Chapter 3

LITERATURE REVIEW

Neural networks have been applied to an increasing number of prediction and classification problems in recent years, and there is some interest in how neural network can be used to predict survival times. A brief outline of survival analysis, neural networks and the application of neural networks in medicine in general are given in this chapter.

3.1 Survival Analysis

Survival analysis describes the analysis of data that corresponds to the time from when an individual enter a study until the occurrence of some particular event or end-point. It is concerned with the comparison of survival curves for different combinations of risk factors and commonly uses statistical models to facilitate the comparison (Don McNeil, 1996).

In survival analysis, an individual with cancer cannot be observed for the same length of time, this is because some individual are diagnosed at the beginning of the period under study; some near the end and others may be diagnosed at any time of the study. Basically, survival data contains uncensored and censored observations. Uncensored observations involved patients who are observed until they reach the end point. Censored observations on the other hand, involved only patients who survive beyond the end or who are lost to follow-up at some point.

Right censoring is the actual survival time of an individual that exceeds the length of the study. Left censoring is its contrary, accounting for the actual survival time of an individual that is less than the predetermined observed period.

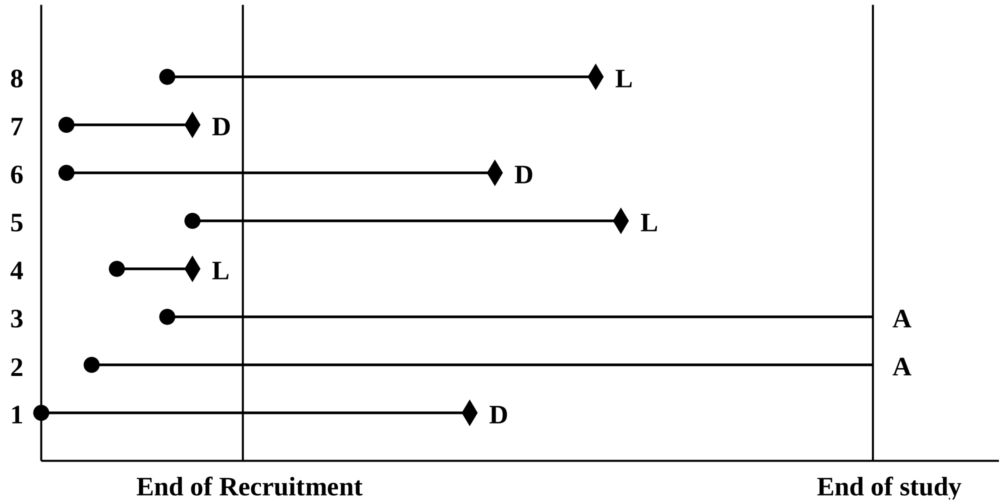


Figure 3.1: Study time for eight individuals
(*Illustration from Collet, 1994*)

Figure 3.1 illustrates the study time for eight individual in a clinical trial. The "•" is the time an individual enters the study. Individuals 1, 6 and 7 die (D) during the study, individuals 4,5 and 8 are lost to follow-up (L) while individuals 2 and 3 are still alive (A) at the end of the observation time. Individuals 2,3,4,5 and 8 are all right-censored cases.

Statistical methods are commonly used in the analysis of survival data. They can be found in a number of medical textbooks and statistical journals (Collet, 1994; David & Stanley, 1998; Lee, 1992). Statistical methods such as Kaplan-Meier methods and regression models such as the Cox Proportional Hazard are usually used to explain the data and to model the disease progression with the ability to handle censored data.

ANN is a viable alternative tool for the analysis of survival data, and recently became very popular in medical survival predictions. Neural networks gather their knowledge by detecting the pattern and relationships in the data, learning from the relationships and adapting; this knowledge is then used to predict the outcome for new combinations of data. 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.

3.2 Neural Network Background

An artificial neural network (ANN) is defined as an information processing system inspired by the structure of the human brain (Caudill & Butler, 1990). ANN gathers its knowledge by detecting a common pattern and relationships in raw data, then learning from such relationships and adapting the results as required. ANNs have been applied in business, medicine, robotics, manufacturing, industrial, as solutions to a variety of problems such as forecasting, decision making, speech recognition, classification of text, signal processing and controllers, image processing, pattern recognition, neurological and cognitive modelling.

The business community uses ANNs in predicting stock prices and currency exchange rate, market analysis, credit card fraud detection, real estate appraisal, corporate loan approval, planning and management. In the industry, ANN applications are used in product design and analysis, manufacturing process control, engine fault detection and diagnosis, visual quality inspection system and others. ANNs have also been applied in the medical field, namely in, disease diagnosis, drug reactions, medical image analysis and patients' mortality prediction.

The biological brain is a complex system with more than a hundred billions of different types of interconnected neurons. A single neuron as shown in Figure 3.2 (Simon & Devin, 2002), consists of a cell body called soma, a number of spine extensions called dendrites, and a single nerve fibre called an axon, which branches out from the soma and connects to other neurons. The axons of one cell connect to the dendrites of another via a synapse.

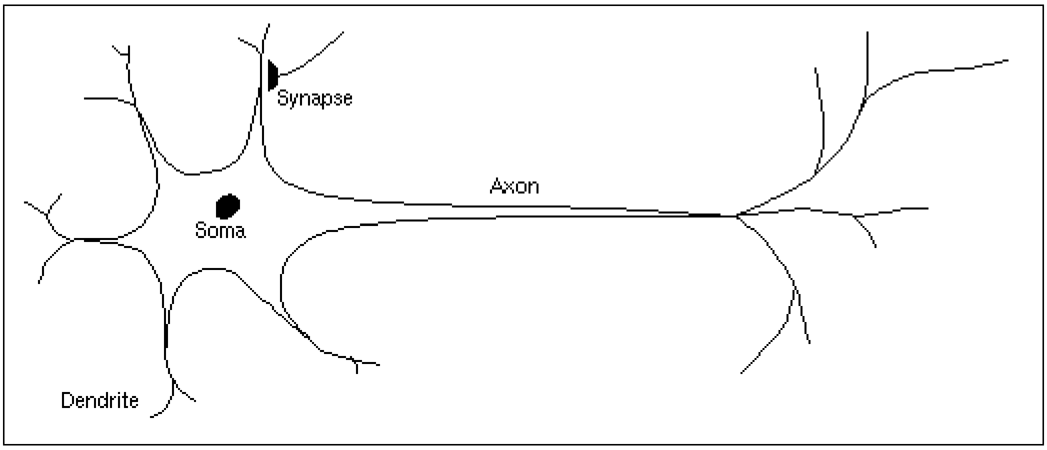


Figure 3.2: The biological neuron

3.2.1 Neural Network History

Research on ANN has been carried out for more than 50 years. The history of ANNs begins with the earliest model of a neuron given by McCulloch & Pitts (1943). This model describes a neuron as a linear threshold computing unit with multiple inputs and a single output of either 0, if the nerve cell remains inactive, or 1, if the cell fires. These ideas are now widely used in logical circuits.

The first approach to artificial neurons was the perceptron, invented by Frank Rosenblatt (1958). The perceptron was composed of processing units that transmit specific signals. These units can also adjust their interconnecting weights in order to achieve results in particular patterns, which are distinguishable from one another and ultimately learnt. The perceptron rule proved that the learning stage could be made a reality by adjusting the interconnecting weights accordingly. From the perceptron rule, Widrow & Hoff (1960) introduced the Mean Squared Error (MSE) and used it to formulate the Adaptive Linear Neuron (ADALINE), then improved it to Many Adaptive Linear Neuron (MADALINES) that could be utilised in multi layer networks.

The new generation of ANNs with several remarkably different architecture, which appeared in the early eighties, consists of elaborations of the simple perceptron rule. In addition, other ANNs have been created by integrating ideas from probability and fuzzy logic (Schalkoff, 1997).

3.2.2 Neuron

A Neuron is an information-processing unit that is fundamental to the operation of the neural network. The basic artificial neuron is shown in Figure 3.3.

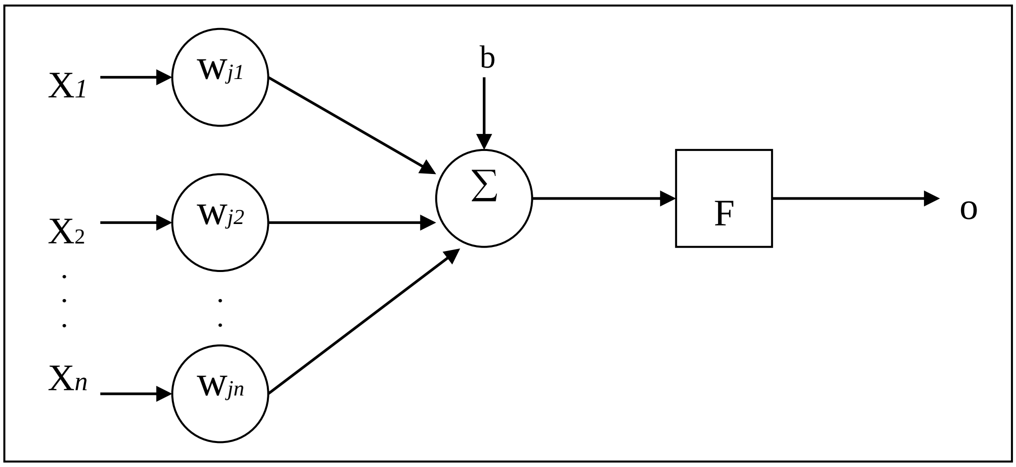


Figure 3.3: Model of a neuron

Each neuron has a certain number of inputs, each of which has a weight (w) assigned to them. These weights correspond to synaptic efficiency in biological neurons. Each neuron also has a single threshold value. The activation of the neuron is accomplished by forming the total weighted sum of the inputs and subtracting them from the threshold value of the respective neuron. The activation signal is passed through an activation function (F) to produce the output (o) of the neuron. Sometime a bias (b) is also added to the networks. The output is defined as follows:

$$o = f\left(\sum_{i} w_{ij}x_i\right) \qquad \text{or} \qquad o = f\left(\sum_{i} w_{ij}x_i + b\right)$$

3.2.3 Neural Network Architecture

The architecture of an ANN is the way the neurons are divided into layers in a network. The manner in which the neurons are structured is intimately linked with the learning algorithm used to train the network. The neural network architecture has at least an input layer and the output layer and most have one or more hidden layers between the input and output layer. The activity of the input units represents the raw information that is fed into the network. The activity of each hidden unit is determined by the activities of the input units and the weights in the connections between the input and the hidden units. The behaviour of the output units depends on the activity of the hidden units and the weights between the hidden and output units.

Individual nodes are linked together in different ways to create a neural network. In a feed-forward neural network, signals travel in one direction from input neuron to output neuron without returning or no feedback. There are two types of feed-forward architecture namely, single layer perceptron and multi layer perceptron (Haykin, 1999). A single layer perceptron consists of an input layer and an output layer only while a multi layer perceptron consists of an input layer, at least one hidden layer and an output layer. The backpropagation neural network is an example of feed-forward architecture.

Backpropagation network consists of at least three layers of units: an input layer, at least one hidden layer and an output layer. The output from the input layer is connected as an input into the hidden layer. The output from the hidden layer however is connected as an input into the output layer to produce output. Since backpropagation networks have attained good results, we have chosen this network to begin our training of the neural network. On the other hand, recurrent networks allow signals to travel in both directions by introducing a loop in the existing network. Hopfield network is an example of recurrent networks.

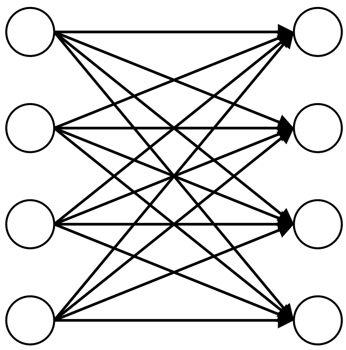


Figure 3.4 (a): Single layer perceptron

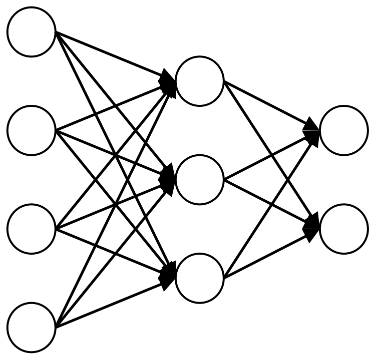


Figure 3.4 (b): Multi layer perceptron

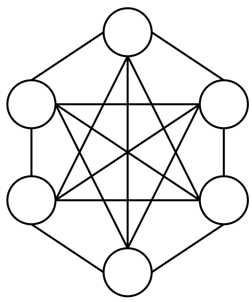


Figure 3.5: Recurrent networks

3.2.4 Activation Functions

An activation function is a cohesive value that is used to transform the neuron activation level into an output signal representative of the value itself. The behaviour of a neural network depends on both the weights and the activation function that is specified for the neuron. Some examples of common activation functions are the identity function, binary step function, binary sigmoid or logistic sigmoid and bipolar sigmoid (Fausett, 1994).

The binary step function is often used in a single layer network to convert the net input, which is a continuously valued variable, to an output node that is a binary (1 or 0) or bipolar (-1 to 1). It is defined as follows:

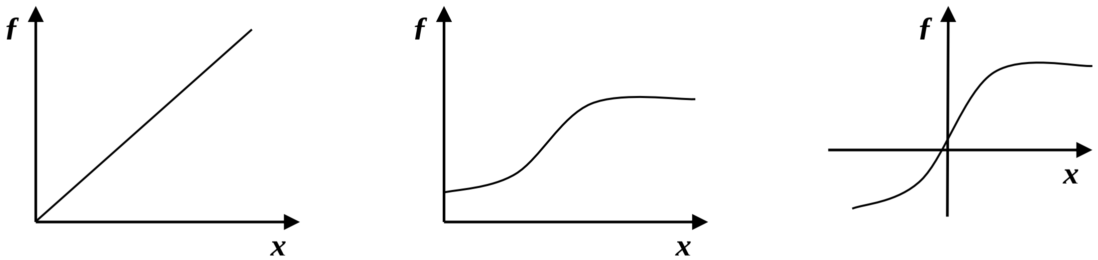
$$f(x) = \begin{cases} 1 & \text{if } x >= \theta \\ 0 & \text{if } x < \theta \end{cases}$$

Common sigmoid functions are the logistic function and hyperbolic tangent functions. A sigmoid function is often used as an activation function for neural networks in which the desired output values are binary or in the interval between 0 and 1. The binary sigmoid function is defined as follows:

$$f(x) = \frac{1}{1 + \exp(-\sigma x)}$$

The bipolar sigmoid is often used as the activation function when the desired range of output values is between –1 and 1. It is defined as follows:

$$f(x) = \frac{1 - \exp(-\sigma x)}{1 + \exp(-\sigma x)}$$



Identity Function   Binary Sigmoid   Bipolar Sigmoid

Figure 3.6: Activation Function

In each case, the x-axis is the value of the inputs and the y-axis is the output from the neurons.

3.2.5 Neural Network Learning

Neural networks have the ability to learn from its environment and improve its performance through learning. A neural network gathers its knowledge by detecting a common pattern and relationships in raw data, then learning from such relationships and adapting the results as per requirement. Coming with an apt model to reflect such patterns will enable it to assign the correct strategy to new combinations of data. The two main learning methods for neural networks are known as supervised and unsupervised learning.

Supervised learning is a process of training a neural network by giving an example of the task to learn. It provides steps to follow and concrete learning results from such supervision. From the derived algorithms, weights can be suitably adjusted and measured. The desired output is a representative of each set of the training vector. An example of supervised learning is the backpropagation algorithm, which is one of the most popular training methods used in neural network (Mc Clelland et al., 1986).

Unsupervised learning is essentially a training model where the network is only given a set of input data. The network itself must seek out shared properties of the data set, and correlate these similarities into a comprehensive algorithm as its output. This involves weight modifications without representation of the desired output for each input vector. Essentially, the network learns to recognize patterns on its own. Examples of unsupervised learning are Kohonen Self-organizing Map and Discrete Hopfield.

3.3 Application of Neural Network in Medicine

The applications of neural network in medicine can be observed through the number of research papers and journals published. Neural networks have been actively utilized in modelling, bioelectric signal processing, diagnostic and prognostics in the medical field (Konel et al., 1997). The vast potential for neural network in learning from the patterns assemble, extract the subtle irregularities and classifying patterns in their respective categories, has led to an incredible surge of usage in the medical fields (Syed, 1996). Neural networks have been applied in many medical disciplines such as cardiology, oncology, biochemistry, radiology, ophthalmology, clinical chemistry, pathology and cancer.

The applications of neural network have been applied in modelling blood pressure and volume (Allen & Murray, 1999), modelling and treating speech and hearing impairments (Guenther, 1995), classification of cornea topography (Maeda et al., 1995) and modelling the human cardiovascular system (Dimitrios, 2002). In the diagnostic fields, neural network applications have been applied in tumour diagnosis (Lixing et al., 1998), cardiopulmonary diagnosis (Tourassi et al., 1995), diagnosing heart attacks (Furlong et al., 1991) and cancer diagnosis.

Some of the most influential and interesting neural network applications seem to be in signal processing such as segmenting prostate ultrasound images (Prater & Richard, 1992) and electro-encephalography (EEG) analysis (LCE, 1997). In the prognostics area, neural networks have been used in predicting the risk of coronary artery disease (Lapuerta et al., 1995) and survival predictions of acquired immune deficiency syndrome (AIDS) patients (Ohno-Machado & Musen, 1997). Most of the areas that applied neural networks produced excellent, reliable results.

3.3.1 Application of Neural Network in Cancer

ANNs have been used in cancer as an aid to identification, prediction and diagnosis. A literature search in the National Library of Medicine database (PUBMED) of "neural network in cancer" yields 413 citations.

Lung cancer is the number one disease most feared by people from around the world. Detection at its early stage is the key to its cure. Zhi-Hua et al. (2002) used ANN to identify whether the lung cell from the images of the specimens of needle biopsies obtained from patients is benign or malignant. The cancer cell is further classified into four different types of lung cancer.

Burke et al. (1997) compares the prediction accuracy of TNM staging system with ANN models on breast cancer patients. They used backpropagation training with the number of patients' variables representing an input node while the number of hidden layer nodes ranged from three to five, besides inclusive of one output node.

Backpropagation neural network also has been applied in skin cancer to diagnose normal tissues, benign tumour and melanoma (Lixing et al., 1998).

In medical research, ANNs have become increasingly popular as a prognostic tool, that is, the prediction of survival and treatment outcome. It is a good alternative for the prediction of survival of individual patients and it offers no obstacle to handling censored data. The description on censored data can be found in the next section. In our research we used neural network to predict the survival of individuals with breast cancer.

3.3.2 Neural Network on Survival Prediction

Neural network technology has recently risen in stature in the field of medical applications. It is an alternative method for prediction of survival of individual patients and it offers no obstacle to handling censored data. The following are some of the successful work done using neural network applications in prognosis and outcome prediction.

Ohno-Machado et al. (1994) compared the prediction of the survival of AIDS patients using hierarchical neural network models and non-hierarchical models. The hierarchical neural network model predicted survival more accurately than the non-hierarchical model. Ohno-Machado and Musen (1996) also used backpropagation neural network to predict development of coronary heart disease.

Marc Theeuwen et al. (1995) used a neural network to predict the treatment outcome in patients of ovarian cancer. They trained ovarian cancer patients' data for each year of the study to classify the patients into alive and dead by reducing the set of prognostic factors until they obtain a maximal predictive value and thereby drew up a cohesive model of survival probability.

Abdul-Kareem et al. (2001) used ANN technology to predict the prognosis of nasopharyngeal carcinoma (NPC). Several neural networks were created and trained using various training algorithms. Several experiments were carried out using different data pre-processing techniques, different prognostic factors and different architectures. Part of the work done was also in using statistical methods and comparing the results obtained between the statistical method and the neural network.

3.3.3 Neural Network on Survival Prediction in the domain of Breast Cancer.

Breast cancer is the most common of all cancers amongst women; it is second after lung cancer in causing fatality among women. Many researches have been done in survival prediction using neural network in breast cancer. These are namely, accuracy of survival prediction, relapse-free survival after surgery and probability of disease free survival (Ravdin et al. 1992; Street 1996; Ruth M Ripley 1998; Harbeck et al. 2000). The following are some of the work done using neural network on survival predictions in the domain of breast cancer.

Ravdin et al. (1992) used ANN to generate survival curves, which plotted the probability of disease free survival against time. The censored survival data from a group of patients with node-positive breast cancer was trained, tested and validated using time as an input variable and approximation of recurrence probability as an output.

Street (1998) applied ANN to predict how long after surgery the disease recurred for breast cancer patients by using standard back propagation neural network coupled with a hidden layer and ten units of output layers as the year's recurrence after surgery. Three predictive models were created: the ability to separate good and poor prognoses, generate an individual disease-free survival curve, and the accuracy of predicted recurrent rates.

Harbeck et al. (2000) used ANN to model relapse-free survival of follow up data in primary breast cancer patients by comparing the output of high-risk and low-risk group using different input variables in the neural networks which incorporates time as a function of risk factors.

3.4 Summary

The structure of the human brain inspired the creation of artificial neural networks. They mimic the brain's mechanism, becoming information processing systems capable of logic learning from any raw data acquired. Neural networks provide solutions to a variety of problems in many areas such as business, industry and medicine. Neural network contains input, output and hidden layers with specific weights assigned to them. Neural networks are increasingly popular as an aid to identification, prediction and diagnosis on a variety of medical ailments including cancer. Several researchers applied neural networks as prognostic tools on a variety of diseases. In this section, we presented the ANN concept and how the ANN applications are applied as a tool to model survival data and to predict the prognosis of cancer. 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 cancer cases.