

Fetal Status Classification with Heart Rate Tracker and CTG by using Decision Tree

การจำแนกสถานะทารกในครรภ์โดยใช้เครื่องมือติดตามอัตราการเต้นของหัวใจ และการหดตัวของมดลูกด้วยเทคนิคต้นไม้ตัดสินใจ

Received	16 Jul 21
Reviewed	14 Aug 21
Revised	29 Sep 21
Accepted	8 Oct 21

Narumol Chumuang*

นฤมล ชูเมือง*

Department of Digital Media Technology, Faculty of Industrial Technology, Muban Chombueng Rajabhat University, Ratchaburi, Thailand.

สาขาวิชาเทคโนโลยีดิจิทัลมีเดีย คณะเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏหมู่บ้านจอมบึง จังหวัดราชบุรี

Anchana Muankid

อัญชณา เหมือนคิด

UB (Thailand) Co., Ltd.

บริษัท ยูบี (ไทยแลนด์) จำกัด

*Corresponding Author, Tel. +661 916-4155, E-mail: lecho20@hotmail.com

*ผู้พิมพ์ประสานงาน โทรศัพท์. 061-9164155 อีเมล: lecho20@hotmail.com

Abstract

The important to care for the pregnant women in both the prenatal and labor stages. This is the period for which the mortality rate is approximately one third of all births. This article purposes a primary classification of fetal state and create a fetal status classification model to help monitoring the risk that may occur with the fetus. Fetal health checks are necessary in the prenatal period to find the fetus at risk for neurological injury and death of the baby. A condition that can be prevented will be prevented from occurring. In this article, we present an intelligent model for identifying fetal status. The cardiotocography dataset by J. Bernardes which consists of 23 attributes and 2,126 samples and divided into 3 classes as normal, suspect and pathologic. In the experimental for the best. We designed the experimental with 10-fold cross validation and compared the group of decision tree classifier as Decision Stump, Hoeffding Tree, J48, Logistic Model Tree, Random Forest, Random Tree and REP Tree. The results show J48 algorithm, which gives the highest accuracy of 92.79%. The results of the experiment indicate that this algorithm is easy to understand. The operation and the process are not complicated but they can effectively classification fetal status from dataset.

Keywords: Fetal status, Classification, CTG, Heart rate, Contraction of the Uterus, Decision tree

บทคัดย่อ

งานวิจัยชิ้นนี้มีวัตถุประสงค์เพื่อจำแนกสถานะทารกในครรภ์เบื้องต้น และสร้างตัวแบบการจำแนกสถานะทารกในครรภ์ เพื่อช่วยในการเฝ้าระวังความเสี่ยงที่อาจเกิดขึ้นกับทารกในครรภ์ เนื่องจากการตรวจสอบสุขภาพของทารกในครรภ์เป็นสิ่งจำเป็นในระยะก่อนคลอดเพื่อการค้นหาทารกในครรภ์ที่มีความเสี่ยงในการเกิดการบาดเจ็บต่อระบบประสาทใน

ครรภ์ (Neurologic injury) และการเสียชีวิตของทารก ซึ่งหากเป็นภาวะที่สามารถป้องกันได้ จะได้เฝ้าระวังมิให้เกิดขึ้น ในการทำวิจัยครั้งนี้ได้นำเสนอวิธีการจำแนกสถานะของทารกในครรภ์โดยใช้ชุดข้อมูลคาร์ดิโอโตกราฟีของ J.Bernardes จากมหาวิทยาลัยปอร์โต สาธารณรัฐโปรตุเกสซึ่งประกอบด้วยจำนวนคุณลักษณะ 23 คุณลักษณะ และกลุ่มตัวอย่างจำนวนทั้งสิ้น 2,126 ตัวอย่าง เพื่อให้การทดลองได้ค่าความถูกต้องสูงสุด ผู้วิจัยได้ออกแบบการทดลอง 10-fold cross validation โดยนำเทคนิคในกลุ่มของต้นไม้ตัดสินใจอันได้แก่ Decision Stump, Hoeffding Tree, J48, Logistic Model Tree, Random Forest, Random Tree และ REP Tree มาเปรียบเทียบกับ และผลการทดลองปรากฏว่าการจำแนกสถานะจะใช้การสร้างต้นไม้ตัดสินใจด้วยขั้นตอนวิธี J48 เป็นวิธีที่ทำให้ค่าความถูกต้องอยู่ที่ร้อยละ 92.79

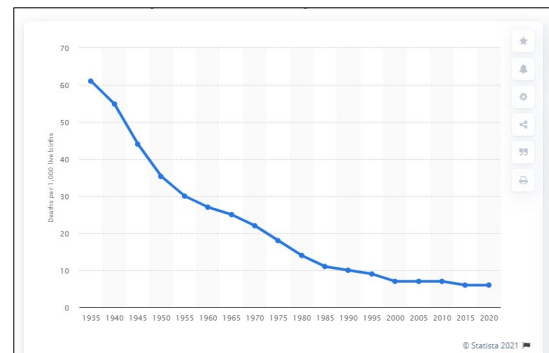
คำสำคัญ: สถานะทารกในครรภ์ การจำแนก การหดตัวของมด เครื่องมือติดตามอัตราการเต้นของหัวใจ ต้นไม้ตัดสินใจ

1. บทนำ

1.1 ความเป็นมาและความสำคัญของปัญหา

ในปัจจุบันวิทยาการทางการแพทย์ได้พัฒนาขึ้นมากแล้ว การดูแลสตรีตั้งครรภ์ทั้งในระยะก่อนคลอดและระยะคลอดจึงเป็นหน้าที่ของแพทย์โดยตรงซึ่งในระยะคลอดเป็นระยะที่มีอัตราเสียชีวิตของทารกถึงประมาณหนึ่งในสามของการคลอดทั้งหมด [1] จากสถิติแสดงอัตราการรอดชีวิตของทารกในสหรัฐอเมริกาแสดงในภาพที่ 1 การตรวจสุขภาพของทารกในครรภ์ในระยะก่อนคลอดเพื่อที่จะทราบสถานะของทารกในครรภ์ว่ามีความปกติ ควรเฝ้าระวังหรือมีความเสี่ยงที่จะเกิดการเสียชีวิต และช่วยลดอัตราการเสียชีวิตของทารกได้อย่างมาก [2], [3] ซึ่งวิธีการตรวจสุขภาพของทารกในครรภ์โดยการใช้เครื่องมือติดตามอัตราการเต้นของหัวใจและการหดตัวของมดลูก (Cardiotocography : CTG) [4] เป็นการตรวจจับเพื่อดูความสัมพันธ์ระหว่างอัตราการเต้นของหัวใจทารกในครรภ์ (fetal heart rate : FHR) อย่างต่อเนื่อง และวัดการหดตัวของมดลูกไปพร้อมกัน เพื่อให้สามารถประเมินหัวใจและกิจกรรมของทารกโดยแพทย์จะนำผลข้อมูลการตรวจที่ได้ใช้ในการวินิจฉัยสถานะของทารก

การจำแนกสถานะของทารกในครรภ์โดยการใช้ข้อมูลที่ได้จากเครื่อง CTG จึงเป็นวิธีที่ช่วยจำแนกสถานะของทารกในครรภ์เบื้องต้นโดยมีสองวิธีในการทำ CTG ของทารกในครรภ์ระหว่างการตั้งครรภ์ คือการตรวจสอบภายนอกและภายในซึ่งมี CTG ภายนอกบริเวณท้องของหญิงตั้งครรภ์สร้างเซ็นเซอร์อัลตราซาวด์ที่จับจังหวะของอัตราการเต้นของหัวใจของทารกในครรภ์และช่วยในการเฝ้าระวังความเสี่ยงจากการบาดเจ็บต่อระบบประสาทในครรภ์ (neurologic injury) และการเสียชีวิตที่อาจเกิดขึ้นกับทารกในครรภ์ได้



ดังนั้นผู้วิจัยจึงมีแนวคิดในการจำแนกสถานะของทารกในครรภ์ เพื่อใช้ในการช่วยเฝ้าระวังความเสี่ยงที่อาจเกิดขึ้นกับทารกในครรภ์ โดยการเลือกใช้อัลกอริทึม J48 ในการสร้างต้นไม้ตัดสินใจ ซึ่งเทคนิคต้นไม้ตัดสินใจ (Decision Tree) เป็นกระบวนการหนึ่งที่ใช้ในการจำแนกประเภทข้อมูล มีลักษณะเป็นโครงสร้างต้นไม้หัวกลับที่มีรากอยู่ด้านบนและใบอยู่ด้านล่างสุด [5] โดยที่ภายในต้นไม้จะประกอบไปด้วยโหนด (Node) ซึ่งแต่ละโหนดนั้นจะแสดงถึงการตัดสินใจบนข้อมูลของคุณสมบัติต่าง ๆ กิ่งของต้นไม้แสดงถึงค่าหรือผลลัพธ์ที่ได้จากการทดสอบ และใบซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจจะแสดงถึงกลุ่มของข้อมูล [6]

ในการวิจัยครั้งนี้ได้นำชุดข้อมูล “Cardiotocography” ซึ่งเป็นข้อมูลเกี่ยวกับค่าที่วัดได้จากเครื่อง CTG มีจำนวนคุณลักษณะ 23 คุณลักษณะ และมีกลุ่มตัวอย่างจำนวน 680 ตัวอย่าง มาใช้ในการจำแนกสถานะของทารกในครรภ์ว่ามีความปกติ ควรเฝ้าระวัง หรือ

มีความเสี่ยงที่จะเกิดการเสียชีวิต บทความวิจัยนี้ประกอบไปด้วยหัวข้อต่อไป ดังนี้ หัวข้อที่ 2 กล่าวถึง วัตถุประสงค์ของการวิจัย หัวข้อที่ 3 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง วิธีการดำเนินการวิจัย หัวข้อที่ 4 ชุดข้อมูล (Dataset) หัวข้อที่ 5 การทดลองและผลลัพธ์ และ หัวข้อที่ 6 เป็นบทสรุปผลการวิจัย

1.2 วัตถุประสงค์ของการวิจัย

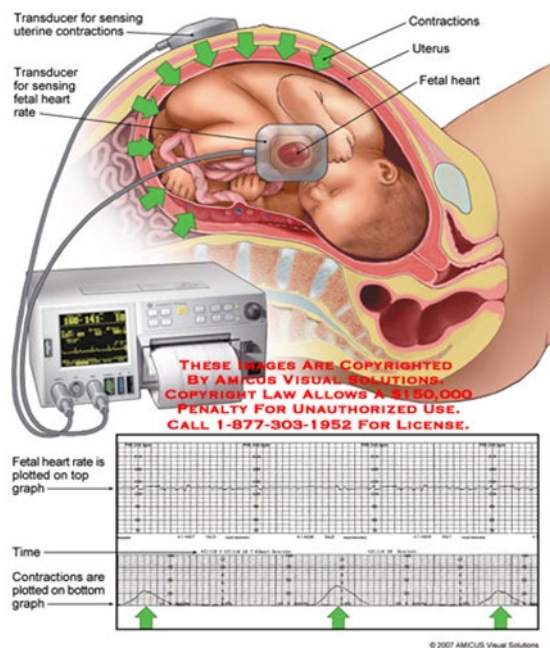
1.2.1 คาร์ดิโอโตกราฟี Cardiotocography (CTG) เป็นการตรวจสอบสุขภาพของทารกในครรภ์โดยใช้เครื่องมืออิเล็กทรอนิกส์ [7] ดังแสดงในรูปที่ 2 ตรวจสอบการเต้นของหัวใจทารกในครรภ์อย่างต่อเนื่อง และ วัดการหดตัวของมดลูกไปพร้อมกัน เพื่อความสัมพันธ์ระหว่าง FHR (Fetal heart rate) [8] กับ การหดตัวของมดลูกแสดงผลเป็นลักษณะกราฟในกระดาษ หรือ ผ่านจอภาพอิเล็กทรอนิกส์ โดยสามารถแบ่งตามลักษณะของเครื่องได้ 2 ชนิด คือ

1) External (indirect) monitoring [9] เป็นการตรวจวัดการเต้นของหัวใจทารกในครรภ์ และการหดตัวของมดลูกผ่านทางหัวตรวจที่วางอยู่บนหน้าท้องแล้วแสดงออกเป็นเส้นกราฟ มีข้อดีคือ non-invasive ใช้น้ำคร่ำไม่จำเป็นต้องแตก ไม่เพิ่มความเสี่ยงต่อการติดเชื้อ วิธีนี้ใช้หัวตรวจแบบ Doppler เพื่อตรวจวัด FHR ดังนั้นการเคลื่อนไหวอื่น ๆ ของมารดาและทารก จึงมีผลกระทบต่อค่าความแปรปรวนของอัตราการเต้นของหัวใจทารก (fetal heart rate variability) นอกจากนั้นการหดตัวของมดลูก(uterine contractility) เป็นการวัดความตึงตัวของผนังมดลูกผ่านทางหน้าท้อง ไม่ได้วัดที่ความดันภายในโพรงมดลูกอย่างแท้จริง ทำให้มีความคลาดเคลื่อนได้สูง รวมทั้งสัญญาณอาจขาดหายได้เมื่อมีการขยับตัวของมารดาและทารก การตรวจวัดอัตราการเต้นของหัวใจของมารดาในช่วง 9 นาทีของการตรวจติดตามและตรวจสอบ FHR ภายนอก (External monitoring) ดังแสดงในรูปที่ 3 ที่ระดับ 1 ชั่วโมง/นาที 2 ชั่วโมง/นาที และ 3 ชั่วโมง/นาที [10]

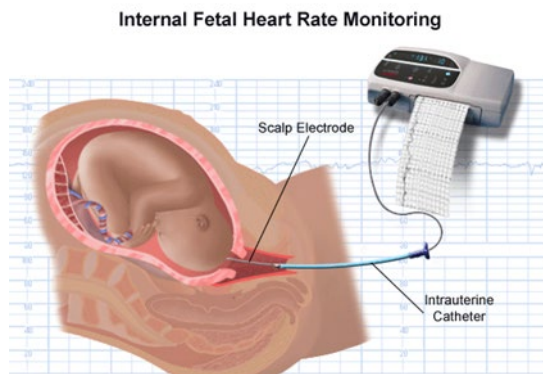


รูปที่ 2 เครื่องบันทึกการบีบตัวของมดลูก และอัตราการเต้นของหัวใจทารกในครรภ์ (Fetal Monitor)

2) Internal (direct) monitoring [11] เป็นการใช้อุปกรณ์สอดผิวหนังทารกเพื่อวัด FHR จากคลื่นไฟฟ้าหัวใจ (Electronic cardiography หรือ ECG) และวัดการหดตัวของมดลูกดังภาพที่ 4 โดยสอดเครื่องมือผ่านเข้าไปในโพรงมดลูก เพื่อวัดการเปลี่ยนแปลง ความดันในโพรงมดลูกโดยตรง วิธีนี้มีข้อดี คือ สามารถวัดความแปรปรวนของอัตราการเต้นของ หัวใจทารก (fetal heart rate variability) และความดัน ภายในโพรงมดลูก (intrauterine pressure) ได้อย่างแม่นยำรวมทั้งสัญญาณจะไม่ขาดหายไป เมื่อมีการขยับตัวของมารดาแต่มีข้อเสียคือเพิ่มความเสี่ยงต่อการติดเชื้อในน้ำคร่ำ และไม่เหมาะสมในรายที่มารดาเป็นพาหะของ group B streptococcus มารดาติดเชื้อ HIV เป็นต้น



รูปที่ 3 ลักษณะตรวจติดตามและตรวจสอบ FHR ภายนอก
(External Fetal Monitor)



รูปที่ 4 ลักษณะตรวจติดตามและตรวจสอบ FHR ภายใน
(Internal Fetal Monitor)

1.2.2 ต้นไม้ตัดสินใจ (Decision Tree) ต้นไม้ตัดสินใจ [12] เป็นเทคนิคที่ให้ผลลัพธ์ในลักษณะของโครงสร้างต้นไม้ ซึ่งเมื่อมีข้อมูลที่ต้องการจัดกลุ่มก็จะนำคุณลักษณะต่างๆ ของข้อมูลนั้นไปเทียบกับเส้นทางในต้นไม้จนกระทั่งคลาสนั้นไปเทียบกับเส้นทางในต้นไม้เหมือนกัน ภายในต้นไม้จะประกอบไปด้วยโหนด (node) ซึ่งแต่ละโหนดจะมีคุณลักษณะเป็นตัวทดสอบกิ่งของต้นไม้ (branch) แสดงถึงค่าที่เป็นไปได้ของคุณลักษณะที่ถูกเลือก

ทดสอบ และใบ (leaf) ซึ่งเป็นสิ่งที่อยู่ล่างสุดของต้นไม้ตัดสินใจแสดงถึงกลุ่มของข้อมูล (class) ก็คือผลลัพธ์ที่ได้จากการทำนาย โหนดที่อยู่บนสุดของต้นไม้เรียกว่าโหนดราก (root node) ต้นไม้ตัดสินใจมีความสามารถในการจัดกลุ่มของแต่ละคุณลักษณะหรือปัจจัย [14]

โดยปกติมักประกอบด้วยกฎในรูปแบบ “ถ้าเงื่อนไขแล้วผลลัพธ์” เช่น

“If Income= High and Married = No THEN Risk = Poor”

“If Income = High and Married = Yes THEN Risk = Good”

Gini Index ค่าที่บ่งบอกว่าคุณลักษณะหรือปัจจัยใดควร นำ มาใช้ของคุณลักษณะในการแบ่งกลุ่มของอัลกอริทึม J48 และ CART

การสร้างต้นไม้ตัดสินใจ หลักการพื้นฐานของการสร้างต้นไม้ตัดสินใจเป็นการสร้างในลักษณะจากบนลงล่าง (Top-Down) คือเริ่มจากการสร้างรากของต้นไม้ก่อนแล้วจึงแตกกิ่งไปจนถึงใบ โดยแสดงขั้นตอนการสร้างต้นไม้ตัดสินใจได้ดังนี้ [13]

1) ต้นไม้เริ่มต้น โดยมีโหนดเพียงโหนดเดียวแสดงถึงชุดข้อมูลฝึก

2) ถ้าข้อมูลทั้งหมดอยู่ในกลุ่มเดียวกันแล้ว ให้โหนดนั้นเป็นใบและตั้งชื่อแยกตามกลุ่มของข้อมูลนั้น

3) ถ้าในโหนดมีข้อมูลหลายกลุ่มปะปนอยู่จะต้องวัดค่าเกณฑ์ (Gain) ของแต่ละแอททริบิวต์เพื่อที่จะใช้เป็นเกณฑ์ในการคัดเลือกแอททริบิวต์ที่มีความสามารถในการแบ่งแยกข้อมูลออกเป็นกลุ่มต่าง ๆ ได้ดีที่สุด โดยแอททริบิวต์ที่มีค่าเกณฑ์มากที่สุดจะถูกเลือกให้เป็นตัวทดสอบหรือแอททริบิวต์ที่ใช้ในการตัดสินใจ โดยแสดงในรูปของโหนดบนต้นไม้

4) กิ่งของต้นไม้ถูกสร้างขึ้นจากค่าต่าง ๆ ที่เป็นไปได้ของโหนดทดสอบ และข้อมูลจะถูกแบ่งออกตามกิ่งต่าง ๆ ที่สร้างขึ้น

5) ทำการวนซ้ำเพื่อหาแอททริบิวต์ที่มีค่าเกณฑ์มากที่สุด สำหรับข้อมูลที่ถูกรับแยกออกมาในแต่ละกิ่งเพื่อนำแอททริบิวต์นั้นมาสร้างเป็นโหนดตัดสินใจต่อไป โดยที่แอททริบิวต์ที่ถูกเลือกมาเป็นโหนดแล้วจะไม่ถูกเลือกมาอีก สำหรับโหนดในระดับต่อไป

6) ทำการวนซ้ำเพื่อแบ่งข้อมูลและแตกกิ่งของต้นไม้ไปเรื่อย ๆ โดยการวนซ้ำจะสิ้นสุดก็ต่อเมื่อเงื่อนไขข้อใดข้อหนึ่งต่อไปนี้จริง

การคำนวณค่าเกณฑ์ของข้อมูลของต้นไม้ตัดสินใจเป็นโครงสร้างที่ใช้สำหรับแสดงกฎที่ได้จากเทคนิคการจำแนกประเภทข้อมูล [14] โดยต้นไม้ตัดสินใจจะมีลักษณะคล้ายโครงสร้างต้นไม้ที่แต่ละโหนดแสดงคุณลักษณะ (Attribute) ในการสร้างต้นไม้ตัดสินใจ ปัญหาสำคัญที่ต้องพิจารณาคือ ควรจะตัดสินใจเลือกแอตทริบิวต์ใดมาทำหน้าที่เป็น [15] โหนดรากในแต่ละขั้นตอนของการสร้างต้นไม้และต้นไม้ย่อย (Subtree) ของต้นไม้ตัดสินใจ เกณฑ์ที่ใช้ช่วยประกอบการเลือกแอตทริบิวต์ คือการคำนวณค่ามาตรฐานเกณฑ์ (Gain Criterion) [16] ซึ่งเป็นค่าที่บ่งบอกว่าแอตทริบิวต์นั้นสามารถจำแนกประเภทของข้อมูลได้เพียงใด โดยทดสอบเลือกค่าของแต่ละแอตทริบิวต์ที่เป็นไปได้จากชุดข้อมูลมาทำหน้าที่เป็นโหนดราก หากแอตทริบิวต์ใดให้ค่าเกณฑ์สูงสุด แสดงว่า แอตทริบิวต์นั้นสามารถจำแนกกลุ่มของข้อมูลได้ดีที่สุด การใช้ค่า เกณฑ์ของข้อมูลจะช่วยลดจำนวนครั้งของการทดสอบในการแยกแยะข้อมูลอีกทั้งยังรับประกันว่า ต้นไม้ตัดสินใจที่ได้ไม่มีความซับซ้อนมากเกินไป ซึ่งค่าเกณฑ์ของข้อมูลนั้น สามารถคำนวณได้จากสมการดังสมการที่ (1)

$$I(S_1, S_2, \dots, S_n) = \sum_{i=1}^n \frac{S_i}{S} \log_2 \frac{S_i}{S} \quad (1)$$

เมื่อ S เป็นเซตของข้อมูลซึ่งประกอบด้วยข้อมูล S เรคอร์ด

n เป็นจำนวนกลุ่มทั้งหมดที่ต่างกันของข้อมูลชุดนั้น

C_i แทนกลุ่มในลำดับที่ i โดย ที่ i มีค่าระหว่าง 1 ถึง n

S_i แทนจำนวนข้อมูลที่เป็นสมาชิกของ S และอยู่ใน กลุ่ม C_i

S_{ij} แทนจำนวนข้อมูลที่เป็นสมาชิกของ S ในกลุ่ม C_i จากการแบ่งข้อมูลด้วยค่าที่เป็นไปได้ของแอตทริบิวต์ A_j ค่าระหว่าง 1 ถึง v

ค่าเอ็นโทรปีของแอตทริบิวต์ A ซึ่งมีค่าของแอตทริบิวต์เป็น $(a_1, a_2, a_3, \dots, a_v)$ หาได้ในสมการที่ (2)

$$E(A) = \sum_{j=1}^v \frac{S_{1j} + \dots + S_{nj}}{S} I(S_{1j}, \dots, S_{nj}) \quad (2)$$

ดังนั้นจะสามารถพิจารณาค่ามาตรฐานเกณฑ์ได้ดังสมการที่ (3)

เมื่อ (p, n) = เซตของ Training Set

(A) = แทน แอตทริบิวต์ที่ถูกเลือกให้เป็นตัวจำแนกข้อมูล

1.2.3 อัลกอริทึมแบบ J48 [17] เป็นอัลกอริทึมที่ใช้ในการสร้างการตัดสินใจแบบโครงสร้างต้นไม้ที่ใช้หลักการของการใช้ทฤษฎีข่าวสาร (Information Theory) และค่าที่วัดได้จะนำมาใช้ตัดสินใจว่าจะใช้ตัวแปรใดใช้ในการทำนาย หรือแบ่งประเภทของข้อมูล โดยที่ชุดตัวอย่าง (Sample) คือชุดของข้อมูลที่ใช้ในการเรียนรู้ (Training Sample) ตัวแปรเป้าหมาย (Target Attribute) คือตัวแปรที่นำไปใช้ในการทำนายผลในโครงสร้างต้นไม้และแอตทริบิวต์ คือตัวแปรอื่น ๆ ที่ใช้ในการสร้างโหนดในต้นไม้และไม่ใช่ตัวแปรเป้าหมาย (Target Attribute) ซึ่งมีลักษณะของอัลกอริทึมดังรูปที่ 5

```

ID3(Examples, Target_Attribute, Attributes)
- Create a root node for the tree
- If all examples are positive, Return the single-node tree Root label = +.
- If all examples are negative, Return the single-node tree, with label = -.
- If number of predicting attributes is empty, then Return the single node tree Root, with label = most common value of the target attribute in the examples.
- Otherwise Begin
- A = The Attribute that best classifies examples.
- Decision Tree attribute for Root = A.
- For each possible value  $v_i$  of A,
- Add a new tree branch below Root, corresponding to the test  $A = v_i$ 
- Let Examples  $(v_i)_j$  be the subset of examples that have the value  $v_i$  for A
- If Examples  $(v_i)_j$  is empty
- Then below this new branch add a leaf node with label = most common target value in the examples
- Else below this new branch add the subtree ID3 (Examples  $(v_i)_j$ , Target_Attribute, Attributes - {A})
- End
- Return Root
  
```

รูปที่ 5 แสดงกระบวนการทำงานของอัลกอริทึมแบบ ID3

1.2.4 อัลกอริทึมแบบ C4.5 (J48) [18] เป็นอัลกอริทึมที่ใช้ในการสร้างกฎจากต้นไม้ตัดสินใจ โดยมีพื้นฐานเดียวกับอัลกอริทึมแบบ ID3 ถูกออกแบบโดยควินแลน (Quinlan) ในปี 2535 ซึ่งได้พัฒนาเพิ่มเติมจากแบบ ID3 [19] ดังนี้

1) สามารถหลีกเลี่ยงการสร้างโครงสร้างต้นไม้ที่ใหญ่เกินไป เนื่องจากมีข้อมูลจำนวนมาก อย่างไรก็ตามขึ้นอยู่กับ การกำหนดความลึกของการตัดสินใจแบบต้นไม้

2) มีความผิดพลาดน้อยลง เพราะมีการตัดทอนส่วนที่ผิดพลาดออกไป (Pruning node)

3) มีการสร้างกฎหลังการตัดทอนข้อมูลส่วนที่ผิดพลาดออก

4) สามารถใช้กับข้อมูลได้อย่างต่อเนื่อง (Continuous Attributes) รวมถึงข้อมูลตัวเลขได้เช่น ค่าอุณหภูมิ จำนวนเงิน เป็นต้น

5) การเลือกแอททริบิวต์ที่วัดการเลือกให้เหมาะสม

6) สามารถใช้กับชุดข้อมูลเรียนรู้ที่มีค่าความผิดพลาด (Missing Attribute)

7) สามารถใช้สำหรับแอททริบิวต์กับคอส (Costs) ที่แตกต่างกันได้ได้รับการยอมรับประสิทธิภาพของกระบวนการคำนวณที่ถูกปรับปรุง

1.3 ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้อง ในหัวข้อนี้จะรวบรวมผลงานวิจัยโดดเด่นที่เกี่ยวข้องกับงานวิจัยนี้ ซึ่งมีดังต่อไปนี้

1.3.1 Helen Cooper และคณะผู้จัดทำ [11] งานวิจัยการจดจำภาษามือโดยใช้หน่วยย่อยทางภาษาศาสตร์ ถูกนำเสนอเพื่อพิจารณาและเรียนรู้จากข้อมูลในรูปแบบของ 2D และ 3D เป็นการพัฒนาจากโมเดลโดยใช้ระบบKinect TM demonstration แบบ Real time ซึ่งจะประกอบกับสัญญาณหลายแบบเพื่อจับการเปลี่ยนแปลงเคลื่อนไหวของ handshape

1.3.2 Alina Kuznetsova และคณะ [12] นำเสนอการรู้จำท่ามือที่มีความแม่นยำสูงโดยใช้กล้องเดียวแบบ RGB ในงานวิจัยนี้จะนำเสนอความแม่นยำและวิธีการจดจำจากข้อมูลเชิงลึกเพื่อให้เกิดคุณสมบัติที่สามารถแปลได้แบบ multi-layered random forest (MLRF) เพื่อให้ได้ความแม่นยำโดยจะใช้กับภาษามือ ASL (American Sign Language)

1.3.3 Thad Stamer และคณะ [13] ได้ออกแบบระบบการจำแนก Sign Language ในภาษามือของ ASL (American Sign Language) โดยใช้โมเดลของ Markov โดยใช้ระบบกล้องเดียวในการติดตามมือของผู้ใช้งานที่ปราศจากการตกแต่ง โดยในงานวิจัยมีการระบุได้ว่า

ทดสอบความแม่นยำของการใช้งานประสบความสำเร็จ 92% และระบบที่สองคือให้ผู้ใช้งานสวมหมวกได้บรรลุความถูกต้องแม่นยำถึง 98%

1.3.4 Ruslan Kurdyumov และคณะ [14] นำเสนอการแยกแยะตัวอักษรที่ใช้จากภาพเว็บแคมโดยปรับขนาดเป็น 20x20 พิกเซล และกำหนดคุณสมบัติเป็น binary pixels ซึ่งสามารถจำแนกประเภทของสัญญาณได้ถึง 92-93% โดยใช้ linear or Gaussian kernel SVM ได้ดีกว่าการทำ k-nearest neighbor ประมาณ 10% ซึ่งจากงานวิจัยสามารถจำแนกประเภทของสัญญาณได้อย่างมีประสิทธิภาพด้วย SVM 81-82%

2.วัสดุ อุปกรณ์ และวิธีการวิจัย

2.1 ชุดข้อมูล

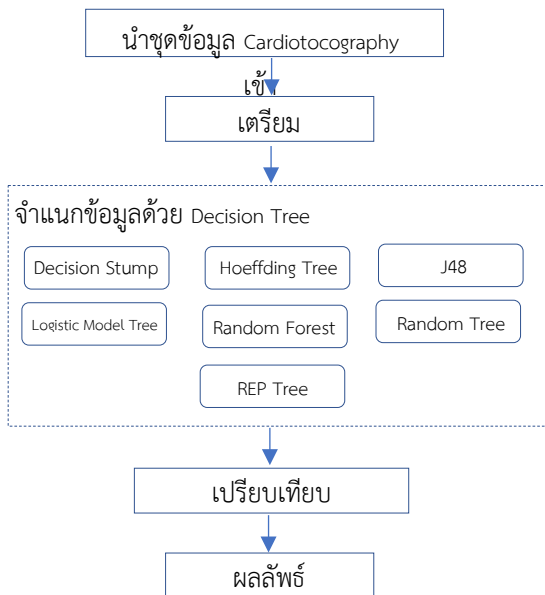
งานวิจัยนี้ใช้ชุดข้อมูลที่มีคุณสมบัติเหมือนกันมาจัดเป็นชุดให้ถูกต้องตามลักษณะโครงสร้างข้อมูลมาตรฐานที่สามารถนำมาประมวลผลทางดิจิทัลได้ โดยชุดข้อมูล cardiotocography ประกอบด้วยการวัดอัตราการเต้นของหัวใจของทารกในครรภ์ (Fetal Heart Rate: FHR) และลักษณะการหดตัวของมดลูก (Uterine Contraction: UC) เกี่ยวกับคาร์ดิโอโทแกรมจำแนกตามสถิติแพทย์ผู้เชี่ยวชาญซึ่งประกอบไปด้วยจำนวน 2,126 ตัวอย่างของทารกในครรภ์ (CTGs) ได้รับการประมวลผลโดยอัตโนมัติและวัดคุณสมบัติการวินิจฉัยตามลำดับ CTG ยังถูกจัดประเภทโดยสถิติแพทย์ผู้เชี่ยวชาญในการจำแนกที่เป็นเอกลักษณ์ที่กำหนดโดยแบ่งประเภทของสถานะเป็น 3 ประเภท (Class) คือสถานะปกติ (Normal), สถานะต้องสงสัย (Suspect) และสถานะเสี่ยง (Pathologic)

2.2 ขั้นตอน วิธีการวิจัย

เพื่อให้การอธิบายวิธีการดำเนินการวิจัยเพื่อให้เข้าใจง่าย ผู้วิจัยได้ออกแบบวิธีการดำเนินการวิจัยดังแสดงในรูปที่ 6 จากรูปที่ 6 อธิบายขั้นตอนในการดำเนินการการจำแนกสถานะทารกในครรภ์โดยการใช้เครื่องมือติดตามอัตราการเต้นของหัวใจ และการหดตัวของมดลูกด้วยเทคนิคต้นไม้ตัดสินใจตามลำดับดังต่อไปนี้

2.2.1 นำชุดข้อมูล Cardiotocography เข้ารวบรวมชุดข้อมูล Cardiotocography สำหรับการจำแนกสถานะทารกในครรภ์โดยการใช้เครื่องมือติดตามอัตราการเต้นของหัวใจ และการหดตัวของมดลูกโดยการทำควม

สะอาดข้อมูล (Clean data) โดยการทำให้ข้อมูลถูกต้อง เช่น การจัดการกับข้อมูลประเภท None หรือ NULL หรือ จัดการกับ Outlier data ซึ่งชุดข้อมูล Cardiotocography ที่ใช้ทดลองในงานวิจัยนี้ประกอบด้วย 23 คุณลักษณะ (features) ตามตารางที่ 1



รูปที่ 6 แสดงภาพรวม (Overview) ในการดำเนินการวิจัย

ตารางที่ 1 แสดงคุณลักษณะข้อมูลและความหมาย

ที่	คุณลักษณะ	ความหมาย
1	LB	FHR baseline (beats per minute)
2	AC d	# of accelerations per second
3	FM	#of fetal movements per second
4	UC	# of uterine contractions per second
5	DL	# of light decelerations per second
6	DS	# of severe decelerations per second
7	DP	# of prolonged decelerations per second
8	ASTV	percentage of time with abnormal short term variability
9	MSTV	mean value of short term variability
10	ALTV	percentage of time with abnormal long term variability
11	MLTV	mean value of long term variability

ที่	คุณลักษณะ	ความหมาย
12	Width	width of FHR histogram
13	Min	minimum of FHR histogram
14	Max	Maximum of FHR histogram
15	Nmax	# of histogram peaks
16	Nzeros	- # of histogram zeros
17	Mode	histogram mode
18	Mean	histogram mean
19	Median	histogram median
20	Variance	histogram variance
21	Tendency	histogram tendency
22	CLASS	FHR pattern class code (1 to 3)
23	NSP	fetal state class code (N=normal; S=suspect; P=pathologic)

ซึ่งชุดข้อมูลจำนวน 2,126 ตัวอย่างและจำแนกประเภท (Class) ทั้งหมด 3 ประเภท คือ สถานะปกติ (Normal), สถานะต้องสงสัย (Suspect) และ สถานะผิดปกติ (Pathologic) โดยแสดงตัวอย่างชุดข้อมูลที่ใช้ในการจำแนกตามตารางที่ 2

ตารางที่ 2 แสดงตัวอย่างชุดข้อมูล Cardiotocography

No	LB	AC	MSTV	...	NSP
1	132	4	2.1	...	N
2	133	2	2.1	...	N
3	134	2	2.4	...	N
4	132	4	2.4	...	N
5	131	4	1.4	...	N
6	131	6	1.5	...	S
7	130	7	2.3	...	S
8	130	4	2.3	...	N
9	130	4	2.1	...	N
10	130	2	1.9	...	N
...	...	...	...	...	...
2126	140	0	4.1	...	P

2.2.2 การเตรียมข้อมูล

ขั้นตอนนี้เป็นขั้นตอนหลังจากทำความสะอาดชุดข้อมูล (Clean data) โดยเป็นการเตรียมข้อมูลให้พร้อมนำไปใช้กับอัลกอริทึมต้นไม้ตัดสินใจ เช่น การเลือกคุณลักษณะ (Feature selection) หรือ การแปลงข้อความ

เป็นตัวเลข ซึ่งอัลกอริทึมต้นไม้ตัดสินใจต้องการชุดข้อมูลนำเข้า (Input) ที่เป็นตัวเลขเท่านั้น ซึ่งตารางที่ 3 แสดงตัวอย่างชุดข้อมูล Cardiotocography ที่ผ่านการเตรียมข้อมูลแล้ว

ตารางที่ 3 แสดงตัวอย่างชุดข้อมูล Cardiotocography ผ่านการเตรียมข้อมูลแล้ว

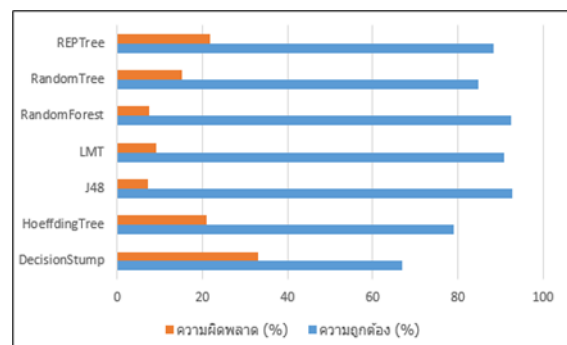
No	LB	AC	MSTV	...	Class
1	132	4	2.1	...	Normal
2	133	2	2.1	...	Normal
3	134	2	2.4	...	Normal
4	132	4	2.4	...	Normal
5	131	4	1.4	...	Normal
6	131	6	1.5	...	Suspect
7	130	7	2.3	...	Suspect
8	130	4	2.3	...	Normal
9	130	4	2.1	...	Normal
10	130	2	1.9	...	Normal
...	...	...	...	...	...
2126	140	0	4.1	...	Pathologic

2.2.3 การสร้างโมเดลจำแนกสถานะทารก

สำหรับขั้นตอนนี้จะกล่าวถึงการออกแบบการทดลองคณะผู้วิจัยได้ใช้ชุดข้อมูล “Cardiotocograph” จากฐานชุดข้อมูลที่มีความนิยมในการทดลองทางด้านการสร้างการเรียนรู้ของเครื่องโมเดล (Machine learning) เช่น UCI [1] ซึ่ง ได้ยกกลุ่มตัวอย่างจำนวน 2,126 ตัวอย่าง จาก คุณลักษณะ 23 คุณลักษณะ โดยการทดลองกับ อัลกอริทึมในกลุ่มต้นไม้ตัดสินใจ ประกอบด้วย โดยนำเทคนิคในกลุ่มของต้นไม้ตัดสินใจอันได้แก่ DecisionStump, Hoeffding Tree, J48, LMT (Logistic Model Tree), RandomForest, RandomTree และ REPTree มาทำการเปรียบเทียบกัน และผลการทดลองปรากฏว่าการจำแนกสถานะจะใช้การสร้างต้นไม้ตัดสินใจด้วยขั้นตอนวิธี J48 เป็นวิธีที่ให้ค่าความถูกต้องสูงที่สุดที่ร้อยละ 92.79 ตามตาราง 4 และสามารถแสดงแผนภูมิเปรียบเทียบประสิทธิภาพความถูกต้องดังรูปที่ 7 โดยใช้สมการคำนวณค่าความถูกต้อง (Accuracy rate) ดังนี้

ตารางที่ 4 ผลการทดลองของทั้ง 7 อัลกอริทึม

อัลกอริทึม	ความถูกต้อง (%)	ความผิดพลาด (%)
DecisionStump	66.91	33.09
HoeffdingTree	79.12	20.88
J48	92.79	7.21
LMT	90.74	9.26
RandomForest	92.50	7.50
RandomTree	84.85	15.15
REPTree	88.25	21.75

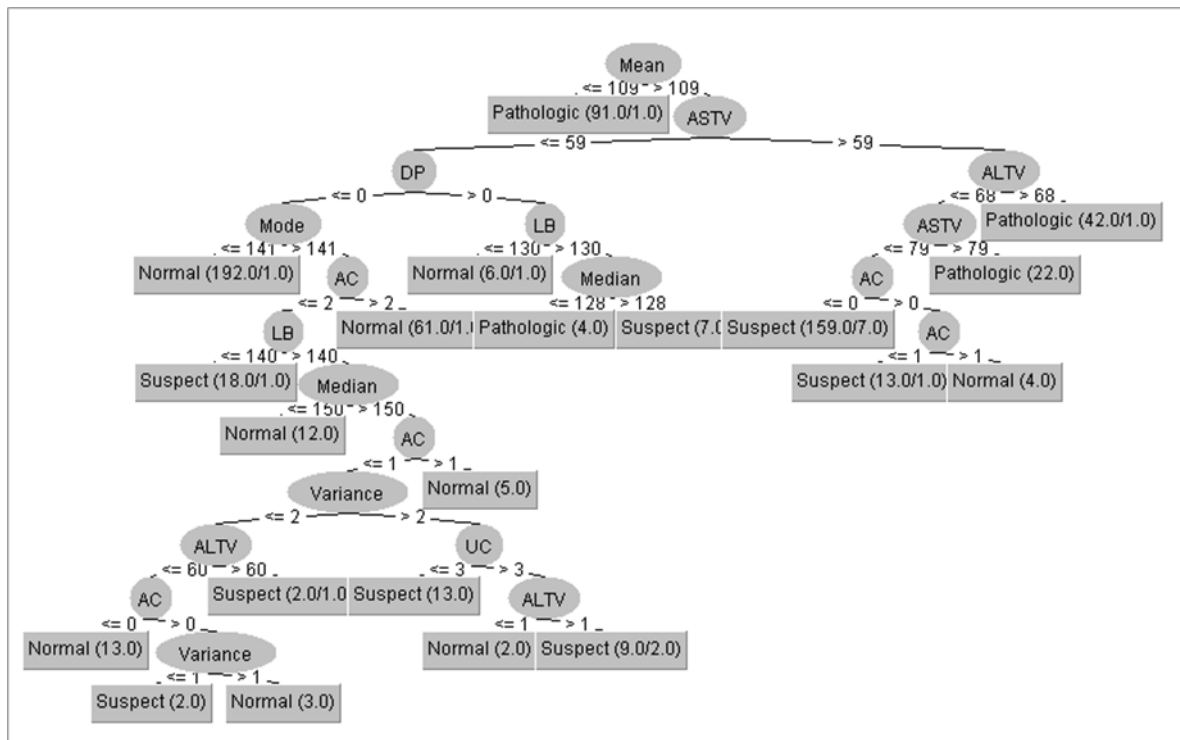


รูปที่ 7 แผนภูมิเปรียบเทียบประสิทธิภาพความถูกต้อง

3.ผลการทดลอง

จากการจำแนกสถานะทารกในครรภ์โดยการใช้เครื่องมือติดตามอัตราการเต้นของหัวใจ และการหดรัดตัวของมดลูกมาทำการทดลองแบบ 10-folds cross validation กับกลุ่มเทคนิคต้นไม้ตัดสินใจได้แก่ DecisionStump, Hoeffding Tree, J48, LMT (Logistic Model Tree), RandomForest, RandomTree และ REPTree กับชุดข้อมูล Cardiotocography จำนวน 2,126 ตัวอย่าง ผลการทดลองปรากฏผลการจำแนกตัวอย่างโดยแบ่งเป็น ได้ 3 class คือ Normal, Suspect และ Pathologic เมื่อพิจารณาแล้วผลลัพธ์ที่ได้ให้ค่าความถูกต้องสูงที่สุดจากการทดลอง ซึ่งพบว่า J48 แสดงอัตราการจำแนกถูกต้องร้อยละ 92.79 รองลงมา RandomForest แสดงอัตราการจำแนกถูกต้องร้อยละ 92.50 และตามมด้วย LMT แสดงอัตราการจำแนกถูกต้องร้อยละ 90.74 REPTree แสดงอัตราการจำแนกถูกต้องร้อยละ 88.25 RandomTree แสดงอัตราการจำแนกถูกต้องร้อยละ 84.85 HoeffdingTree แสดงอัตราการจำแนก

ถูกต้องร้อยละ 79.12 และ DecisionStump แสดงอัตรา
การจำแนกถูกต้องร้อยละ 66.91 ตามลำดับ ซึ่งโมเดล
จำแนกสถานะทารกด้วยเทคนิค J48 แสดงดังภาพที่ 8



J48

4.อภิปรายผลและสรุปผลการศึกษา

ในการจำแนกสถานะทารกในครรภ์โดยการใช้
เครื่องมือติดตามอัตราการเต้นของหัวใจ และการหดรัดตัว
ของมดลูกด้วยเทคนิคต้นไม้ตัดสินใจนี้ คณะผู้วิจัยได้นำชุด
ข้อมูลตัวอย่าง Cardiotocograph จำนวน 2,126 ตัวอย่าง
มาสร้างโมเดลการจำแนกสถานะทารกในครรภ์ด้วยเทคนิค
ต้นไม้ตัดสินใจ ซึ่งได้เปรียบเทียบประสิทธิภาพการจำแนก
จาก 7 เทคนิคอันได้แก่ DecisionStump, HoeffdingTree,
J4 8 , LMT, RandomForest, RandomTree และ
REPTree และออกแบบการทดลองแบบ 10-fold cross
validation โดยอัตราร้อยละของความถูกต้องในการ
จำแนกสถานะทารกในครรภ์พบว่าเทคนิค J48 สามารถ
จำแนกได้ถูกต้องสูงถึงร้อยละ 92.79 กับชุดข้อมูล
Cardiotocograph นี้ สาเหตุที่เป็นเช่นนี้ก็เพราะเทคนิค
J48 เป็นอัลกอริทึมที่มีการปรับปรุงให้สามารถจัดการ
คุณลักษณะที่มีมูลค่าตัวเลขเป็นทศนิยม นอกจากนี้

อัลกอริทึม J48 เองยังสามารถลดความผิดพลาดในการ
จำแนกข้อมูลลงได้เพราะมีการตัดทอนความผิดพลาด
ออกไป (Pruning node)

จากการวิจัยและทดลองในงานนี้ทำให้ผู้วิจัยพบ
แนวทางที่สามารถพัฒนาต่อยอดได้สองแนวทางคือ สามารถ
เพิ่มประสิทธิภาพการจำแนกให้สูงขึ้นได้จากการอัลกอริทึม
อื่น และพัฒนาต่อยอดเป็นอุปกรณ์รองรับโมเดลในการ
จำแนกสถานะทารกในครรภ์ให้เป็นอุปกรณ์ช่วยคัดกรอง
สถานะทารกในครรภ์เบื้องต้นเพื่อช่วยในการลดอัตราการ
เสียชีวิตของทารก และช่วยแบ่งเบาภาระงานของบุคลากร
ทางการแพทย์

5.องค์ความรู้ใหม่ที่ได้จากงานวิจัย

สำหรับองค์ความรู้ใหม่ที่เกิดขึ้นจากการทำงาน
วิจัยชิ้นนี้เป็นการนำเทคนิคต้นไม้ตัดสินใจ (Decision
Tree) ซึ่งเป็นองค์ความรู้ทางด้านเทคโนโลยีปัญญาประดิษฐ์

มาสร้างโมเดลการจำแนกสถานะของทารก โดยสามารถนำโมเดลต้นแบบนี้ไปพัฒนาต่อยอดในการพัฒนาเป็นแอปพลิเคชันเพื่อช่วยคัดกรองสถานะทารกเบื้องต้น แบ่งเบาภาระบุคลากรทางการแพทย์ในสถานพยาบาล และสะดวกในการใช้ประโยชน์

6.เอกสารอ้างอิง

- [1] A. H. Gee, R. Barbieri, D. Paydarfar and P. Indic, "Predicting Bradycardia in Preterm Infants Using Point Process Analysis of Heart Rate," in IEEE Transactions on Biomedical Engineering, vol. 64, no. 9, pp. 2300-2308, Sept. 2017, doi: 10.1109/TBME.2016.2632746.
- [2] McKinney D, House M, Chen A, et al. The influence of interpregnancy interval on infant mortality. Am J Obstet Gynecol 2017;216:316.e1-9.
- [3] R. Ramos, C. Silva, M. W. L. Moreira, J. J. P. C. Rodrigues, M. Oliveira and O. Monteiro, "Using predictive classifiers to prevent infant mortality in the Brazilian northeast," 2017 IEEE 19th International Conference on e-Health Networking, Applications and Services (Healthcom), 2017, pp. 1-6, doi: 10.1109/HealthCom.2017.8210811.
- [4] P. A. Warrick, E. F. Hamilton, D. Precup and R. E. Kearney, "Classification of Normal and Hypoxic Fetuses From Systems Modeling of Intrapartum Cardiotocography," in IEEE Transactions on Biomedical Engineering, vol. 57, no. 4, pp. 771-779, April 2010, doi: 10.1109/TBME.2009.2035818.
- [5] N. Chamidah and I. Wasito, "Fetal state classification from cardiotocography based on feature extraction using hybrid K-Means and support vector machine," 2015 International Conference on Advanced Computer Science and Information Systems (ICACSIS), 2015, pp. 37-41, doi: 10.1109/ICACSIS.2015.7415166.
- [6] A. M. Posonia, S. Vigneshwari and D. J. Rani, "Machine Learning based Diabetes Prediction using Decision Tree J48," 2020 3rd International Conference on Intelligent Sustainable Systems (ICISS), 2020, pp. 498-502, doi: 10.1109/ICISS49785.2020.9316001.
- [7] U. Akkanapalli, M. Mudigonda and A. Palakurti, "Digitization and Analysis of Cardiotocography Records," 2020 IEEE-HYDCON, 2020, pp. 1-3, doi: 10.1109/HYDCON48903.2020.9242781.
- [8] M. Ramla, S. Sangeetha and S. Nickolas, "Fetal Health State Monitoring Using Decision Tree Classifier from Cardiotocography Measurements," 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS), 2018, pp. 1799-1803, doi: 10.1109/ICCONS.2018.8663047.
- [9] N. Sevani, I. Hermawan and W. Jatmiko, "Feature Selection based on F-score for Enhancing CTG Data Classification," 2019 IEEE International Conference on Cybernetics and Computational Intelligence (CyberneticsCom), 2019, pp. 18-22, doi: 10.1109/CYBERNETICSCOM.2019.8875656.
- [10] K. Agrawal and H. Mohan, "Cardiotocography Analysis for Fetal State Classification Using Machine Learning Algorithms," 2019 International Conference on Computer Communication and Informatics (ICCCI), 2019, pp. 1-6, doi: 10.1109/ICCCI.2019.8822218.
- [11] P. Hamelmann et al., "Doppler Ultrasound Technology for Fetal Heart Rate Monitoring: A Review," in IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control, vol. 67, no. 2, pp. 226-238, Feb. 2020, doi: 10.1109/TUFFC.2019.2943626.
- [12] N. Chumuang et al., "An Efficiency Random Forest Algorithm for Classification of Patients with Kidney Dysfunction," 2020 15th International Joint Symposium on Artificial Intelligence and Natural Language Processing

- (iSAI-NLP), 2020, pp. 1-6, doi: 10.1109/iSAI-NLP51646.2020.9376785.
- [13] หิรัญจันทร์ศ., ปันดี ญ., และ เกตุฉ่ำม., “ระบบตรวจจับและแจ้งเตือนข้อมูลทะเบียนรถ สำหรับกล้องวงจรปิดบนทางเท้า”, STIJ, ปี 1, ฉบับที่ 1, น. 1-7, ก.พ. 2020.
- [14] ว่องวุฒิไกรธ., “ระบบตรวจจับบุคคลเพื่อเฝ้าระวังในพื้นที่ที่กำหนด”, STIJ, ปี 1, ฉบับที่ 1, น. 39-46, ก.พ. 2020.
- [15] สุนนพันธ์., ธงชัยร., คำชื่นช., และ วิริยะรัตนกุล., “ระบบควบคุมและแจ้งเตือนการเจริญเติบโตของเห็ดนางฟ้าภูฐาน ด้วยการประมวลผลภาพและอินเทอร์เนตในทุกสิ่ง”, STIJ, ปี 2, ฉบับที่ 3, น. 1-10, ก.ค. 2021.
- [16] ใจเพชรก. และ ตั้งตราชูพ., “ระบบคัดกรองผู้ที่มีอาการป่วยในที่สาธารณะ”, STIJ, ปี 1, ฉบับที่ 4, น. 16-19, พ.ย. 2020.
- [17] S. Thiennoi, N. Chumuang and C. Siladech, "An Efficiency Comparison for Predicting of Educational Achievement Based on LMT," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045703.
- [18] P. Mangalkar and V. Barkade, "Efficient Categorization of Document with J48 Multi-Class Classifiers," 2019 5th International Conference On Computing, Communication, Control And Automation (ICCUBEA), 2019, pp. 1-5, doi: 10.1109/ICCUBEA47591.2019.9128665.
- [19] W. Ninphet, N. Chumuang, C. Siladech and M. Ketcham, "The Development of a Model to Predict Marbling Score for fattening Kamphaeng Saen Beef Breed Using Data Mining," 2019 14th International Joint Symposium on Artificial Intelligence and Natural Language Processing (iSAI-NLP), 2019, pp. 1-6, doi: 10.1109/iSAI-NLP48611.2019.9045289.