In vitro osteosarcoma biosensing using THz time domain spectroscopy

Bradley Ferguson[a,b,c,d], Haibo Liu[a], Shelley Hay[e], David Findlay[e], X.-C. Zhang[a] and Derek Abbott[b,c]

[a] Center for Subsurface Sensing and Imaging Systems, Rensselaer Polytechnic Institute, Troy, NY 12180, USA

[b] Centre for Biomedical Engineering, The University of Adelaide, SA 5005, Australia

[c] Department of Electrical & Electronic Engineering, The University of Adelaide, SA 5005, Australia

[d] CRC for Sensor, Signal and Information Processing, Technology Park, Mawson Lakes Boulevard, Mawson Lakes, SA 5095, Australia

[e] Department of Orthopaedics and Trauma, The University of Adelaide and Hanson Institute, SA 5005, Adelaide.

ABSTRACT

Terahertz time domain spectroscopy (THz-TDS) has a wide range of applications from semiconductor diagnostics to biosensing. Recent attention has focused on bio-applications and several groups have noted the ability of THz-TDS to differentiate basal cell carcinoma tissue from healthy dermal tissue *ex vivo*.[1]

The contrast mechanism is unclear but has been attributed to increased interstitial water in cancerous tissue. In this work we investigate the THz response of human osteosarcoma cells and normal human bone cells grown in culture to isolate the cells' responses from other effects. A classification algorithms based on a frequency selection by genetic algorithm is used to attempt to differentiate between the cell types based on the THz spectra. Encouraging preliminary results have been obtained.

Keywords: Terahertz imaging, genetic algorithm, feature selection, osteosarcoma

1. INTRODUCTION

T-ray imaging for medical applications is limited to the surface of the human body, for example for scrotal, corneal and dermatological/cutaneous imaging, but could find important use in estimating the depth of burns[2] and early warning signs of skin cancer. The level of image differentiation at these shallow depths is more precise than with the use of X-rays. Unlike X-rays, T-rays have the advantage of being a non-ionising radiation and thus represent a totally non-invasive diagnostic technique.[3,4] Due to these significant advantages, cutaneous imaging for biomedical diagnostics is a significant application focus.

Terahertz systems have been proposed for a number of biomedical applications including the detection of tooth cavities[5] and burn severity diagnosis.[2] One of the most potentially significant biomedical applications is in the detection of skin cancers such as malignant melanoma and basal cell carcinoma. The incidence of these cancers continue to escalate and in the advanced stages of melanoma there is no curative therapy available,[6] early detection is therefore of prime importance.

Currently most dermatologists rely on a visual assessment of the patient for diagnosis. This diagnosis is not straight-forward and results in a large number of cases being treated inappropriately or inadequately. Hence there

Further author information: (XCZ) E-mail: zhangxc@rpi.edu, Telephone: 1 518 276 3079,
(DA) E-mail: dabbott@eleceng.adelaide.edu.au,

BioMEMS and Nanotechnology, edited by Dan V. Nicolau, Uwe R. Muller, John M. Dell, Proceedings of SPIE Vol. 5275 (SPIE, Bellingham, WA, 2004)
0277-786X/04/$15 · doi: 10.1117/12.527880

is considerable interest in the development of a non-invasive technique to improve clinician's diagnostic accuracy. Recently researchers have investigated the use of white light reflectance spectroscopy for this purpose.[7] They have found significant differences in the spectra for malignant and benign lesions which would allow automated diagnosis. These studies have focussed on calculating metrics such as the tissue hemoglobin concentration, water concentration and hemoglobin oxygen saturation.[8,9] A similar system based on the principle of terahertz time-domain spectroscopy has the potential to perform the same task with the added advantages of reduced Rayleigh scattering and finer frequency resolution.

1.1. Previous Work

The most extensive study of the application of THz-TDS to cancer detection has been conducted by researchers are the University of Cambridge and TeraView Limited.[1,10,11] They have focused on Basal Cell Carcinoma (BCC), a form of skin cancer, which seldom metastasises and is therefore not life-threatening, but can be disfiguring and is often misdiagnosed. Woodward *et. al.* presents a clinical trial on excised (*ex vivo*) diseased and healthy tissue from 21 patients diagnosed with BCC.[1] The excised tissue was imaged using a reflective mode scanning THz imaging system. A 20 Hz scanning delay line enabled an acquisition rate of 20 pixels/second and their quoted SNR was 6000:1. The THz radiation was focused on the tissue samples and the reflected radiation collected using parabolic mirrors and detected using electro-optic sampling. To differentiate between cancerous and non-cancerous tissue a technique termed *time post pulse* (TPP) was used.[11] In this method the THz signal is deconvolved and converted back to the time domain using the inverse Fourier transform. TPP normalises the response by dividing each sample by the peak in the time domain. TPP essentially quantifies the absorption and broadening of the THz pulse after reflection from the tissue sample. 20 of the 21 samples considered were accurately diagnosed using the THz imaging results. They show that cancerous tissue results in increased absorption and broadening of the THz response. This change is attributed to 'an increase in the interstitial water within the diseased tissue, or a change in the vibrational modes of water molecules with other functional groups.'[1]

Other groups have investigated THz imaging of cancerous tissue.[12] It has been shown that *dark* imaging, where only the portion of the illuminating THz radiation that is scattered beyond the ray path is detected, has promise in this context.[13]

These previous studies have focused on excised tissue. In these cases a large number of factors influence the measured THz response, and changes in the concentration of water molecules and its bonds are likely to dominate the response. In this case study we sought to investigate the problem at a cellular level and isolate the cells' responses from that of bulk tissue properties.

1.2. Paper Overview

Accordingly THz spectroscopy was performed on human cells grown in culture. The cells were grown in transparent plastic flasks which enabled spectroscopic measurement of the live cells. The cells considered were normal human bone (NHB) cells and human osteosarcoma (HOS) cells. A classification framework was then developed to allow automatic differentiation between the two cell types. Section 2 describes the experimental procedure adopted and the THz imaging system. Statistical analysis was applied to analyse the THz spectral results and the classification framework is described in Sec. 3. These techniques were able to differentiate between the THz responses of cancerous and normal human bone cells with high accuracy. The classification results are presented in Sec. 4. Finally Sec. 5 summarises this paper and discusses future directions.

2. EXPERIMENTAL PROCEDURE

2.1. Sample Preparation

The NHB cells were grown from patients undergoing a hip replacement. The cells were cultured from small pieces of trabecular bone for 4-6 weeks to obtain a confluent culture. The HOS cells were cultured from an immortalised cell line and a confluent culture was obtained within 1 week. The cells were cultured in 5 mL of Dulbecco's Modified Eagle Media (DMEM) supplemented with L-Glutamine (0.29 g/l) , 10% foetal bovine serum and Gentamicin (16 μg/ml) as an antibiotic. The same media was used for both types of cells to remove the

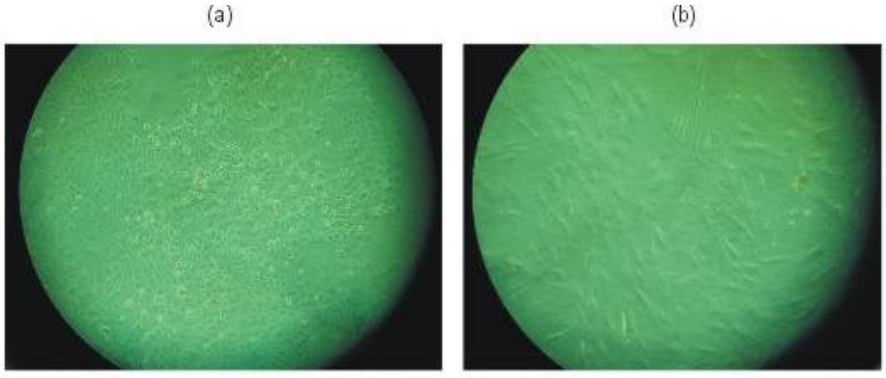

Figure 1. Normal and cancerous cells viewed under a microscope. (a) Human osteosarcoma cells. (b) Normal human bone cells. This image was taken several days before the spectroscopy results were obtained and the cells are not yet confluent.

possibility of different media influencing the results. Figure 1 shows a microscope image of each of the different cell types.

The cells were cultured in 25 ml polyethylene flasks, in a 5% carbon dioxide environment at a temperature of 37°C. The cells form a thin layer attached to the bottom of the flask, and once they are confluent they cover the entire bottom surface with a dense layer of cells. The flasks are transparent to THz radiation. To perform spectroscopy the flasks were tipped on their side to allow the media to run off and the flasks were placed in the THz beam path as illustrated in Fig. 2. Three 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 10 different positions, separated by 50 μm each. This was performed using a standard scanning THz imaging system that has been described previously.[14] A lock in amplifier time constant of 10 ms was used. The laser was a regeneratively amplified Ti:sapphire laser producing 130 fs pulses with a 1 kHz repetition rate and an average power of 0.7 W. The THz emitter was a 2 mm thick $< 110 >$ oriented ZnTe crystal and the THz beam was detected using electro-optic sampling in a 4 mm thick $< 110 >$ ZnTe crystal. The image took less than 3 minutes to acquire. After this time the flask was removed from the spectroscopy system to ensure that the cells did not die from lack of media.

This process was performed for each of the three flasks and then iterated a further 5 times, thereby providing the THz response at 50 pixels for each of the flasks. This then provided sufficient data to allow statistical classification algorithms to be used to attempt to differentiate the cells.

2.2. Imaging Results

The average time and frequency domain responses of the three flasks are shown in Figs. 3 and 4. The error bars indicate one standard deviation either side of the mean for the 50 samples measured. From the THz spectrum it is clear that the cells result in additional absorption compared to the reference. However, the problem of distinguishing between the two types of cells is non-trivial because the difference between the two signals is less than the standard deviation (ie: the noise level) of the data.

Further understanding of the responses is provided by deconvolving the signals from the average reference. Deconvolution is performed by taking the ratio of the cells' THz response to the empty flask response in the frequency domain. It allows the frequency domain amplitude and phase response of the cells to be determined without the influence of the flask. These deconvolved responses are shown in Figs. 5 and 6.

The cells do not have sharp resonant absorption peaks. This is expected for any solid material, at room temperature, due to the lattice (phonon) modes in the THz region causing spectral broadening. However there is sufficient difference between the two responses to allow classification algorithms to distinguish between them.

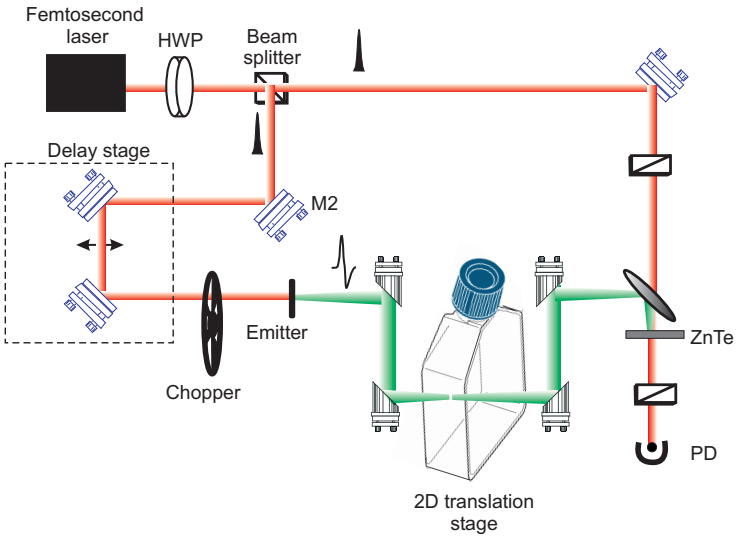

Figure 2. Scanned THz imaging system used to image cell flasks. A scanned THz imaging system uses the traditional THz time domain spectroscopy (THz-TDS) system to obtain the THz response and raster scans the target to measure the response at all pixels. The cells are cultured in a polyethylene flask and a thin layer of cells grows on the bottom surface of the flask. The flask is tipped on its end and inserted into the THz beam such that the focal point of the parabolic mirrors is on the cell layer. The flask is raster scanned using a 2D translation stage.

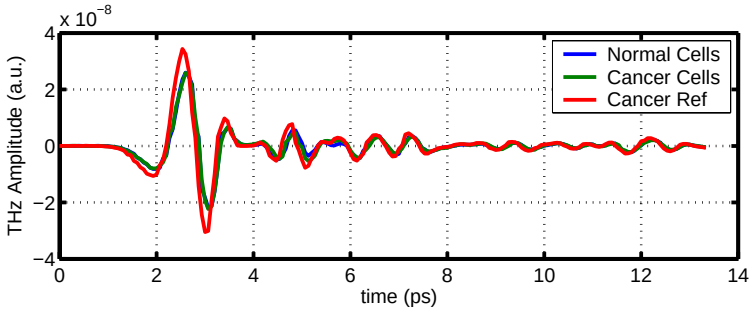

Figure 3. THz pulses after transmission through the cells. Three flasks were considered. One containing cultured normal human bone cells, another containing cultured human osteosarcoma and a third reference containing only the culturing Dulbecco's Modified Eagle Media (DMEM).

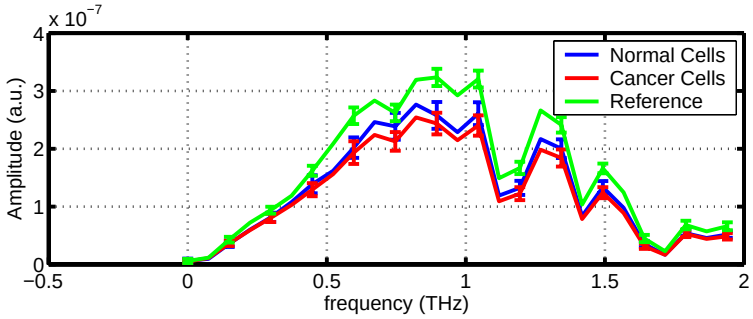

Figure 4. THz amplitude spectra after transmission through the three flasks. This plot shows the amplitude of the Fourier transform of the time domain pulses shown in Fig. 3. The error bars indicate one standard deviation either side of the mean.

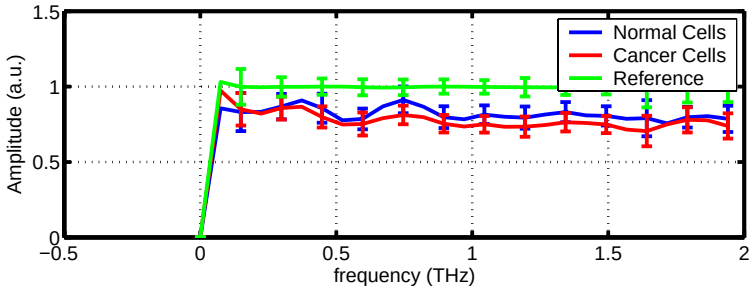

Figure 5. Deconvolved THz amplitude spectra for the three flasks. The cell responses were deconvolved using the average reference flask response. The error bars indicate one standard deviation either side of the mean.

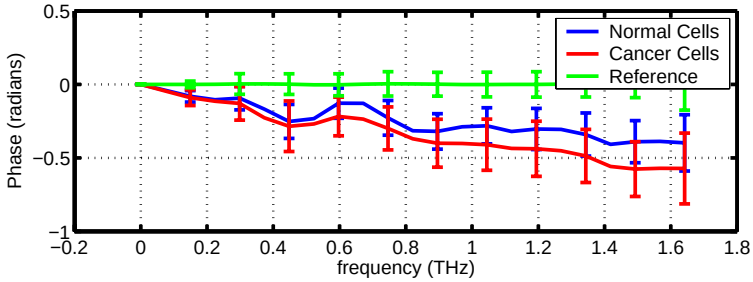

Figure 6. Deconvolved THz phase spectra for the three flasks. The error bars indicate one standard deviation either side of the mean.

3. CELL CLASSIFICATION

3.1. Feature Selection by Genetic Algorithm

This sections describes a feature extraction method based on the deconvolved frequency dependent transmission function of the target. Unlike gas analysis considered in,[15] most solid materials do not exhibit visibly significant sharp absorption resonances. A typical THz-TDS system provides a frequency resolution of 37.5 GHz and a useful bandwidth of approximately 2.5 THz. Therefore there are up to 67 ($\frac{2500}{37.5}$) usable frequency coefficients. In practice these coefficients will contain much redundant and even misleading information and a classifier based on all 67 frequency components will prove computationally inefficient and will have poor generalisation performance as a result of the Curse of Dimensionality.[16, 17]

Given a set of d possible features we desire to select a subset of these such that the classification error for the sample data is minimised. The *only* guaranteed method of selecting the optimal subset is to sequentially consider every possible combination of features.[18] This requires that a prohibitively large set of features be considered and is generally not feasible for realistic pattern recognition problems where very large numbers of potential features may be available.

A number of iterative feature selection methods have been devised to avoid the necessity of an exhaustive search and while they cannot guarantee optimality they have shown to provide good performance in most practical situations.[19] An example is the Sequential Forward Selection (SFS) algorithm.[19] In this method the single best feature is selected, and features are added one at a time, where the added feature in combination with the previously selected features minimises the classification error. This method is computationally attractive but has the disadvantage that features may never be discarded once selected. Floating search methods, which iteratively add and discard features, overcome this disadvantage at the cost of additionally computational complexity.[20, 21]

We adopt an alternative approach to these methods, and use a genetic algorithm to search the feature space and identify optimal feature sets.

3.1.1. Genetic Algorithms

Genetic algorithms are inspired by the process of biological evolution and follow the principles of mutation and survival of the fittest.[22] Genetic algorithms are an efficient method of iteratively searching a large sample space and arriving at near-optimal solutions based on some *fitness* function. This is achieved by defining a mathematical representation of each possible solution as a binary string, which is referred to as a *chromosome*. Several potential solutions are created and defined as the initial *population*. The fitness of each member of this population is calculated and the members are ranked according to their fitness score. Only the fittest members are retained for the next generation. Thus simulating the principle of survival of the fittest. The surviving members are then stochastically altered using several evolutionary rules to derive new potential solutions. The fitness of these solutions are calculated and the process repeated until a certain fitness limit is reached or a specified number of generations are tested.

There are three principle genetic operators, which take members of a previous generation and derive *offspring*, or child members for the subsequent generation. These operators, depicted in Fig. 7, are replication, crossover (or mating) and mutation:

Replication A chromosome from generation k is simply reproduced, unchanged, in generation $k+1$,

Crossover Two parent chromosomes are combined (or *mated*) by choosing a split point randomly along the length of the chromosome. The two chromosomes are split at this point and the sections recombined with the matching section from the other parent. This step is one of the keys to the strengths of genetic algorithms as it allows two strong solutions to be combined. The probability that crossover is performed is denoted p_{co}.

Mutation Each bit of a parent chromosome has a small probability p_{mut} of being flipped to derive a mutated child chromosome.

Other operators and variations on these operators abound, however these three form the core of most genetic algorithms.[23, 24]

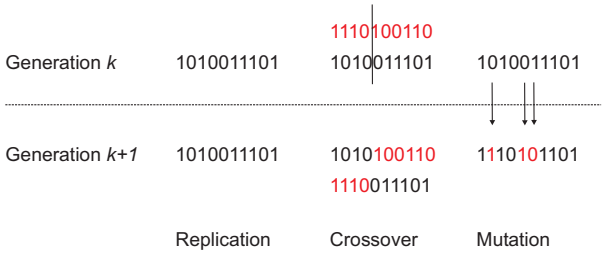

Figure 7. Example of the three main genetic operators. These operators govern the chromosomes present in the child generation $(k+1)$ based on those in the parent generation (k). Replication simply copies the parent chromosome. Crossover occurs with probability p_{co} and it splits two parent chromosomes at a random point and joins the matching segments to form two child chromosomes. Finally the mutation operator flips a single bit of the parent chromosome with probability p_{mut}.

In the context of the feature selection problem the n potential features are mapped onto an n bit chromosome where each bit (or *gene*) represents a specific feature. If the gene $= 1$ then that feature is used, otherwise it is not. For example given a frequency resolution of 37.5 GHz the chromosome $[0001010000100\cdots0]$ indicates that the amplitude and phase coefficients at frequencies of 0.15 THz, 0.225 THz and 0.412 THz are used by the classifier and all others are discarded.

The fitness function is the classification accuracy using the simple classifier discussed in Sec. 3.2 when trained with a representative set of training data and tested on a separate set of test data. Given these definitions the full genetic algorithm can now be described.

Initially a population of chromosomes is randomly selected from the possible combinations. N chromosomes are selected. Each chromosome is then assessed by the fitness function to calculate its fitness f_i. The chromosomes are ranked in order of fitness and the maximum fitness is denoted $f_{\max}$. The genetic operators are then applied to the population. Each chromosome is replicated in the next generation with a probability of $f_i/f_{\max}$. Each chromosome is also input to either the crossover or the mutation operator. The probability that the crossover operator is selected is given by p_{co}. Otherwise the mutation operator is selected and each bit has a probability of p_{mut} of being mutated.

The fitness of the new generation of chromosomes is then calculated and the population re-ranked by fitness. A terminating criteria may be defined in terms of an acceptable fitness value or a maximum number of generations. If the terminating criteria is satisfied the algorithm returns the chromosome with the highest fitness. Otherwise any duplicates in the population are removed and the fittest N elements are retained while the remainder are discarded. The generation is incremented and the genetic operators are again applied to the new generation of chromosomes. This procedure is repeated until the terminating criteria is satisfied. This algorithm is summarised in the flow chart shown in Fig. 8.

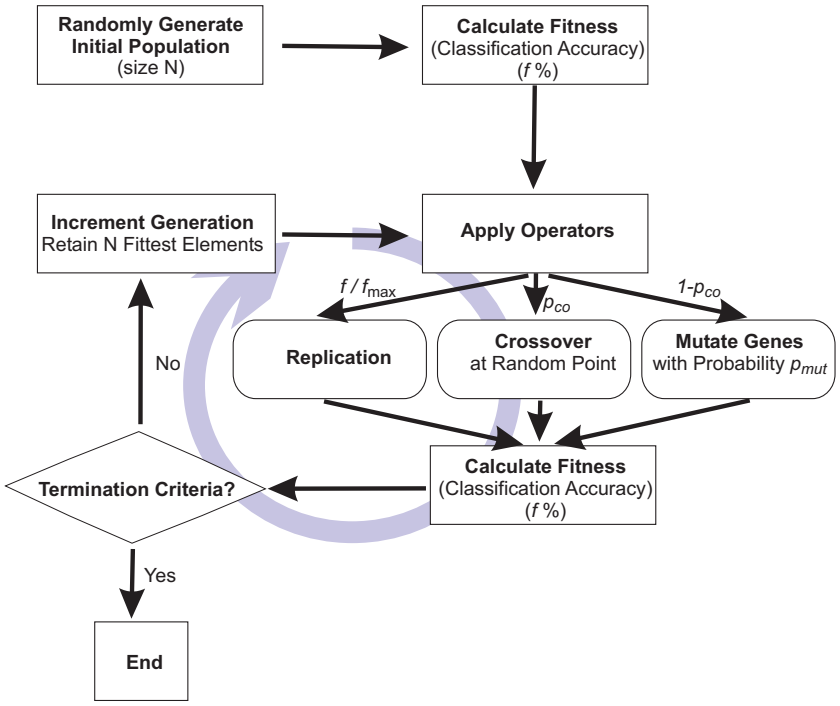

Figure 8. Flow chart of a genetic algorithm for feature selection. The algorithm commences by randomly selecting an initial population of size N. The fitness of each element is assessed and the chromosomes ranked according to fitness f. The genetic operators are applied to the population. Each chromosome is replicated with probability $f/f_{\max}$. The chromosomes also pass through either the crossover or mutation operator. The crossover operation takes two parent chromosomes, splits them at a random position and combines the matching sections to produce two child chromosomes, the mutation operator has a low probability of mutating each bit of the chromosome. The fitness of the resulting chromosomes is calculated using the fitness function and the results compared against the terminating criteria, which may be a minimum fitness or a maximum generation. If the criteria is satisfied the highest ranking chromosome is returned otherwise the least fit members of the population are discarded and the process repeats.

Genetic algorithms have been used for feature selection by several authors.[25, 26] These authors used the interclass separation of the training data as the fitness function with a penalty term for long feature vectors. The method developed above uses the actual classification performance on realistic test data as the fitness criteria. This fitness function measures the ability of the chosen features to allow the classifier to generalise and classify unknown data. It can therefore be expected to result in greater overall classifier performance.

3.2. Mahalanobis Distance Classifier

One of the most common and versatile classifiers is the *Mahalanobis distance* classifier.[27] It is one of a class of minimum distance classifiers. It assumes that the data for each class are normally distributed, thus the samples drawn from each class will form a cluster in k dimensions, with a centre given by the mean vector, μ, and shape dependent on the covariance matrix, Σ. Estimates are formed for these parameters for each class i, using the training vectors,

$$\mu_i = E[\boldsymbol{x}_i], \quad (1)$$

$$\Sigma_i = E[(\boldsymbol{x}_i - \mu_i)(\boldsymbol{x}_i - \mu_i)^T]. \quad (2)$$

The *Mahalanobis distance* calculates the distance of a given point from the mean value for a given class normalised by the variance of the training vectors in that direction. For a given class, i, the distance is defined as,

$$d_i(\boldsymbol{x}) = (\boldsymbol{x} - \mu_i)^T \Sigma_i^{-1} (\boldsymbol{x} - \mu_i)^T. \quad (3)$$

Classification is then performed by selecting the class for which the *Mahalanobis distance* is minimised. This classifier is optimal for normally distributed classes with equal covariance matrices and equal *a priori* probabilities.

This classification scheme was chosen because it is simple to implement and it provides reasonable results for a variety of statistical properties, thereby highlighting the performance of the feature selection technique. More complicated classification algorithms abound and the appropriate choice for this application is an open research area.

4. CLASSIFICATION RESULTS

The Mahalanobis distance was used to classify the responses. The input features to the classifier were the deconvolved amplitude and phase at specific frequencies. It was not possible to manually choose optimal frequencies, so they were chosen using a genetic algorithm that identified near-optimal frequencies. The genetic algorithm described in Sec. 3 was applied to the THz responses for the HOS and NHB cells. The parameters of the genetic algorithm were $p_{\text{mut}} = 0.01$ and $p_{\text{co}} = 0.7$. The termination criteria was after 50 generations, and the initial population consisted of 100 randomly chosen feature vectors.

The genetic algorithm resulted in a set of near-optimum training features consisting of 6 different frequencies: 0.22, 0.37, 1.12, 1.27, 1.34, and 1.57 THz. This provided a classification accuracy of 98.6%. The *confusion matrix* for this classifier is given below where the element, $[i, j]$, shows the relative proportion of samples belonging to class i that were recognised as class j. The classes were NHB cells (1), HOS cells (2), and Reference flask (3):

$$\mathbf{X} = \begin{bmatrix} 1 & 0.042 & 0 \\ 0 & 0.958 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4)$$

To visually illustrate the strength of this method the genetic algorithm was programmed to choose the optimum 2 frequencies (note that this provides a poorer classification than using 6 frequencies as less information is available to the classifier, but it allows the results to be easily visualised). It determined that 1.00 THz and 1.57 THz provided a classification accuracy of 0.889. The scatter plots for both amplitude and phase at these frequencies are shown in Figs. 9 and 10.

The frequencies chosen by genetic algorithm were compared with two randomly chosen frequencies. The frequencies 0.52 THz and 2.00 THz were randomly chosen and the deconvolved amplitude and phase of the cell responses were used to train the Mahalanobis classifier. The scatterplots for both the amplitude and phase of the training data are shown in Figs. 11 and 12. The resulting classifier had a very poor accuracy of 0.670 when tested using 100 test responses, and is obviously inferior to the genetic algorithm result.

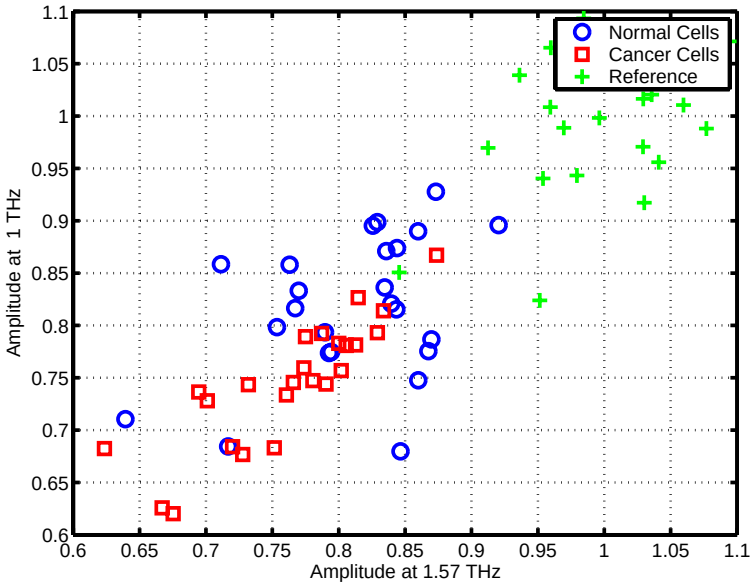

Figure 9. Scatterplot of the deconvolved THz amplitude at the optimum two frequencies. The optimal 2 frequencies were chosen using genetic algorithm to be 1.00 and 1.57 THz. The deconvolved amplitude of the cell responses are shown at these frequencies.

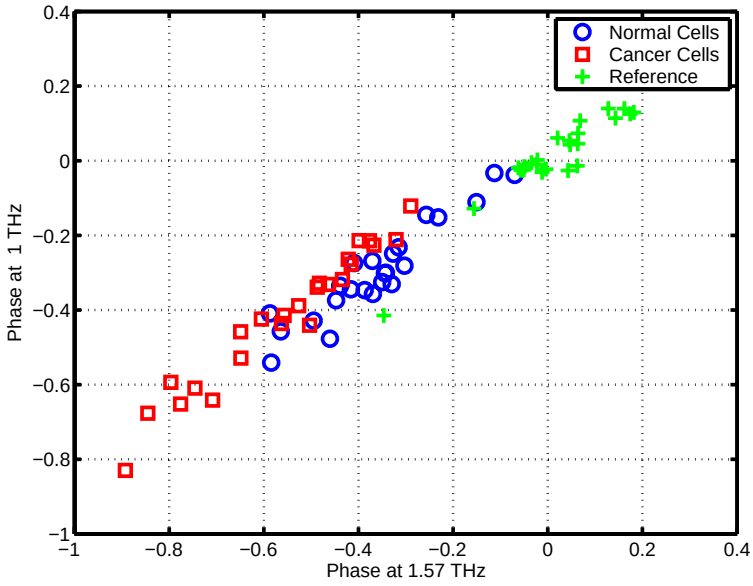

Figure 10. Scatterplot of the deconvolved THz phase at the optimum two frequencies. The deconvolved phase of the human osteosarcoma (HOS) cells, normal human bone (NHB) cells and the reference flask are shown at frequencies of 1.00 and 1.57 THz. A Mahalanobis classifier was trained using this data and the amplitude data in Fig. 9. When tested on 100 random responses the classification accuracy was 0.889.

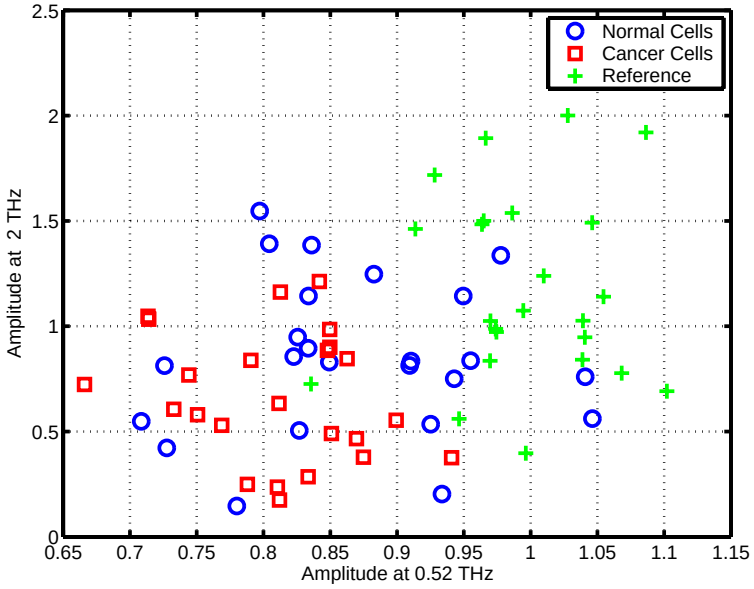

Figure 11. Scatterplot of the deconvolved THz amplitude at two random frequencies. Two random frequencies (0.52 and 2.00 THz) were chosen and the deconvolved amplitude at these frequencies are plotted for the HOS and NHB cells and for the reference flask. There is poor separation between the two types of cells.

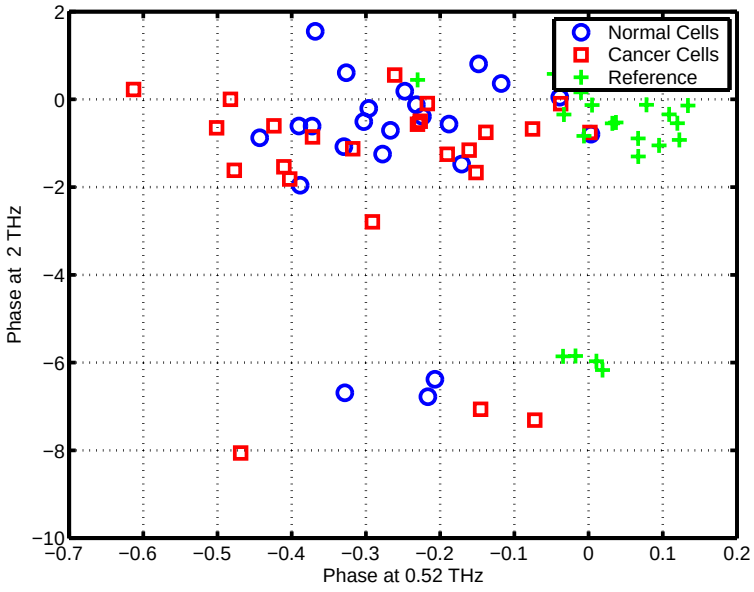

Figure 12. Scatterplot of the deconvolved THz phase at two random frequencies. The deconvolved phase of the HOS and NHB cell responses are plotted at the randomly chosen frequencies of 0.57 and 2.00 THz. As with the amplitude data there is a poor separation between the two classes. A Mahalanobis distance classifier was trained using this data and the amplitude data shown in Fig. 11 and tested on a separate set of responses. The classifier yielded an accuracy of 0.670.

5. CONCLUSIONS AND FUTURE DIRECTIONS

The results of this preliminary case study are promising. They show that THz-TDS can detect the response of a thin layer of cells with a thickness of under 100 μm. They also show that there is sufficient spectral signature information to allow a classifier to be trained to recognise specific types of cells. This would appear to be a significant result however a number of questions remain to be answered. The most significant problem is transitioning the results from the meticulously controlled environment of this case study to a more general setting. This study only considered one flask of each cell type. Further work is required to confirm the results and conclusively show that the detected differences are in fact due to the cellular responses and not other potential experimental variations such as long term laser drift, variations in flask thickness and degree of cell confluency. More exhaustive trials should be able to establish this fact by culturing several flasks containing each of the cell types and obtaining THz data using multiple THz-TDS systems.

It is also important to verify that the observed classification accuracy can be maintained in the presence of standard environmental variations. These include the relative humidity, the specific THz emitter and detector characteristics and the concentration and type of cell media solution. Similar experiments should be conducted with other cell types, particularly skin cancer cells cultured *in-vitro* to investigate the scope of THz-TDS in cellular identification.

A further issue that must be addressed in the future relates to the resolution of THz-TDS. With a focal diameter of greater than 1 mm the THz spectral response is averaged over an extremely large area relative to the size of the cells under investigation. In the IR band microspectroscopic techniques are available, and are capable of acquiring spectral responses with a spatial resolution of 18 μm. This is sufficient to differentiate between the response of a cell nucleus and cytosol.[28–30] Lasch *et. al.*[31] studied skin fibroblasts and sarcoma cells and concluded that previous FTIR results identifying spectral lines corresponding to cancer[32,33] may instead be due to differences in the cells' divisional and metabolic activity rather than signatures specific to cancer. FTIR microspectroscopic studies have also been performed to identify a host of potentially obscuring variables in screening for cervical cancer.[34] It was demonstrated that variables such as leukocytes, endocervical cells, seminal fluids, thrombocytes, bacteria and nylon threads all exhibit spectral responses that can potentially obscure the response of cervical malignancies. Similar problems are likely to hinder *in-vivo* THz cancer diagnosis.

ACKNOWLEDGMENTS

This work was supported by the Australian Research Council, the U.S. National Science Foundation and the U.S. Army Research Office. This work was also supported in part by the Center for Subsurface Sensing and Imaging Systems, under the Engineering Research Centers Program of the National Science Foundation (award number EEC-9986821)

Bradley Ferguson would like to thank the Australian-American Fulbright Commission.

The authors would like to acknowledge Professor's Keese and Giaever for the use of their facilities and the assistance of Guo Chen and Dong Chunzhi.

REFERENCES

1. R. M. Woodward, V. P. Wallace, R. Pye, B. Cole, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging of *ex vivo* Basal Cell Carcinoma," *Journal of Investigative Dermatology* **120**(1), pp. 72–78, 2003.
2. D. M. Mittleman, M. Gupta, R. Neelamani, R. G. Baraniuk, J. V. Rudd, and M. Koch, "Recent advances in terahertz imaging," *Applied Physics B: Lasers and Optics* **68**(6), pp. 1085–1094, 1999.
3. S. W. Smye, J. M. Chamberlain, A. J. Fitzgerald, and E. Berry, "The interaction between terahertz radiation and biological tissue," *Physics in Medicine and Biology* **46**(9), pp. R101–R112, 2001.
4. A. J. Fitzgerald, E. Berry, N. N. Zinovev, G. C. Walker, M. A. Smith, and J. M. Chamberlain, "An introduction to medical imaging with coherent terahertz frequency radiation," *Physics in Medicine and Biology* **47**(7), pp. 67–84, 2002.

5. C. M. Ciesla, D. D. Arnone, A. Corchia, D. Crawley, C. Longbottom, E. H. Linfield, and M. Pepper, "Biomedical applications of terahertz pulse imaging," in *Proc. SPIE - Commercial and Biomedical Applications of Ultrafast Lasers II*, J. Neev and M. K. Reed, eds., **3934**, pp. 73–81, (San Jose, CA, U.S.A.), 2000.
6. P. Altmeyer, K. Hoffmann, and M. Stücker, eds., *Skin Cancer and UV Radiation*, Springer-Verlag, Berlin, 1997.
7. V. P. Wallace, D. C. Crawford, P. S. Mortimer, R. J. Ott, and J. C. Bamber, "Specrophotometric assessment of pigmented skin lesions: methods and feature selection for evaluation of diagnostic performance," *Physics in Medicine and Biology* **45**(3), pp. 735–751, 2000.
8. J. B. Fishkin, O. Coquoz, E. R. Anderson, M. Brenner, and B. J. Tromberg, "Frequency-domain photon migration measurements of normal and malignant tissue optical properties in a human subject," *Applied Optics* **36**(1), pp. 10–20, 1997.
9. D. L. Conover, B. M. Fenton, T. H. Foster, and E. L. Hull, "An evaluation of near infrared spectroscopy and cryospectrophotometry estimates of haemoglobin oxygen saturation in a rodent mammary tumour model," *Physics in Medicine and Biology* **45**(9), pp. 2685–2699, 2000.
10. R. M. Woodward, B. Cole, V. P. Walace, D. D. Arnone, R. Pye, E. H. Linfield, M. Pepper, and A. G. Davies, "Terahertz pulse imaging of in-vitro basal cell carcinoma samples," in *Conference on Lasers and Electro-Optics 2001*, pp. 329–330, SPIE, 2001.
11. R. M. Woodward, B. Cole, V. P. Wallace, R. Pye, D. D. Arnone, E. H. Linfield, and M. Pepper, "Terahertz pulse imaging in reflection geometry of human skin cancer and skin tissue," *Physics in Medicine and Biology* **47**(21), pp. 3853–3863, 2002.
12. P. Knobloch, C. Schildknecht, T. Kleine-Ostmann, M. Koch, S. Hoffmann, M. Hofmann, E. Rehberg, M. Sperling, K. Donhuijsen, G. Hein, and K. Pierz, "Medical THz imaging: an investigation of histopathological samples," *Physics in Medicine and Biology* **47**, pp. 3875–3884, 2002.
13. T. Loffler, T. Bauer, K. J. Siebert, H. G. Roskos, A. Fitzgerald, and S. Czasch, "Terahertz dark-field imaging of biomedical tissue," *Optics Express* **9**(12), 2001.
14. B. Ferguson, S. Wang, D. Gray, D. Abbott, and X.-C. Zhang, "Identification of biological tissue using chirped probe thz imaging," *Microelectronics Journal (Elsevier)* **33**(12), pp. 1043–1051, 2002.
15. D. M. Mittleman, R. Neelamani, R. G. Baraniuk, and M. C. Nuss, "Applications of terahertz imaging," in *Nonlinear Optics '98: Materials, Fundamentals and Applications Topical Meeting*, pp. 294–296, IEEE, (Princeville, HI, U.S.A.), 1998.
16. S. J. Raudys and V. Pikelis, "On dimensionality, sample size, classification error and complexity of classification algorithms in pattern recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **2**, pp. 243–251, 1980.
17. G. V. Trunk, "A problem of dimensionality: A simple example," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **1**(3), pp. 306–307, 1979.
18. T. M. Cover and J. M. Van Campenhout, "On the possible orderings in the measurement selection problem," *IEEE Transactions on Systems, Management and Cybernetics* **7**(9), pp. 657–661, 1977.
19. A. K. Jain and D. Zongker, "Feature selection: Evaluation, application and small sample performance," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **19**(2), pp. 153–158, 1997.
20. P. Pudil, J. Novovicova, and J. Kittler, "Floating search methods in feature selection," *Pattern Recognition Letters* **15**(11), pp. 1119–1125, 1994.
21. P. Somol, P. Pudil, J. Novovicova, and P. Paclik, "Adaptive floating search methods in feature selecton," *Pattern Recognition Letters* **20**(11), pp. 1157–1163, 1999.
22. L. J. Fogel, A. J. Owens, and M. J. Walsh, *Artificial Intelligence through Simulated Evolution*, Wiley, New York, 1966.
23. D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, Addison-Wesley Publishing Company, Reading, MA, U.S.A., 1989.
24. B. P. Buckles and F. E. Petry, *Genetic Algorithms*, The IEEE Computer Society Press, Los Alamos, CA, U.S.A., 1992.

25. P. Corcoran, J. Anglesea, and M. Elshaw, "The application of genetic algorithms to sensor parameter selection for multisensor array configuration," *Sensors and Actuators A - Physical* **76**, pp. 57–66, 1999.
26. P. Biolot, E. L. Hines, M. A. Gongora, and R. S. Folland, "Electronic noses inter-comparison, data fusion and sensor selection in discrimination of standard fruit solutions," *Sensors and Actuators B (Chemical)* **88**, pp. 80–88, 2003.
27. J. Schürmann, *Pattern Classification - A Unified View of Statistical and Neural Approaches*, John Wiley & Sons, Inc., New York, 1996.
28. M. Diem, L. Chiriboga, and H. Yee, "Infrared spectroscopy of human cells and tissue. VIII. Strategies for analysis of infrared tissue mapping data and applications to liver tissue," *Biopolymers (Biospectroscopy)* **57**, pp. 282–290, 2000.
29. M. Diem, L. Chiriboga, P. Lasch, and A. Pacifico, "IR spectra and IR spectral maps of individual normal and cancerous cells," *Biopolymers (Biospectroscopy)* **67**, pp. 349–353, 2002.
30. E. Gazi, J. Dwyer, P. Gardner, A. Ghanbari-Siahkali, A. P. Wade, J. Miyan, N. P. Lockyer, J. C. Vickerman, N. W. Clarke, J. H. Shanks, L. J. Scott, and M. Brown, "Applications of Fourier transform infrared microspectroscopy in studies of benign prostate and prostate cancer. A pilot study," *Journal of Pathology* **201**, pp. 99–108, 2003.
31. P. Lasch, A. Pacifico, and M. Diem, "Spatially resolved IR microspectroscopy of single cells," *Biopolymers (Biospectroscopy)* **67**, pp. 335–338, 2002.
32. P. G. L. Andrus and R. D. Strickland, "Cancer grading by Fourier Transform Infrared Spectroscopy," *Biospectroscopy* **4**, pp. 37–46, 1998.
33. L. M. McIntosh, M. Jackson, H. H. Mantsch, M. F. Stranc, D. Pilavdzic, and A. N. Crowson, "Infrared spectra of Basal Cell Carcinomas are distinct from non tumor-bearing skin components," *Journal of Investigative Dermatology* **112**, pp. 951–956, 1999.
34. B. R. Wood, M. A. Quinn, B. Tait, M. Ashdown, T. Hislop, M. Romeo, and D. McNaughton, "FTIR microspectroscopic study of cell types and potential confounding variables in screening for cervical malignancies," *Biospectroscopy* **4**(75–91).