Chapter 6

CONCLUSION

The main objective of this research is to gain an understanding on the process of ANN design and how it can be used in predicting the survival of breast cancer. The model of predicting survival of breast cancer was developed successfully using feed forward backpropagation neural network. This chapter summarizes the findings made with respect to predicting outcomes in the domain of breast cancer and some suggestions for extended work.

6.1   Neural Network Model

Several neural networks were created and trained using the same number of neurons, layers, epochs, performance goal and various training algorithms. A 17-15-1 network, with tansig-tansig-purelin was used. Several experiments using various training algorithms, different data pre-processing techniques, and survival intervals are done to investigate the potential of neural network as outcome predictors. The results of the experiments have been discussed in section 5.2. From the results obtained, we have shown that ANNs can be used to model breast cancer survival.

The ability of neural networks to generalise to new cases based on existing patterns is used as a basis to compute and predict the survival of individual cases. Neural networks are able to predict new cases of breast cancer, in which the survival time is yet unknown. The ability of the network to correctly classify new unseen data is important as it shows that a trained neural network has the ability to generalise and not memorize its training.

"One of N" representation is considered as the best pre-processing technique with the Trainrp algorithm and backpropagation model (see Figure 5.1, 5.2, 5.3) used as the starting point for training the survival data. However, the differences between the results obtained using different pre-processing techniques; different algorithms and also different neural network models are not significant as discussed in 5.2.1, 5.2.2 and 5.2.3. We would therefore suggest that the PCA pre-processing technique, the Elman network model and also some of the other algorithms (Trainoss, Trainscg, and Trainlm) could be used in other similar experiments.

6.2   Future Work

Future enhancements that can be brought into this research are by modelling the impact of the input variables, which are the prognostic factors of breast cancer on the prediction. The effect of the input variables on the outcome prediction could be observed by removing each variable individually from the model and examining the results thus obtained. The approach in modelling variable has been done by Abdul Kareem et al. (2001) and Harbeck et al. (2000). Genetic algorithms are a class of stochastic search algorithms based on biological evolution (Michael, 2002). Genetic algorithm is another option in determining the variables prior to training the neural network as done by Vinterbo & Ohno-Machado (1999).

The data set can be further investigated for predictions on whether or not relapse occurred within a few years or predict how long after surgery the disease will recur for breast cancer patients as done by Street (1998). The risk of breast cancer returning can occur in an individual even after surgery has been done. It would be useful to be able to

identify patients with the potential risk of recurrence, so that they are aware of their situation and can plan their future treatment.

Although most work in survival analysis has been involved with the back propagation neural networks, based on the results obtained using backpropagation and the recurrent network we would suggest either one of two networks (backpropagation or recurrent network) could be used in other similar experiments.

6.3   Contributions

Medical informatics in Malaysia is a relatively young field and the government has encouraged Malaysians to increase their participation in it. This research will hopefully start as a step towards generating more research in the medical informatics field and involving the application of neural networks in various types of medical applications. The government is also focusing on the bioinformatics field; we believe some of the methods in this research can contribute in the field of bioinformatics such as identifying or classifying the groups of genes, accuracy of prediction of drug response and others by using the neural network approach.

Finally, the results of this work will help local medical organizations in their research on accurate cancer predictions. Artificial neural networks seem to be the best model that can be used by a researcher to make predictions for the general population or by a physician who want to make prediction for an individual case. As we have shown, the percent accuracy of the predictive performance produces reasonably good results in the pre-processing techniques, neural network models and algorithms of the neural network.