

Classification of Osteosarcoma T-ray Responses Using Adaptive and Rational Wavelets for Feature Extraction

Desmond Ng, Wong Fu Tian, Withawat Withayachumnankul, David Findlay, Bradley Ferguson and Derek Abbott

Centre for Biomedical Engineering (CBME) and Department of Electrical & Electronic Engineering, The University of Adelaide, SA 5005, Australia

ABSTRACT

In this work we investigate new feature extraction algorithms on the T-ray response of normal human bone cells and human osteosarcoma cells. One of the most promising feature extraction methods is the Discrete Wavelet Transform (DWT). However, the classification accuracy is dependant on the specific wavelet base chosen. Adaptive wavelets circumvent this problem by gradually adapting to the signal to retain optimum discriminatory information, while removing redundant information. Using adaptive wavelets, classification accuracy, using a quadratic Bayesian classifier, of 96.88% is obtained based on 25 features. In addition, the potential of using rational wavelets rather than the standard dyadic wavelets in classification is explored. The advantage it has over dyadic wavelets is that it allows a better adaptation of the scale factor according to the signal. An accuracy of 91.15% is obtained through rational wavelets with 12 coefficients using a Support Vector Machine (SVM) as the classifier. These results highlight adaptive and rational wavelets as an efficient feature extraction method and the enormous potential of T-rays in cancer detection.

Keywords: Terahertz time-domain spectroscopy, T-rays, osteosarcoma, adaptive wavelets, rational MRA, feature extraction, signal classification, cancer detection, Bayesian classifier

1. INTRODUCTION

Cancer is one of the most potent killers of humanity. A common form of bone cancer is *osteosarcoma*¹ which is a condition where malignant cancer cells are found in the bone. The probability of recovering from such a cancer is dependent on the stage at which the cancer is diagnosed. Therefore early detection could be a matter of life and death.

In the early stages of cancer, malignant cells are hard to distinguish from normal cells. Common diagnosis techniques in practice now are X-Rays, Magnetic Resonance Imaging (MRI), and Computerized Tomography (CT) scans. Using these imaging methods, a biopsy of suspected cancer cells is then performed. Biopsy is an invasive technique where surgery is performed to obtain suspected bone cells or tissues for further analysis. Current treatments for bone cancer are chemotherapy or complete surgical removal of cancerous cells. For an effective treatment using both these methods, cancerous cells need to be detected in the early stages and a clear distinction from normal cells is required. Therefore, an accurate and quick method of detecting cancerous cells is required to increase the patients chances of survival.

THz-TDS is an emerging technology, which gives rise to a whole range of applications and still has huge potential to be further explored and improved.² T-rays generated and detected by THz-TDS are unique in biosensing applications due to the information they provide on a macromolecular level as compared to the individual stretching of atomic bonds and does not suffer much from Rayleigh scattering as compared to infrared. Hence, these qualities form the motivation for applying THz-TDS to classify between human osteosarcoma cells (HOS) and normal human bone (NHB) cells.

Through T-ray imaging, it is possible to be able to detect the existence of these malignant cancer cells. T-rays provide a more precise level of image differentiation than X-rays at shallow depths. T-rays are also non-ionizing, therefore making it appealing for a non-invasive diagnostic technique.

Further authors information:

D. Ng, E-mail: des_n@yahoo.com; F. T. Wong, E-mail: dannywft@gmail.com

W. Withayachumnankul, E-mail:withawat@eleceng.adelaide.edu.au; D. Abbott, E-mail: dabbott@eleceng.adelaide.edu.au

Complex Systems II, edited by Derek Abbott, Tomaso Aste, Murray Batchelor, Robert Dewar,
Tiziana Di Matteo, Tony Guttman, Proc. of SPIE Vol. 6802, 680211, (2008)
0277-786X/08/\$18 · doi: 10.1117/12.753026

Proc. of SPIE Vol. 6802 680211-1

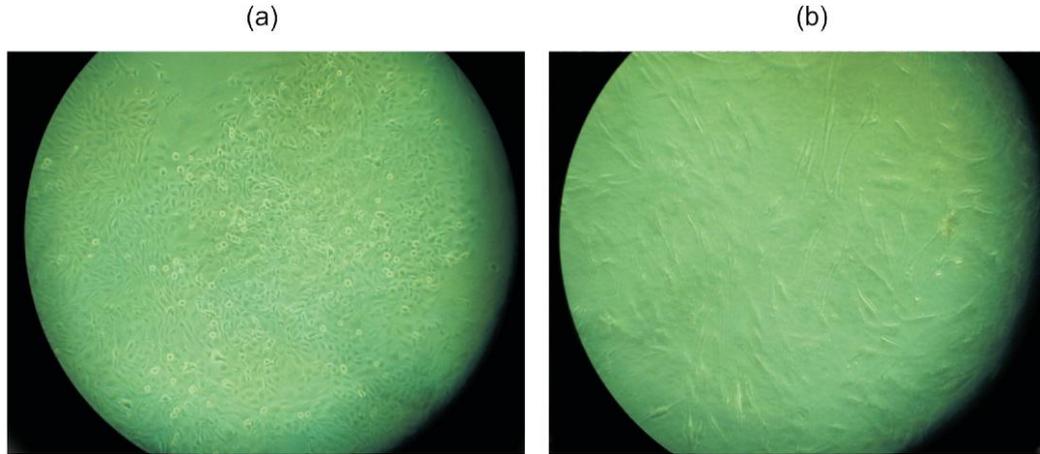


Figure 1. Optical microscope images of (a) Human osteosarcoma cells (b) Normal human bone cells. These images are of cells grown on a petri dish, and can be visually distinguished. However, optical discrimination is not so clear⁶ during surgery where there is a need to ensure all cancerous regions are truly removed. T-rays are a possible modality that may allow a surface scan during the process of surgical removal.

2. PREVIOUS WORK

The application of T-rays as a cancer detection modality has been explored mainly on bulk tissue samples.³⁻⁵ Ferguson *et al.*⁶ tackled this problem from a cellular level and promising results were obtained. The motivation for investigating the HOS and NHB cells at a cellular instead of bulk excised tissue is because of a large number of factors influencing the measured terahertz response and changes in the concentration of water molecules and its bonds are likely to dominate the response.

Ferguson *et al.*⁶ used the deconvolved magnitude and phase frequency components as features for classification. Genetic algorithms (GA) were employed as a feature selection method to select the optimal features. The classification accuracies were used as a fitness function, and the GA iteratively searches features which gives the highest accuracy. However this computation was done at the expense of time.

Withayachumnankul *et al.*⁷ extended this work by improving the generalization ability of classifiers and the speed of feature selection process, using sophisticated feature subset selection methods. However, previous works so far utilizes features from the time and frequency domain. Also, Yin *et al.*⁸ utilized wavelet coefficients as input features for classification and successfully obtained a high degree of accuracy. However, the best wavelet base was selected using a ‘trial and error’ method. This motivates the work on a feature extraction method base on wavelets which automatically chooses the best wavelet base.

3. CELL PREPARATION AND MEASUREMENT

The NHB cells were cultured from small pieces of trabecular bone, which is the spongy interior of a bone for 4-6 weeks. The HOS cells were cultured from an immortalized cell line and a confluent culture was obtained within 1 week. The cells were cultured in 5 mL of Dulbeccos Modified Eagle Media (DMEM) supplemented with L-Glutamine (0.29 g/l), 10% foetal bovine serum and Gentamicin (16 $\mu\text{g}/\text{ml}$) as an antibiotic in 25 ml polystyrene flasks. Both of the cell types contain the same media and cultured in a 5% carbon dioxide environment at 37°C to ensure a constant environment setting. Optical microscope images of the cells are shown in Figure 1. Once confluent, the cells form a thick dense layer on the bottom of the flask.

Using a scanned THz imaging system,⁹ T-ray pulses are incident on HOS and NHB cells grown in transparent plastic flasks. A lock-in amplifier is set with a time constant of 10 ms. The ultrafast pulse laser was a regeneratively amplified Ti:sapphire laser producing 130 fs pulses at 0.7 W with a 1 kHz repetition rate. Also, 2 and 4 mm thick <110> oriented ZnTe crystals were used as the THz emitter and detector respectively. Three

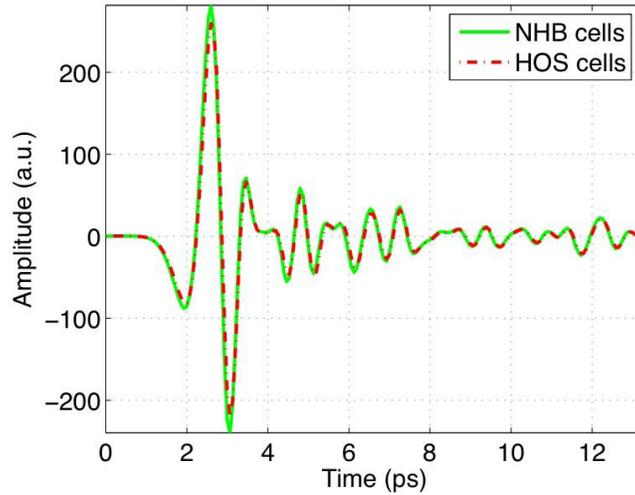


Figure 2. The average T-ray time domain waveform of HOS and NHB cells.

identical flasks were used. The first two contained confluent HOS cells and confluent NHB cells in cell media. The third flask was used as a reference and contained only the cell media solution.

A THz image was obtained providing spectroscopic information at 48 different positions or pixels. This then provided sufficient data to allow statistical classification algorithms to be used to attempt to differentiate the cells. Each waveform at a location contained 200 data points sampled every 0.067 ps for a duration of 13.33 ps. The average time domain waveforms for HOS and NHB cells are shown in Figure 2.

4. FEATURE EXTRACTION

4.1. Adaptive Wavelets

Classical Discrete Wavelet Transforms are carried out by firstly selecting a particular wavelet base such as Daubechies, Coiflets, and Haar. Mallet *et al.*¹⁰ proposed a new form of wavelets, which gradually adapts to the signal relative to a discriminant criterion. This section presents a succinct description of the algorithm used to generate adaptive wavelets.

4.1.1. m-band Discrete Wavelet Transform (DWT)

For a general m-band DWT, there is one low pass filter and $m - 1$ high pass filters. The number of coefficients in each band is $1/m$ the previous level. The low pass filter are denoted $\ell = (\ell_0, \dots, \ell_{N_f-1})$ and high pass filters $\mathbf{h}^{(s)} = (h_0^{(s)}, \dots, h_{N_f}^{(s)})$ where $N_f = m(q + 1)$. A choice of q determines the length of the filters, which relates to the smoothness of the wavelets. The filtering operations from level l to $l + 1$ are given by

$$c_{l+1,k} = \sum_{i=0}^{N_f-1} \ell_i c_{l,mk+i}, \quad (1)$$

$$d_{l+1,k}^{(s)} = \sum_{i=0}^{N_f-1} h_i^{(s)} c_{l,mk+i} \quad s = 1, \dots, m - 1. \quad (2)$$

The m-band DWT can be expressed in matrix form by

$$x^{[l+1]} = W^{[l]} x^{[l]}(0), \quad (3)$$

where W is the block-Toeplitz structure of the low and high pass filter coefficients.¹¹

The low and high pass filters can be represented by the wavelet matrix, A . For example $m = 3, q = 1, N_f = 6$ the wavelet matrix has the form

$$A = \begin{pmatrix} \ell_0 & \ell_1 & \ell_2 & \ell_3 & \ell_4 & \ell_5 \\ h_0^{(1)} & h_1^{(1)} & h_2^{(1)} & h_3^{(1)} & h_4^{(1)} & h_5^{(1)} \\ h_0^{(2)} & h_1^{(2)} & h_2^{(2)} & h_3^{(2)} & h_4^{(2)} & h_5^{(2)} \end{pmatrix}.$$

4.1.2. Filter Coefficient Conditions

For a MRA and wavelet basis to exist, the following conditions must be satisfied

Shifted Orthogonality $\sum_k A_k A_{k+i}^T = \delta_{0i} I$

Basic Regularity Condition $\sum_k \ell_k = \sqrt{m}$

Lawton Matrix Condition $M_{ij} = \sum_k \ell_k \ell_{k+j-mi}$, where the M has to have an eigenvalue and its multiplicity equal to 1.

4.1.3. Parameterization of the Filter Coefficient Matrix

Optimizing over all filters coefficients (mN_f parameters) would be a computationally demanding task. Therefore, the wavelet matrix A can be parameterized so that the optimization parameters are reduced to $(N_f - 1)$.

Given that the condition stated in Section 4.1.2 are satisfied, the wavelet matrix can be factorized into

$$A = Q \square F_1 \square \dots \square F_q, \quad (4)$$

where \square denotes the 'polynomial product'. Here, $F_i = (R_i I - R_i)$, where the symmetric projection matrix $R_i = u_i u_i^T$.

The orthogonal matrix Q is composed of

$$Q = \begin{pmatrix} \frac{1/\sqrt{m}\mathbf{1}}{(I - 2vv^T)TD} \end{pmatrix}, \quad (5)$$

where T is an upper triangular matrix with diagonal elements $T_{ii} = i - m$ and off-diagonal elements equal to 1. Also, D is a matrix that normalizes the rows of T . Thus, the wavelet matrix can be constructed by normalized vectors u_1, \dots, u_q and v . Initially these vectors are randomly assigned from the uniform distribution.

4.1.4. Discriminant Criterion

The adaptive wavelets are used to represent the signal such that the discriminatory information is optimized. Therefore the adaptive wavelets are optimized relative to a measure which reflects differences among classes.

The criterion is based on posterior probabilities known as the Cross Validated Quadratic Probability Measure (CVQPM) which can be computed by

$$\text{CVQPM}(X^{[l](t)}) = \frac{1}{n} \sum_{i=1}^n a_{Q/i}, \quad (6)$$

$$\text{where } a_{Q/i} = \frac{1}{2} + P(r|X_{i(r)}^{[l]}(t)) - \frac{1}{2} \sum_{r=1}^R P_{/i}(r|X_i^{[l]}(t))^2, \quad (7)$$

where $P(r|X_{i(r)}^{[l]}(t))$ is the posterior probability for the true class of $X_i^{[l]}(t)$ and $P_{/i}(r|X_i^{[l]}(t))$ is the posterior probability for $X_i^{[l]}(t)$ belonging to class r . These posterior probabilities can be calculated as outlined in Section 5.1.1. The CVQPM has a range of values between 0 and 1. A high CVQPM value implies that the classes can be well differentiated. Figure 3 shows the flowchart for adaptive wavelets algorithm.

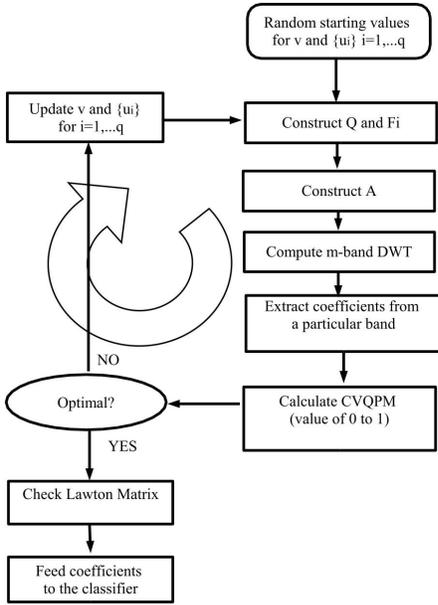


Figure 3. The Adaptive Wavelet algorithm.

4.2. Rational Wavelets

The classical Multiresolution Analysis (MRA) introduced by Mallat¹² is a special case of the rational MRA, whereby the scaling factor used was 2, and thus called dyadic scaling. However dyadic MRA may not provide a good enough separation of signal components for certain cases. This motivates the need to explore the application of rational wavelets in classification processes, where rational MRA may be used to allow a better adaptation of the scale factor according to the signal.¹³ The scaling factor of rational wavelets is p/q . The dyadic case is when $p = 2, q = 1$. By having a fractional dilation factor, it may provide a sharper frequency localization.¹⁴ As classical MRA is a specific case of the rational MRA, thus the same theorems that govern classical dyadic MRA apply to rational MRA too.

4.2.1. Algorithm

The section briefly covers the main algorithms used in this project which are based on the work by Baussard *et al.*¹³ For analysis (i.e. decomposition of the signal), a pyramidal scheme is used. Figure 4 shows that the signal is decomposed into its approximate and detail coefficients. For each level, the number of detailed coefficients is $(1 - p/q)$ of the approximate signal. Note that the original signal is treated as an approximate coefficient at level 0, (i.e. cA0).

To obtain the approximate coefficients of the other levels, the following equation is used,

$$a_{j,sq+i} = \sum_r \tilde{h}_i[ps - r]a_{j-1,r} , \quad (8)$$

where $s \in \mathbb{Z}, i = 1, \dots, q - 1, h_i[n] = \langle \varphi_{0,i}, \varphi_{-1,n} \rangle$ is the impulse response of the q numerical filters and $\tilde{h}_i[n] = h[-n]$. This equation shows the decomposition from a higher level approximation a_{j-1} to a coarser approximation a_j .

Similarly, the detail coefficients of the signal are obtained by,

$$d_{j,n}^m = \sum_r \tilde{g}_m[np - r]a_{j-1,r} , \quad (9)$$

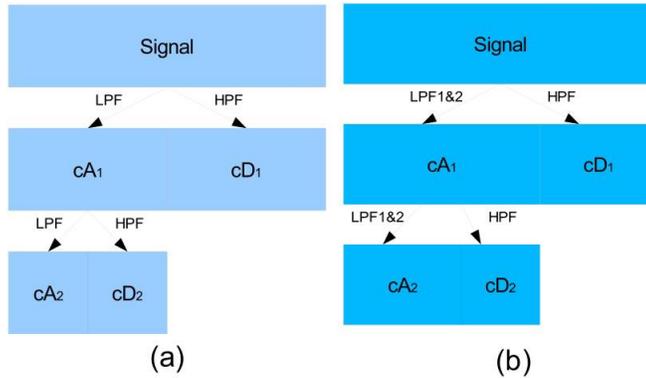


Figure 4. Decomposition of a signal into its approximate and detail coefficients. (a) The classical scaling (dyadic) with $p = 2$, $q = 1$. (b) Rational scaling of $p = 3$, $q = 2$.

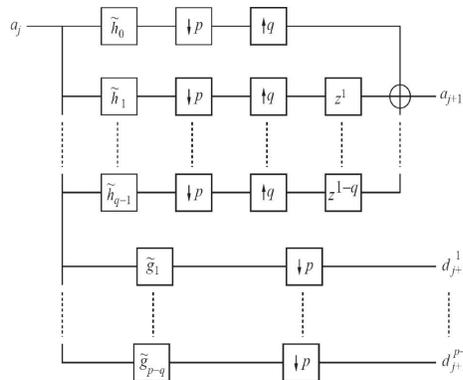


Figure 5. Pyramidal Analysis Algorithm. After Baussard *et al.*¹³

which is a convolution between a_{j-1} and the filter \tilde{h}_m and a down-sampling of p . Figure 5 is a block diagram describing the structure of the algorithm for a general case the rational number of $M = p/q$ is used.

To practically implement rational wavelets, the filters need to be appropriately designed. This is because the filters used determine the shape of the wavelets that are used. As will be seen later, that rational orthogonal wavelet bases are defined in the Fourier domain, therefore it would be good to define the filters in the Fourier domain instead. Furthermore, implementation in the Fourier domain avoids truncation.

$$\tilde{h}_n(\omega) = \sqrt{M} \frac{\hat{\varphi}(M\omega)}{\hat{\varphi}(\omega)} e^{-inM\omega}, \quad (10)$$

$$\tilde{g}_m(\omega) = \sqrt{M} \frac{\hat{\psi}^m(M\omega)}{\hat{\varphi}(\omega)}. \quad (11)$$

These rational wavelet bases are based on Meyer construction extended to the rational case, as proposed by Auscher¹⁵ and it requires to fulfil certain conditions to be considered as an Auscher basis for the rational case. The scaling function $\Phi(t)$ and wavelet function $\Psi(t)$ used is¹⁶

$$\Phi(\omega) = \begin{cases} (2\pi)^{\frac{1}{2}}, & |\omega| \leq \omega_1 \\ (2\pi)^{\frac{1}{2}} \cos\left(\frac{\pi}{2}\beta\left(\frac{q}{\omega_1}|\omega| - q\right)\right), & \omega_1 \leq |\omega| \leq \omega_2 \\ 0, & |\omega| > \omega_2, \end{cases}$$

$$\Psi(\omega) = \begin{cases} (2\pi)^{\frac{1}{2}} e^{j\frac{\omega}{2}} \sin(\frac{\pi}{2}\beta(\frac{q}{\omega_1}|\omega| - q)) & \omega_1 \leq |\omega| \leq \omega_2 \\ (2\pi)^{\frac{1}{2}} e^{j\frac{\omega}{2}} \cos(\frac{\pi}{2}\beta(\frac{q}{\omega_2}|\omega| - q)) & \omega_2 \leq |\omega| \leq \omega_3 \\ 0 & |\omega| \notin [\omega_1, \omega_3] \end{cases}$$

where $\omega_1 = (q - \frac{q}{2q+1})\pi$, $\omega_2 = a\omega_1$, $\omega_3 = a\omega_2 = a^2\omega_1$ and $\beta(t) = t^4(35 - 84t + 70t^2 - 20t^3)$ is the construction function.

5. CLASSIFICATION AND VALIDATION

The problem of classification is a complex one because of the No Free Lunch theorem.¹⁷ This theorem states that, given an ideal general set of data, no classifier is better than another. A highly accurate result from a particular classification algorithm is due to the physical properties of the problem that include data distribution, prior knowledge, and size of training data, rather than the classification algorithm itself. The main focus of this project is the utilization of good feature extraction techniques, which result in low dimensions. Hence, low dimensional Bayesian classifiers are explored.

5.1. Bayesian Classification

Bayesian classifiers present a statistical viewpoint to solving classification problems.¹⁸ The main approach relies on the Bayes decision rule with assigns test features to a corresponding target or class such that it maximizes the posterior probability. Using Bayes theorem, the posterior probability is written as

$$P(\omega_i|x) = \frac{p(x|\omega_i)P(\omega_i)}{p(x)}, \quad (12)$$

where $p(x|\omega_i)$ is the class conditional probability which is the probability density function for x given that it belongs to class ω_i . Here, $P(\omega_i)$ is the prior probability of each class, and it is usually assumed to be distributed uniformly. Here, $p(x)$ is the probability density of x .

The ideal Bayes classifier relies on the full knowledge of the probability density functions for each class, which results in maximal theoretical classification accuracy. However, this is usually not the common case, and therefore it relies on estimates based on the finite training features. Consequently, in order to further improve accuracy, explicit assumptions are made on the class conditional probability densities. A common assumption is that the class probability densities are from a multivariate normal distribution and expressed as

$$p(\mathbf{x}|\omega) = (2\pi)^{-d/2} |\mathbf{S}_\omega|^{-0.5} \exp(-0.5(\mathbf{x} - \bar{\mathbf{x}}_\omega)\mathbf{S}_\omega^{-1}(\mathbf{x} - \bar{\mathbf{x}}_\omega)^T), \quad (13)$$

where S_ω and \bar{x}_ω are the covariance matrices and class mean vectors respectively and calculated using maximum likelihood estimates.

The Bayesian classifier is a simple and robust classifier suited for low dimensional data relative to the data set size. However, it relies heavily on estimates and assumptions on the class conditional probabilities. Thus, good estimates require an ample amount of training data and the validity of the assumptions made are vital for the classifier's performance.

5.1.1. Bayesian Linear Discriminant Analysis

Assuming that the covariance matrices, \mathbf{S}_ω from the previous section are equal, the Bayesian Classifier can be described as a Bayesian Linear classifier. Then \mathbf{S}_ω is replaced with a pooled covariance matrix. From Equation 13, replacing \mathbf{S}_ω with $\mathbf{S}_{\text{pooled}}$, taking the logarithmic scale and ignoring the constants, it becomes

$$g_\omega(x) = -0.5(\mathbf{x} - \bar{\mathbf{x}}_\omega)\mathbf{S}_{\text{pooled}}^{-1}(\mathbf{x} - \bar{\mathbf{x}}_\omega)^T + \ln \mathbf{P}(\omega). \quad (14)$$

Consequently, the decision boundaries that separate the classes are linear.

5.1.2. Bayesian Quadratic Discriminant Analysis

If the assumption of equal covariance matrices, \mathbf{S}_ω is not valid, the Bayesian Classifier can be described as a Bayesian Quadratic classifier. From equation 13, taking the logarithmic scale and ignoring the constants, it becomes

$$g_\omega(x) = -0.5(\mathbf{x} - \bar{\mathbf{x}}_\omega)\mathbf{S}_\omega^{-1}(\mathbf{x} - \bar{\mathbf{x}}_\omega)^T + \ln \mathbf{P}(\omega) . \quad (15)$$

Consequently, the decision boundaries that separate the classes are non-linear (quadratic).

5.2. Leave-one-out cross validation

There are many validation methods to assess the accuracy of classifiers namely holdout, leave-one-out cross validation and, k -fold cross validation.

For the leave-one-out cross validation method, one data set is held from the combined data set and the remaining data are used to train the classifier. For example, if there are n vectors, the remaining $n - 1$ vectors are used as training data. The classifier is then tested with that one vector withheld. The process is repeated n times until each data are tested. The accuracy is then the mean of these n results. This method enables nearly an unbiased validation of the classification accuracy. However, it is computationally expensive.

All the work in this project utilizes the leave-one-out cross validation method to assess the accuracy of the classifier, because the number of data points available is relatively small. Small data sets do not present a significant computational expense and other methods tend to produce biased results especially with small data sets.

6. OSTEOSARCOMA CLASSIFICATION RESULTS

6.1. Adaptive Wavelet Features Classification

DWT entails choosing a specific wavelet family and by using adaptive wavelets, this step can be circumvented. The linear and quadratic discriminant criterion were explored in this work.

The linear discriminant criterion makes the assumption that the class covariance matrices are equal as described in Section 5.1.1. However, upon inspection of the class covariance matrices, the assumption is not valid. Only the first 4 out of 48 data are shown for brevity sake. The results are biased since fewer parameters are estimated by replacing the class covariance matrices with a pooled covariance matrix.¹⁸ Therefore, a quadratic discriminant criterion would be more suitable.

$$S_{\text{HOS}} = \begin{pmatrix} 46.18 & -74.76 & -17.60 & 20.82 \\ -74.76 & 151.07 & 4.75 & -23.73 \\ -17.60 & 4.75 & 48.07 & -22.61 \\ 20.82 & -23.73 & -22.61 & 36.46 \end{pmatrix}$$
$$S_{\text{NHB}} = \begin{pmatrix} 49.18 & -95.04 & -4.94 & 24.14 \\ -95.04 & 194.86 & -0.57 & -49.58 \\ -4.94 & -0.57 & 20.56 & 4.47 \\ 24.14 & -49.58 & 4.47 & 34.00 \end{pmatrix}$$

Table 1 and 2 compare the classification accuracies using Daubechies order 5 features and adaptive wavelet features (before and after optimization). Optimization is using the unconstrained MATLAB optimizer. Both the approximate and detailed bands at level 3 were investigated.

Further analysis is carried out to interpret the wavelet coefficients in the Fourier domain. The coefficients are converted to the Fourier domain to provide insight to the range of frequencies that a particular band encompasses. Figure 6 illustrates the average spectral band of HOS cells and NHB cells with the spectral band of the approximate and detailed coefficients at level 3 respectively. Although optimum discriminatory information is obtained, only 40% of the signal energy are retained.

Table 1. Comparison of accuracies using the quadratic classifier for Daubechies 5 and Adaptive wavelets (before and after optimization) of the detailed coefficients at level 3.

Wavelet	m	q	l	d	Nf	% (Quad classifier)	CVQPM
Daubechies 5	2	9	3	25	20	86.46	0.884
Adaptive (before opt.)	2	9	3	25	20	76.04	0.773
Adaptive (after opt.)	2	9	3	25	20	96.88	0.967

Table 2. Comparison of accuracies using quadratic classifier for Daubechies 5 and Adaptive wavelets (before and after optimization) of the approximate coefficients at level 3.

Wavelet	m	q	l	d	Nf	% (Quad classifier)	CVQPM
Daubechies 5	2	9	3	25	20	77.08	0.781
Adaptive (before opt.)	2	9	3	25	20	73.96	0.764
Adaptive (after opt.)	2	9	3	25	20	92.71	0.928

To inspect the separation of classes visually, each features are assigned a Fisher Score given by¹⁹

$$c_i = \frac{(\mu_i^+ - \mu_i^-)^2}{(\sigma_i^+)^2 + (\sigma_i^-)^2}, \quad (16)$$

where μ_i^+ and σ_i^+ are the mean and the standard deviation of feature i for all training vectors in the positive class and μ_i^- and σ_i^- are the mean and the standard deviation of feature i for all training vectors in the negative class. This score provides a quantitative evaluation of a feature with a higher score implying a higher relevancy.

A scatter plot of the two features with the highest Fisher score is shown in Figure 7. From 25 coefficients, the two selected features have a score of 1.23 and 1.61 respectively. From the plot, it can be seen that the responses are well separated with little overlap and is more distinct in higher dimensions but that is beyond our visualization ability.

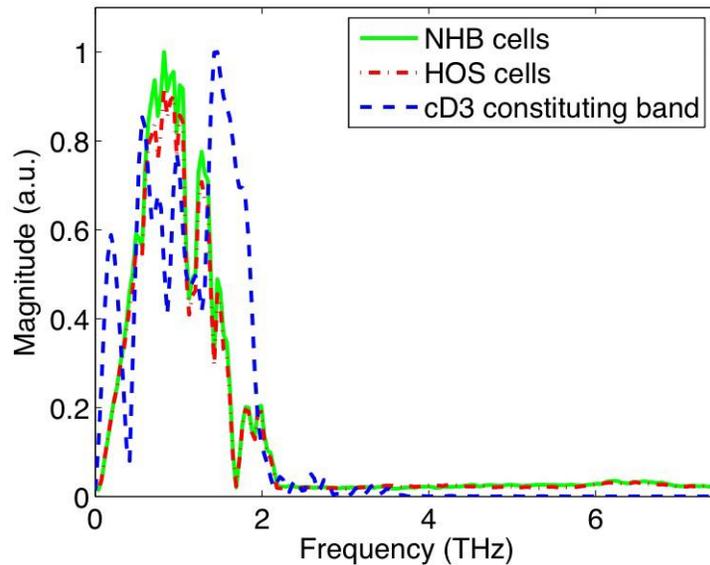


Figure 6. Average spectras of the HOS and NHB cells' responses in comparison to the frequency band of cD3.

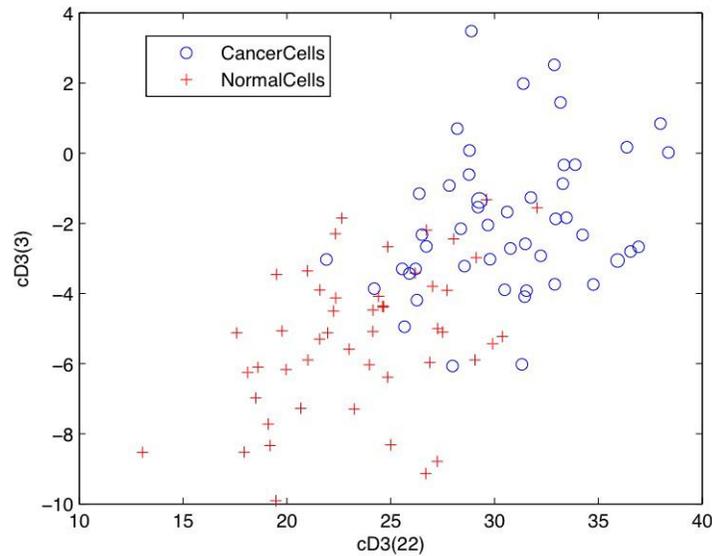


Figure 7. Scatter plot of features.

6.2. Rational Wavelet Features Classification

Three classifiers were used, namely the Mahalanobis Distance Classifier, and Support Vector Machines with a linear kernel and one with a Gaussian radial basis function (RBF) kernel. SVM using Gaussian radial basis function²⁰ is a type of SVM which uses a kernel function given as $K(x, y) = e^{-\|x-y\|/2\sigma^2}$. A kernel function is used to map the original data into a higher-dimensional non-linear space. This is necessary to classify non-linear data using linear classification methods.

Table 3 shows that most of the discriminant information is within the cD5 band. A frequency plot of the cD3 band is shown in Figure 9. Rational wavelets provide a finer resolution compared to dyadic wavelets and therefore capture relevant discriminant information while removing redundant information.

Table 3. Classification accuracy obtained using rational wavelets with a 3/2 scaling.

Level	Mahalanobis(5 top Fisher features)	SVM(Linear)	SVM(RBF)
1	56.25%	50%	50%
2	58.33%	50%	50%
3	66.67%	57.71%	59.27%
4	72.92%	64.90%	73.54%
5	89.58%	88.54%	91.15%
6	70.83%	55.00%	50%

7. CONCLUSION AND FUTURE WORK

Adaptive and Rational wavelets are applied to reduce the dimensionality of classification features to a point where optimum discrimination information is retained while removing redundant or adverse information, which impacts negatively the classification accuracy.

Adaptive wavelets gradually adapts to the signal relative to a discriminant criterion. The linear and quadratic discriminant criterion were investigated and a classification accuracy of 96.88% for both cases are achieved. The quadratic discriminant criterion provides a more unbiased result. The linear discriminant criterion case assumes

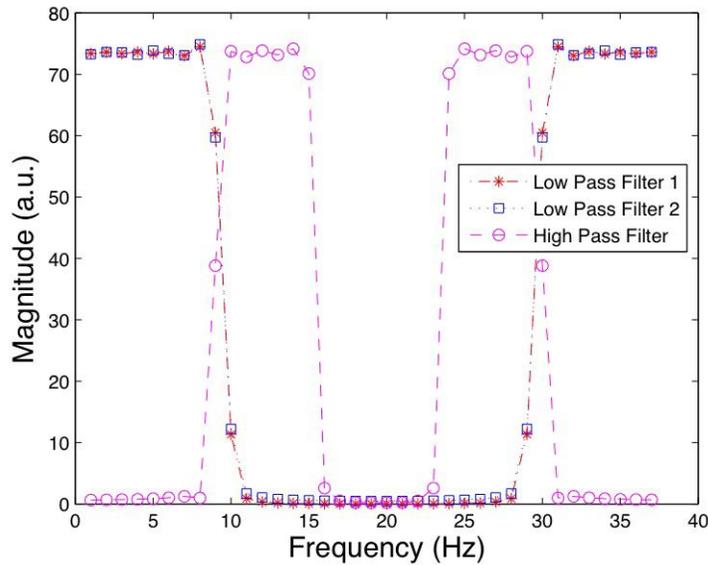


Figure 8. Rational wavelet filters in the frequency domain where $M = p/q = 3/2$.

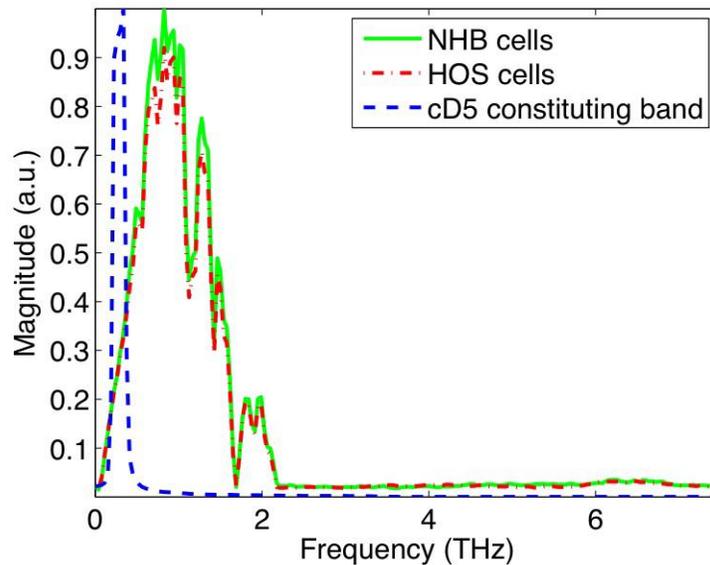


Figure 9. Frequency band of the detailed coefficients of the rational Meyer wavelet at level 5.

equal covariance matrices for both the signals, which are invalid and therefore lead to biased results. In addition, adaptive wavelets compute the best basis iteratively instead of choosing a specific wavelet base manually. The classification accuracy of adaptive wavelets features is remarkably higher compared to Daubechies wavelets features (86.5%).

From the results observed, we can conclude that rational wavelets do offer good classification accuracy. However, wavelet packet analysis should be able to perform as well as rational wavelets, the only difference is that wavelet packet analysis do not provide embedded subspaces.

Future work into adaptive wavelets as a feature extraction tool involves comparisons of other discriminant criteria, such as regularized discriminant analysis, penalized discriminant analysis and flexible discriminant analysis. The features extracted using adaptive wavelets can be fed into more robust classifiers such as Support Vector Machines.

ACKNOWLEDGMENTS

Useful discussions with Langford White, Limin Yu, and Brian Ng from University of Adelaide, Australia are gratefully acknowledged.

REFERENCES

1. J. C. Wittig, J. Bickels, D. Priebat, J. Jelinek, K. Kellar-Graney, B. Shmookler, and M. M. Malawer, "Osteosarcoma: a multidisciplinary approach to diagnosis and treatment," *Am. Fam. Physician* **65**(6), pp. 1123–1132, 2002.
2. W. Withayachumnankul, G. M. Png, X. X. Yin, S. Atakaramians, I. Jones, H. Lin, B. S. Y. Ung, J. Balakrishnan, B. W.-H. Ng, B. Ferguson, S. P. Mickan, B. Fischer, and D. Abbott, "T-ray sensing and imaging," *Proceedings of the IEEE* **95**(8), pp. 1592–1603, 2007.
3. J. W. Handley, A. J. Fitzgerald, E. Berry, and R. D. Boyle, "Wavelet compression in medical terahertz pulsed imaging," *Phys. Med. Biol.* **47**, pp. 3885–3892, 2002.
4. E. Pickwell, B. Cole, A. Fitzgerald, V. Wallace, and M. Pepper, "Simulation of terahertz pulse propagation in biological systems," *Applied Physics Letters* **84**(12), pp. 2190–2192, 2004.
5. R. M. Woodward, V. P. Wallace, R. J. Pye, B. E. Cole, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging of ex vivo basal cell carcinoma," *Journal of Investigative Dermatology* **120**, pp. 72–78, 2003.
6. B. Ferguson, H. Liu, S. Hay, D. Finlay, X.-C. Zhang, and D. Abbott, "In vitro osteosarcoma biosensing using thz time domain spectroscopy," *Proc. of SPIE BioMEMS and Nanotechnology* **5275**, pp. 316–329, 2004.
7. W. Withayachumnankul, B. Ferguson, T. Rainsford, D. Findlay, S. P. Mickan, and D. Abbott, "T-ray relevant frequencies for osteosarcoma classification," *Proc. of SPIE Device and Process Technologies for Microelectronics, MEMS, and Photonics IV* **6038**, pp. 321–335, 2005.
8. X. X. Yin, K. M. Kong, J. W. Lim, B. W. Ng, B. Ferguson, S. P. Mickan, and D. Abbott, "Enhanced t-ray signal classification using wavelet preprocessing," *Medical & Biological Engineering & Computing* **45**(6), pp. 611–616, 2007.
9. B. Ferguson, *Three Dimensional T-Ray Inspection Systems*. PhD thesis, PhD Thesis, University of Adelaide, Australia., 2004.
10. Y. Mallet, D. Coomans, J. Kautsky, and O. D. Vel, "Classification using adaptive wavelets for feature extraction," *IEEE Trans. Pattern Analysis and Machine Intelligence* **19**(10), pp. 1058–1066, 1997.
11. T. A. Ridsdill-Smith and M. C. Dentith, "The wavelet transform in aeromagnetic processing," *Geophysics* **64**(4), pp. 1003–1013, 1999.
12. S. G. Mallat, "A theory of multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**(7), pp. 674–693, 1989.
13. A. Baussard, F. Nicolier, and F. Truchetet, "Rational multiresolution analysis and fast wavelet transform: application to wavelet shrinkage denoising," *Signal Processing* **84**(10), pp. 1735–1747, 2004.
14. I. Daubechies, *Ten Lectures on Wavelets*, 1992.
15. P. Auscher, *Ondelettes Fractales et Applications*. PhD thesis, Universite IX Paris Dauphine, 1989.
16. L. Yu and L. White, "Complex rational orthogonal wavelet and its application in communications.," *Signal Processing Letters* **13**(8), pp. 477–480, 2006.
17. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, John Wiley and Sons, Inc., New York, 2nd ed., 2001.
18. Y. Mallet, D. Coomans, and O. D. Vel, "Recent developments in discriminant analysis on high dimensional data," *Chemometrics and Intelligent Laboratory Systems* **35**, pp. 157–173, 1996.

19. P. Pavlidis, J. Weston, J. Cai, and W. N. Grundy, "Gene functional classification from heterogeneous data," in *Proceedings of the Fifth Annual International Conference on Computational Biology*, pp. 249–255, 2001.
20. J. Schramm, *Pattern classification: a unified view of statistical and neural approaches*, John Wiley & Sons, Inc, New York, NY, USA., 1996.