A Cancer Detection Device Utilizing Multi-tiered Neural Networks for Improved Classification

John Shell, PhD1; William D. Gregory, PhD, PE1
1 NovaScan LLC

Abstract

We introduce a multi-tiered neural network architecture that accurately classifies malignant breast tissue from benign breast tissue.

The data were collected utilizing a NovaScan cancer detection prototype device in an approved IRB study at Aurora Medical Center, Milwaukee.

The analysis autonomously selected 4 of 6 neural networks to determine a malignant or benign classification by majority consensus. The analysis results were compared to histology results with consistent sensitivity of 100 percent and a specificity of 100 percent.

Neural network feature selection relied solely on statistical variation of histologically confirmed benign and malignant impedance data.

Data Apparatus

• The NovaScan device with laptop computer and camera.
• A tissue cassette in position atop 9 parallel 8pt electrodes with 0.25 mm spacing.
• Device outputs optically transmitted 47 samples of complex impedance data acquired logarithmically between 1 Hz and 32 MHz.

Data Collection

• 232 scans each of adipose, skin, benign, and malignant tissue were collected.
• Pathologically confirmed malignant tissue included IDC, ILC, DCIS, and LCIS.
• Depicted are the averages of 180 malignant and 180 benign impedance files displaying potential conflicts in imaginary component differentiation.

Data Analysis

• Smooth dataset with Loess nonparametric smoothing, a locally weighted polynomial regression.
• Perform a two-sided two-sample Kolmogorov-Smirnov test.
• Perform Kruskal-Wallis test, a nonparametric one-way ANOVA.
• Perform a balanced one way ANOVA.
• Depicted are the averaged 33 sample dataset showing the Cole Fc for the imaginary BE and CA components.

Neural Network Feature Set Design

• Based on statistical results use 33 frequency sample dataset for benign (BE) and malignant (CA) resulting in 4-180x33 datasets.
• Each real and imaginary component of the CA and BE dataset were divided into a training 126x33 array (70%), a validation 27x33 array (15%), and a test 27x33 array (15%).
• Feature vectors were formed by: subtracting the group mean and STD from each respective 1x33 row for CA-2', CA-2', BE-2', and BE-2' to address excursions or noisy data that could contribute to error in the NN analysis.
• A final feature was the gradient or derivative of each set to identify the rate of change of CA and BE over the entire frequency range.
• Combining the three 126x33 feature sets including the actual 126x33 raw data set to arrive at 504x33 feature set vectors for CA-2', CA-2', BE-2', and BE-2' respectively.

Neural Network Implementation

• Six backpropagation neural networks implemented to take advantage of different optimization techniques.
• Training set consisted of 504x66 concatenated imaginary components of BE and CA data.
• Target vector consisted of a 1x66 array of 33(1)…33(1).
• Symmetric hard limit chosen for network target output to apply software analysis to discern correlated BE and CA data.

Neural Network Analysis

• Training data input to each NN to optimize convergence based on parameters specific to each NN.
• The majority rule calling for 4 of the 6 networks to agree was applied to ensure optimal sensitivity and specificity.
• Validation data input after the training phase to test and allow further NN optimization.
• Introduce CA and BE data not seen by the NNs for input to trained and validated NNs.

Conclusions

• The results suggest that a sufficient amount of statistical variation in impedance data provide an excellent basis to discriminate malignant from benign tissue.
• Backpropagation neural networks can with prudent and judicious parameter optimization be highly predictive and accurate as a classifier.
• Recent successful testing of basal cell and squamous cell carcinoma data used as NN input suggests that the networks are robust, not prone to overfitting and a potential platform that can mediate different types of malignant tissue.

References