

## สัญญาเลขที่ RMU5080010

โครงการวิจัย: ตัวแบบการวินิจฉัยโรคทางการแพทย์โดยการประมวลผลภาพโดยเซลล์ดาร์  
ออโตมาตาและเครือข่ายความเชื่อเบย์

### **Medical Diagnosis Model Using Cellular Automata-based Image Processing and Bayesian Belief Network**

#### **Abstract**

Cellular Automata (CA) are discrete spatiotemporal systems with local interactions. A behavior of cellular automata is fascinating not only from a theoretical perspective but also from an experimental perspective, especially for digital image processing. This research aims to present a medical diagnosis model. CA-based algorithms for mammogram image processing are proposed. A statistical influence diagram, called Bayesian Belief Network (BBN), was investigated in modeling the medical breast cancer diagnosis under supervision by medical experts. Four types of datasets, namely, historic biodata, physical findings, indirect and direct mammographic findings were taken into consideration for modeling the BBN. Biodata are comprised of age, number of relatives having breast cancer, age at first live birth and age at menarche. Physical findings consist of pain, axilla, inflame and nipple discharge. Indirect mammographic data are breast composition. Direct mammographic findings were information obtained by mammogram image processing using the proposed cellular automata algorithms. In this regard, pectoral muscle representing a predominant density region must be removed by the proposed cellular automata algorithm prior to processing to determine the presence of mass and calcification. A dataset in real case of the breast cancer patients who came to get serviced at Srinakarind Hospital, Khon Kaen University, Thailand was collected. In this respect, a dataset consisting of 500 cases is used for performance evaluation of the proposed model. For training propose, an 80% of data was used for training the model, while the rest, 20%, was for testing. The trained BBN model is tested on 100 patients consisting of 50, 25 and 25 for normal, benign and malignant patients, respectively. It reports the promising result of 96.4% for accuracy. In addition, 5-fold and 10-fold cross-validation are implemented. The proposed BBN reports the promising results providing 96.2 and 97.4 percentages of accuracy, respectively.

## บทคัดย่อ

เซลล์ลาร์อโตมาตาเป็นระบบซับซ้อนเชิงพลวัตที่มีการเปลี่ยนแปลงสถานะเชิงพื้นที่และเวลาแบบไม่ต่อเนื่องโดยอาศัยความสัมพันธ์แบบท้องถิ่น พฤติกรรมของเซลล์ลาร์อโตมาตาได้รับความสนใจไม่เพียงแต่ในเชิงทฤษฎีเท่านั้น แต่ได้รับความสนใจเพื่อนำไปประยุกต์ใช้งานในด้านต่างๆ โดยเฉพาะอย่างยิ่งทางด้านการประมวลผลภาพ โครงการวิจัยนี้นำเสนอตัวแบบการวินิจฉัยโรคมะเร็งเต้านมโดยส่วนประกอบหลัก 2 ส่วน ได้แก่ 1) ขั้นตอนวิธีการประมวลผลภาพเอกซเรย์เต้านม (ภาพแมมโมแกรม) โดยใช้ตัวแบบเซลล์ลาร์อโตมาตา 2) ขั้นตอนวิธีการวินิจฉัยการเกิดโรคมะเร็งเต้านมโดยใช้ตัวแบบเครือข่ายความเชื่อเบย์ สำหรับขั้นตอนที่ 2) จะนำข้อมูลผลลัพธ์จากขั้นตอนที่ 1) และ ข้อมูลอื่นๆ ได้แก่ ข้อมูลคลินิก ข้อมูลเชิงกายภาพ มาเป็นตัวแปรสำหรับการสร้างเครือข่ายเบย์ ข้อมูลคลินิกประกอบ ได้แก่ อายุ จำนวนญาติพี่น้องที่ป่วยเป็นมะเร็งเต้านม อายุการมีประจำเดือนครั้งแรกของผู้ป่วย อายุการมีบุตรครั้งแรก ข้อมูลเชิงกายภาพ ได้แก่ อาการที่ได้จากการสังเกตทางกายภาพ เช่น อาการปวดเต้านม การเกิดแผลที่เต้านม การมีของเหลวไหลออกมาจากหัวนม สำหรับข้อมูลจากแมมโมแกรมทางอ้อมได้แก่ องค์ประกอบเต้านม ส่วนข้อมูลจากภาพแมมโมแกรมทางตรงได้แก่ผลจากการประมวลผลภาพตามที่นำเสนอในขั้นตอนที่ 1) ซึ่งแยกเป็นขั้นตอนวิธีการคัดค้านเนื้อหน้าอกจากภาพแมมโมแกรมออกจากเต้านมก่อนการนำภาพมาประมวลผลการหาบริเวณที่สงสัยว่าจะเป็นเนื้องอก การหาค่าคุณสมบัติของเนื้องอกจากการเก็บรวบรวมข้อมูลคนไข้ที่เข้ารับการตรวจมะเร็งเต้านมที่โรงพยาบาลศรีนครินทร์ มหาวิทยาลัยขอนแก่น จำนวน 500 ราย เพื่อนำมาฝึกสอนตัวแบบเครือข่ายความเชื่อเบย์ และใช้ข้อมูลทดสอบตัวแบบ จำนวน 100 ราย แยกเป็น คนไข้ปกติ 50 ราย คนไข้ที่พบเนื้องอกในเต้านมแต่ไม่เป็นมะเร็ง (benign) จำนวน 25 คน และคนไข้ที่พบเนื้องอกในเต้านมและเป็นมะเร็ง (malignant) จำนวน 25 ราย ระบบเครือข่ายเบย์ที่นำเสนอรายงานความถูกต้องสูงถึงร้อยละ 96.4 และนอกจากนี้จากการวัดประสิทธิภาพด้วยวิธี k-fold cross validation ที่ระดับ 5-fold และ 10-fold ระบบที่นำเสนอสามารถรายงานความถูกต้องถึงร้อยละ 96.2 และ 97.4 ตามลำดับ