11

Classification of Breast Cancer Cases

Breast cancer has become a major cause of death among women in developed countries (Senapati et al., 2011). As the causes of breast cancer remain unknown, early detection is crucial to reduce the death rate. However, early detection requires accurate and reliable diagnosis (Cheng et al., 2010). A diagnostic tool should distinguish between benign and malignant tumors while producing low false-positive (rate of missing chances) and false-negative (rate of failure) rates.

Mammography is probably the most effective method for breast tumor detection. However, the technique has limitations in cancer detection. For example, due to its low specificity, many unnecessary biopsy operations are performed, increasing the cost, the emotional pressure, and in some cases the risk to the patient (Zainuddin and Ong, 2010).

In this chapter a wavelet network is constructed to classify breast cancer based on various attributes. Hence, a computer-aided system is developed and proposed to provide additional accuracy in the classification of benign and malignant cases of breast tumors. The Wisconsin breast cancer (WBC) data set was obtained by the UCI Machine Learning Repository and was provided by Mangasarian and Wolberg (1990). In this particular case study we are more interested in producing fewer false negatives. Whereas a false positive will result in extra cost for additional clinical tests, a false negative may result in the death of the patient.

Wavelet Neural Networks: With Applications in Financial Engineering, Chaos, and Classification, First Edition. Antonios K. Alexandridis and Achilleas D. Zapranis.
© 2014 John Wiley & Sons, Inc. Published 2014 by John Wiley & Sons, Inc.

Cheng et al. (2010) detected breast cancer based on ultrasound images. They employed a variety of classifiers, such as wavelet networks, and neural networks, and support vector machines to construct a computer-aided diagnostic system.

Similarly, Zainuddin and Ong (2010) used microarray slides as inputs to a wavelet neural network for cancer diagnosis. El-Sebakhy et al. (2006) proposed and evaluated functional networks from the WBC data set. Their results indicate that their proposed classifier is reliable and efficient.

Senapati et al. (2011) proposed a local linear wavelet neural network for breast cancer recognition. Their model was evaluated on the WBC data set and compared with methods already developed. Their results indicate that the local linear wavelet neural network performs better and has a higher level of generalization than those of common existing approaches. Methods other than artificial intelligence have been proposed for breast cancer classification: for example, linear programing by Mangasarian and Wolberg (1990) and fuzzy logic by Hassanien and Ali (2006).

In the remainder of the chapter we evaluate the classification ability of a wavelet network using two methods. In the first, the wavelet network is trained on the training sample and then evaluated out-of-sample in the validation sample. In the second, a cross-validation technique is utilized for training and forecasting evaluation of the wavelet network. Moreover, the model identification method is used to find the optimal set of input variables and the optimal structure of the wavelet network.

Data

The Wisconsin breast cancer (WBC) data set contains 699 samples. However, 16 values are missing, reducing the sample to 683 values. Each instance has one of two possible classes: benign or malignant. There are 239 (35%) malignant cases and 444 (65%) benign cases. The aim is to construct a wavelet network that classifies each clinical case accurately. The classification is based on nine attributes: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses.

PART A: CLASSIFICATION OF BREAST CANCER

In this case the data set was split into training and validation samples. The training sample consists of 478 (70%) cases. The validation sample, consisting of 205 (30%) cases, is used to evaluate the predictive and classification power of the trained wavelet network. We assume that all nine variables are statistically significant, and they will be used as predictors. Hence, the variable selection algorithm is omitted in this section.

Model Selection

To construct the wavelet network, first the optimal number of hidden units must be found. To do so, the minimum prediction risk criterion is applied and the bootstrap method will be used. From the training sample we created 50 bootstrapped samples.

The prediction risk was minimized when only 1 hidden unit was used. The prediction risk for the full model was 0.2796 and the empirical loss was 0.0862.

Initialization and Training

The BE method is used to initialize the wavelet network. A wavelet basis is constructed by scanning the first four levels of the wavelet decomposition of the data. The wavelet basis is very large and consists of 719 wavelets. However, not all wavelets in the wavelet basis contribute to the approximation of the original time series. The wavelets that contain fewer than 11 sample points of the training data in their support are removed. Seven hundred and seven wavelets that do not contribute significantly to the approximation of the original time series were identified. The truncated basis contains only 12 wavelet candidates. Applying the BE method, the wavelets are ranked in order of significance. Since only 1 hidden unit is used on the architecture of the model, only the wavelet with the highest ranking is used to initialize the wavelet network. The initialization was very good and the MSE after the initialization was only 0.1905. The time needed for the initialization was 0.1692 second. The training stopped after 135 iterations, when the minimum velocity was reached and the MSE was 0.1725. The complete training time (initialization and training) was only 0.7572 second.

Classification

In-Sample In the training sample there are 284 benign cases given the value −1 and 194 malignant cases given the value 1. The cutting score is −0.189. The classification matrix of the training sample is presented in Table 11.1. A closer inspection of the table reveals very good classification rates. More specifically, the wavelet network classified correctly 274 benign cases and 194 malignant cases. Hence, the wavelet network classified correctly 461 of 478 cases (96.44%). The specificity of the models is 96.39% and the sensitivity is 96.48%. On the other hand, the rate of failure and the rate of missing chances are very low, 3.61% and 3.52%, respectively. Finally, the fitness function is 0.5643.

Evaluation of the classification ability of the wavelet network is presented in Table 11.2. The maximum chance criterion is 59.41%, while in the heuristic method presented by Hair et al. (2010) it is 74.27%. Finally, the proportional chance criterion is 51.77%. The hit ratio is 96.44%, significantly larger than the various chance criteria.

TABLE 11.1 In-Sample Classification Matrix

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	274	10	284	96.48%	96.39%
Malignant	7	187	194	Rate of Missing Chances	Rate of Failure
Total	281	197	478	3.52%	3.61%

TABLE 11.2 Evaluation of the Classification Ability of the Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
59.41%	74.27%	51.77%	412.41	96.44%

[a]Press's Q critical values at the confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.3 Out-of-Sample Classification Matrix

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	159	1	160	99.38%	100%
Malignant	0	45	45	Rate of Missing Chances	Rate of Failure
Total	159	46	205	0.63%	0.00%

Hence, the model is predicting significantly better than chance. This is also confirmed by the large value of Press's Q statistic, which is greater than the critical values in the 0.1, 0.05, and 0.01 confidence levels.

Out-of-Sample Next, the forecasting and classification ability of the trained wavelet network is evaluated in the validation sample. The data in the validation sample were not used during the training phase. Hence, these are new data that were never presented to the wavelet network. In the validation sample there are 160 benign cases given the value -1, and 45 malignant cases given the value 1. The classification matrix of the training sample is presented in Table 11.3. A close inspection of the table reveals the very good predictive ability and classification rates. More specifically, the wavelet network classified correctly 159 benign cases and 45 malignant cases. Hence, the wavelet network classified correctly 204 of 205 cases (99.51%). The specificity of the model is 100% and the sensitivity is 99.38%. Note that in this application the rate of failure is significantly more important than the rate of missing chances. If the network classifies a benign case as malignant, it is just a false alarm; however, classifying a malignant case as benign is lethal. Our results indicate that the wavelet network has very strong classification ability since the rate of failure and the rate of missing chances are very low: 0% and 0.63%, respectively. Finally, the fitness function is 0.5964.

Evaluation of the classification ability of the wavelet network is presented in Table 11.4. The maximum chance criterion is 78.05%, while in the heuristic method presented by Hair et al. (2010) it is 97.56%. Finally, the proportional chance criterion

TABLE 11.4 Out-of-Sample Evaluation of the Classification Ability of the Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
78.05%	97.56%	65.73%	201.02	99.51%

[a]Press's Q critical values at the confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

is 65.73%. The hit ratio, 99.51%, is significantly larger than the maximum chance and the proportional chance criteria. Hence, the model is predicting significantly better than chance. This is also confirmed by the large value of Press's Q-statistic, which is greater than the critical values at the 0.1, 0.05, and 0.01 confidence levels.

PART B: CROSS-VALIDATION IN BREAST CANCER CLASSIFICATION IN WISCONSIN

In this section a different approach is followed. Instead of splitting the data into training and validation samples, cross-validation methods are used. In addition, we employ the variable selection algorithm to test if some of the attributes can be removed. More precisely, the model identification algorithm is utilized, and at each step the significant variables and the optimal structure of the wavelet network are estimated.

As mentioned earlier, cross-validation is used to assess the predictive power of the wavelet network. Hold-one-out cross-validation is used in each step of our algorithm. One training pattern of the data set will be removed from the training sample in each step. Then a wavelet network is trained on the remaining data. Finally, the trained network is evaluated on the pattern that was removed from the sample. The procedure is repeated 683 times, once from each training pattern.

Variable Selection

The target values of the wavelet network are the two possible classes. The explanatory variables are the nine attributes named earlier. To construct an accurate wavelet network classifier, the contribution of each attribute to the predictive power of the classifier must be tested. First, the significance of each attribute is examined. Hence, the initial training set consists of 9 inputs, 1 output, and 683 training samples. Again, the relevance of each attribute is quantified by the SBP criterion. Applying the variable selection proposed, the final model has only six variables, while the predictive power of the model remains almost unchanged.

Table 11.5 summarizes the results of the model identification algorithm for the WBC data. Both the model selection and variable selection algorithm are included

TABLE 11.5 Variable Selection with Backward Elimination in Wisconsin Breast Cancer Data Set[a]

Step	Variable to Remove	Variable to Enter (Lag)	Variables in Model	Hidden Units (Parameters)	n/p Ratio	Empirical Loss	Prediction Risk
—	—	—	9	1 (29)	23.6	0.0713	0.1488
1	X_9	—	8	1 (26)	26.7	0.0713	0.1485
2	X_4	—	7	3 (53)	12.9	0.0404	0.1136
3	X_3	—	6	3 (46)	14.8	0.0426	0.1135

[a]The algorithm concluded in four steps. In each step the following are presented: which variable is removed, the number of hidden units for the particular set of input variables and the parameters used in the wavelet network, the empirical loss, and the prediction risk.

TABLE 11.6 Step-by-Step Variable Selection[a]

	Full Model		Step 1		Step 2		Step 3	
Variable	SBP	p-Value	SBP	p-Value	SBP	p-Value	SBP	p-Value
1	0.0323	0.0000	0.0328	0.0000	0.0369	0.2695	0.0511	0.0000
2	0.0188	0.0000	0.0198	0.0000	0.0255	0.5068	0.0501	0.0000
3	0.0099	0.0873	0.0091	0.0000	0.0067	**0.6547**		
4	0.0025	0.0000	0.0023	**0.2166**				
5	0.0021	0.0985	0.0021	0.1995	0.0100	0.1655	0.0359	0.0000
6	0.1087	0.0000	0.1101	0.0000	0.1636	0.0000	0.01915	0.0000
7	0.0084	0.0831	0.0084	0.0945	0.0050	0.6080	0.0200	0.0000
8	0.0126	0.0000	0.0123	0.0298	0.0318	0.0000	0.0828	0.0000
9	0.0001	**0.9133**						
MAE	0.2474		0.2481		0.1473		0.2510	
MaxAE	1.6811		1.6481		1.9735		1.6857	
NMSE	0.1566		0.1569		0.0935		0.1597	
MAPE	24.73%		24.81%		14.73%		25.10%	
$\bar{R}^2$	83.67%		83.71%		89.88%		82.90%	
Empirical loss	0.0713		0.0713		0.0404		0.0426	
Prediction risk	0.1488		0.1485		0.1136		0.1135	
Iterations	149		119		11,283		264	

[a]SBP is the average SBP for each variable of 50 bootstrapped samples, the standard deviation, and the p-value. SBP, sensitivity-based pruning; MAE, mean absolute error; MaxAE, maximum absolute error; NMSE, normalized mean squared error; MSE, mean squared error; MAPE, mean absolute percentage error.

in the table. The algorithm was concluded in three steps and the final model consists of six variables only. A closer inspection of the table reveals that the empirical loss decreased from 0.0713 in the full model to 0.0426 in the reduced and simpler model. In addition, the prediction risk decreased from 0.1488 to 0.1135, indicating that the reduced model provides a better fitting to the data but also has better forecasting ability. The results of the variable significance algorithm indicate that the uniformity of cell shape, marginal adhesion, and mitoses should be removed from the input of the training sample in breast tumor classification. On the other hand, the attributes clump thickness, uniformity of cell size, single epithelial cell size, bare nuclei, bland chromatin, and normal nucleoli are statistically significant predictors.

Finally, the reduced model needed more hidden units in order to train the wavelet network. More precisely, in the full model only 1 hidden unit was used, corresponding to a 23.6 n/p ratio, while in the reduced model 3 hidden units were used, corresponding to a ratio of 14.8.

The statistics for the wavelet network model at each step are given in Table 11.6. The first part of the table reports the value of the SBP and its p-value. Various fitting criteria are also reported: MAE, MaxAE, NMSE, MAPE, $\bar{R}^2$, empirical loss, and prediction risk.

In the full model it is clear that the value of the SBP for the last variable (mitoses) is very high compared to the remaining variables. Observing the p-values, we conclude that the p-value of mitoses is 0.9133 and is greater than 0.1, strongly indicating a "not significant" variable. The wavelet network was converged after 149 iterations. In

general, a very good fit was obtained. The empirical loss is 0.0713 and the prediction risk is 0.1488. MaxAE is 1.6811, MAE is 0.2474, and NMSE is 0.1566. MAPE is 24.74%. Finally, $\bar{R}^2 = 84.34\%$.

The statistics for the wavelet network at step 1 are also presented in Table 11.6. The network had 8 inputs, one wavelet was used to construct the wavelet network, and 26 weights were adjusted during the training phase. The wavelet network converged after 119 iterations. By removing X_9 from the model, we observe from Table 11.6 that the p-value of X_5 and X_4 became 0.1995 and 0.2166, respectively. The empirical loss remained the same. However, MAE and NMSE were increased slightly to 0.2481 and 0.1569, respectively. Similarly, the remaining error criteria were increased. Next, the decision to remove X_9 is tested. The new prediction risk was reduced to 0.1485, while the explained variability adjusted for degrees of freedom increased to 83.71%. Hence, the removal of X_6 reduced the complexity of the model while its predictive power increased.

At step 2, X_4(marginal adhesion), which had the largest p-value, 0.2166, at step 1, was removed from the model. The new wavelet network had 7 inputs, 3 hidden units were used for the architecture of the wavelet network, and 53 weights were adjusted during the training phase. The wavelet network converged after 11,283 iterations. All error criteria were reduced significantly and the new $\bar{R}^2$ is 89.88%.The new prediction risk is reduced significantly, to 0.1136. Hence, removing X_4, we obtain a better fit and better potential forecasting ability. Finally, observing the p-values, we conclude that at step 2, X_3 has the higher p-value, 0.6547.

In the final step the variable X_3 (uniformity of cell shape) was removed from the model. The network had 6 inputs, three wavelets were used for the construction of the wavelet network, and 46 weights were adjusted during the training phase. The wavelet network converged after 264 iterations. The new empirical loss was increased slightly to 0.0424, compared to 0.0404 in the previous step. Similarly, all error criteria were increased slightly. However, the explained variability adjusted for degrees of freedom was reduced to 82.90%. Hence, removing X_3, a slightly poorer fit was obtained. On the other hand, the prediction risk decreased further, to 0.1135.

The p-values of the remaining variables are zero, indicating that the remaining variables are characterized as very significant variables. Hence, the algorithm stops.

Model Selection

In each step of the algorithm the optimal number of hidden units is determined by applying the model selection algorithm. The results of the model selection algorithm are presented in Table 11.7. In the full model a wavelet network with 1 hidden unit was constructed. Applying the model selection algorithm using 50 bootstrapped samples of the initial training set, the prediction risk was minimized when only 1 hidden unit was used. The prediction risk for the full model was 0.1488 and the empirical loss was 0.0713. In the second step, the prediction risk increases monotonically again as the complexity of the wavelet network increases. Hence, the prediction risk is minimized when only 1 hidden unit is used and is 0.1485. Similarly, in the third step the prediction risk is minimized when 3 hidden units are used. In the final step,

TABLE 11.7 Prediction Risk at Each Step of the Variable Selection Algorithm for the First 5 Hidden Units

	Hidden Units				
Step	1	2	3	4	5
0	**0.1488**	0.1495	0.1504	0.1520	0.1519
1	**0.1485**	0.1550	0.1624	0.1699	0.1706
2	0.1526	0.1672	**0.1136**	0.1146	0.1232
3	0.1490	0.1573	**0.1135**	0.1157	0.1151

the reduced model needed 3 hidden units and the prediction risk was 0.1135. The empirical loss was 0.0426, indicating that the reduced model provides a better fit to the data but also has better forecasting ability.

Initialization and Training

After the training set and the correct topology of the wavelet network are selected, the wavelet network can be constructed and trained. The BE method is used to initialize the wavelet network. A wavelet basis is constructed by scanning the first four levels of the wavelet decomposition of the data set.

The initial wavelet basis consists of 675 wavelets. However, not all wavelets in the wavelet basis contribute to the approximation of the original time series. The wavelets that contain fewer than eight sample points of the training data in their support are removed. The truncated basis contains 28 wavelet candidates. The MSE after the initialization was 0.173170 and the initialization needed 0.23 second to finish. The initialization is very good and the wavelet network converged after only 264 iterations. The training stopped when the minimum velocity, 10^{-5}, of the training algorithm was reached. The MSE error after the training is 0.145352, and the total amount of time needed to train the network (initialization and training) was 1.53 seconds.

Classification Power of the Full and Reduced Models

In this section the predictive and classification power of the wavelet network are evaluated. More precisely, first the full model, including all nine attributes, is tested using the leave-one-out cross-validation. Then a comparison is made against the reduced model, which uses only six attributes.

The full model is first trained using all training examples. The classification matrix of the wavelet network in-sample is presented in Table 11.8. The wavelet network accuracy in the sample is 97.66%. The sensitivity is 97.07% and the specificity is 98.74%. Also, the wavelet network classified the malignant tumors incorrectly only three times, indicating a rate of failure of only 1.26%. Finally, in Table 11.9 we observe that the wavelet network classification ability is significantly greater than chance.

TABLE 11.8 In Sample Classification Matrix of the Full Model

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	431	13	444	97.07%	98.74%
Malignant	3	236	239	Rate of Missing Chances	Rate of Failure
Total	434	249	683	2.93%	1.26%

TABLE 11.9 Evaluation of the Classification Ability of the Full Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
65%	81.25%	54.50%	620.49	97.66%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.10 Out-of-Sample Classification Matrix of the Full Model

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	431	13	444	97.07%	98.32%
Malignant	4	235	239	Rate of Missing Chances	Rate of Failure
Total	435	248	683	2.93%	1.68%

Next, the predictive power of the wavelet network is evaluated out-of-sample using the leave-one-out cross-validation method. Each time a validation sample is created that consists of only one observation, with the remaining pairs (**x**, y) used for the training of a wavelet network. In the next step, another validation sample is created and a new wavelet network is trained. The procedure is repeated until the wavelet network classifies all pairs (x, y). Table 11.10 presents the out-of-sample classification matrix of the full model. The accuracy of the full model out-of-sample is 97.51%, while the sensitivity and specificity are 97.07% and 98.32%, respectively. The misclassification of malignant cases is only 4, indicating a rate of failure of only 1.68%. Again, the very high hit ratio and the high Press's Q-statistic presented in Table 11.11 indicate that the wavelet network's classification ability is statistically significantly better than chance.

Next, the predictive power of the reduced model is evaluated in-sample and out-of-sample. The classification matrix is presented in Table 11.12. In-sample the sensitivity

TABLE 11.11 Evaluation of the Classification Ability of the Full Wavelet Network Out-of-Sample

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
65%	81.25%	54.50%	616.29	97.51%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.12 In-Sample Classification Matrix of the Reduced Model

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	431	13	444	97.07%	98.32%
Malignant	4	235	239	Rate of Missing Chances	Rate of Failure
Total	435	248	683	2.93%	1.68%

TABLE 11.13 Evaluation of the Classification Ability of the Reduced Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
65%	81.25%	54.50%	616.29	97.51%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.14 Out-of-Sample Classification Matrix of the Full Model

	Forecast				
Target			Total	Sensitivity	Specificity
Benign	431	13	444	97.07%	97.91
Malignant	5	234	239	Rate of Missing Chances	Rate of Failure
Total	436	237	683	2.93%	2.09%

and specificity of the wavelet network are 97.07% and 98.32%, respectively. The malignant cases were misclassified only four times, indicating a very small rate of failure of 1.68%. Finally, in Table 11.13 we observe that Press's Q-statistic is higher than the critical values. Finally, the hit ratio is 97.51%, so the forecasting ability of the wavelet network is significantly better than chance.

Finally, the classification matrix of the reduced model out-of-sample is presented in Table 11.14. A closer inspection of the table reveals that the sensitivity and specificity are 97.07% and 97.91%. Also, there are five misclassified malignant cases, indicating a rate of failure of only 2.09%. Finally, examining Table 11.15, we conclude that the wavelet network has better forecasting ability than chance, with a hit ratio of 97.36%. Hence, the full model outperforms the reduced model by only one correct classification.

TABLE 11.15 Evaluation of the Classification Ability of the Reduced Wavelet Network Out-of-Sample

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
65%	81.25%	54.50%	612.90	97.36%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.16 Classification Power of the Full and Reduced Models[a]

Model	HU	Accuracy	Epochs	Correct	Wrong	B/B	M/B	M/M	B/M
Full (out)	1	97.51%	146	666	17	431	4	235	13
Full (in)	1	97.66%	146	667	16	431	3	236	13
Reduced (out)	3	97.36%	265	665	18	431	5	234	13
Reduced (in)	3	97.51%	264	666	17	431	4	235	13

[a]The algorithm concluded in four steps. In each step the following are presented: which variable is removed, the number of hidden units for the particular set of input variables and the parameters used in the wavelet network, the empirical loss and the prediction risk. (in), in-sample; (out), out-of-sample using leave-one-out cross-validation; B/B, case is B/WN predicts B; B/M, case is B/WN predicts M; M/M, case is M/WN predicts M; M/B, case is M/WN predicts B.

It is clear that the accuracy of the network remains practically the same even though three classifiers were removed from the data. Hence, we can conclude that the information that comes from the uniformity of cell shape, marginal adhesion, and mitoses does not contribute significantly toward classifying breast tumors, since the additional accuracy is only 0.15%. A summary of our results of the full and reduced models is presented in Table 11.16.

Our results indicate that a wavelet network can be used successfully in breast cancer classification, providing high classification accuracy. Moreover, the accuracy of the wavelet network is higher than those presented in relevant studies (Duch and Adamczak, 1998; Hassanien and Ali, 2006; Senapati et al., 2011; Setiono and Liu, 1997; Wei and Billings, 2007).

PART C: CLASSIFICATION OF BREAST CANCER (CONTINUED)

Two different methods were used in parts A and B to build a model for classifying breast cancer. In the first part, the data set were split into two samples, the training sample and the test sample, while in the second part, the cross-validation method was used to create additional sample to train and evaluate our model.

Another difference between the two methodologies was the corresponding input set of variables. In part A we assumed that all variables were statistically significant and used as predictors, while in part B we applied the variable selection algorithm to find which of the explanatory variables are statistically significant. Our results indicate that the uniformity of cell shape, marginal adhesion, and mitoses should be removed from the input of the training sample in breast cancer classification.

In this section we again split the data into training and validation samples, as in part A, but only the statistically significant variables will be used for construction of the wavelet networks. Hence, the attributes clump thickness, uniformity of cell size, single epithelial cell size, bare nuclei, bland chromatin, and normal nucleoli are used as input variables.

Classification

In-Sample The classification matrix of the training sample is presented in Table 11.17. Close inspection of the table reveals very good classification rates. More specifically, the wavelet network classified correctly 274 benign cases and 182

TABLE 11.17 In-Sample Classification Matrix

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	274	10	284	96.48%	93.81%
Malignant	12	182	194	Rate of Missing Chances	Rate of Failure
Total	286	182	478	3.52%	6.19%

malignant cases. Hence, the wavelet network classified correctly 456 of 478 cases (95.40%). The specificity of the models is 93.81% and the sensitivity is 96.48%. On the other hand, the rate of failure and the rate of missing chances are very low, 6.19% and 3.52%, respectively. Finally, the fitness function is 0.5503. Comparing our results to those in part A, we observe that the wavelet network with the reduced set of input variables misclassified 12 malignant cases, whereas when the full set of input variables was used, the wavelet network misclassified only 7 cases.

Evaluation of the classification ability of the wavelet network is presented in Table 11.18. The maximum chance criterion is 59.41%, whereas in the heuristic method presented by Hair et al. (2010) it is 74.27%. Finally, the proportional chance criterion is 51.77%. The hit ratio, 90.40%, is significantly larger than the various chance criteria. Hence, the model is predicting significantly better than chance. This is confirmed by the large value of Press's Q-statistic, which is greater than the critical values at the 0.1, 0.05, and 0.01 confidence levels.

Out-of-Sample Next, the forecasting and classification ability of the trained wavelet network are evaluated in the validation sample. The data of the validation sample were not used during the training phase. Hence, these are new data that were never presented to the wavelet network.

In the validation sample there are 160 benign cases given the value -1 and 45 malignant cases given the value 1. The classification matrix of the training sample is presented in Table 11.19. Close inspection of the table reveals perfect prediction ability and classification rates. More specifically, the wavelet network classified correctly 160 benign cases and 45 malignant cases. Hence, the wavelet network classified all 205 cases correctly (100%). Hence, both the sensitivity and the specificity of the model are 100%, whereas the rate of failure and the rate of missing chances are 0%. Finally, the fitness function, 0.6, is the maximum value possible. Comparing our results to those of part A, we conclude that although a poorer fit was obtained to the data in-sample when the truncated set of input variables was used, the predictive

TABLE 11.18 Evaluation of the Classification Ability of the Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
59.41%	74.27%	51.77%	394.05	95.40%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

TABLE 11.19 Out-of-Sample Classification Matrix.

	Forecast				
Target	Benign	Malignant	Total	Sensitivity	Specificity
Benign	160	0	160	100%	100%
Malignant	0	45	45	Rate of Missing Chances	Rate of Failure
Total	160	45	205	0.00%	0.00%

TABLE 11.20 Out-of-Sample Evaluation of the Classification Ability of the Wavelet Network

Maximum Chance	1.25% Max. Chance	Pro	Press's Q^a	Hit Ratio
78.05%	97.56%	65.73%	205	100%

[a]Press's Q critical values at confidence levels 0.1, 0.05, and 0.01: 2.71, 3.84, 6.63.

power of the wavelet network increased. The out-of-sample results provided a perfect classification.

Evaluation of the classification ability of the wavelet network is presented in Table 11.20. The maximum chance criterion is 78.05%, whereas in the heuristic method presented by Hair et al. (2010) it is 97.56%. Finally, the proportional chance criterion is 65.73%. The hit ratio, 100%, is significantly larger than the maximum chance and the proportional chance criteria. Hence, the model is predicting significantly better than chance. This is confirmed by the large value of Press's Q statistic, which is greater than the critical values at the 0.1, 0.05, and 0.01 confidence levels.

CONCLUSIONS

Mammography is probably the most effective method for breast tumor detection. In this chapter a computer-aided system for breast cancer classification was proposed. More precisely, in this chapter a nonlinear nonparametric wavelet neural network was constructed and trained to identify and classify benign and malignant breast cancer cases correctly. The data set were obtained by the UCI Machine Learning Repository and corresponds to clinical cases of breast cancer in Wisconsin. The classification was based on nine attributes: clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single epithelial cell size, bare nuclei, bland chromatin, normal nucleoli, and mitoses.

Two modeling approaches were presented. In the first case the data were split into training and validation samples. The first set was used for the model selection training of the wavelet network; the second set was used to assess its classification power. The procedure was applied in the full set of explanatory variables and on the truncated set after the variable selection algorithm was employed. Our results indicate that the wavelet network can classify the clinical case with very high accuracy. When the full set of explanatory variables was used as an input, only one case was misclassified.

On the other hand, on the reduced input of variables the wavelet network was able to classify all cases perfectly.

In the second case, the hold-one-out cross-validation was used to create additional training and validation samples. At the same time, the model identification algorithm was used. Our results indicate that only six input variables—clump thickness, uniformity of cell size, single epithelial cell size, bare nuclei, bland chromatin, and normal nucleoli—can offer the same level of classification accuracy as the full set of input variables.

In every case the classification ability of the wavelet network was very high. The wavelet network had high generalization ability and produced robust and reliable results both in-sample and out-of-sample, indicating that wavelet networks can be accurate nonlinear nonparametric estimators for breast cancer recognition.

REFERENCES

Cheng, H. D., Shan, J., Ju, W., Guo, Y., and Zhang, L. (2010). "Automated breast cancer detection and classification using ultrasound images: a survey." *Pattern Recognition*, 43(1), 299–317.

Duch, W., and Adamczak, R. (1998). "Statistical methods for construction of neural networks." *Proceedings of 5th International Conference on Neural Inf*, 639–642.

El-Sebakhy, E. A., Faisal, K. A., Helmy, T., Azzedin, F., and Al-Suhaim, A. (2006). "Evaluation of breast cancer tumor classification with unconstrained functional networks classifier." *IEEE International Conference on Computer Systems and Applications*, 281–287.

Hair, J., Black, B., Babin, B., and Anderson, R. (2010). *Multivariate Data Analysis*, 7th ed. Prentice Hall, Upper Saddle River, NJ.

Hassanien, A., and Ali, J. (2006). "Rough set approach for classification of breast cancer mammogram images. In *Fuzzy Logic and Applications*, V. Di Gesú, F. Masulli, and A. Petrosino, eds. Springer-Verlag, Berlin, 224–231.

Mangasarian, O. L., and Wolberg, W. H. (1990). "Cancer diagnosis via linear programming." *SIAM News*, 23(5), 1–18.

Senapati, M., Mohanty, A., Dash, S., and Dash, P. (2011). "Local linear wavelet neural network for breast cancer recognition." *Neural Computing and Applications*, 1–7.

Setiono, R., and Liu, H. (1997). "NeuroLinear: from neural networks to oblique decision rules." *Neurocomputing*, 17(1), 1–24.

Wei, H.-L., and Billings, S. A. (2007). "Feature subset selection and ranking for data dimensionality reduction." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(1), 162–166.

Zainuddin, Z., and Ong, P. (2010). "Improved wavelet neural network for early cancer diagnosis using clustering algorithms." *International Journal of Information and Mathematical Sciences*, 6(1), 30–36.