degree in Electronics and Communication Engineering in 1998 and the M.E. degree in Communication Systems in 2000, both from Madurai Kamaraj University, India. He has been awarded Doctoral degree in the SRM University in the year 2009 for his research work on Testing the Stability of two dimensional recursive filters. He is currently working as a Professor in the Department of Electronics and Communication Engineering at Saveetha Engineering College, Chennai. India. His current research interests concern digital signal processing particularly to find the stability of two dimensional recursive digital filters.

J. Anita Christaline, was born in Dindigul, India in 1974. She completed her B.E. Degree in Electronics and Communication Engineering in 1996 from Bharathidasan University, Trichy. She is recipient of GATE score during 2003. She received her M.E. Degree in Applied Electronics in the year 2006 from Anna University, Chennai. She is currently working as Assistant Professor in SRM University, Chennai where she is pursuing her Doctoral Degree. Her area of interest is Steganography and Steganalysis by computational intelligence.