Optimizing multi-dimensional terahertz imaging analysis for colon cancer diagnosis

Leila H. Eadie [a,b,*], Caroline B. Reid [c], Anthony J. Fitzgerald [d], Vincent P. Wallace [d]

[a] Centre for Computational Intelligence, De Montfort University, The Gateway, Leicester, UK
[b] Centre for Rural Health, Aberdeen University, Centre for Health Science, Old Perth Road, Inverness, UK
[c] Department of Medical Physics and Bioengineering, University College London, London WC1E 6BT, UK
[d] School of Physics, University of Western Australia, Crawley, 6009, Australia

ARTICLE INFO

ABSTRACT

Terahertz reflection imaging (at frequencies ∼0.1–10 THz/10^{12} Hz) is non-ionizing and has potential as a medical imaging technique; however, there is currently no consensus on the optimum imaging parameters to use and the procedure for data analysis. This may be holding back the progress of the technique. This article describes the use of various intelligent analysis methods to choose relevant imaging parameters and optimize the processing of terahertz data in the diagnosis of *ex vivo* colon cancer samples. Decision trees were used to find important parameters, and neural networks and support vector machines were used to classify the terahertz data as indicating normal or abnormal samples. This work reanalyzes the data described in Reid et al. (2011) (*Physics in Medicine and Biology,* 56, 4333–4353), and improves on their reported diagnostic accuracy, finding sensitivities of 90–100% and specificities of 86–90%. This optimization of the analysis of terahertz data allows certain recommendations to be suggested concerning terahertz reflection imaging of colon cancer samples.

1. Introduction

Terahertz (THz) imaging (imaging at frequencies around 10^{12} Hz) is a novel technique for medical imaging. It uses non-ionizing radiation and can safely be used for imaging different types of tissue, such as normal cells and tumors; the contrast between tissue types is thought to occur due to differences in water absorption, protein density or cellular structure. However, THz imaging is still relatively new in the field of medical diagnosis. The optimum methods for using the various data aspects recorded by THz systems are still being determined. This research examines the use of 'intelligent' methods of data analysis to optimize THz imaging for cancer diagnosis, and determine which parameters and analysis method are most useful.

1.1. THz imaging

THz imaging uses electromagnetic waves with frequencies between 0.1 and 10 THz, and associated wavelengths between 0.3 and 3 mm. Within the electromagnetic spectrum, THz waves are situated between infrared and microwaves. They are low energy, non-ionizing and are therefore not harmful to living tissue (Fitzgerald et al., 2002). Penetration of tissue depends on the fat and water content and can reach a depth ranging from several hundred microns to several centimeters (Fitzgerald et al., 2005). They are, however, absorbed by polar molecules such as those that make up water and can be reflected by metal.

In reflectance imaging, the THz waves are reflected off a sample and detected. Waves reflect from the sample surface but also from other interfaces present in the sample within the penetration depth of the radiation (e.g., tissue type boundaries). This allows imaging beneath the surface of optically opaque samples. Depth information can be determined from the timing and amplitude of the detected reflected waves (Sun et al., 2011). Structural images can be created from various parameters associated with the measured THz waves, both in the frequency and time domains, at each pixel in the sample area measured. These parameters change as a result of differing wave delay, broadening and attenuation, as it is absorbed, reflected, scattered, and phase-shifted by the material imaged (Reid et al., 2011).

Reference pulses are used in THz imaging to correct the measured waves: air is a standard reference when imaging biological tissue and a water reference is also used because the optical properties of water are similar to that of tissue, which typically has a high water content (Reid et al., 2011).

The main areas of THz imaging development currently are chemical spectroscopy for the pharmaceutical industry, security measures for airports, and medical imaging. When choosing potential medical imaging areas, the properties of THz waves such as the limited penetration depth and tissue water content must be con-

* Corresponding author. Present address: Centre for Rural Health, Aberdeen University, Centre for Health Science, Old Perth Road, Inverness, UK.
  E-mail address: l.eadie@abdn.ac.uk (L.H. Eadie).

sidered. As such, surface imaging of tissue with relatively low water content was initially attempted, such as investigating skin, bone and tooth imaging (e.g., Ciesla et al., 2000; Wallace et al., 2004, 2006). Similarly, tissue hydration studies were proposed as an initial medical application (Arnone et al., 1999; Mittleman et al., 1999). It has also been tested in skin, lung, pancreas and breast cancer applications (e.g., Woodward et al., 2003; Joseph, Yaroslavsky, Neel, Goyette, & Giles, 2011; Brun et al., 2010; Fitzgerald et al., 2006), and suggested for a number of other diagnostic purposes, such as ophthalmic imaging (Bennett et al., 2011), assessing burn severity (Arbab et al., 2011), liver cirrhosis (Sy et al., 2010) and wound healing (Mittleman et al., 1999).

The THz wave parameters most commonly used in previous medical imaging studies are refractive index, reflectivity and absorption coefficient, although there are many more parameters that can potentially be calculated. There is currently no consensus about the most useful THz parameters for medical imaging and further exploration of this topic is needed.

1.2. Colon cancer

The UK figures from 2009 reveal that approximately 41,000 cases of colorectal cancer are diagnosed each year, making it the second most common cancer in women (after breast cancer) and the third most common cancer in men (after prostate and lung cancer) (Cancer Research UK website, 2012). It is estimated that around 1.24 million new cases were diagnosed and 647,000 people died from colorectal cancer worldwide in 2008 (WHO Global Health Observatory data, 2008), despite the disease being generally curable if found and removed at an early stage. Following surgery, recovery rates are estimated to range from 40% to 90%, depending on the cancer stage at treatment (NHS website, 2012). The majority of colorectal cancer starts in the lining of the colon, initially appearing as a non-cancerous polyp. These slowly develop and can grow into cancer over a period of years. Colorectal cancer screening aims to find and then remove polyps before they become cancerous. The UK screening programme for people aged 60 and over uses a fecal occult blood test to detect blood in the stool to determine who should attend clinics for further investigations. Colorectal cancer is usually diagnosed by sigmoidoscopy (visualizing the lower third of the colon) or colonoscopy (viewing the entire colon). Both procedures involve using a long flexible camera inserted through the anus to view the colon lining. If abnormal areas are located, tissue samples can be taken, or the whole area can be removed.

THz imaging has been suggested for colon cancer imaging via an endoscope, and work is progressing in creating and testing endoscopic THz systems (Wang & Mittleman, 2004; Ji, Lee, Kim, Son, & Jeon, 2009; Chen, Lu, & You, 2011). Such a system could be more convenient than current operator-dependent visual endoscope systems, as they could employ a computer-aided diagnosis system and would therefore not require highly trained operators. In this way, biopsies could be undertaken by staff who are less skilled and therefore less expensive, and less in demand (Reid, 2009).

1.3. Previous research

At the time of writing, the only previous published work on THz imaging of colon cancer is that of Reid et al. (2011), whose data is reanalyzed in the current research. Using logistic regression to analyze the 17 THz imaging parameters measured, they found that their model produced the best classification results when just nine parameters were used. These are not listed here but full details can be found in their publication. Using these parameters, they reported a sensitivity of 82% and a specificity of 77% for distinguishing between healthy and abnormal tissues (cancerous and dysplastic tissues), and 89% sensitivity and 71% specificity for distinguishing between healthy and dysplastic tissues.

1.4. Research aims

THz imaging of human tissue with the aim of detecting abnormal areas is not yet in widespread mainstream use. This may be due to the fact that there is currently no standardized data analysis methodology that allows a simple clinically useful result to be determined. There are differences in thresholds between normal and abnormal tissue both within and between patients. Different researchers have suggested different data parameters, in both the time and frequency domains to create and interpret THz imaging results, and this lack of consensus does not help promote THz as a useful medical imaging modality. Therefore this research aimed to determine which THz imaging parameters are useful in distinguishing normal from abnormal tissue in the specific problem of colon cancer using different intelligent analysis techniques. It attempted to find a standardized method of data analysis for THz imaging output for this disease. It also aimed to find out whether a generalized method can be proposed, or whether data differs between and within patients so greatly that each patient requires individualized analysis methods.

2. Methods

2.1. Data collection

The data were collected by Reid and colleagues, and the collection procedure is described in full in Reid et al. (2011). A brief summary is given here: 30 patients undergoing colonic resection surgery consented to allow tissue samples to be collected during the procedure. Immediately after excision, the resected colon tissue was assessed by pathologists to histologically identify diseased and normal areas. One or more normal and diseased samples were taken from each patient. The samples were imaged using a stand-alone portable THz system, the TPIimaga1000 (TeraView Ltd., Cambridge, UK) in reflection mode. Full details of the system can be found in Wallace et al. (2004). At each point required for the image, a complete THz waveform was recorded, and the image was raster-scanned in the x–y plane to collect the full grid of data points. Air and water reference images, used to normalize the sample data, were taken with the sample holder containing air or water, and baseline or background measurements were taken before all sample measurements. This allowed the data to be corrected, removing both the background and the reference waveforms to leave only the sample function (Reid et al., 2011).

The 17 THz parameters measured were:
Air referenced parameters (AP):

1. E_{min}: minimum value of THz pulse
2. $-E_{min}/E_{max}$
3. $A(t)/E_{min}$: normalized amplitude at time = 0.33 ps before E_{min}
4. $A(t)/E_{min}$: normalized amplitude at time = 2 ps after E_{min} time
5. Integral of values in THz pulse from 0.33 ps before E_{min} to 2 ps after
6. $P(f)$: power in spectrum at frequency = 0.6 THz
7. Integral of power spectrum over frequency range 0.44–0.60 THz
8. Full width at half maximum
9. $E_{max} - E_{min}$

Water referenced parameters (WP):

1. E_{max}: Maximum value of THz pulse
2. $A(t)$: Amplitude of pulse at time = 0.26 ps before E_{max}
3. $A(t)$: Amplitude of pulse at time = 0.26 ps after E_{max}

4. Integral in THz pulse from 0.26 ps before E_{max} time to 0.26 ps after
5. $P(f)$: Power in spectrum at frequency = 0.6 THz
6. Integral of power spectrum over frequency range 0.44–0.60 THz
7. Log of Integrated power
8. Full width at half maximum

These parameters were calculated for regions of interest defined using histopathology of the tissue samples. Within the defined regions of abnormal and normal tissue, the THz waveforms were averaged to generate a single wave, which was used to calculate the parameters.

2.2. Parameter analysis

In order to discover which parameters and analysis methods are best at discriminating between normal and abnormal tissue a range of 'intelligent analysis' techniques were used. The only previous analysis of this data used regression (Reid et al., 2011). The current research built on this by using wider data mining techniques to discover whether the reported sensitivity and specificity results could be improved upon.

Decision trees (DT) were used to identify the most useful parameters, and neural networks (NN) and support vector machines (SVM) were used to classify the data for diagnosis. The results produced by these machine learning techniques were analyzed to determine which technique produces the most accurate results – using which parameters – and whether anything could be learned about the THz technique to guide future imaging protocols.

2.3. The data

There were 31 tumor samples, 20 dysplastic samples and 42 normal samples (taken from a healthy area in each patient providing an abnormal sample) imaged using the THz system. Parameters were calculated for each of these samples and then normalized to lie within a range of 0 to 1. The data was grouped into two non-exclusive data sets for analysis: normal versus abnormal (with dysplastic and cancerous grouped together) and normal versus dysplasia (potentially useful for a screening application for THz imaging).

Three patients were held out from the data as test cases in the generalization testing (described in Section 2.6). These samples were removed from the data set, and the remainder used for training the analysis algorithms. The available data therefore contained 34 normal, 15 dysplasia and 23 cancer samples (N = 72 samples in total). The data were then divided into training and test sets. The number of samples available for use was not very large, at 34 normal and 38 abnormal, and so cross-validation was used to maximize the use of the available data.

2.4. Reduced input data sets

With 17 different parameters measured for each sample, the possibility of reducing the number of parameters/inputs used was considered in order to reduce model complexity thereby increasing model stability and helping to prevent overfitting. This was done in two ways: using principal component analysis (PCA) to create a range of principal uncorrelated inputs that accounted for the majority of the variability of the data; and using DT to determine the most important variables in classifying the data and which therefore should be used as inputs for further analysis. The PCA method used was the processpca function built in to Matlab (Version R2010b). This processes matrices using PCA so that each row is uncorrelated, and is in the order of the amount they contribute to total variation. Rows whose contribution to total variation is less than a specified value (set to 5% in this analysis) were removed.

DTs were generated using the Matlab function classregtree. This creates a regression or classification tree depending on the input data (in this analysis, it created a classification tree). The resulting tree uses binary splits based on the input variables. DTs had a minimum leaf size of five cases and were pruned to the point representing the minimum error found plus one standard error. An automated 10-fold cross-validation method was used when creating the trees.

2.5. Intelligent data analysis

The full and reduced input data sets were analyzed using NN and SVM. The NN used a Bayesian regulation algorithm (trainbr). This updates the weights and bias according to Levenberg–Marquardt optimization, minimizing combinations of errors and weights. Default values used were 5000 maximum epochs, 0.001 maximum error (performance goal), with no time limit set. The NN topology for the analysis of the full data set involved 17 input neurons (one for each parameter), a hidden layer containing 25 neurons and a single output neuron. Because the NN would also be using reduced input data sets and there would be far fewer inputs, an additional NN topology was tested: with varying numbers of input neurons (depending on the input reduction method used), a hidden layer containing 12 neurons and a single output neuron. The NN analysis used a 6-fold cross-validation method for the normal versus abnormal task and a 4-fold cross-validation method for the normal versus dysplasia task due to there being fewer cases available.

The output from the NN was a single number and this was interpreted as follows: if output <0.5 then output = normal; if output $\geqslant$0.5 then output = abnormal or dysplastic.

The SVM were created using the Matlab functions svmtrain and svmclassify. Various kernels were tested, but the linear kernel provided the best results throughout and was used to produce the results reported here.

2.6. Generalization testing

The optimum analysis methods found were tested to determine whether they could be generalized to new patients. This was done by using the three patients held out as test cases: when an optimum trained system was calculated, the three test cases were analyzed to determine how well these algorithms generalized to new cases.

2.7. Outcome measures

The results were assessed in terms of classification accuracy. This is commonly expressed as sensitivity (the ability to correctly identify positive cases) and specificity (the ability to correctly identify negative cases). These are calculated as follows:

Sensitivity : number of true positive results/(true positives
$$+ \text{ false negatives}).$$

Specificity : number of true negative results/(true negatives
$$+ \text{ false positives}).$$

A percentage measure of accuracy (correctly classified samples) was also calculated.

3. Results

3.1. Decision tree analysis

A DT was created for the normal versus abnormal diagnosis task and used only five of the 17 parameters, listed here in order
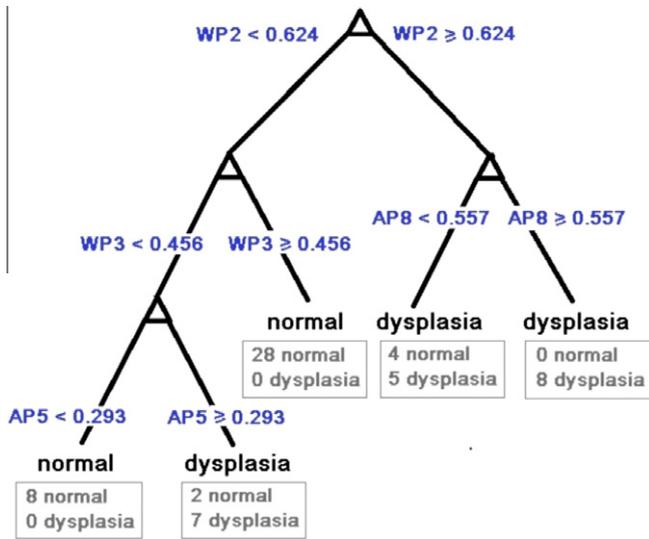
Fig. 1. Decision tree to classify the colon data into two classes: normal and dysplastic. The boxes list the data cases that landed in each leaf.
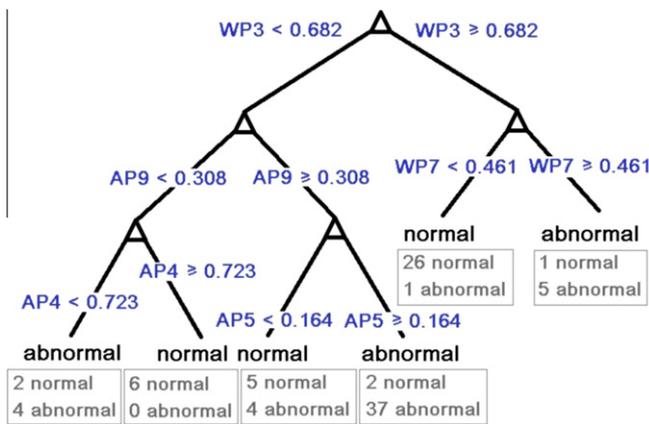

Fig. 2. Decision tree to classify the colon data into two classes: normal and abnormal. The boxes list the data cases that landed in each leaf.

Table 1
Decision tree classification results.

Classification tree	% Correct	Sensitivity	Specificity
Normal ($n = 42$) vs. abnormal ($n = 51$)	89.2%	90.2	88.1
Normal ($n = 42$) vs. dysplastic ($n = 20$)	90.3%	100.0	85.7

of importance (i.e., from the root downwards): WP3, AP9, WP7, AP4, AP5 (see Fig. 1). For the normal versus dysplasia diagnosis task, the DT used only four of the 17 parameters, listed here in order of importance: WP2, WP3, AP8, AP5 (see Fig. 2). It should be noted that AP8 was not used to differentiate between classes, instead providing a split between definitely dysplasia (100% correctly classified) and mostly dysplasia (56% correctly classified). It can therefore be seen that parameters WP3 and AP5 in particular are important in classifying the colon data, as these appear in both trees. WP3 occurs near or at the root of the trees, so is possibly the most important parameter of the 17 for classification.

The results from the DTs are shown in Table 1. It should also be noted that the DT to classify normal versus dysplastic samples produced no false negative results.

The parameters identified by the DT as important (plus an additional parameter that was found to be useful for a 3-way (normal versus dysplasia versus cancer) classification task: WP4) were used in a reduced input data set for the NN and SVM analysis.

3.2. Neural network analysis

NN were used to analyze three different data sets: the full data set (containing all 17 parameters), a PCA reduced input data set, and the reduced input data set suggested by the DT results (8 parameters).

Table 2 displays the results for the analyses determining whether a sample is normal or abnormal. The best overall results were attained when using the PCA-reduced input data set. Both the NN topology with 12 hidden neurons and that with 25 hidden neurons performed well, with the 12 hidden neuron NN producing better specificity and the 25 hidden neuron NN producing slightly better sensitivity. However, these were very similar to the results from the full data set with 12 hidden neurons.

Table 3 displays the results for the analyses determining whether a sample is normal or dysplastic. The best overall results were attained using the full data set, although the DT-reduced input data set also produced good results (with better sensitivity and worse specificity than the full set). It is notable that the PCA-reduced data set did not produce good results for this classification task.

3.3. Support vector machine analysis

Table 4 displays the results for both the analyses determining whether a sample is normal or abnormal, and for normal versus dysplasia. The best overall results for the normal versus abnormal task were attained when using the DT-reduced input data set, and for the normal or dysplasia task the best overall results were attained when using the full data set.

3.4. Generalization testing

The trained NN and SVM algorithms were tested to determine whether they could generalize to new patients. The results are shown in Tables 5 and 6. The results from the DT are also shown for comparison, even though the three test patients were not held out for DT analysis.

It can be seen that patient A is in general diagnosed less well using the trained models than the other two patients tested. For the normal versus abnormal classification task, the best results are produced by the NN using the full data set with 12 hidden neurons. For the normal versus dysplasia classification task, the best results are produced by a linear SVM using DT-reduced data set. The NNs using the full data set perform well, but not perfectly in patient A in particular.

Table 2
Results from NN analyses distinguishing normal from abnormal cases. The best overall paired results are highlighted in bold type.

NN topology	% Correct	Sensitivity	Specificity
Full data set			
12 Hidden neurons	89.02	91.67	86.31
25 Hidden neurons	87.19	87.73	86.61
PCA reduced data set			
12 Hidden neurons	**89.02**	**92.13**	**87.14**
25 Hidden neurons	86.80	93.98	79.20
Decision tree reduced data set			
12 Hidden neurons	82.75	79.33	84.66
25 Hidden neurons	86.93	85.88	88.99

Table 3
Results from NN analyses distinguishing normal from dysplasia cases. The best overall paired results are highlighted in bold type.

NN topology	% correct	Sensitivity	Specificity
Full data set			
12 Hidden neurons	88.92	92.86	87.59
25 Hidden neurons	**90.39**	**92.26**	**90.37**
PCA reduced data set			
12 Hidden neurons	63.63	42.26	68.98
25 Hidden neurons	58.63	32.44	67.31
Decision tree reduced data set			
12 Hidden neurons	**93.33**	**93.75**	**86.01**
25 Hidden neurons	88.53	94.10	82.89

Table 4
Results from SVM analyses. The best overall paired results are highlighted in bold type.

	% Correct	Sensitivity	Specificity
Full data set			
Normal vs. abnormal	77.32	74.21	81.85
Normal vs. dysplasia	**90.39**	**95.83**	**87.17**
PCA reduced data set			
Normal vs. abnormal	77.32	74.21	81.85
Normal vs. dysplasia	77.84	77.98	79.26
Decision tree reduced data set			
Normal vs. abnormal	**86.93**	**90.28**	**84.76**
Normal vs. dysplasia	87.25	93.75	79.02

Fig. 3 displays the parameter values for normal, dysplastic and cancerous samples for one of the held-out patients. It can be seen that the different sample diagnoses are, for almost all parameters, mixed and not easily separable.

4. Discussion

This reanalysis of the data collected by Reid et al. (2011) has shown that the tissue classification results gained by simple regression analysis can be improved upon using 'intelligent analysis' techniques.

4.1. Comparison with previous results

The DTs each used only four or five of the 17 parameters available to classify the data, yet produced impressive results, with 100% sensitivity and 86% specificity for distinguishing dysplastic from normal samples, and 90% sensitivity and 88% specificity in distinguishing abnormal from normal samples. This suggests that the THz imaging process could be streamlined for more efficient data measurement and/or processing.

Further testing using NN and SVM to analyze the data produced classification rates of 92% sensitivity and 90% specificity (using the full dataset with a NN containing a layer of 25 hidden neurons) and 96% sensitivity and 87% specificity (using a linear SVM on the full dataset) for distinguishing dysplastic from normal samples. For determining whether a sample is normal or
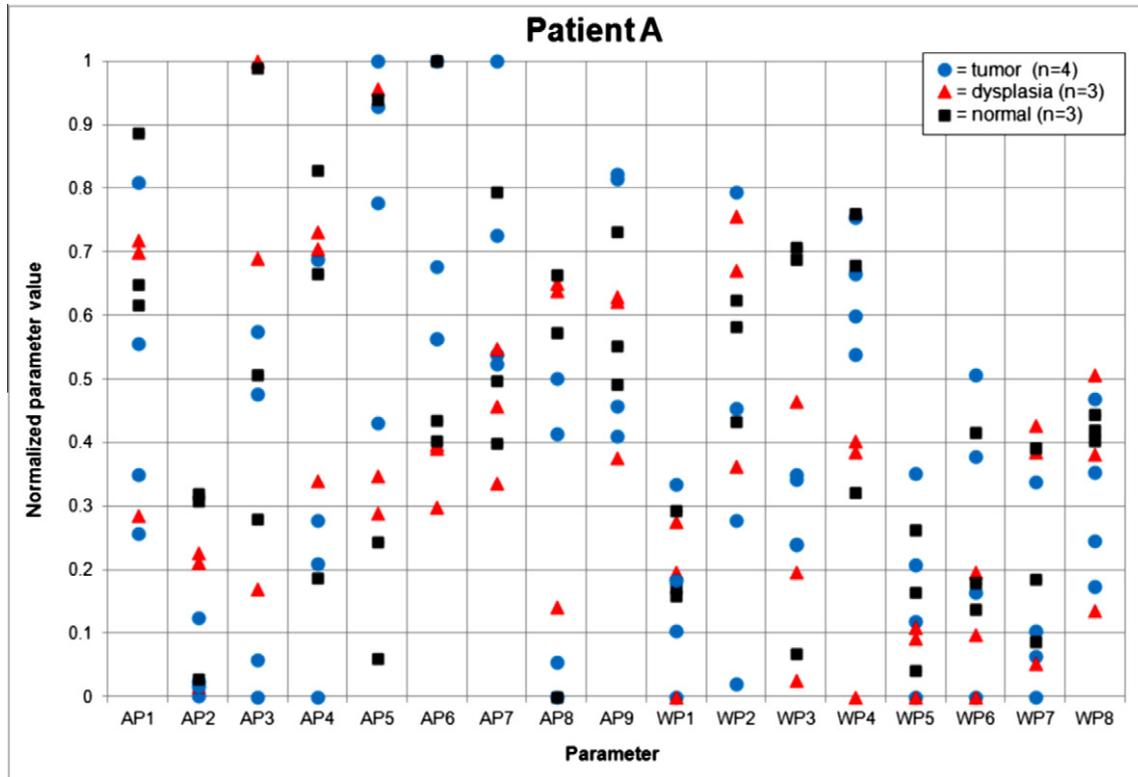
Table 5
Results from generalization analyses distinguishing normal from abnormal cases. The best paired results are highlighted in bold type.

Method	Patient A (n = 10)		Patient B (n = 7)		Patient C (n = 4)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Decision tree	**100**	**100**	50	100	100	50
Full data set						
12 Hidden neurons	**85.71**	**100**	**100**	**100**	**100**	**100**
25 Hidden neurons	100	66.67	**100**	**100**	**100**	**100**
Linear SVM	100	66.67	75	100	50	100
PCA reduced						
12 Hidden neurons	85.71	66.67	**100**	**100**	50	100
25 Hidden neurons	85.71	66.67	**100**	**100**	50	100
Linear SVM	85.71	33.33	50	100	50	100
DT reduced						
12 Hidden neurons	**85.71**	**100**	75	66.67	100	50
25 Hidden neurons	**85.71**	**100**	100	66.67	100	50
Linear SVM	100	66.67	75	100	**100**	**100**

Table 6
Results from generalization analyses distinguishing normal from dysplasia cases. The best results are highlighted in bold type.

Method	Patient A (n = 10)		Patient B (n = 7)		Patient C (n = 4)	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Decision tree	**100**	**100**	100	33.3	**100**	**100**
Full data set						
12 Hidden neurons	**100**	**100**	100	33.3	**100**	**100**
25 Hidden neurons	100	66.67	**100**	**100**	**100**	**100**
Linear SVM	100	66.67	**100**	**100**	**100**	**100**
PCA reduced						
12 Hidden neurons	100	66.67	**100**	**100**	0	100
25 Hidden neurons	**100**	**100**	**100**	**100**	100	50
Linear SVM	66.67	100	**100**	**100**	**100**	**100**
DT reduced						
12 Hidden neurons	100	66. 67	0	66. 7	100	0
25 Hidden neurons	66.67	66.67	0	100	**100**	**100**
Linear SVM	**100**	**100**	**100**	**100**	**100**	**100**

Fig. 3. Graph showing the parameter value spread within a single patient. Each column in the graph shows the values for that parameter for the different samples taken from the patient.

Table 7
Comparison of prior and current results.

	Reid et al. (2011) best results		Current analysis best results	
	Sensitivity	Specificity	Sensitivity	Specificity
Normal vs. abnormal	82%	77%	92%[a]	87%[a]
			90%[b]	88%[b]
Normal vs. dysplastic	89%	71%	100%[b]	86%[b]
			96%[c]	87%[c]
			92%[d]	90%[d]

[a] Using PCA-reduced input data set and NN with 12 hidden neurons.
[b] Using a DT.
[c] Using a linear SVM.
[d] Using the full data set and NN with 25 hidden neurons.

abnormal, results were 92% sensitivity and 87% specificity (using the PCA-reduced dataset with a NN containing a hidden layer of 12 neurons) and 90% sensitivity and 85% specificity (using a linear SVM on the full dataset). All of these results compare well to the results found by Reid et al. (2011). The comparison is summarized in Table 7.

4.2. Useful THz parameters

Of the 17 different THz imaging parameters, DT analysis found 4 air referenced and 4 water referenced parameters to be useful in classifying the data:

- AP4: $A(t)/E_{min}$: normalized amplitude at time = 2 ps after E_{min} time.
- AP5: integral of values in THz pulse from 0.33 ps before E_{min} to 2 ps after.
- AP8: full width at half maximum.
- AP9: $E_{max} - E_{min}$
- WP2: $A(t)$: amplitude of pulse at time = 0.26 ps before E_{max}.

- WP3: $A(t)$: amplitude of pulse at time = 0.26 ps after E_{max}.
- WP4: integral of values in THz pulse from 0.26 ps before E_{max} to 0.26 ps after.
- WP7: log of Integrated power.

These are different to the parameters identified by Reid et al. (2011), which were: AP1, AP3, AP7, AP9, and WP1, WP4, WP5, WP6, WP8 – there are only two parameters in common (AP9 and WP4).

The DTs found WP3, the amplitude of the THz pulse at a specified time-point after the maximum amplitude, to be a particularly important parameter, appearing at or close to the root of both trees.

4.3. Using reduced data sets

This testing investigated the use of reduced input data sets; the 17 THz imaging parameters were not totally independent of each other and it was possible that some could be removed to help reduce the processing power and/or time required without affecting the analysis results. In the normal versus abnormal diagnostic

task, both of the top performing analysis methods did indeed use reduced data inputs: the best NN used the PCA-reduced data set, and the best SVM used the DT-reduced data set. For the normal versus dysplasia diagnostic task, the NN using the full data set performed best, although the DT-reduced input data set also produced good results. The PCA-reduced data set did not produce good results for this classification task. The possibility of using a reduced data set should therefore not be dismissed, but further testing, with greater sample numbers would be needed to confirm whether they can be reliably used.

Two different topologies of NN were tested: one with a hidden layer containing 12 neurons, and one with 25 neurons. This could also have an impact on the processing power and/or time required. In both of the diagnosis tasks, both NN topologies performed well, with 12 hidden neurons producing slightly better sensitivity, but 25 hidden neurons producing marginally better specificity. Therefore, the NN containing fewer neurons could potentially be used in order to improve the algorithm efficiency.

4.4. THz imaging optimization and colon cancer

While colon cancer is a serious problem (currently a relatively common cause of cancer death, despite being readily treatable if diagnosed early enough), it is a slow developing disease. Early diagnosis and removal of potentially cancerous regions – and ideally dysplastic, pre-cancerous regions also – can greatly improve survival rates. Because of this, colon cancer screening is undertaken in the UK. For such screening applications, sensitivity is generally considered to be more important than specificity (i.e., not missing a potential cancer is considered more important than biopsying people who turn out to be healthy). The intelligent analysis methods used in this research all produced high sensitivity (between 90% and 100%), but also good specificity (between 85% and 90%), which would mean that relatively few healthy people would be asked back for further testing following initial THz screening. This would be beneficial for patient morale and confidence in the screening tests. Of particular note in this research, the DT distinguishing between normal and dysplasia samples produced no false negative (missed abnormality) results, which is exactly what a screening programme would be looking for.

4.5. A patient-specific or generalized analysis method?

Reid et al. (2011) stated that their method showed the potential for "a THz-based parametric technique where a normal tissue measurement is taken well away from suspected regions of disease, calibrating the system to the patient. All parametric analysis of suspected diseased regions would then be performed relative to the patient's own normal tissue area." Examining the raw data illustrated the problem: the individual parameter results for the different sample types were not easily separable or interpretable, and lay in different configurations for different patients, even having widely different values for different samples of the same diagnosis from the same patient, as seen in Fig. 1. The current research tested the generalization of the analysis algorithms between and within different patients, to attempt to determine whether such patient-specific calibration was required, or whether a single analysis method could be applied to any patient without needing new calibration each time. When applied to new data from three patients who had not been used in the previous cross-validation testing, the NN analyzing the full datasets with a layer of 12 hidden neurons produced good results for distinguishing between normal and abnormal samples and normal and dysplastic samples. A linear SVM (using the DT-reduced dataset) also produced good results for classifying normal versus dysplastic samples, and the DT for this classification task did not produce any false negative results, a factor, as mentioned

above, that is important for screening applications. These results suggest that patient-specific calibration may not be necessary; instead a more time efficient "one size fits all" analysis might be possible – although in this investigation, one patient was less accurately diagnosed than the others, suggesting individual differences may still potentially cause analysis problems.

4.6. The best analysis method

So which is the best analysis method to use for interpreting THz imaging of colon samples? The results found in this research vary depending on the exact diagnostic question asked. For distinguishing normal from abnormal (dysplastic and tumor), a NN working with a PCA-reduced dataset produces the best results (92% sensitivity, 87% specificity with a NN with 12 hidden neurons or 94% sensitivity, 79% specificity with a NN with 25 hidden neurons). Using the simpler DT method produced surprisingly similar results: 90% sensitivity, 88% specificity. Indeed, for the normal versus dysplasia classification, the DT produces the best sensitivity (100%, specificity 86%). This is similar to the results produced by a linear kernel SVM (sensitivity 96%, specificity 87%). The specificity can be improved slightly (90%) at a slightly lower sensitivity (92%) by using a NN containing 25 hidden neurons on the full dataset, but this is only a small change.

For the purpose of generalizing the method to new cases, testing showed that a NN analyzing the full dataset might produce the best results for normal versus abnormal or versus dysplasia classification.

4.7. Implications and recommendations for THz imaging

This research potentially has implications for future investigation of THz imaging for colon cancer screening. Points of interest and recommendations arising from them include:

- Both air- and water-referenced parameters were found useful by DTs in this research, so both types of reference should continue to be measured when performing THz imaging.
- A relatively simple DT using as few as four THz imaging parameters classified the data well, providing 100% normal versus dysplasia classification accuracy in two of the three patients held out from training for testing purposes. This is a simple and readily comprehensible method of data analysis, easy to automate. Use of this method would simplify the THz imaging data analysis, as only four parameters would have to be calculated, instead of the full 17 investigated here.
- An alternative would be to use a NN with the full data set, as this performed best when generalizing methods to new patients. Once trained, NN are easy to use, and the ones used in this research were quick to make calculations and report results. However, they are 'black box' in nature and may not find favor with some physicians, although others may be willing to use the results as a diagnostic aid.
- An individualized analysis involving recording both definitely normal and suspected abnormal areas in each patient may not be necessary. The analysis methods trained on a mixed sample, mixed patient data set produced good results in several previously unseen samples from new patients. This would save time and processing when imaging.

4.8. Limitations

This research used a limited data set, containing relatively few samples. Ideally, the results and ideas arising from this work should be tested with a larger data set, containing a greater number of all three types of samples. This could provide more reliable

results to confirm which analysis method(s) are best used with THz imaging, and whether efficiency measures such as reduced input data sets can be used without decreasing the analysis accuracy. Reliable and repeatable results could be a valuable guide for physicians using THz imaging, allowing them to understand the sensitivity and specificity they can expect from the analysis and therefore the trustworthiness of the results.

The NN results reported here were absolute, either normal or abnormal with nothing in between, rounded to the nearest numerical result for interpretation. This could potentially be limiting the usefulness of the results; it would be interesting to consider a gradation scale of NN results. It could be investigated whether comparing the exact results (i.e., the decimal output of the NN) with the grade of the sample (i.e., high or low grade dysplasia) allows sensible connections to be made. This could help account for areas of possible overlap between normal and low grade dysplasia, and high grade dysplasia and tumor. This would require great care to be taken when recording the gold standard result for the experiments, possibly recruiting more than one pathologist to examine the samples and provide an opinion.

5. Conclusions

This research tested different intelligent analysis techniques to discover which THz data parameters and techniques provided the best diagnostic accuracy between normal and abnormal tissue samples in colon cancer. Suggestions for useful parameters were determined by DT, and the optimum analysis methods for the various relevant diagnostic tasks were found and discussed. Results also suggest that a non-patient-specific, generalized analysis method could be possible, although further testing with a larger data set is advised. The contribution of this research is to produce recommendations for THz imaging methodology and data analysis, hopefully helping to bring it closer to being a clinically useful technology.

Author Contributions and Acknowledgements

References

Arbab, M. H., Dickey, T. C., Winebrenner, D. P., Chen, A., Klein, M. B., & Mourad, P. D. (2011). Terahertz reflectometry of burn wounds in a rat model. *Biomedical Optics Express, 2*(8), 2339–2347.

Arnone, D. D., Ciesla, C. M., Corchia, A., Egusa, S., Pepper, M., Chamberlain, J. M., et al. (1999). Applications of terahertz (THz) technology to medical imaging. *Proceedings of SPIE, 3828*, 209–219.

Bennett, D. B., Taylor, Z. D., Tewari, P., Singh, R. S., Culjat, M. O., Grundfest, W. S., et al. (2011). Terahertz sensing in corneal tissues. *Journal of Biomedical Optics, 16*(5), 057003-1–057003-8.

Brun, M.-A., Formanek, F., Yasuda, A., Sekine, M., Ando, N., & Eishii, Y. (2010). Terahertz imaging applied to cancer diagnosis. *Physics in Medicine and Biology, 55*, 4615–4623.

Cancer Research UK website (2012). Bowel (colorectal) cancer – UK incidence statistics, available from: <http://info.cancerresearchuk.org/cancerstats/types/bowel/incidence/uk-bowel-cancer-incidence-statistics> Accessed July 2012.

Chen, H.-Z., Lu, J.-Y., & You, B. (2011) Terahertz endoscope based on anti-resonant reflecting hollow core waveguides. In *Proceedings of the optical society of America conference on lasers and electro-optics (CLEO)*, May 2011, 1–2.

Ciesla, C. M., Arnone, D. D., Corchia, A., Crawley, D., Longbottom, C., Linfield, E. H., et al. (2000). Biomedical applications of terahertz pulse imaging. *Proceedings of SPIE, 3934*, 73–81.

Fitzgerald, A. J., Berry, E., Zinovev, N. N., Walker, G. C., Smith, M. A., & Chamberlain, J. M. (2002). An introduction to medical imaging with coherent terahertz frequency radiation. *Physics in Medicine and Biology, 47*(7), 67–84.

Fitzgerald, A., Pickwell, E., Wallace, V., Purushotham, A., Pinder, S., Linan, M., Pye, R., & Ha, T. (2005) Medical applications of broadband terahertz pulsed radiation. In *The 18th annual meeting of the IEEE lasers and electro-optics society* (pp. 120–121). 22–28 Oct. 2005.

Fitzgerald, A. J., Wallace, V. P., Jimenez-Linan, M., Bobrow, L., Pye, R. J., Purushotham, A. D., et al. (2006). Terahertz pulsed imaging of human breast tumors. *Radiology, 239*, 533–540.

Ji, Y. B., Lee, E. S., Kim, S., Son, J., & Jeon, T. (2009). A miniaturized fiber-coupled terahertz endoscope system. *Optics Express, 17*(19), 17082–17087.

Joseph, C. S., Yaroslavsky, A. N., Neel, V. A., Goyette, T. M., & Giles, R. H. (2011). Continuous wave terahertz transmission imaging of nonmelanoma skin cancers. *Lasers in Surgery and Medicine, 43*, 457–462.

Mittleman, D. M., Gupta, M., Neelamani, R., Baraniuk, R. G., Rudd, J. V., & Koch, M. (1999). Recent advances in terahertz imaging. *Applied Physics B, 68*, 1085–1094.

NHS colon cancer website (2012). <http://www.nhs.uk/Conditions/Cancer-of-the-colon-rectum-or-bowel/Pages/Introduction.aspx> Accessed March 2012.

Reid, C. (2009). *Spectroscopic methods for medical diagnosis at terahertz wavelengths*. PhD thesis, University College, London.

Reid, C. B., Fitzgerald, A., Reese, G., Goldin, R., Tekkis, P., O'Kelly, P. S., et al. (2011). Terahertz pulsed imaging of freshly excised human colonic tissues. *Physics in Medicine and Biology, 56*(14), 4333–4353.

Sun, Y., Sy, M. Y., Wang, Y. J., Ahuja, A. T., Zhang, Y.-T., & Pickwell-MacPherson, E. (2011). A promising diagnostic method: Terahertz pulsed imaging and spectroscopy. *World Journal of Radiology, 3*(3), 55–65.

Sy, S., Huang, S., Wang, Y. J., Yu, J., Ahuja, A. T., Zhang, Y., et al. (2010). Terahertz spectroscopy of liver cirrhosis: investigating the origin of contrast. *Physics in Medicine and Biology, 55*, 7587–7596.

Wallace, V. P., Taday, P. F., Fitzgerald, A. J., Woodward, R. M., Cluff, J., Pye, R. J., et al. (2004). Terahertz pulsed imaging and spectroscopy for biomedical and pharmaceutical applications. *Faraday Discussions, 126*, 255–263.

Wallace, V. P., Fitzgerald, A. J., Pickwell, E., Pye, R. J., Taday, P. F., Flanagan, N., et al. (2006). Terahertz pulsed spectroscopy of human basal cell carcinoma. *Applied Spectroscopy, 60*(10), 1127–1133.

Wang, K., & Mittleman, D. M. (2004). Metal wires for terahertz wave guiding. *Nature, 432*, 376–379.

Woodward, R. M., Wallace, V. P., Pye, R. J., Cole, B. E., Arnone, D. D., Linfield, E. H., et al. (2003). Terahertz pulse imaging of ex vivo basal cell carcinoma. *Journal of Investigative Dermatology, 120*, 72–78.

World Health Organisation Global Health Observatory Data (2008). Available from: <http://apps.who.int/ghodata>/ Accessed July 2012.