Support Vector Machine-Assisted Diagnosis Method for Human Malignant Gastric Tissue Based on Dielectric Properties (Postprint)

Authors: Zhang Sa, Li Zhou, Xin Xuegang

Date: 2018-06-15T00:00:00+00:00

Abstract

Objective: To automatically differentiate between normal and malignant gastric tissues based on differences in their dielectric property values using support vector machine (SVM). Methods: The open-ended coaxial probe method was used to measure the dielectric properties of normal and malignant gastric tissues within the frequency range of 42.58–500 MHz, and the measured dielectric property data were fitted using the Cole-Cole model. Receiver operating characteristic (ROC) curve analysis was employed to evaluate the discriminative capability of permittivity, conductivity, and Cole-Cole fitting parameters at each frequency point. SVM was utilized to differentiate between normal and malignant gastric tissues, with classification accuracy calculated via k-fold cross-validation. Results: Within the measured frequency range, the area under the ROC curve for permittivity at five low-end frequency points exceeded 0.8. The combination of permittivity values at these five frequencies, used as feature values in conjunction with SVM, achieved the highest classification accuracy of 84.38%, with a MATLAB runtime of 3.40 s. Conclusion: The SVM-assisted diagnostic method for malignant human gastric tissue based on dielectric properties proposed in this study is feasible.

Full Text

Preamble

ZHANG Sa[1], LI Zhou[2], XIN Xuegang[1]
[1]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China
[2]Department of General Surgery, Zhujiang Hospital, Southern Medical University, Guangzhou 510280, China

Abstract

Objective: To achieve automatic discrimination between normal and malignant gastric tissues based on discrepancies in their dielectric properties using support vector machine (SVM). **Methods:** The dielectric properties of normal and malignant gastric tissues were measured across the frequency range of 42.58–500 MHz using the open-ended coaxial probe method, and the measured data were fitted using the Cole-Cole model. Receiver operating characteristic (ROC) curve analysis was employed to evaluate the discriminative capability of permittivity, conductivity, and Cole-Cole fitting parameters at each frequency point. SVM was used to classify normal and malignant gastric tissues, with discrimination accuracy calculated via k-fold cross-validation. **Results:** Within the measured frequency range, the area under the ROC curve (AUC) for permittivity at five lower-frequency points exceeded 0.8. The combination of permittivity values at these five frequencies as feature vectors, together with SVM, achieved the highest discrimination accuracy of 84.38% with a MATLAB runtime of 3.40 s. **Conclusion:** The proposed SVM-assisted diagnostic method for human malignant gastric tissues based on dielectric properties is feasible.

Keywords: dielectric properties; malignant gastric tissues; discrimination; ROC curve; support vector machine

Introduction

Dielectric properties are inherent physical attributes of human tissues, comprising permittivity ( ) and conductivity ( ) [1]. These properties are associated with tissue water content, protein concentration, ion concentration, and other factors [2–4], as well as the microenvironment including microcirculation, and exhibit frequency dependence. When tissues undergo malignant transformation, these dielectric property-related factors change, leading to significant differences in dielectric values between normal and malignant tissues [2, 4]. Theoretically, these differences exist prior to morphological changes, potentially enabling early detection of suspicious lesions. This could provide a novel technical approach for cancer diagnosis, where suspicious areas identified through dielectric analysis can be targeted for biopsy.

Gabriel et al. [5–7] systematically investigated the dielectric spectra of healthy human tissues and established a comprehensive database. Subsequent simultaneous measurements of normal and malignant tissues have demonstrated significant dielectric differences [8–12], confirming the potential for tissue characterization based on these properties. However, no clear criteria currently exist regarding the magnitude of dielectric differences required to classify tissue as suspiciously malignant. Therefore, developing discrimination methods based on the degree of dielectric property differences is of significant importance. Several studies have pursued this direction. Truong et al. [13, 14] employed machine learning methods to discriminate normal from malignant breast and skin tissues at THz frequencies, achieving classification accuracies above 90% for both tissue types. Yilmaz et al. [15] measured the dielectric properties of normal and malignant rat liver tissues across 500 MHz–6 GHz and achieved over 90% classification accuracy using machine learning. However, no reported studies have applied this approach for automatic discrimination of human normal and malignant gastric tissues based on dielectric differences. This study measures the dielectric properties of human gastric normal and cancerous tissues intraoperatively across 42.58–500 MHz using the open-ended coaxial probe method, and explores the feasibility of automatic discrimination using machine learning. ROC curve analysis is utilized to evaluate the discriminative capability of various parameters, from which high-performance features are selected for machine learning-based classification. Given that our data are numerical and represent a binary classification problem, we selected support vector machine (SVM) for discriminating normal and malignant gastric tissues.

1.1 Data Collection and Cole-Cole Modeling

This study employed the open-ended coaxial probe method [16–18] to measure the dielectric properties of freshly excised malignant gastric tissues and adjacent normal tissues in the operating room. The measurement frequency range was 42.58–500 MHz, encompassing 43 frequency points. A total of 290 samples from 26 patients were measured, including 100 normal samples and 190 malignant samples. Patient ages ranged from 36 to 80 years, and ex vivo measurement temperatures ranged from 23.8–32.8 °C.

The measured dielectric properties were fitted using the Cole-Cole model [19–21]. The single-pole Cole-Cole fitting formula is:

$$\varepsilon^*(\omega) = \varepsilon'(\omega) - j\varepsilon''(\omega) = \varepsilon_\infty + \frac{\Delta\varepsilon}{1 + (j\omega\tau)^{1-\alpha}} + \frac{\sigma_i}{j\omega\varepsilon_0}$$

where ω is the angular frequency, $\varepsilon^*(\omega)$ is the frequency-dependent complex permittivity, $\varepsilon'(\omega)$ is the frequency-dependent permittivity, $\varepsilon''(\omega)$ is the frequency-dependent dielectric loss (convertible to conductivity: $\sigma(\omega) = \omega\varepsilon_0\varepsilon''(\omega)$), and ε_∞, σ_i, $\Delta\varepsilon$, τ, and α are the Cole-Cole model fitting parameters estimated from experimental data. The Cole-Cole fitting was performed using a software package from the literature [22]. Figure 1 [Figure 1: see original paper] shows the raw measurement data and Cole-Cole fitting curves for two different samples (normal and malignant tissues). Thus, each sample includes permittivity and conductivity at 43 frequencies, plus five Cole-Cole fitting parameters. These dielectric values and fitting parameters were used for subsequent ROC curve analysis to evaluate their discriminative capability.

1.2 ROC Curve Analysis

ROC curve analysis [23, 24] is a valuable technique for characterizing and visualizing classifier performance, commonly employed in medical decision-making and increasingly utilized in machine learning and data mining. The ROC curve is a two-dimensional plot with false positive rate on the x-axis and true positive rate on the y-axis, where each point corresponds to a decision threshold. A key metric of ROC analysis, the area under the curve (AUC), is threshold-independent and ranges from 0.5 to 1.0, where values closer to 1.0 indicate higher discriminative capability.

This study employed ROC curve analysis to evaluate the discriminative capability of permittivity and conductivity at each measurement frequency, as well as the five Cole-Cole fitting parameters, to identify high-performance features for subsequent machine learning discrimination.

1.3 Support Vector Machine Classification

SVM [25–28] is a machine learning method based on VC dimension theory and structural risk minimization principles from statistical learning theory. Numerous studies have reported using SVM to discriminate normal and malignant tissues [13–15]. This study employed binary classification SVM to identify an optimal hyperplane that separates normal and malignant gastric tissues, which was then used to classify test samples.

Let $X = (x_1, x_2, \dots, x_L)^T$ represent the training sample set, where L is the number of samples, $x_l \in \mathbb{R}^n$ $(l = 1, 2, \dots, L)$ denotes a sample feature vector, and $y_l \in \{-1, 1\}$ represents class labels, where $y_l = 1$ indicates malignant tissue and $y_l = -1$ indicates normal tissue. The distance between the hyperplanes $w \cdot x + b = 1$ and $w \cdot x + b = -1$, determined by support vectors, is $2/\|w\|$, and maximizing this distance is the primary objective. Additionally, to ensure the hyperplane is robust to noise, slack variables ζ_l are introduced. The problem of finding the optimal hyperplane can thus be formulated as the following constrained optimization problem:

$$\min_{w,b} \frac{1}{2}\|w\|^2 + C \sum_{l=1}^{L} \zeta_l$$
$$\text{subject to } y_l(w \cdot x_l + b) \geq 1 - \zeta_l, \quad \zeta_l \geq 0, \quad l = 1, 2, \dots, L$$

where C is the penalty factor determining the penalty for misclassified samples and affecting the classification accuracy on test samples. This constrained optimization problem is solved using Lagrangian functions [29, 30].

After comparing several common kernel functions including linear, polynomial, and radial basis function (RBF), this study selected the Gaussian RBF kernel [30], expressed as:

$$K(x_l, x) = \exp(-g\|x - x_l\|^2)$$

The penalty factor C and kernel parameter g are critical parameters requiring optimization. This study employed grid search method [31] for parameter optimization. Classification accuracy was calculated using k-fold cross-validation [32, 33].

Results

ROC curve analysis and SVM classification were implemented using SPSS 21.0 and MATLAB R2014a, respectively. This section describes the experimental results in detail.

Figures 2 [Figure 2: see original paper]A and B present ROC curves for permittivity and conductivity at all measured frequency points. The diagonal line represents an AUC of 0.5, indicating no discriminative capability. All ROC curves for permittivity and conductivity lie above the diagonal. Figure 3 [Figure 3: see original paper] shows the AUC values for all ROC curves, with AUCs for permittivity and conductivity exceeding 0.7 at all frequencies. Notably, only permittivity at five lower frequencies (42.58, 53.23, 63.87, 74.52, and 85.16 MHz) achieved AUC values above 0.8.

Figure 4 [Figure 4: see original paper] displays ROC curves for the five Cole-Cole fitting parameters. All five curves lie near the diagonal, with AUC values of 0.629, 0.671, 0.616, 0.676, and 0.544 for parameters ε_∞, $\Delta\varepsilon$, τ, α, and σ_i, respectively.

Based on these ROC analysis results, we selected permittivity at the five frequencies with AUC > 0.8 (denoted as $\varepsilon_{f1}, \varepsilon_{f2}, \varepsilon_{f3}, \varepsilon_{f4}, \varepsilon_{f5}$) and their combinations as classification features for SVM-based discrimination of normal and malignant gastric tissues. Using the combination of permittivity at all five frequencies:

Traditional grid search was performed with C and g ranging in $[2^{-10}, 2^{10}]$ at step size 0.5, requiring 14.88 s. The improved grid search method involved: (1) coarse search with C and g in $[2^{-10}, 2^{10}]$ at step size 2, yielding local optimum $(C, g) = (256, 1)$ with 5-fold cross-validation accuracy of 82.81%; (2) fine search near this optimum with C in $[2^5, 2^{10}]$ at step size 0.5 and g in $[2^{-3}, 2^3]$ at step size 0.5, yielding optimal parameters $(C, g) = (362.0387, 8)$ with 5-fold cross-validation accuracy of 84.38%. The entire process required 3.40 s. Figures 5 [Figure 5: see original paper]A and B illustrate the coarse and fine search results, respectively.

Table 1 presents SVM discrimination accuracy and MATLAB runtime for different parameter combinations. The improved grid search achieved the same accuracy as traditional grid search for each parameter combination while reducing runtime. The combination of permittivity at five frequencies achieved the highest accuracy (84.38%) without increasing MATLAB runtime compared to using individual frequency permittivity values.

Discussion

Dielectric properties of biological tissues vary with physiological and pathological states [34]. Numerous studies have reported measurements of dielectric properties for normal and malignant tissues, demonstrating significant differences [8-12, 35, 36]. However, few have analyzed the discriminative capability of permittivity, conductivity, and Cole-Cole fitting parameters across different frequencies. Recently, a limited number of studies have applied machine learning algorithms based on dielectric properties for tissue discrimination [13-15], but none have reported automatic diagnostic assistance for malignant human gastric tissues. This study measured dielectric properties of human normal and malignant gastric tissues, employed ROC curve analysis to evaluate parameter diagnostic performance, selected high-performance parameters for SVM-assisted diagnosis, and analyzed both discrimination accuracy and computational time.

We measured permittivity and conductivity at 43 frequency points across 42.58–500 MHz for normal and malignant gastric tissues, and fitted the data using the Cole-Cole model. ROC analysis revealed that AUC values for permittivity and conductivity exceeded 0.7 at all frequencies, while AUC values for all five Cole-Cole fitting parameters remained below 0.7, indicating inferior discriminative capability. Furthermore, only permittivity at five lower frequencies (42.58, 53.23, 63.87, 74.52, and 85.16 MHz) achieved AUC > 0.8, demonstrating high discriminative potential. Based on these findings, we selected the permittivity values at these five frequencies and their combination as classification features for SVM. The SVM classification achieved highest accuracy of 84.38% using the combined five-frequency permittivity feature vector, with improved grid search requiring only 3.40 s—comparable to runtime for individual parameters. Future research could focus on measuring permittivity at these high-discrimination frequencies or within a narrower bandwidth, combined with improved grid search SVM, to assist in malignant gastric tissue diagnosis while avoiding broadband measurements and Cole-Cole fitting, thereby saving time and enabling real-time intraoperative diagnostic assistance.

The final discrimination accuracy and MATLAB runtime demonstrate the feasibility of SVM-assisted diagnosis of human malignant gastric tissues based on dielectric properties as a novel approach for major disease diagnosis. However, factors including patient sex, age, tumor stage, and ex vivo measurement effects have not yet been considered, limiting the achieved accuracy. Future work will expand the sample size, incorporate these factors, and identify superior parameter combinations to further improve diagnostic accuracy and efficiency.

References

[1] Xin XG. Advances in magnetic resonance electrical properties tomography (MR EPT) for human tissues [J]. Chinese Journal of Biomedical Engineering, 2015, 34(1): 83-90.
[2] Pethig R. Dielectric properties of body tissues [J]. Clin Phys Physiol Meas,

1987, 8(4A): 5-12.
[3] Schwan HP. Electrical properties of tissues and cell suspensions: mechanisms and models [C]. Engineering Advances: New Opportunities for Biomedical Engineers. Proceedings of the 16th Annual International Conference of the IEEE. IEEE, 1994, 1(1): A70-A71.
[4] Sha L, Ward ER, Stroy B. A review of dielectric properties of normal and malignant breast tissue [C]. Southeast Con, 2002. Proceedings IEEE, 2002: 457-62.
[5] Gabriel C, Gabriel S, Corthout E. The dielectric properties of biological tissues: I. Literature survey [J]. Phys Med Biol, 1996, 41(11): 2231-49.
[6] Gabriel S, Lau R, Gabriel C. The dielectric properties of biological tissues: II. Measurements in the frequency range 10 Hz-20 GHz [J]. Phys Med Biol, 1996, 41(11): 2251-69.
[7] Gabriel S, Lau R, Gabriel C. The dielectric properties of biological tissues: III. Parametric models for the dielectric spectrum of tissues [J]. Phys Med Biol, 1996, 41(11): 2271-93.
[8] Smith SR, Foster KR, Wolf GL. Dielectric properties of VX-2 carcinoma versus normal liver tissue [J]. IEEE Trans Biomed Eng, 1986, 33(5): 522-4.
[9] Surowiec AJ, Stuchly SS, Barr JR, et al. Dielectric properties of breast carcinoma and the surrounding tissues [J]. IEEE Trans Biomed Eng, 1988, 35(4): 257-63.
[10] Joines WT, Zhang Y, Li C, et al. The measured electrical properties of normal and malignant human tissues from 50 to 900 MHz [J]. Med Phys, 1994, 21(4): 547-50.
[11] Haemmerich D, Staelin ST, Tsai JZ, et al. In vivo electrical conductivity of hepatic tumours [J]. Physiol Meas, 2003, 24(2): 251-60.
[12] Stauffer PR, Rossetto F, Prakash M, et al. Phantom and animal tissues for modelling the electrical properties of human liver [J]. Int J Hyperthermia, 2003, 19(1): 89-101.
[13] Truong BCQ, Tuan HD, Fitzgerald AJ, et al. A dielectric model of human breast tissue in Terahertz regime [J]. IEEE Trans Biomed Eng, 2015, 62(2): 699-707.
[14] Truong BCQ, Tuan HD, Wallace VP, et al. The potential of the double debye parameters to discriminate between basal cell carcinoma and normal skin [J]. IEEE Trans Terahertz Sci Technol, 2015, 5(6): 990-8.
[15] Yilmaz T, Kılıç MA, Erdoğan M, et al. Machine learning aided diagnosis of hepatic malignancies through in vivo dielectric measurements with microwaves [J]. Phys Med Biol, 2016, 61(13): 5082-97.
[16] Li Z, Deng G, Li Z, et al. A large-scale measurement of dielectric properties of normal and malignant colorectal tissues obtained from cancer surgeries at Larmor frequencies [J]. Med Phys, 2016, 43(11): 5832.
[17] Bobowski JS, Johnson T. Permittivity measurements of biological samples by an open-ended coaxial line [J]. Prog Electromagn Res B, 2012, 40: 159-83.
[18] Abdilla L, Sammut C, Mangion LZ. Dielectric properties of muscle and liver from 500 MHz-40 GHz [J]. Electromagn Biol Med, 2013, 32(2): 244-52.
[19] C. Gabriel. Compilation of the dielectric properties of body tissues at RF

and microwave frequencies [R]. King' s Coll London (United Kingdom) Dept of Physics, 1996.
[20] Peyman A, Gabriel C. Cole-Cole parameters for the dielectric properties of porcine tissues as a function of age at microwave frequencies [J]. Phys Med Biol, 2010, 55(15): N413.
[21] Martellosio A, Pasian M, Bozzi M, et al. Dielectric properties characterization from 0.5 to 50 GHz of breast cancer tissues [J]. IEEE Trans Microw Theory Tech, 2017, 65(3): 998-1011.
[22] Grosse C. A program for the fitting of Debye, Cole-Cole, Cole-Davidson, and Havriliak-Negami dispersions to dielectric data [J]. J Colloid Interface Sci, 2014, 419: 102-6.
[23] Fawcett T. ROC graphs: Notes and practical considerations for researchers [J]. Mach Learn, 2004, 31(1): 1-38.
[24] Hoo ZH, Candlish J, Teare D. What is an ROC curve? [J]. Emerg Med J, 2017, 34(6): 357-9.
[25] Cortes C, Vapnik V. Support-vector networks [J]. Mach Learn, 1995, 20(3): 273-97.
[26] Burges CJC. A tutorial on support vector machines for pattern recognition [J]. Data Min Knowl Discov, 1998, 2(2): 121-67.
[27] Chen P, Yuan L, He Y, et al. An improved SVM classifier based on double chains quantum genetic algorithm and its application in analogue circuit diagnosis [J]. Neurocomputing, 2016, 211: 202-11.
[28] Gao X, Hou J. An improved SVM integrated GS-PCA fault diagnosis approach Tennessee Eastman process [J]. Neurocomputing, 2016, 174: 906-11.
[29] Wu Y, Tang Z, Xu Y, et al. Support vector regression for measuring electromagnetic parameters of magnetic thin-film materials [J]. IEEE Trans Magn, 2007, 43(12): 4071-5.
[30] Phan AV, Le Nguyen M, Bui LT. Feature weighting and SVM parameters optimization based genetic algorithms classification problems [J]. Appl Intell, 2017, 46(2): 455-69.
[31] Li AM, Liu JY, Hao DM. Motor imagery EEG recognition method based on improved CSP algorithm [J]. Chinese Journal of Biomedical Engineering, 2009, 28(2): 161-5.
[32] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection [C]. IJCAI, 1995, 14(2): 1137-45.
[33] Galdi C, Wechsler H, Cantoni V, et al. Towards demographic categorization using gaze analysis [J]. Pattern Recognit Lett, 2016, 82: 226-31.
[34] Wang H, He Y, Yang M, et al. Dielectric properties of human liver from 10Hz to 100MHz: normal liver, hepatocellular carcinoma, hepatic fibrosis and liver hemangioma [J]. Biomed Mater Eng, 2014, 24(6): 2725-32.
[35] Wang JR, Wang HX, Xu X. Experimental study on dielectric properties of human lung tissues [J]. Chinese Journal of Biomedical Engineering, 2013, 32(2): 178-83.
[36] Hesabgar SM, Sadeghi-Naini A, Czarnota G, et al. Dielectric properties of the normal and malignant breast tissues in xenograft mice at low frequencies (100 Hz-1 MHz) [J]. Measurement, 2017, 105: 56-65.