2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)
October 11-14, 2020. Toronto, Canada

Tissue Discrimination from Impedance Spectroscopy as a Multi-objective Optimisation Problem with Weighted Naïve Bayes Classification

Brayden Kent, and Carlos Rossa
Faculty of Engineering and Applied Science
Ontario Tech University, Oshawa, ON, Canada
brayden.kent@ontariotechu.net; carlos.rossa@ontariotechu.ca

***Abstract*—Tissue classification from electrical impedance spectroscopy has several applications in diagnosis, surgical planning, and minimally invasive surgery. The method involves applying an alternating current to the sample and measuring its electric impedance at various frequencies. The spectrum is fit to a equivalent electric circuit that mimics the shape of the tissue's impedance spectrum. The model parameters are then used for classification.**

This paper proposes a new solution to decompose the model fitting problem into a form suitable for multi-objective optimisation, from which all the non-dominated solutions are used to form the database of parameters for a given tissue, as opposed to a single solution that is typically seen in impedance spectroscopy. The solution explores the use of the reference point dominance condition within Non-dominated Sorting Genetic Algorithm II to fit the data to the double dispersion Cole model. Each non-dominated solution contain values for the dispersion model elements. The multiple parameter value solutions from the optimiser are used as features in a weighted Naïve Bayes classifier to identify a new tissue sample. Experiments results in 3 different tissue samples shows that the method is successful in correctly labelling the data with an average accuracy of 89%.

***Index Terms*—Electric Impedance Spectroscopy, Weighted Naïve Bayes Classifier, Multi-objective Optimisation**

I. Introduction

Over the past few decades there has been increasing interest in new technologies to aid in the discrimination of healthy, benign and malignant tissues in cancer screenings. One such technology that has been developing steadily is the method of electric impedance spectroscopy. The method involves applying an alternating current across the sample and measuring the resulting voltage, at a wide range of frequencies. This method shows promise as it has been reported to observe differences in the electric impedance of tissues; including healthy, benign and malignant tissues in prostate [1], breast [2], bladder [3] and skin [4].

We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canadian Institutes of Health Research (CIHR), and the Social Sciences and Humanities Research Council of Canada (SSHRC), [funding reference number NFRFE-2018-01986].

Cette recherche a été financée par le Conseil de recherches en sciences naturelles et en génie du Canada (CRSNG), par les Instituts de recherche en santé du Canada (IRSC), et par le Conseil de recherches en sciences humaines du Canada (CRSH), [numéro de référence NFRFE-2018-01986].

There have been many attempts at extracting information from the impedance spectrum of a tissue to classify for malignancy. Such analysis may involve fitting the measured electric impedance to a model, and scrutinising the parameters of the model [5] [6]. An intuitive approach would be to explore a circuit comprised of an assortment of components that mimic the shape of the tissue's impedance spectrum. If the values for the circuit elements were known for a variety of tissue, one can attempt to classify a new tissue sample, with the circuit parameters known, using methods such as Naïve Bayes [7].

There is a long history detailing the development of such equivalent circuit models, and such models are better suited to different types of biological matter or for the range of frequencies sampled. An equivalent circuit model deemed the double dispersion Cole model, refer to Figure 1, has shown promise of fitting several tissue types across a wide range of frequencies [5].

There have been a few approaches to extracting the parameters of the double dispersion model from spectroscopy data. Freeborn has utilised the trust-region-reflective method [6] and Liu et al. have used modified particle swarm optimisation to solve for these parameters [8]. The aforementioned have explicitly only fit the model based on the magnitude data of the impedance, claiming hardware and cost restrictions. Recently, low cost devices have entered the market that are able to measure the magnitude and phase of a tissue sample across a wide range of frequencies [9]. Whereas in the single dispersion Cole models, several papers have been published where one considers both the real and imaginary parts of the impedance in fitting the parameters [10], as opposed to working the realm of magnitude and phase.

There are multiple ways to fit the measured electric impedance to the double dispersion Cole model. The issue seen across the deterministic methods is the convergence to local minima traps; which stems, in part, from the selection of the initial guess of the parameters. To overcome this, stochastic methods could be employed, in which the landscape of model parameter combinations are more widely explored. However, the randomness of the method does not guarantee convergence to the same set of parameters in a multi-modal landscape. The challenge of classifying tissue based on the

circuit parameters is made worse when an optimiser converges to a set of parameters that do not match the recognised set in a database. One could alter the initial parameters of these single objective methods to provide variance in the training data-set solutions, but the problem then becomes manipulated by the lens of the programmer, and could lead to other problems such as over-fitting.

If one were to consider a multi objective optimisation algorithm, one could fit the magnitude and the phase to the model concurrently. In addition this approach can also allow for other additional cost functions to be used, such as the desired location of cut-off frequencies for example. As seen in many multi-objective problems, as the number of objectives increases the problem becomes more challenging given that the inverse problem is ill posed, i.e., there may be more than a single combination of the circuit parameters that is able to satisfy the optimisation problem. Therefore, there is a need to address how to handle multiple solutions in this multi-objective problem. Common optimisation algorithms like particle swarm and simulated annealing are better suited for single objective functions, and are therefore not suited for this application. The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is a well known optimisation method that can handle multiple objective functions, and is explored here in this application.

This paper proposes a new solution to establishing the training data-set of model parameters and classifying the tissue. The suggested approach is to decompose the model fitting problem into a form suitable for multi-objective optimisation, from which all the non-dominated solutions are used to form the database of parameters for a given tissue, as opposed to a single solution and it is typically seen in impedance spectroscopy. Each of these non-dominated solutions contain values for the circuit model elements, and represent the Pareto front from NSGA-II. When a new impedance sample is to be tested, another set of potential solutions are obtained from its spectrum. All solutions are then fed into a weighted Naïve Bayes classifier where, based on the training data, a predicted label is given to each individual solution. The labels are counted, and the label that occurs the most within the test solutions is used to label the test sample.

The contributions in this paper are the decomposition of electric impedance into magnitude and phase to be used in a multi-objective optimisation approach for fitting the double dispersion Cole model. The NSGA-II method is used in conjunction with the reference point dominance (RPD) condition to ensure a diverse set of possible solutions that satisfy the objectives. A procedure is described on the handling of multiple non-dominated solutions and how a new sample is classified using these solutions through a weighted Naïve Bayes classifier.

II. Equivalent Circuit Model

As mentioned in the introduction, this paper utilises the double dispersion Cole model, see Figure 1. The model has been known to capture the resistive and psuedo-capacitive behaviour of biological tissues. Furthermore, it has shown to fit tissue impedance well over a wide range of frequency [5] and the additional number of parameters in the model, compared to the more simple models, potentially yield more markers to use when classifying tissue.

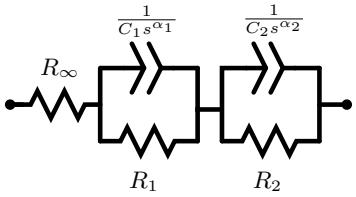

Fig. 1. The double-dispersion Cole model, commonly used to fit bioimpedance spectra. The model consists of resistors and constant phase elements (CPE).

The impedance of the double dispersion model is

$$Z(\omega) = R_\infty + \frac{R_1}{R_1 C_1 s^{\alpha_1} + 1} + \frac{R_2}{R_2 C_2 s^{\alpha_2} + 1} \tag{1}$$

where $s = j\omega$, with $j = \sqrt{-1}$ and ω as the frequency of the excitation signal. The capacitive behaviour of the tissue is represented by the constant phase elements (CPE), with C_1 and C_2. $0 \leq \alpha_{1,2} \leq 1$ are dimensionless coefficients. In the event either $\alpha_{1,2} = 1$ the CPE behaves like an ideal capacitor.

In this paper, the resistive nature of the tissue is represented by R_∞, R_1, R_2, where R_∞ represents the magnitude of the impedance as $\omega \to \infty$, R_1 (in combination with the other resistive elements) ensures the high magnitude impedance as $\omega \to 0$, and R_2 aids in the positioning of the cutoff frequencies.

The impedance in (1) can then be converted into the magnitude and phase as follows, refer to Table I,

$$\mathrm{Re}\Big(Z\Big) - j\,\mathrm{Im}\Big(Z\Big) = (z_1 + z_2) - j(z_3 + z_4) \tag{2}$$

$$|Z(\omega)| = \sqrt{(z_1 + z_2)^2 + (z_3 + z_4)^2} \tag{3}$$

$$\angle Z(\omega) = \frac{180^\circ}{\pi} \arctan\left(\frac{z_3 + z_4}{z_1 + z_2}\right) \tag{4}$$

Let $\mathbf{p} \in \mathbb{R}^+$ be the vector containing the 7 unknown model parameters that form the double dispersion model,

$$\mathbf{p} = [R_\infty, \quad R_1, \quad R_2, \quad C_1, \quad C_2, \quad \alpha_1, \quad \alpha_2] \tag{5}$$

The goal is, for a given measured impedance spectrum, to find the parameter values that would result in similar calculated impedance from (1) to (4). Thus, one needs an optimisation method that determines these parameter values such that the calculated impedance matches the measured impedance at various frequencies.

III. Fitting EIS Data to Equivalent Circuit

The purpose of the optimisation method is to minimise the error between the measured impedance and the calculate impedance of the model. As seen in (2), the impedance is complex, and as a result many researchers approach this

TABLE I
THE IMPEDANCE MODEL PARAMETERS FOR THE DOUBLE DISPERSION MODEL WITH CPE AS DEFINED IN EQUATIONS (2) TO (4)

Var.	Equation
$z_1 =$	$\frac{R_\infty + R_1(R_1C_1e^{a_2\ln(w)}\cos(a_2\frac{\pi}{2})+1)}{(R_1C_1e^{a_2\ln(w)}\cos(a_2\frac{\pi}{2})+1)^2+R_1^2C_1^2e^{(a_2*\ln(w))^2}\sin(a_2\frac{\pi}{2})^2}$
$z_2 =$	$\frac{R_2(R_2C_2e^{a_3\ln(w)}\cos(a_3\frac{\pi}{2})+1)}{(R_2C_2e^{a_3\ln(w)}\cos(a_3\frac{\pi}{2})+1)^2+R_2^2C_2^2e^{(a_3\ln(w))^2}\sin(a_3\frac{\pi}{2})^2}$
$z_3 =$	$\frac{R_1^2C_1e^{a_1\ln(w)}\sin(a_1\frac{\pi}{2})}{(R_1C_1e^{a_1\ln(w)}\cos(a_1\frac{\pi}{2})+1)^2+R_1^2C_1^2e^{(a_1\ln(w))^2}\sin(a_1\frac{\pi}{2})^2}$
$z_4 =$	$\frac{R_2^2C_2e^{a_2\ln(w)}\sin(a_2\frac{\pi}{2})}{(R_2C_2e^{a_2\ln(w)}\cos(a_2\frac{\pi}{2})+1)^2+R_2^2C_2^2e^{(a_2\ln(w))^2}sin(a_2\frac{\pi}{2})^2}$

problem using complex non-linear least squares methods. In this paper however, a different approach is taken.

Consider that the electric impedance of the sample tissue is measured at n frequencies. The impedance data is organised into $n\times 1$ column vectors, containing the measured impedance magnitude $\mathbf{z}_m$ and phase $\mathbf{z}_p$ separately. If the parameters of the circuit $\mathbf{p}$ were known, then equations (3) and (4) both form $n \times 1$ column vectors as well, and yield the estimated impedance magnitude $\hat{\mathbf{z}}_m$ and phase $\hat{\mathbf{z}}_p$ respectively.

There exists two desired tasks: to minimise the error in the estimated and measured magnitude $\hat{\mathbf{z}}_m - \mathbf{z}_m$, and phase $\hat{\mathbf{z}}_p - \mathbf{z}_p$, by only altering $\mathbf{p}$. This can be formed into a multi-objective problem as [11],

$$\begin{aligned} \textbf{Minimise:} \quad & \mathbf{F}(\mathbf{p}) = (f_1(\mathbf{p}), f_2(\mathbf{p}))^T, \mathbf{p} \in \Omega \\ \textbf{Subject to:} \quad & g_k(\mathbf{p}) \geq 0 \quad k = 1 \dots P \\ & h_l(\mathbf{p}) = 0 \quad l = 1 \dots Q \end{aligned}$$

where P and Q are the numbers of inequality and equality constraints, respectively. Ω is the variable space for $\mathbf{p}$ as a candidate solution. $\mathbf{F}$ houses the objective functions,

$$f_1(\mathbf{p}) = \min_{\mathbf{p}\in\mathbb{R}} \left(\hat{\mathbf{z}}_m(\mathbf{p}) - \mathbf{z}_m \right) \tag{6}$$

$$f_2(\mathbf{p}) = \min_{\mathbf{p}\in\mathbb{R}} \left(\hat{\mathbf{z}}_p(\mathbf{p}) - \mathbf{z}_p \right) \tag{7}$$

This may seem unintuitive to decompose the impedance problem, but this enables the problem to include more objective functions in the future.

During optimisation with more than one objective, there is not a guaranteed global solution that minimises all objective functions. More often, there will be a set of solutions that are nondominated, see Figure 2. The challenge that typically comes with finding the multi-objective solutions is that, as the number of objectives increase, the more likely the solutions are to become nondominated, resulting in poor convergence or representation of solutions. Thus, this paper includes the approach presented by Elarbi et al. [11] to prevent this from occurring.

The solution proposed in [11] is to define a set of evenly distributed reference points across the hyperplane of the search space [11]. These reference points are used to evaluate the convergence and diversity of nearby candidate solutions. The reference point set W is generated by,

$$W = \begin{pmatrix} m+v-1 \\ v \end{pmatrix} \tag{8}$$

where m is the number of objective functions, and v is the number of divisions along the objective. As previously mentioned, the distance from the solution candidates to the reference points are determined as,

$$d_1(\mathbf{p}) = \frac{\left\| \tilde{f}(\mathbf{p})^T R_k \right\|}{\|R_k\|} \tag{9}$$

$$d_2(\mathbf{p}) = \left\| \tilde{f}(\mathbf{p}) - d_1(\mathbf{p}) \left(\frac{R_k}{\|R_k\|} \right) \right\| \tag{10}$$

with $\tilde{f}(\mathbf{p})$ being the normalised objective function. R_k is an M-dimensional direction vector [11].

The method selected to solve the multi-objective problem is the wildly popular Non-dominated Sorting Genetic Algorithm II (NSGA-II) [12], but with the additional consideration of reference point dominance (RPD) [11]. There are several parameters the user needs to define within RPD-NSGA-II. Namely, the number of reference points, the population size, search space bounds and the crossover and mutation rates. Generally speaking, increasing the number of reference points will result in a better representation of the Pareto front, but at the cost of increasing the number of computations per reference point, see (9) and (10). A population member in NSGA-II is simply a combination of values for $\mathbf{p}$. Increasing the size of the population will enable exploring multiple solutions to the multi-objective problem. Similar to the reference points, increasing the size of the population increases the number of computations significantly. The user must also define bounds of the search space in each dimension. In other words, upper and lower limits for each value within $\mathbf{p}$. Lastly, as part of NSGA-II, the population members undergo operations known as crossover and mutation. These operations each require a modifying value that is used in part to determine how the population children for the generation of the optimiser are created [13]. Selecting values for these rates will vary for the application. An emphasis on mutation can help solutions escape from local minima in multi-modal objective problems.

The method described above attempts to minimise both d_1 and d_2, which will ensure good convergence and diversity of the potential solutions respectively.The RPD-NSGA-II algorithm is well documented in [11], and the reader is directed there for further reading.

IV. TISSUE CLASSIFICATION

The optimiser will produce a set of equivalent circuit model parameters that best matches the impedance of the tissue at all measured frequencies. It is desired to use these parameters as features to determine what the tissue is. Due to the variability

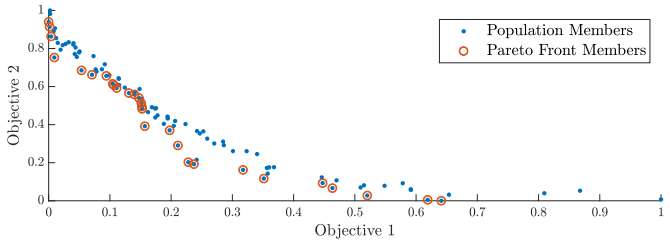

Fig. 2. Normalised fitness values for the population members after 50 generations of the optimisation method. Encircled members constitute the non-dominated population, which are used to represent the final set of parameter combinations for the sample.

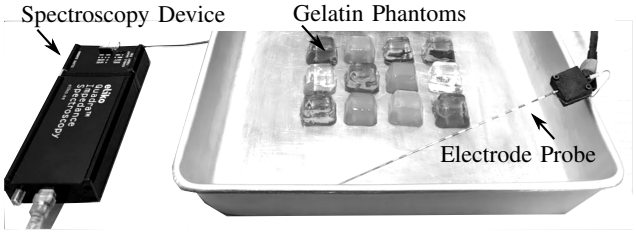

Fig. 3. The Quadra spectroscopy system was used with a bipolar electrode embedded probe [18] to measure the electric impedance of gelatin phantoms.

in tissue samples and the stochastic nature of the optimiser, it is not guaranteed that features represent the physical values of the tissue, and will be identical from one sample to the next. Furthermore, the relationship that exists between the parameters and impedance may yield overlapping regions from differing tissues.

Thus, when it comes to classifying the tissue based on these parameter values a probability based approach is considered. The Naïve Bayes classifier is well suited for these types of problems. One defines a set of features from the circuit parameters $\mathbf{p}$ that belongs to a tissue type class c. The probability of a test sample belonging to a class given a set of features is defined as [14],

$$P(c_k|\mathbf{p}_i) = \frac{P(\mathbf{p}_i|c_k)P(c_k)}{P(\mathbf{p}_i)} \tag{11}$$

where $\mathbf{p}_i$ denotes the i^{th} parameter of $\mathbf{p}$ and c_k refers to a specific tissue type class. $P(c_k|\mathbf{p}_i)$ is then a probability density function for a class given a feature, assuming a Gaussian distribution. In essence, from this formula, one can determine which class a given parameter will belong to based on which has the largest probability [15], [16],

$$\text{predicted label} \leftarrow \underset{i=1\ldots7}{\arg\max}\, P(c_k|\mathbf{p}_i) \tag{12}$$

The combined probability density function for the set of parameters is written as,

$$P(\mathbf{p}|\mathbf{c_k}) = \prod_{i=1}^{7} P(\mathbf{p}_i|c_k) \tag{13}$$

the predicted label can be found by considering the highest probability from multiple parameters,

$$\text{predicted label} \leftarrow \underset{i=1\ldots7}{\arg\max}\, P(c_k)\prod_{i=1}^{7} P(\mathbf{p}_i|c_k) \tag{14}$$

While it is desirable to consider all of the model parameters, some model parameters may be better markers for classification. For example, R_∞ may be distinct among the different tissue samples, whereas values for α_1 might be similar across all tissues. For this reason, a weighted classifier is considered:

$$\text{predicted label} \leftarrow \underset{i=1\ldots7}{\arg\max}\, P(c_k)\prod_{i=1}^{7} P(\mathbf{p}_i|c_k)^{\mathbf{w}_i} \tag{15}$$

where $\mathbf{w}_i$ is the weight attributed to parameter $\mathbf{p}_i$ [17]. Based on a training set of parameter data, a new tissue sample can be labelled based on the probability of it belonging to a type of tissue with similar circuit parameters.

Assume one has taken impedance measurements for different tissue samples and has extracted multiple unique sets of parameters $\mathbf{p}$ that satisfy the multi-objective optimisation for each tissue, i.e., the Pareto-front solutions from the multi-objective optimisation. Each individual set of parameters is identified with their corresponding tissue type and sample number and used in the classifier to determine the probability density functions. This is the training data set.

With a new sample to be labelled, the optimisation will again yield multiple sets of potential solutions $\mathbf{p}$ for it. Each of these new sets of parameters is then classified individually, using the training data described earlier, with the weighted classifier. Once all solutions to the new sample have been assigned predictive labels, the labels are tallied. The label with the highest recurrence is then assigned to the test sample. A flow chart of this process is presented in Figure 4.

V. Experiment Setup

The proposed method for tissue classification is experimentally tested to determine its efficacy. This section describes the experimental validation setup and protocol. Gelatin, with varying concentrations of salt, is used as the tissue for the experiments. The following 3 recipes are used: **Tissue 1** (T1) has a ratio of 6.7 grams porcine gelatin powder in 80 mL deionized water; **Tissue 2** (T2) has a ratio of 6.7 grams porcine gelatin powder in 80 mL deionized water with 2.0 grams of ionized salt. A small drop of green food colouring is used to tint the gel; and **Tissue 3** (T3) has a ratio of 6.7 grams porcine gelatin powder in 80 mL deionized water with 5.0 grams of ionized salt. A small drop of red food colouring is used to tint the gel.

To ensure variability in the data, 4 batches are made separately for each phantom, where each batch yielded 4 samples. Thus, 16 phantoms are created for each tissue type, which gives a total of 48 samples. The phantoms are separated into two categories for the training data set and the test data set. There are 4 samples of each tissue type set aside to form the training data (12 total samples). This leaves 36 samples to be used in the test data set, 12 samples of each tissue type. In

TABLE II
PARAMETER INITIAL VALUE BOUNDS (TOP), AND OPTIMISER PARAMETERS FOR TRAINING AND TESTING DATA (BOTTOM)

Var. Symbol	Lower Bound	Upper Bound
R_∞	1.0e0	2.0e3
R_1	1.0e5	2.0e7
R_2	1.0e0	2.0e3
C_1	1.0e−9	1.0e−4
C_2	1.0e−9	1.0e−4
α_1	0.40	1.0
α_2	0.40	1.0

Variable	Symbol	Training	Test
Number of Objectives	m	2	2
Dimension of Problem	P	7	7
Population Size	N	100	30
Crossover Rate	p_c	20	20
Mutation Rate	p_m	20	20
Number of Divisions	v	20	20
Number of Generations		5000	2000

the figures and tables of this paper, the training data samples are denoted TxSx, for brevity. For example, the second sample of T3 is labelled as T3S2.

The experiment included measuring the electric impedance of each phantom tissue sample with a spectroscopy system (Quadra, Eliko, Tallin, Estonia) [9]. The electric impedance is determined at $n = 23$ frequencies, spanning 10.4 Hz to 349 kHz, $\boldsymbol{\omega} = [10.4,\ 20.8,\ 31.2, \cdots\ 179000,\ 251000,\ 349000]^\top$ Hz. The device presents the impedance as the magnitude and phase (degrees) at each frequency (Hz).

Twelve gelatin samples are used as the training data for the experiment. First, the magnitude and phase of the impedance for each of the these samples is recorded. The double dispersion model parameters that fit the impedance for each tissue were desired. The parameters are determined using the RPD-NSGA-II method with the considerations shown in Table II.

For a given sample, $N = 100$ random combinations of parameters $\mathbf{p}$ are generated using the bounds in Table II. The initial parameters are chosen randomly, within reasonable bounds, as to remove the dependency on being near a solution that is typically required in the deterministic methods. The reference point set is created with $v = 20$ divisions, linearly spanning [1,0] to [0,1] in the normalised objective space. The optimiser is run for 5000 generations using crossover and mutation rates of $p_c = 20$ and $p_m = 20$ respectively [13]. The fitness function for each population member is detailed in (6) (7). With the optimisation complete the Pareto front is extracted, consisting of 100 combinations of parameters that minimise the objective functions while being diverse due to the RP diversity condition. This process is repeated for each of tissue samples in the training data set. Examples of the measured impedance and the calculated impedance from a set of optimised parameters are shown in Figure 3.

The test data is created in a similar fashion to the training data but with fewer population members and fewer generations, see Table II (top). The initial values for the model parameters in forming the test data are also randomly generated

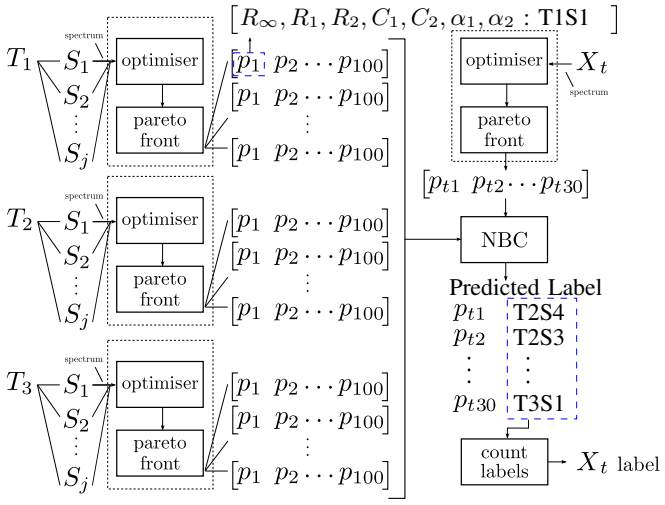

Fig. 4. A flow diagram of the classification method with multiple solutions. The left half of the figure depicts the formation of the training data-set. Tissues (T_1, T_2, T_3) each have multiple samples S_j that house the impedance magnitude and phase spectrum. Using this spectrum, a set of 100 solutions $\mathbf{p}$ are created. Each of these solutions contain values for the 7 double dispersion model elements, as well as its supervisor assigned label. On the right side of the figure, a new impedance sample X_t is to be tested to determine its label. 30 solutions for the parameters are determined and are fed into the weighted Naïve Bayes classifier where, based on the training data, a predicted label is given to each test solution. The labels are counted, and the label that occurs the most within the test solutions is used to label the test sample.

using the bounds in Table II (bottom). When the optimisation for the each of the test samples is complete, each Pareto front member is labelled based on the training data and classifier described earlier in this paper. With each Pareto front solution labelled, the occurrence of each label are counted. The label with the largest number is then used as the label for the sample. A flow chart summarising the proposed classification method is shown in Figure 4.

The weights assigned to the classifier are selected as, $w = [4.40, 0.85, 0.10, 0.65, 0.85, 2.00, 1.85]$. These values are determined using the random-walk method described in [14] and empirical testing with the training data.

VI. EXPERIMENT RESULTS AND DISCUSSION

With all of the test data classified, the results are compared to the ground-truth. The comparison of the actual label against the classified label is shown in Table 6. The classifier is able to correctly identify T1 and T2 perfectly. While the majority is correct, the classifier struggled in correctly labelling T3. The total accuracy of test data was 89%. The optimisation of the test data parameters and its classification was run multiple times to account for the stochastic nature of the method. The method is consistent in its accuracy, ranging from 86-92%.

The results indicate that the method is capable of correctly identifying the tissue when it has a distinctly different impedance spectrum from other tissues. It is seen in Table 6 that the method struggled with classifying one of the tissue types. Further analysis of the data reveals the same T3 samples in the test are consistently mislabelled as T2S4 from the

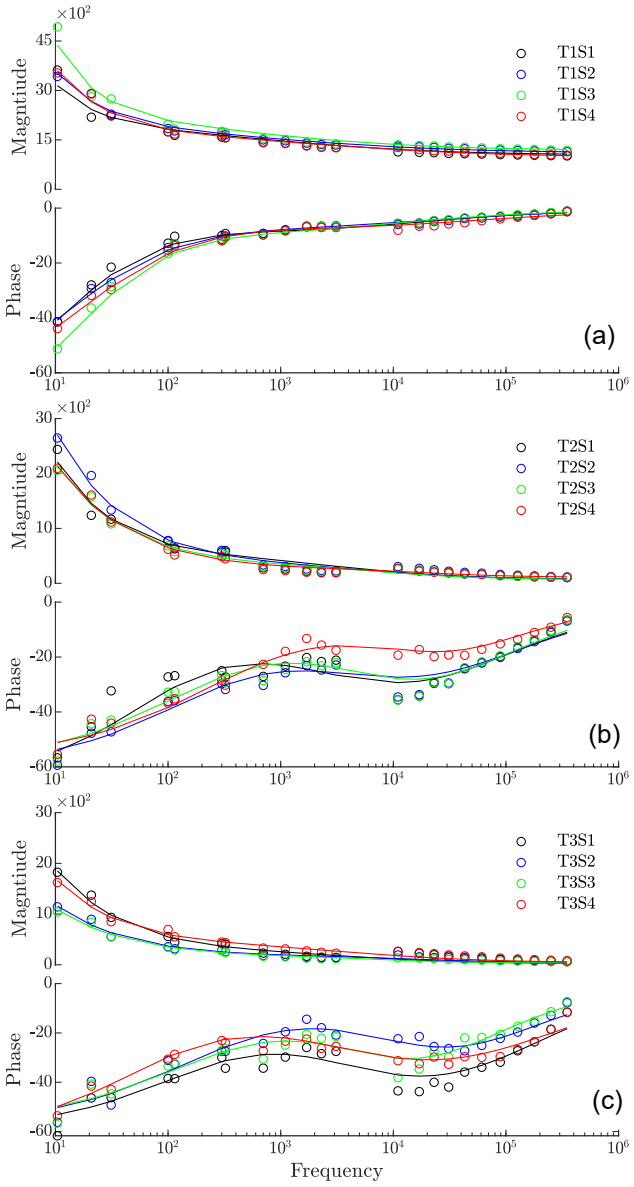

Fig. 5. Comparison of the measured impedance (dots) from the training data set and the calculated impedance (line) from the double dispersion model using one set of the optimised parameters. Magnitude and phase presented as Ω and degrees respectively.

Predicted Class

True Class	T1	T2	T3
T1	1.00	0.00	0.00
T2	0.00	1.00	0.00
T3	0.00	0.33	0.67

Fig. 6. A confusion matrix representing the actual tissue and the classifier label. Data is shown as percentages of 5 runs of the optimiser and classifier, with the same 36 samples per run. Average correct accuracy is 89%, most accurate run is 92%, least accurate run is 86%.

training data set. Interestingly, the average parameters of T2S4 are quite different from that of any of T3 samples, seen in Table III. However, the standard deviation of R_2 in T2S4 is also abnormally large, which may have also contributed to this issue. The root cause may stem from the impedance spectra itself, the magnitude of the impedance is significantly greater at all frequencies relative to correctly identified T3 samples. Therefore these samples may have more similar impedance to the T2 training data samples than the T3 samples. Lastly, as shown in Figure 5, the phase for the T2S4 sample deviates from the other samples at higher frequencies, which may have also been a source of the over-representation.

From Table III it is evident that the most unique parameter across the tissue types is R_∞, which is correlated to the magnitude of the impedance at higher frequencies. However, should one simply classify the tissue based off of this information alone they ignore the intricacies that happen along the spectrum. It may be worthwhile to investigate a decision tree based approach, where tissues are initially classified by the high frequency magnitude, then further classified by additional model parameters in the event of multiple possible labels. Furthermore, additional metrics could be considered in the classification process, such as cutoff frequencies.

The large deviation of R_1, which in this application correlates to the magnitude of the impedance at low frequencies, is a result of limited information available regarding the impedance at low (near DC) frequencies due to hardware limitations of the spectroscopy device. As seen in Figures 5(a) and 5(c), the magnitude of the impedance appears to have asymptotic behaviour as the frequency approaches DC. If this were true it would suggest open-circuit behaviour typically seen in capacitive elements, thus R_1 would effectively be removed from the circuit model. Therefore, reliable information about the impedance at low frequencies is critical in determining an accurate value for R_1.

Taking into account the importance of R_∞ and the uncertainty of R_1, among the characteristics of the other parameters, it is evident that the selection of the classifier weights is crucial to its success. The random-walk method proved to yield acceptable values for this application, but further refinement of these weights may improve performance.

Furthermore, the properties of the optimiser including population size, number of generations, mutation and crossover rates directly impact the performance of the method. While not covered in the scope of this paper, one should be cognizant of the impact altering these parameters will have on the accuracy of the method.

To summarise the above, further development is needed in the method to handle cases like T3, where classification accuracy was low. Future work should focus on improvements to the classifier or refinement of the model.

VII. Conclusion

This paper presented a novel method for extracting the parameters from the double dispersion model using a multi-objective technique, which are used to classify new samples.

TABLE III
AVERAGE AND STANDARD DEVIATION OF PARAMETER VALUES FROM THE TRAINING DATA SAMPLES ACROSS ALL SOLUTIONS

Sample	R_∞	R_1	R_2	C_1	C_2	α_1	α_2
T1S1	1.0e3 ± 9.2e0	9.9e6 ± 1.4e6	1.0e3 ± 2.9e1	8.1e−6 ± 1.4e−7	1.0e−5 ± 9.2e−7	0.82 ± 5.3e−3	0.44 ± 8.0e−3
T1S2	8.8e2 ± 1.2e1	7.1e6 ± 1.6e6	6.0e2 ± 6.1e1	1.7e−5 ± 1.2e−6	1.0e−5 ± 1.5e−6	0.73 ± 1.7e−2	0.48 ± 1.0e−2
T1S3	1.1e3 ± 1.3e1	9.2e6 ± 1.2e6	8.8e2 ± 3.2e1	6.2e−6 ± 1.3e−6	5.3e−6 ± 5.9e−7	0.77 ± 3.7e−3	0.53 ± 1.1e−2
T1S4	1.0e3 ± 1.5e1	1.1e7 ± 1.5e6	7.5e2 ± 5.3e1	1.7e−5 ± 1.2e−7	1.5e−5 ± 1.7e−6	0.79 ± 2.5e−3	0.49 ± 1.2e−2
T2S1	1.3e3 ± 4.3e0	1.2e7 ± 1.5e6	7.1e2 ± 2.2e1	8.5e−7 ± 3.8e−7	3.2e−6 ± 2.7e−7	0.80 ± 6.8e−2	0.57 ± 7.3e−2
T2S2	9.8e2 ± 3.5e0	9.9e6 ± 1.8e6	4.1e2 ± 3.9e1	3.2e−5 ± 3.6e−6	3.0e−6 ± 4.2e−7	0.67 ± 2.7e−2	0.61 ± 1.5e−2
T2S3	1.3e3 ± 2.3e1	9.5e6 ± 1.1e6	3.3e2 ± 4.9e1	2.7e−5 ± 2.6e−6	7.7e−6 ± 5.8e−6	0.73 ± 1.9e−2	0.59 ± 3.1e−2
T2S4	1.1e3 ± 1.9e1	9.6e6 ± 1.5e6	3.9e2 ± 3.4e2	5.8e−5 ± 2.9e−5	7.7e−6 ± 7.9e−6	0.57 ± 6.5e−2	0.80 ± 1.4e−1
T3S1	2.4e1 ± 7.7e0	1.1e7 ± 8.8e5	1.9e3 ± 2.2e1	6.8e−5 ± 9.4e−7	1.8e−5 ± 1.9e−6	0.40 ± 1.4e−3	0.99 ± 3.1e−3
T3S2	3.8e1 ± 2.8e0	1.1e7 ± 1.1e6	2.7e2 ± 3.1e1	4.9e−5 ± 5.1e−6	4.2e−6 ± 1.3e−6	0.63 ± 2.6e−2	0.64 ± 2.0e−2
T3S3	3.7e1 ± 1.2e0	1.0e7 ± 2.6e6	1.4e2 ± 7.3e0	3.4e−5 ± 4.7e−6	2.3e−6 ± 3.9e−7	0.73 ± 2.3e−2	0.76 ± 1.5e−2
T3S4	3.0e1 ± 3.5e0	1.1e7 ± 9.6e5	1.5e2 ± 3.9e1	1.6e−5 ± 5.4e−6	6.4e−6 ± 6.0e−6	0.81 ± 6.1e−2	0.77 ± 1.0e−1

The optimisation method is implemented using NSGA-II with the reference point dominance constraint, which yielded a diverse set of possible solutions that fit the measured impedance.

The method is successful in its classification, but struggled with similar tissue samples. Further refinement of the technique is required and additional methods of classifying tissue from equivalent circuit parameters should be investigated to improve the algorithm. Nevertheless, the technique showcased the ability to find multiple solutions that fit the double dispersion model without having to be near a solution, as seen in the deterministic methods.

Lastly, while the method was executed multiple times to confirm reputability, future development should consider a larger data-set where cross validation can be performed.

With further development, this method may yet be useful in the field of impedance spectroscopy, and by extension, aid in the discrimination of healthy, benign and malignant tissues during cancer screenings.

REFERENCES

[1] V. Mishra, A. Schned, A. Hartov, J. Heaney, J. Seigne, and R. Halter, "Electrical property sensing biopsy needle for prostate cancer detection," *The Prostate*, vol. 73, no. 15, pp. 1603–1613, 2013.

[2] J. Jossinet, "The impedivity of freshly excised human breast tissue," *Physiological measurement*, vol. 19, no. 1, p. 61, 1998.

[3] A. Keshtkar, A. Keshtkar, and R. H. Smallwood, "Electrical impedance spectroscopy and the diagnosis of bladder pathology," *Physiological Measurement*, vol. 27, no. 7, p. 585, 2006.

[4] P. Åberg, U. Birgersson, P. Elsner, P. Mohr, and S. Ollmar, "Electrical impedance spectroscopy and the diagnostic accuracy for malignant melanoma," *Experimental dermatology*, vol. 20, no. 8, pp. 648–652, 2011.

[5] T. J. Freeborn, "A survey of fractional-order circuit models for biology and biomedicine," *IEEE Journal on emerging and selected topics in circuits and systems*, vol. 3, no. 3, pp. 416–424, 2013.

[6] T. J. Freeborn, B. Maundy, and A. S. Elwakil, "Extracting the parameters of the double-dispersion cole bioimpedance model from magnitude response measurements," *Medical & biological engineering & computing*, vol. 52, no. 9, pp. 749–758, 2014.

[7] S. M. Moqadam, P. K. Grewal, Z. Haeri, P. A. Ingledew, K. Kohli, and F. Golnaraghi, "Cancer detection based on electrical impedance spectroscopy: A clinical study," *Journal of Electrical Bioimpedance*, vol. 9, no. 1, pp. 17–23, 2018.

[8] L. Liu, L. Shan, C. Jiang, Y.-W. Dai, C.-L. Liu, and Z.-D. Qi, "A modified pso algorithm for parameters identification of the double-dispersion cole model," *Journal of Circuits, Systems and Computers*, vol. 27, no. 13, p. 1850210, 2018.

[9] M. Min, M. Lehti-Polojärvi, J. Hyttinen, M. Rist, R. Land, and P. Annus, "Bioimpedance spectro-tomography system using binary multifrequency excitation," *International Journal of Bioelectromagnetism*, vol. 209, pp. 76–79, 05 2018.

[10] D. Ayllon, F. Seoane, and R. Gil-Pita, "Cole equation and parameter estimation from electrical bioimpedance spectroscopy measurements-a comparative study," in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2009, pp. 3779–3782.

[11] M. Elarbi, S. Bechikh, A. Gupta, L. B. Said, and Y.-S. Ong, "A new decomposition-based nsga-ii for many-objective optimization," *IEEE transactions on systems, man, and cybernetics: systems*, vol. 48, no. 7, pp. 1191–1210, 2017.

[12] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[13] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, and A. Fast, "Nsga-ii," *IEEE transactions on evolutionary computation*, vol. 6, no. 2, pp. 182–197, 2002.

[14] K. B. Korb and A. E. Nicholson, *Bayesian artificial intelligence*. CRC press, 2010.

[15] S. Raschka, "Naive bayes and text classification i-introduction and theory," *arXiv preprint arXiv:1410.5329*, 2014.

[16] S. Theodoridis, *Machine learning: a Bayesian and optimization perspective*. Academic Press, 2015.

[17] H. Zhang and S. Sheng, "Learning weighted naive bayes with accurate ranking," in *Fourth IEEE International Conference on Data Mining (ICDM'04)*. IEEE, 2004, pp. 567–570.

[18] B. Kent, A. Cusipag, and C. Rossa, "Tissue discrimination through force-feedback from impedance spectroscopy in robot-assisted surgery," *International Conference on Smart Multimedia, San Diego, Dec 2019*, pp. 274–285, 2019, doi: 10.1007/978-3-030-54407-2_23.