EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

ORAYA INTEM

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF ENGINEERING
(BIOMEDICAL ENGINEERING)
FACULTY OF GRADUATE STUDIES
MAHIDOL UNIVERSITY
2010

Thesis
entitled

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

……..……..……

Miss Oraya Intem
Candidate

……..……..……

Lect. Panrasee Ritthiphavat,
D.Eng. (Mechanical Engineering)
Major-advisor

……..……..……

Asst.Prof. Thongchai Bhongmakapat,
M.D. (Otolaryngology)
Co-advisor

……..……..……

Prof. Banchong Mahaisavariya,
M.D.,Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……..……..……

Asst.Prof.Jackrit Suthakorn,
Ph.D. (Robotics)
Chair
Master of Engineering Programme in
Biomedical Engineering
Faculty of Engineering

Thesis
entitled

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

was submitted to the Faculty of Graduate Studies, Mahidol University
for the degree of Master of Engineering (Biomedical Engineering)
on
June 11, 2010

……
Miss Oraya Intem
Candidate

……
Lect.Yodchanan Wongsawat,
Ph.D.(Electrical Engineering)
Chair

……
Lect.Panrasee Ritthipravat,
D.Eng.(Mechanical Engineering)
Member

……
Lect. Thavida Maneewarn,
Ph.D.(Electrical Engineering)
Member

……
Asst.Prof. Thongchai Bhongmakapat,
M.D.(Otolaryngology)
Member

……
Prof. Banchong Mahaisavariya,
M.D.,Dip Thai Board of Orthopedics
Dean
Faculty of Graduate Studies
Mahidol University

……
Asst.Prof. Rawin Raviwongse,
Ph.D. (Engineering Management)
Dean
Faculty of Engineering
Mahidol University

ACKNOWLEDGEMENTS

I would like to thank my major advisor, Dr.Panrasee Ritthipravat. She always gives me a kindness, academic assistance, valuable guidance, encouragement and advice throughout my study. I would like to thank my co-advisor, Associate Professor Thongchai Bhongmakapat, for his time and effort in providing the author with his continual guidance and valuable recommendation which make me achieving this success.

I also would like to express my sincere gratitude to the committee members: Dr. Yodchanan Wongsawat and Dr. Thavida Maneewarn for their comments and kind suggestions.

Furthermore, I would like to thank all of the service staffs at Biomedical Engineering Department, Faculty of Engineering, Mahidol University for their kindness supports. I am particularly indebted to Ramathibodi Hospital for their co-operation and generous assistance.

I would like to sincerely thank to my family for all kinds of their support and entirely care which made this thesis possible and enabled me to undertake this thesis successfully. Additionally, I also wish to thank all of my friends for their encouragement.

Finally, I would like to thank the funding from TRFMAG Window II for providing partial financial resources.

Oraya Intem

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

ORAYA INTEM 4936227 EGBE/M

M.Eng. (BIOMEDICAL ENGINEERING)

THESIS ADVISORY COMMITTEE: PANRASEE RITTHIPRAVAT, D.Eng., THONGCHAI BHONGMAKAPAT, M.D.

ABSTRACT

This thesis aims to study various censored data techniques for effective prediction of nasopharyngeal carcinoma (NPC) recurrence. Clinical data and time to recurrence of NPC patients were collected from Ramathibodi Hospital, Thailand. Recurrence factors were then selected by univariate and multivariate analysis. The results showed that only 9 factors related to the NPC recurrence. They were N stage, smoking, alcohol, family history of cancer, neck fibrosis, IgG, IgA, radiation dose, and KPS. These factors were used in predictive model development. In the study, three ANN based censored data techniques are mainly investigated. They are Street, PLANN, and our proposed technique. The results showed that our proposed technique provided the highest predictive performances compared with the other two techniques and Cox regression model. All ANN based techniques outperformed the Cox model. Hosmer-Lemeshow goodness-of-fit test was then applied. The results showed that the chi-square statistic for all models was less than 15.51. This means that every model fitted well.

Survival curves for each predictive model were then generated and compared. The results showed that the curve of the Cox model was obviously different from the others. This is confirmed by the log-rank test in which only the Cox model is significantly different from the Kaplan-Meier model. With the proposed technique, four main problems existing in the previous censored data techniques can be handled. The problems include scalability, generation of a non-monotonic survival curve, specification of unknown recurrence status, and data replication problems.

KEY WORDS: SURVIVAL ANALYSIS/ NPC/ PREDICTION/ NEURAL NETWORK

45 pages

การนำข้อมูลขาดหายมาใช้ในการทำนายการกลับมาเป็นซ้ำของผู้ป่วยโรคมะเร็งช่องคอหลังโพรงจมูกอย่างมีประสิทธิภาพ

EFFECTIVE USE OF CENSORED DATA FOR NPC RECURRENCE PREDICTION

อรยา อินเต็ม 4936227 EGBE/M

วศ.ม. (วิศวกรรมชีวการแพทย์)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์ : ปัณรสี ฤทธิประวัติ, D.Eng., ธงชัย พงศ์มฆพัฒน์, M.D.

บทคัดย่อ

งานวิจัยนี้ได้ศึกษาเทคนิคในการแก้ปัญหาข้อมูลขาดหายเพื่อใช้ในการทำนายการกลับมาเป็นซ้ำของผู้ป่วยโรคมะเร็งช่องคอหลังโพรงจมูก ซึ่งข้อมูลทางคลินิกและเวลาถูกเก็บรวบรวมมาจากโรงพยาบาลรามาธิบดี ประเทศไทย โดยตัวแปรเหล่านี้จะถูกเลือกให้เหลือเพียงตัวแปรที่สำคัญด้วยวิธีการวิเคราะห์ตัวแปรเดี่ยว (Univariate analysis) และการวิเคราะห์ตัวแปรเชิงพหุ (Multivariate analysis) ผลจากการวิเคราะห์ทำให้เหลือตัวแปรที่สำคัญ 9 ตัวได้แก่ N stage, ประวัติการสูบบุหรี่, ประวัติการดื่มสุรา, ประวัติการเป็นมะเร็งในครอบครัว, การปรากฏของพังผืดบริเวณคอหลังการฉายรังสี, IgG, IgA, ปริมาณรังสีรักษาและ KPS งานวิจัยนี้ได้ทำการทดลองเปรียบเทียบประสิทธิภาพการพยากรณ์ของโมเดลพยากรณ์ด้วยเครือข่ายระบบประสาทเทียมสามรูปแบบ ได้แก่ โมเดลพยากรณ์ของ Street โมเดลพยากรณ์ของ PLANN และโมเดลการพยากรณ์ที่เสนอขึ้นมาใหม่ ซึ่งผลการศึกษาพบว่าโมเดลพยากรณ์ที่เสนอขึ้นมาใหม่ให้ประสิทธิภาพการทำนายที่ดีกว่าโมเดลพยากรณ์อื่นๆ รวมทั้งดีกว่าการใช้โมเดล Cox regression ซึ่งเป็นโมเดลทางสถิติ และในการทดลองยังพบว่าการพยากรณ์ของเครือข่ายระบบประสาทเทียมทั้งสามรูปแบบมีประสิทธิภาพสูงกว่าโมเดล Cox regression เมื่อทดสอบความสมบูรณ์ของโมเดลด้วยค่าไคสแควร์จาก Hosmer-Lemeshow พบว่าทุกโมเดลให้ค่าไคสแควร์น้อยกว่า 15.51 ซึ่งแปลว่าทุกโมเดลมีความเหมาะสม

ในการทดสอบประสิทธิภาพของโมเดลเชิงกลุ่มโดยการพล๊อตกราฟการรอดชีวิตที่ได้จากแต่ละวิธี พบว่ากราฟของ Cox โมเดลให้การทำนายที่แตกต่างจากกราฟการรอดชีวิตของข้อมูลจริง และเมื่อทดสอบความแตกต่างด้วย Log-rank พบว่า Cox โมเดลให้การทำนายเชิงกลุ่มที่แตกต่างจาก Kaplan-Meier อย่างมีนัยสำคัญทางสถิติ จากงานวิจัยนี้พบว่าเทคนิคที่นำเสนอสามารถแก้ไข 4 ปัญหาที่มักเกิดขึ้นในวิธีการแก้ปัญหาข้อมูลขาดหายในอดีต โดยปัญหาทั้ง 4 ได้แก่ปัญหาในการ Scalable ของโมเดล, ปัญหาเส้นโค้งการรอดชีวิตที่ไม่ลดลงตามเวลา, ปัญหาข้อมูลขาดหายและปัญหาการ Replication ของข้อมูล

45 หน้า

CONTENTS

CONTENTS (cont.)

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATION

T	Survival time
t	Specific time
$S(t)$	Survival function
$h(t)$	Hazard function
$H(t)$	Cumulative hazard function
$h_0(t)$	Baseline hazard function
β_i	Coefficients
X_i	Set of covariates
x_i^k	The input variables
i, o, k, h	Iteration number
y_o^k	Output
z_h	The induced local fields of hidden nodes
z_o	The induced local fields of output nodes
w_{ih}	Weights between input to hidden layer
w_{ho}	Weights between hidden to output layer
ϕ	The transfer function
η	Learning rate
α	Momentum
λ	Regularization parameter
H	Approximated inverse Hessian matrix

CHAPTER I
INTRODUCTION

1.1 Problem Statement

Nasopharyngeal carcinoma (NPC) is frequently discovered in Thailand, China, Hong Kong, and Taiwan. Though it is in the hidden location; nasopharynx, NPC is curable when detected early. After a treatment, each patient must be followed up regularly because it can redevelop. Early detection of the recurring cancer can reduce patient mortality and costs. For low-risk patients, however, frequent examination may be excessive and increase unnecessary expenses. Prediction of NPC recurrence for each patient is thus required in which follow-up time can be set appropriately. In addition, hospital resources, time and cost can be effectively managed.

Survival analysis has been extensively used to predict the presence of redeveloping cancer. In the analysis, Kaplan-Meier and Cox proportionally hazard model are mainly employed [1]-[2]. The Kaplan-Meier technique can only predict the recurring cancer in a group manner. This is different from the Cox model in which individual prediction can perform. However, for the Cox regression analysis, strong assumption with the proportional hazard ratio exists. ANN based prediction technique is proposed to overcome those limitations [3]-[4]. It has been shown that time to cancer recurrence was efficiently predicted for each patient [4]. In the prediction, relevant clinical data and recurring time must be collected. Due to various reasons, after a treatment, some patients may withdraw from the follow-up before the end of the study. Some of the other patients may not have redeveloping cancer within the study period. For those cases, the recurring time cannot be correctly specified. When the missing information is time to an event of interest, the missing data is called the censored data. In general, there are 3 types of censored data, i.e. types I – III [5]. Type I censoring arises when the cancer does not redevelop during the study period. For type II censoring, it occurs when patients withdraw from the follow-up before the

study ends. When the collected data set contains both type I and type II censoring observations, the missing data are called type III censoring. Though ANN based predictive model is effective, censored data cannot be directly utilized.

ANN based on censored data technique can be categorized into 3 types. There are single-point [6], multiple-point [7] and time-coded models [4,8], as explained in chapter II. For the single-point model, a neural network which predicts the presence or absence of cancer recurrence within a specific time point is generated. For prediction of several time points, multiple models must be used. Scalability problem thus arises. This is different from the multiple-point model in which the number of outputs is equal to the number of time points of interest. Type II censoring data cannot be used in this model because the status of cancer recurrence after the censoring time cannot be correctly specified. Street [7] assigns survival probabilities from the Kaplan-Meier model to the unknown statuses. Though previously proposed models provide efficient prediction, generated survival probabilities are not monotonically decreasing with time. Monotonically decreasing survival curve is necessary in the survival analysis because it truly represents the actual survival phenomena. For the time-coded model, it is purposed by Ravdin [4]. Time is used as an additional input of the neural network. A single output represents survival status at a given time (0 is being alive, 1 is death). Since time is used as an input, each data is replicated until the most recent follow-up time. This technique suffers from the size of the training data which will grow enormously when the number of records is large. Overtraining problem may arise. Biganzoli [8] proposes PLANN model which is based on Ravdin's technique. Hazard rate is used as the target in order to guarantee monotonically decreasing survival curve. However, the data replication problem still exists.

From the previous studies, censoring data techniques encounter with four main problems, i.e., scalability problem, unknown statuses for type II censoring data when used with the multiple-point model, non-monotonic survival curve, and data replication problem. This thesis introduces a new censored data technique that can handle those problems simultaneously. The proposed model is based on the multiple-point model. Hazard rate is used as a target in order to provide monotonically decreasing survival curve as in the PLANN model.

1.2 Thesis Objectives

1.2.1 To investigate performances of previously proposed censored data in tackling the following problems: (1) Monotonic survival curve, (2) replication problem, (3) type II censoring data, and (4) scalability.

1.2.2 To develop a new censored data technique that can handle all above problems simultaneously.

1.2.3 To compare performances of the proposed technique with the previous ones in prediction of nasopharyngeal carcinoma recurrence problem.

1.3 Thesis Scope

This thesis will examine performances of previously proposed censored data techniques in prediction of nasopharyngeal carcinoma recurrence. A new technique that can handle 4 important issues is proposed. The issues compose of (1) monotonic survival curve, (2) replication problem, (3) type II censoring data, and (4) scalable problem.

1.4 Expected Results

Outcome of this thesis is to gain a new censored data technique that can tackle 4 problems existing in the previously proposed techniques. The problems include scalability problem, type II censoring data, non-monotonic survival curve, and replication problem. The predictive model generated from the proposed technique can provide monotonic survival curve and use type II censoring data. It can scale well in a more complicated problem and do not possess replication problem.

CHAPTER II
LITERATURE REVIEWS

2.1 Nasopharyngeal Carcinoma

Nasopharyngeal carcinoma (NPC) is one of head and neck cancers frequently discovered in Southeast Asia including China, Hong Kong, and Thailand etc. In 2005, it is in the sixth place of common male cancers and approximately 5% of all cancers found in Thailand [9].

NPC arises when the epithelium cells of the nasopharynx are out-of-control or grow in the abnormal way. Normally, NPC is difficult to early detect because it appears in the hidden location – nasopharynx. Anatomy of nasopharynx [10] is illustrated in Figure 2.1. The nasopharyngeal cavity is a cuboidal structure covered by mucociliary columnar epithelium. The posterior and superior borders are formed by bone structure of the basiocciput, basisphenoid, and the first two cervical vertebrae. The inferior and anterior boundaries are the upper surface of soft palate and the posterior choanae respectively. The lateral walls are connected with Torus tubarii and Rosenmuller's fossa which is the most common NPC site.

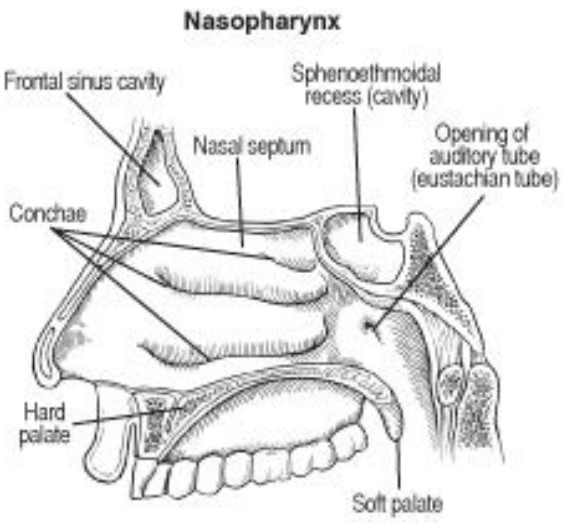

Figure 2.1 Anatomy of nasopharynx

Diagnosis of this cancer is based on clinical examination and histological confirmation. Both computed tomography (CT) and magnetic resonance imaging (MRI) are generally used to detect local and regional extension of the tumor which is important in the cancer staging. MRI provides better information about extension and intracranial involvement than CT scan. To investigate the bone invasion, however, CT scan is more suitable [10,11].

Nasopharynx is difficult to access. In addition, it is close to blood vessels and nerves. Surgery is thus rarely performed. Radiotherapy and chemotherapy are normally applied. Because this cancer can recur, after a treatment, the patient must be followed up for checking the cancer recurrence continually.

2.2 Prognostic Factors of NPC Development

From the literature survey, the following factors have been reported that they related to NPC development.

1. **Age** [12,13]: In general, NPC patients can be categorized into two different groups according to their incident rate. The first group is the bimodal age distribution. It is found in countries that have low to medium incident rates. The age distribution is composed of two peaks belonging to late childhood (aged 10-20 years) and patients aged between 55 and 65 years. The second group is the plateau age distribution. It is found in countries that have high incident rates such as China [9].
2. **Sex**: NPC is discovered in male more than female with the ratio of 2-3:1[13].
3. **Cancer genetics**: People with a family history of NPC have higher risk of cancer development [9].
4. **Nationality**: Chinese people who live in Southeast Asia and Eskimo have higher risk of NPC [9].
5. **Epstein-Barr virus (EBV)**: Nasopharyngeal cancer cells usually contain EBV. It is thus one of the suspicious recurrence factors [9, 13].

6. **TNM staging**: Many studies reported that TNM stage associated with nasopharyngeal carcinoma recurrence [12 to 15]. According to AJCC system, TNM stage can be defined as follows.

   T Stage:

   T1 Tumor confined to the nasopharynx

   T2 Tumor extends to soft tissue of oropharynx and/or nasal fossa

   T2a Without parapharyngeal extension

   T2b With parapharyngeal extension

   T3 Tumor invades bone structures and/or paranasal sinuses

   T4 Tumor with intracranial extension and/or involvement of the cranial nerves, infra- temporal fossa, hypopharynx, orbit

   N Stage:

   NX Regional lymph nodes cannot be assessed

   N0 No regional lymph nodes metastasis

   N1 Unilateral metastasis in lymph node (s), 6 cm or less in greatest dimension, above the clavicular fossa

   N2 Bilateral metastasis in lymph none (s), 6 cm or less in greatest dimension, above the clavicular fossa

   N3 Metastasis in a lymph node (s)

   N3a Greater than 6 cm in dimension

   N3d In the supraclavicular fossa

   M Stage:

   MX Present of distant metastasis cannot be assessed

   M0 No distant metastasis

   M1 Distant metastasis present

7. **Cancer Stage**s: NPC stages can be determined from tumor size (T), regional lymph node involvement (N), and distant metastasis (M). Many research projects found that the cancer stages related to developing of NPC [12,13,15]. In this thesis, AJCC system will be used for NPC staging. With

this system, there exist 4 cancer stages, i.e., stage I to stage IV. Patients with the stage IV have the highest severity. Each stage can be determined from

Stage 0: Tis N0 M0

Stage I: T1 N0 Mo

Stage II: T2 N0 M0

Stage III: T3 N0, T1-3 N1, M0

Stage IV: T4 N0-1 M0 or Any T N2-3 M0 or Any T any N M1

8. **Histological cell type**: From Thongchai [9], there are 3 cell types of NPC. They are well- differentiated, poor-differentiated, and undifferentiated carcinoma. From the review of Hwang J et al [16], undifferentiated carcinoma type provided higher recurrence rate than the others within 5-year study period.
9. **Duration time**: Duration in days, between date that the first symptom is observed and date that a patient is firstly diagnosed with NPC, related to NPC development [13].
10. **Treatment type**: with or without chemotherapy also related to NPC recurrence [12].
11. **Dose of radiation**: Dose of radiotherapy measured in Gy is a recurrence factor [12].
12. **IgG and IgA quantity**: Serum VCA/IgA titers before treatment is the prognostic factors of NPC development [12].
13. **Hemoglobin quantity**: HUA Yijun et al [12] reviewed that hemoglobin concentration before and after treatments (g/L) were also prognostic factors of NPC development.
14. **Other suspicious factors from specialist physicians**: Alcohol, smoking, neck fibrosis, and Karnofsky Performance Scale index (KPS) are suspicious factors of NPC recurrence.

2.3 Censored and Uncensored Data

Clinical data and time to recurrence are used for generating a predictive model. When these data are collected, censored data may be found. The censored data is one type of incomplete data in that the time to an event of interest is not known exactly. Here, the event is NPC recurrence. Therefore, the censored data focusing in this thesis are the data that their time to recurrence cannot be specified correctly. This may arise from various reasons, such as withdrawing from a treatment, death, or no cancer recurrence within the study period. On the contrary, if the recurring time can be measured, such data are called the “uncensored data”. Since the time to recurrence is not known precisely, the censored data cannot be used directly in the prediction particularly when machine learning approaches are used for developing predictive models. Removing all censored data from the analysis may reduce the data size drastically. The residual data may be inadequate for providing accurate predictive models. Moreover, some deleted censored data may contain relevant information, excluding them from the analysis can cause biased results. Effective use of censored data is important in the prediction of cancer recurrence in which it can improve the prediction performances.

In general, there are 3 types of censored data, i.e., type I – type III [5]. Type I censoring arises when patients do not have redeveloping cancer within the study period. For type II censoring, it occurs when the patients withdraw from the follow-up before the study ends. When the collected data contain both type I and type II censoring observations, all censored data are alternatively called type III censoring. Examples of censored data and uncensored data can be shown in Figure 2.2.

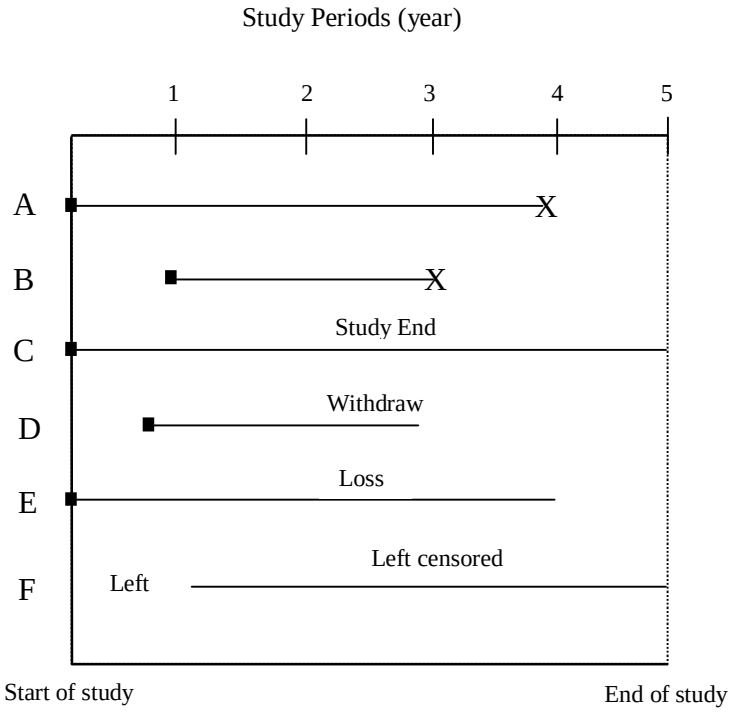

Figure 2.2 Types of censored data

From the figure, there are six NPC patients. The study period is set at 5 years. The time of entry to the study is presented by ' ■' . The 'X' symbol represents cancer recurrence status. Within this study periods, patients A and B suffer with cancer recurrence. Their recurrence times are 4 and 2 years respectively. These records are examples of uncensored data. Patients C, D, and E are censored data with the censoring times of 5, 2.5, and 4 years respectively. Record of patients C is type I censoring while type II censoring data arise with patients D and E. These types of censored data are examples of *right censored*. Right censored data is defined as the data that exact survival time becomes incomplete at the right side of the study period. *Left censored data*, on the contrary, can occur when the survival time becomes incomplete at the left side. It may arise from the patient is referred from another hospital. His exact entry time point cannot be known. In Figure 2.2, patient F is an example of left censored data. This thesis will only focus on right censored data.

2.4 Survival Analysis: Statistical Approach

Survival analysis is a statistical technique in which time to an event of interest, so called survival time, is the output of the analysis. The time can be in terms of years, months, weeks, or days. For the event, it can be the death, relapse of a disease etc. In survival analysis, any event of interest is usually called "failure". In this thesis, the failure is NPC recurrence.

Prior to describe the survival analysis, related terminology is presented in the next subsection.

2.4.1 Survival Function

Survival function, $S(t)$: It may be called as survival probability or survivorship function. It is the probability that an individual survives longer than a specific time *t* and can be represented as

$$S(t) = P(\textit{an individual survives longer than } t)$$

$$S(t) = P(T > t) \qquad (2.1)$$

where *T* denoted the survival time.

Theoretically, $S(t)$ is a monotonically decreasing function over time *t*. At $t = 0$, the probability of survival is assumed to be 1. The probability approaches to zero when the time goes infinity. If there are no censored observations, the survival probability is estimated from the proportion of patients surviving longer than *t* [17].

$$\hat{S}(t) = \frac{\textit{number of patients surviving longer than } t}{\textit{total number of patients}} \qquad (2.2)$$

An example of $\hat{S}(t)$ calculation is presented in Example 1.

In addition, $\hat{S}(t)$ can be calculated by the nonparametric approach, called Kaplan-Meier method. It can be used regardless the presence of censored data.

2.4.2 Kaplan-Meier method

For this method, the survival probability can be determined from

$$\hat{S}(k) = p_1 \times p_2 \times p_3 \times \ldots \times p_t \qquad (2.3)$$

where p_{t_j} is the proportion of patients surviving at least time t_j. The survival probability calculated from this method involves the product of terms p_{t_j}.

Example 2.1 shows calculation of the survival probability using Kaplan-Meier model.

Example 2.1 Calculation of survival probability using Kaplan-Meier model

In this example, two groups of NPC patients who receive different treatments are used in the study. The first group is treated by chemotherapy while the second group uses a combination of chemotherapy and radiotherapy. Survival time of each patient is presented in Table 2.1.

Table 2.1 Survival times for two groups of NPC patients

Group1 Chemotherapy	Group2 Chemotherapy and radiotherapy
1,1,1,1,1,1.5,1.5,2,2,2,2.5,2.7,2.7,2,7, 2.7,3.4,3.4,3.4,3.4,3.6,3.6	1, 1, 1, 1.5, 2, 2.5, 2.7, 3.4, 3.6, 1^+, 1.6^+, 2^+, 2.2^+, 2.8^+, 2.9^+, 3.1^+, 3.7^+, 4.1^+, 4.1^+, 4.5^+, 4.6^+

Note: + represents censored case

From Table 2.1, there is no censored data in the group 1. This is different from group 2 in which twelve patients withdraw or they are lost to follow before the end of study. Table 2.2 shows the survival probability determined from the Kaplan Meier model. In Table 2.2, t presents the survival times in ascending order. n is the number of relapse-free patients before time t. m is the number of patients who have cancer recurrence at time t. q is the number of patients who withdraw from the study at time t. $\hat{S}(t)$ presents survival probability at the time t.

Table 2.2 The example to calculation the survival function using Kaplan-Meier.

t	Group 1				Group 2			
	n^{G1}	m^{G1}	q^{G1}	$\hat{S}(t)^{G1}$	n^{G2}	m^{G2}	q^{G2}	$\hat{S}(t)^{G2}$
0	21	0	0	1	21	0	0	1
1	21	5	0	0.762	21	3	1	0.857
1.5	16	2	0	0.667	17	1	1	0.807
2	14	3	0	0.524	15	1	2	0.753
2.5	11	1	0	0.476	12	1	0	0.690
2.7	10	4	0	0.286	11	1	3	0.627
3.4	6	4	0	0.095	7	1	0	0.538
3.6	2	2	0	0.000	6	1	5	0.448

As seen in Table 2.2, the survival probability of group 1 at time t = 1.5 years is determined from Eq. (2.3) as $\hat{S}(1.5)^{G1} = 1 \times \frac{16}{21} \times \frac{14}{16} = 0.667$. It is the product of the proportion terms. For group 2, the survival probability is $\hat{S}(1.5)^{G2} = 1 \times \frac{18}{21} \times \frac{16}{17} = 0.807$ at time t = 1.5 years. For both groups, the proportion of recurrence-free patients at time t = 0 is 1. This means every patient is relapse-free at the beginning of the study. However, the proportion of recurrence-free patients at least time t = 1 is different for both group. For group 1, there exist 5 relapsed patients from 21 patients who do not have cancer recurrence before this time. Therefore, the proportion of recurrence-free patients at least time t = 1 is $\frac{16}{21}$. For group 2, the proportion is $\frac{18}{21}$ because there are only 3 relapsed patients. At this time, 1 patient withdraws from the study. Therefore, after time t = 1, only 17 patients exists in the study. At time t = 1.5, there is 1 relapsed patient. The proportion of recurrence-free patients at time t = 1.5 is thus equal to $\frac{16}{17}$. After determining survival probability, survival curve can be generated as shown in Figure 2.3.

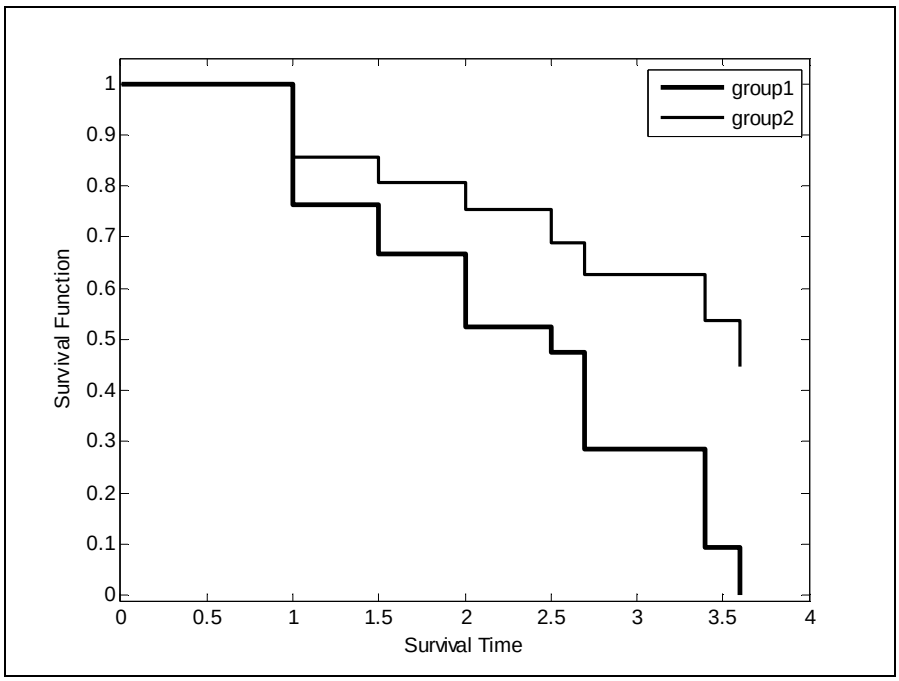

Figure 2.3 Survival curves for groups 1 and 2.

As seen in the figure, group 2 has higher recurrence-free probabilities. Therefore, combination of treatments can reduce cancer recurrence for this example.

2.4.3 Hazard Function

Hazard function, $h(t)$: Hazard function or hazard rate [17] is the conditional probability that the failure occurs within the interval $(t, t+\Delta t)$ given that it does not occur before time *t* [17].

$$h(t) = \lim_{\Delta t \to 0} \frac{P\{\text{an individual fails within the time interval}(t, t+\Delta t)\}}{\Delta t} \quad (2.4)$$

Equation (2.4) presents a hazard function which can be estimated from

$$\hat{h}(t) = \frac{\text{the number of patients failing in the interval begining at time t}}{(\text{number of patients surviving at t})(\text{interval width})} \quad (2.5)$$

An example of hazard function calculation can be presented in Table 2.3. In this example, the data from example 1 are employed to estimate the hazard function. The estimated hazard is plotted as shown in Figure 2.4. In general Hazard function can also be derived from the survival function as given by

$$S(t) = \prod_{i=0}^{t-1} [1 - h(t-i)] \qquad (2.6)$$

Table 2.3 An example of hazard function calculation

	Group 1				Group 2			
t	n^{G1}	m^{G1}	q^{G1}	$\hat{h}(t)^{G1}$	n^{G2}	m^{G2}	q^{G2}	$\hat{h}(t)^{G2}$
0	21	0	0	0.000	21	0	0	0.000
1	21	5	0	0.238	21	3	1	0.143
1.5	16	2	0	0.125	17	1	1	0.059
2	14	3	0	0.214	15	1	2	0.067
2.5	11	1	0	0.091	12	1	0	0.083
2.7	10	4	0	0.400	11	1	3	0.091
3.4	6	4	0	0.667	7	1	0	0.143
3.6	2	2	0	1.000	6	1	5	0.167

The estimated hazard functions, $\hat{h}(t)$, of both groups are computed from Eq. (2.5). For example, at the time 1.5 years, the hazard function is $\hat{h}(1.5) = \frac{1}{17 \times 1} = 0.059$ for group 2. The corresponding survival function is estimated from Eq.(2.6) as $\hat{S}(1.5) = (1-0) \times (1-0.143) \times (1-0.059) = 0.807$. After the hazard functions are determined. They can be plotted as shown in Figure 2.4.

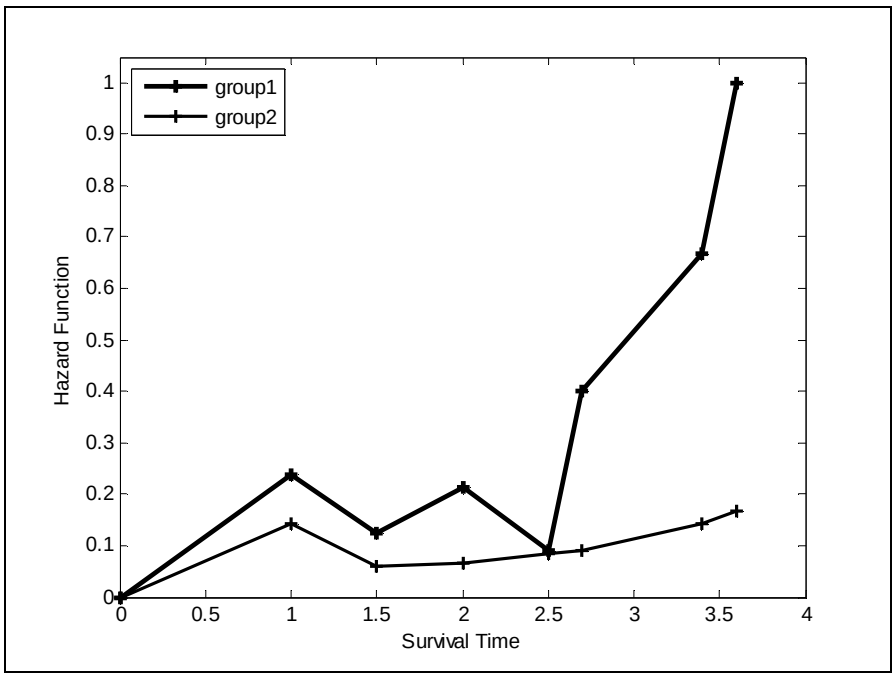

Figure 2.4 The hazard functions for groups1 and 2

From Figure 2.4, patients who received only chemotherapy have higher hazard rates than patients who were treated with both chemotherapy and radiotherapy.

2.4.4 Cumulative Hazard

Cumulative hazard, $H(t)$: It is the cumulative sum of hazard rate and it is a probability of failure at time *t* given that the patient survives until time *t* [17]. Cumulative hazard can be determined from

$$H(t) = \int_0^t h(x)dx \quad . \tag{2.7}$$

$H(t)$, $h(t)$ can be determined from $S(t)$ by

$$H(t) = -\log_e S(t) \tag{2.8}$$

$$h(t) = -\frac{d}{dt}\log_e S(t) \tag{2.9}$$

2.4.5 Cox Proportional Hazard Regression Model

Normally, Kaplan-Meier and Cox proportional hazard models are employed to estimate survival function. KM model describes the overall survival function for a group of patients. Cox model, on the contrary, can predict survival probability for an individual patient. In the Cox model, survival function is estimated based on a set of given covariates, $X \in [X_1, \ldots, X_2]$.

The hazard function of an individual given a set of covariates X at a specific time *t*, $h(t, X)$ can be approximated from

$$h(t, X) = h_0(t) e^{\sum_{i=1}^{p} \beta_i X_i} \qquad (2.10)$$

where $h_0(t)$ is the baseline hazard function. β_i is the i^{th} coefficient and X_i is the i^{th} covariate. Though Cox model is widely used in survival analysis, it has strong assumption with the proportional hazard ratio.

2.5 Artificial Neural Network (ANN)

ANN is a computational model that simulates human brain's function. It can approximate a function that expresses relation of training input-output pairs through a learning process. In this thesis, multilayer perceptron (MLP) which is one of artificial neural networks is focused. The architecture of MLP with single hidden layer and multiple outputs is presented in Figure 2.5.

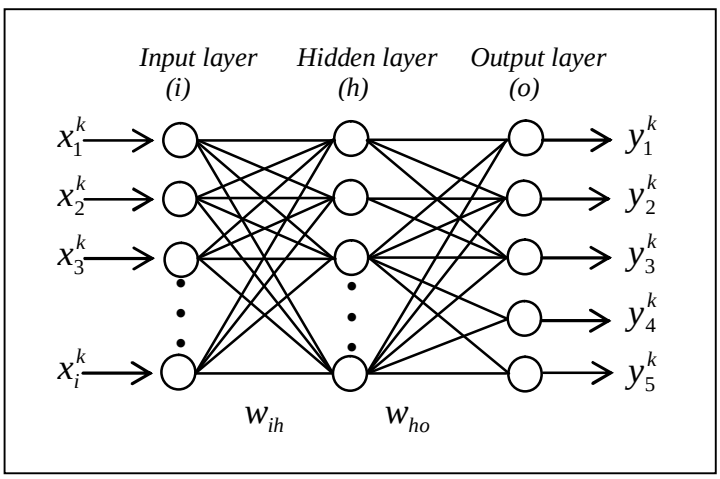

Figure 2.5 Multilayer perceptron with single hidden layer

From Figure 2.5, ANN model consists of three layers, i.e., input-layer, hidden-layer, and output layer. x_i^k is the i^{th} input variable of patient *k*, and y_o^k is the corresponding output. The neural networks model will estimate the output by learning process that consists of two processes, forward and backward passes.

For the forward pass, the neuronal signals are computed from:

$$z_h = \sum_{i=0}^{I} w_{ih} x_i^k \qquad (2.11)$$

$$z_o = \sum_{h=0}^{H} w_{ho} \phi(z_h) \qquad (2.12)$$

where z_h and z_o are the induced local fields of hidden nodes and output nodes respectively. w_{ih} and w_{ho} represent weights between input to hidden layer and hidden to output layer respectively. *h* is the number of hidden nodes and *o* is the number of output nodes. ϕ is the transfer function. In this thesis, the sigmoid function is used as the transfer function for both hidden and output layers. It can be presented as in Equation 2.13.

$$\phi(u) = \frac{1}{1+e^{(-u)}} \quad . \qquad (2.13)$$

The neuronal output $\hat{y}_o^k$ can be estimated from

$$\hat{y}_o^k = \phi(z_o) \quad . \qquad (2.14)$$

For the backward pass, the quadratic function error is formed as given by

$$E = \frac{1}{2} \sum_{k=1}^{K} \sum_{o=1}^{O} (\hat{y}_o^k - y_o^k)^2 \quad . \qquad (2.15)$$

Weight updating equations are derived from the error gradient as presented in Equations 2.16 to 2.19. The weight updating equations are expressed as in Equations 2.20 to 2.21.

$$\frac{\partial E}{\partial w_{ho}} = \frac{\partial E}{\partial \phi(z_o)} \frac{\partial \phi(z_o)}{\partial z_o} \frac{\partial z_o}{\partial w_{ho}} \tag{2.16}$$

$$\frac{\partial E}{\partial w_{ho}} = -(y_o^k - \phi(z_o))\phi(z_o)(1-\phi(z_o))\phi(z_h) \tag{2.17}$$

$$\frac{\partial E}{\partial w_{ih}} = \frac{\partial E}{\partial \phi(z_h)} \frac{\partial \phi(z_h)}{\partial z_h} \frac{\partial z_h}{\partial w_{ih}} \tag{2.18}$$

$$\frac{\partial E}{\partial w_{ih}} = -(y_o^k - \phi(z_o))\phi(z_o)(1-\phi(z_o))w_{ih}\phi(z_h)(1-\phi(z_h))x_i \tag{2.19}$$

$$w_{ih}^{n+1} = w_{ih}^n - \eta(\frac{\partial E}{\partial w_{ih}}) + \alpha \Delta w_{ih} \tag{2.20}$$

$$w_{ho}^{n+1} = w_{ho}^n - \eta(\frac{\partial E}{\partial w_{ho}}) + \alpha \Delta w_{ho} \tag{2.21}$$

where η and α are the learning rate and momentum term respectively.

The applications of ANN techniques for predictive model are summarized as follow.

ANN has been used in classification, regression or prediction problems. For cancer management, Cruz et al. [18] described that ANN could be used for predictions of cancer susceptibility, cancer recurrence, and cancer survivability.

M.D. Laurentiis et al. [19] simulated dataset which consisted of complex variables, and censored data. They predicted outcome of these dependent variables using three different ANN models, i.e., time coded model, single-time point model, and multiple time-coding model. The results were compared with that of Cox and logistic regression models. ANN provided better prediction than the statistical methods in terms of ROC and global chi square. The time-coded model was superior to the other techniques.

J.M. Jerez et al. [20] compared the predictive performances between ANN and Cox model in the prediction of breast cancer relapse. ANN selected relevant prognostic factors which correlated with that of Cox model. The results showed that ANN with time as an additional input could provide better prediction than the Cox model.

Kareem et al. [13] compared two types of ANNs, i.e., multilayer perceptron and recurrent network in survival prediction of NPC. The prognostic variables included age, sex, race, dialect, date of firstly observed symptoms, cancer type, biopsy, diagnosis, symptoms, tumor extent, nerve involvement, distant metastasis, WHO type, TNM classification, and cancer stage. The performances of both models were not significantly different.

Baker et al. [21] reviewed soft computing techniques for cancer prognosis. The techniques include ANN, Genetic Algorithms and Fuzzy logic. ANN had been used for prediction in breast, ovary, bladder, and prostate cancers

H. Yijun et al. [12] used support vector machine (SVM) technique to predict 5 year survival status of NPC. Twenty five variables initially were selected for the prediction. The relevant factors were extracted from the use of logistic regression model. The selected factors were then used in the predictive model development. The performances were compared with the model with full factors. Their performances were not different in terms of accuracy, sensitivity, and specificity.

R.N.G. Naguib et al. [22] utilize the Radial Basis Function (RBF) neural network to predict the recurrence of oesophago-gastric junction cancer at 12, 18 and 24 months. Two predictive models were developed and compared. Input variables of first model consisted of pre-operative data. For the second model, its input variables composed of both pre- and post- operative data. The results showed that performances of both models were not different. The paper concluded that using only pre-operative information could generate reliable predictive model.

Jones et al. [23] studied survival prediction of laryngeal squamous carcinoma. ANN, Cox, and Kaplan-Meier were used in the predictive model development. Each input variable was separated into low risk and high risk groups and exploited in the prediction. Kaplan-Meier curves were plotted to represent survival probability. ANN could generate better predictive model than Cox model when age and N stage were separately used as the input of the model. In addition, ANN could efficiently handle a complex interaction existing between input variables.

M. Theeuwen et al. [24] used Boltzman perceptron to analyze survival of ovarian cancer. Predictive performances were compared with that of the Cox model. ANN performed slightly better than another model.

R. Mofidi et al. [25] compared ANN with Union International Contra Cancrum (UICC) TNM classification in the prediction of 1 and 3 years survival for oesophagus- and oesophago-gastric junction (OG junction) cancer. ANN was significantly better than TNM classification for both 1 and 3 years predictive models. They presented that ANN had become a valuable tool in oesophargeal carcinoma management.

2.6 Censored Data Techniques: ANN based approaches

Censored data can be directly used in survival analysis based on statistical approaches. However, they suffer from various limitations. Kaplan-Meier model can provide prediction only in a group manner. Cox regression model, on the other hand, can predict individually. However, linear relationship among prognostic factors is assumed. Moreover, proportional hazard assumption exists. Though ANN based approaches do not suffer from those assumptions, censored data cannot be used directly. Some modifications are required. From the previous research, various ANN based approaches for survival analysis had been proposed. They are

2.6.1 Single point model

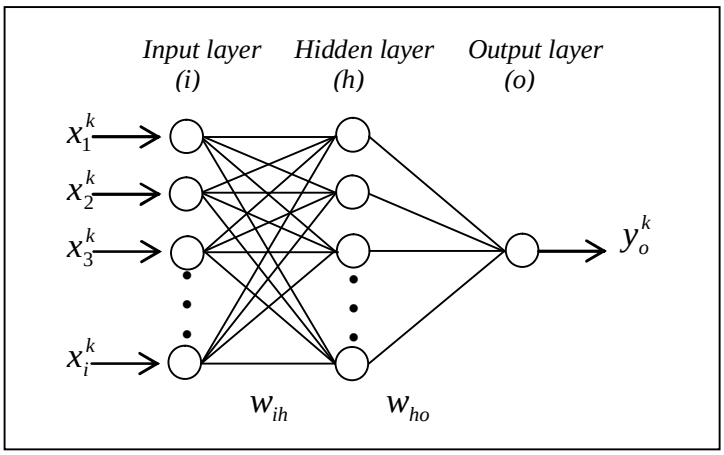

Figure 2.6 Single-point model

For the single-point model, a neural network which predicts the presence or absence of cancer recurrence within a specific time point is generated. For

prediction of several time points, multiple models must be used. Scalability problem thus arises. Censored data could be directly used in the single point model.

Mofidi R. el al [25] used the single-point model for cancer survival prediction at 1 and 3 years. The model can use all types of censored data. However, multiple models were created for prediction in various time points. The technique thus suffered from multiple model generation.

Kumdee et al [26] studied the prediction of nasopharyngeal carcinoma recurrence using five single-point models for 5-year prediction. Type II censoring data were excluded for their analysis. Though deletion method was simple, biased analysis might be arisen.

2.6.2 Multiple-point model

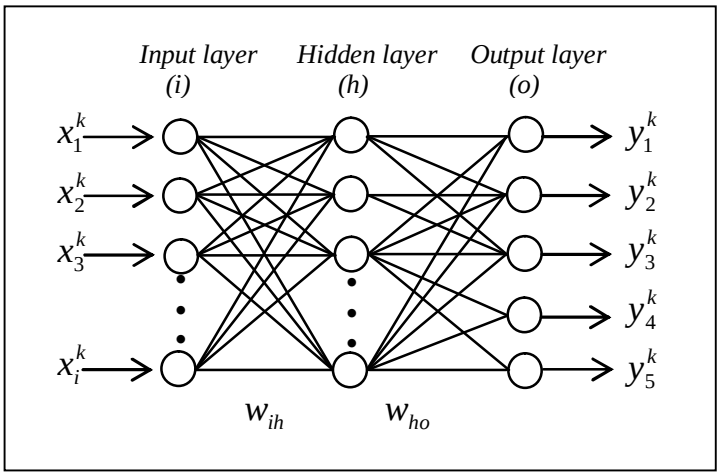

Figure 2.7 Multiple-point model

For multiple-point model, the number of outputs is equal to the number of time points of interest as shown in Figure 2.7. Type II censoring data cannot be used in this model directly because the status of cancer recurrence after the censoring time cannot be correctly specified. Imputation technique can be used for estimating unknown targets. However, the survival probability generated by this model may not be monotonically decreasing function.

2.6.3 Time-coded model

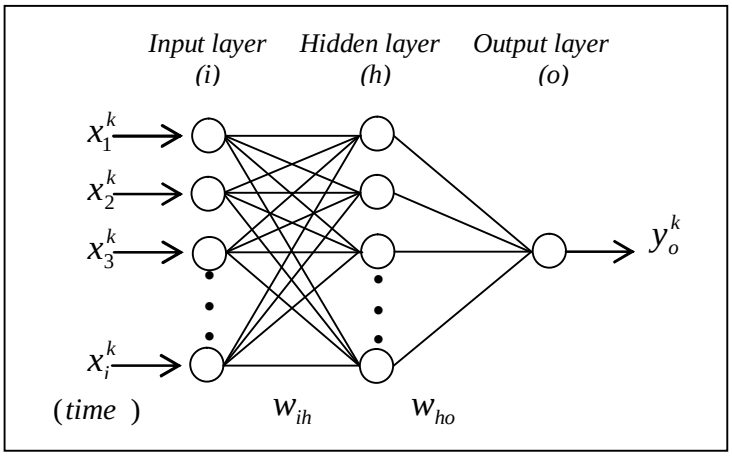

Figure 2.8 Time-coded model

This model is purposed by Ravdin [4]. Time is used as one of the input variables. A single output represents survival status at a specified time point (0 is alive, 1 is death). Since time is used as an input, survival status at a time point must be provided. For example, in 5-year survival analysis, input of type I censoring observation is in the form of [data, time]. Its target is set as 0 for each time. For type II censoring records, the data are reproduced and fed into the network until their censoring time. This technique suffers from the size of a training data which will grow enormously when the number of records is large. Overtraining problem may arise. In addition, non-monotonic survival function may be generated.

2.6.4 PLANN

Biganzoli [8] propose PLANN technique. It is based on time-coded model. For this technique, time is used as an additional input to an artificial neural network. Hazard rate is used as the target in order to guarantee monotonically decreasing survival curve. A single output represents conditional failure probability ranging from 0 and 1. Initially the study period is divided into several time intervals such as [1 2 3 4 5] for 5-year study period. Inputs to the network compose of all prognostic factors and a time interval. The output is a probability that an individual will have cancer recurrence at a given time, conditioned on disease-free up to that time. As an example, there are i-1 prognostic factors, $[x_1^k,\ldots,x_{i-1}^k]$. Inputs vectors for 5-year study period are

$[x_1^k,\ldots,x_{i-1}^k,1]$, $[x_1^k,\ldots,x_{i-1}^k,2]$,…, $[x_1^k,\ldots,x_{i-1}^k,5]$. Targets of these inputs vectors are 0, 0, 0, 0, 0 respectively when the observation is type I censoring. For a patient having the redeveloping cancer at 2.5 years, its input vectors are $[x_1^k,\ldots,x_{i-1}^k,1]$,$[x_1^k,\ldots,x_{i-1}^k,2]$, $[x_1^k,\ldots,x_{i-1}^k,3]$. The corresponding targets are 0, 0, and 1 respectively. For a type II censoring data, the input vectors and their targets are given until the censoring time. PLANN model can be shown in Figure 2.8.

From the figure, the model composes of 3 layers, i.e., input, hidden, and output layers. Logistic function is used as an activation function for both hidden and output layers. The cost function being minimized E^* is formed from cross-entropy error function E_c including weight decay term as shown in Equation 2.22.

$$E^* = E_c + \lambda \sum w^2 \quad , \tag{2.22}$$

where the cross entropy is defined from

$$E_c = -\sum_{k=1}^{K}\sum_{l=1}^{l_k}\left\{ d_{kl} \log \hat{y}_o^k(x_i^k, w) + (1-d_{kl}) \log[1-\hat{y}_o^k(x_i^k, w)] \right\} \tag{2.23}$$

where K represents the number of training records before data replication, l_k is the number of time intervals. d_{kl} is target of the k[th] patient at an l_{th} time interval. λ is the regularization parameter ranging between 0.01 - 0.1 [8]. w can be either weight between input and hidden layers, w_{ih}, or weight between hidden and output layers, w_{ho}.

Quasi-Newton algorithm which is the second order optimization method is used for training. Weight update equations are derived from

$$w_{ih}^{n+1} = w_{ih}^n - \eta H^n \left(\frac{\partial E^*}{\partial w_{ih}}\right) \tag{2.24}$$

$$w_{ho}^{n+1} = w_{ho}^n - \eta H^n \left(\frac{\partial E^*}{\partial w_{ho}}\right) \tag{2.25}$$

where η is the learning rate and H represents the approximated inverse Hessian matrix. Since output of the PLANN model represents the hazard rate, survival

probabilities generated are always monotonically decreasing with time, as shown in Equation 2.6. However, the data replication problem still exists. In addition, overtraining problem may arise. Therefore, generalization capability may be lacked.

2.6.5 Street

Street [7] proposed a technique based on the multiple time-point model as shown in Figure 2.7. Inputs to the model are all relevant prognostic factors. The outputs ranging from 0 and 1 represent probabilities of disease-free at different time points. Multilayer perceptron with backpropagation training is used to provide mapping between inputs and outputs. The cost function is formed from quadratic error function. Momentum term is added to enhance predictive performance.

For uncensored data, targets can be simply set. For example, the target vector of a patient having cancer recurrence at 3.5 years from 5-year study period is [1 1 1 0 0]. For one who does not have the redeveloping cancer within the study period, the target is set as [1 1 1 1 1]. For type II censoring data, actual recurrence-free status after the censoring time cannot be correctly provided. In this case, survival probabilities from the Kaplan-Meier model are used in the imputation. As an example, a target vector of a patient who withdraws from the follow-up at 3.5 years is [1 1 1 0.8 0.7] where 0.8 and 0.7 represent the survival probabilities from the Kaplan-Meier model. Though this model can utilize all censored data in the analysis, survival probabilities generated may not be monotonically decreasing with time.

2.6.6 Other Censored Data Techniques

Faraggi et al. [27] used outputs of neural networks as coefficients of Cox proportional hazard model. This method had advantages over the classical proportional hazard model. Though it provided monotonically decreasing survival curve, and could handle interaction among variables, proportional hazard assumption existed.

Lapuerta et al. [28] developed multiple neural networks to predict the survival probability at various time points. Single model served the prediction at a time point. For type II censoring data, unknown target after the censoring time was estimated from output of the neural network. The technique was superior to the Cox

proportional hazard model in terms of accuracy. However, it could not provide monotonically decreasing survival curve. Since many neural networks were generated, scalability property was deteriorated.

From the previous research, there are many censored data techniques. Their characteristics and abilities can be summarized in Table 2.4. In general, there are four main problems that the censored data techniques may encounter. They are scalability problem in the single point model, non-monotonic survival curve generation problem, unknown recurrence status specification problem in the multiple-point model, data replication problem in the time-coded model. This thesis proposes new censored data technique that can handle all problems simultaneously. The proposed technique combines Street and Biganzoli methods. Multiple-point model is used. Hazard rate is used as a target for unknown recurrence status after the censoring time. This can provide monotonically decreasing survival curve.

Table 2.4 Characteristics of censored data techniques: ANN based methods

Methods	Multiple NN	Single Output	Censored type II	Replication problem	Monotonic survival curve	Scalable property
Mofidi R.	Y	Y	Y	N	N	N
Street.	N	N	Y	N	N	Y
Ravdin.	N	Y	Y	Y	N	N
Biganzoli	N	N	Y	N	Y	Y
Lapuerta.	Y	N	Y	N	N	N
Faraggi.	N	Y	Y	Y	Y	Y

Note : N=No, Y=Yes

CHAPTER III
METHODOLOGY

In this chapter, research methodology is summarized. It can be presented in Figure 3.1.

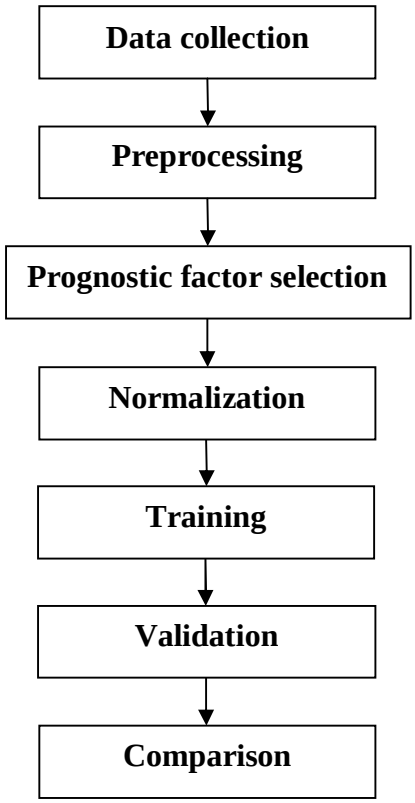

Figure 3.1 Research methodology

3.1 Data Collection

Clinical data and time to recurrence of NPC patients are collected from Ramathibodi Hospital, Thailand. This research has been approved by the ethics committee of Mahidol University. This research is an extended study from

Ritthipravat [29] in which several prognostic factors are added. The factors analyzed in this thesis include 1) patient's age, 2) duration from the date that the first symptom is observed to the date that the patient is diagnosed as having NPC, 3) IgG quantity, 4) IgA quantity, 5) dose of radiation, 6) patient's sex, 7-9) TNM staging, 10) cancer stage, 11) cell type, 12) present of chemotherapy, 13) presence of neck fibrosis, 14) alcohol history, 15) smoking history, 16) hemoglobin quantity before a treatment, 17) family history, and 18) Karnofsky Performance Scale index (KPS). The factors 1 to 13 are gained from Ritthipravat [29]. The additional factors obtain from specialist doctor suggestion are in 14 to 18.

3.2 Preprocessing

3.2.1 Data Representation

The qualitative data, such as sex (male, female), tumor stages (stage1 to stage 4), smoking history (yes, no) are converted into numerical value.

3.2.2 Missing Value

Missing data are imputed by expectation maximization (EM) technique. Kumdee et al [26] used the EM imputation technique for completing NPC data. EM imputation technique provided highest predictive performance when compared with mean imputation, K-nearest neighbor and deletion techniques.

3.3 Prognostic Factor Selection

The prognostic factors are selected by univariate and multivariate analysis [30].

3.3.1 Univariate Analysis

The univariate analysis is employed to select related recurrence factors. Its analysis independently considers each variable in the data set. The categorical variables are tested by Kaplan-Meier survival curve and the p-value from log-rank test. Cox proportional hazard model and the p-value from partial likelihood ratio test

are used for continuous variable. In the analysis, the factors that have p-value less than 0.25 are considered to relate with NPC recurrence.

3.3.2 Multivariate Analysis

The multivariate analysis analyzes relevant recurrence factors by controlling other factors. Cox proportional hazard model and the purposeful selection of covariates with backward selection technique are used in the analysis. This method is composed of 3 steps. Initially, all factors are considered in the Cox proportional hazard model. The factor that has p-value above 0.05 from the Wald test is removed from the model. The p-value from the partial likelihood ratio test indicates whether the removed factor is relevant to the model. If the p-value is above 0.05, that factor can be removed. Before discarding the factor, coefficients of new model should not be changed more than 20%. Otherwise, the factor cannot be removed.

3.4 Normalization

The normalization is to scale each variable into the similar range in order to reduce the biasing problem. For ANN models developed in this thesis, all input variables are normalized by the use of standardization technique.

3.5 Predictive Models

Three predictive models based on multilayer perceptrons with backpropagation training are mainly investigated. They are presented as follows.

3.5.1 Street Model [7]

Street's technique is based on the multiple-point model. For type II censoring data, unknown recurrence statuses are imputed by survival probabilities from the Kaplan-Meier model. Quadratic error is the cost function used for deriving weight updating equations. Momentum term is added to enhance the predictive performances. Though this model provides efficient prediction, generated survival probabilities are not monotonically decreasing with time.

3.5.2 PLANN Model [8]

PLANN technique is based on the time-coded model. Hazard rate is used as the target in order to guarantee monotonically decreasing of the survival curve. For this technique, cross-entropy error adding with the weight decay term is used as the cost function. Quasi-Newton algorithm is used for deriving weight updating equation. Since output of PLANN model represents the hazard rate, survival probabilities generated are always monotonically decreasing with time. However, this model suffers from data replication. It can cause high computational expenses. In addition, overtraining problem may arise leading to the lack of generalization capability.

3.5.3 Our Purposed Model

From limitations of both Street and PLANN models, this thesis introduces a new censored data technique that combines their advantages. It is based on the multiple-point model similar to Street's technique. The target is set as conditional failure probability in order to guarantee monotonically decreasing survival curve in a similar manner to PLANN. In case of type I censoring observations, the target vector is simply set as [0 0 0 0 0] for 5-year study period. For type II censoring, unknown targets are assigned with conditional failure probability or hazard rate. For example, the target vector of a patient who withdraws from the follow up at 3.5 years is set as [0 0 0 0.1 0.7] where 0.1 and 0.7 are hazard rates at years 4 and 5 respectively. The hazard rates are computed from the training data set. In the similar manner, the target vector of a patient who has cancer recurrence after 4 years is set as [0 0 0 1 0.7]. Quadratic error function is used as the cost function. Gradient descent is employed in minimization. Momentum term is added to enhance learning capability.

After training, survival probability at time *t* can be determined from Equation 2.6. Monotonically decreasing of *S(t)* can be guaranteed since the hazard rate is a positive value ranging from 0 and 1.

In the experiments, model parameters of Street, PLANN and our technique are varied in order to gain the best predictive performances. Hidden nodes are adjusted from 1-20 nodes. Learning rate is set as 0.001,0.01,0.05,0.1,0.5. Regularization parameter in PLANN model is varied from 0.001,0.01,0.05,0.1. Finally, the

momentum term is set at 0.01,0.05,0.1,0.5,0.7. The number of epochs for learning is set at 150,000.

3.6 Validation

The completed data are separated into two groups, 80% for model generation and 20% for model validation. Ten-fold cross validation is used for searching the optimal parameter.

3.7Comparison

Performances of each predictive model are evaluated from

3.7.1 Model discrimination: Area under the receiver operator characteristic curve (AUC) is used for evaluating model discrimination. The receiver operator characteristic curve, basically, is a plot of sensitivity versus 1-specificity. AUC is 1 representing a perfect model and 0.5 representing an unreliable case.

3.7.2 Model calibration: Chi-square statistic from Hosmer-Lemeshow goodness-of-fit test is used for model calibration. If it is below 15.51 corresponding to the p-value > 0.05, the model is fit.

3.7.3 Percent of non-monotonic prediction: It is determined from

$$\frac{number\ of\ \ patients\ who\ has\ non\ monotonically\ survival\ curve}{number\ of\ \ all\ \ patients} \times 100 \qquad (3.1)$$

3.7.4 Survival curve comparison: Survival curves are plotted and compared with the Log-rank test. This comparison represents predictive performance of an entire group of patients. Statistical difference between the generated survival curve of a model and the Kaplan-meier survival curve is investigated by the log-rank test. If the p-value is above 0.05, it represents these curves are not significantly different.

CHAPTER IV
RESULTS

4.1 Data

After the data collection process, there existed 495 records for the analysis. 348 records were censoring cases (70.30%) and the other 147 (29.70%) cases were uncensored data. The censored cases occurred from either the patients withdrew from the follow-up or they did not have recurring cancer within the 5-year study period. The numbers of patients who had cancer recurrence or withdrew from the study in each year are summarized in Table 4.1.

Table 4.1 The number of patients who have recurrence and lost to follow in each year

Number of Patients	Year periods						Total
	[0-1)	[1-2)	[2-3)	[3-4)	[4-5]	>5	
Recurrence	65	31	16	6	7	22	147
Lost to follow / withdrawn	51	48	23	18	19	189	348
Total	116	79	39	24	26	211	495

From Table 4.1, the number of uncensored and censored observations varied with time.

From the survival analysis, cancer recurrence of all patients can be presented in Table 4.2.

Table 4.2 The overall cancer recurrence

	Time at risk	Incidence rate	No. of all subject	No. of recurrence	Survival time		
					25%	50%	75%
Total	1518.725	0.082306	495	125	2.852778	.	.

From Table 4.2, the total number of patients who had cancer recurrence is 125 persons. Time at risk, time obtained from the summation of both censoring times or recurrence times of all patients, is 1518.725 years. The incident rate is 0.0823 persons per year. Within 2.852778 years, 25% of all patients had redeveloping cancer. This is called 25% survival time. The median survival time exceeded the 5-year study period. Survival curve of overall patients is shown in Figure 4.1.

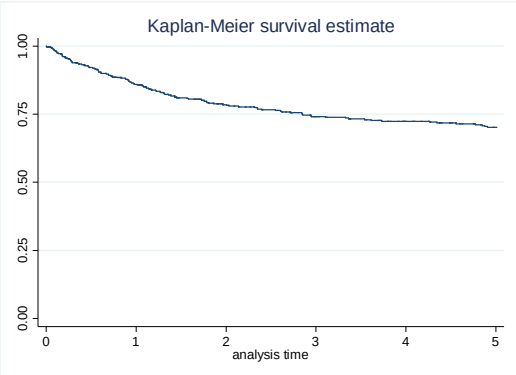

Figure 4.1 Kaplan-Meier survival curve

The Kaplan-Meier survival curve shown in Figure 4.1 is compared with the curves generated from censored data techniques.

4.2 Preprocessing

All categorical factors are encoded into numerical values. Ranging, coding and detail of each variable are presented in Table 4.3.

Table 4.3 Description of prognostic factors

Factors	Range/Coding	Descriptions
Sex	0 : Male 1 : Female	Patient's sex
T Stage (AJCC1997)	1 : Stage T1	Tumor confined to the nasopharynx
	2 : Stage T2	Tumor extends to soft tissue of oropharynx and/or nasal fossa
	3 : Stage T3	Tumor invades bony structures and/or paranasal sinuses
	4 : Stage T4	Tumor with intracranial extension and/or involvement of the cranial nerves, infra-temporal fossa, hypopharynx, orbit

Factors	Range/Coding	Descriptions
N Stage (AJCC1997)	0 : Stage N0	No regional lymph nodes metastasis
	1 : Stage N1	Unilateral metastasis in lymph node (s), 6 cm or less in greatest dimension, above the clavicular fossa
	2 : Stage N2	Bilateral metastasis in lymph none (s), 6 cm or less in greatest dimension, above the clavicular fossa
	3 : Stage N3	Metastasis in a lymph node (s)
M Stage (AJCC1997)	0 : Stage M0	No distant metastasis
	1 : Stage M1	Distant metastasis present
	2 : Stage Mx	Present of distant metastasis cannot be assessed
Staging (AJCC1997)	1 : Stage1	T1 N0 Mo
	2 : Stage2	T2 N0 M0
	3 : Stage3	T3 N0, T1-3 N1, M0
	4 : Stage4	T4 N0-1 M0, Any T N2-3 M0, Any T any N M1
Cell Type	1 : Type1	Well differentiate: Squamous cell carcinoma (SCC)
	2 : Type2	Poor differentiate: Non-keratinized carcinoma
	3 : Type3	Undifferentiated carcinoma
	4 : Type other	Other cell type
Chemotherapy	0 : No 1 : Yes	Presence of chemotherapy
Neck fibrosis	0 : No 1 : Yes	Presence of neck fibrosis after radiotherapy
Smoke	0 : No 1 : Yes	Smoking history
Alcohol	0 : No 1 : Yes	Drinking alcohol history
Family	0 : No 1 : Yes	Family history of having Cancer
Age	10-84	Age (year) at diagnostic date of NPC
Duration time	0-3600	Time interval (day) between first symptom time until time to conclusion that patient is NPC at Ear Nose Throat (ENT) department.
IgG	10-2560	Epstein-Barr Virus (EBV) antibody titer type IgG (Immunoglobulin G)
IgA	10-640	Epstein-Barr Virus (EBV) antibody titer type IgA (Immunoglobulin A)
Dose of radiation	0-8000	Dose of radiation at NPC area (cGy)
KPS	60-100	Karnofsky performance Scale Index (KPS)
Hb	6.3-17	Hemoglobin quantity before treatment (g/dl)
Status	0 1	Disease free Recurrence

4.3 Prognostic Factor Selection

4.3.1 Univariate analysis: the p-values from log-rank test and partial likelihood ratio test are presented in Tables 4.4 to 4.5.

Table 4.4 Categorical variable

Factors	25% Survival Times	Time at risk (person-times)	Incident rate*100	Event observed	p-value
Sex					0.4575
Male	2.3972	930.6056	8.70401	314	
Female	3.6139	588.1194	7.48147	181	
T					0.4815
T1	2.9472	256.1528	7.41745	74	
T2	4.8750	436.7139	6.86949	141	
T3	2.9417	424.9917	8.47075	128	
T4	1.5833	400.8667	9.97838	152	
N					**0.0189**
N0	.	365.6083	4.64978	101	
N1	2.8528	310.8083	7.07832	93	
N2	2.8417	614.0833	9.44497	215	
N3	1.4278	228.2250	12.26859	86	
M					0.9125
M0	2.8444	1382.0917	8.24837	440	
M1	0.6889	24.0889	12.45387	22	
Mx	4.9083	112.5444	7.1083	33	
Staging					**0.0069**
1	.	77.2917	2.5876	19	
2	.	149.8944	4.66995	41	
3	.	325.9083	4.90936	85	
4	1.8056	965.6306	10.35593	350	
Cell Type					0.3266
1	.	73.5806	2.71811	22	
2	2.8528	953.2278	7.97291	290	
3	2.3194	472.2944	9.73969	176	
4	.	19.6222	5.09626	7	
Chemotherapy					**0.0335**
No	.	344.1306	4.93999	90	
Yes	2.3972	1174.5944	9.19466	405	
Neck fibrosis					0.4023
No	2.8528	96.3444	11.41737	47	
Yes	2.8444	1422.3806	8.01473	448	
Smoke					**0.031**
No	4.7750	872.1306	6.76504	263	
Yes	1.8056	646.5944	10.20733	232	
Alcohol					**0.1228**
No	4.3472	839.8000	7.14456	256	
Yes	2.0333	678.9250	9.57396	239	
Family history					**0.0503**
No	2.9472	1352.9306	7.6131	432	
Yes	2.6056 / 4.9083	165.7944	13.26944	63	

Table 4.5 Continuous variables

Variables	Coefficient	95%CI		p-value
Age	0.0077726	-0.0063497	0.0218949	0.2787
Duration time	0.000055	-0.0005778	0.0006878	0.8672
IgG	0.0022772	0.0017341	0.0028202	**0.0000**
IgA	0.0027632	0.0004047	0.0051216	**0.0453**
Dose of radiation	-0.0003071	-0.0005532	-0.000061	**0.0316**
KPS	-0.0604818	-0.0888349	-0.0321287	**0.0001**
Hb	-0.0692452	-0.1836242	0.0451338	**0.2370**

Tables 4.4 and 4.5 show that N stage, staging, presence of chemotherapy, smoking, alcohol, family history, IgG quantity, IgA quantity, dose of radiation, KPS, and hemoglobin before treatment are important factors and relate to NPC recurrence (p-value <0.25). However, staging is excluded from the model because it relates to TNM staging. Neck fibrosis is included to the model because the medical doctor suggests that it should be one of the relevant factors. The factors from this step are used in multivariate analysis.

4.3.2 Multivariate analysis: prognostic factors from multivariate analysis are presented in Table 4.6.

Table 4.6 Prognostic factors from multivariate analysis.

Variables	Coef.	Std. Err.	z	P>z	[95% Conf. Interval]	
N1	0.377089	0.32537	1.16	0.2460	-0.26063	1.014802
N2	0.404927	0.284287	1.42	0.1540	-0.15226	0.962119
N3	0.684283	0.31682	2.16	**0.0310**	0.063327	1.30524
Smoke	0.383921	0.348085	1.1	0.2700	-0.29831	1.066156
Alcohol	-0.10319	0.338781	-0.3	0.7610	-0.76719	0.560807
Family history	0.393148	0.242281	1.62	0.1050	-0.08171	0.86801
Neck fibrosis	-0.22074	0.345411	-0.64	0.5230	-0.89773	0.456252
IgG	0.002211	0.000305	7.24	**0.0000**	0.001612	0.002809
IgA	0.000609	0.001378	0.44	0.6590	-0.00209	0.003309
Dose of radiation	-0.00021	0.000136	-1.58	0.1150	-0.00048	0.000052
KPS	-0.04783	0.015211	-3.14	**0.0020**	-0.07764	-0.01802

The relevant factors obtained from multivariate analysis are N3 stage, IgG, and KPS. These factors have the p-value below 0.05. Other factors do not relate to NPC recurrence but they cannot remove from the model because reduction of coefficients was greater than 20%

4.4 Predictive Performances

Tables 4.7 summarizes the model parameters of the best model after testing with the testing set.

Table 4.7 Best parameters of each technique

Techniques	Hd*	Lr*	Mo*	Rg*
Our technique	1	0.1	0.1	-
PLANN	1	0.1	-	0.01
Street	2	0.01	0.7	-

*Hd = The number of hidden node, Lr = Learning rate,
Mo = Momentum, Rg = Regularization parameter

Average AUCs from 10 fold cross validation are presented in Table 4.8.

Table 4.8 AUC of average test set

Time	Our study	PLANN	Street
1 year	0.7816	0.743	0.8092
2 year	0.7902	0.7932	0.8057
3 year	0.7991	0.7926	0.813
4 year	0.8163	0.8418	0.8324
5 year	0.8474	0.8858	0.8534
Average	0.8069	0.8113	**0.8227**

From Table 4.8, Street model provided highest average AUC in every time period. Independent sample t-test was applied. The results showed that they were not statistically different as presented in Table 4.9.

Table 4.9 P-value from independent sample t-test

	PLANN	Street
Our study	0.877	0.315
Street	0.677	

Table 4.10 AUC of the validation set

Time	Our Technique	PLANN	Street	Cox
1 year	0.7899	0.7294	0.7244	0.6655
2 year	0.7356	0.6978	0.6972	0.6667
3 year	0.7421	0.7064	0.7191	0.7116
4 year	0.7454	0.6933	0.7338	0.6681
5 year	0.761	0.7225	0.7491	0.7134
Average	**0.7548**	0.7099	0.7247	0.683

From Table 4.10, our proposed technique provided the best predictive performances (average AUC was 0.7548). All ANN based techniques outperformed the Cox model. Independent sample t-test was also employed. The results are summarized in Table 4.11.

Table 4.11 P-value from independent sample t-test for validation set

Techniques	Our Technique	PLANN	Street
PLANN	0.006*	-	-
Street	0.049*	0.216	-
Cox	0.002*	0.104	0.023*

*p-value < 0.05

The results showed that average AUC of our proposed technique was significantly higher than the other techniques.

For model calibration, chi-square from Hosmer-Lemeshow goodness-of-fit test was used. All models have chi-square lessen than 15.51 as shown in Table 4.12.

Table 4.12 Hosmer-Lemeshow goodness-of-fit test of validation set

Time	Our study	PLANN	Street	Cox
1 year	11.01	5.59	8.44	7.08
2 year	13.28	10.88	6.67	8.68
3 year	7.46	12.43	1.51	10.91
4 year	6.92	8.66	9.95	9.74
5 year	9.3	9.45	12.49	10.11

In the experiments, every model had 0% non-monotonically survival probability decreasing. Survival curve of all models are plotted versus Kaplan-Meier survival curve as presented in Figure 4.2.

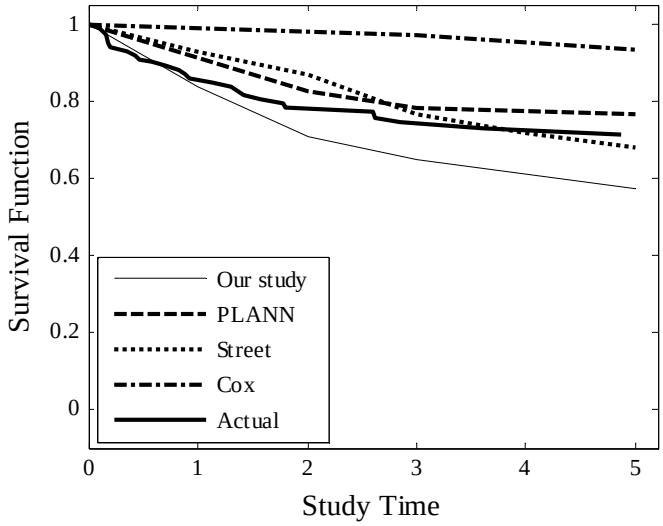

Figure 4.2 Survival curve of all models

From Figure 4.3, Cox model was obviously different from Kaplan-Meier survival curve. This was different from the curve of Street and PLANN models. The log-rank test showed that our proposed technique, PLANN, and Street provided the survival curves similar to that of Kaplan-Meier model as shown in Table 4.13.

Table 4.13 P-value from log-rank test

Our study	PLANN	Street	Cox
0.2518	0.1458	0.6883	0*

*p-value < 0.05

From Table 4.13, survival curve of Cox model was significantly different from the Kaplan-Meier model.

CHAPTER V
DISCUSSION

After univariate and multivariate analysis were applied for the factor selection, recurrence factors were N stage, family history, dose of radiation, smoking, alcohol, neck fibrosis, IgG quantity, IgA quantity, and KPS. Some of these factors, i.e. N stage, family history, and dose of radiation, support the previous studies in which they related to NPC development [15], [9], [12]. Though the other factors do not have any supporting medical report, the statistical analysis indicated that they related to NPC recurrence. The selected recurrence factors were used for generating the predictive models. In this thesis, four censored techniques are investigated. They are Street, PLANN, Cox model, and our purposed technique. The experimental results showed that average AUC of Street was slightly higher than PLANN and our purposed technique when testing with the test set as seen in Table 4.8. However, they were not statistically different when testing with the independent sample t-test as presented in Table 4.9. When the validation set was applied, average AUC of our proposed technique was highest. This means that the proposed technique can provide accurate prediction with unseen inputs. Due to the fact that the numbers of patients who have cancer recurrence and who withdrew from the study were different in each year as presented in Table 4.1, the developed predictive models might be less reliable for the later years.

For the model calibration, all predictive models had chi-square statistic less than 15.51 as seen in Table 4.12. This indicates that every model fitted well with the patients' data.

For survival curve comparison, all survival curves are plotted as shown in Figure 4.2. Cox model was obviously different from the others. This is confirmed by the log-rank test. Cox model is only significantly different from the Kaplan-Meier curve. This presents that linear relationship among data cannot be assumed.

The last issue is related to monotonically decreasing survival curve. All curves provided zero percent of non-monotonic prediction. This is not surprised for our proposed technique, PLANN, and Cox model because these techniques guaranteed monotonically decreasing survival curves. However, for the Street technique, monotonic survival curve may occur from handling censored data. With this technique, unknown targets of censored data were imputed by survival probabilities derived from Kaplan-Meier model. In the study, there exist a large number of censored data (70.30%) within the study period. The imputed targets which are monotonically decreasing values over time thus influence the predictive model to generate monotonic survival curve.

From the experiments, the investigated censored data techniques possess different advantages and disadvantages. PLANN uses time information as one of the input variables. Therefore, each data must be replicated before training. This requires high computational resources and time. In addition, the overtraining problem may arise. PLANN, however, guarantees monotonically decreasing survival curve. This is different from Street technique. Street technique uses multiple-point model. Therefore, only single model is sufficient for predicting several time points.

Cox proportional hazard model suffers from various limitations such as linear relationship among data and proportional hazard rates. However, this technique is generally used in medical prognosis.

Our purposed technique does not suffer from the replication problem. It can only use a single model for predicting various time points. Therefore, it is scalable well in a more complicated problem. In addition, it can guarantee monotonic survival curve generation and can efficiently handle unknown recurrence statuses when type II censoring data are observed. These reasons confirm that the proposed technique outperforms the other censored data techniques.

CHAPTER VI
CONCLUSION

New censored data technique is proposed in this thesis. It is applied to the prediction of nasopharyngeal carcinoma recurrence. The data were collected from Ramathibodi Hospital, Thailand. All missing data were imputed by EM imputation. Univariate and multivariate analysis were used for selecting recurrence factors. Four censored data techniques, i.e., PLANN, Street, our purposed technique, and Cox proportional hazard regression techniques were applied. In model development, the data were separated into two groups (80% for model generation and 20% for model validation). For ANN techniques, ten-fold cross validation is applied for searching the best training parameters. Our purposed technique combines advantages of PLANN and Street techniques. The technique is based on multiple-point model similar to Street leading to less computational expenses. Data replication is not required as in PLANN model. Overtraining problem can be relieved. Conditional failure probabilities are used for unknown targets as in PLANN. By doing so, survival probabilities generated are guaranteed to be monotonically decreasing with time. Experimental results showed that our proposed technique can handle four main problems, i.e., scalability problem in the single point model, non-monotonic survival curve generation problem, unknown recurrence status specification problem in the multiple-point model, data replication problem in the time-coded model. Additionally, our technique provides the highest predictive performances in terms of model discrimination, model calibration.

REFERENCES

1 Catalona wj, Smith ds. Cancer recurrence and survival rates after anatomic radical retropubic prostatectomy for prostate cancer: intermediate-term results. The Journal of Urology. 1998;160(6):2428-34.2

2 Shukla-Dave A, Hricak H, Ishill N, Moskowitz CS, Drobnjak M, Reuter VE, et al. Prediction of prostate cancer recurrence using magnetic resonance imaging and molecular profiles. Clinical Cancer Research. 2009;15(11):3842-9.

3 Laurentiis MD, Ravdin PM. Survival analysis of survival data: neural network analysis detection of complex interactions between variables. Breast Cancer Research and Treatment. 1994;32:113-8.

4 Ravdin PM, Clark GM. A practical application of neural network analysis for predicting outcome of individual breast cancer patients. Breast Cancer Research and Treatment. 1992;22:285-93.

5 Ohno-Machado L. Methodological review, modeling medical prognosis: survival analysis techniques. Journal of Biomedical Informatics. 2001;34:428-39.7

6 Mofidi R, Deans C, Duff MD, De Beaux AC, Paterson Brown S. Prediction of survival from carcinoma of oesophagus and oesophago-gastric junction following surgical resection using an artificial neural network. European journal of surgical oncology 2006;32:533-9.

7 Street WN, editor. A neural network model for prognostic prediction. Proceedings of the Fifteenth International Conference on Machine Learning; 1998; SanFrancisco.

8 Biganzoli E, Boracchi P, Mariani L, E M. Feed forward neural networks for the analysis of censored survival data: a partial logistic regression approach. Statistic in Medicine. 1998;17:1169-86.

9 พงศ์มฆพัฒน์ ธ, กุลประดิษฐารมน์ บ. มะเร็งช่องคอหลังโพรงจมูก. เอกสารประกอบการเรียนการสอน คณะแพทย์ศาสตร์ โรงพยาบาลรามาธิบดี.

10 Zahra T. Nasopharyngeal carcinoma: past, present and future directions: University of Gothenburg Sahlgrenska Academy; 2007.

11 Anita J, Todd M. B, Alwin J, Timothy D. Review of nasopharyngeal carcinoma. Ear, Nose and Throat Journal 2006.

12 H. Yijun, Y. Shu, H. Minghuang, Y. Xiaowei, Q. Fang, G. Ling, H. Peiyu, Z Guoyi: Application of support vector machine to predict 5 year survival status of patients with nasopharyngeal carcinoma after treatment. The Chinese-German Journal of Clinical Oncology 2006; 5:8-12.

13 Sameem Abdul-Kareem, Sapiya Baba, Yong Zulina Zubaibi, U Prasad, Mohd Ibrahim A Wahid. Prognosis system for NPC: A comparison of the multilayer perceptron model and the recurrent model. Proceeding of the 9th international conference on Neural Information Proceeding (ICONIP'02), Vol.1, 2002.

14 Tang S SL, Chen W, Tsang S, Chang J, Hong J. . The effect of nodal status on determinants of initial treatment response and patterns of relapse-free survival in nasopharyngeal carcinoma. International Journal of Radiation Oncology. 2000;47(4):867-73.15

15 Terence P, Fernando L, Roberto AL, Jacob K, Geraldo M, Mauro MB, et al. Prognostic Factors and Outcome for Nasopharyngeal Carcinoma Arch Otolaryngol Head Neck Surg. 2003;129(7):794-9.

16 Hwang J, Fu K, Phillips T. Results and prognostic factors in the retreatment of locally recurrent nasopharyngeal carcinoma. International Journal of Radiation Oncology. 1998;41(5):1099-111.

17 Elisa TL. Statistical methods for survival data analysis. 2, editor: John Wiley & Sons, Inc.; 1992.

18 Joseph A Cruz, David S Wishart. Application of machine learning in cancer prediction and prognosis. Cancer Informatics 2006;2:59-78.

19 M.D. Laurentiis, P.M. Ravdin: Survival analysis of censored data : neural network analysis detection of complex interactions between variables. Breast Cancer Research and Treatment 1994;32:113-118.

20 J.M. Jerez, I. Molina, J.L. Subirats, L. Franco: Missing data imputation in breast cancer prognosis. Proceeding of the 24th IASTED International Multi-

Conference BIOMEDICAL ENGINEERING 2006.

21 Bakar OF, Sameem AK. Soft computing in medicine: Nasopharyngeal carcinoma prognosis. The Internet Journal of Medical Informatics. 2005;2(1).

22 R.N.G. Naguib, J. Wayman, M.K. Bennett, S.A. Raimes, S.M. Griffin: Pre and post-operative prediction of recurrence in patients with cancer of the oesophago-gastric junction using radial basis function artificial neural network. Proceeding of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society 1998;20 No.6.

23 Jones AS, Taktak AGF, Helliwell TR, Fenton JE, Birchall MA, Husband DJ, et al. An artificial neural network improves prediction of observed survival in patients with laryngeal squamous carcinoma. Eur Arch Otorhinolaryngol. 2006;263:541-7.

24 M. Theeuwen, B. Kappen, J. Neijt: Neural network analysis to predict treatment outcome in patients with ovarian cancer.

25 R. Mofidi, C. Deans, M.D. Duff, A.C. de Beaux, S. P. Brown: Prediction of survival from carcinoma of oesophagus and oesophago-gastric junction following surgical resection using an artificial neural network. EJSO the Journal of Cancer Surgery 2006;32:533-539.

26 Kumdee O, Ritthipravat P, Bhongmakapat T, Cheewaruangroj W. Dealing with missing values for effective prediction of NPC recurrence. SICE Annual Conference; 2008; Japan. 2008. p. 1290-4.

27 David Faraggi and Richard Simon. A neural network model for survival data. Statistics in Medicine 1995;14:73-82.

28 P. Lapuerta, S.P. Azen and L. Labree. Use of neural networks in predicting the risk of coronary artery diseases. Computer and Biomedical Research 1995;28:38-52.

29 P. Ritthipravat, “Recurrent prediction software for patients with nasopharyngeal carcinomas,” research report for a young researcher grant program, Mahidol University, 2008

30 Hosmer DW JR, Lemeshow S. Applied survival analysis: Regression modeling of time to event data. New York: John Wiley & Sons; 1999.

BIOGRAPHY

NAME	Miss Oraya Intem
DATE OF BIRTH	1 August 1981
PLACE OF BIRTH	Phayao, Thailand
INSTITUTIONS ATTENDED	Nurasuan University, 2000-2003 Bachelor of Science (Radiological Technology) Mahidol University, 2006-2010 Master of Engineering (Biomedical Engineering)
RESEARCH GRANTS	TRFMAG-WII
HOME ADDRESS	159 M. 6 Thungkluai, Phusang, Phayao 56110 Tel. 087-1667510 E-mail : koysth@yahoo.com
PUBLICATION / PRESENTATION	Intem O, Ritthipravat P, Bhongmakapat T. "Comparison of censored data technique for NPC recurrence prediction". ISBME 2009; BKK2009..