บทความวิจัย

การจำแนกผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียม

วงกต ศรีอุไร*

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อสร้างโมเดลในการจำแนกข้อมูลผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และเพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูล 3 วิธี ได้แก่ 1) โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนเพียงอย่างเดียว 2) วิธี Correlation–based Feature Selection ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และ 3) วิธี Information Gain ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน สำหรับข้อมูลที่ใช้ได้มาจากการเก็บข้อมูลโดยใช้แบบสอบถามจากบุคลากรในโรงพยาบาลค่ายสรรพสิทธิประสงค์ตั้งแต่ปี พ.ศ. 2550–2554 ผลการวิจัยพบว่าการใช้วิธีการคัดเลือกคุณลักษณะด้วย Correlation–based Feature Selection ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนในการสร้างโมเดลการจำแนกข้อมูลจะให้ค่าความถูกต้องในการจำแนกข้อมูลสูงที่สุดซึ่งมีค่าเท่ากับ 92.56%

คำสำคัญ: การจำแนกข้อมูล การคัดเลือกคุณลักษณะ โครงข่ายประสาทเทียม

ภาควิชาคณิตศาสตร์ สถิติและคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยอุบลราชธานี

*ผู้นิพนธ์ประสานงาน, email: scwongsr@ubu.ac.th

Patients Classification of Metabolic Syndrome Using Feature Selection and Artificial Neural Network

Wongkot Sriurai*

ABSTRACT

The objectives of this research are to create a model of metabolic syndrome classification using Feature Selection and Multi-layer Perceptron Neural Network and to compare the efficiency of three classification methods: 1) Multi-layer Perceptron Neural Network, 2) Correlation-based Feature Selection with Multi-layer Perceptron Neural Network and 3) Information Gain with Multi-layer Perceptron Neural Network. The data is collected from the medical records of the Fort Sumpasitthiprasong hospital, from 2007 to 2011, by using questionnaire. The results showed that the using Correlation-based Feature Selection with Multi-layer Perceptron Neural Network for creating classification model yielded the best performance based on accuracy equal to 92.56%.

Keywords: Classification, Feature Selection, Artificial Neural Network

Department of Mathematics Statistics and Computer, Faculty of Science, Ubon Ratchathani University
*Corresponding author, e-mail: scwongsr@ubu.ac.th

บทนำ

อ้วนลงพุง (metabolic syndrome) [1] เป็นกลุ่มอาการที่มีความผิดปกติทางเมแทบอลิซึม ทำให้เกิดการสะสมของไขมันในช่องท้อง ซึ่งเป็นปัจจัยเสี่ยงต่อการเกิดโรคความดันโลหิตสูง โรคเบาหวาน โรคหัวใจและหลอดเลือด สาเหตุของภาวะอ้วนลงพุงเป็นผลมาจากโรคอ้วนและภาวะดื้อต่ออินซูลิน คนที่มีภาวะดื้อต่ออินซูลินเสี่ยงต่อการเกิดโรคหลอดเลือด ซึ่งข้อมูลสถิติจากกระทรวงสาธารณสุขในปี พ.ศ. 2551 พบว่าเด็กไทยอายุ 2–18 ปี จำนวน 17.6 ล้านคน เป็นโรคอ้วนประมาณร้อยละ 8 และกลุ่มวัยรุ่นอายุ 13–18 ปี เป็นวัยที่เป็นโรคอ้วนมากที่สุดคิดเป็นร้อยละ 9 [2] โดยสาเหตุของโรคเกิดจากพฤติกรรมการขาดการออกกำลังกาย ชอบนั่งดูโทรทัศน์ หรือเล่นเกมคอมพิวเตอร์ทั้งวัน และยังบริโภคอาหารประเภทขนม น้ำหวาน อาหารสำเร็จรูปเกินความจำเป็น จากสถิติดังกล่าวทำให้มีจำนวนผู้ป่วยที่เป็นโรคอ้วนลงพุงเพิ่มขึ้นอย่างต่อเนื่อง แม้ว่าจะมีการรณรงค์ เพื่อกระตุ้นให้มีการตื่นตัวให้คนควบคุมปริมาณอาหารที่กินในแต่ละวัน พร้อมทั้งการออกกำลังกายร่วมด้วยเพื่อลดภาวะที่เสี่ยงต่อการเกิดโรค แต่วิธีนี้ไม่สามารถลดจำนวนผู้ป่วยที่มีความเสี่ยงต่อการเกิดโรคนี้ได้ ปัจจุบันพบว่าจำนวนผู้ที่เข้ารับการตรวจวินิจฉัยโรคอ้วนลงพุงเพิ่มขึ้นอย่างต่อเนื่อง ซึ่งการวินิจฉัยโรคนี้กระทำโดยแพทย์ ซึ่งแพทย์จะต้องนำผลการตรวจจากห้องปฏิบัติการ เช่น ระดับน้ำตาลในเลือด ไขมันไตรกลีเซอร์ไรด์ และคอเลสเตอรอลชนิดดี (HDL cholesterol) เป็นต้น มาวิเคราะห์ร่วมกับการตรวจวินิจฉัยจากแพทย์ ซึ่งวิธีการตรวจวินิจฉัยแบบนี้จะต้องใช้เวลาในการวินิจฉัยค่อนข้างมาก เนื่องจากแพทย์จะต้องเป็นผู้วิเคราะห์ผลตรวจจากห้องปฏิบัติการของผู้ป่วยในแต่ละรายด้วยตนเอง ดังนั้นการคัดกรองผู้ป่วยเบื้องต้นก่อนจึงเป็นสิ่งจำเป็นเพื่อจะทำให้ได้ข้อมูลเบื้องต้นประกอบการวินิจฉัยของแพทย์ซึ่งบางโรงพยาบาลจะให้ผู้ป่วยตอบแบบสอบถามเกี่ยวกับพฤติกรรมการกินอาหารของผู้ป่วย จากนั้นนำคำตอบที่ได้จากแบบสอบถามมาประเมินในเบื้องต้นว่าผู้ป่วยมีความเสี่ยงต่อการเกิดโรคอ้วนลงพุงหรือไม่ หากผลการประเมินออกมาว่าผู้ป่วยมีความเสี่ยงก็จะส่งผู้ป่วยไปตรวจกับแพทย์ผู้เชี่ยวชาญต่อไป จะเห็นได้ว่าวิธีการนี้เป็นวิธีที่ช่วยในการคัดกรองผู้ป่วยในเบื้องต้น ข้อมูลที่ได้จากการตอบแบบสอบถามจะเป็นข้อมูลที่ใช้ประกอบการวินิจฉัยของแพทย์ วิธีการคัดกรองผู้ป่วยโดยใช้แบบสอบถามนี้เป็นวิธีที่นิยมใช้ แต่ผลการตอบแบบสอบถามจะต้องใช้คนในการตรวจสอบ ซึ่งหากในโรงพยาบาลที่มีผู้ป่วยจำนวนมาก การวิเคราะห์ผลจากแบบสอบถามอาจทำได้ไม่ทั่วถึง ดังนั้นหากมีการนำเทคนิคหรือวิธีการที่สามารถนำมาใช้วิเคราะห์ข้อมูลผู้ป่วยและคัดกรองผู้ป่วยที่มีความเสี่ยงต่อการเกิดโรคอ้วนลงพุงได้จะช่วยให้การตรวจวินิจฉัยของแพทย์สะดวกและรวดเร็วยิ่งขึ้น

จากปัญหาจำนวนผู้ป่วยโรคอ้วนลงพุงที่มีจำนวนเพิ่มขึ้น ทำให้แพทย์จะต้องใช้เวลามากในการตรวจวินิจฉัย แนวทางในการแก้ไขปัญหานี้ คือ จะต้องสร้างโมเดลเพื่อใช้ในการวิเคราะห์ข้อมูลของผู้ป่วยในเบื้องต้น ซึ่งโมเดลนี้จะต้องสามารถนำไปคัดกรองผู้ป่วยที่มีความเสี่ยงต่อการเกิดโรคอ้วนลงพุงได้และข้อมูลที่ได้จากโมเดลสามารถนำมาใช้ประกอบการวินิจฉัยของแพทย์ได้ ดังนั้นงานวิจัยนี้จึงนำเสนอโมเดลเพื่อใช้ในการวิเคราะห์ข้อมูลของผู้ป่วยที่มีความเสี่ยงต่อการเกิดโรคอ้วนลงพุงโดยประยุกต์ใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียม

วัตถุประสงค์ของการวิจัย

1. เพื่อสร้างโมเดลในการจำแนกข้อมูลผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะและโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน

2. เพื่อเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลผู้ป่วยโรคอ้วนลงพุงใน 3 วิธี ได้แก่ 1) โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนเพียงอย่างเดียว 2) วิธีการคัดเลือกคุณลักษณะโดยใช้ Correlation–based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และ 3) วิธีการคัดเลือกคุณลักษณะโดยใช้ Information Gain (IG) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน

ทฤษฎีที่เกี่ยวข้อง

1. โรคอ้วนลงพุง (metabolic syndrome) [1, 3] เป็นภาวะอ้วนโดยเฉพาะส่วนเอวซึ่งเป็นความผิดปกติของกระบวนการเผาผลาญอาหารในร่างกาย หรืออาจเรียกอีกอย่างได้ว่าอาการดื้อต่ออินซูลิน โดยปกติฮอร์โมนอินซูลินมีหน้าที่ในการนำน้ำตาลจากกระแสเลือดเข้าสู่เซลล์เพื่อเปลี่ยนให้เป็นพลังงาน หากร่างกายดื้อต่ออินซูลินจะมีผลทำให้น้ำตาลไม่ถูกนำไปใช้ น้ำตาลในเลือดจึงมีระดับสูง หากสูงมากจะส่งผลให้เกิดโรคเบาหวาน โรคหัวใจและหลอดเลือด

เกณฑ์การตัดสินว่าเป็นโรคอ้วนลงพุง [1, 3] คือ

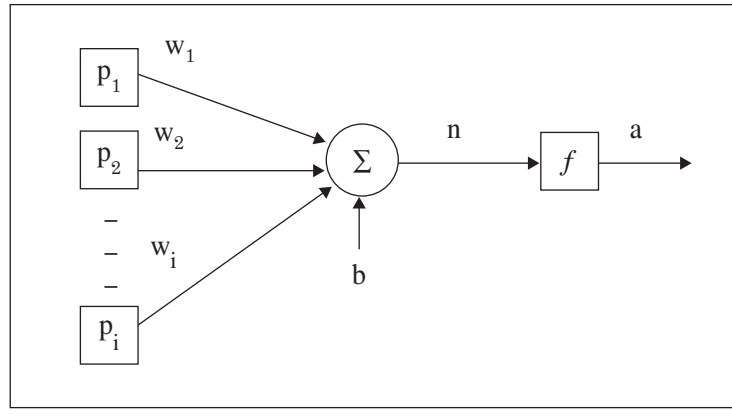
1. เส้นรอบเอวเกินมาตรฐาน (ผู้ชายมากกว่า 90 เซนติเมตร ผู้หญิงมากกว่า 80 เซนติเมตร)
2. ความดันโลหิตมีค่ามากกว่า 130/85 มิลลิเมตรปรอท
3. มีคอเลสเตอรอลชนิดดี (HDL cholesterol) น้อยกว่า 40 มิลลิกรัม/เดซิลิตรในผู้ชายและน้อยกว่า 50 มิลลิกรัม/เดซิลิตรในผู้หญิง
4. มีไขมันไตรกลีเซอไรด์ตั้งแต่ 150 มิลลิกรัมต่อเดซิลิตรขึ้นไป ระดับน้ำตาลในเลือดเริ่มสูงขึ้นคือ ตั้งแต่ 100 มิลลิกรัมต่อเดซิลิตรขึ้นไป

การป้องกันและรักษาโรคอ้วนลงพุง [1, 3] ทำได้โดยการปรับเปลี่ยนพฤติกรรมการกินและการออกกำลังกาย การออกกำลังกายจะช่วยป้องกันการเพิ่มของน้ำหนักตัวและช่วยให้ร่างกายใช้ฮอร์โมนอินซูลินได้อย่างมีประสิทธิภาพ ดังนั้นควรออกกำลังกายให้ได้วันละ 30 นาที โดยการเดินเร็ว ว่ายน้ำ หรือเต้นแอโรบิค การควบคุมคุณภาพและปริมาณของคาร์โบไฮเดรตเป็นสิ่งสำคัญในการป้องกันและรักษาโรคอ้วนลงพุง พลังงานประมาณ 40–50 เปอร์เซ็นต์จากคาร์โบไฮเดรตควรมาจากธัญพืชไม่ขัดสี ผักและผลไม้ พลังงานอีก 40 เปอร์เซ็นต์ ควรได้จากไขมันดีจากปลาและพืชเช่น น้ำมันมะกอก น้ำมันปลา พลังงานที่เหลืออีก 10–20 เปอร์เซ็นต์ ควรมาจากโปรตีนไขมันต่ำ อาหารทะเล เนื้อสัตว์ไม่ติดมัน

2. การจำแนกประเภทข้อมูล (Data Classification) [4, 5] เป็นเทคนิคที่ใช้ในการจำแนกกลุ่มข้อมูล โดยใช้กระบวนการสร้างโมเดลจำแนกประเภทข้อมูล (Data Classification Model) เพื่อทำนายกลุ่มของข้อมูลใหม่ (unseen data) การสร้างโมเดลจำแนกประเภทข้อมูลเกิดขึ้นจากการหาความสัมพันธ์ของข้อมูลในฐานข้อมูลขนาดใหญ่ โดยข้อมูลทั้งหมดมีการแบ่งออกเป็น 2 กลุ่ม คือ กลุ่มข้อมูลที่ใช้ในการฝึกฝน (training dataset) เป็นชุดข้อมูลที่ใช้ในการสร้างโมเดลจำแนกประเภทข้อมูล

และกลุ่มข้อมูลที่ใช้ในการทดสอบ (testing dataset) เป็นชุดข้อมูลที่ใช้ในการประเมินความถูกต้องของโมเดลจำแนกประเภทข้อมูล เทคนิคที่ใช้ในการจำแนกมีหลากหลายเทคนิค ได้แก่ ต้นไม้ตัดสินใจ (Decision Tree) โครงข่ายประสาทเทียม (Artificial Neural Network) และซัพพอร์ตเวคเตอร์แมชชีน (Support Vector Machine) [4, 5] เป็นต้น ปัจจุบันได้มีการนำเทคนิคดังกล่าวมาใช้ในการจำแนกข้อมูลในด้านต่างๆ โดยเฉพาะทางการแพทย์ที่มีการนำเทคนิคโครงข่ายประสาทเทียมมาใช้ในการจำแนกข้อมูลผู้ป่วยเพื่อทำนายโรคต่างๆ ที่อาจจะเกิดขึ้นกับผู้ป่วย

โครงข่ายประสาทเทียม (Artificial Neural Network) [4, 6] เป็นกระบวนการที่คอมพิวเตอร์จำลองการทำงานโดยเลียนแบบการทำงานของสมองมนุษย์ วัตถุประสงค์ของโครงข่ายประสาทเทียม คือให้คอมพิวเตอร์มีความฉลาดในการเรียนรู้เหมือนที่มนุษย์มีการเรียนรู้ และสามารถนำความรู้ไปใช้แก้ไขปัญหาต่างๆ ซึ่งในปัจจุบันบางปัญหาที่พบในชีวิตประจำวันเป็นปัญหาที่มีความซับซ้อนสูง ไม่สามารถที่จะแก้ปัญหาได้โดยใช้วิธีการดั้งเดิมได้ จึงจำเป็นที่จะต้องพัฒนาโครงข่ายประสาทเทียมขึ้นมาเพื่อใช้ในงานที่มีการตัดสินใจได้เทียบเท่ากับมนุษย์ เช่น การพยากรณ์สภาพภูมิอากาศ การจำแนกเซลล์มะเร็ง และการรู้จำใบหน้ามนุษย์ เป็นต้น การเรียนรู้ของนิวรอลเน็ตเวิร์กทำได้โดยการส่งข้อมูลเข้ามายังส่วนที่เรียกว่าเพอร์เซ็ปตรอน (perceptron) ซึ่งสามารถเทียบได้กับเซลล์สมองของมนุษย์ โดยเพอร์เซ็ปตรอนจะรับข้อมูลที่เป็นตัวเลขในรูปของเมทริกซ์เข้ามาคำนวณ ดังรูปที่ 1



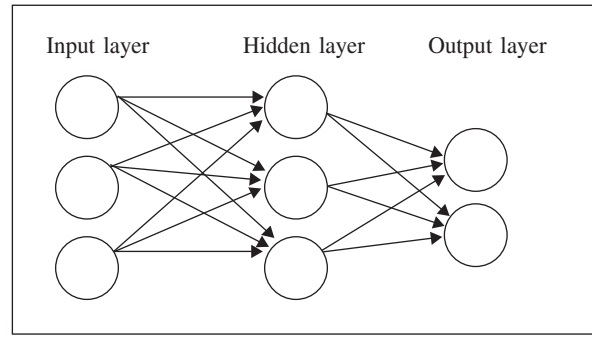
รูปที่ 1 โครงข่ายประสาทเทียมเพอร์เซ็ปตรอน

ฟังก์ชันผลรวม (summation function)

$$n = \sum_{i=1}^{k} p_i w_i + b \qquad \text{(eq.1)}$$

โดยที่ n คือ ผลรวมที่ได้จากฟังก์ชันผลรวม ตัวแปร p_i เป็นค่าข้อมูลเข้าหน่วยประสาทเทียมตัวที่ i ตัวแปร w_i เป็นค่าถ่วงน้ำหนักของนิวรอลตัวที่ i ตัวแปร k เป็นจำนวนนิวรอลชั้นข้อมูลเข้า ตัวแปร b เป็นค่าความโน้มเอียง และตัวแปร i มีค่าตั้งแต่ 1 ถึง k

โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi–layer Perceptron: MLP) [4, 6] เป็นโครงข่ายประสาทเทียมที่มีโครงสร้างแบบหลายชั้น โดยประกอบด้วย 3 ชั้นหลัก คือ ชั้นรับข้อมูลเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นส่งข้อมูลออก (output layer) โดยมีกระบวนการเรียนรู้แบบมีผู้สอน (Supervised Learning) ดังรูปที่ 2



รูปที่ 2 โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน

ตัวอย่างของการประยุกต์ใช้เทคนิคโครงข่ายประสาทเทียม เช่น ในงานวิจัยของภรัณยา อำมฤครัตน์และพยุง มีสัจ [7] ได้จำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีโครงข่ายประสาทเทียม ในงานวิจัยนี้ได้ทำการเปรียบเทียบวิธีการจำแนกข้อมูลโรคลมร้อนด้วยซัพพอร์ตเวกเตอร์แมชชีนกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ซึ่งผลการทดลองพบว่า การใช้แบบจำลองโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนให้ประสิทธิภาพในการจำแนกข้อมูลได้ดีกว่าการใช้ซัพพอร์ตเวกเตอร์แมชชีน และงานวิจัยของวิชุดา ไชยศิวามงคล และคณะ [8] ได้นำเสนอตัวแบบในการจำแนกผู้ป่วยโรคมะเร็งตับโดยประยุกต์ใช้โครงข่ายประสาทเทียม ซึ่งคณะผู้วิจัยได้สร้างตัวแบบในการจำแนกจำนวน 18 ตัวแบบ จากนั้นทำการทดสอบตัวแบบแต่ละตัวโดยใช้ข้อมูล training set, test set และ validation set แบ่งข้อมูลออกเป็น 70 : 20 : 10 ตามลำดับ ผลการทดสอบพบว่าตัวแบบโครงข่ายประสาทเทียมที่ได้มีความถูกต้องสูงและสามารถนำไปใช้เพื่อประกอบการตัดสินใจของแพทย์ในเบื้องต้นได้

3. การคัดเลือกคุณลักษณะ

การคัดเลือกคุณลักษณะ (Feature Selection) [4] เป็นขั้นตอนการเลือกคุณลักษณะของข้อมูลก่อนการนำข้อมูลเข้าสู่กระบวนการจำแนกข้อมูลโดยใช้โมเดลของโครงข่ายประสาทเทียม ซึ่งแต่ละวิธีการคัดเลือกข้อมูลก็มีวิธีที่แตกต่างกัน โดยวัตถุประสงค์ของการคัดเลือกข้อมูลเพื่อคัดเลือกเฉพาะข้อมูลที่มีความสำคัญเท่านั้น ซึ่งข้อมูลเหล่านี้จะช่วยให้สังเคราะห์โมเดลได้อย่างรวดเร็ว และช่วยเพิ่มประสิทธิภาพในการจำแนกข้อมูล ในงานวิจัยนี้ได้เลือกวิธีการคัดเลือกคุณลักษณะมา 2 วิธี ได้แก่ 1) Correlation–based Feature Selection (CFS) เป็นวิธีคัดเลือกคุณลักษณะที่ได้รับการประเมินค่าจากความสามารถในการคาดการณ์ ซึ่งคุณลักษณะที่ถูกคัดเลือกมาใช้ในการจำแนกข้อมูลจะมีระดับความสัมพันธ์ที่เกี่ยวข้องกันภายใน [9] และ 2) Information Gain (IG) เป็นวิธีคัดเลือกข้อมูลโดยการวัดค่า Gain ของแต่ละโหนด หากโหนดใดที่มีค่า Gain สูงสุดจะถูกเลือกเป็นโหนดรากและนำข้อมูลที่เหลือมาหาค่า Gain อีกครั้งเพื่อให้ได้โหนดต่อไป [10]

วิธีดำเนินการวิจัย

1. รวบรวมข้อมูล

งานวิจัยนี้รวบรวมข้อมูลโดยใช้แบบสอบถามที่เป็นแบบบันทึกการตรวจคัดกรองยืนยันความเสี่ยงต่อภาวะโรคอ้วนลงพุงจากบุคลากรในโรงพยาบาลค่ายสรรพสิทธิประสงค์ตั้งแต่ปี พ.ศ. 2550–2554 จำนวนทั้งหมด 265 ราย โดยข้อมูลเหล่านี้จะถูกนำไปกำหนดค่าของคุณลักษณะในขั้นตอนต่อไป

2. เตรียมข้อมูลและกำหนดค่าของคุณลักษณะ

จากการรวบรวมข้อมูลผู้ป่วยโรคอ้วนลงพุงในข้อที่ 1 เรียบร้อยแล้ว ก่อนที่จะนำข้อมูลชุดนี้ไปสร้างโมเดล จะต้องมีการกำหนดค่าข้อมูลตามคุณลักษณะ (Attribute) ที่กำหนดขึ้น ดังแสดงในตารางที่ 1 และเมื่อกำหนดค่าข้อมูลแล้วข้อมูลนี้จะถูกนำไปใช้ในการสร้างโมเดล

ตารางที่ 1 คุณลักษณะ (Attribute) และค่าของคุณลักษณะสำหรับใช้ในการสร้างโมเดล

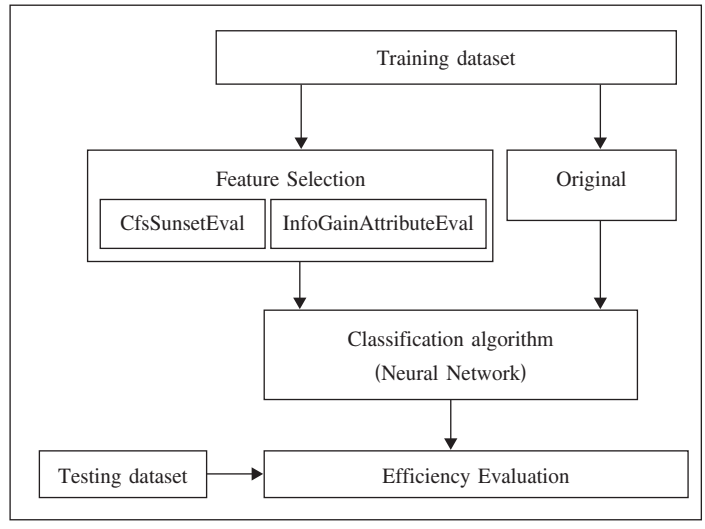
ลำดับที่	คุณลักษณะ (Attribute)	ค่าของคุณลักษณะ
1	ประวัติการเจ็บป่วยของบิดาหรือมารดา และพี่น้อง (สายตรง)	– โรคเบาหวาน – โรคความดันโลหิตสูง – โรคเกาต์ – โรคไตวายเรื้อรัง – โรคกล้ามเนื้อหัวใจตาย – โรคเส้นเลือดสมอง
2	ประวัติการเจ็บป่วยด้วยโรคหรืออาการต่างๆ ของผู้รับการตรวจ	– โรคเบาหวาน – โรคความดันโลหิตสูง – โรคตับ – โรคอัมพาต – โรคหัวใจ – ไขมัน/คอเลสเตอรอลสูง
3	การสูบบุหรี่	– สูบ – ไม่สูบ – เคยสูบแต่เลิกแล้ว
4	การออกกำลังกาย/เล่นกีฬา	– ทุกวัน ครั้งละ 30 นาที – สัปดาห์ละมากกว่า 3 ครั้ง ครั้งละ 30 นาที สม่ำเสมอ – สัปดาห์ละ 3 ครั้ง ครั้งละ 30 นาที สม่ำเสมอ – น้อยกว่าสัปดาห์ละ 3 ครั้ง – ไม่ออกกำลังกายเลย

ตารางที่ 1 (ต่อ)

ลำดับที่	คุณลักษณะ (Attribute)	ค่าของคุณลักษณะ
5	รสอาหารที่ชอบรับประทาน	– หวาน – เค็ม – มัน
6	น้ำหนัก	น้ำหนักของผู้ป่วย (ตัวเลขจำนวนเต็มในหน่วย กิโลกรัม)
7	ส่วนสูง	ส่วนสูงของผู้ป่วย (ตัวเลขจำนวนเต็มในหน่วยเซนติเมตร)
8	เส้นรอบเอว	เส้นรอบเอวของผู้ป่วย (ตัวเลขจำนวนเต็มในหน่วยเซนติเมตร)
9	ผลการวิเคราะห์โรคอ้วนลงพุง	– เป็นโรคอ้วนลงพุง – ไม่เป็นโรคอ้วนลงพุง

3. สร้างโมเดลและวัดประสิทธิภาพของโมเดล

ขั้นตอนนี้เป็นขั้นตอนการสร้างโมเดลสำหรับวิเคราะห์ข้อมูลของผู้ที่มีความเสี่ยงต่อการเกิดโรคอ้วนลงพุง โดยเริ่มจากนำชุดข้อมูลมาแบ่งเป็น 2 ชุด ได้แก่ ชุดข้อมูลที่ใช้ในการฝึกฝน (training dataset) และชุดข้อมูลทดสอบ (testing dataset) ในอัตราส่วน 80% : 20% ตามลำดับ จากนั้นนำชุดข้อมูลที่ใช้ในการฝึกฝนเข้าสู่กระบวนการสร้างโมเดลโดยใช้โปรแกรม Weka เวอร์ชัน 3.7.5 [5, 11] ซึ่งก่อนสร้างโมเดลจะต้องนำชุดข้อมูลนี้ไปผ่านกระบวนการคัดเลือกคุณลักษณะ (Feature Selection) โดยใช้วิธี Correlation-based Feature Selection (CFS) และใช้วิธี Information Gain (IG) ก่อน โดยวิธีการคัดเลือกคุณลักษณะทั้งสองวิธีนี้จะช่วยลดจำนวนคุณลักษณะ (attribute) ลง ซึ่งแต่ละวิธีจะคัดเลือกคุณลักษณะที่สำคัญสำหรับนำไปใช้สร้างโมเดลเท่านั้น หลังจากผ่านกระบวนการคัดเลือกคุณลักษณะแล้วจึงใช้อัลกอริทึมโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi–layer Perceptron: MLP) ในการสร้างโมเดล เมื่อได้โมเดลเรียบร้อยแล้วนำชุดข้อมูลทดสอบมาใช้ทดสอบกับโมเดลที่ได้เพื่อวัดประสิทธิภาพของโมเดล รายละเอียดดังรูปที่ 3



รูปที่ 3 ขั้นตอนการสร้างโมเดล

ในงานวิจัยนี้ประเมินประสิทธิภาพของโมเดลโดยใช้วิธีการวัดค่าความถูกต้อง (Accuracy) ในการจำแนกข้อมูล สามารถคำนวณได้จาก

$$\text{เปอร์เซ็นต์ความถูกต้องในการจำแนก} = \frac{\text{จำนวนตัวอย่างที่โมเดลทำนายถูกต้อง}}{\text{จำนวนตัวอย่างทั้งหมด}} \times 100 \quad \text{(eq.2)}$$

ผลการดำเนินการวิจัย

งานวิจัยนี้ใช้โปรแกรม Weka เวอร์ชัน 3.7.5 ในการสร้างโมเดลสำหรับวิเคราะห์ข้อมูลของผู้ที่มีความเสี่ยงต่อการเกิดโรคอ้วนลงพุง ซึ่งผู้วิจัยเลือกใช้โมดูล Multilayer Perceptron สำหรับเทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน โดยทำการปรับค่าพารามิเตอร์จำนวนโหนดของชั้นซ่อน (hidden layer) จาก 3 โหนดไปจนถึง 5 โหนด และปรับอัตราการเรียนรู้ (learning rate) เป็น 0.1 ไปจนถึง 0.3 และกำหนดค่าโมเมนตัม (momentum) เท่ากับ 0.3 เพื่อเปรียบเทียบค่าความถูกต้องของโมเดล ผลลัพธ์ที่ได้ดังแสดงในตารางที่ 2

ตารางที่ 2 ผลการเปรียบเทียบประสิทธิภาพของการสร้างโมเดลโดยใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน

Type	Model	Learning rate	Accuracy (%)
Original	8:3:2	0.1	85.28
	8:3:2	0.2	87.05
	8:3:2	0.3	86.92
	8:4:2	0.1	88.67
	8:4:2	0.2	89.18
	8:4:2	0.3	87.55
	8:5:2	0.1	88.06
	8:5:2	0.2	87.10
	8:5:2	0.3	86.62
CFS	8:3:2	0.1	90.19
	8:3:2	0.2	90.22
	8:3:2	0.3	89.81
	8:4:2	0.1	90.32
	8:4:2	**0.2**	**92.56**
	8:4:2	0.3	90.17
	8:5:2	0.1	90.21
	8:5:2	0.2	90.35
	8:5:2	0.3	90.19
IG	8:3:2	0.1	89.43
	8:3:2	0.2	87.54
	8:3:2	0.3	88.26
	8:4:2	0.1	88.67
	8:4:2	0.2	88.69
	8:4:2	0.3	87.64
	8:5:2	0.1	88.30
	8:5:2	0.2	88.34
	8:5:2	0.3	87.67

จากผลการเปรียบเทียบประสิทธิภาพของการสร้างโมเดลโดยใช้เทคนิคโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนในตารางที่ 2 แสดงให้เห็นว่าการสร้างโมเดลในการจำแนกข้อมูลด้วยวิธีการคัดเลือกคุณลักษณะด้วย Correlation–based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนให้ค่าความถูกต้อง (accuracy) ในการจำแนกเท่ากับ 92.56% ซึ่งให้ค่าความถูกต้องสูงกว่าวิธีคัดเลือกคุณลักษณะด้วย Information Gain (IG) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และสูงกว่าวิธีที่ใช้โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนเพียงอย่างเดียว โดยพบว่าโมเดลการเรียนรู้ที่ดีที่สุด คือ 8 : 4 : 2 ซึ่งหมายถึง ชั้นรับข้อมูลเข้าจำนวน 8 โหนด ชั้นซ่อนจำนวน 4 โหนดและชั้นส่งข้อมูลออกจำนวน 2 โหนด โดยใช้อัตราการเรียนรู้เท่ากับ 0.2 และค่าโมเมนตัมเท่ากับ 0.3

สรุปและอภิปรายผลการวิจัย

บทความนี้นำเสนอโมเดลในการจำแนกข้อมูลผู้ป่วยโรคอ้วนลงพุงโดยใช้วิธีการคัดเลือกคุณลักษณะร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และทำการเปรียบเทียบประสิทธิภาพในการจำแนกข้อมูลใน 3 วิธี ได้แก่ 1) โครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอนเพียงวิธีเดียว 2) วิธีการคัดเลือกคุณลักษณะด้วย Correlation–based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน และ 3) วิธีการคัดเลือกคุณลักษณะด้วย Information Gain (IG) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน จากผลการวิจัยพบว่าการสร้างโมเดลจำแนกข้อมูลโดยใช้วิธีการคัดเลือกคุณลักษณะด้วย Correlation–based Feature Selection (CFS) ร่วมกับโครงข่ายประสาทเทียมแบบมัลติเลเยอร์เพอร์เซ็ปตรอน ให้ค่าความถูกต้องในการจำแนกสูงที่สุด โดยผลการวิจัยนี้สอดคล้องกับงานวิจัยของ Lu และคณะ [12] ที่ได้นำเสนอวิธีการคัดเลือกคุณลักษณะของข้อมูลโรคมะเร็งโดยมีพื้นฐานจากวิธี Correlation–based Feature Selection (CFS) และได้เปรียบเทียบวิธีการคัดเลือกคุณลักษณะแบบต่างๆ ได้แก่ Correlation–based Feature Selection (CFS) Information Gain (IG) Principal Components Analysis (PCA) มาใช้คัดเลือกคุณลักษณะของข้อมูลโรคมะเร็ง แล้วนำคุณลักษณะที่ผ่านการคัดเลือกแล้วมาสร้างโมเดลในการจำแนกข้อมูล ผลของงานวิจัยนี้พบว่าวิธีการคัดเลือกคุณลักษณะด้วย Correlation–based Feature Selection (CFS) ช่วยลดจำนวนคุณลักษณะของข้อมูลและช่วยเพิ่มประสิทธิภาพในการจำแนกข้อมูล โดยมีประสิทธิภาพดีกว่าวิธี Principal Components Analysis (PCA) และวิธี Information Gain (IG)

จากผลลัพธ์ที่ได้จากการจำแนกข้อมูลนี้แสดงให้เห็นว่าวิธีการคัดเลือกคุณลักษณะของข้อมูลมีส่วนสำคัญที่จะเข้ามาช่วยเพิ่มประสิทธิภาพในการจำแนกข้อมูลได้ และจากค่าความถูกต้องของโมเดลในการจำแนกข้อมูลแสดงให้เห็นว่าโมเดลนี้เหมาะที่จะนำไปใช้งาน การวิจัยในครั้งต่อไปควรพิจารณานำผลการตรวจจากห้องปฏิบัติการมาใช้ในการสร้างโมเดลในการจำแนกด้วย เนื่องจากผลการตรวจจากห้องปฏิบัติการบางตัวอาจเป็นปัจจัยหนึ่งที่บ่งบอกถึงโรคแทรกซ้อนอื่นๆ ที่อาจเกิดขึ้นกับผู้ป่วยได้ ดังนั้นการสร้างโมเดลในอนาคตนอกจากจะนำมาใช้ในการจำแนกข้อมูลในโรคนั้นๆ แล้ว อาจจะเป็นโมเดลที่จะใช้คาดการณ์ได้อีกว่าผู้ป่วยน่าจะเกิดโรคแทรกซ้อนอะไรบ้าง ซึ่งข้อมูลเหล่านี้จะเป็นประโยชน์ต่อการรักษาโรคต่อไป

เอกสารอ้างอิง

1. ธงชัย ประฏิภาณวัตร. 2550. กลุ่มอาการอ้วนลงพุง (Metabolic syndrome). *วารสารอายุรศาสตร์อีสาน* 6(3): 207–217.
2. สำนักงานกองทุนสนับสนุนการสร้างเสริมสุขภาพ. 2551. วิกฤติโรคอ้วนป่วนไทย. ได้จาก http://www.thaihealth.or.th/healthcontent/article/2098.17 ตุลาคม 2556.
3. ชัยชาญ ดีโรจนวงศ์. 2550. Metabolic syndrome. *วารสารคลินิก* 276: 11–17.
4. Tan, P. N., Steinbach, M., and Kumar, V. 2005. Introduction to Data Mining. Boston. Pearson.
5. Ian, H. W., Eibe, F., and Mark, A. H. 2011. Data Mining: Practical Machine Learning Tools and Techniques. 3th Edition. Burlington. Morgan Kaufmann, Burlington, MA.
6. พยุง มีสัจ. 2551. ระบบฟัซซีและโครงข่ายประสาทเทียม. เอกสารประกอบการสอน. คณะเทคโนโลยีสารสนเทศ. มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ. กรุงเทพฯ. มหาวิทยาลัยพระจอมเกล้าพระนครเหนือ.
7. ภรัณยา อำมฤครัตน์ และ พยุง มีสัจ. 2553. การเปรียบเทียบประสิทธิภาพการจำแนกกลุ่มข้อมูลของโรคลมร้อนด้วยวิธีการทางเครือข่ายประสาทเทียม. ใน: การประชุมวิชาการระดับชาติด้านคอมพิวเตอร์และเทคโนโลยีสารสนเทศ ครั้งที่ 6. 3–5 มิถุนายน 2553. มักกะสัน กรุงเทพฯ. หน้า 116–121.
8. วิชุดา ไชยศิวามงคล เปรม จันทร์สว่าง จุฬาพรรณ อิ้งจะนิล ประจวบ ชัยมณี วรุฒ ชัยวงษ์ และ วิราวรรณ พุทธ มาตย์. 2556. การจำแนกผู้ป่วยโรคมะเร็งตับโดยการประยุกต์ใช้โครงข่ายประสาทเทียม. *วารสารวิจัย มข.* 18 (4): 585–593.
9. สมเกียรติ โกศลสมบัติ และ รัชฎา คงคะจันทร์. 2554. การคัดเลือกคุณลักษณะที่เหมาะสมในการทำเหมือง ข้อมูลเพื่อสร้างระบบตรวจสอบมะเร็งเต้านม. ใน: การประชุมวิชาการข่ายงานวิศวกรรมอุตสาหการ ประจำปี 2554. 20–21 ตุลาคม 2554. โรงแรมแอมบาสเดอร์ซิตี้ จอมเทียน จังหวัดชลบุรี. หน้า 2017–2021.
10. จิราภรณ์ ถมแก้ว และศรัณย์ อินทโกสุม. 2554. การจำแนกข้อมูลโดยการคัดเลือกคุณลักษณะที่สำคัญ. ใน: การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 23. 23–24 ธันวาคม 2554. คณะวิทยาศาสตร์และศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีราชมงคลอีสาน จังหวัดนครราชสีมา. หน้า 7–12.
11. Machine Learning Group at the University of Waikato. 2013. Data Mining with Weka. Available from URL: http://www.cs.waikato.ac.nz/ml/weka/.17 October 2013.
12. Lu, X., Peng, X., Liu, P., Deng, Y., Feng, B., and Liao, B. 2012. A Novel Feature Selection Method Based on CFS in Cancer Recognition. Proceedings of the IEEE 6th International Conference on Systems Biology (ISB 2012). 18–20 August 2012. Xian, China. p. 226-231.

ได้รับบทความวันที่ 15 พฤศจิกายน 2556
ยอมรับตีพิมพ์วันที่ 21 มกราคม 2557