



## การจำแนกผู้ป่วยโรคมะเร็งตับโดยการประยุกต์ใช้โครงข่ายประสาทเทียม Patients Classification of Liver Cancer by Applying Artificial Neural Networks

วิชุดา ไชยวิมมงคล<sup>1\*</sup>, เปรม จันทร์สว่าง<sup>1</sup>, จุลพวรรณ อังจะนิส<sup>2</sup>, ประจวบ ชัยมณี<sup>3</sup>, วรุฒ ชัยวงษ์<sup>4</sup> และ  
 วิราวรรณ พุทธมาตย์<sup>4</sup>

Wichuda Chaisiwamongkol<sup>1\*</sup>, Prem Junsawang<sup>1</sup>, Julapan Engchanil<sup>2</sup>, Prajuab Chaimanee<sup>3</sup>,  
 WarutChaiwong<sup>4</sup> and Wirawan Puttamat<sup>4</sup>

<sup>1</sup>ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

<sup>2</sup>ภาควิชาจุลชีววิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

<sup>3</sup>งานห้องปฏิบัติการเวชศาสตร์ชั้นสูง คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

<sup>4</sup>บัณฑิตสาขาสารสนเทศสถิติ หลักสูตรวิทยาศาสตรบัณฑิต คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

\* Correspondent author: [wichuda@kku.ac.th](mailto:wichuda@kku.ac.th)

### บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์เพื่อนำเสนอการสร้างตัวแบบในการจำแนกผู้ป่วยโรคมะเร็งตับเพื่อประกอบการสนับสนุนการวินิจฉัยของแพทย์ในเบื้องต้นในการจำแนกผู้ป่วยโรคมะเร็งตับ โดยการประยุกต์ใช้วิธีโครงข่ายประสาทเทียม (ANN) สำหรับข้อมูลได้จากการบันทึกข้อมูลประวัติการรักษาจากงานเวชระเบียนและสถิติและงานห้องปฏิบัติการเวชศาสตร์ชั้นสูง ของโรงพยาบาลศรีนครินทร์ จังหวัดขอนแก่น ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2553 ถึงวันที่ 31 ธันวาคม พ.ศ. 2554 โดยข้อมูลจะถูกนำมาจัดการเบื้องต้นและแปลงให้อยู่ในรูปแบบที่เหมาะสม ข้อมูลที่ใช้ในการทดลองมีทั้งหมด 6,499 ราย โดยเป็นผู้ป่วยมะเร็งตับ 3,412 ราย ตัวแปรที่ศึกษารวม 8 ตัวแปรได้แก่ อายุ เพศ อาชีพ ค่าซีรัมอัลบูมิน ค่าซีรัมไคเร็กบิรูบิน ค่าซีรัมเอนไซม์อะสปาทเทอะมิโนทรานสเฟอเรส ค่าซีรัมอัลคาไลด์ฟอสฟาเทส และค่าซีรัมอะลานินอะมิโนทรานสเฟอเรส ในงานนี้ผู้วิจัยทำการศึกษาและทดสอบโครงสร้างตัวแบบ 18 ตัวแบบ โดยจำแนกจากจำนวนโหนดที่ใช้ใน Hidden Layer พร้อมทั้งทดสอบอัตราส่วนของข้อมูลที่เหมาะสมของ Training set, Testing set และ Validation set เพื่อให้ได้ตัวแบบที่มีความถูกต้องในการจำแนกผู้ป่วยสูงที่สุด กระบวนการสร้างตัวแบบ ANN ดำเนินการโดยใช้โปรแกรม MATLAB

ผลการศึกษาพบว่าประสิทธิภาพการจำแนกมีค่าสูงสุดเมื่อจำนวน Hidden Layer เป็น 5 และอัตราส่วนของข้อมูล Training set, Testing set และ Validation set เป็น 70:20:10 ตามลำดับ จะได้ตัวแบบที่ให้ค่าความถูกต้องของ Test set 95.39% และค่าเฉลี่ยความถูกต้องโดยรวม 94.91% โดยตัวแบบโครงข่ายประสาทเทียมที่ได้มีความถูกต้องสูง สามารถนำไปใช้เพื่อประกอบการสนับสนุนการตัดสินใจของแพทย์ได้ในเบื้องต้น และจากการประเมินการนำไปใช้ในกลุ่มผู้ให้พบว่า คุ่มีอง่ายต่อการทำความเข้าใจและใช้งานได้ง่าย

## Abstract

The objective of this research is to create a model of liver cancer classification using Artificial Neural Network (ANN) for supporting physicians in the primary diagnosis of liver cancer classification. In this work, the experimental data were obtained from patient registration section and Clinical Laboratory section, Srinagarind Hospital, Faculty of Medicine, Khon Kaen University. This research studied patients who have risk of liver cancer and had been physical examination from Srinagarind Hospital from 1 January 2010 to 31 December 2011. Data were cleaned and transformed into the appropriate format. There were totally 6,499 patient records for experiment. The 3,412 cases are considered as liver cancer patients. Eight relevant variables, including age, gender, occupation, albumin, direct bilirubin, aspartate aminotransferase, alkaline phosphatase, and alanine aminotransferase, were used to construct a classification model by ANN. In this work, the 18 ANN structures, identified by the number of nodes in hidden layer, with various ratios of training, test and validation sets were evaluated to achieve the ANN model structure with highest classification accuracy. The ANN model construction was executed by MATLAB programming.

The results showed that classification performance was highest when the number of nodes in hidden layer was 5, and the suitable ratio of the training set, testing set and validation set was 70:20:10. The accuracy of the test set was 95.39% and the overall average of accuracy 94.91%. The obtained Artificial Neural Network model in this research had high classification accuracy. The ANN model could be applied to support physicians in the primary diagnosis of liver cancer classification. Furthermore, the satisfaction evaluation of the ANN model was made by the user group. The evaluation results indicated that the model was manually simple and easy to understand.

**คำสำคัญ :** มะเร็งตับ โครงข่ายประสาทเทียม การจำแนกประเภท

**Keyword:** liver cancer, neural network, classification.

## 1. บทนำ

ปัจจุบันประเทศไทยมีผู้ป่วยเป็นโรคมะเร็งตับจำนวนมาก โดยสถิติมะเร็งตับเป็นสาเหตุการเสียชีวิตอันดับ 1 ในผู้ชาย และเป็นอันดับ 3 ในผู้หญิง โรคมะเร็งตับที่พบมากในประเทศไทยมี 2 ชนิด คือ มะเร็งตับชนิดเซลล์ตับ พบได้ทั่วทุกภาคของประเทศ ซึ่งเกิดจากไวรัสตับอักเสบบี ซี และสารอัลฟาที่ออกซิน สำหรับมะเร็งตับชนิดเซลล์ท่อน้ำดี คือ มะเร็งที่เกิดจากเซลล์ที่อยู่ภายในท่อน้ำดี ส่วนมะเร็งตับที่อยู่ภายในตับ เกิดจากพยาธิใบไม้ตับ ซึ่งเป็นพยาธิที่มีอยู่ในปลาน้ำจืดตามหนองบึง (1)

ในภาคตะวันออกเฉียงเหนือ พบการเกิดโรคมะเร็งตับในผู้ชายอัตราเฉลี่ย 80 คน และผู้หญิงอัตราเฉลี่ย 40 คนต่อประชากร 100,000 คน โดยเฉพาะอย่างยิ่งจังหวัดขอนแก่น จัดเป็นเขตที่มีอุบัติการณ์ของโรคมะเร็งตับสูงที่สุดในโลก ทั้งเพศชายและเพศหญิง จะพบอุบัติการณ์เท่ากับ 190.6 คนต่อแสนประชากรในเพศชาย และ 132.4 คนต่อแสนประชากรในเพศหญิง (2)

สาเหตุประการสำคัญที่ทำให้ผู้ป่วยมะเร็งตับมีอัตราการอยู่รอดต่ำคือ ผู้ป่วยมะเร็งตับในระยะแรกสามารถรักษาให้หายขาดได้แต่มักจะไม่ได้แสดงอาการที่ชัดเจนออกมา เมื่อมีอาการที่ชัดเจนมะเร็งก็อยู่ในระยะลุกลามหรือมีขนาดใหญ่และไม่สามารถจะรักษาได้แล้ว

โดยผู้ป่วยจะมีการคลุ้มเคลือ เช่น เสียคท้องด้านขวา มีอาการจุกแน่นในบางครั้ง คลื่นไส้อาเจียนบ่อย ปวดท้องเรื้อรัง ตัวเหลือง ตาเหลือง อาการของตับอักเสบเรื้อรังหรือตับแข็งแย่ง ซึ่งอาการเหล่านี้จะทำให้แพทย์วินิจฉัยถึงโรคทางระบบประสาทหรือต่อมไร้ท่อมากกว่าโรคมะเร็งตับ (1)

งานวิจัยของนางเยาว์ นอรุณ และ พรณี สิทธิเดช (3) ได้ทำการจำแนกข้อมูลผู้ป่วยโรคหัวใจขาดเลือด และผู้ป่วยโรคหัวใจรูปแบบอื่นๆ โดยใช้เทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม (Artificial Neural Network: ANN) และวิธีเค-เนียร์เรสเนเบอร์ (K-Nearest Neighbor) ซึ่งทำให้แพทย์สามารถรักษาผู้ป่วยได้ถูกต้องทันเวลาและจะช่วยลดอัตราการเสียชีวิตของผู้ป่วยได้ โดยที่แพทย์ผู้เชี่ยวชาญสามารถนำข้อมูลด้านปัจจัยเสี่ยงต่างๆ มาใช้ทำนายกลุ่มผู้ป่วย ณ ปัจจุบันได้ว่าจะจัดผู้ป่วยให้อยู่ในผู้ป่วยกลุ่มใด

Chien-Yeh Hsu และคณะ (4) ได้ศึกษาเพื่อทำนายการรอดชีวิตในผู้ป่วยมะเร็งระดับโดยใช้โครงข่ายประสาทเทียม (ANN) และ classification and regression tree (CART) ด้วยข้อมูลจากฐานข้อมูลทะเบียนมะเร็งในทางการแพทย์ศูนย์ในภาคเหนือของไต้หวันระหว่างปี 2004 และ 2008 รวม 227 จากตัวแปร 9 ตัวคือ age, tumor size, tracking period, gender, clinical stage, undergoing surgery, radiotherapy, chemotherapy, and transcatheter arterial embolization (TAE) ผลการศึกษาพบว่า แบบจำลอง ANN สามารถทำนายการอยู่รอดได้ดีกว่า ( $p < 0.001$ )

ดังนั้นผู้วิจัยจึงสนใจที่จะนำเทคนิคการทำเหมืองข้อมูลด้วยวิธีโครงข่ายประสาทเทียม (Artificial Neural Network) มาใช้เพื่อหาตัวแบบในการจำแนกผู้ป่วยโรคมะเร็งตับของโรงพยาบาลศรีนครินทร์ จังหวัดขอนแก่น โดยคาดว่าจะได้ตัวแบบที่มีประสิทธิภาพการทำนายสูงพอที่จะนำมาใช้ให้เป็นประโยชน์เพื่อประกอบการวินิจฉัยของแพทย์ได้เร็วขึ้น ซึ่งจะช่วยลดอัตราการเสียชีวิตของผู้ป่วยได้

## 2. วิธีการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยที่ใช้ข้อมูลจาก การตรวจร่างกายและวินิจฉัยโรคมะเร็งตับของผู้ป่วย ที่ผ่านการคัดกรองว่ามีความเสี่ยงที่จะเป็นโรคมะเร็งตับ และมารับการตรวจร่างกายจากแผนกเวชศาสตร์ชันสูตรโรงพยาบาลศรีนครินทร์ จังหวัดขอนแก่น ตั้งแต่ 1 มกราคม 2553 ถึง 31 ธันวาคม 2554 ทั้งนี้ผู้วิจัยได้ดำเนินการขอความเห็นชอบในการทำวิจัย โดยผ่านคณะกรรมการจริยธรรมการวิจัยในมนุษย์มหาวิทยาลัยขอนแก่น เมื่อวันที่ 12 มกราคม 2555 หมายเลข HE541388 โดยมีขั้นตอนการดำเนินการดังนี้

### 2.1 การเตรียมข้อมูล (Data Preparation)

เริ่มจากการทำความเข้าใจกับกระบวนการตรวจคัดกรองผู้ป่วยโรคมะเร็งตับ ณ แผนกเวชศาสตร์ (Business Understanding) ชันสูตร ด้วยการสัมภาษณ์บุคลากรที่เกี่ยวข้องทำให้ทราบถึงกระบวนการในการตรวจคัดกรองโรคมะเร็งตับ และตัวแปรที่แพทย์ใช้ในการประกอบการวินิจฉัยโรคมะเร็งตับ ตลอดจนกระบวนการจัดเก็บข้อมูลในระบบดังกล่าว ซึ่งแหล่งข้อมูลที่เกี่ยวข้องกับงานวิจัยนี้มี 2 แหล่ง คือ งานเวชระเบียนและสถิติที่ประกอบไปด้วยแฟ้มข้อมูลทะเบียนผู้ป่วยและแฟ้มข้อมูลห้องตรวจทั่วไป และงานห้องปฏิบัติการเวชศาสตร์ชันสูตร คณะแพทยศาสตร์ประกอบด้วยแฟ้มข้อมูลการตรวจทางห้องปฏิบัติการ จากนั้นศึกษา Metadata ของแฟ้มข้อมูลเหล่านั้น และทำการเลือกตัวแปรที่ต้องการมาไว้ที่พิกข้อมูล (staging area) จากนั้นทำการตรวจสอบคุณภาพข้อมูล (clean and validation) และแปลงข้อมูล (transform) ตามตารางที่ 1 การตรวจสอบความสัมพันธ์ระหว่างตัวแปรต่างๆ กับตัวแปรเป้าหมาย ใช้ตัวสถิติทดสอบ Pearson Chi-square เนื่องจากข้อมูลเป็นข้อมูลเชิงคุณภาพ ทั้งนี้ผู้วิจัยได้มีการวิเคราะห์เชิงสถิติโดยคำนึงถึงเงื่อนไขของค่าคาดหวัง (Expected count :  $E_i$ ) ตามทฤษฎีอย่างครบถ้วน สุดท้ายได้ตัวแปรที่จะนำมาใช้ในงานวิจัยนี้ทั้งสิ้น 9 ตัวแปรรวมตัวแปรเป้าหมายดังตารางที่ 2

ตารางที่ 1. การตรวจสอบคุณภาพข้อมูลและการแปลงข้อมูล

| ประเด็น                | วิธีการตรวจสอบและการจัดการข้อมูล                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                    |
|------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| การตรวจสอบคุณภาพข้อมูล | <ol style="list-style-type: none"> <li>1. หมายเลขประจำตัวผู้ป่วย (HN) 6 หลักประกอบด้วย 2 หลักแรก จะเป็นตัวอักษรภาษาอังกฤษตัวพิมพ์ใหญ่และ 4 หลักสุดท้ายจะเป็นตัวเลขจำนวนเต็มบวก หากข้อมูลใดไม่มีรูปแบบดังกล่าวแสดงว่าผู้ป่วยนั้น เป็นผู้ป่วยของโรงพยาบาลอื่น จะไม่นำมาใช้ในการวิเคราะห์</li> <li>2. ตรวจสอบค่าว่าง (Missing) ของตัวแปรแต่ละตัว หากใน 1 record มีจำนวนตัวแปรที่ว่าง จะไม่นำมาใช้ในการวิเคราะห์</li> <li>3. ตรวจสอบช่วง (range) ของข้อมูลที่เป็นไปได้ในตัวแปรแต่ละตัวตามรหัสที่ใช้ในโรงพยาบาลฯ เช่นอายุของผู้ป่วยมีค่า 0 ถึง 150 ปี อาชีพ 0100-1014 และค่าผลตรวจจากห้องปฏิบัติการต่างๆ ก็เช่นกัน</li> <li>4. ตรวจสอบความสอดคล้องของตัวแปร โดยการเปรียบเทียบค่าตัวแปรระหว่าง 2 ตัวเช่น เพศกับค่าน้ำหนัก เช่นเดียวกับอายุและวันเดือนปีเกิด หากพบว่าข้อมูลไม่สอดคล้องกันก็จะทำการแก้ไขให้ถูกต้องตามความเป็นจริง</li> <li>5. ตรวจสอบความสัมพันธ์ระหว่างตัวแปรต่างๆ กับตัวแปรเป้าหมาย เนื่องจากข้อมูลเป็นเชิงคุณภาพจึงทดสอบด้วย ตัวสถิติทดสอบ Pearson Chi-square</li> </ol> |
| การแปลงข้อมูล          | <ol style="list-style-type: none"> <li>1. ตัวแปร อายุ แปลงให้อยู่ในรูป 16 ช่วงอายุ ตามรหัสที่ใช้ในโรงพยาบาล</li> <li>2. ค่าผลตรวจจากห้องปฏิบัติการต่างๆ ได้แก่ ค่าซีรั่มอัลบิวมิน ค่าซีรั่มไคเรคิบริวบิน ค่าซีรั่มแอสปาเตอะมิโนทรานสเฟอร์เรส ค่าซีรั่มอัลคาไลด์ฟอสฟาเทส และค่าซีรั่มอะลานิน อะมิโนทรานสเฟอร์เรส ถูกแปลงเป็นระดับต่างๆ ตามตารางที่ 2</li> </ol>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                      |

ตารางที่ 2. ชื่อตัวแปรและคำอธิบายตัวแปรในการวิจัย

| ชื่อตัวแปร (attribute)                | คำอธิบายตัวแปร                                                                                                                                                                              |
|---------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 เพศของผู้ป่วย                       | 1 = ชาย 2 = หญิง                                                                                                                                                                            |
| 2 อายุจริงของผู้ป่วย                  | อายุจริงของผู้ป่วยขณะที่ได้รับการวินิจฉัยครั้งแรก จัดเป็น 16 ช่วงอายุ                                                                                                                       |
| 3 อาชีพของผู้ป่วย                     | 12 กลุ่มอาชีพ 0-11                                                                                                                                                                          |
| 4 ค่าซีรั่มอัลบิวมิน                  | 0=ปริมาณของอัลบิวมินที่วัดได้น้อยกว่า 3.8 กรัมเปอร์เซ็นต์<br>1=ปริมาณของอัลบิวมินที่วัดได้มากกว่าหรือเท่ากับ 3.8 กรัมเปอร์เซ็นต์<br>2=ปริมาณของอัลบิวมินที่วัดได้มากกว่า 5.4กรัมเปอร์เซ็นต์ |
| 5 ค่าซีรั่มไคเรคิบริวบิน              | 0=ปริมาณของไคเรคิบริวบินที่วัดได้น้อยกว่าหรือเท่ากับ 0.5 ร้อยละมิลลิกรัม<br>1=ปริมาณของไคเรคิบริวบินที่วัดได้มากกว่า 0.5 ร้อยละมิลลิกรัม                                                    |
| 6 ค่าซีรั่มแอสปาเตอะมิโนทรานสเฟอร์เรส | 0=ปริมาณของแอสปาเตอะมิโนทรานสเฟอร์เรสที่วัดได้น้อยกว่าหรือเท่ากับ 32 ยูนิตต่อลิตร                                                                                                           |

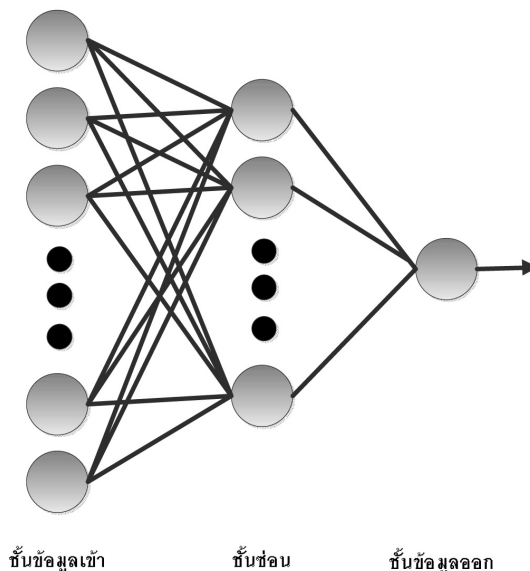
ตารางที่ 2. ชื่อตัวแปรและคำอธิบายตัวแปรในการวิจัย (ต่อ)

| ชื่อตัวแปร (attribute)               | คำอธิบายตัวแปร                                                                       |
|--------------------------------------|--------------------------------------------------------------------------------------|
|                                      | 1=ปริมาณของแอสปาเทอะมิโนทรานสเฟอร์เรสที่วัดได้มากกว่า 32 ยูนิทต่อลิตร                |
| 7 ค่าซีรัมอัลคาไลน์ฟอสฟาเทส          | 0=ปริมาณของอัลคาไลน์ฟอสฟาเทสที่วัดได้น้อยกว่าหรือเท่ากับ 109 ยูนิทต่อลิตร            |
|                                      | 1=ปริมาณของอัลคาไลน์ฟอสฟาเทสที่วัดได้ตั้งแต่ 110 ถึง 162 ยูนิทต่อลิตร                |
|                                      | 2=ปริมาณของอัลคาไลน์ฟอสฟาเทสที่วัดได้มีค่าตั้งแต่ 163 ถึง 235 ยูนิทต่อลิตร           |
|                                      | 3=ปริมาณของอัลคาไลน์ฟอสฟาเทสที่วัดได้มีค่าตั้งแต่ 236 ถึง 1229 ยูนิทต่อลิตร          |
|                                      | 4=ปริมาณของอัลคาไลน์ฟอสฟาเทสที่วัดได้มีค่ามากกว่า 1229 ยูนิทต่อลิตร                  |
| 8 ค่าซีรัมอะลานินอะมิโนทรานสเฟอร์เรส | 0=ปริมาณของอะลานินอะมิโนทรานสเฟอร์เรสที่วัดได้น้อยกว่าหรือเท่ากับ 36 ยูนิทต่อลิตร(5) |
|                                      | 1=ปริมาณของอะลานินอะมิโนทรานสเฟอร์เรสที่วัดได้มากกว่า 36 ยูนิทต่อลิตร(5)             |
| 9 ผลการวินิจฉัย                      | 0= ไม่เป็นมะเร็งตับ 1= เป็นมะเร็งตับ                                                 |

## 2.2 การสร้างตัวแบบANN (Establishing models)

โครงข่ายประสาทเทียม (Artificial Neural Network) คือแบบจำลองทางคณิตศาสตร์ที่จำลองการทำงานของเซลล์ประสาทของมนุษย์ในการจำแนกหรือรู้จำวัตถุหรือสิ่งของซึ่งถือเป็นความสามารถเบื้องต้นของมนุษย์ โดยโครงข่ายประสาทเทียมจะจำลองการทำงานของสมองมนุษย์ 2 ลักษณะได้แก่ 1) การได้มาซึ่งองค์ความรู้ของโครงข่ายประสาทเทียมซึ่งได้มาโดยผ่านกระบวนการเรียนรู้ที่ตอบสนองต่อสิ่งรอบข้างที่เกี่ยวข้อง 2) การจัดเก็บองค์ความรู้ที่ได้มาเช่นเดียวกับ

สมองมนุษย์โดยองค์ความรู้ที่ได้ของโครงข่ายประสาทเทียมจะเก็บอยู่ในรูปของค่าที่ที่เรียกว่า ค่าน้ำหนัก (6) โดยโครงข่ายประสาทเทียมที่เรียกว่า มัลติเลเยอร์เพอเซ็ปตรอน (Multi-layer Perceptron, MLP) จะนำมาประยุกต์ใช้ในงานนี้ โครงสร้างมัลติเลเยอร์เพอเซ็ปตรอนประกอบด้วยชั้นของโหนด 3 ประเภท ได้แก่ 1) ชั้นข้อมูลเข้า (Input Layer) ประกอบด้วยโหนดต่างๆ ที่ได้จากตัวแปรต้น 2) ชั้นซ่อน (Hidden Layer) และ 3) ชั้นข้อมูลออก (Output Layer) สำหรับชั้นซ่อนและชั้นข้อมูลออกภายในจะประกอบด้วยโหนดสำหรับประมวลผลรูปที่ 1 แสดงลักษณะโครงสร้างมัลติเลเยอร์เพอเซ็ปตรอน



รูปที่ 1. โครงสร้างมัลติเลเยอร์เพอร์เซปตรอน (Multi-layer Perceptron, MLP)

แต่ละโหนดจะมีค่าน้ำหนักซึ่งแทนด้วยเส้นเชื่อมระหว่างโหนด  $i$  กับทุกโหนดในชั้นก่อนหน้า ในชั้นซ่อนและชั้นข้อมูลออกสามารถคำนวณค่าผลลัพธ์ของโหนดจากฟังก์ชันซิกมอยด์ (sigmoid function) ดังนี้

$$f(n_i) = \frac{1}{1 + e^{-n_i}} \quad (1)$$

โดยที่ข้อมูลเข้าของโหนด  $i$ ,  $n_i = \sum_{j=1}^L w_{ij}x_j + b_i$ ,  $w_{ij}$  คือค่าน้ำหนัก (weight) ที่เชื่อมระหว่างโหนดกับโหนด  $j$  ในชั้นก่อนหน้า,  $b_i$  คือค่าไบแอส (bias) ของโหนด  $i$  โดยค่าน้ำหนักและค่าไบแอสทั้งหมดสามารถคำนวณได้จากวิธีการ Levenberg-Marquardt ซึ่งเป็นวิธีที่พัฒนาจากวิธีการนิวตัน (Newton's Method) ในการประมาณค่าเฮเซียนเมทริกซ์ (Hessian Matrix) ในรูปของจาโคเบียนเมทริกซ์ ซึ่งวิธีการ Levenberg-Marquardt เป็นวิธีการประมาณค่าพารามิเตอร์สูงสุดที่มีประสิทธิภาพมากกว่าวิธีการ gradient descent (7)

สำหรับการสร้างตัวแบบ ANN ในการจำแนกผู้ป่วยโรคมะเร็งตับในงานนี้ ผู้วิจัยใช้ Neural Network Toolbox ในโปรแกรม MATLAB ในที่นี้เลือกใช้ตัวแบบที่เรียกว่า Multilayer Perceptron (MLP) โดยใช้ Levenberg-Marquardt algorithm สำหรับการหาค่า weight ที่เหมาะสม

และใช้ฟังก์ชันซิกมอยด์ (sigmoid function) ในการหาค่าผลลัพธ์ของแต่ละโหนด สำหรับงานนี้คณะผู้วิจัยเลือกใช้ฟังก์ชันค่าเฉลี่ยความผิดพลาดกำลังสอง (Mean Square Error, MSE) เป็น Cost Function สำหรับกระบวนการ Minimization โดยกำหนดค่า  $MSE = 0.01$  จำนวนรอบสูงสุดในการปรับค่าน้ำหนักเท่ากับ 5,000 รอบ และเลือกใช้ฟังก์ชัน Nguyen-Widrow layer initialization function เป็นฟังก์ชันในการกำหนดค่าน้ำหนักเริ่มต้น (Weight Initialization) โดยในการจำแนก คณะผู้วิจัยศึกษาประสิทธิภาพการจำแนกของตัวแบบ 18 ตัวแบบ โดยพิจารณาอัตราส่วนของข้อมูล Training set: Testing set: Validation set แตกต่างกัน ใน 3 กรณีคือ 1) 70 : 20 : 10 2) 65 : 20 : 15 และ 3) 65 : 15 : 20 ภายใต้จำนวน Hidden Layer ที่แตกต่างกัน 6 ระดับ ได้แก่ 5, 10, 15, 20, 25 และ 30 ทั้งนี้ในแต่ละตัวแบบทำการทดลอง 30 ครั้ง โดยในงานนี้เนื่องจากจำนวนของผู้ที่เป็นมะเร็งในตับกับผู้ที่ไม่เป็นมะเร็งมีจำนวนเท่ากันดังนั้นผู้วิจัยจึงใช้วิธีการคำนวณเปอร์เซ็นต์ความถูกต้องในการจำแนก (Percentage of classification accuracy) มาเป็นตัววัดประสิทธิภาพของตัวแบบโดยพิจารณาจากค่าความถูกต้องเมื่อทดสอบกับ Test set และค่าเฉลี่ยความถูกต้องโดยรวม (คำนวณ

จากค่าเฉลี่ยความถูกต้องจาก Training set Test set และ Validation set ของแต่ละตัวแบบ) จากนั้นเลือก โครงสร้างตัวแบบที่เหมาะสม โดยพิจารณาตัวแบบที่มีค่า

เปอร์เซ็นต์ความถูกต้องทั้งสองในการจำแนกค่าสูงที่สุด ซึ่งเปอร์เซ็นต์ความถูกต้องในการจำแนก (Percentage of classification accuracy) สามารถคำนวณได้จาก

$$\text{เปอร์เซ็นต์ความถูกต้องในการจำแนก} = \frac{\text{จำนวนตัวอย่างที่ตัวแบบทำนายถูกต้อง}}{\text{จำนวนตัวอย่างทั้งหมด}} \times 100 \quad (2)$$

### 2.3 การนำไปใช้ (Deployment)

เมื่อได้ตัวแบบที่ให้ค่าความถูกต้องมากที่สุดแล้ว ผู้วิจัยมีการจัดทำคู่มือสำหรับการใช้งานตัวแบบการจำแนกผู้ป่วยโรคมะเร็งตับ โดยให้เตรียมข้อมูลผู้ป่วยรายใหม่ที่จะนำมาจำแนกไว้ในไฟล์ Excel และเขียนแนะนำการนำข้อมูลเข้าตัวแบบการจำแนกในโปรแกรม MATLAB พร้อมแนะนำการแปลผล

## 3. ผลการวิจัยและอภิปรายผล

### 3.1 ผลการวิจัย

เมื่อทำการตรวจสอบคุณภาพข้อมูล (clean and validation) และแปลงข้อมูล (transform) สุดท้ายได้ตัวแปรที่จะนำมาใช้ในงานวิจัยนี้ทั้งสิ้น 8 ตัวแปรแล้วได้ข้อมูลที่สมบูรณ์รวมทั้งสิ้น 6,499 รายเป็นผู้ป่วยมะเร็งตับจำนวน 3,412 ราย โดยเป็นเพศชายจำนวน 1,281 คน และเป็นเพศหญิงจำนวน 2,131 คนดังตารางที่ 3

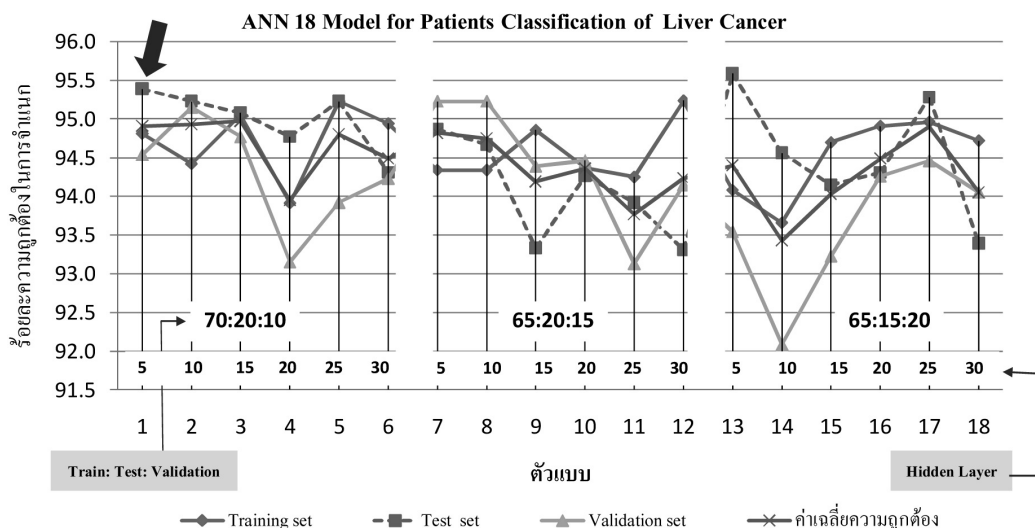
ตารางที่ 3. จำนวน record ที่นำมาศึกษาจำแนกตาม เพศ อาชีพ และผลการวินิจฉัย (เป็น/ไม่เป็น)

| อาชีพ                                              | ชาย  |             | หญิง |             | รวมทั้งสิ้น |       |             |        | รวมทั้งสิ้น |
|----------------------------------------------------|------|-------------|------|-------------|-------------|-------|-------------|--------|-------------|
|                                                    | เป็น | ไม่<br>เป็น | เป็น | ไม่<br>เป็น | เป็น        | %     | ไม่<br>เป็น | %      |             |
| 1. รับราชการ                                       | 442  | 256         | 470  | 36          | 912         | 75.75 | 292         | 24.25  | 1204        |
| 2. ครู                                             | 115  | 70          | 206  | 14          | 321         | 79.26 | 84          | 20.74  | 405         |
| 3. งานบ้านที่ไม่ได้ไปรับจ้างผู้อื่น                | 17   | 13          | 255  | 95          | 272         | 71.58 | 108         | 28.42  | 380         |
| 4. เกษตรกร                                         | 49   | 1079        | 205  | 462         | 254         | 14.15 | 1541        | 85.85  | 1795        |
| 5. งานรับจ้างทั่วไป ที่ไม่ใช่งานประจำหรือข้าราชการ | 175  | 215         | 122  | 55          | 297         | 52.38 | 270         | 47.62  | 567         |
| 6. นักเรียน หรือนักศึกษา                           | 46   | 3           | 31   | 6           | 77          | 89.53 | 9           | 10.47  | 86          |
| 7. เด็กในปกครองได้แก่เด็กหรือผู้พิการและทุพพลภาพ   | -    | 7           | -    | 7           | 0           | 0.00  | 14          | 100.00 | 14          |
| 8. นักบวช                                          | 10   | 36          | 2    | -           | 12          | 25.00 | 36          | 75.00  | 48          |
| 9. ค้าขายหรือประกอบกิจการส่วนตัว                   | 47   | 98          | 125  | 55          | 172         | 52.92 | 153         | 47.08  | 325         |
| 10. คนชราที่มีอายุ 60 ปีขึ้นไป                     | 325  | 286         | 641  | 181         | 966         | 67.41 | 467         | 32.59  | 1433        |
| 11. บุคคลทางสาธารณสุข                              | 23   |             | 39   | 1           | 62          | 98.41 | 1           | 1.59   | 63          |
| 12. อื่นๆ                                          | 32   | 86          | 35   | 26          | 67          | 37.43 | 112         | 62.57  | 179         |
| รวมทั้งสิ้น                                        | 1281 | 2149        | 2131 | 938         | 3412        | 52.50 | 3087        | 47.50  | 6499        |

เมื่อทำการทดสอบความสัมพันธ์ระหว่างการวินิจฉัยกับตัวแปรเพศ อายุ ค่าซีรัมอัลบิวมิน ค่าซีรัมไคเร็กบิรูบิน ค่าซีรัมแอสปาเทอะมิโนทรานสเฟอเรส ค่าซีรัมอัลคาไลด์ฟอสฟาเทส และค่าซีรัมอะลานินอะมิโนทรานสเฟอเรส ด้วยสถิติทดสอบ Pearson Chi-square พบว่ามีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติ ( $p < 0.001$ )

ผลจากการสร้างตัวแบบ 18 ตัวแบบ ภายใต้การทดลองที่กำหนดอัตราส่วนของข้อมูล 3 set ใน 3 กรณีและจำนวน Hidden Layer ที่แตกต่างกัน 6 ระดับ ได้ผลดังรูปที่

1 โดยตัวแบบที่ 1 ให้ค่าความถูกต้องในการจำแนกสูงที่สุด คือ จำนวนโหนดใน Hidden Layer เป็น 5 และอัตราส่วนของข้อมูล Training set, Testing set และ Validation set เป็น 70:20:10 ตามลำดับ มีค่าเปอร์เซ็นต์ความถูกต้องของ Test set เท่ากับ 95.39 และค่าเฉลี่ยเปอร์เซ็นต์ความถูกต้องเท่ากับ 94.91 นอกจากนี้จากรูปที่ 2 จะพบว่าไม่ว่าจะเป็นอัตราส่วนของข้อมูล 3 set แบบใดก็ตาม จำนวน Hidden Layer ที่ 5 จะให้ค่าเปอร์เซ็นต์ความถูกต้องของ Test set สูงที่สุดเสมอ



รูปที่ 2. เปรียบเทียบค่าเปอร์เซ็นต์ความถูกต้องของทั้ง 18 ตัวแบบ

สำหรับการประเมินผลการนำไปใช้ ด้านคู่มือการใช้งานพบว่า อธิบายการทำงานได้เข้าใจง่ายสามารถปฏิบัติตามขั้นตอนได้จริง รวมทั้งการใช้ภาพประกอบต่างๆ ทำให้ปฏิบัติตามได้ง่าย แต่คำอธิบายสั้นเกินไป ส่วนด้านตัวแบบการจำแนกผู้ป่วยโรคมะเร็งตับ พบว่ามีการเขียนโปรแกรมเพื่อช่วยในการแปลงข้อมูลผู้ป่วยใหม่ ก่อนนำเข้าไปทำนายในตัวแบบ ทำให้ง่ายต่อการใช้งาน ตัวแบบสามารถตอบสนองความต้องการของผู้ใช้งานได้ดี เพียงแต่ในการนำไปใช้จริงอาจต้องมีผู้วิเคราะห์ข้อมูลเหล่านี้ในเบื้องต้น ก่อนให้กับแพทย์ใช้ประกอบการวินิจฉัย

### 3.2 อภิปรายผลและข้อเสนอแนะ

ข้อมูลที่นำมาประมวลผลในครั้งนี้เป็นข้อมูลของผู้ป่วยที่มีรักษาตัวที่ โรงพยาบาลศรีนครินทร์ จังหวัดขอนแก่น ในช่วงระยะเวลา ตั้งแต่วันที่ 1 มกราคม พ.ศ. 2553 ถึงวันที่ 31 ธันวาคม พ.ศ. 2554 เท่านั้น ดังนั้นเมื่อวันเวลาเปลี่ยนไปนวัตกรรมใหม่ๆ ในการรักษาอาจมีการปรับเปลี่ยน ซึ่งอาจส่งผลกระทบต่อปัจจัยในการเกิดโรค ดังนั้นควรประมวลผลเพื่อหาตัวแบบที่เหมาะสมใหม่อย่างต่อเนื่อง และหากมีการนำผู้ป่วยจากพื้นที่อื่นมาประมวลผลร่วมด้วยอาจได้ตัวแบบที่ทำนายได้ครอบคลุมยิ่งขึ้น

สำหรับตัวแปรต่างๆ ที่นำมาใช้ในการประมวลผลครั้งนี้ นับว่าเป็นตัวแปรที่สามารถนำมาใช้ในการทำนาย

โรคได้ดี ดังจะเห็นได้จากค่าความถูกต้องในการทำนายของ Test set สูงถึง 95.39 % ค่าเฉลี่ยของความถูกต้องโดยรวมในการทำนายสูงถึง 94.91 % ซึ่งเป็นไปได้ที่ในอนาคตอาจจะมีตัวแปรอื่นเพิ่มเติม เพื่อให้การทำนายถูกต้องยิ่งขึ้น

#### 4. สรุปผลการวิจัย

จากการทดลองตัวแบบโครงข่ายประสาทเทียม ที่มีค่า Performance of accuracy สูง เมื่อกำหนดจำนวน Hidden Layer เป็น 5 และอัตราส่วนของข้อมูล Training set, Testing set และ Validation set เป็น 70:20:10 ตามลำดับ จะได้ตัวแบบที่ให้ค่าความถูกต้องของ Test set 95.39% และค่าเฉลี่ยความถูกต้องโดยรวม 94.91% และไม่ว่าจะเป็นอัตราส่วนของข้อมูล 3 set แบบใดก็ตาม จำนวน Hidden Layer ที่ 5 จะให้ค่าเปอร์เซ็นต์ความถูกต้องของ Test set สูงที่สุดเสมอสามารถนำไปใช้เพื่อประกอบการสนับสนุนการตัดสินใจของแพทย์ได้ในเบื้องต้น และจากการประเมินทดลองนำไปใช้ในกลุ่มผู้ใช้ (เจ้าหน้าที่ห้องปฏิบัติการ แผนกเวชศาสตร์ชันสูตร) พบว่าความพึงพอใจด้านคู่มือการใช้งานและการใช้งานตัวแบบในการจำแนกผู้ป่วยรายใหม่อยู่ในระดับดี คู่มือง่ายต่อการทำความเข้าใจและใช้ได้ง่าย

#### 5. กิตติกรรมประกาศ

ขอขอบคุณผู้อำนวยการโรงพยาบาลศรีนครินทร์จังหวัดขอนแก่น ที่อนุญาตให้ข้อมูล และขอขอบคุณบุคลากรของโรงพยาบาลศรีนครินทร์ ที่ให้ความสะดวกในการเข้าศึกษาข้อมูลต่างๆ และสุดท้ายนี้ขอขอบคุณผู้ป่วยทุกท่านที่เป็นเจ้าของข้อมูลเหล่านี้

#### 6. เอกสารอ้างอิง

- (1) Thai Society of Clinical Oncology. Liver cancer. Bangkok; 2010.
- (2) Department of Disease Control. Eating medium rare foods. It's risk for liver cancer. Bangkok; 2011.
- (3) Nongyao N, Punnee S. Classification of patients with ischemic heart disease and other forms of heart disease patients by using data mining techniques. Khon Kaen; 2011.
- (4) Cheng-Mei C, Chien-Yeh H, Hung-Wen C, Hsiao-Hsien R. Prediction of survival in patients with liver cancer using artificial neural networks and classification and regression trees. Proceedings of the 7th International Conference on Natural Computation, IEEE; 2011, P. 811-815.
- (5) Suwannee M, A Prognostic model to predict the bile duct cancer. Khon Kaen; 2010.
- (6) Simon H. Neural network, a comprehensive foundation. 2nd ed. New York: Macmillan College Publishing Company; 1994.
- (7) Sirilak A, Prem J, Auttapon P. Prediction of dissolved oxygen using artificial neural network. Proceedings of International Conference on Computer Communication and Management; 2011 [cited 2011 Aug 30]. Available from: [http://www.researchgate.net/publication/228533526\\_Prediction\\_of\\_Dissolved\\_Oxygen\\_Using\\_Artificial\\_Neural\\_Network](http://www.researchgate.net/publication/228533526_Prediction_of_Dissolved_Oxygen_Using_Artificial_Neural_Network).