

ภาคผนวก

บทความสำหรับการเผยแพร่

การวิเคราะห์ความเสี่ยงของการเกิดโรคมะเร็งด้วยโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่
แผ่รัศมีและโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

Analysis Risk for Cancer Incident Using Radial Basis Function Neural Network
and Multi-Layer Perceptron

อัจฉรา มหาวีรวัฒน์

ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์, คณะวิทยาศาสตร์, มหาวิทยาลัยอุบลราชธานี

scatchma@ubu.ac.th

บทคัดย่อ

ปัจจุบันผู้ป่วยมะเร็งมีแนวโน้มเพิ่มจำนวนขึ้นเรื่อยๆ เนื่องจากปัจจัยเสี่ยงที่ก่อให้เกิดโรคมะเร็งมากขึ้น การตรวจวินิจฉัยโรคมะเร็งมีกระบวนการและขั้นตอนที่ซับซ้อนโดยส่วนใหญ่เป็นการตรวจทางห้องปฏิบัติการซึ่งต้องใช้ระยะเวลาและค่าใช้จ่ายสูง ส่งผลให้ผู้ป่วยชะลอการตัดสินใจเพื่อเข้ารับการตรวจวิเคราะห์จนกว่าจะพบความผิดปกติในร่างกายตนเองอย่างเด่นชัดซึ่งมักอยู่ในระยะที่อันตรายและรักษาให้หายขาดได้ยาก

ดังนั้นงานวิจัยนี้จึงได้นำเสนอการประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Neural Network) และโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptrons) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) เพื่อวิเคราะห์การเป็นโรคมะเร็งและระบุชนิดโรคมะเร็ง โดยชนิดมะเร็งที่สนใจคือ มะเร็งปอด มะเร็งตับ มะเร็งเต้านม และมะเร็งลำไส้ใหญ่ จากผลการวิเคราะห์พบว่าวิธีการนี้สามารถวิเคราะห์การเป็นโรคมะเร็งและชนิดมะเร็งได้ถูกต้องมากกว่าร้อยละ 90 ซึ่งสามารถนำไปใช้ประกอบการตัดสินใจของผู้ป่วยเพื่อเข้ารับการตรวจรักษาได้

คำสำคัญ : โครงข่ายประสาทเทียม, โรคมะเร็ง, โครงข่ายเพอร์เซปตรอนแบบหลายชั้น, โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี

Abstract

In current day, number of cancer patients is growing due to the increasing of cancer risk factors. Cancer diagnosis is based on laboratory test, so cancer patients delay in taking medical examination or treatment. Consequently, most cancer patients realize their illness during dangerous phase of disease which is difficult to be completely cured.

Therefore, this research introduces an approach applying Radial Basis Function Neural Network (RBFN) and Multi-Layer Perceptron (MLP) with Back-Propagation Learning Algorithm to analyze cancer incidence in 4 cancer types: lung cancer, liver cancer, breast cancer and colon cancer. From the experimental results, the approach yields more than 90 percentage of accuracy and could be a decision support tool for patient to take medical examination or treatment.

Key Words : Artificial intelligence, cancer, Multi-Layer Perceptrons, Radial Basis Function Neural Network

1. บทนำ

ปัจจุบันผู้ป่วยมะเร็งมีแนวโน้มเพิ่มจำนวนขึ้นเรื่อยๆ เนื่องจากปัจจัยเสี่ยงที่ก่อให้เกิดโรคมะเร็งมากขึ้น ซึ่งเป็นปัจจัยทั้งจากสิ่งแวดล้อมภายนอกในร่างกายและความผิดปกติภายในร่างกาย การตรวจวินิจฉัยโรคมะเร็งมีกระบวนการและขั้นตอนที่ซับซ้อนโดยส่วนใหญ่เป็นการตรวจทางห้องปฏิบัติการซึ่งต้องใช้ระยะเวลานานและค่าใช้จ่ายสูง ส่งผลให้ผู้ป่วยชะลอการตัดสินใจเพื่อเข้ารับการตรวจวิเคราะห์จนกว่าจะพบความผิดปกติในร่างกายตนเองอย่างเด่นชัดซึ่งมักอยู่ในระยะที่อันตรายและรักษาให้หายขาดได้ยาก

ดังนั้นงานวิจัยนี้จึงได้นำเสนอวิธีการจำแนกข้อมูลผู้ป่วยเพื่อวิเคราะห์ความเสี่ยงของการเกิดโรคมะเร็งโดยประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมซึ่งเป็นศาสตร์ในสาขาวิทยาการคอมพิวเตอร์เพื่อประมวลผลข้อมูลเบื้องต้นของผู้ป่วยซึ่งได้แก่ อาการ พฤติกรรม ประวัติการเป็นมะเร็ง เพศ และอายุ โดยให้ผลลัพธ์ว่าผู้ป่วยมีแนวโน้มที่จะเป็นโรคมะเร็งหรือไม่

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 ความรู้พื้นฐานเกี่ยวกับโรคมะเร็ง

มะเร็ง คือ กลุ่มของโรคที่เกิดเนื่องจากเซลล์ของร่างกายมีความผิดปกติที่สารพันธุกรรม ส่งผลให้เซลล์มีการเจริญเติบโตและแบ่งตัวเพื่อเพิ่มจำนวนเซลล์ที่รวดเร็วและมากกว่าปกติ [14] จึงอาจทำให้เกิดก้อนเนื้อผิดปกติและในที่สุดจะเกิดการตายของเซลล์ในก้อนเนื้อนั้นเนื่องจากขาดเลือดไปเลี้ยง ถ้าเซลล์พวกนี้เกิดอยู่ในอวัยวะใดจะเรียกชื่อมะเร็งตามอวัยวะนั้น เช่น มะเร็งปอด มะเร็งเต้านม เป็นต้น โดยมะเร็งที่พบในร่างกายมนุษย์มีหลายชนิด แต่ละชนิดมีการดำเนินของโรคที่ต่างกันเช่น มะเร็งปอดและมะเร็งสมองมีการดำเนินของโรคที่รุนแรงผู้ป่วยมีชีวิตการอยู่รอดสั้นกว่าผู้ป่วยมะเร็งผิวหนังและมะเร็งชนิดอื่นๆ ดังนั้นวิธีการรักษามะเร็งแต่ละชนิดจึงต่างกัน ทั้งนี้ขึ้นอยู่กับอวัยวะที่เป็นมะเร็ง ระยะของมะเร็ง สภาพร่างกายและความเหมาะสมของผู้ป่วย รวมทั้งการดำเนินของโรคมะเร็ง

สาเหตุและปัจจัยเสี่ยงของการเกิดมะเร็งประกอบด้วยปัจจัยที่เกิดจากสิ่งแวดล้อมภายนอกในร่างกาย เช่น สารก่อมะเร็งที่ปนเปื้อนในอาหาร และปัจจัยที่เกิดจากความผิดปกติภายในร่างกาย เช่น ความผิดปกติทางพันธุกรรม เป็นต้น การรักษามะเร็งเป็นแบบผสมผสานหลายวิธีได้แก่ ศัลยกรรมคือการผ่าตัดเอาก้อนมะเร็งออก รังสีรักษา คือการฉายแสงบริเวณที่มีเซลล์มะเร็งอยู่ และเคมีบำบัดคือ การรักษาหรือการทำลายเซลล์มะเร็งทั้งที่ต้นตอและที่กระจายไปตามทางเดินน้ำเหลือง กระแสเลือดหรืออวัยวะอื่นของร่างกาย โดยการรับประทานยาที่มีความสามารถในการฆ่าหรือทำลาย เซลล์มะเร็ง ฉีดยาทางหลอดเลือดดำหรือแดง เป็นต้น

2.2 โครงข่ายประสาทเทียม (Artificial Neural Networks)

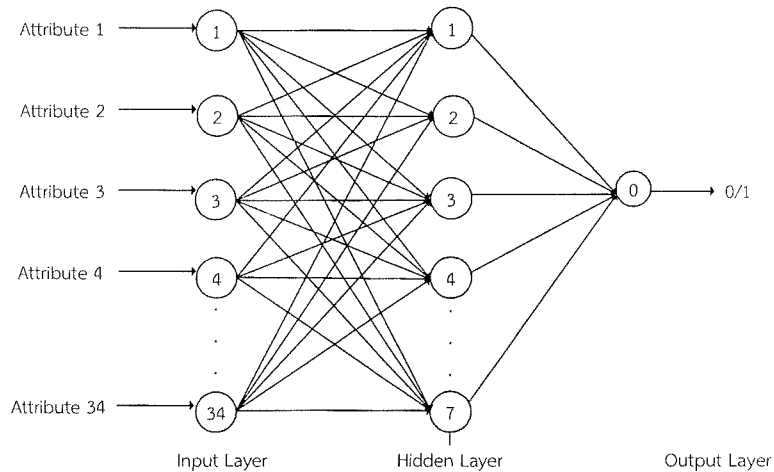
โครงข่ายประสาทเทียม (Artificial neural network) หรือมักเรียกสั้น ๆ ว่าข่ายงานประสาท (neural network หรือ neural net) คือโมเดลทางคณิตศาสตร์ สำหรับประมวลผลสารสนเทศด้วยการคำนวณแบบคอนเนคชันนิสต์ (connectionist) เพื่อจำลองการทำงานของเครือข่ายประสาทในสมองมนุษย์สำหรับการประมวลผลทางคอมพิวเตอร์ หน่วยประสาทในคอมพิวเตอร์

▪ โครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layers Perceptrons)

โครงข่ายเพอร์เซปตรอนแบบหลายชั้นคือโครงข่ายประสาทเทียมที่มีโครงสร้างเป็นแบบหลาย ดังนี้

- ชั้นนำเข้า (Input Layer) ทำหน้าที่รับข้อมูลจากภายนอกแล้วส่งต่อไปยังชั้นซ่อน (Hidden Layer)
- ชั้นซ่อน (Hidden Layer) ทำหน้าที่ประมวลผลข้อมูลที่ได้จากชั้นนำเข้าแล้วส่งต่อไปยังชั้นผลลัพธ์ (Output Layer)
- ชั้นผลลัพธ์ (Output Layer) ทำหน้าที่ประมวลผลครั้งสุดท้ายแล้วแสดงเป็นผลลัพธ์ ของโครงข่าย

โครงข่ายเพอร์เซปตรอนแบบหลายชั้นสามารถนำมาแก้ปัญหาที่ซับซ้อนมากได้ ซึ่งความสามารถและประสิทธิภาพของโครงข่ายขึ้นอยู่กับวิธีการเลือกรูปแบบการต่อโครงข่าย วิธีการฝึกสอน และการปรับค่าน้ำหนัก(Weight) ของแต่ละโครงข่าย ตัวอย่างโครงสร้างของโครงข่ายดังแสดงในรูปที่ 1



รูปที่1 โครงสร้างโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

▪ ขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm)

ขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back propagation Learning Algorithm) เป็นวิธีการเรียนรู้แบบมีการสอน (Supervised Learning) โดยชุดข้อมูลที่ใช้สอนโครงข่ายจะมีค่าตอบอยู่แล้วเพื่อให้โครงข่ายปรับตัวให้ได้คำตอบที่ดีที่สุด ขั้นตอนวิธีนี้จะถูกใช้ในการเรียนรู้ของโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptron) ซึ่งโหนดในแต่ละชั้นจะถูกเชื่อมต่อกับค่าน้ำหนัก (weight) ที่แตกต่างกัน เมื่อป้อนชุดข้อมูลเข้าไปในชั้นนำเข้า (input layer) จะมีการคำนวณแล้วส่งผ่านผลลัพธ์ที่ได้ไปยังชั้นซ่อน (hidden layer) จนถึงชั้นข้อมูลผลลัพธ์ (output layer) ซึ่งในขั้นนี้จะคำนวณค่าคลาดเคลื่อนของผลลัพธ์และปรับค่าน้ำหนักใหม่แล้วส่งกลับไปยังชั้นก่อนหน้าจนถึงชั้นนำเข้า และทำซ้ำจนกระทั่งได้ค่าคลาดเคลื่อนที่ยอมรับได้ [7] โดยการคำนวณผลลัพธ์จากชั้นซ่อนและชั้นผลลัพธ์จะใช้ฟังก์ชันแปลงถ่ายทอดข้อมูล (Transfer function) หรือฟังก์ชันกระตุ้น (Activation function) ซึ่งมีหลายรูปแบบให้เลือกใช้ เช่น ซิกมอยด์ฟังก์ชัน (Sigmoid function)

ขั้นตอนวิธีในการคำนวณโดยเริ่มจากป้อนชุดข้อมูลนำเข้าเข้าสู่โครงข่าย ข้อมูลนำเข้าทั้งหมดจะถูกคูณด้วยค่าน้ำหนักและนำมารวมกันดังสมการที่ (1)

$$h = \sum_{i=0}^n x_i w_i \tag{1}$$

โดยที่ h เป็นค่าผลรวมของผลคูณระหว่างค่าข้อมูลนำเข้าจากชั้นนำเข้าและค่าน้ำหนัก x_i เป็นค่าของข้อมูลนำเข้าตำแหน่งที่ i โดย x_0 คือค่าเอนเอียง (bias) ซึ่งมีค่าเป็น 1 w_i เป็นค่าน้ำหนักตำแหน่งที่ i

จากนั้นโครงข่ายจะนำค่าผลรวมที่ได้ (h) มาผ่านฟังก์ชันแปลงถ่ายทอดข้อมูลแบบซิกมอยด์ดังสมการที่ (2)

$$O = g(h) = \frac{1}{1 + e^{-h}} \tag{2}$$

โดยที่ h เป็นค่าผลรวมของผลคูณระหว่างค่าข้อมูลนำเข้าจากชั้นนำเข้าและค่าน้ำหนัก

O เป็นผลลัพธ์ของชั้นซ่อน

เมื่อการคำนวณที่ชั้นซ่อนเสร็จสิ้นแล้ว ผลลัพธ์ที่ได้จะเป็นข้อมูลนำเข้าสำหรับชั้นผลลัพธ์เพื่อหาค่าคำตอบสุดท้ายซึ่งเป็นผลลัพธ์ของโครงข่าย และเมื่อข้อมูลซึ่งเป็นผลลัพธ์จากชั้นซ่อนถูกส่งมาถึงชั้นผลลัพธ์ก็จะเข้าสู่กระบวนการคำนวณเช่นเดียวกันกับในชั้นซ่อน จากนั้นจึงเปรียบเทียบค่าผลลัพธ์ที่ได้ (actual output) กับผลลัพธ์เป้าหมาย (target) ถ้าค่าต่างกันจะเข้าสู่กระบวนการ Backward Computation ซึ่งเป็นการทำงานลักษณะแพร่ย้อนกลับเพื่อหาค่าความผิดพลาดของผลลัพธ์ โดยจะมีการคำนวณสองส่วนคือช่วงชั้นซ่อนกับชั้นผลลัพธ์ และช่วงชั้นนำเข้ากับชั้นซ่อนดังสมการต่อไปนี้ [10]

$$\delta_k = O_k(1 - O_k)(T_k - O_k) \quad (3)$$

$$\delta_l = O_l(1 - O_l) \left(\sum_{k \in O} w_{kl} \delta_k \right) \quad (4)$$

กำหนดให้

δ_k คือ ค่าความผิดพลาดของผลลัพธ์ที่คำนวณจากชั้นผลลัพธ์

δ_l คือ ค่าความผิดพลาดของผลลัพธ์ที่คำนวณจากชั้นซ่อน

O คือ ค่าผลลัพธ์จากโครงข่าย

T_k คือ ค่าผลลัพธ์เป้าหมาย

w คือ ค่าน้ำหนัก

k เป็นดัชนีของโหนดในชั้นผลลัพธ์

l เป็นดัชนีของโหนดในชั้นซ่อน

เมื่อผ่านขั้นตอน Backward Computation แล้วจะทำการปรับค่าน้ำหนัก (Weight Adjustment) ขั้นตอนนี้เป็นการทำงานเพื่อปรับปรุงค่าน้ำหนักให้ได้ค่าน้ำหนักที่เหมาะสมสำหรับคำนวณผลลัพธ์ของโครงข่ายที่ต้องการและคำนวณหาความผิดพลาดของโครงข่ายโดยดำเนินการกับค่าน้ำหนักที่อยู่ระหว่างชั้นซ่อนกับชั้นผลลัพธ์ และระหว่างชั้นนำเข้ากับชั้นซ่อนตามลำดับดังสมการที่ (5)

$$w_{ji} = w_{old} + \Delta w_{ji} \quad (5)$$

กำหนดให้

w_{ji} คือค่าน้ำหนักใหม่

w_{old} คือค่าน้ำหนักเดิม

Δw_{ji} คือส่วนต่างของค่าน้ำหนัก

$$\Delta w_{ji} = \eta \delta_j x_{ji} \quad (6)$$

กำหนดให้

η คือ ค่าอัตราการฝึกสอน (Learning rate)

δ_j คือ ค่าความผิดพลาดของผลลัพธ์ที่คำนวณจากชั้นผลลัพธ์สำหรับกรณีที่ปรับค่าน้ำหนักระหว่าง ชั้นซ่อนกับชั้นผลลัพธ์ หรือค่าความผิดพลาดของผลลัพธ์ที่คำนวณจากชั้นซ่อนสำหรับกรณีที่ปรับค่าน้ำหนักระหว่างชั้นนำเข้ากับชั้นซ่อน

x_{ji} คือ ค่าข้อมูลเข้าของชั้นนำเข้าสำหรับกรณีที่ปรับค่าน้ำหนักระหว่างชั้นนำเข้าและชั้นซ่อน หรือค่าที่คำนวณได้ซึ่งเป็นผลลัพธ์ของชั้นซ่อนสำหรับกรณีที่ปรับค่าน้ำหนักระหว่างชั้นซ่อนและชั้นผลลัพธ์

จากนั้นคำนวณหาค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean square error) ของโครงข่ายโดยใช้สมการดังนี้

$$E = \frac{1}{2N} \sum_p \sum_k (T_k - O_k)^2 \quad (7)$$

กำหนดให้

- E คือค่าความผิดพลาดของโมเดล,
- N คือจำนวนรูปแบบ (pattern) หรือข้อมูลที่ผ่านเข้าโครงข่าย
- T_k คือค่าเป้าหมาย
- O_k คือผลลัพธ์ที่คำนวณได้จากชั้นผลลัพธ์
- k เป็นดัชนีของโหนดในชั้นผลลัพธ์
- p คือดัชนีระบอบที่นำรูปแบบหรือข้อมูลเข้าโครงข่าย

เมื่อกระบวนการฝึกสอนโครงข่ายดำเนินไปจนกระทั่งค่าความผิดพลาดของโครงข่ายน้อยกว่าหรือเท่ากับเกณฑ์ที่กำหนดแล้ว จะแปลงค่าผลลัพธ์จากชั้นผลลัพธ์เป็นค่าคำตอบของโครงข่ายในรูปของ 0 และ 1 ด้วยฟังก์ชันการแปลงถ่ายเทแบบ Hard Limit ดังสมการที่ 8

$$f(x) = \begin{cases} 1 & \text{if } x \geq T \\ 0 & \text{if } x < T \end{cases} \quad (8)$$

กำหนดให้ x คือค่าผลลัพธ์ที่คำนวณได้จากชั้นผลลัพธ์ และ T คือค่าช่วงผลลัพธ์

▪ โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Network)

โครงสร้างของโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีมีลักษณะคล้ายกับโครงข่ายเพอร์เซปตรอนแบบหลายชั้น แต่แตกต่างตรงที่ไม่มีค่าน้ำหนักเชื่อมระหว่างชั้นนำเข้าและชั้นซ่อน มีเพียงค่าน้ำหนักที่เชื่อมระหว่างชั้นซ่อนและชั้นผลลัพธ์เท่านั้น โดยที่โหนดในชั้นซ่อนจะมีการคำนวณผลลัพธ์จากชั้นซ่อนด้วยฟังก์ชันฐานหลักที่แผ่รัศมีแล้วส่งต่อค่าที่คำนวณได้ไปยังชั้นผลลัพธ์เพื่อคำนวณผลลัพธ์ของโครงข่ายด้วยฟังก์ชันกระตุ้นโดยโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีนี้มี 2 ชนิดได้แก่

3. โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิด Regularization (Regularization Radial-Basis Function Network) ซึ่งโครงสร้างดังนี้

- ชั้นนำเข้า (Input layer) ประกอบด้วยโหนดนำเข้า (input node) ที่มีจำนวนเท่ากับจำนวนมิติ (dimension) m_0 ของเวกเตอร์ข้อมูลนำเข้า x
- ชั้นซ่อน (Hidden layer) ประกอบด้วยโหนดที่เชื่อมกับโหนดนำเข้า
- ชั้นผลลัพธ์ (Output layer) ประกอบด้วยโหนดที่เชื่อมกับโหนดในชั้นซ่อน

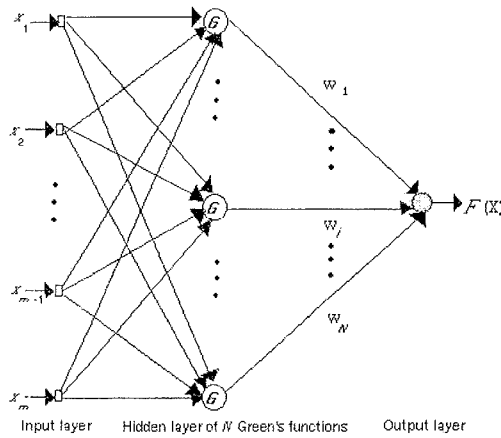
จำนวนโหนดในชั้นซ่อนเท่ากับจำนวนจุดข้อมูลในชุดข้อมูลฝึกสอน และใช้ฟังก์ชันกรีนเป็นฟังก์ชันกระตุ้นสำหรับโหนดในชั้นซ่อน ดังนั้นผลลัพธ์จากโหนดในชั้นซ่อนลำดับที่ i คือ $G(x, x_i)$ และโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีประเภทนี้เป็นการจำลองฟังก์ชันการประมาณค่าให้อยู่ในรูปแบบดังสมการที่ 9-10

$$F(x) = \sum_{i=1}^N w_i G(x, x_i) \quad (9)$$

$$F(x) = \sum_{i=1}^N w_i \exp\left(-\frac{\|x - x_i\|^2}{2\sigma_i^2}\right) \quad (10)$$

เมื่อ w_i คือค่าน้ำหนักที่เชื่อมจากโหนดในชั้นซ่อนลำดับที่ i ไปยังโหนดในชั้นผลลัพธ์

ตัวอย่างโครงสร้างโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิดRegularization ดังแสดงในรูปที่ 2



รูปที่ 2 โครงสร้างโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิดRegularization

4. โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิด Generalized (Generalized Radial-Basis Function Network)

จำนวนโหนดในชั้นซ่อนของโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิดRegularization เท่ากับจำนวนจุดข้อมูลในชุดข้อมูลฝึกสอนซึ่งถือว่าต้องใช้ระยะเวลาและทรัพยากรในการคำนวณมาก เพื่อแก้ไขข้อเสียดังกล่าวจึงต้องประมาณวิธีการแก้ปัญหาด้วยโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิด Regularization ด้วย $F^*(x)$ ดังสมการที่ 11

$$F^*(x) = \sum_{i=1}^{m_1} w_i \varphi_i(x) \tag{11}$$

เมื่อ $\{\varphi_i(x) | i = 1, 2, \dots, m_1\}$ คือชุดของฟังก์ชันฐานหลักที่แผ่รัศมีใหม่ โดยจำนวนฟังก์ชันที่ได้จะน้อยกว่าหรือเท่ากับจำนวนจุดข้อมูล ($m_1 \leq N$) และ w_i เป็นชุดค่าน้ำหนักใหม่

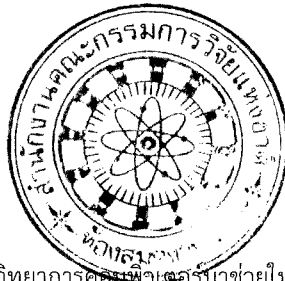
$$\varphi_i(x) = G(\|x - t_i\|), \quad i = 1, 2, \dots, m_1 \tag{12}$$

เมื่อ $\{t_i | i = 1, 2, \dots, m_1\}$ คือชุดของจุดศูนย์กลางซึ่งมีจำนวนน้อยกว่าหรือเท่ากับจำนวนจุดข้อมูล ซึ่งก็คือจำนวนโหนดในชั้นซ่อนโดยจะลดลงจากเดิมในโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีชนิด Regularization และสามารถเขียนในรูปแบบใหม่ดังสมการที่ 13

$$F^*(x) = \sum_{i=1}^{m_1} w_i G(x, t_i) = \sum_{i=1}^{m_1} \hat{w}_i G(\|x - t_i\|) \tag{13}$$

2.4 งานวิจัยที่เกี่ยวข้อง

การตรวจผู้ป่วยเพื่อวิเคราะห์โรคมะเร็งกระทำได้หลายวิธีเช่น การสอบถามประวัติโดยละเอียด การตรวจทางห้องปฏิบัติการ การตรวจเอ็กซเรย์ การตรวจทางเซลล์วิทยาและพยาธิวิทยา [11-13] เป็นต้น โดยในการตรวจด้วยวิธีการเหล่านี้ผู้ป่วยต้องเข้าพบแพทย์ผู้เชี่ยวชาญเท่านั้น ซึ่งผู้ป่วยอาจไม่สะดวกในการเข้าพบแพทย์และปล่อยให้อาการโรครุนแรงส่งผลให้การรักษามีให้หายขาดได้ยากหรืออาจถึงขั้นเสียชีวิต



ปัจจุบันมีการนำความรู้ทางด้านวิทยาการคอมพิวเตอร์มาช่วยในกระบวนการวิเคราะห์โรคเพื่อลดระยะเวลาการวินิจฉัยและคัดกรองผู้ป่วย เช่น การคัดเลือกยีนและจำแนกโรคมะเร็งด้วยโดยประยุกต์ใช้ตรรกศาสตร์คลุมเครือและโครงข่ายประสาทเทียม [1-5] การตรวจหามะเร็งเต้านมจากภาพด้านการแพทย์โดยใช้เทคนิคการประมวลผลภาพ (Image Processing) [10] การวิเคราะห์โรคตับโดยการคัดเลือกคุณลักษณะ (Feature Selection) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) และโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptron) ประมวลผลกับข้อมูลผลการตรวจจากห้องปฏิบัติการ [9] ซึ่งงานวิจัยเหล่านี้มักเลือกใช้เทคนิคโครงข่ายประสาทเทียมแบบมีผู้สอน (Supervised approach) และข้อมูลที่ใช้ในงานวิจัยได้มาจากการตรวจทางห้องปฏิบัติการซึ่งเป็นขั้นตอนที่ซับซ้อนและเสียค่าใช้จ่ายสูง

3. ระเบียบวิธีวิจัย / วิธีการ

3.1 การรวบรวมและเตรียมข้อมูล

3.1.1 กำหนดขนาดตัวอย่างข้อมูล

งานวิจัยนี้ได้สำรวจข้อมูลผู้ป่วยที่เข้ารับการตรวจรักษาที่โรงพยาบาลสรรพสิทธิประสงค์ จังหวัดอุบลราชธานี ในปี.ศ. 2552 ดังแสดงในตารางที่ 1

ตารางที่ 1 จำนวนข้อมูลตัวอย่างสำหรับวิเคราะห์การเป็นโรคมะเร็ง

| ชนิดโรค | จำนวนข้อมูลสำหรับวิเคราะห์การเป็นโรคมะเร็ง (คน) | จำนวนข้อมูลสำหรับวิเคราะห์ชนิดโรคมะเร็ง (คน) |
|-----------------|---|--|
| มะเร็งปอด | 60 | 203 |
| มะเร็งตับ | 150 | 329 |
| มะเร็งเต้านม | 90 | 263 |
| มะเร็งลำไส้ใหญ่ | 45 | 154 |
| โรคอื่น | 330 | - |
| รวม | 675 | 949 |

3.1.2 กำหนดคุณสมบัติข้อมูล

ข้อมูลที่รวบรวมได้ประกอบด้วยข้อมูลหลัก 5 ชนิด คืออาการ พฤติกรรม ประวัติการเป็นมะเร็งชนิดต่างๆ เพศและอายุ ซึ่งถูกแจกแจงเป็น 278 คุณสมบัติเพื่ออธิบายข้อมูลผู้ป่วยโดยกำหนดให้ค่าข้อมูลเป็นตัวเลขที่สามารถคำนวณได้ ดังตัวอย่างในตารางที่ 2

ตารางที่ 2 ตัวอย่างคุณสมบัติสำหรับอธิบายข้อมูลผู้ป่วย

| ชื่อคุณสมบัติ | ค่าคุณสมบัติ |
|---------------|-----------------------------|
| ท้องผูก | 0 = ไม่มีอาการ, 1 = มีอาการ |
| ตัวเหลือง | 0 = ไม่มีอาการ, 1 = มีอาการ |
| ไอมีเสมหะ | 0 = ไม่มีอาการ, 1 = มีอาการ |

3.2 การสร้างแบบจำลองวิเคราะห์โรคมะเร็ง

การวิเคราะห์โรคมะเร็งประกอบด้วยการวิเคราะห์การเป็นโรคมะเร็งและการวิเคราะห์ชนิดโรคมะเร็ง โดยใช้แบบจำลองที่สร้างด้วยโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-layer perceptrons) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) และโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผรัศมี (Radial Basis Function Network)

3.3 การวัดประสิทธิภาพ

แบบจำลองจะถูกวัดประสิทธิภาพในรูปของความแม่นยำในการจำแนกข้อมูลผู้ป่วยว่ามีแนวโน้มหรือไม่มีแนวโน้มเป็นโรคมะเร็ง(หรือชนิดมะเร็งที่สนใจ)ได้แก่

- ร้อยละความถูกต้องในการวิเคราะห์ข้อมูล

$$p = \frac{cn}{n} \times 100 \quad (14)$$

เมื่อ p คือร้อยละความถูกต้องในการวิเคราะห์โรคมะเร็ง(หรือชนิดมะเร็งที่สนใจ)
 cn คือจำนวนข้อมูลตัวอย่างที่ถูกวิเคราะห์ได้ถูกต้อง
 n คือจำนวนข้อมูลตัวอย่างทั้งหมดที่ถูกวิเคราะห์ด้วยแบบจำลอง

- ร้อยละของความผิดพลาดแบบ False positive นั้นคือร้อยละของจำนวนคนที่ไม่เป็นโรคมะเร็งแต่แบบจำลองวิเคราะห์หว่านโรคมะเร็ง ซึ่งใช้วัดประสิทธิภาพเฉพาะแบบจำลองวิเคราะห์การเป็นโรคมะเร็ง โดยเป็นการจำแนกข้อมูลผู้ป่วยออกเป็นสองกลุ่มคือ กลุ่มที่เป็นโรคมะเร็งและกลุ่มที่ไม่เป็นโรคมะเร็ง

$$fp = \frac{fpn}{n} \times 100 \quad (15)$$

เมื่อ fp คือร้อยละของความผิดพลาดแบบ False positive
 fpn คือจำนวนคนที่ไม่เป็นโรคมะเร็งแต่แบบจำลองวิเคราะห์หว่านโรคมะเร็ง
 n คือจำนวนข้อมูลตัวอย่างทั้งหมดที่ถูกวิเคราะห์ด้วยแบบจำลอง

- ร้อยละของความผิดพลาดแบบ False negative นั้นคือร้อยละของจำนวนคนที่เป็นโรคมะเร็งแต่แบบจำลองวิเคราะห์หว่านไม่เป็นโรคมะเร็ง ซึ่งใช้วัดประสิทธิภาพเฉพาะแบบจำลองวิเคราะห์การเป็นโรคมะเร็ง โดยเป็นการจำแนกข้อมูลผู้ป่วยออกเป็นสองกลุ่มคือ กลุ่มที่เป็นโรคมะเร็งและกลุ่มที่ไม่เป็นโรคมะเร็ง

$$fn = \frac{fnn}{n} \times 100 \quad (16)$$

เมื่อ fn คือร้อยละของความผิดพลาดแบบ False negative
 fnn คือจำนวนคนที่เป็นโรคมะเร็งแต่แบบจำลองวิเคราะห์หว่านไม่เป็นโรคมะเร็ง
 n คือจำนวนข้อมูลตัวอย่างทั้งหมดที่ถูกวิเคราะห์ด้วยแบบจำลอง

4. แบบจำลองวิเคราะห์การเป็นโรคมะเร็ง

การวิเคราะห์การเป็นโรคมะเร็งดำเนินการโดยสร้างแบบจำลองซึ่งประยุกต์ใช้โครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptrons Network) และเปรียบเทียบประสิทธิภาพการทำงานกับแบบจำลองที่สร้างด้วยโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Network) ซึ่งทั้งสองแบบจำลองนี้ถูกสร้างโดยใช้ข้อมูลที่รวบรวมได้ดังตารางที่ 1 โดยแบ่งข้อมูลออกเป็นสองชุดคือชุดข้อมูลฝึกสอน (training data set) และชุดข้อมูลทดสอบ (testing data set) ด้วยสัดส่วน 80% ต่อ 20% นั่นคือจำนวนข้อมูลสำหรับฝึกสอนแบบจำลองเท่ากับ 540 คน และอีก 135 คนสำหรับทดสอบแบบจำลอง

4.1 แบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptron Network Model) เพื่อวิเคราะห์การเป็นโรคมะเร็ง

แบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นถูกสร้างขึ้นโดยขั้นแรกกำหนดโครงสร้างและพารามิเตอร์ต่างๆ ของแบบจำลองได้แก่

- จำนวนโหนดข้อมูลนำเข้าในชั้นนำเข้า ซึ่งกำหนดให้เท่ากับ 278 โหนดตามคุณสมบัติที่อธิบายข้อมูล
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นซ่อน ซึ่งมีจำนวนขึ้นอยู่กับการทดลองและประสิทธิภาพการทำงาน ของแบบจำลอง โดยเริ่มต้นให้เท่ากับ 30 โหนดแล้วเพิ่มครั้งละหนึ่งโหนด
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นผลลัพธ์ซึ่งให้เท่ากับ 1 โหนด โดยผลลัพธ์ได้แก่ 1 คือเป็นโรคมะเร็ง หรือ 0 ไม่เป็นโรคมะเร็ง
- ฟังก์ชันกระตุ้นที่ใช้ประมวลผลในโหนดเซลล์ประสาทเทียม ซึ่งใช้ฟังก์ชันซิกมอยด์ (Sigmoid function)
- อัตราการเรียนรู้ (Learning rate) เท่ากับ 0.3
- จำนวน epoch เท่ากับ 30000
- เป้าหมายค่าความผิดพลาดกำลังสอง (Mean square error) เท่ากับ 0.001

ตารางที่ 2 ความถูกต้องของการวิเคราะห์การเป็นโรคมะเร็ง ด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

| จำนวนโหนดในชั้นซ่อน | ข้อมูลชุดฝึกสอน | ข้อมูลชุดทดสอบ |
|---------------------|-----------------|----------------|
| 30 | 94.81% | 89.63% |
| 31 | 97.59% | 90.37% |
| 32 | 97.59% | 93.33% |
| 33 | 97.22% | 93.33% |

ตารางที่ 3 ความผิดพลาดของการวิเคราะห์การเป็นโรคมะเร็งจากชุดข้อมูลฝึกสอน ด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

| จำนวนโหนดในชั้นซ่อน | False positive | False negative |
|---------------------|----------------|----------------|
| 30 | 0.56% | 4.63% |
| 31 | 0.93% | 1.48% |
| 32 | 0.19% | 2.22% |
| 33 | 0.37% | 2.41% |

ตารางที่ 4 ความผิดพลาดของการวิเคราะห์การเป็นโรคมะเร็งจากชุดข้อมูลทดสอบ
ด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

| จำนวนโหนดในชั้นซ่อน | False positive | False negative |
|---------------------|----------------|----------------|
| 30 | 1.48% | 8.89% |
| 31 | 0.74% | 8.89% |
| 32 | 0% | 6.67% |
| 33 | 0% | 6.67% |

เมื่อกำหนดโครงสร้างโครงข่ายแล้ว จากนั้นเป็นการฝึกสอนแบบจำลองด้วยขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับกับข้อมูลชุดฝึกสอนและทดสอบการทำงานของแบบจำลองด้วยชุดข้อมูลทดสอบและวัดประสิทธิภาพการทำงานดังแสดงในตารางที่ 2 – 4 โดยพบว่าแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นที่มีจำนวนโหนดในชั้นซ่อนเท่ากับ 32 โหนดสามารถวิเคราะห์การเป็นโรคมะเร็งได้ถูกต้องมากที่สุดทั้งจากชุดข้อมูลฝึกสอนและชุดข้อมูลทดสอบ

4.2 แบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Network Model) เพื่อวิเคราะห์การเป็นโรคมะเร็ง

การสร้างแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีนี้ เริ่มด้วยการกำหนดโครงสร้างของแบบจำลองเช่นเดียวกับแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น โดยต่างกันที่จำนวนโหนดในชั้นซ่อนและฟังก์ชันที่ประมวลผลในเซลล์ประสาทเทียมรวมถึงค่าพารามิเตอร์อื่นๆ ดังนี้

- จำนวนโหนดข้อมูลนำเข้าในชั้นนำเข้า ซึ่งกำหนดให้เท่ากับ 278 โหนดตามคุณสมบัติที่อธิบายข้อมูล
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นซ่อน ซึ่งมีจำนวนขึ้นอยู่กับการทดลองและประสิทธิภาพการทำงานของแบบจำลอง โดยเริ่มต้นให้เท่ากับ 1 โหนดแล้วเพิ่มครั้งละหนึ่งโหนด จนกว่าจะจนกว่าจะได้โครงข่ายที่ทำงานได้ประสิทธิภาพสูงสุด
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นผลลัพธ์ซึ่งให้เท่ากับ 1 โหนด โดยผลลัพธ์ได้แก่ 1 คือเป็นโรคมะเร็งหรือ 0 ไม่เป็นโรคมะเร็ง
- ฟังก์ชันกระตุ้นที่ใช้ประมวลผลในโหนดเซลล์ประสาทเทียม ซึ่งใช้ฟังก์ชันกรีน (Green's function) สำหรับโหนดในชั้นซ่อน และฟังก์ชันเชิงเส้น (Linear function) สำหรับโหนดในชั้นผลลัพธ์
- ค่ารัศมีฟังก์ชันฐานหลักที่แผ่รัศมี (spread) เริ่มต้นเท่ากับ 1 แล้วปรับขึ้นหรือลงครั้งละ 0.2 จนกว่าจะได้โครงข่ายที่ทำงานได้ประสิทธิภาพสูงสุด
- เป้าหมายค่าความผิดพลาดกำลังสอง (Mean square error) เท่ากับ 0.001

ตารางที่ 5 ความถูกต้องของการวิเคราะห์การเป็นโรคมะเร็ง
ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี

| จำนวนโหนดในชั้นซ่อน | ค่ารัศมีฟังก์ชันฐานหลักที่แผ่รัศมี (spread) | ข้อมูลชุดฝึกสอน | ข้อมูลชุดทดสอบ |
|---------------------|---|-----------------|----------------|
| 540 | 1 | 99.63% | 97.78% |
| 540 | 0.8 | 99.63% | 94.07% |
| 540 | 1.2 | 99.63% | 77.04% |

ตารางที่ 6 ความผิดพลาดของการวิเคราะห์การเป็นโรคมะเร็งจากชุดข้อมูลฝึกสอน
ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี

| จำนวนโหนดในชั้น ซ่อน | ค่ารัศมีฟังก์ชันฐานหลักที่ แผ่รัศมี (spread) | False positive | False negative |
|-------------------------|---|----------------|----------------|
| 540 | 1 | 0.19% | 0.18% |
| 540 | 0.8 | 0% | 0.37% |
| 540 | 1.2 | 0% | 0.37% |

ตารางที่ 7 ความผิดพลาดของการวิเคราะห์การเป็นโรคมะเร็งจากชุดข้อมูลทดสอบ
ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี

| จำนวนโหนดในชั้น ซ่อน | ค่ารัศมีฟังก์ชันฐานหลักที่ แผ่รัศมี (spread) | False positive | False negative |
|-------------------------|---|----------------|----------------|
| 540 | 1 | 0% | 2.22% |
| 540 | 0.8 | 5.19% | 0.74% |
| 540 | 1.2 | 0% | 22.96% |

เมื่อปรับโครงสร้างและคำนวณผลลัพธ์ของโครงข่ายแล้วได้ผลการดำเนินงานดังแสดงในตารางที่ 5-7 โดยพบว่าแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีที่กำหนดค่ารัศมีเท่ากับ 1 ให้ความถูกต้องในการวิเคราะห์การเป็นโรคมะเร็งมากที่สุด

5. แบบจำลองวิเคราะห์ชนิดโรคมะเร็ง

การวิเคราะห์ชนิดโรคมะเร็งดำเนินการด้วยแบบจำลองโครงข่ายประสาทเทียมเช่นเดียวกับการวิเคราะห์การเป็นโรคมะเร็ง โดยประยุกต์ใช้โครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptrons Network) เปรียบเทียบกับโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Network) และใช้จำนวนข้อมูลตามที่แสดงในตารางที่ 1 พร้อมทั้งแบ่งข้อมูลออกเป็นสองชุดคือชุดข้อมูลฝึกสอน (training data set) จำนวน 758 คน และชุดข้อมูลทดสอบ (testing data set) จำนวน 191 คน ดังแสดงในตารางที่ 8

ตารางที่ 8 จำนวนข้อมูลฝึกสอนและทดสอบสำหรับสร้างแบบจำลอง
วิเคราะห์ชนิดโรคมะเร็ง

| ชนิดโรค | ชุดข้อมูลฝึกสอน | ชุดข้อมูลทดสอบ |
|-----------------|-----------------|----------------|
| มะเร็งปอด | 162 | 41 |
| มะเร็งตับ | 263 | 66 |
| มะเร็งเต้านม | 210 | 53 |
| มะเร็งลำไส้ใหญ่ | 123 | 31 |
| รวม | 758 | 191 |

5.1 แบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptron Network Model) เพื่อวิเคราะห์ชนิดโรคมะเร็ง

การกำหนดโครงสร้างและพารามิเตอร์สำหรับแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นเพื่อวิเคราะห์ชนิดโรคมะเร็งเป็นดังนี้

- จำนวนโหนดข้อมูลนำเข้าในชั้นนำเข้า ซึ่งกำหนดให้เท่ากับ 278 โหนดตามคุณสมบัติที่อธิบายข้อมูล
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นซ่อน ซึ่งมีจำนวนขึ้นอยู่กับการทดลองและประสิทธิภาพการทำงานของแบบจำลอง โดยเริ่มต้นให้เท่ากับ 30 โหนด แล้วเพิ่มครั้งละหนึ่งโหนด
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นผลลัพธ์ซึ่งให้เท่ากับ 4 โหนด โดยผลลัพธ์และการแปลผลได้แก่
 - 1000 คือโรคมะเร็งปอด
 - 0100 คือโรคมะเร็งตับ
 - 0010 คือโรคมะเร็งเต้านม
 - 0001 คือโรคมะเร็งลำไส้ใหญ่
- ฟังก์ชันกระตุ้นที่ใช้ประมวลผลในโหนดเซลล์ประสาทเทียม ซึ่งใช้ฟังก์ชันซิกมอยด์ (Sigmoid function)
- อัตราการเรียนรู้ (Learning rate) เท่ากับ 0.3
- จำนวน epoch เท่ากับ 30000
- เป้าหมายค่าความผิดพลาดกำลังสอง (Mean square error) เท่ากับ 0.001

ผลการวิเคราะห์ชนิดโรคมะเร็งดังแสดงในตารางที่ 9-11 โดยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นที่มี 34 โหนดในชั้นซ่อนมีความแม่นยำในการวิเคราะห์ชนิดมะเร็งจากข้อมูลชุดทดสอบมากที่สุด โดยวิเคราะห์มะเร็งเต้านมได้ถูกต้องมากที่สุดและวิเคราะห์มะเร็งลำไส้ใหญ่ได้ถูกต้องน้อยที่สุด

ตารางที่ 9 ความถูกต้องของการวิเคราะห์ชนิดโรคมะเร็งด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้น

| จำนวนโหนดในชั้นซ่อน | ข้อมูลชุดฝึกสอน | ข้อมูลชุดทดสอบ |
|---------------------|-----------------|----------------|
| 30 | 92.88% | 85.86% |
| 31 | 91.16% | 86.91% |
| 32 | 92.74% | 86.91% |
| 33 | 90.37% | 87.43% |
| 34 | 91.95% | 90.52% |
| 35 | 91.69% | 89.01% |

ตารางที่ 10 ผลการวิเคราะห์ชนิดโรคมะเร็งจากชุดข้อมูลฝึกสอนด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นด้วย 34 โหนดซ่อน

| ชนิดโรค | ผลการวิเคราะห์ชนิดโรคมะเร็ง | | | |
|-----------------|-----------------------------|-----------|--------------|-----------------|
| | มะเร็งปอด | มะเร็งตับ | มะเร็งเต้านม | มะเร็งลำไส้ใหญ่ |
| มะเร็งปอด | 158 | 2 | 2 | 0 |
| มะเร็งตับ | 10 | 238 | 0 | 15 |
| มะเร็งเต้านม | 0 | 0 | 208 | 2 |
| มะเร็งลำไส้ใหญ่ | 1 | 28 | 1 | 93 |

ตารางที่ 11 ผลการวิเคราะห์ชนิดโรคมะเร็งจากชุดข้อมูลทดสอบ
ด้วยแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นแบบ 34 โหนดซ่อน

| ชนิดโรค | ผลการวิเคราะห์ชนิดโรคมะเร็ง | | | |
|-----------------|-----------------------------|-----------|--------------|-----------------|
| | มะเร็งปอด | มะเร็งตับ | มะเร็งเต้านม | มะเร็งลำไส้ใหญ่ |
| มะเร็งปอด | 36 | 4 | 0 | 1 |
| มะเร็งตับ | 1 | 61 | 0 | 4 |
| มะเร็งเต้านม | 0 | 0 | 53 | 0 |
| มะเร็งลำไส้ใหญ่ | 0 | 6 | 3 | 22 |

5.2 แบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Network Model) เพื่อวิเคราะห์ชนิดโรคมะเร็ง

การกำหนดโครงสร้างและพารามิเตอร์สำหรับแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีเพื่อวิเคราะห์ชนิดโรคมะเร็งเป็นดังนี้

- จำนวนโหนดข้อมูลนำเข้าในชั้นนำเข้า ซึ่งกำหนดให้เท่ากับ 278 โหนดตามคุณสมบัติที่อธิบายข้อมูล
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นซ่อน ซึ่งมีจำนวนขึ้นอยู่กับการทดลองและประสิทธิภาพการทำงานของแบบจำลอง โดยเริ่มต้นให้เท่ากับ 1 โหนด แล้วเพิ่มครั้งละหนึ่งโหนดจนกว่าจะได้โครงข่ายที่ทำงานได้ประสิทธิภาพสูงสุด
- จำนวนโหนดเซลล์ประสาทเทียมในชั้นผลลัพธ์ซึ่งให้เท่ากับ 4 โหนด โดยผลลัพธ์และการแปลผลได้แก่
 - 1000 คือโรคมะเร็งปอด
 - 0100 คือโรคมะเร็งตับ
 - 0010 คือโรคมะเร็งเต้านม
 - 0001 คือโรคมะเร็งลำไส้ใหญ่
- ฟังก์ชันกระตุ้นที่ใช้ประมวลผลในโหนดเซลล์ประสาทเทียม ซึ่งใช้ฟังก์ชันกรีน (Green's function) สำหรับโหนดในชั้นซ่อน และฟังก์ชันเชิงเส้น (Linear function) สำหรับโหนดในชั้นผลลัพธ์
- ค่ารัศมีฟังก์ชันฐานหลักที่แผ่รัศมี (spread) เริ่มต้นเท่ากับ 1 แล้วปรับขึ้นหรือลงครั้งละ 0.2 จนกว่าจะได้โครงข่ายที่ทำงานได้ประสิทธิภาพสูงสุด
- เป้าหมายค่าความผิดพลาดกำลังสอง (Mean square error) เท่ากับ 0.001

ตารางที่ 12 ความถูกต้องของการวิเคราะห์ชนิดโรคมะเร็ง
ด้วยแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี

| จำนวนโหนดในชั้นซ่อน | ค่ารัศมีฟังก์ชันฐานหลักที่แผ่รัศมี (spread) | ข้อมูลชุดฝึกสอน | ข้อมูลชุดทดสอบ |
|---------------------|---|-----------------|----------------|
| 758 | 1 | 98.55% | 87.43% |
| 758 | 1.5 | 98.55% | 79.06% |
| 758 | 1.3 | 98.55% | 79.58% |
| 758 | 0.8 | 98.55% | 85.34% |

ตารางที่ 13 ผลการวิเคราะห์ชนิดโรคมะเร็งจากชุดข้อมูลฝึกสอนด้วยแบบจำลอง
โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี ด้วย spread = 1

| ชนิดโรค | ผลการวิเคราะห์ชนิดโรคมะเร็ง | | | |
|-----------------|-----------------------------|-----------|--------------|-----------------|
| | มะเร็งปอด | มะเร็งตับ | มะเร็งเต้านม | มะเร็งลำไส้ใหญ่ |
| มะเร็งปอด | 162 | 0 | 0 | 0 |
| มะเร็งตับ | 1 | 259 | 0 | 3 |
| มะเร็งเต้านม | 2 | 0 | 208 | 0 |
| มะเร็งลำไส้ใหญ่ | 2 | 3 | 0 | 118 |

ตารางที่ 14 ผลการวิเคราะห์ชนิดโรคมะเร็งจากชุดข้อมูลทดสอบด้วยแบบจำลอง
โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี ด้วย spread = 1

| ชนิดโรค | ผลการวิเคราะห์ชนิดโรคมะเร็ง | | | |
|-----------------|-----------------------------|-----------|--------------|-----------------|
| | มะเร็งปอด | มะเร็งตับ | มะเร็งเต้านม | มะเร็งลำไส้ใหญ่ |
| มะเร็งปอด | 38 | 3 | 0 | 0 |
| มะเร็งตับ | 1 | 64 | 0 | 1 |
| มะเร็งเต้านม | 0 | 3 | 50 | 0 |
| มะเร็งลำไส้ใหญ่ | 1 | 14 | 1 | 15 |

ผลการวิเคราะห์ชนิดโรคมะเร็งดังแสดงในตารางที่ 12-14 ซึ่งพบว่าแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีที่กำหนดค่ารัศมีฟังก์ชันฐานหลักที่แผ่รัศมี เท่ากับ 1 ให้ผลการวิเคราะห์ชนิดโรคมะเร็งถูกต้องที่สุด โดยมะเร็งตับถูกวิเคราะห์ได้ถูกต้องที่สุดในขณะที่มะเร็งลำไส้ใหญ่ถูกวิเคราะห์ได้ถูกต้องน้อยที่สุดจากชุดข้อมูลทดสอบ

6. บทสรุปและข้อเสนอแนะ

งานวิจัยนี้ได้ประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียมซึ่งได้แก่ โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Neural Network) เปรียบเทียบกับโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptrons) ร่วมกับขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm) เพื่อวิเคราะห์การเป็นโรคมะเร็งและระบุชนิดโรคมะเร็ง โดยชนิดมะเร็งที่สนใจคือ มะเร็งปอด มะเร็งตับ มะเร็งเต้านม และมะเร็งลำไส้ใหญ่

จากผลการทดลองวิเคราะห์การเป็นโรคมะเร็งพบว่าแบบจำลองทั้งสองแบบสามารถวิเคราะห์การเป็นโรคมะเร็งได้ถูกต้องมากกว่าร้อยละ 90 โดยเฉพาะแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี สามารถวิเคราะห์การเป็นโรคมะเร็งได้ถูกต้องถึงร้อยละ 97.78 โดยความผิดพลาดส่วนมากเป็นแบบ false negative คือคนที่ เป็นโรคมะเร็งแต่แบบจำลองวิเคราะห์หาว่าไม่เป็นโรคมะเร็ง

สำหรับการวิเคราะห์ชนิดโรคมะเร็งนั้น แบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี สามารถวิเคราะห์ชนิดโรคมะเร็งได้ถูกต้องร้อยละ 87.43 โดยวิเคราะห์มะเร็งตับได้ดีที่สุดเมื่อเทียบกับมะเร็งชนิดอื่น ส่วนแบบจำลองโครงข่ายเพอร์เซปตรอนแบบหลายชั้นสามารถวิเคราะห์ชนิดโรคมะเร็งได้ถูกต้องมากกว่าแบบจำลองโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี โดยมีค่าร้อยละความถูกต้องเท่ากับร้อยละ 90.52 ซึ่งวิเคราะห์มะเร็งเต้านมได้ดีที่สุดเมื่อเทียบกับมะเร็งชนิดอื่น อย่างไรก็ตามแบบจำลองวิเคราะห์ชนิดมะเร็งทั้งสองแบบนี้ยังมี

ความผิดพลาดซึ่งระบุชนิดมะเร็งไม่ถูกต้องโดยเฉพาะมะเร็งลำไส้ใหญ่ถูกวิเคราะห์ผิดว่าเป็นมะเร็งตับซึ่งสาเหตุคือผู้ป่วยที่เป็นมะเร็งลำไส้ใหญ่นั้นมักจะมีสาเหตุคือการลุกลามจากมะเร็งตับซึ่งเป็นอวัยวะที่อยู่ใกล้กับลำไส้ใหญ่ นั่นคือผู้ที่เป็นมะเร็งลำไส้ใหญ่มักเป็นมะเร็งตับอยู่ก่อนแล้ว ดังนั้นการรวบรวมข้อมูลควรพิจารณาด้วยว่าเป็นมะเร็งที่ลุกลามหรือไม่

เอกสารอ้างอิง

- [1] Y. Yuhui, C. Lihui, A. Goh and A. Wogn, "Clustering Gene Data via Associative Clustering Neural Network", Proceedings of the 9th International Conference on Neural Information Processing – ICONIP'02, November, 2002, Vol. 5, pp. 2228-2232.
- [2] A. Farag, "Computer Based Acute Leukemia Classification", Proceedings of the 46th International Midwest Symposium on Circuits and Systems 2003—MWSCAS'03, December, 2003, Volume 2, pp. 701-703
- [3] F. Chu, W. Xie and L. Wang, "Gene Selection and Cancer Classification Using a Fuzzy Neural Network", Proceedings of IEEE Annual Meeting of the Fuzzy Information, June 27th -30th, 2004, Volume 2, pp. 555-559.
- [4] C. Lee, H. Chiue ad H. Yang, "A platform of Biomedical Literal Mining for Categorization of Cancer Related Abstracts", Proceedings of the Second International Conference on Innovative Computing, Information and Control 2007 – ICICIC'07, September, 2007, pp. 174-177.
- [5] U. Kitkana, W. Sairatanathongkham and A. Mahaweerawat, "Leukemia Classification using Codon and Neural Networks", Proceedings of the 12th National Computer Science and Engineering Conference, November 20th - 21st, 2008, pp. 293-299
- [6] ผศ.นพ. วิเชียร ศรีมนินทรนิมิต. เอกสารเผยแพร่ความรู้เกี่ยวกับโรคมะเร็ง. กรุงเทพฯ : บริษัท โรช ไทยแลนด์ จำกัด.
- [7] S. Haykin, Neural Networks a comprehensive foundation, Prentice Hall, 1999, pp. 50-312.
- [8] T. M. Mitchell, Machine Learning, First Edition, McGraw-Hill, 1997, pp. 95-112.
- [9] R. Kullawong and A. Mahaweerawat, "Hepatobiliary Disorders Classification using Neural Network and Feature Selection", Proceedings of the 13th National Computer Science and Engineering Conference (NCSEC 2009), November 4th-6th, 2009, pp. 192-197.
- [10] M. S. Sharif, A. N. Sazish and A. Amira, "An Efficient Algorithm and Architecture for Medical Image Segmentation and Tumour Detection", Proceedings of IEEE Biomedical Circuits and Systems Conference, November 20th-22nd, 2008, pp. 157-160.
- [11] K. Belkic, "Magnetic resonance spectroscopic imaging in breast cancer detection: possibilities beyond the conventional theoretical framework for data analysis", Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, Vol. 525, Issues 1-2, 2004, pp. 313-321.
- [12] Z. Herced and P. Hainaut, "Genetic and epigenetic alterations as biomarkers for cancer detection, diagnosis and prognosis", Molecular Oncology, Vol. 1, Issue 1, June 2007, pp. 26-41.
- [13] T. Bauer, "Lung Cancer Screening", Hematology/Oncology Clinics of North America, Vol. 19, Issue 2, 2005, pp. 209-217.

- [14] สถาบันมะเร็งแห่งชาติ(National Cancer Institute).
<http://www.nci.go.th/knowledge/whatis.html>(Accessed March 15, 2010)
- [15] มะเร็ง(Cancer).<http://th.wikipedia.org/wiki> (Accessed March 15, 2010)

ตารางเปรียบเทียบผลดำเนินงานตามแผน

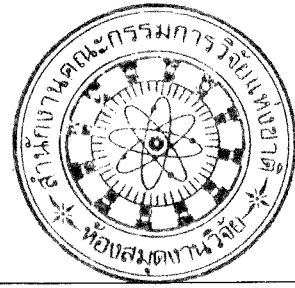
| กิจกรรม(ตามแผน) | ผลที่คาดว่าจะได้รับ (ตามแผน) | ผลการดำเนินงาน | หมายเหตุ |
|--|---|--|----------|
| <p>1. การเตรียมงานวิจัย/ ทบทวนศึกษาเอกสาร ศึกษาและทบทวน ทฤษฎี เทคนิค และ ขั้นตอนวิธีต่างๆ สำหรับพัฒนา แบบจำลองโครงข่ายประสาทเทียมและการวิเคราะห์ผลลัพธ์ รวมทั้งงานวิจัยอื่นๆที่เกี่ยวข้อง</p> | <p>ทฤษฎี เทคนิคและ ขั้นตอนวิธีที่เหมาะสมกับการพัฒนา แบบจำลองโครงข่ายประสาทเทียมและการวิเคราะห์ผลลัพธ์</p> | <p>ทฤษฎี เทคนิคและ ขั้นตอนวิธีที่เหมาะสมกับการพัฒนา แบบจำลองโครงข่ายประสาทเทียมและการวิเคราะห์ผลลัพธ์ โดยพบว่าควรใช้ เทคนิคโครงข่ายประสาทเทียมแบบมี ผู้สอนซึ่งได้แก่ โครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมี (Radial Basis Function Neural Network) เปรียบเทียบกับโครงข่ายเพอร์เซปตรอนแบบหลายชั้น (Multi-Layer Perceptrons) ร่วมกับ ขั้นตอนวิธีการเรียนรู้แบบแพร่ย้อนกลับ (Back Propagation Learning Algorithm)</p> | |
| <p>2. กำหนดแผนการ/ วิธีการดำเนินงานวิจัย</p> | <p>แผนงานและขั้นตอน การดำเนินงานวิจัย</p> | <p>แผนงานและขั้นตอน การดำเนินงานวิจัย</p> | |

| | | | |
|--|---|--|--|
| วางแผนและกำหนด ขั้นตอนการดำเนินงาน วิจัย | | | |
| 3. ดำเนินงานวิจัย/เก็บ ข้อมูล | | | |
| 3.1 รวบรวมข้อมูล ผู้ป่วย | ข้อมูลผู้ป่วยตามชนิด โรคที่สนใจ | ข้อมูลผู้ป่วยตามชนิด โรคที่สนใจดังแสดง ในตารางที่ 3.1 | |
| 3.2 กำหนดคุณสมบัติ สำหรับอธิบายข้อมูล ผู้ป่วยที่รวบรวมได้จาก ข้อ 3.1 และแปลงให้ อยู่ในรูปแบบที่ ประมวลผลด้วย ซอฟต์แวร์ คอมพิวเตอร์ได้ | คุณสมบัติสำหรับ อธิบายข้อมูลผู้ป่วย และข้อมูลที่อยู่ในรูปที่ สามารถประมวลผล ด้วยซอฟต์แวร์ คอมพิวเตอร์ได้ | คุณสมบัติสำหรับ อธิบายข้อมูลผู้ป่วย และข้อมูลที่อยู่ในรูปที่ สามารถประมวลผล ด้วยซอฟต์แวร์ คอมพิวเตอร์ได้ ดัง แสดงในตารางที่ 3.4- 3.22 | |
| 3.3 ทำการทดลองเพื่อ พัฒนาแบบจำลอง โครงข่ายประสาท เทียมที่เหมาะสมใน การวิเคราะห์ความ เสี่ยงของการเกิด โรคมะเร็งด้วยทฤษฎี เทคนิค และขั้นตอนวิธี ที่ได้จากข้อ 1 | แบบจำลองโครงข่าย ประสาทเทียม | แบบจำลองโครงข่าย ประสาทเทียมแบบ แบบฟังก์ชันฐานหลัก ที่แพร่กระจาย และ แบบจำลองโครงข่าย เพอร์เซปตรอนแบบ หลายชั้น | |
| 4.การวิเคราะห์ | | | |
| วิเคราะห์ผลลัพธ์ที่ได้ จากการทดลอง เพื่อหา คุณสมบัติที่สำคัญต่อ | ผลการวิเคราะห์ ผลลัพธ์ ประสิทธิภาพ ของแบบจำลอง และ | ผลการวิเคราะห์ ผลลัพธ์ ประสิทธิภาพ ของแบบจำลอง และ | |

| | | | |
|---|-------------------|-------------------|--|
| การวิเคราะห์ความ เสี่ยงของการเกิด โรคมะเร็ง | คุณสมบัติที่สำคัญ | คุณสมบัติที่สำคัญ | |
| 5. จัดทำรายงานการ วิจัยฉบับสมบูรณ์ | รายงานฉบับสมบูรณ์ | รายงานฉบับสมบูรณ์ | |

ผลที่ได้รับตลอดโครงการ

1. ได้แบบจำลองวิเคราะห์การเป็นโรคมะเร็งและแบบจำลองการวิเคราะห์ชนิดโรคมะเร็ง
2. ได้บทความวิจัยเรื่องการวิเคราะห์ความเสี่ยงของการเกิดโรคมะเร็งด้วยโครงข่ายประสาทเทียมแบบฟังก์ชันฐานหลักที่แผ่รัศมีและโครงข่ายเพอร์เซปตรอนแบบหลายชั้น



รายงานการเงิน

| รายการ | จำนวนเงิน (บาท) |
|---|------------------|
| 1. หมวดค่าตอบแทน | |
| 1.1 ค่าตอบแทนนักวิจัย | 9,000 |
| 1.2 ค่าตอบแทนผู้เชี่ยวชาญ | 3,000 |
| 2. หมวดค่าใช้จ่าย | |
| 2.1 ค่าจ้างเหมาพิมพ์เอกสาร ถ่ายเอกสารและทำรูปเล่ม | 767 |
| 2.2 ค่าเดินทาง ไปพบผู้เชี่ยวชาญ | 16,800 |
| 2.3 ค่าจ้างเหมาผู้ช่วยวิจัย | 1,000 |
| 3. หมวดค่าวัสดุ | |
| 3.1 วัสดุสำนักงาน | 2,007 |
| 3.2 วัสดุหนังสือ วารสารและตำรา | 1,054.50 |
| 3.3 วัสดุคอมพิวเตอร์ | 14,748 |
| รวมงบประมาณทั้งสิ้น | 48,376.50 |

