

แบบจำลองการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูก ด้วยโครงข่ายประสาทเทียม

Model for Cervical Cancer Result Prediction Using Artificial Neural Network

เชาวนนท์ โสโธ (Chaowan Soto)* ดร.พุดษดี ศิริแสงตระกูล (Dr.Pusadee Seresangtakul)^{1**}
วรชัย ตั้งวรพงศ์ชัย (Vorachai Tangvoraphonkchai)^{***}

บทคัดย่อ

งานวิจัยนี้นำเสนอแบบจำลองเพื่อทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสี โดยการประยุกต์ใช้โครงข่ายประสาทเทียมแบบแพร่ย้อนกลับ ในการศึกษาผู้วิจัยได้รวบรวมปัจจัยต่างๆ ที่มีความสัมพันธ์ต่อผลการรักษาจากงานวิจัยทางการแพทย์ที่เกี่ยวข้อง และได้ทำการวิเคราะห์ความสัมพันธ์ของปัจจัยด้วยวิธีแบบลำดับชั้น ผลการศึกษาพบว่าปัจจัยที่มีความสัมพันธ์ต่อผลการรักษาประกอบด้วย ระยะของโรค อายุ ขนาดของก้อนเนื้อ ประเภทของเซลล์มะเร็ง และน้ำหนักตัว จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่าระดับฮีโมโกลบินมีผลตอบสนองต่อการรักษา ดังนั้นจึงได้เพิ่มระดับฮีโมโกลบินร่วมกับปัจจัยที่ได้จากการวิเคราะห์รวมเป็น 6 ปัจจัย ในการสร้างแบบจำลองด้วยตัวแบบโครงข่ายประสาทเทียม ผลการศึกษาพบว่าตัวแบบจำลองที่นำเสนอถึงแม้จะมีประสิทธิภาพสูง แต่ค่าความจำเพาะมีค่าเป็น 0% ซึ่งจากการวิเคราะห์พบว่าปัญหาเกิดจากความไม่สมดุลของข้อมูลซึ่งมีข้อมูลกลุ่มหนึ่งมากกว่าอีกกลุ่มเป็นจำนวนมาก ดังนั้นผู้วิจัยจึงได้ทำการแก้ปัญหาปรับความไม่สมดุลของข้อมูลด้วยวิธี Cost-Sensitive Learning: CSL และวิธี Synthetic Minority Over-sampling Technique: SMOTE และทำการสร้างแบบจำลองข้อมูลโครงข่ายประสาทเทียมและแบบจำลองด้วยการถดถอยแบบลอจิสติก เมื่อเปรียบเทียบประสิทธิภาพการทำนายพบว่าโครงข่ายประสาทเทียมที่มีการแก้ปัญหาค่าความไม่สมดุลของข้อมูลด้วยวิธีการสุ่มเพิ่มชุดข้อมูลตัวอย่างมีประสิทธิภาพการทำนายที่ ด้วยค่าความถูกต้อง 81.71% ค่าความไว 94.47% และค่าความจำเพาะ 55.47% สูงกว่าวิธีการถดถอยแบบลอจิสติกที่มีการแก้ปัญหาค่าความไม่สมดุลของข้อมูลด้วยวิธี CSL ซึ่งมีค่าความถูกต้อง 81.00% ค่าความไว 84.52% และค่าความจำเพาะ 30.66%

ABSTRACT

This paper presents models to predict the result of cervical cancer treatment by radiotherapy using Back Propagation Artificial Neural Network (BPANN). In order to select appropriate input factors for the model, the researchers studied factors that had an effect on

¹ Correspondent author: pusadee@kku.ac.th

* นักศึกษา หลักสูตรวิทยาศาสตรมหาบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

** ผู้ช่วยศาสตราจารย์ ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

*** รองศาสตราจารย์ นพ. ภาควิชารังสีวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

cervical cancer. Stepwise factor analysis was used to analyze the relationship between the factors and the result of cervical cancer treatment. The results show that there are five factors that are significant in the result of cervical cancer treatment by radiotherapy; these include the cancer stage, age, tumor size, type of cancer cell, and body weight. Medical research has shown that hemoglobin is also related to the treatment of cervical cancer. Therefore, hemoglobin was also used as an additional input factor of the prediction model. The artificial neural network (ANN) and logistic regression models were constructed and used to predict the results of cervical cancer treatment. In order to analyze for the appropriate models, the factors data from 1994 to 1998 were used in the study. The results showed that the accuracy was very high but the models could not identify the specificity rate. The analysis results showed that the problem came from imbalanced data sets. In order to improve the efficiency of the models, Cost Sensitive Learning (CSL) and Synthetic Minority Over-sampling Techniques (SMOTE) were addressed to resolve the imbalance in the data sets. The output data sets were used to construct the artificial neural network and logistic regression models. The experimental results show that the accuracy, sensitivity and specificity of the ANN with SMOTE are 81.71%, 94.47% and 55.47% compared to the Logistic Regression with CSL are 81.00%, 84.52% and 30.66%, respectively. The results showed that the ANN with imbalanced data by SMOTE was more accurate than the logistic regression with imbalanced data by CSL.

คำสำคัญ: การทำนายผลการรักษามะเร็งปากมดลูก โครงข่ายประสาทเทียม ความไม่สมดุลข้อมูล

Key Words: Cervical cancer result prediction, Artificial neural network, Imbalance data

บทนำ

ปัจจุบันโรคมะเร็งปากมดลูกเป็นมะเร็งที่พบมากที่สุดเป็นอันดับสองของโลกที่พบในเพศหญิง รองลงมาจากมะเร็งเต้านม พบประมาณ 500,000 รายต่อปี และมีอัตราการเสียชีวิตประมาณ 250,000 รายต่อปี โดยจะพบในประเทศที่กำลังพัฒนามากกว่าประเทศที่พัฒนาแล้ว [1]

เนื่องจากผู้ป่วยมะเร็งปากมดลูกจะเข้ารับการรักษาเมื่อมีระยะของโรคเป็นมากแล้ว ตั้งแต่ระยะที่ 2 ขึ้นไป ส่วนใหญ่จึงเข้ารับการรักษาด้วยรังสีรักษา [12] การรักษาโรคมะเร็งปากมดลูกด้วยวิธีการฉายรังสี สามารถรักษาได้ทุกช่วงระยะของโรค การประเมินผลการรักษาแบ่งเป็น 2 รูปแบบคือ การศึกษาอัตราการรอดชีพ (Survival rate) หมายถึง อัตราการอยู่

รอดของผู้ป่วยภายหลังการรักษาจนกระทั่งผู้ป่วยเสียชีวิต แต่การศึกษาต้องใช้เวลานานและยังไม่อาจสรุปได้ว่าผู้ป่วยเสียชีวิตด้วยสาเหตุใด และการศึกษาอัตราการหายของโรค (Disease free rate) ซึ่งจะตรวจการหายของโรค ณ เวลาต่างๆ ที่กำหนด การประเมินแบบหลังนี้เป็นดัชนีทางอ้อมในการทำนายถึงการรอดชีพของผู้ป่วยในอนาคตได้

ดังนั้นหากมีระบบที่ใช้ปัจจัยที่เกี่ยวข้องกับผู้ป่วยช่วยคาดการณ์แนวโน้มว่าผู้ป่วยมีโอกาสที่จะกลับมาเป็นซ้ำหรือมีการแพร่กระจายของมะเร็ง ก็จะเป็นข้อมูลช่วยแพทย์ในการวางแผนการรักษาที่เหมาะสมกับผู้ป่วยแต่ละรายได้

การวิเคราะห์โรคหรือจำแนกข้อมูลทางการแพทย์ ส่วนใหญ่แล้วนิยมใช้โครงข่ายประสาทเทียม

[2, 16] เนื่องจากมีความสามารถในการจำแนกข้อมูลอยู่ในระดับที่ดี เช่น Yan *et al.* [25] วินิจฉัยโรคหัวใจโดยใช้โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น พบว่าผลการวินิจฉัยมีความถูกต้องมากกว่าร้อยละ 90 งานวิจัยของ Botoa *et al.* [2] ทำนายมะเร็งต่อมลูกหมาก โดยใช้โครงข่ายประสาทเทียมแบบ MLP, RBF-MLP และ RNN เปรียบเทียบกับการถดถอยแบบโลจิสติก พบว่าโครงข่ายประสาทเทียมให้ผลการทำนายดีกว่าการถดถอยแบบโลจิสติก และ Chowdhury *et al.* [4] ใช้โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนเพื่อวินิจฉัยโรคที่เกิดในเด็กทารกแรกเกิดผลการศึกษา มีความถูกต้องในการวินิจฉัยถึง 75%

ดังนั้นงานวิจัยนี้จึงได้นำเอาวิธีโครงข่ายประสาทเทียมมาประยุกต์ใช้ในการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสี พร้อมทั้งได้นำวิธีการถดถอยโลจิสติกเข้ามาเปรียบเทียบประสิทธิภาพในการทำนาย

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

1. โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น (Multilayer Perceptron Artificial Neural Network: ANN)

โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น [9] ประกอบด้วยชั้นอินพุต (Input Layer) ชั้นซ่อน (Hidden Layer) และชั้นเอาต์พุต (Output Layer) แสดงตามภาพที่ 1 ในแต่ละชั้นของโครงข่ายประสาทจะประกอบด้วยหน่วยประมวลผลหรือโหนด [8] ขั้นตอนวิธีแบบแพร่ย้อนกลับ (back propagation algorithm) [19] เป็นขั้นตอนวิธีที่ใช้ในการฝึกสอนโครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น ใช้สำหรับปรับค่าน้ำหนักเส้นเชื่อมโยงของโหนดต่างๆ ให้มีค่าที่เหมาะสม โดยแบ่งเป็น 2 ส่วน คือ ส่วนป้อนไปข้างหน้า (Feed forward) เริ่มต้นจากข้อมูลจะถูกส่งไปยังชั้นถัดไปของโครงข่ายประสาท โหนดแต่ละโหนดในชั้นถัด

ไปจะมีการคำนวณค่าเอาต์พุตที่ได้จากชั้นก่อนหน้าตามสมการที่ (1)

$$u_j^{(l)} = \sum_{i=0}^p w_{ji}^{(l)} \square y_i^{(l-1)} \quad (1)$$

โดยที่ $y_i^{(l-1)}$ คือ ค่าเอาต์พุตโหนดที่ i ในชั้นก่อนหน้า ($l-1$)

$W_{ji}^{(l)}$ คือ ค่าถ่วงน้ำหนักที่เชื่อมโยงระหว่างโหนดที่ j ชั้นที่ l กับโหนดที่ i ชั้นที่ ($l-1$)

p คือ จำนวนโหนดชั้นที่ ($l-1$)

$u_j^{(l)}$ คือ ค่าผลลัพธ์ที่ได้จากการคำนวณ และจะต้องนำค่าดังกล่าวไปผ่านฟังก์ชันกระตุ้นเพื่อให้ได้เอาต์พุตที่ต้องการ $y_j^{(l)}$

สมการฟังก์ชันกระตุ้น (Activation Function) [11] จะแสดงคำตอบอยู่ในช่วง -1 ถึง 1 ได้แก่ ฟังก์ชันไบนารี (binary function) เป็นฟังก์ชันเชิงตรรกศาสตร์ที่จะให้ค่าเป็นจริงหรือเท็จเท่านั้น ฟังก์ชันซิกมอยด์ แบ่งเป็น 2 ประเภทคือ Log-sigmoid ค่าที่ได้จะอยู่ในช่วง 0 ถึง 1 ตามสมการที่ (2) และ Tan-sigmoid ค่าที่ได้จะอยู่ในช่วง -1 ถึง 1 ตามสมการที่ (3)

$$y = \frac{1}{1 + \exp(-v)} \quad (2)$$

$$y = \frac{1 - \exp(-v)}{1 + \exp(-v)} \quad (3)$$

คำนวณค่าความผิดพลาดระหว่างเอาต์พุตที่ต้องการ (desire output) และเอาต์พุตที่ชั้นเอาต์พุต ($y_j^{(L)}$) ตามสมการที่ (4)

$$e_j = d_j - y_j^{(L)} \quad (4)$$

โดยที่ e_j คือค่าความผิดพลาด และ d_j คือเอาต์พุตที่ต้องการ (desire output)

ส่วนแพร่ย้อนกลับ (Back Propagation) เริ่มจากการคำนวณหาค่าความชันเฉพาะที่ (δ) ย้อน

หลังจากชั้นปัจจุบันกลับไปยังชั้นก่อนหน้า ตามสมการที่ (5)

$$\delta_j^{(l)} = y_j^{(l)} [1 - y_j^{(l)}] \sum_k \delta_k^{(l+1)} w_{kj}^{(l+1)} \quad (5)$$

โดยที่ $\delta_k^{(l+1)}$ คือค่าความชันเฉพาะที่โหนดที่ k ชั้นที่ $l+1$

$w_{kj}^{(l+1)}$ คือค่าถ่วงน้ำหนักที่เชื่อมโยงระหว่างโหนดที่ k ชั้นที่ $l+1$ กับโหนดที่ j ชั้นที่ l

หลังจากคำนวณค่าความชันเฉพาะที่แล้วคำนวณปรับค่าถ่วงน้ำหนัก ตามสมการที่ (6)

$$w_{ji}^{(l)}(new) = w_{ji}^{(l)}(n) + \alpha [w_{ji}^{(l)}(n)] + \eta \delta_j^{(l)}(n) y_i^{(l-1)} \quad (6)$$

โดยที่ η คืออัตราการเรียนรู้ (Learning Rate)

α คือค่าโมเมนตัม (Momentum)

n คือจำนวนรอบ

2. การถดถอยโลจิสติก (Logistic Regression)

การถดถอยโลจิสติก มีวัตถุประสงค์เพื่อศึกษาความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระ และนำสมการที่ได้ไปพยากรณ์ตัวแปรตามเมื่อมีการกำหนดค่าตัวแปรอิสระ การวิเคราะห์การถดถอยแบบโลจิสติกแบ่งออกเป็น 2 ประเภทคือ [23]

2.1 Binary Logistic ใช้ในกรณีที่ตัวแปรตามเป็นตัวแปร 2 ค่าเท่านั้น

2.2 Multinomial Logistic ใช้ในกรณีที่ตัวแปรตามเป็นตัวแปรที่มีค่ามากกว่า 2 ค่า

ตัวแบบการวิเคราะห์การถดถอยโลจิสติกสามารถเขียนได้ตามสมการที่ (7)

$$P(event) = \frac{e^z}{1 + e^z} \quad (7)$$

โดยที่ $z = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p$

β คือ ค่าสัมประสิทธิ์ที่ประมาณได้จากข้อมูล

X คือ ค่าตัวแปรอิสระ

e คือ ค่าลอการิธึมธรรมชาติ (natural logarithm) มีค่าประมาณ 2.718

โอกาสที่จะไม่เกิดเหตุการณ์สามารถประมาณได้จากสมการ (8)

$$P(noevent) = 1 - P(event) \quad (8)$$

3. การจัดการความไม่สมดุลของข้อมูล (Imbalance Data)

การจำแนกข้อมูลที่มีข้อมูลคลาสหนึ่งมากกว่าอีกคลาสหนึ่งมาก จะนำมาสู่ปัญหาความไม่สมดุลของข้อมูล (Imbalanced data) ทำให้ผลลัพธ์ที่ได้จากการจำแนกข้อมูลมีความโน้มเอียงไปทางคลาสที่มีจำนวนมาก [7] วิธีการแก้ปัญหาคือการปรับสมดุลของข้อมูลมีหลายวิธี เช่น วิธี Cost-Sensitive Learning [10, 14] และการสุ่มเพิ่มชุดข้อมูลตัวอย่าง [27]

3.1 Cost-Sensitive Learning

การปรับความสมดุลของข้อมูลด้วย Cost-Sensitive Learning [10, 14] วัตถุประสงค์ของวิธีการจะพิจารณาค่าใช้จ่ายในการจำแนกผิด (misclassification cost) คือ ฟังก์ชันระหว่างคลาสค่าความจริง (actual class) และคลาสค่าทำนาย (predicted class) แทนด้วย $cost(actual\ class, predicted\ class)$ หรือเรียกว่า cost matrix [13] โดยที่ค่าใช้จ่าย (cost) ไม่ได้หมายความว่าเงินแต่เป็นเวลาที่ใช้ไป [5] ค่าที่ได้จากการสร้าง cost matrix จะใช้ในกระบวนการเรียนรู้ของตัวแบบการทำนายเพื่อปรับค่า threshold ในการจำแนกข้อมูลให้ดีขึ้น [26]

3.2 การสุ่มเพิ่มชุดข้อมูลตัวอย่าง (Over-Sampling)

การปรับความสมดุลของข้อมูลด้วยวิธีการสุ่มเพิ่มชุดข้อมูลตัวอย่าง จะสุ่มเพิ่มชุดข้อมูลที่มีคลาสน้อยจนกว่าจะได้จำนวนของชุดข้อมูลมีจำนวน

ใกล้เคียงหรือเท่ากับชุดข้อมูลอื่น ข้อเสียคือใช้เวลาในการฝึกสอนเพิ่มมากขึ้นกว่าเดิม เนื่องจากมีการเพิ่มจำนวนชุดตัวอย่างเพิ่มมากขึ้น [15] ได้นำเสนอวิธี Synthetic Minority Over-Sampling Technique (SMOTE) เพื่อเพิ่มชุดข้อมูล โดยวิธีนี้จะสุ่มเพิ่มชุดข้อมูลของคลาสที่มีจำนวนน้อย มีวิธีการดังนี้

- 1) คำนวณหาผลต่างระหว่างชุดข้อมูลที่พิจารณาและชุดข้อมูลที่ใกล้เคียง
- 2) คำนวณหาค่าข้อมูลใหม่ด้วยการคูณผลลัพธ์จากข้อ 1 ด้วยค่าที่ได้จากการสุ่มตัวเลขที่อยู่ในช่วง 0 ถึง 1

4. การจัดการค่าข้อมูลที่ขาดหาย (Missing Value)

ในการวิเคราะห์ข้อมูล การขาดหายของข้อมูลจะส่งผลกระทบต่อประสิทธิภาพในการจำแนกข้อมูล [20] วิธีการที่ใช้ในการจัดการกับปัญหาดังกล่าว ได้แก่ การตัดกลุ่มข้อมูลที่สูญหายทิ้งไป (ignoring and discarding data) การแทนที่ค่าที่หายไป (imputation) ด้วยกระบวนการประมาณค่า [20] วิธีการทางสถิติ วิธีการเรียนรู้ของเครื่อง (machine learning) [6] ซึ่งในงานวิจัยนี้เลือกใช้การเติมค่าข้อมูลด้วยวิธีการทางสถิติ เนื่องจากวิธีดังกล่าวเหมาะสมสำหรับชุดข้อมูลที่มีค่าต่อเนื่องและข้อมูลที่มีค่าแบบจัดกลุ่ม [6]

5. การคัดเลือกตัวแปรแบบลำดับขั้น (Stepwise Selection)

การคัดเลือกตัวแปรแบบลำดับขั้น [18] เป็นเทคนิคทางสถิติเพื่อวิเคราะห์ความสัมพันธ์ระหว่างตัวแปรตามและตัวแปรอิสระหลายตัวแปร วิธีการเริ่มจากคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ r ระหว่างตัวแปรตามและตัวแปรอิสระแต่ละตัว ดังแสดงตามสมการที่ (10)

$$r = \frac{\sum xy - (\sum x)(\sum y)/n}{\sqrt{\sum x^2 - (\sum x)^2/n} \sqrt{\sum y^2 - (\sum y)^2/n}} \quad (10)$$

โดยที่ x คือ ตัวแปรอิสระและ y คือ ตัวแปรตาม

n คือ จำนวนข้อมูลของแต่ละตัวแปร เลือกตัวแปรอิสระที่มีค่า r สูงสุด เข้าสมการ ตามสมการที่ (11)

$$Y = \beta_0 + \beta_p X_p (\pm \epsilon)$$

โดยที่ X_p คือ ตัวแปรอิสระที่ถูกเลือกเข้าสมการ

β_0 คือ ค่าคงที่ (Constant) ของสมการถดถอย

β_p คือ ค่าสัมประสิทธิ์การถดถอย (Regression Coefficient) ของตัวแปรอิสระตัวที่ $p (X_p)$

ϵ คือ ค่าความคลาดเคลื่อนระหว่างค่าจริง y และค่าที่ได้จากสมการ \hat{y}

นำตัวแปรที่ไม่ได้ถูกเลือกเข้าสมการหาค่า F-test หรือ t-test แล้วนำค่าที่มากที่สุดมาพิจารณาค่าความมีนัยทางสถิติ ถ้ามีค่าน้อยกว่าหรือเท่ากับ 0.05 ให้นำตัวแปรดังกล่าวเข้าสมการ แต่ถ้ามีค่ามากกว่า 0.05 จะหยุดการคำนวณและได้สมการที่เหมาะสม ตัวแปรอิสระตั้งแต่ตัวที่ 2 เป็นต้นไปถูกเลือกเข้าสู่สมการ จะมีการตรวจสอบตัวแปรที่อยู่ในสมการว่าตัวแปรอิสระใดควรถูกคัดเลือกเข้าสมการ

6. การวัดประสิทธิภาพ

6.1 ครอสวาไลเดชัน (Cross Validation)

ครอสวาไลเดชัน [8] เป็นการวัดประสิทธิภาพตัวแบบในการจำแนก โดยการแบ่งข้อมูลออกเป็นกลุ่มย่อยขนาดเท่ากัน (folds) จำนวน k กลุ่ม เพื่อให้ข้อมูลทุกตัวมีโอกาสเป็นชุดทดสอบและชุดสอน เพื่อป้องกันปัญหาการเลือกข้อมูลที่ดีและง่ายมาเป็น

ข้อมูลชุดทดสอบ การฝึกสอนและทดสอบตัวแบบ จะทำทั้งหมด k รอบ โดยรอบแรกใช้ข้อมูลชุดที่ 1 เป็นข้อมูลทดสอบ ส่วนข้อมูลชุดที่เหลือเป็นข้อมูลชุดฝึกสอน จากนั้นจะสลับข้อมูลชุด 2 มาเป็นชุดทดสอบและข้อมูลกลุ่มอื่น ๆ ที่เหลือเป็นชุดสอน สลับอย่างนี้ไปเรื่อย ๆ จนครบ k ชุด และคำนวณหาประสิทธิภาพเฉลี่ย

6.2 Confusion Matrix

Confusion Matrix [21] คือตารางสรุปจำนวนข้อมูลที่ตัวแบบมีการจำแนกได้ถูกต้องและไม่ถูกต้อง ดังแสดงในตารางที่ 1 จากตารางที่ 1 สามารถนำข้อมูลในตารางมาใช้ในการคำนวณการวัดประสิทธิภาพของตัวแบบการทำนาย ด้วยวิธีการหาค่าความแม่นยำ (Accuracy) อัตราความถูกต้องเชิงบวก (True Positive Rate: TP) หรือค่าความอ่อนไหว (Sensitivity) หรือค่าความระลึก (Recall) และอัตราความถูกต้องเชิงลบ (True Negative Rate: TN) หรือค่าความจำเพาะ (Specificity)

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (12)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

7. งานวิจัยที่เกี่ยวข้อง

Padmavathi [17] ศึกษาเปรียบเทียบประสิทธิภาพการทำนายมะเร็งเต้านมโดยใช้ตัวแบบที่ใช้ในการทำนายคือ โครงข่ายประสาทเทียมเพอร์เซปตรอนหลายชั้น (MLP) โครงข่ายประสาทเทียมเรเดียลเบสิสฟังก์ชัน (RBF) และวิธีโลจิสติกส์เกรสชัน พบว่า วิธี MLP, RBF, วิธีโลจิสติกส์เกรสชันสามารถทำนายว่าเป็นโรคได้ถูกต้อง 92.1%, 97.3% และ 75.5% ตามลำดับ

สุพจน์ และคณะ [12] ได้ศึกษาหาโมเดลทำนายการหายจากการรักษาผู้ป่วยมะเร็งปากมดลูกที่ได้รับการรักษาโดยวิธีรังสีรักษา ที่เข้ารับการรักษา

ที่โรงพยาบาลศรีนครินทร์ โดยใช้เทคนิคทางสถิติวิเคราะห์การถดถอยพหุคูณโลจิสติก (Multiple logistic regressions) ผลการวิจัยพบว่าโมเดลสามารถทำนายการหายได้ถูกต้องร้อยละ 78.6 และทำนายการไม่หายได้ถูกต้อง ร้อยละ 43.2

วรชัย [22] ศึกษาปัจจัยที่เกี่ยวข้องกับการตอบสนองต่อการรักษามะเร็งปากมดลูกด้วยรังสี พบว่าปัจจัยที่มีส่วนสัมพันธ์กับการตอบสนองของมะเร็งปากมดลูกต่อการรักษาด้วยรังสี ดังนี้ ระยะของโรค ปริมาณรังสีที่ใช้ ขนาดของเนื้องอก และภาวะโลหิตจาง

Winter *et al.* [24] ศึกษาปัจจัยที่มีผลต่อการอยู่รอดของผู้ป่วยมะเร็งปากมดลูกระยะที่ II III และ IV ที่รับการรักษาด้วยวิธีการฉายรังสีพร้อมเคมีบำบัด พบว่า ปัจจัยที่มีผลต่อการอยู่รอด ได้แก่ อายุ เชื้อชาติ ความสมบูรณ์ของร่างกาย ระยะของโรค ขนาดของเซลล์มะเร็ง ชนิดเซลล์มะเร็ง ระยะเวลาในการรักษา และระดับฮีโมโกลบินเฉลี่ย

Chang *et al.* [3] ศึกษาวิธีการคัดเลือกปัจจัยที่เหมาะสมสำหรับการพยากรณ์มะเร็งช่องปาก เปรียบเทียบกับใช้ปัจจัยทั้งหมดในการพยากรณ์ พบว่าการพยากรณ์ร่วมกับวิธีการคัดเลือกปัจจัยนั้นมีประสิทธิภาพสูงกว่าการใช้ปัจจัยทั้งหมดในการพยากรณ์

วิธีการดำเนินการวิจัย

งานวิจัยนี้มีขั้นตอนในการดำเนินการ ประกอบด้วย การรวบรวมข้อมูล การวิเคราะห์ปัจจัยเพื่อหาปัจจัยที่มีผลต่อการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูก และการสร้างตัวแบบการทำนายด้วยวิธีการโครงข่ายประสาทเทียม และโลจิสติก ดังปรากฏในภาพที่ 2

1. กระบวนการเตรียมข้อมูลก่อนการประมวลผล (Pre-processing)

1.1 การรวบรวมข้อมูล (Data collection)

ในงานวิจัยนี้ ใช้ข้อมูลทำการเก็บรวบรวมข้อมูลแบบย้อนกลับจากฐานข้อมูลผู้ป่วยมะเร็งปาก

มดลูก หน่วยรังสีรักษา โรงพยาบาลศรีนครินทร์ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น ระหว่างวันที่ 1 มกราคม พ.ศ. 2537 ถึงวันที่ 31 ธันวาคม พ.ศ. 2552 ซึ่งมีปัจจัยที่รวบรวมได้ทั้งสิ้น 17 ปัจจัยดังแสดงในตารางที่ 2

1.2 การตรวจสอบความถูกต้องของข้อมูล (Data cleaning)

การตรวจสอบความถูกต้องของข้อมูลเป็นการทำให้ข้อมูลมีคุณภาพในการนำไปวิเคราะห์ซึ่งมีขั้นตอนดังนี้

(1) ข้อมูลที่ไม่สมบูรณ์ (Missing value) ในงานวิจัยนี้จะทำการเติมค่าข้อมูลที่ขาดหาย ซึ่งข้อมูลที่มีค่าเป็นค่าต่อเนื่อง (Continuous Data) จะใช้วิธีคำนวณค่าเฉลี่ย (Mean) และข้อมูลที่มีค่าแบบจัดกลุ่มจะใช้วิธีแทนค่าด้วยการคำนวณฐานนิยม (Mode)

(2) จะทำการตัดข้อมูลของผู้ป่วยทั้ง ในกรณีที่ผู้ป่วยคนนั้นได้รับการผ่าตัดมาก่อนที่จะได้รับการรักษาด้วยรังสีรักษาเนื่องจากไม่สามารถติดตามผลขึ้นเนื้อที่มดลูกได้

หลังจากที่ผ่านกระบวนการตรวจสอบความถูกต้องของข้อมูลแล้วมีจำนวนข้อมูลทั้งสิ้น 2,236 ชุดข้อมูล ในการสร้างตัวแบบผู้วิจัยได้ทำการแบ่งข้อมูลโดยใช้ข้อมูลผู้ป่วยระหว่างปี 2537 ถึง 2551 จำนวน 2,107 ชุด เพื่อใช้สร้างตัวแบบและใช้ข้อมูลผู้ป่วยปี 2552 จำนวน 129 ชุดเพื่อใช้ทดสอบความถูกต้องตามลำดับ

2. การวิเคราะห์และคัดเลือกปัจจัย

จากข้อมูลปัจจัยที่มีการรวบรวมในฐานข้อมูลทั้ง 17 ปัจจัยผู้วิจัยได้ทำการคัดเลือกปัจจัยที่เกี่ยวข้องกับทางการแพทย์มาทำการวิเคราะห์โดยวิธีการคัดเลือกตัวแปรแบบลำดับขั้นผลการทดลองพบว่า ปัจจัยที่มีนัยสำคัญทางสถิติได้แก่ อายุ ระยะของโรค ขนาดเซลล์มะเร็ง กลุ่มเซลล์มะเร็ง และน้ำหนักของผู้ป่วย ซึ่งสอดคล้องกับงานวิจัยของ สุพจน์และคณะ [12] และงานวิจัยของวรชัย [22] นอกจากนี้จากงานวิจัยของวรชัย [22] พบว่า ภาวะโลหิตจางหรือระดับค่าเฉลี่ยของฮีโมโกลบิน (Hbaverage)

เป็นปัจจัยหนึ่งที่มีผลต่อการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาโดยการฉายรังสี ดังนั้นในงานวิจัยนี้จึงได้เพิ่มปัจจัยดังกล่าวเป็นปัจจัยนำเข้าตัวแบบดังปรากฏในตารางที่ 3

3. การสร้างตัวแบบโครงข่ายประสาทเทียม

ในงานวิจัยนี้ใช้โครงข่ายประสาทเทียมแบบเพอร์เซปตรอนหลายชั้น มีอินพุต 6 ปัจจัย ทดลองด้วยวิธี K-Fold Cross Validation ปรับค่าตั้งแต่ 3K ถึง 10K ปรับค่าโมเมนตัม (Momentum) ค่าอัตราการเรียนรู้ (Learning Rate) ตั้งแต่ 0.1-0.9 ปรับเวลาฝึกสอน (Training Time) และ Hidden node จนกระทั่งได้ค่าพารามิเตอร์ที่ให้ค่าของการพยากรณ์ที่แม่นยำโดยมีค่าความผิดพลาดน้อยที่สุดในการพิจารณาจำนวนโหนดในชั้นซ่อน [16] มีดังนี้

- 1) จำนวนโหนดในชั้นซ่อนควรจะมีค่าในช่วงจำนวนโหนดชั้นอินพุตและจำนวนโหนดชั้นเอาต์พุต
- 2) จำนวนโหนดในชั้นซ่อนควรมีจำนวนสองในสาม (2/3) ของจำนวนโหนดชั้นอินพุตรวมกับจำนวนโหนดชั้นเอาต์พุต
- 3) จำนวนโหนดในชั้นซ่อนควรมีจำนวนไม่เกินสองเท่าของจำนวนโหนดในชั้นอินพุต

จากการทดลองพบว่าค่าพารามิเตอร์ที่เหมาะสมสามารถแสดงในตารางที่ 4

4. สร้างตัวแบบการทำนายด้วยวิธีการถดถอยโลจิสติก

การสร้างตัวแบบการทำนายด้วยวิธีการถดถอยโลจิสติก กำหนดตัวแปรดังต่อไปนี้ ระยะของโรค อายุ กลุ่ม เซลล์มะเร็ง น้ำหนักตัว ขนาดเซลล์มะเร็ง และระดับฮีโมโกลบิน เป็นตัวแปรอิสระ (Independent) และกำหนดตัวแปรผลการทำนาย (overall) เป็นตัวแปรตาม (Dependent) สร้างสมการโลจิสติกตรวจสอบเงื่อนไขการวิเคราะห์ความถดถอยโลจิสติกโดยพิจารณาค่าความน่าจะเป็นที่ใช้เป็น Cutting Score เท่ากับ 0.5 พิจารณาดังนี้

- $P(\text{เกิดเหตุการณ์}) < 0.5$ จะได้ $Y = 0$ หรือไม่เกิดเหตุการณ์

- $P(\text{เกิดเหตุการณ์}) \geq 0.5$ จะได้ $Y = 1$ หรือเกิดเหตุการณ์

ผลการทดสอบประสิทธิภาพของตัวแบบโครงข่ายประสาทเทียมและวิธีถดถอยโลจิสติกสามารถแสดงในตารางที่ 5 ซึ่งพบว่าผลการทำนายของแบบจำลองโครงข่ายประสาทเทียม และวิธีการถดถอยแบบลอจิสติก อยู่ในระดับที่สูงมาก แต่เมื่อพิจารณาค่าความจำเพาะ พบว่ามีค่าที่ต่ำมาก (0.0%) เหมือนกันทั้ง 2 ตัวแบบ กล่าวคือตัวแบบทั้งสองไม่สามารถทำนายคลาสที่เป็นลบได้ถูกต้องเลย ซึ่งจากการวิเคราะห์ข้อมูลพบว่าปัญหาเกิดจากความไม่สมดุลของข้อมูล (Imbalance data) โดยมีข้อมูลในคลาสหนึ่งมากกว่าอีกคลาสมาก ดังนั้นผู้วิจัยจึงได้ศึกษาและทำการปรับความไม่สมดุลของข้อมูลก่อนนำมาทำการสร้างโมเดลใหม่

5. การปรับความไม่สมดุลของข้อมูล

ในงานวิจัยนี้ได้ทำการปรับความไม่สมดุลของข้อมูลด้วยวิธี CSL และวิธี SMOTE

การปรับความไม่สมดุลด้วยวิธี SMOTE ผู้วิจัยได้ปรับค่าการเพิ่มจำนวนชุดข้อมูลที่มีจำนวนคลาสน้อย โดยทำการกำหนดค่าพารามิเตอร์ของตัวแบบให้คงที่ และเริ่มเพิ่มขนาดของข้อมูลไปจนกว่าจะได้ค่าประสิทธิภาพที่ดีที่สุด ผลการทดลองพบว่าขนาดชุดข้อมูลที่เหมาะสมคือ 600%

การปรับความไม่สมดุลของข้อมูลด้วยวิธี CSL จะทำการกำหนดค่าพารามิเตอร์ของตัวแบบการทำนายให้คงที่ การทดลองนี้จะสร้างตาราง Cost matrix โดยการสุ่มปรับค่า False Negative (FN) และค่า False Positive (FP) ค่าที่สุ่มจะอยู่ระหว่าง 1.0 ถึง 10.0 จะปรับค่าจนกว่าค่าใช้จ่าย (cost) ลดลง

ผลการวิจัยและการอภิปรายผล

ในงานวิจัยนี้จึงได้นำวิธีการ CSL และ SMOTE มาทำการปรับสมดุลของข้อมูลแล้วทำการสร้างแบบจำลองด้วยโครงข่ายประสาทเทียมเปรียบเทียบกับวิธีถดถอยโลจิสติก และทดลองด้วยวิธี K-Fold Cross Validation สามารถสรุปผลได้ดังตารางที่ 6

ผลการทดลองพบว่า โครงข่ายประสาทเทียมที่มีการปรับความสมดุลของข้อมูลด้วยวิธีการ SMOTE มีประสิทธิภาพในการทำนายที่ดีกว่าวิธีอื่น โดยมีค่าความถูกต้องสูงที่สุดอยู่ที่ 5K เท่ากับ 81.70% ค่าความไว หรือ True Positive Rate เท่ากับ 94.47% และ ค่าความจำเพาะ หรือ True Negative Rate เท่ากับ 55.47% จากตัวแบบที่ได้ในตารางที่ 6 ผู้วิจัยได้นำตัวแบบมาทำนายผู้ป่วยที่เข้ารับการรักษาในปี พ.ศ. 2552 จำนวน 129 คน พบว่า โครงข่ายประสาทเทียมที่ปรับความไม่สมดุลด้วยวิธี SMOTE มีค่าความถูกต้องสูงที่สุดเท่ากับ 96.12% ค่าความไวและค่าความจำเพาะเท่ากับ 97.62% และ 33.33% ตามลำดับ ดังแสดงในตารางที่ 7 ส่วนของตัวแบบโลจิสติกที่ปรับความไม่สมดุลด้วยวิธี SMOTE มีค่าความถูกต้องอยู่ในระดับที่สูง เมื่อเทียบกับวิธีโครงข่ายประสาทเทียมก็ยังมีประสิทธิภาพการทำนายต่ำกว่า ซึ่งสอดคล้องกับงานวิจัยของ Padmavathi [17]

สรุปผลการวิจัย

ในงานวิจัยนี้ได้ศึกษามีวัตถุประสงค์เพื่อสร้างตัวแบบในการทำนายผลการรักษาผู้ป่วยมะเร็งปากมดลูกที่เข้ารับการรักษาด้วยวิธีการฉายรังสี จากผลการวิจัยสามารถสรุปได้ว่า ตัวแบบโครงข่ายประสาทเทียมที่สร้างจากข้อมูลผู้ป่วยที่ได้ทำการปรับความไม่สมดุลของข้อมูล (Imbalance Data) ด้วยวิธี SMOTE มีตัวแปรอินพุตได้จากการวิเคราะห์คัดเลือกตัวแปรแบบลำดับขั้นและศึกษาจากงานวิจัยทางการแพทย์ มีทั้งหมด 6 ตัวแปรได้แก่ อายุ ระยะของโรค ขนาดของเซลล์มะเร็ง กลุ่มเซลล์มะเร็ง ปากมดลูก น้ำหนักของผู้ป่วย และระดับค่าเฉลี่ยของฮีโมโกลบิน จะมีประสิทธิภาพในการทำนายดีกว่าวิธีอื่น โดยมีค่าความถูกต้องเท่ากับ 81.70% เมื่อนำมาทดสอบกับชุดข้อมูลทดสอบจำนวน 129 ชุดข้อมูล พบว่ามีค่าความถูกต้อง 96.12% ค่าความไวเท่ากับ 97.62% และค่าความจำเพาะเท่ากับ 33.33%

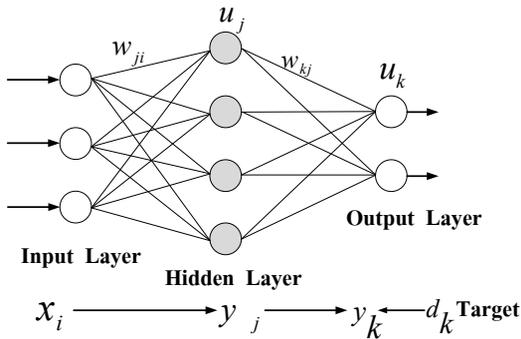
กิตติกรรมประกาศ

งานวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยความกรุณาอย่างยิ่งจาก หน่วยรังสีรักษา ภาควิชารังสีวิทยา คณะแพทยศาสตร์โรงพยาบาลศรีนครินทร์ มหาวิทยาลัยขอนแก่น ที่ให้อนุเคราะห์ข้อมูลสำหรับการทำวิจัยในครั้งนี้ และสนับสนุนและส่งเสริมงานวิจัยนี้ให้สำเร็จลุล่วงได้ตามวัตถุประสงค์ รวมทั้งคำแนะนำที่เป็นประโยชน์ในการทำวิจัย

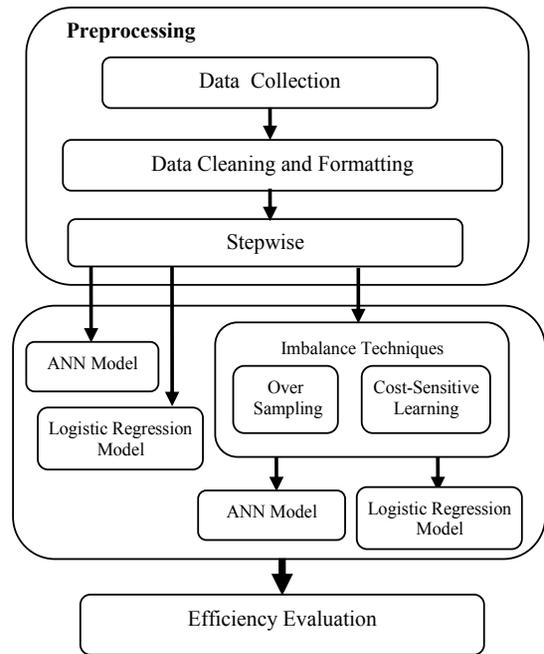
เอกสารอ้างอิง

- Asakit T. The preliminary results of High-Dose-Rate intracavitary brachytherapy with small-sized cobalt-60 radioisotope in carcinoma of cervix uteri: The first results of Thailand from Lampang regional cancer center. *Journal of Thai Society of Therapeutic Radiology and Oncology*. 2011; 17(1): 13-24. Thai.
- Botoa C, Bardan R, Botoca M, Alexz F. Prostate cancer prognosis evaluation assisted by neural networks. *WSEAS Transactions on Computers*. 2010; 2(9): 164-173.
- Chang SW, Kareem SA, Kallarakal TG, Merican AF, Abraham MT, Zain RB. Feature selection methods for optimizing clinicopathologic input variables in oral cancer prognosis. *Asian Pacific Journal of Cancer Prevention*. 2011; 12(10):2659-2664.
- Chowdhury DR, Chatterjee M, Samanta RK. An artificial neural network model for neonatal disease diagnosis. *International Journal of Artificial Intelligence and Expert Systems (IJAE)*. 2011; 2(3): 96-106.
- Elkan C. The foundation of cost-sensitive learning. *Proceedings of IJCAI*. 2011; 17: 973-978.
- Farhangfar A, Kurgan LA, Pedryez W. A Novel Framework for imputation of missing values in databases. *IEEE Transactions on systems, Man, and Cybernetics-Part A: Systems and Humans*. 2007; 37(5): 692-709.
- Fernandez A, Garcia S, Herrera F. Addressing the classification with imbalanced data: Open problem and new challenges on class distribution. *HAIS'11 Proceedings of the 6th international conference on Hybrid artificial intelligent systems*. 2011; 1: 1-10.
- Han J, Kamber M. *Data mining concepts and techniques*. 2nd ed. United States of America: Morgan Kaufmann Publishers. 2006.
- Haykin S. *Neural networks*. 2nd ed. United States of America: United States of America. 1999.
- He H, Garcia EA. Learning from Imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*. 2009; 21(9):1263-1284.
- Jarunratanakul C. Prediction of compressor performance map by using neural network and beta line [M Eng thesis]. Bangkok: Kasetsart University; 2009. Thai.
- Kamsaard S, Tangvorapongchai V, Krusun S, Sriamporn S, Suwanrungruang K. A prognostic model for prediction the immediate result in cervical cancer by radiotherapy; 2007. Thai.

13. Kukar M, Kononenko I. Cost-sensitive learning with neural networks. Proceedings of the 13th European Conference on Artificial Intelligence. 1998; 445-449.
14. Nguyen GH, Bouzerdoum A, Phung SL. Learning pattern classification tasks with imbalance data sets. In: Pattern recognition, Yin, P.Y., editors. 2009. p.193-207.
15. Nitesh VC, Kevin WB, Lawrence OH, W PK. SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research. 2006; 16: 321-357.
16. Panchal G, Ganatra A, Kosta YP, Panchal D. Behaviour analysis of multilayer perceptrons with multiple hidden neurons and hidden layers. International Journals of Computer Theory and Engineering. 2011; 3(2): 332-337.
17. Padmavathi J. A comparative study on breast cancer prediction using RBF and MLP. International Journal of Scientific & Engineering Research. 2011; 2(1): 1-5.
18. Pongwichai S. Statical analysis of data by computer. 16th ed. Bangkok: V. Print Publisher. 2009. Thai.
19. Sandberg IW, Lo JT, Fancourt CL, Principe JC, Katagiri S, *et al.* Nonlinear Dynamical systems feedforward neural network perspectives. United Kingdom:Canada. 2001.
20. Su X, Khoshgoftaar TM, Greiner R. Using imputation techniques to help learn accurate classifiers. International Conference on Tools with Artificial Intelligence. 2008; 20: 437-444.
21. Suwanarattaphoom P. An approach for improving associative classification in imbalanced datasets [M Eng Thesis]. Bangkok: Kasetsart University; 2010.Thai.
22. Tangvorapongchai V. Factor influencing the responsiveness in radiotherapy of cervical cancer. Thai Cancer Journal. 1987; 13(1): 11-15. Thai.
23. Vanichbuncha K. Advanced analysis using SPSS for Windows. 5th ed. Bangkok: Thammasarn co.,Ltd Publisher. 2006. Thai.
24. Winter WE III, Maxwell GL, Tian C, Sobel E, Rose GS, Thomas G, *et al.* Association of hemoglobin level with survival in cervical carcinoma patients treated with concurrent cisplatin and radiotherapy: a gynecologic oncology group study. Gynecologic Oncology. 2004; 94(2): 495-501.
25. Yan H, Jiang Y, Zheng J, Peng C, Li Q. A multilayer perceptron-based medical decision support system for heart disease diagnosis. Expert Systems with Applications. 2006; 30(2): 272-281.
26. Zhao H. Instance weighting versus threshold adjusting for cost-sensitive classification. Knowledge and Information Systems. 2008; 15(3): 321-324.
27. Zhou ZH, Liu XY. Training Cost-Sensitive Neural networks with methods addressing the class imbalance problem. IEEE Transactions on Knowledge and Data Engineering. 2006; 18(1): 63-77.



ภาพที่ 1 สถาปัตยกรรมโครงข่ายประสาทเทียม [2]



ภาพที่ 2 วิธีการดำเนินการ

ตารางที่ 1 ตาราง Confusion Matrix

ค่าความจริง (Actual)	ค่าทำนาย (Predicted)	
	บวก (Positive)	ลบ (Negative)
บวก (Positive)	True positive (TP)	False negative (FN)
ลบ (Negative)	False positive (FP)	True negative (TN)

ตารางที่ 2 ปัจจัยที่เกี่ยวข้องกับป่วยมะเร็งปากมดลูก

ลำดับ	ปัจจัยอินพุต
1	อายุ (Age)
2	แพทย์ผู้ทำการรักษา (Doctor)
3	กลุ่มเซลล์มะเร็งปากมดลูก (PathoGroup)
4	เซลล์มะเร็งกลุ่มย่อย (M_code)
5	เซลล์ประเภท (Squamous cell carcinoma:SCC)
6	ระดับความตุงของเซลล์มะเร็ง (Grade_DIS)
7	ระยะของโรค (StagingGrouping)
8	จำนวนการตั้งครรภ์ (Gravidity)
9	จำนวนของการคลอดบุตร (Parity)

ลำดับ	ปัจจัยอินพุต
10	จำนวนการแท้งบุตร (Abortion)
11	ระดับค่าเฉลี่ยของฮีโมโกลบิน (Hbaverage)
12	กลุ่มจังหวัดภูมิภาค (Provice_Grouping)
13	ขนาดของเซลล์มะเร็ง (Tumor_Size)
14	น้ำหนักของผู้ป่วย (Body_Weigh)
15	ช่วงอายุที่เป็นมะเร็งปากมดลูก (Mentru)
16	เป้าหมายการรักษา (Aim_Treat)
17	ลักษณะการลุกลามของเซลล์มะเร็ง (Center)

ตารางที่ 3 ผลการวิเคราะห์โดยการคัดเลือกตัวแปรแบบลำดับขั้น

ลำดับ	ปัจจัย
1	อายุ (Age)
2	ระยะของโรค (StagingGrouping)
3	ขนาดของเซลล์มะเร็ง (Tumor_Size)
4	กลุ่มเซลล์มะเร็งปากมดลูก (PathoGroup)
5	น้ำหนักของผู้ป่วย (Body_Weigh)
6	ระดับค่าเฉลี่ยของฮีโมโกลบิน (Hbaverage)

ตารางที่ 4 ค่าพารามิเตอร์ที่มีความผิดพลาดน้อยที่สุด

การกำหนดค่าพารามิเตอร์	ANN+CSL	ANN+SMOT
จำนวน Input	6	6
จำนวน Output	1	1
จำนวนชั้นซ่อน	7	7
ค่าอัตราการเรียนรู้	0.1	0.1
ค่าโมเมนตัม	0.2	0.1
จำนวนรอบ	2500	5000

ตารางที่ 5 ประสิทธิภาพการทำนายด้วยโครงข่ายประสาทเทียม (ANN) และวิธีถดถอยโลจิสติก (Logistic)

Method	K-Fold	TP (%)	TN (%)	Accuracy (%)
ANN	ทุก K	100.00	0.00	93.50
Logistic	ทุก K	100.00	0.00	93.50

ตารางที่ 6 ประสิทธิภาพการทำนายด้วยโครงข่ายประสาทเทียม (ANN) เปรียบเทียบกับวิธีถดถอยโลจิสติก (Logistic) หลังจากปรับความไม่สมดุลของข้อมูลด้วยวิธี CSL และ SMOTE

Method	K-Fold	TP (%)	TN (%)	Accuracy (%)
ANN + CSL	4K	76.70	40.88	74.37
ANN + SMOTE	5K	94.47	55.47	81.70
Logistic + CSL	4K	84.52	30.66	81.02
Logistic + SMOTE	5K	90.71	16.27	66.34

ตารางที่ 7 ประสิทธิภาพการทำนายผลการรักษา

Method	TP (%)	TN (%)	Accuracy (%)
ANN + CSL	98.41	0.00	96.12
ANN + SMOTE	97.62	33.33	96.12
Logistic + CSL	87.30	66.67	86.82
Logistic + SMOTE	93.65	33.33	92.25