ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล

Screening of Chronic Kidney Disease Patients System Based on Data Mining Techniques

ณัฐพล แสนคำ[1*]
ทิพวัลย์ แสนคำ[1]
ธนากร ปุรารัมย์[2]

[1]อาจารย์ สาขาวิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
E-mail : nuttapol.sk@gmail.com
E-mail : thipphawand@hotmail.com
[2]นักศึกษาหลักสูตรปริญญาวิทยาศาตรบัณฑิต สาขาวิชาวิทยาการคอมพิวเตอร์
คณะวิทยาศาสตร์ มหาวิทยาลัยราชภัฏบุรีรัมย์
E-mail : tanakorn.pur@bru.ac.th

บทคัดย่อ : งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล เปรียบเทียบประสิทธิภาพของอัลกอริทึมสำหรับทำนายโรคไตเรื้อรัง และประเมินประสิทธิภาพของระบบทำนายโรคไต การทำงานหลักของระบบ ประกอบด้วย การจัดการข้อมูลผู้ป่วย บุคลากรทางการแพทย์ การตรวจร่างกาย การทำนายโรคไต การตรวจสอบผลการทำนายโรค การจัดการตัวแบบ และออกรายงานต่าง ๆ

จากการวัดประสิทธิภาพของการวิจัยครั้งนี้ สรุปได้ว่า เทคนิค Random Forest ที่ใช้ชุดข้อมูลที่มีการเพิ่ม (Oversampling Data) มีประสิทธิภาพค่าความถูกต้อง (Accuracy) สูงที่สุดจากทุกตัวแบบที่ค่า 97.29% ค่าความแม่นตรง (Precision) ที่ 95.76% และค่าวัดประสิทธิภาพโดยรวม (F-Measure) เท่ากับ 97.44% และนำเทคนิคเหมืองข้อมูลนี้มาพัฒนาเป็นตัวแบบในการทำนายโรคไต ผลการประเมินประสิทธิภาพในการทำนายโรคไตของระบบพบว่าสามารถทำนายโรคไตของข้อมูลใหม่ได้ถูกต้อง 95.71% ทั้งนี้ เทคนิคต่าง ๆ และตัวแบบที่ได้พัฒนาขึ้นจะสามารถนำไปต่อยอด เพื่อพัฒนาระบบสนับสนุนทางการแพทย์ที่มีประสิทธิภาพในอนาคต

คำสำคัญ : คัดกรองผู้ป่วย โรคไตเรื้อรัง เหมืองข้อมูล ทำนายโรคไต

ABSTRACT : The purposes of the research were to developed a Screening of Chronic Kidney Disease Patients System Based on Data Mining Techniques, compare the performance of algorithm of predicting disease and evaluate the performance of predicting disease system. The main functions of this system included the management of information of patients and medical personnel, physical examination, prediction of kidney disease, assessment of predictive result, model management and other various reports.

The research findings showed that the Random Forest Technique that utilized the dataset containing Oversampling Data had the highest accuracy value of all the information was 97.29%, precision value was 95.76% and F-Measure value was equal to 97.44% and this (particular) data mining technique can be used to develop a model to predict the occurrence of kidney disease. The result of performance evaluation for the prediction of kidney disease by the system found that (the system) is able to predict the new data as correct as correct as 95.71%. It can be concluded that the techniques and model can be applied in the development of medical support systems effectively in the future.

KEYWORDS : Screening patients, Chronic Kidney Disease, Data mining, Predicting kidney disease.

1. บทนำ

ปัจจุบันมีข้อมูลจำนวนมหาศาลที่ได้จากการดำเนินกิจกรรมทางการแพทย์ ซึ่งได้มีการนำข้อมูลเหล่านี้มาวิเคราะห์เพื่อนำไปใช้ประโยชน์ โดยอาศัยวิธีการทำเหมืองข้อมูล ได้แก่เทคนิคแขนงตัดสินใจ การแบ่งกลุ่ม การจัดกลุ่ม การหาความสัมพันธ์ แล้วนำเอารูปแบบความสัมพันธ์ที่เป็นผลลัพธ์มาสรุปหาความรู้ใหม่ๆ อาทิเช่น การวินิจฉัยโรค การพยากรณ์โรค การรักษาด้วยวิธีที่เหมาะสม เป็นต้น ทำให้สามารถตอบคำถามที่อาจจะเกิดขึ้นในอนาคตได้ ทั้งนี้ บุคคลทั่วไปสามารถค้นคว้า หาความรู้เกี่ยวกับโรคต่างๆ เพื่อนำไปเป็นแนวทางในการรักษาเบื้องต้น รวมถึงการป้องกันโรคต่างๆ ได้

โรคไตเรื้อรัง คือภาวะที่ความสามารถการทำงานของไตลดน้อยลง โดยเป็นผลจากไตถูกทำลาย [1] และมีอัตราการกรองของไตน้อยกว่า 60 ml/min/173m^2 ปัจจุบันพบว่าคนไทยป่วยเป็นโรคไตเรื้อรังประมาณ 8 ล้าน เป็นผู้ป่วยระยะสุดท้าย 2 แสนคน ป่วยเพิ่มปีละกว่า 7,800 ราย [2] ส่วนใหญ่จะไม่แสดงอาการ จนเมื่อหน้าที่ไตเสียเกือบหมด อาการจะรุนแรงมากขึ้น จึงจำเป็นต้องได้รับการวินิจฉัยที่รวดเร็ว และเหมาะสม

จากการศึกษา ณ โรงพยาบาลห้วยราช จังหวัดบุรีรัมย์ พบว่า ในการคัดกรองผู้ป่วยเบื้องต้นทำโดยวิธีการคำนวณหาค่า GFR เพื่อเทียบค่าหาระยะโรคไต โดยพิจารณาตั้งแต่ระยะที่ 3 ขึ้นไป ในการวินิจฉัยโรคไตเรื้อรัง และถ้าไม่มีหลักฐานของภาวะไตผิดปกติระยะที่ 1-2 จะไม่เข้าเกณฑ์การวินิจฉัยโรคไตเรื้อรัง

จากข้อมูลข้างต้น ผู้พัฒนาจึงต้องการพัฒนาระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยเรื้อรังโดยใช้เทคนิคเหมืองข้อมูลในรูปแบบ Web Application เพื่อช่วยคัดกรองผู้ป่วยก่อนถึงการรักษาของแพทย์ผู้เชี่ยวชาญ

ให้สามารถทำนายโรคไตได้อย่างแม่นยำและสามารถตรวจพบความผิดปกติของไต เพื่อให้ผู้ป่วยในระยะที่ 1-2 เข้าสู่กระบวนการรักษาในขั้นตอนต่อไป อีกทั้งเพื่อเพิ่มประสิทธิภาพในการทำงานและลดโอกาสการคัดกรองหรือวินิจฉัยผิดลาดที่อาจจะเกิดขึ้น

2. ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

2.1 โรคไตเรื้อรัง

โรคไตเรื้อรัง เป็นภาวะที่ความสามารถในการทำงานของไตลดลงอย่างต่อเนื่อง ทำให้ไตสูญเสียหน้าที่ในการขจัดของเสีย การควบคุมน้ำและแร่ธาตุต่างๆ ในเลือด โดยค่าอัตราการกรองของไตมีค่าน้อยกว่า 60 ml/min/173m^2 โรคไตเรื้อรังสามารถแบ่งออกได้ 5 ระยะ ดังนี้ [3]

ระยะ	ค่า GFR	คำอธิบาย
1	>=90	ปกติ
2	60-89	ไตผิดปกติ และอัตรากรองไตลดลง
3A	45-59	อัตรากรองไตลดลงปานกลางและมีความเสี่ยงต่อไตล้มเหลวต่ำ
3B	30-44	อัตรากรองไตลดลงปานกลางและมีความเสี่ยงต่อไตล้มเหลวต่ำ
4	15-29	อัตรากรองไตลดลงมากและมีความเสี่ยงต่อไตล้มเหลวสูง
5 5D 5T	<15	ไตล้มเหลว (โรคไตเรื้อรังระยะสุดท้าย)

ภาพที่ 1 ระยะโรคไตเรื้อรัง

2.2 เทคนิคการสุ่มเพิ่มตัวอย่างกลุ่มน้อย (SMOTE)

เทคนิค SMOTE เป็นเทคนิคที่ใช้ในการแก้ปัญหาที่ต้องการจำแนกข้อมูลไม่สมดุล (Imbalance Data) [4] ซึ่งข้อมูลมีจำนวนตัวอย่างแตกต่างกันมากในแต่ละคลาส เมื่อทำการจำแนกประเภท จะทำให้มีการเรียนรู้แต่ข้อมูลกลุ่มที่มาก ผลที่ได้ก็จะจำแนกไปในข้อมูลกลุ่มมาก เทคนิค SMOTE เป็นเทคนิคการเพิ่มจำนวนข้อมูลประเภทที่มีข้อมูลน้อย ให้เพิ่มปริมาณข้อมูลใกล้เคียงกับประเภทที่มีมากที่สุด โดยสุ่มค่าขึ้นมาหนึ่งค่า และหาค่าระยะห่างระหว่างค่าที่เลือกกับทุกๆ ค่า เลือกค่าที่ใกล้เคียงที่สุด เช่น กำหนดไว้ 5 ค่า สุ่มค่าจากที่เลือก 1 ใน 5 หาค่าอยู่ระหว่างค่าที่เลือกตอนแรกและค่าที่สุ่มตอนหลัง เพื่อนำค่าที่ได้มาเพิ่มจำนวนข้อมูลซึ่งมีวิธีการดังนี้ [5]

1. คำนวณหาผลต่างระหว่างชุดข้อมูลที่พิจารณาและชุดข้อมูลที่ใกล้เคียง

2. คำนวณหาค่าข้อมูลใหม่ด้วยการคูณผลลัพธ์จากข้อ 1. ด้วยค่าที่ได้จากการสุ่มตัวเลขที่อยู่ในช่วง 0 ถึง 1 และสามารถเขียนเป็นสมการได้ดังนี้

$$Xnew = xi + (-xi) \times \delta \qquad (1)$$

โดยที่

X_{new} คือ ข้อมูลใหม่

xi คือ ข้อมูลที่สุ่มในตอนแรก

$-x_i$ คือ ข้อมูลที่สุ่มมาอีก เช่น สุ่มมาอีก 5 จุด

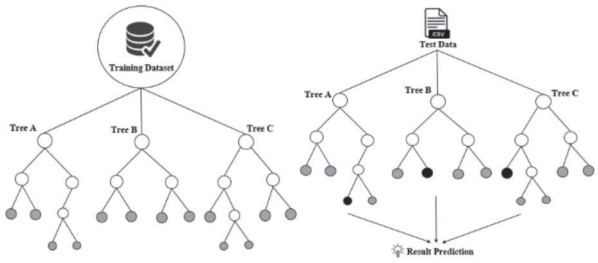
δ คือ ค่าสุ่มตั้งแต่ 0-1

2.3 Decision Tree

Decision Tree เป็นวิธีการทำนายข้อมูลชนิดหนึ่ง ใช้ในการสร้างแบบจำลองการจำแนกข้อมูลหรือการถดถอยของข้อมูลในประเภทโครงสร้างแบบต้นไม้ [6] แต่ละโหนดของต้นไม้ทำการทดสอบด้วยคุณลักษณะเรียนรู้และแสดงผลทดสอบด้วยการแบ่งแต่ละโหนดเป็นชั้น โดยวิธีการจำแนกข้อมูล เป็นการหาเส้นทางจากโหนดรากถึงโหนดใบตามการวัดด้วยคุณลักษณะทดสอบและคุณลักษณะบนโหนดใบคือคำตอบของเทคนิคนี้ [7]

2.4 Random Forest Ensemble

Random Forest Ensemble เป็นเทคนิคที่ใช้ทำตัวแบบ (Model) การจำแนกข้อมูล (Classification) หลายๆ ตัวแบบ มาช่วยในการค้นหาคำตอบซึ่งจะทำให้มีประสิทธิภาพมากยิ่งกว่าการใช้ตัวแบบเดียว ซึ่งเป็นการสุ่มตัวอย่างข้อมูลในชุดข้อมูล Training Data ออกเป็นหลายๆ ชุด [8] ในแต่ละชุดมีการสุ่มแอตทริบิวส์มาใช้สร้างตัวแบบโดยเทคนิคที่ใช้ในการสร้างตัวแบบเป็น Decision Tree เดียวกันทั้งหมด ถึงแม้ว่าจะเป็นเทคนิคเดียวแต่ข้อมูลและแอตทริบิวส์ที่ใช้ในการสร้างตัวแบบต่างกัน ก็ทำให้โมเดลที่สร้างนั้นมีลักษณะที่ต่างกันออกไป ดังภาพ (ก)



(ก) (ข)

ภาพที่ 2 Random Forest การสร้างและการเลือกคำตอบของ model

เมื่อสร้างตัวแบบเสร็จสิ้น จากนั้นนำไปทำนายข้อมูลใหม่ (Testing Data) ทุกตัวแบบ โดยผลลัพธ์สุดท้ายที่ได้นั้นมาจากคำตอบที่มีเหมือนกันมากที่สุด จะถือว่าเป็นคำตอบของข้อมูลใหม่ที่เข้าไปทดสอบ ดังภาพ (ข)

2.5 งานวิจัยที่เกี่ยวข้อง

ผู้วิจัยทำการค้นคว้างานวิจัยด้านเหมืองข้อมูลที่เกี่ยวข้องวิธีการทำนำมาใช้ในการวิจัย เพื่อนำมาเป็นข้อมูลในการศึกษาและอ้างอิง ได้แก่

งานวิจัยของ วีระยุทธ มายุศิริ (2557) วิจัยเรื่องแบบจำลองเพื่อการพยากรณ์การกลับมารักษาซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูล โดยทำการปรับความสมดุลของข้อมูลด้วยวิธี SMOTE และใช้ 10-fold cross validation ในการแบ่งข้อมูลเพื่อประเมินประสิทธิภาพแบบจำลองซึ่งพบว่าช่วยแก้ปัญหาข้อมูลเกิดความไม่สมดุลของคลาสในข้อมูลและมีประสิทธิภาพมากยิ่งขึ้น [4]

Subasi A., Alickovic E., Kevric J. วิจัยเรื่อง Diagnosis of Chronic Kidney Disease by Using Random Forest ทำการศึกษาวิธี

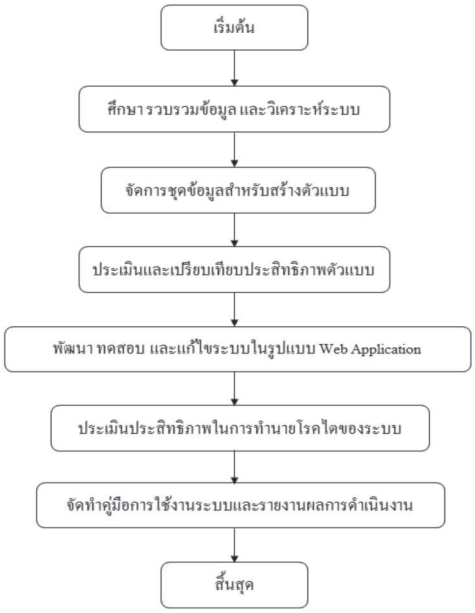
การวินิจฉัยโรค CKD โดยใช้เทคนิคการเรียนรู้ของเครื่อง ซึ่งใช้เทคนิค ANNs, SVM, K-NN, Decision Tree, Random Forest ในการตรวจสอบความผิดปกติและจำแนกข้อมูลโรคไต ผลการศึกษาพบว่า Random Forest สามารถจำแนกโรคได้ใกล้เคียงและสามารถนำมาวินิจฉัยโรคที่คล้ายกันได้ [9]

ภรัณยา ปาลวิสุทธิ์ วิจัยเรื่อง การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต โดยมีการพัฒนาตัวแบบโดยใช้ J48, ID3, LMT, CART และ Random Forest ทำการปรับความสมดุลข้อมูลด้วย SMOTE ซึ่งผลการทดลองพบว่า Random Forest สามารถพยากรณ์ได้เทคนิคอื่นๆ ซึ่งมีค่า accuracy 87.15% [5]

จากการศึกษางานวิจัยพบว่าเทคนิค Random Forest สามารถจำแนกข้อมูลจากชุดข้อมูล Chronic Kidney Disease และชุดข้อมูลอื่นได้อย่างมีประสิทธิภาพ ผู้วิจัยจึงเลือกเทคนิค Random Forest และ Decision Tree เพื่อเปรียบเทียบประสิทธิภาพตัวแบบจากเทคนิคกลุ่ม Tree ที่มีโครงสร้างแตกต่างกันกับชุดข้อมูลนี้

3. วิธีการดำเนินงานวิจัย

ขั้นตอนการดำเนินการวิจัยประกอบด้วย 6 ขั้นตอน ดังนี้



ภาพที่ 3 ขั้นตอนการดำเนินการวิจัยและพัฒนา

3.1 ศึกษา รวบรวมข้อมูลและวิเคราะห์ระบบ

วางแผนขั้นตอนในการทำงาน ศึกษาภาษาที่ใช้ในการพัฒนาระบบ คือ ภาษา PHP ใช้ Scikit-learn Python Library ซึ่งเป็นโมดูลของภาษา Python ที่ช่วยจัดการด้านการทำเหมืองข้อมูล (Data Mining) โดยเลือกใช้ประเภทการจำแนกประเภทข้อมูล (Classification) ในการสร้างระบบทำนายโรค และเครื่องมือจัดการข้อมูล ได้แก่โปรแกรม Rapidminer ใช้ในการเตรียมข้อมูล สร้างตัวแบบและวัดประสิทธิภาพเพื่อเปรียบเทียบหาค่าความถูกต้องของตัวแบบในการทำนายผล และโปรแกรม Weka ใช้จัดการข้อมูลสูญหาย (Missing Data) และทำ Oversampling ด้วยเทคนิค SMOTE

3.2 จัดการชุดข้อมูลสำหรับสร้างตัวแบบ

ผู้วิจัยได้ใช้ชุดข้อมูลจาก UCIMachineLearning Repository ได้แก่ ChronicKidneyDisease [10] ประกอบด้วยข้อมูล 400 รายการ 25 แอตทริบิวต์ โดยมีรายละเอียดดังนี้

ตารางที่ 1 รายละเอียดชุดข้อมูล

ลำดับ	ชื่อคุณลักษณะ	คำอธิบาย
1	Age	อายุ (ปี)
2	Blood Pressure	ความดันโลหิต
3	Specific Gravity	ค่าความถ่วงจำเพาะ
4	Albumin	ค่าโปรตีนอัลบูมิน
5	Sugar	ระดับน้ำตาล
6	Red Blood Cells	เซลล์เม็ดเลือดแดง
7	Pus Cell	เซลล์หนอง
8	Pus Cell clumps	ก้อนเซลล์หนอง
9	Bacteria	แบคทีเรีย
10	Blood Glucose Random	ค่าน้ำตาลในเลือดแบบสุ่ม
11	Blood Urea	ค่ายูเรียในเลือด
12	Serum Creatinine	ค่าซีรั่มครีเอตินีน
13	Sodium	ค่าโซเดียม
14	Potassium	ค่าโพแทสเซียม
15	Hemoglobin	ค่าสารสีของเม็ดเลือดแดง
16	Packed Cell Volume	ร้อยละของเม็ดเลือดแดงต่อปริมาณเลือดทั้งหมด
17	White Blood Cell Count	จำนวนเซลล์เม็ดเลือดขาว
18	Red Blood Cell Count	จำนวนเซลล์เม็ดเลือดแดง
19	Hypertension	โรคความดันโลหิตสูง
20	Diabetes Mellitus	โรคเบาหวาน
21	Coronary Artery Disease	โรคหลอดเลือดหัวใจ
22	Appetite	ความอยากอาหาร
23	Pedal Edema	อาการบวมเท้า
24	Anemia	โรคโลหิตจาง
25	Class	ผลตรวจโรค ckd = เป็นโรคไต notckd = ไม่เป็นโรคไต

จากตารางที่ 1 คุณลักษณะลำดับ 25 เป็นคุณลักษณะที่ใช้จำแนกข้อมูล มีจำนวน 2 กลุ่ม คือ กลุ่ม CKD จำนวน 250 รายการ และกลุ่ม notCKD จำนวน 150 รายการ ซึ่งทำการจัดการข้อมูลดังนี้ แทนที่ข้อมูลที่ว่าง โดยใช้หลักการคือ ข้อมูลประเภท Nominal แทนด้วยค่าฐานนิยม (Mode) และข้อมูลประเภท Numerical แทนด้วยค่าเฉลี่ย (Mean) จากข้อมูลทั้งหมด จัดการ Outlier Data โดยการตัดค่าเหล่านั้นทิ้งไปซึ่งกำหนด 10 รายการข้อมูลและจัดการความไม่สมดุลของข้อมูล ผู้วิจัยจึงนำวิธี SMOTE มาใช้เพื่อปรับสมดุลข้อมูล แสดงรายละเอียดดังนี้

ตารางที่ 2 จำนวนข้อมูลแต่ละชุดข้อมูล

Dataset	CKD class	NotCKD class
Original Data	250	150
Undersampling Data	140	140
Oversampling Data	250	249

3.3 ประเมินและเปรียบเทียบประสิทธิภาพตัวแบบ

ผู้วิจัยทำการทดลองเปรียบเทียบประสิทธิภาพ 3 ชุดข้อมูล ได้แก่ ชุดข้อมูล Original ชุดข้อมูล Undersampling และชุดข้อมูล Oversampling โดยใช้เทคนิค ได้แก่ Random Forest และ Decision Tree รวมทั้งหมด 6 ตัวแบบ ทำการประเมินและเปรียบเทียบประสิทธิภาพตัวแบบด้วยค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าประสิทธิภาพโดยรวม (F-measure) และนำตัวแบบที่มีค่าความถูกต้องสูงสุดไปพัฒนาระบบทำนายโรค

3.4 พัฒนา ทดสอบ และแก้ไขระบบในรูปแบบ Web Application

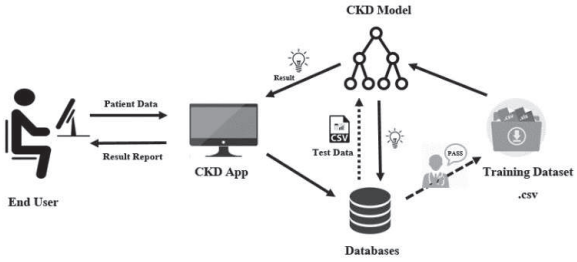
พัฒนาระบบทำนายโรคไตขึ้นโดยใช้เทคนิคและชุดข้อมูลที่ดีที่สุด พัฒนาระบบจัดการข้อมูล ตรวจสอบผลการทำนายโรค จัดการตัวแบบ และออกรายงานต่างๆ

3.5 ประเมินประสิทธิภาพในการทำนายโรคไตของระบบ

ดำเนินการโดยทำการแบ่งข้อมูลด้วยวิธี Spilt Test จำนวน 35% ของข้อมูลทั้งหมดในการศึกษานี้ ประกอบด้วย CKD 94 ข้อมูลและ notCKD46 ข้อมูล รวม 140 ข้อมูล เป็นข้อมูลทดสอบการทำนายโรคว่าระบบให้ผลลัพธ์ถูกต้องหรือไม่ และคำนวณร้อยละ (%)

3.6 จัดทำคู่มือการใช้งานระบบและรายงานผลการดำเนินงาน

ขั้นตอนการทำงานของระบบทำนายโรคไต



ภาพที่ 4 ขั้นตอนการทำงานของระบบทำนายโรคไต

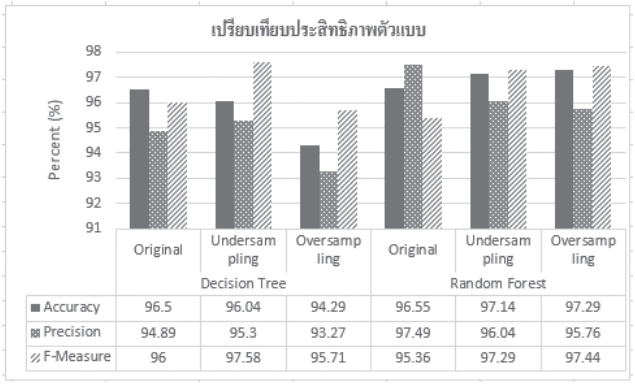
4. ผลการดำเนินงานวิจัย

4.1 ผลการเปรียบเทียบประสิทธิภาพอัลกอริทึมของตัวแบบ 6 ตัวแบบจาก 3 ชุดข้อมูล 2 เทคนิค

ตารางที่ 3 ผลการวัดประสิทธิภาพตัวแบบ

Techniques	Dataset	Accuracy	Precision	F-Measure
Decision Tree	Original	96.50	94.89	96.00
	Undersampling	96.04	95.30	97.58
	Oversampling	94.29	93.27	95.71
Random Forest	Original	96.55	97.49	95.36
	Undersampling	97.14	96.04	97.29
	Oversampling	**97.29**	95.76	97.44

นำผลค่าวัดประสิทธิภาพในแต่ละตัวแบบมาพิจารณาเปรียบเทียบโดยแสดงผลในรูปแบบกราฟ ได้ดังภาพที่ 5



ภาพที่ 5 แผนภูมิเปรียบเทียบประสิทธิภาพแต่ละตัวแบบ

จากตาราง 3 และภาพ 5 การวัดค่าความถูกต้อง (Accuracy) นับว่าเป็นประสิทธิภาพโดยรวมในการทำนายโรคของตัวแบบ โดยผลการวัด

ประสิทธิภาพทั้ง 6 ตัวแบบ ด้วยการแบ่งข้อมูลวัดประสิทธิภาพวิธี 10-fold Cross Validation ปรากฏว่า เทคนิค Decision Tree ชุดข้อมูล Original Data ให้ค่าความถูกต้อง 96.50% ชุดข้อมูล Undersampling ให้ค่าความถูกต้อง 96.04% และชุดข้อมูล Oversampling มีค่าความถูกต้อง 94.29% และเทคนิค Random Forest ชุดข้อมูล Original Data ให้ค่าความถูกต้อง 96.55% ชุดข้อมูล Undersampling ให้ค่าความถูกต้อง 97.14% และชุดข้อมูล Oversampling มีค่าความถูกต้อง 97.29%

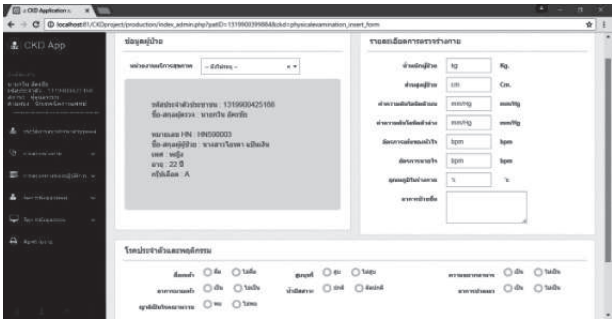
ดังนั้น เมื่อพิจารณาค่าความถูกต้องจากทั้ง 6 ตัวแบบ พบว่า ตัวแบบที่ใช้เทคนิค Random Forest และชุดข้อมูล Oversampling ให้ประสิทธิภาพสูงสุดที่สุดโดยมีค่าความถูกต้องที่ 97.29% ค่าความแม่นตรง (Precision) 95.76% และค่าวัดประสิทธิภาพโดยตรง (F-Measure) 97.44% จึงนำเทคนิคและชุดข้อมูลนี้มาพัฒนาระบบทำนายโรคไต

4.2 ผลการประเมินประสิทธิภาพการทำนายโรคของระบบ

ผลการประเมินประสิทธิภาพในการทำนายโรคไตเรื้อรังของระบบ จากทำการแบ่งข้อมูลด้วยวิธี Spilt Test ซึ่งในงานวิจัยนี้ได้แบ่งข้อมูลผู้ป่วยจำนวน 35% ของข้อมูลทั้งหมดในการศึกษานี้มาทดสอบประสิทธิภาพการทำนายโรคไตของระบบพบว่า ระบบสามารถทำนายข้อมูลใหม่ได้ถูกต้องจำนวน 132 รายการ จากข้อมูล 140 รายการ หรือคิดเป็น 95.71%

4.3 ผลการพัฒนาระบบ

จากการพัฒนาระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูล มีรายละเอียด ดังนี้

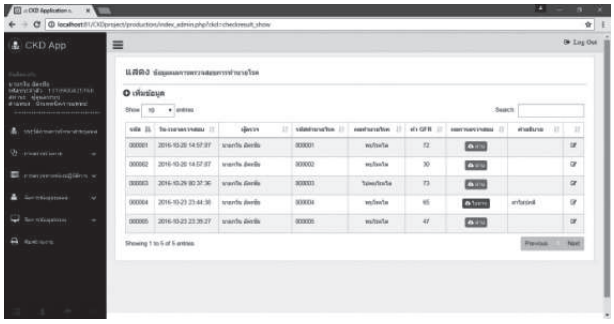


ภาพที่ 6 หน้าจอบันทึกข้อมูลการตรวจร่างกายและผลการตรวจจากห้องปฏิบัติการ

เมื่อผู้ใช้บันทึกข้อมูลเรียบร้อย ทำการทำนายโรคไต ซึ่งระบบทำการประมวลผลส่งผลลัพธ์ทางหน้าจอ ดังภาพที่ 7

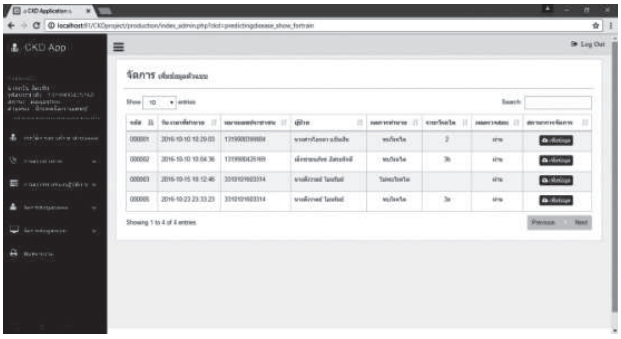
ภาพที่ 7 หน้าจอแสดงผลการทำนายโรคไต

แพทย์ผู้เชี่ยวชาญสามารถตรวจสอบผลการทำนายโรคไต บันทึกความเห็นและผลการตรวจสอบในระบบได้และแสดงผลการตรวจสอบผ่านหน้าจอให้ทุกผู้ใช้งานเห็น ดังภาพที่ 8



ภาพที่ 8 หน้าจอแสดงผลตรวจสอบผลการทำนายโรคไต

ระบบสามารถเพิ่มข้อมูลลงใน Dataset ได้ โดยต้องเป็นข้อมูลที่ผ่านการตรวจสอบจากแพทย์แล้วเท่านั้น



ภาพที่ 9 หน้าจอการจัดการตัวแบบ

5. สรุปผล

ผลศึกษาและพัฒนาระบบสารสนเทศสำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูลสามารถสรุปผลได้ ดังนี้

ระบบสนับสนุนทางการแพทย์สำหรับคัดกรองผู้ป่วยโรคไตเรื้อรังโดยใช้เทคนิคเหมืองข้อมูลที่จัดทำขึ้นสามารถจัดเก็บข้อมูลทางการแพทย์ ทำนายโรคไตของผู้ป่วย และสามารถขยายชุดข้อมูลเพื่อใช้ในการสร้างตัวแบบทำนาย ทั้งนี้เพื่อให้เกิดประสิทธิภาพสูงสุดในการทำนายโรคไต จึงเปรียบเทียบประสิทธิภาพตัวแบบจำนวน 6 ตัวแบบ ได้แก่ ชุดข้อมูล 3 ชุด คือ ชุดข้อมูล Original ข้อมูล Undersampling ข้อมูลที่ Oversampling และเทคนิค 2 เทคนิค คือ Random Forest และ Decision Tree ซึ่งผลการเปรียบเทียบประสิทธิภาพพบว่า ชุดข้อมูล Oversampling กับเทคนิค Random Forest มีประสิทธิภาพสูงสุดด้วยค่าความถูกต้องสูงสุด 97.29% ค่าความแม่นตรง (Precision) 95.76% และค่าวัดประสิทธิภาพโดยตรง (F-Measure) 97.44% จากนั้นนำชุดข้อมูลและเทคนิคดังกล่าวมาพัฒนาเป็นระบบทำนายโรคไต และทดสอบระบบด้วยข้อมูลจำนวน 140 รายการ พบว่า สามารถทำนายโรคไตได้ถูกต้องจำนวน 132 รายการ คิดเป็นร้อยละ 95.71 ถือว่าอยู่ในระดับที่มีความแม่นยำและสามารถนำไปใช้งานได้จริง ทั้งนี้ ระบบที่พัฒนาขึ้นสามารถนำไปเป็นต้นแบบในการพัฒนาระบบทำนายโรคอื่นๆ ได้ต่อไป

บรรณานุกรม

(1) โรงพยาบาลบำรุงราษฎร์, 2553. โรคไตเรื้อรัง. โรงพยาบาลบำรุงราษฎร์. (Online) https://www.bumrungrad.com/th/nephrology-kidney-center-bangkokthailand/ckd-chronic-kidney-disease-1.

(2) Gidanan ganghair, 2559. คนไทยป่วยโรคไตติดอันดับ 3 ของอาเซียน. (Online) http://www.thaihealth.or.th/Content/30963-คนไทยป่วยโรคไตติดอันดับ 3 ของอาเซียน.html.

(3) Gregory D. Krol, 2011.Chronic Kidney Disease Staging and Progression (Online) https://www.asn-online.org/education/training/fellows/HFHS_CKD_V6.pdf.

(4) วีรยุทธ มายุศิริ และคณะ, 2555. การพัฒนาแบบจำลองเพื่อการพยากรณ์การรักซ้ำของผู้ป่วยโรคจิตเภทโดยเทคนิคเหมืองข้อมูล" การประชุมวิชาการ มหาสารคามวิจัย ครั้งที่ 10, 144-153.

(5) ภรันยา ปาลวิสุทธิ์, 2559. การเพิ่มประสิทธิภาพเทคนิคต้นไม้ตัดสินใจบนชุดข้อมูลที่ไม่สมดุลโดยวิธีการสุ่มเพิ่มตัวอย่างกลุ่มน้อยสำหรับข้อมูลการเป็นโรคติดอินเทอร์เน็ต. วารสารเทคโนโลยีสารสนเทศ 12(1): 54-63.

(6) K. R. Anantha Padmanaban and G. Parthiban., 2016. Applying Machine Learning Techniques for Predicting the Risk of Chronic Kidney Disease. Indian Journal of Science and Technology, 9(29).

(7) T.Miranda Lakshmi, A.Martin, R.Mumtaj Begum, V.Prasanna Venkatesan., 2013. An Analysis on Performance of Decision Tree Algorithms using Student's Qualitative Data. I.J.Modern Education and Computer Science, 18-27.

(8) Data Mining Trend บล็อก., 2559. การสร้างโมเดล Ensembleแบบต่างๆ. (Online) http://dataminingtrend.com/2014/data-mining-techniques/ensemble-model/.

(9) Subasi A., Alickovic E., Kevric J., 2017. Diagnosis of Chronic Kidney Disease by Using Random Forest. In: BadnjevicA. (eds) CMBEBIH 2017. IFMBE Proceedings, 62. Springer, Singapore.

(10) UCI Machine Learning Repository, 2015. Chronic_Kidney_Disease Data Set. (Online) https://archive.ics.uci.edu/ml/datasets/Chronic_Kidney_Disease.