Quarter Plane ARMA Model for Analysis and Classification of Histopathology Images: Application to Cancer Detection

D. Vaishali¹, R. Ramesh² and J. Anita Christaline¹

¹Department of Electronics and Communication Engineering, SRM University, Chennai – 603203, Tamil Nadu, India; saivaishali@gmail.com, ani_christ@yahoo.co.uk
²Department of Electronics and Communication Engineering, Saveetha Engineering College, Thandalam – 602105, Tamil Nadu, India; ramesh@saveetha.ac.in

Abstract

Objective: Cancer diagnostic using clinical pathology have been proved as a standard method in which histologist/pathologist examines biopsy sample for cell morphology and tissue distribution. Pathologist detects random growth and random placements in tissue samples. These diagnostics are very subjective and based on experience/knowledge base of pathologists. This work presents the use of 2D Autoregressive And Moving Average (ARMA) model in computer assisted automatic cancer detection. Analysis: ARMA model parameters have been considered for representing entire histopathology image. These features have further used for analysis and classification. Parameter estimation has been carried out by Yule walker Least Square (LS) method. Histology images have been classified into healthy and malignant images according to ARMA parameters. K-Fole cross validation has been performed with Linear Kernel support vector machine classifier for classification. Findings: As an outcomes of this experimentation, it is proven that ARMA model parameters works as an excellent discriminating features. These ARMA features are capable of extracting hidden information of the underlying cancer decease. This study also presents the role of neighborhood pixel in image analysis and classification. Improvement: This work have described innovative way of using ARMA features in histopathology imagery and can be implemented in computer assisted diagnosis.

Keywords: Autoregressive Model and Moving Average (ARMA), Markov Random Field model (MRF), Model Based Study, Quarter Plane (QP), Support Vector Machine Minimum (SVM), Texture Analysis, Yule Walker Least Square (YWLS)

1. Introduction

1.1 Background

Study of cell structures, cell morphology and tissues deformities for disease diagnosis have been carried out under the branch called histopathology.¹ As we know mitosis is the most common form of cell division leading to cell multiplication which allows human body to grow. A small change in (mutation) cell's DNA causes cell to become cancerous and this affects the entire life cycle of a cell. Further it causes uncontrolled mitosis activity. Uncontrolled mitosis produces more and more damage cells which further accumulate at one region to produce tumor. Detection/study of uncontrolled mitosis becomes complex as entire process becomes random.² Detection and examination of mitotic figures in cancer screening is very important. Histologist used to do this cancer screening in early days but then the reliability of cancer diagnosis have been restricted/subjective to knowledge base of histologists. To eliminate this computer assistance have been taken and have developed Computer Assisted Diagnosis (CAD). It has been proved that CAD reduces time, cost and errors in diagnosis.
1.2 Recent Developments
In past few years with advancement in technology, we have seen advancement in biomedical instrumentation, in scanning techniques, in staining protocols. Newly introduced Tissue Micro Array (TMA) technology in pathology labs have proven simple but revolutionary. The high end computer resources making computation simpler and faster. High resolution monitors motivate to observe and analyse digitised slides with better accuracy. The introduction of tissue microarray in pathology labs encourages to study hundreds of tissue sample simultaneously which reduces subjective diagnostic errors. In recent past there are many algorithms and many protocols developed to achieve better accuracy in diagnosis. Tissue preparation is the first step in histopathology image analysis. Many standard staining protocols have been introduced. They are Hematoxylin & Eosin (H&E) and Immuno Histo Chemicals (IHC) which gives better visibility of the structures on the slide. In feature analysis, spatial nuclei features, textural features, and spatial arrangements of tubules, stromal, etc. have properly seen with H&E stain. To study histopathology samples for chromatin specific features IHC stain has been used. The design of CAD system takes place in three basic steps, pre-processing, feature extraction and classification. The objective of pre-processing is to enhance image quality by color normalisation, eliminating auto-fluorescence and noise filtering. In next step image has been converted into set of extracted features. The set of features have changed according to method opted to represent an image. They can be statistical features, model based features or transform base features. In third step classification of feature datasets have been performed. The classifier distinguishes healthy tissues from malignant tissues and also detect the different levels of malignancy. Finding best classifier or classifier combination for given feature set is very important to classify sample in correct class.

1.3 Contribution
In cancer, the cells/tissues show randomness in their growth (cell-based features) and in their placement (tissue-based features). That is why researchers have suggested textural feature analysis. The textural properties computed are dependent on cancer type. This work assumes the entire image as a texture and hence it has been processed with random field measures. While processing/analysing textured images, we have to adapt deterministic or stochastic approaches according to degree of randomness involved. In model based study of texture analysis this work presents use of ARMA model parameters to represent entire histopathology image. The main focus of this work is to study ARMA model for histopathology image analysis and design the robust automated diagnostic system for brain cancer histopathology.

1.4 Work Carried Out in Past
Many researchers have acknowledged the importance of histopathological image analysis in cancer diagnosis. With technological advancement in imaging systems, computer resources, and high resolution display systems, analysis of this complex diseases have become relatively simple, easier and faster. The study of features related to cell nuclei, their shape, morphology, cell deformities has been referred as cytopathology. Histopathology mainly classified into analysing object level attribute and spatial attributes. In Object-level features we study size and shape of structures, radiometry and densitometry features, textural features and chromatin specific features. This work has been based on analysing textural features. Many ways have been proposed to study textural properties of an image. They are like co-occurring matrix features, fractal features, run length features, statistical model based features and wavelet features etc. Researchers in biomedical image processing have worked on different ways to analyse textural images. Some of them are Relative Neighbour Graphs, Minimum spanning tree, Connected Graphs, Voronol Tessellation, K-NN Graphs etc. Graph theory plays important role while studying different states of tissue/cell developments. Study of graph consists of 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. The entire study of histopathology images with different approaches have been reviewed by M. Gurucan et.al. We have performed survey of some work done in past with histopathology images for cancer diagnosis. Lymphocytic infiltration is the major activity happens in progress of cancer. Study of this activity in breast cancer carcinomas and fibroadenomas has been carried out by Ajay Basavanahally. The use of standard staining protocols always enhances the disease specific features in an image as stated above. Activity related to P53 protein biomarker has been studied with Immuno Histo Camical (IHC) in breast cancer detection.
of histopathology, use of Autoregressive model for brain cancer diagnosis is given in our previous work.\textsuperscript{24} In study histopathology images for cancer diagnosis selection of right classification strategy is also important. Range of classifier algorithms have been proposed by researchers in biomedical image processing field. Some of them are Support Vector Machine (SVM), neural network Multi-Layer Perceptron (MLP), Bayesian, Fuzzy systems, K-Nearest Neighborhood (KNN) etc. The comparison of their performance for breast cancer detection has been presented.\textsuperscript{25} The use of SVM classifier for diagnosis of oral sub-mucus fibrosis has been given in:\textsuperscript{26} Scott Doyle et al. have described the Bayesian classifier for prostate cancer.\textsuperscript{27} To overcome weaknesses of individual classifier researchers have suggested to work with Multi-Classifiers Systems (MCS) and hybrid way of classification for proper grading of cancer. Use of multi classifier with K-Ratio Super Item Set Findings (KRSIF) and Nearest Neighborhood Classifier (NNC) in cancer grading is given by S. Padmapriya et al.\textsuperscript{28} S. N. Deepa et al. have presented survey on medical image classification using artificial intelligence approach.\textsuperscript{29} Use of expectation maximization and watershed transform and classification in study of intra-ductal breast lesions have been given in this work.\textsuperscript{30} In textural feature analysis Akif Burak et al. have presented treatment for unsupervised segmentation in cancer diagnosis.\textsuperscript{31} In textural feature analysis use of Graph Run Length Matrix (GRLM) and Grey-Level Co-Occurrence Matrix (GLCM) for detection of abnormal tissue presented by K. P. Kannan et al.\textsuperscript{32,33} In transform based study, use Gabor and wavelet filters for textural feature extraction have been discussed for meningioma.\textsuperscript{34}

The rest of the article is organized as follows. Section 2 describes the basics of stochastic models, it also details the representation of ARMA model and parameter estimation procedure and classification. Section 3 gives experimentation results and Section 4 we draw the conclusion.

\section*{2. Methods}

\subsection*{2.1 Stochastic Models}

In model based study of stochastic field, many researchers have studied modelling with autocorrelation function, covariance function. To study linearity and non linearity of pixel grey levels in images they have suggested to use Auto-Regressive (AR) models, Auto-Regressive And Moving Average (ARMA) models, Markov Random Field (MRF) models, Radial Basis Function (RBF) models.\textsuperscript{35} Selection and use of proper neighborhood to describe histopathology image has been discussed by Heralick. He has described stochastic model parameters used for studying underlying image properties and also given their use in image synthesis.\textsuperscript{36} We have encountered few problems while working with histopathology images. They are like selection of proper model, neighborhood, selection of right method to estimate model parameters and finely selecting right classifier.

\subsection*{2.2 ARMA Models}

Autocorrelation function estimated for random field basically describes grey level dependency of one pixel on other pixels in neighbourhood. McCormick and Jayaramamurthy presented work to describe linear dependency with an auto regression model.\textsuperscript{37} Deguchi and Morishita\textsuperscript{38} and Tou and Chang\textsuperscript{39} have also used a similar technique to synthesize textures. The homogeneous random field of cancer lesions in ultrasound images have studied by ARMA model than AR model.\textsuperscript{40} In following section we have presented basic ARMA model extension to our previous work.\textsuperscript{24}

\subsection*{2.2.1 Representation of 2D ARMA Model}

Texture image in two dimensional spatial plane represented as:

\begin{equation}
I = \{x[n,m]: n \leq N1-1 & m \leq N2-1\} \tag{1}
\end{equation}

Grey level value at \([n,m]\) coordinates is represented by \(x[n,m]\). The size of the image is \(N1 \times N2\).

In difference equation form:

\begin{equation}
x[n,m] + \sum_{(i=0,j=0)}^{(i=p1,j=p2)} \sum_{(i=0,j=0)}^{(i=q1,j=q2)} a_{ij}x[n-i,m-j] + \sum_{(i=0,j=0)}^{(i=p1,j=p2)} b_{ij}w[n-i,m-j] \tag{2}
\end{equation}

\begin{equation}
\{w [n,m]\}, \{a_{ij},b_{ij}\} \text{ represents white noise, variance and parameter vectors. If we assume image I is an output of linear system excited by white noise its transfer function}
\end{equation}

\begin{equation}
H(z_1,z_2) = \frac{B(1,z_2)}{A(1,z_2)} = \frac{\sum_{(i=0,j=0)}^{(i=p1,j=p2)} b_{ij}[z_1-i,z_2-j]}{\sum_{(i=0,j=0)}^{(i=p1,j=p2)} a_{ij}[z_1-i,z_2-j]} \tag{3}
\end{equation}
2.2.2 Yule-Walker Least Square parameter Estimation

Yule Walker equations are set of linear system equations normally shown in matrix form to represents linear system with system parameters. Noise sequences w[n, m] is known, and apply as a input to get out as a texture image using estimated ARMA parameters. It has seen that the simplest method of parameter estimation is Least–square method. In the LS method we express:

\[ X[n,m]+\Phi[n, m] \theta = w[n, m] \]  \hspace{1cm} (4)

Writing (4) in matrix form

\[ X + \Phi \theta = W \]  \hspace{1cm} (5)

\[ X = [x[l+1,M+1]……x[N1-1,M-1]]^t \]
\[ W = [w[l+1,M+1]……w[N1-1,M-1]]^t \]

Assume that we know \( \Phi \), then we can obtain a least squares estimate \( \theta^* \) of parameter vector \( \theta \) by

\[ \theta^* = (\Phi^t \Phi)^{-1} \Phi^t X \]  \hspace{1cm} (6)

For ARMA \((p1,p2,q1,q2)\), the parameters estimation of ARMA can be performed with 2D extension of the Yule-Walker equations. \( r[k, l] \) is the autocorrelation values of the \( \{x[n,m]\} \),

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

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

Equation (7) represents linear system.

Apply least square principle for ARMA parameter estimation through Yule walker equations. Repeat the procedure till we achieved least square estimate. In brief it is:

- Read input texture image, resize it.
- Generate random image with Gaussian Distribution and with variance \( \sigma^2 \).
- Use Yule-Walker method in given by Eq(7), estimate AR parameters forgiven neighbourhood \( N \).
- Estimate noise field \( w[n,m]^\wedge \).

\[ w[n,m]^\wedge = x[n,m] + \sum_{i=0}^{p1} \sum_{j=0}^{p2} x[n-i,m-j]a_{ij} \quad n = p1+1 \text{ to } N1, \quad m = p2+1 \text{ to } M1 \]  \hspace{1cm} (9)

Replace \( w[n, m] \) by \( w[n , m]^\wedge \) computed in step 2 and obtain \( \theta^* \) in Eq (6) with \( L = p1+q1 \), \( M = p2+q2 \).

Repeat it to get least square estimate of AR parameters. Substitute AR parameters and repeat the same for MA parameters.

Figure 1. Quarter plane causal and Non Symmetrical Half Plane Semi causal Model (NSHP).

2.3 Classification

We need a proper classifier to classify normal and malignant samples. In some experiments classifier have also been used to grade malignancy. Classification algorithms mainly based on statistical tests or follows machine learning. Simple classification is based on statistical measures, but samples are not fully independent in this study, as they have been taken from same patient. Because of this classification might not be robust or reliable. In supervised machine learning, classifier first learn to classify through training and then classify test data. There are four basic classifier evaluation strategies 1) no separate training and testing data, 2) separate training and testing data, 3) K-fold cross validation technique, 4) leave one out technique. In our study the support vector machine algorithm has been considered to classify the malignant tissues from the healthy ones. We have opted K-fold cross validation technique to evaluate classifier performance. In this method we have divided dataset in to K parts. K-1 parts, have been considered for training and remaining for testing. This process have been repeated for K times to find out final accuracy.

2.4 Model Optimization: Choice of Neighbourhood (N) for ARMA Model

Optimization of ARMA model in this study is done by checking the classification accuracy. Criteria selected are
very simple. We have declared optimum neighborhood selection have been achieved when we get maximum classification accuracy. We have estimated parameters with different model orders and form datasets. These datasets have then given to classifier and accuracies have been found. In simple words 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.

3. Experimentation

3.1 Experiment Setup

We have presented this work in MATLAB, using image processing and statistical toolboxes. We have used Tissue Microarray Images made available by Department of Biochemistry, School of Medicine, Stanford University. This is an open source to do research in cancer diagnosis with pathology. Figure 2 shows sample healthy and malignant brain tissues. As a first step in this work, we have cropped the image of portion of 256x256 pixels and perform noise filtration, color normalisation and removed effect of auto fluorescence. In the next step we have estimated ARMA parameters using Yule-Walker least square method and then formed respective datasets with varying model orders. In the next step K-fold cross validation technique has been implemented with SVM classifier. We have set the value of K as 10 and number of iteration to 50.

3.2 Results

Two dimensional quarter plan neighborhood has been given in Figure 3. Table 1 and Figure 1 represents all the pixel under consideration with varying model orders.

Table 2 shows estimated autoregressive and moving average parameters for both normal and cancerous tissue samples. We observed that as the model order increases, number of parameters for estimation also increased. This leads to estimation process computational intensive. Performance evaluation of ARMA model with different AR and MA model orders has been given in Table 3. Experimental results reveals that accuracy calculated with ARMA (1,1,1,1), first order model is enough to classify histopathology images. Results hows that the higher order models are not appropriate for this work. The better representation of ARMA model have been possible with large neighborhood in synthesis experiments, but at the same time it is harmful for classification as it might include pixels from other classes and this causes misclassification.

3.3 Discussion

In model based study of histopathology images this work shows that ARMA model presents good results in cancer diagnosis. Textural feature analysis has been proved as better tool to analyse random field of histopathology images because random growth in cell and general randomness seen in entire mitotic activity. There have been many more other methods to study textural feature, researchers have attempted work mainly with the co-occurring matrix features, fractal features, run length features, wavelet features. The work presented in this paper proves ARMA model parameters have a good discriminating properties to classify malignant and normal tissue and hence can be used in automated cancer diagnosis process. Ideally it is not recommended to compare numerical results of many approaches as their base cancer images are different, the cell and tissue structures at various parts of the body are not same. So the diagnostic approach which works for one cancer might not work for another. This work proposes a new approach to extract underlying disease
specific features. In the proposed method we encounter some limitations. First limitation is parameter estimation is much complex compare to other methods (AR, MRF). As ARMA model considers linear data, this approach is not suitable to handle non linearity present in histopathology imagery. In future one can implement MRF and RBF models for histopathology images to handle nonlinearity. Many researchers have suggested to using fusion classification instead of single classifier so that weak classifier output can be compensated by good classifier and overall classification result can be improved. To analyse the heterogeneity and long range dependency in histopathology images, one can perform the same work in wavelet domain.

4. Conclusion

The use of ARMA model for cancer diagnosis has been presented in this work. Work performed in past decades have also discussed. The results show that ARMA model

<p>| Table 1. Spatial Quarter Plane Causal Neighborhood for Model. Pixel Under Consideration = X(i,j)=X(0,0) |</p>
<table>
<thead>
<tr>
<th>Model Order 1</th>
<th>Model Order 2</th>
<th>Model Order 3</th>
<th>Model Order 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>(i,j-1), (i-1,j-1), (i-1,j)</td>
<td>(i,j-1),(i-1,j-1),(i-1,j),(i,j-2), (i-2,j-2), (i-2,j-1), (i-2,j), (i-2,j-0)</td>
<td>(i,j-1),(i-1,j-1),(i-1,j),(i,j-2), (i-1,j),(i-1,j-2), (i-2,j-3),(i-3,j-3),(i-3,j-2), (i-3,j-1),(i-3,j)</td>
<td>(i,j-1),(i-1,j-1),(i-1,j),(i,j-2), (i-1,j),(i-1,j-2), (i-2,j),(i-2,j-1), (i-2,j-2),(i-2,j-3), (i-3,j-2),(i-3,j-3),(i-3,j-4), (i-3,j-4),(i-4,j-4), (i-4,j-4),(i-5,j-4)</td>
</tr>
</tbody>
</table>

<p>| Table 2. Estimated Ar and Ma Coefficient for Healthy Brain and Malignant Brain |</p>
<table>
<thead>
<tr>
<th>Histopathology Image</th>
<th>ARMA(1,1,1,1)</th>
<th>ARMA(2,2,2,2)</th>
<th>ARMA(3,3,3,3)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AR Parameters</td>
<td>MA Parameters</td>
<td>AR Parameters</td>
<td>MA Parameters</td>
</tr>
<tr>
<td>Healthy Brain</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.5589</td>
<td>0.2492</td>
<td>0.2138</td>
<td>0.2138</td>
</tr>
<tr>
<td>0.1845</td>
<td>-1138</td>
<td>-1138</td>
<td>-1138</td>
</tr>
<tr>
<td>Malignant Brain</td>
<td>1.0000</td>
<td>1.0000</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.7332</td>
<td>0.4090</td>
<td>0.4090</td>
<td>0.4090</td>
</tr>
<tr>
<td>0.4302</td>
<td>0.1734</td>
<td>0.1734</td>
<td>0.1734</td>
</tr>
<tr>
<td>0.0854</td>
<td>0.0747</td>
<td>0.0747</td>
<td>0.0747</td>
</tr>
</tbody>
</table>

<p>| Table 3. ARMA (P1, P2, Q1, Q2) Model Classification Efficiency with SVM Classifier |</p>
<table>
<thead>
<tr>
<th>Classifier</th>
<th>ARMA(1,1,1,1)</th>
<th>ARMA(2,2,2,2)</th>
<th>ARMA(1,1,2,2)</th>
<th>ARMA(1,1,4,4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Support vector Machine (SVM)</td>
<td>82.78%</td>
<td>64.55%</td>
<td>77.80</td>
<td>62.28</td>
</tr>
</tbody>
</table>
proved to be one of the best approaches to analyse histology imagery. The estimated ARMA coefficients performs well in finding out discriminating features and can be used for second order statistical analysis of histopathology images. The SVM classifier with K-fold validation technique has been proved better in performance evaluation of classifier. To analyse nonlinearity in sample images as a future task, one can perform the same study with MRF and RBF models. For heterogeneity issue same work can be carried out in wavelet domain and to improve classifier performance one can opt fusion classification approach.

5. Acknowledgement

The author would like to thanks management S.R.M. University, Chennai for their continuous support and motivation during this work.

6. References