รายงานวิจัยฉบับสมบูรณ์

โครงการวิจัยตัวแบบการวินิจฉัยโรคทางการแพทย์โดยการประมวลผลภาพโดยเซลลูลาร์ออโตมาตาและเครือข่ายความเชื่อเบย์

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

โดย รองศาสตราจารย์ ดร.ศาสตรา วงศ์ธนวสุ และคณะ

กุมภาพันธ์ 2557

สัญญาทุนเลขที่ RMU5080010

รายงานวิจัยฉบับสมบูรณ์

โครงการวิจัยตัวแบบการวินิจฉัยโรคทางการแพทย์โดยการประมวลผลภาพโดยเซลลูลาร์ออโตมาตาและเครือข่ายความเชื่อเบย์

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

คณะผู้วิจัย	สังกัด
1. รศ ดร. ศาสตรา วงศ์ธนวสุ หัวหน้านักวิจัย	ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น
2. รศ. นพ. วรชัย ตั้งวรพงศ์ชัย นักวิจัย	ภาควิชารังสีวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

สนับสนุนโดยสำนักงานกองทุนสนับสนุนการวิจัย

สารบัญ

หน้า

สัญญาเลขที่ **RMU5080010**

โครงการวิจัย: ตัวแบบการวินิจฉัยโรคทางการแพทย์โดยการประมวลผลภาพโดยเซลลูลาร์ออโตมาตาและเครือข่ายความเชื่อเบย์

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

เซลลูลาร์ออโตมาตาเป็นระบบซับซ้อนเชิงพลวัตรที่มีการเปลี่ยนแปลงสถานะเชิงพื้นที่และเวลาแบบไม่ต่อเนื่องโดยอาศัยความสัมพันธ์แบบท้องถิ่น พฤติกรรมของเซลลูลาร์ออโตมาตาได้รับความสนใจไม่เพียงแต่ในเชิงทฤษฎีเท่านั้น แต่ได้รับความสนใจเพื่อนำไปประยุกต์ใช้งานในด้านต่างๆ โดยเฉพาะอย่างยิ่งทางด้านการประมวลผลภาพ โครงการวิจัยนี้นำเสนอตัวแบบการวินิจฉัยโรคมะเร็งเต้านมโดยส่วนประกอบหลัก 2 ส่วน ได้แก่ 1) ขั้นตอนวิธีการประมวลผลภาพเอกซ์เรย์เต้านม (ภาพแมมโมแกรม) โดยใช้ตัวแบบเซลลูลาร์ออโตมาตา 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) ซึ่งแยกเป็นขั้นตอนวิธีการคัดกล้ามเนื้อหน้าอกจากภาพแมมโมแกรมออกจากเต้านมก่อนการนำภาพมาประมวลผลการหาบริเวณที่สงสัยว่าจะเป็นเนื้องอก การหาค่าคุณสมบัติของเนื้องอก จากการเก็บรวบรวมข้อมูลคนไข้ที่เข้ารับการตรวจมะเร็งเต้านมที่โรงพยาบาลศรีนครินทร์ 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

Executive Summary

I. ชื่อโครงการ (ภาษาไทย) โครงการวิจัยตัวแบบการวินิจฉัยโรคทางการแพทย์โดยการประมวลผลภาพโดยเซลลูลาร์ออโตมาตาและเครือข่ายความเชื่อเบย์

(ภาษาอังกฤษ) Medical Diagnosis Model Using Cellular Automata-based Image Processing and Bayesian Belief Network

II. หัวหน้าโครงการ

ชื่อนักวิจัยที่ได้รับทุน (หัวหน้านักวิจัย)

รศ.ดร.ศาสตรา วงศ์ธนวสุ

ภาควิชาวิทยาการคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

โทร. 0-4336-2189-91 โทรสาร 0-43342-910 email: wongsar@kku.ac.th

ชื่อนักวิจัย

รศ. นพ. วรชัย ตั้งวรพงศ์ชัย

ภาควิชารังสีวิทยา คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น

โทร. 0-8195-40853 email: vorachai@kku.ac.th

III. สาขาวิชาที่ทำวิจัย วิทยาการคอมพิวเตอร์

V. ระยะเวลาดำเนินงาน 3 ปี

VI. ไม่ได้เสนอโครงการนี้ต่อแหล่งทุนอื่น

VII. ปัญหาที่ทำวิจัย และความสำคัญของปัญหา

ในปัจจุบันการตรวจวินิจฉัยโรคทางการแพทย์ได้มีการนำเอา Intelligent System มาประยุกต์ใช้งานเพื่อประสิทธิภาพและเพิ่มมากขึ้นเรื่อยๆ ในการนี้ Computer Aided Diagnosis (CAD) ได้รับความนิยมอย่างแพร่หลายในหลายๆ โดเมนของปัญหา สำหรับทางด้านการแพทย์ได้มีการพัฒนาระบบ CAD เป็นระบบผู้เชี่ยวชาญในการวินิจฉัยโรคทางการแพทย์ ตัวอย่าง เช่น Path Finder System ซึ่งเป็นระบบผู้เชี่ยวชาญทางการแพทย์สำหรับวินิจฉัยกลุ่มโรค lymph-node ที่มีประสิทธิภาพดีที่สุดในโลกในปัจจุบัน สามารถวินิจฉัยโรคทางด้าน lymph-node ที่ให้ความถูกต้อง

ดีกว่า Computer Aided Medical Diagnosis ทุกระบบที่ปรากฏในปัจจุบัน และดีกว่าผู้เชี่ยวชาญโรค lymph-node ที่เป็นมนุษย์

การพัฒนาระบบผู้เชี่ยวชาญ Computer Aided Medical Diagnosis สำหรับวินิจฉัยการเกิดโรคมะเร็งเต้านมซึ่งเป็นหนึ่งในสองอันดับแรกของโรคมะเร็งที่ตรวจพบสูงสุดในประชากรของประเทศไทย โดยมีแนวโน้มที่จำนวนผู้ป่วยจะเพิ่มขึ้นเรื่อยๆ ตามความเจริญของการใช้ชีวิตของมนุษย์ ซึ่งเป็นโจทย์วิจัยที่น่าสนใจเป็นอย่างมาก อย่างไรก็ตามการวินิจฉัยโรคมะเร็งเต้านมซึ่งประเมินผลโดยระบบ Computer Aided Diagnosis ต้องอาศัยข้อมูลในลักษณะเดียวกันกับแพทย์ผู้เชี่ยวชาญที่เป็นมนุษย์ ได้แก่ ข้อมูลเชิงคลินิก และ ภาพเอกซ์เรย์เต้านม (mammogram) เป็นต้น ด้วยเหตุผลดังกล่าวการพัฒนาวิธีการประมวลผลทางการแพทย์โดยอาศัยเทคนิคที่มีประสิทธิภาพโดยตัวแบบเซลลูลาร์ออโตมาตาเป็นสิ่งที่ท้าทาย โครงการวิจัยนี้นำเสนอการพัฒนาระบบตรวจวินิจฉัยโรคมะเร็งเต้านมโดยการประมวลผลภาพ mammogram โดยใช้เซลลูลาร์ออโตมาตาเพื่อหาเนื้องอก และ การเกิดหินปูน (calcification) และหาคุณสมบัติของทั้งสองประกอบกับข้อมูลคลินิก เพื่อนำเข้าเป็นพารามิเตอร์ของตัวแบบเครือข่ายความเชื่อเบย์สำหรับการวินิจฉัยการเกิดมะเร็งเต้านมต่อไป

VIII. วัตถุประสงค์

1) เพื่อนำเสนอขั้นตอนวิธีการประมวลผลภาพ และแบ่งส่วนภาพแมมโมแกรมโดยใช้ตัวแบบเซลลูลาร์ออโตมาตา
2) เพื่อพัฒนาระบบ Computer Aided Medical Diagnosis สำหรับวินิจฉัยการเกิดมะเร็งเต้านมโดยใช้ตัวแบบเครือข่ายความเชื่อเบย์

IX. ระเบียบวิธีวิจัย

ขั้นตอนในการวิจัยประกอบด้วยการประมวลผลภาพและการแบ่งส่วนภาพแมมโมแกรมโดยใช้ตัวแบบเซลลูลาร์ออโตมาตาเพื่อหาเนื้องอก หินปูน และคุณสมบัติของทั้งสอง และสร้างตัวแบบการระบบวินิจโรคมะเร็งเต้านม (Computer Aided Medical Diagnosis) โดยใช้ตัวแบบเครือข่ายความเชื่อเบย์ มีรายละเอียดดังนี้

1) ศึกษาค้นคว้างานวิจัย บทความวารสารที่เกี่ยวข้อง และเข้าร่วมไประชุมวิชาการที่เกี่ยวข้อง

2) ศึกษาแมมโมแกรมระยะต่างๆ จากแพทย์ผู้เชี่ยวชาญ โดยอาศัยข้อมูลจากโรงพยาบาลศรีนครินทร์ คณะแพทยศาสตร์ มหาวิทยาลัยขอนแก่น
3) ศึกษาทฤษฎีและหลักการตัวแบบเซลลูลาร์ออโตมาตา ตัวแบบเครือข่ายความเชื่อเบย์
4) นำเสนอขั้นตอนวิธีการประมวลผลภาพและการแบ่งส่วนภาพแมมโมแกรมโดยตัวแบบเซลลูลาร์ออโตมาตา
5) นำเสนอขั้นตอนวิธีการสร้างระบบวินิจฉัยโรคมะเร็งเต้านมโดยใช้ตัวแบบเครือข่ายความเชื่อเบย์
6) ทดสอบประสิทธิภาพ และปรับระบบเครือข่ายความเชื่อเบย์จากข้อเสนอแนะของ Domain experts
7) ทดลอง และวิเคราะห์ผล

X. ผลการดำเนินงาน

การดำเนินงานประกอบด้วย 4 ขั้นตอนสำคัญ ได้แก่

1) Cellular Automata-based Image Processing
2) Cellular Automata-based Pectoral Muscle Removal
3) Cellular Automata-based Mass Segmentation
4) Breast Cancer Diagnosis Using Bayesian Belief Network

โดยมีรายละเอียดแสดงในหัวข้อ “เนื้อหาของการวิจัย” และ ผลการดำเนินงานมีดังนี้

ขั้นตอนที่ 1 และ 2 นำเสนอขั้นตอนวิธีการประมวลผลภาพเอกซ์เรย์เต้านม (Mammogram) โดยใช้ตัวแบบเซลลูลาร์ออโตมาตา ผลงานวิจัยตีพิมพ์ดังนี้

S. Wongthanavasu and V. Tangvoraphongchai, “CA-based Algorithms and Its Application in Medical Image Processing,” The Proceedings of the 14th IEEE Int. Conference on Image Processing (IEEE-ICIP 2007), IEEE Catalog No.: 07CH37925C, ISBN: 1-4244-1437-7, ISSN: 1522-4880, pp. III-41-III-44, 2007.

S. Wongthanavasu, “Cellular Automata for Medical Image Processing,” Book Chapter of the Book entitled “Cellular Autoamata – Innovative Modelling for Science and Engineering,” ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

ขั้นตอนที่ 3 เป็นการนำเสนอขั้นตอนวิธีการการแบ่งส่วนภาพเอกซ์เรย์เต้านม (Mammogram) โดยใช้ตัวแบบเซลลูลาร์ออโตมาตาโดยใช้ภาพนำเข้าที่ได้จากขั้นตอนที่ 1

S. Wongthanavasu and V. Tangvoraphongchai, "Cellular Automata-based Identification of The Pectoral Muscle in Mammograms," The Proceedings of the 3rd Int. Symposium on Biomedical Engineering (ISBME 2008), November 10-11, 2008, Bangkok, Thailand, pp. 294-298, 2008.

S. Wongthanavasu, "Cellular Automata for Medical Image Processing," Book Chapter of the Book entitled "Cellular Autoamata – Innovative Modelling for Science and Engineering," ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

ขั้นตอนที่ 4 เป็นการนำเสนอ Computer Aided Medical Diagnosis โดยใช้ตัวแบบเครือข่ายความเชื่อเบย์สำหรับการวินิจฉัยโรคมะเร็งเต้านม ผลงานวิจัยตีพิมพ์ดังนี้

S. Wongthanavasu, "A Bayesian Belief Network Model for Breast Cancer Diagnosis," Selected Papers of the 2010 Annual International Conference of German Operations Research Society. ISSN 0721-5924, ISBN 978-3-642-20008-3, e-ISBN 978-3-642-20009-0, Springer-Verlag Berlin Heidelberg, pp. 3-8, 2011.

XI. ประโยชน์ที่คาดว่าจะได้รับ

ผลงานวิจัยจากโครงการสามารถใช้เป็นต้นแบบในการพัฒนา Computer Aided Medical Diagnosis สำหรับการวินิจฉัยการเกิดโรค/กลุ่มอาการของโรคต่างๆ ได้อย่างมีประสิทธิภาพ ในส่วนของระบบที่นำเสนอนักวิจัยสามารถปรับปรุงประสิทธิภาพโดยนำตัวแปรอื่นๆ มาพิจารณาร่วมด้วย เช่น สภาพทางประชากร ได้แก่ ชุมชนเมือง ชุมชนชนบท เป็นต้น

นอกจากนี้ ผลจากการวิจัยสามารถนำไปสร้างงานวิจัยประยุกต์เพื่อสร้างเป็นซอฟท์แวร์เพื่อนำไปใช้ในโรงพยาบาล เพื่อเสนอแนะแพทย์ผู้เชี่ยวชาญในการวิเคราะห์โอกาสการเกิดมะเร็งเต้านม

1. เนื้อหาของการวิจัย

1.1) Cellular Automata-based Image Processing [section 4]
1.2) Cellular Automata-based Pectoral Muscle Removal [section 5]
1.3) Cellular Automata-based Mass Segmentation [section 6]
1.4) Breast Cancer Diagnosis Using Bayesian Belief Network [section 7]

The details are provided as follows:

1. INTRODUCTION

Cellular Automata (CA) are a discrete spatiotemporal system whose behavior is specified in terms of local interactions. They appear as natural tools for image processing due to their local nature and simple parallel computation implementation. In this respect, there are a number of papers which generally discuss cellular automata for image processing. Hernandez et al. [9] presented CA for elementary 2-D image enhancement. Wongthanavasu et al. [1,3] presented 3-D CA for edge detection on binary and grayscale images and compared its performance evaluation to well-known edge operators. Rosin [10] presented algorithms for training cellular automata for image processing. Besides these, there are some papers discussed medical image processing . Cheng et al. [5] presented methods for mass detection and classification. Viher et al. [12] presented cellular automata algorithms for follicle image recognition.

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms [11,13,15]. Its inclusion can affect the results of intensity-based image processing methods. Pectoral muscle removal is necessary for mammogram image processing. There is a number of papers presenting pectoral muscle removals. Most of them presents non-CA algorithms [12,13,16]. An mammographic images analysis society digital mammogram database is popular in testing benchmark [15]. In addition, expert systems to analyze breast cancer were built on several models, such as Support Vector Machines [5,6], Bayesian Belief Networks [7,8,11]. This research aims to provide promising algorithms for mammogram image processing using cellular automata model. In addition, a proposed expert system on the basis of Bayesian Belief Network is modeled to analyze the breast cancer.

2. TERMINOLOGY

C_j is the j^{th} class of the pixel values (states) in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2,..,m$ and $v^j(\alpha+\delta_i) \in C_j$.

$N(C_j)$ is a number of neighbors of α which fall into class C_j.

$sum(v^j(C_{target}))$ is summation of $v^j(C_{target})$.

C_{target} is the majority class containing maximal number of neighbors.

$v^j(C_{target})$ denotes all of $v^j(\alpha+\delta_i) \in C_{target}$.

$E(\alpha)$ is the edge pixel value.

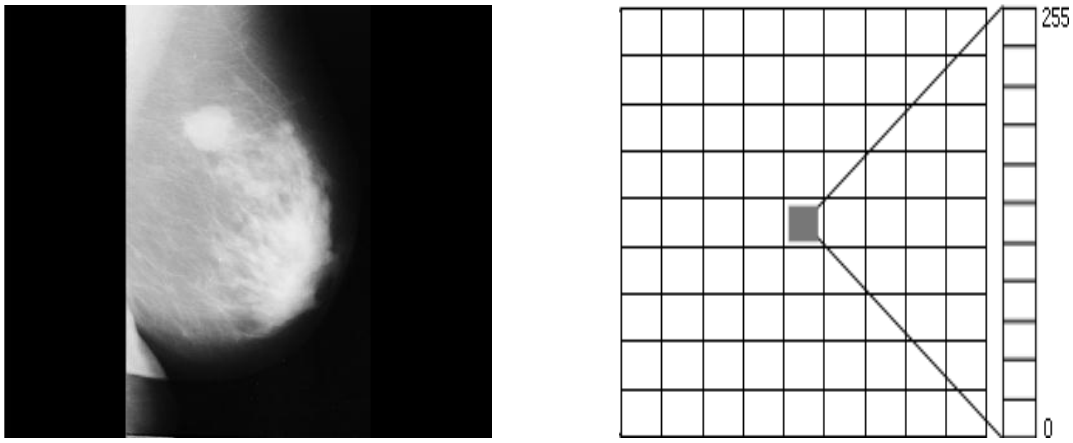
$B(\alpha)$ is the background pixel value.

k is a number of states.

3. CELLULAR AUTOMATA

Let I denote the set of integer. A 2-D cellular space is a 4-tuple, (IxI, V, N, f), where IxI is a set of cartesian product of two integer sets, V is a set of cellular states, N is the type of neighborhood, and f is the local transition function from Vn into V. The relevant neighborhood function is a function from IxI into 2IxI defined by $g(\alpha) = \{\alpha+\delta1, \alpha+\delta2,\ldots, \alpha+\delta n\}$, for all $\alpha \in$ IxI, where δi $(i = 1,2,\ldots,n) \in$ IxI is fixed. The neighborhood state function of a cell α at time t is defined by $h_t(\alpha) = (v_t(\alpha+\delta1), v_t(\alpha+\delta2), \ldots, v_t(\alpha+\delta n))$. For 2-D von Neumann neighborhood, the neighborhood state function of the central cell (α) is defined by: $h_t(\alpha) = (v_t(\alpha+(0,0)), v_t(\alpha+(0,1)), v_t(\alpha+(1,0)), v_t(\alpha+(0,-1)), v_t(\alpha+(-1,0))$, where $v_t(\alpha+(0,0)$ is current state of the central cell, $v_t(\alpha+(0,1)$ $(v_t(\alpha+(0,-1))$ for the north (south) cells, $v_t(\alpha+(1,0)$ $(v_t(\alpha+(-1,0))$ for the east (west) cells.

Now we relate the neighborhood state of a cell α at time t to the cellular state of that cell at time t+1 by $f(h_t(\alpha)) = v_{t+1}(\alpha)$. The function f is referred to as the 2-D CA rule and is usually given in the form of a state table, specifying all possible pairs of the form $(h_t(\alpha), v_{t+1}(\alpha))$. Figure 1 shows 2-D digital image and 2-D CA.



(a) 2-D digital image. (b) 2-D cellular automata.

Figure 1. 2-D digital image vs. 2-D CA.

4. CELLULAR AUTOMATA FOR MEDICAL IMAGE PROCESSING

As stated previously, cellular automata techniques appear as a natural tool for image processing due to their local nature and simple parallel computing implementation. In this section, we present one main algorithm and investigate its variation as methods for processing mammogram images. The methods will correspond to edge detection and noise removal for both binary and grayscale images, while the last one will correspond to spots detection for breast cancer diagnosis. Examples of the application of these cellular automata

techniques to real mammogram images will be presented, which together with the results will show the performance characteristic.

A. CELLULAR AUTOMATA FOR EDGE DETECTION

The main cellular automata algorithm for k gray levels of digital images is on the basis of a bi-dimensional cellular automata (IxI, V, N, f) with V = {0, 1, 2, …, k-1}, where k is a number of states, N is the type of neighborhood (e.g. n neighbors), while the local transition function f is from Vn into V. The proposed algorithm is shown in (1) following:

$$f((v^t(\alpha+\delta_1), v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_n)) = E(\alpha), \quad (1)$$

$$\text{if } \max_{j=0}^{k-1} N(C_j) = C_{target} \text{ and } \text{sum}(v^t(C_{target})) > k\text{-}1$$

$$= B(\alpha), \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values (states) in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2,..,m$ and $v^j(\alpha+\delta_i) \in C_j$.

$N(C_j)$ is a number of neighbors of α which fall into class C_j.

$sum(v^j(C_{target}))$ is summation of $v^j(C_{target})$.

C_{target} is the majority class containing maximal number of neighbors.

$v^j(C_{target})$ denotes all of $v^j(\alpha+\delta_i) \in C_{target}$.

$E(\alpha)$ is the edge pixel value.

$B(\alpha)$ is the background pixel value.

k is a number of states.

For class arrangement, histogram distribution will be utilized to suggest the range of each class.

A.1 GRAYSCALE IMAGES

The objective of the edge detection techniques is to enhance the magnitude of the local differences in gray level values between regions of the images. Over regions which are different, changes must be made to enhance the edges. The proposed variation of (1) which deals with this task is shown in formula (2) for Von Neumann's neighborhood, four classes and 256 gray levels as following:

$$\text{if } ((v^t(\alpha+\delta_1), v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_5)) = 255, \quad (2)$$

$$\text{if } \max_{j=0}^{3} N(C_j) = C_{target} \text{ and } sum(v^t(C_{target})) > 255$$

$$= 0, \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $v^t(\alpha+\delta_i) \in C_j$.

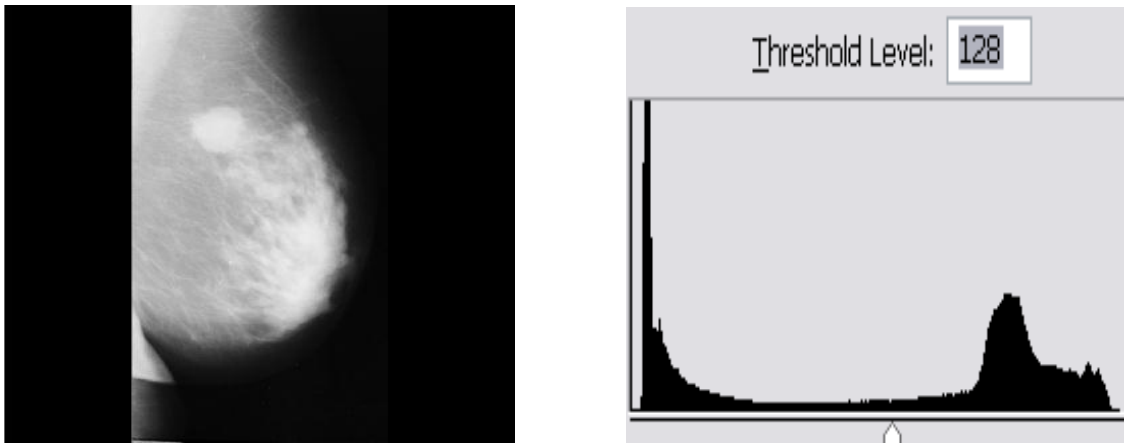
$N(C_j)$ is number of neighbors of α which fall into class C_j.

C_{target} is the majority class containing maximal number of neighbors $(v^t(\alpha+\delta_i) \in C_{target})$.

$sum(v^t(C_{target}))$ is summation of $v^t(C_{target})$.

$v^t(C_{target})$ denotes all of $v^t(\alpha+\delta_i) \in C_{target}$.

In implementing (2) in an original mammogram image being supervised by the histogram information (Fig.2) for the class arrangement, the results were shown in Figure 3.



(a) Original mammogram. (b) Histogram.

Figure 2. Original mammogram and its histogram.

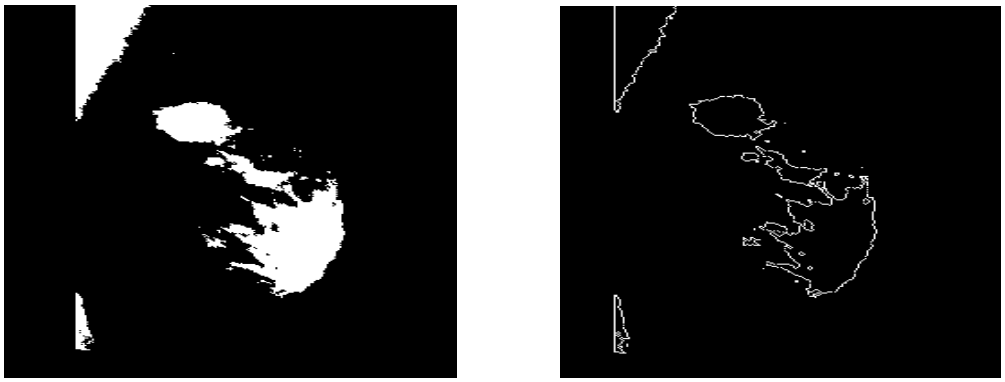
(a) Four class image. (b) Edge map result.

Figure 3. (a) Four classes of image, and (b) result of edge detection.

It is remarkable that the cellular automata algorithm simply provides the best edge maps shown in Figure 3(b).

A.2 BINARY IMAGES

In case of edge detection on binary image, the cellular automata algorithm in (2) is directly applied to binary image efficiently. It is no need to be changed. Figure 4 shows the promising result of the algorithm dealing with binary images. The edge result exhibits the superb quality with one pixel wide and edge has no break.



(a) Original binary image. (b) Edge result.

Figure 4. Edge result as dealing with binary image.

B. CELLULAR AUTOMATA FOR NOISE FILTERING

The objective of the noise filtering is to reduce the local differences in gray level values between regions of the images. Over regions which are similar, no changes must be made, in order to avoid the destruction of the main characteristics of the image. The proposed variation of cellular automata algorithm given in (1) using von Neumann's neighborhood and two states dealing with this task is shown in formula (3) as follow:

$$f((\nu^t(\alpha+\delta_1),\nu^t(\alpha+\delta_2),\ldots,\nu^t(\alpha+\delta_5)) = \max(\nu^t(C_{target})), \quad \text{if } \max_{j=0}^{3} N(C_j) = C_{target} \qquad (3)$$

$$= 0, \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $\nu^t(\alpha+\delta_i) \in C_j$.

$N(C_j)$ is number of neighbors of α which fall into class C_j.

C_{target} is the majority class containing maximal number of neighbors $(\nu^t(\alpha+\delta_i) \in C_{target})$.

$\nu^t(C_{target})$ denotes all of $\nu^t(\alpha+\delta_i) \in C_{target}$.

$max(\nu^t(C_{target}))$ is the maximal state of $\nu^t(C_{target})$.

Figure 5 shows an original binary mammogram with salt and pepper noise at 2%. The noise filtering result using one iteration of implementing such an algorithm is shown in Figure 6. In this respect, the cellular automata algorithm provides the promising result.

(a) (b)

Figure 5. Binary image with 2% salt and pepper noise (a) and the image after noise removal (b).

C. CELLULAR AUTOMATA FOR SPOT DETECTION

The objective of the spot detection is to assist the physicians and doctors in locating the hypothesis spots for breast cancer. The shape and spread region of the spots play a vital role for further steps of analysis and have to be comprehensively taken into account. In this regard, a set operator is presented to provide such an affect as following:

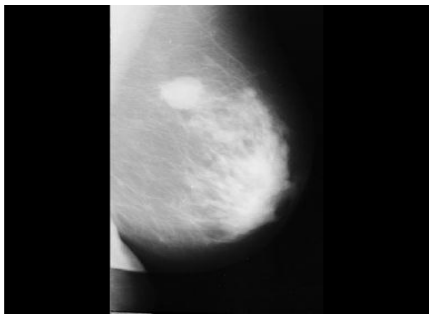
$$W = X - Y \qquad (4)$$

where X denotes an original image investigated,

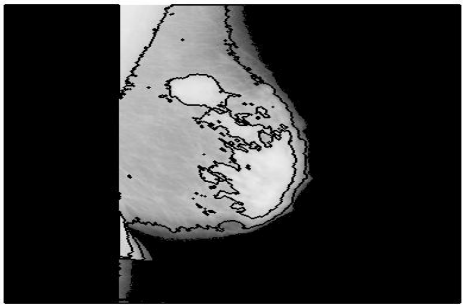
Y denotes an edge map due to formula (2), and

W denotes the resulting image.

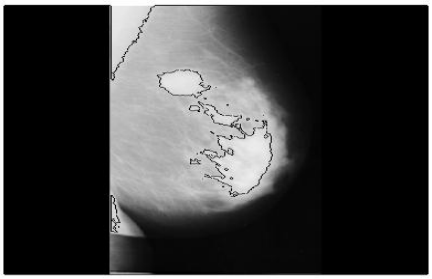
The difference (-) of two sets (images) is simply the subtraction of pixel values between two images (sets) in the same coordinate. By implementing such an operator in an original image (Fig. 7(a)) with respect to Y the results of formula (2) on grayscale and binary images, the resulting white spots detection were shown in Figure 7 (b) and (c), respectively.



(a) Malignant breast cancer



(b) Spots detection due to (2) for grayscale.



(c) Spot detection due to (2) for binary image.

Figure 7. Mammogram image processing with white spot detection: (a) original malignant breast cancer, (b) cancer spot detection due to (4) for grayscale, and (c) due to (4) for binary image.

5. CELLULAR AUTOMATA FOR IDENTIFICATION OF THE PECTORIAL MUSCLE IN MAMMOGRAMS

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms. Its inclusion can affect the results of intensity-based image processing methods. This paper presents a new method on the basis of cellular automata model for the identification of the pectoral muscle in MLO mammograms. A dataset of 84 MLO mammograms from the MIAS (Mammographic Image Analysis Society, London, U.K.) database was implemented throughout for evaluation. In this respect, the pectoral muscle edge detected in the mammograms was carried out based upon the percentage of false-positive (FP) and false-negative (FN) pixels determined by comparison between the numbers of pixels enclosed in the regions delimited by the edges identified by a radiologist and by the proposed method. The proposed CA-based method provides the promising results.

5.1 PECTORAL MUSCLE IDENTIFICATION

The inclusion of pectoral muscle in mammogram can affect the results of intensity-based image processing methods or bias procedures in the detection of breast cancer. The objective of the pectoral muscle identification is to determine the pectoral muscle for the exclusion from the mammograms and being used for further steps of breast cancer diagnosis. In this respect, a number of mammogram processing steps presented earlier and the segmentation algorithm given below were used for carrying out this task.

Algorithm 1: Algorithm for removal of unqualified objects.

1. Implement the formula (1) in a mammogram image P, resulting in the image Q.
2. Filter out the objects consisting of possible concatenated white pixels (edge) in Q that are less than 2,500 pixels, resulting in the image R.
   { An object is defined by contiguous white pixels. }

Algorithm 2: Algorithm for identification of pectoral muscle.

1. FOR $i := 1$ TO n DO
   Find the area of $ROI_r(i)$.
   { $ROI_r(i)$ is a region i inside the image R obtained by algorithm 1 and it is defined by an area of contiguous black pixels surrounded by a closed boundary edge of white pixels. }
2. Pectoral muscle segmentation:

2.1 MUSCLE := NIL {empty set}

2.2 FOR *j:=1* TO *n* DO

IF ROI_r (j) > max_size AND ROI_r (j) ≤ min_size

THEN ROI_r(j) := background.

{ Set the ROI_r(j) to background when it is not of interest regions. In practical uses in this paper, max_size = 95,200 pixles, and min_size = 2,500 pixles. }

ELSE IF min (y, ROI_r (j)) > min_upper

THEN ROI_r(j) := background.

{ min (y,ROI_r(j)) denotes minimal value of y for the coordinates (x,y) of all pixels in the region of ROI_r(j). In the paper, min_upper = 60.}

{ Regions of interest locate at min(y, ROI_r(j)) at the topmost part of the image R. }

ELSE IF mean (y, ROI_r (j)) > min_y_average

THEN ROI_r(j) := background.

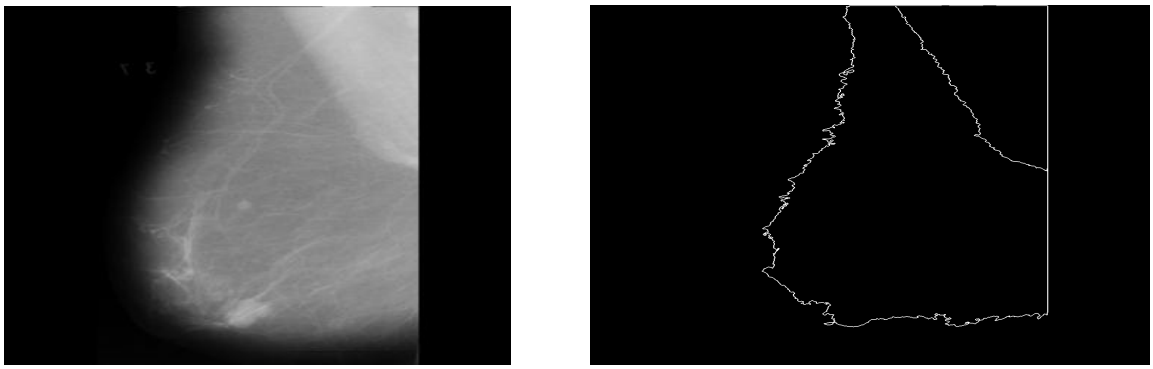
{ mean (y,ROI_r(j)) = average of y for the coordinates (x,y) of all pixels in the region of ROI_r(j). average_y_average = 350. }

ELSE IF mean (I, ROI_r (j)) > min_average_intensity

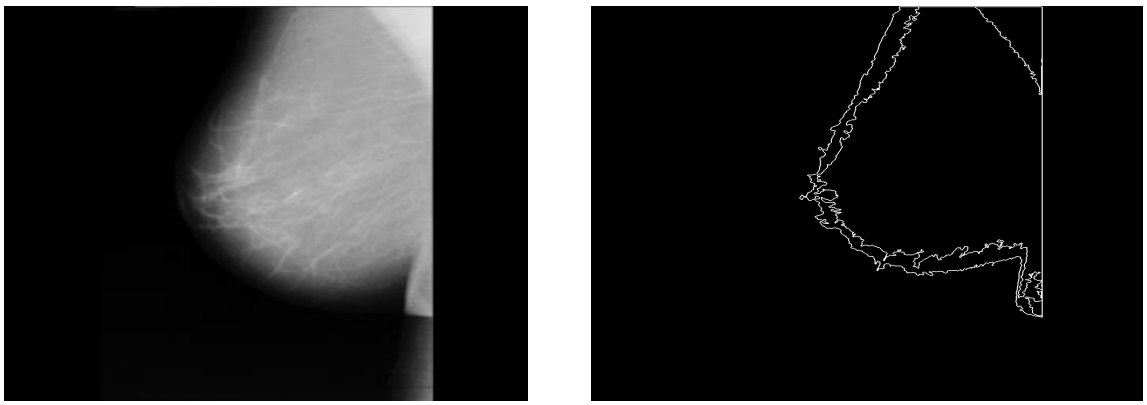
THEN MUSCLE := MUSCLE ∪ ROI_r(i)

{ mean (I, ROI_r(j)) = average intensity of all pixels in the region of ROI_r(j). In the paper, min_average_intensity = 157. }

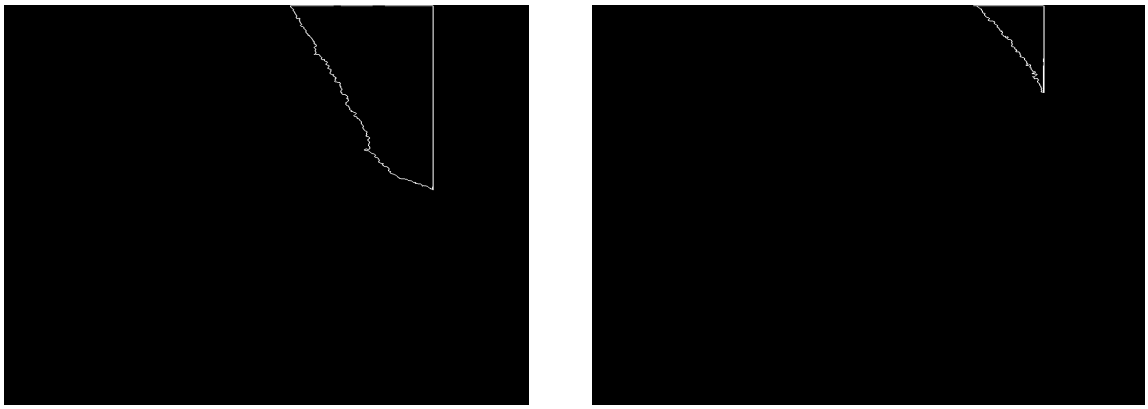
The results of implementing the algorithm 1 on mdb005 and mdb011 and the results of implementing the algorithm 2 were shown in Fig. 2. Fig. 3 shows hand-drawn pectoral muscle edge by radiologist as applied to mdb005 (c), the result obtained by the proposed algorithm (b), and the pectoral muscle removal according to the proposed algorithm (d).



(a) Original image (b) Result obtained by algorithm 1

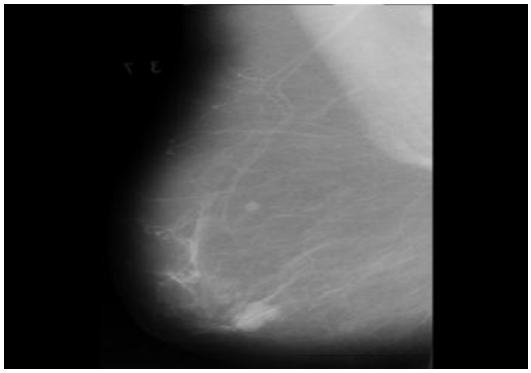


(c) Original image (d) Result obtained by algorithm 1

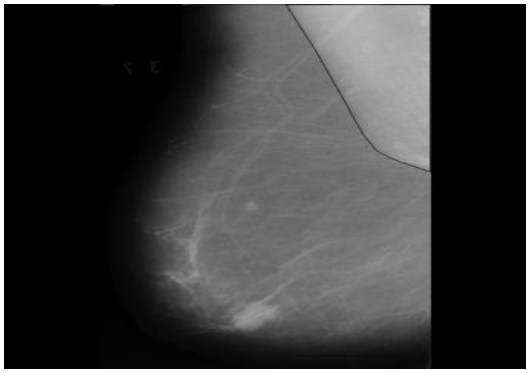


(e) Result obtained by algorithm 2 (f) Result obtained by algorithm 2

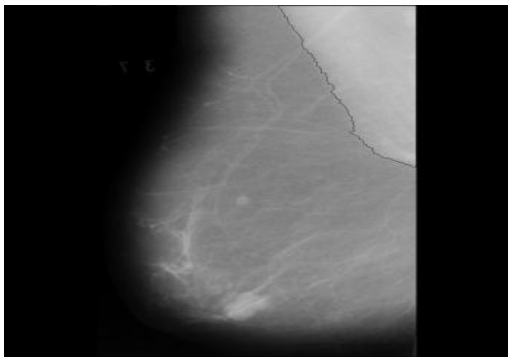
Fig. 2. Results obtained from algorithms 1 and 2. (a) and (c) are original images mdb005 and mdb011, respectively. (b) and (d) are results obtained from algorithm 1 as applied to (a) and (c), respectively. (e) and (f) are results obtained from algorithm 2 as applied to the edged images (b) and (d), respectively.



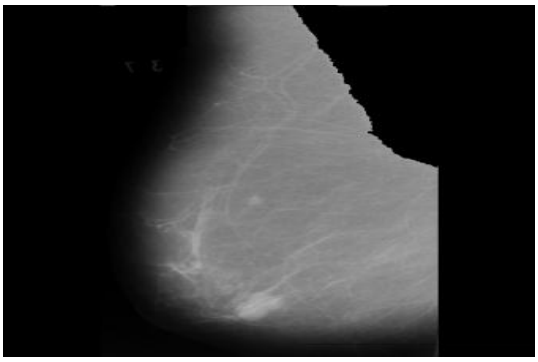
(a) Original image



(b) Hand-drawn pectoral muscle edge by radiologist



(c) Pectoral muscle edges detected by the algorithm.



(d) Pectoral muscle removal by the algorithm.

Fig. 3. Results obtained for the image mdb005. (a) Original image. (b) Hand-drawn pectoral muscle edge by radiologist superimposed on the original image. (c) Pectoral muscle edges detected by the CA-based method superimposed on the original image. (d) Pectoral muscle removal using the CA-based method.

5.2. RESULTS

The total of 84 images, randomly selected from the Mammographic Image Analysis Society, London, U.K. [7] were used in experimentation in this paper. The spatial resolution of these images is 200 μm and depth resolution in 8 bit. The images in the database are 1024x1024 pixels in size. The result obtained from the proposed method was evaluated in consultation with radiologists experienced in mammography. Then, the pectoral muscle edges were manually drawn by one of the authors under the supervision of radiologists, without referring to the results of detection by the proposed method. The segmentation results of the proposed method was evaluated based upon the number of false-positive (FP) and false-negative (FN) pixels in the regions demarcated by the

manually drawn edges. An FP pixels was defined as a pixel outside the reference region that was included in the pectoral region segmented. An FN pixel was defined as a pixel in the reference region that was not present within the segmented region. Table 1 shows mean and standard deviation values of the FP and FN pixels for the result of the proposed method with 84 images.

CA-based algorithm	**Statistics**
Analysis of area enclosed	
FP ± σ	1.99±8.19%
FN ± σ	18.89±14.19%
# images with (FP and FN) < 5%	13
# images with 5% <(FP and FN) <10%	15
# images with (FP and FN) > 10%	56

Table 1. Mean and standard deviation values of the FP and FN pixels for the results of CA-based algorithm with 84 images.

From Table 1, it can be seen that the identification obtained by the CA-based algorithm provides the promising results with minimal FP and FN. In this regard, it is one of promising methods for being used in further steps in breast cancer diagnosis.

6. CELLULAR AUTOMATA FOR MASS SEGMENTATION IN MAMMOGRAMS

6.1 AUTOMATIC MASS SEGMENTATION

This section presents a proposed CA-based algorithm for mass segmentation. To implement the algorithm, a pectoral muscle-removed mammogram shown in Fig. 3 (b) will be manually thresheld to arrive at the hypothesized mass seeds. Then, the seeds will be utilized as parameters in implementing the algorithm. The proposed segmentation algorithm utilizes a concept of two-types bacteria propagation. The first type confiscates a mass seed, which is an object image, while the second one confiscates the background which is declared by a circle surrounding the mass seed. Both types of bacteria propagate to their neighbors in parallel at each time step depending on its and their neighbor's strengths. A circle surrounding a mass seed shown in Fig. 4 (a) represents the background seed being seized by the other type of bacteria. Both types of bacteria continue propagating in parallel in discrete time steps as stated earlier. In order to reduce the computation time, the proposed algorithm can be stopped anytime earlier so far as

the propagation of bacteria taken up the mass seed is consistent. The CA-based segmentation algorithm is given as follow:

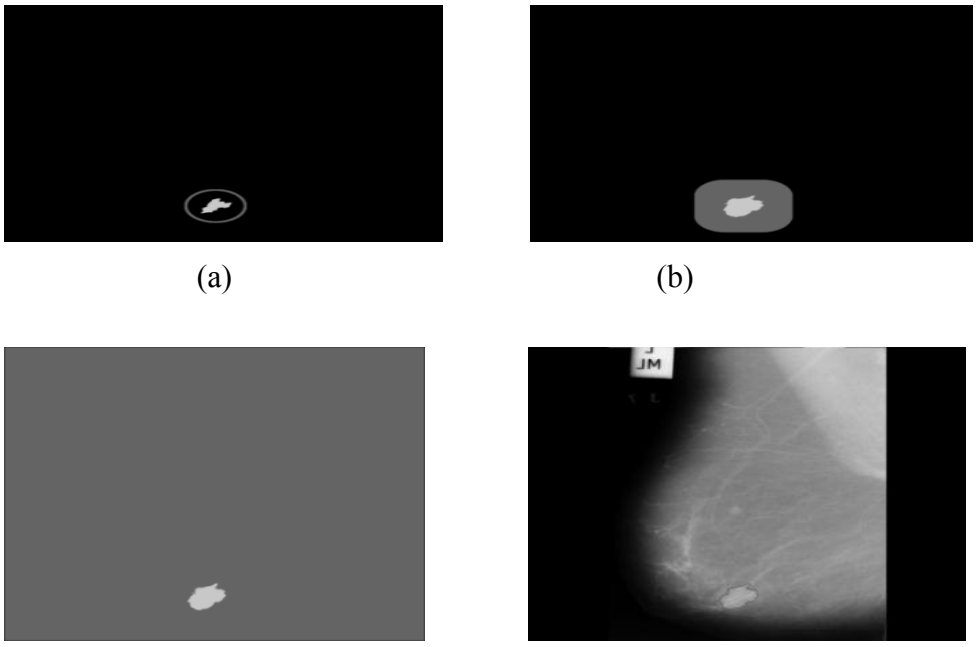
// Identify mass seed from pectoral-removed image
// by binary thresholding
Threshold an image P, resulting in object Q and background B. Given mass seed to object Q.
// Circle the mass seed
Circle the mass seed Q by a thick circle
// For each cell (pixel) in P
for $\forall p \in P$
// copy previous state
$l_p^{t+1} = l_p^t;$
$\theta_p^{t+1} = \theta_p^t;$
// neighbors try to attack current cell
for $\forall q \in N(p)$
if $g(||C_p\text{-}C_q||_2).\theta_q^t > \theta_p^t$ then
$l_p^{t+1} = l_q^t;$
$\theta_p^{t+1} = g(||C_p\text{-}C_q||_2).\theta_q^t$
end if
end for
end for

where P denotes an image, p denotes a pixel.
θ_q^t denotes the strength of a cell $q \in Q$ at time t.
θ_p^t denotes the strength of a p's neighboring cells at time t.
l_p^t, l_p^{t+1} denote the centering cell state of p at time t and $t+1$, respectively.
l_q^t, l_q^{t+1} denote the neighboring cell state of q at time t and $t+1$, respectively.
$N(p)$ denotes neighboring cells of p.
$||C_p\text{-}C_q||_2$ denotes the Euclidian distance between a cell q and a neighboring cell p.

$g(x)$ represents a monotonous decreasing function defined by (2) as following

$$g(x) = 1 - \frac{x}{\max \|C\|_2} \quad (2)$$

Figure 4 shows the results of implementing the proposed mass segmentation algorithm. The hypothesized mass seeds, white object images, and background seeds, which depict in a circle, are represented two types of bacteria as depicted in Fig. 4 (a). Fig. 4 (b) and (c) show a series of propagating results on intermediate and final evolution. Fig. 4 (d) depicts a hypothesized mass superimposed on the original mammogram.



(a) (b)

(c) (d)

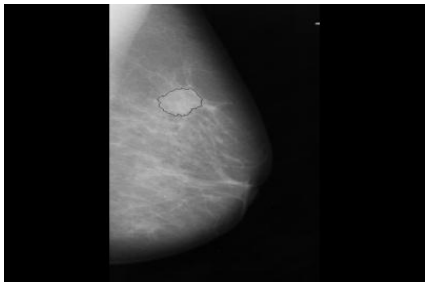
Figure 4. Mass segmentation obtained from the image mdb005 [6]: (a) mass seed initialization, (b) result at 42 steps of evolution, (c) result after completed evolution at 420 steps, and (d) mass segmentation result superimposed on the original mammogram.

For experimental purposes, the mammograms of mdb028 and mdb092 [6] were empirically implemented due to the proposed algorithm resulting in Figures 5 and 6, respectively.

(a) (b)

Figure 5. Mass segmentation results obtained from the image mdb028 [6]: (a) original image, (b) the result after completed evolution.

(a) Original image (b) Result after 442 iterations



(c) Mass segmentation

Figure 6. Mass segmentation results obtained from the image mdb092 [6]: (a) original image, (b) the result after completed evolution, and (c) the result of segmentation.

7. A BAYESIAN BELIEF NETWORK MODEL FOR BREAST CANCER DIAGNOSIS

A statistical influence diagram called Bayesian Belief Network (BBN) was investigated in modeling a mass cancer diagnosis for benign and malignant. BBN is a directed acyclic graph (DAG) in topology. Mammography is a process to obtain mammogram. It is an important tool in early detection of breast cancer. Unfortunately, many mammography

findings cannot be classified easily as malignant or benign. Successful diagnosis depends on the ability of a physician to detect mammographic abnormalities and to integrate clinical information such as risk factors and physical findings to determine the likelihood of breast cancer. In this regard, machine learning [5],[7],[8] are capable to successfully assist physicians in detecting the early detection of breast cancer. A Bayesian Belief Network (BBN) is one of those promising machine learning techniques. This section aims at developing an expert system for breast cancer diagnosis using a BBN.

7.1 MODEL STRUCTURE OF BAYESIAN BELIEF NETWORKS

A Bayesian network — also called a belief network or causal probabilistic network — is a graphical representation of probabilistic information: it is a directed, acyclic graph in which nodes represent random (stochastic) variables, and links between nodes represent direct probabilistic influences between the variables [7]. In this formalism, propositions are given numerical probability values signifying the degree of belief accorded them, and the values are combined and manipulated according to the rules of probability theory. Each node represents a variable and has two or more possible states. For example, the variable "Breast Cancer" has two states: "present" and "absent". Each state is associated with a probability value; for each node, these probability values sum to 1.

For implementing the proposed BBN, nodes and their states are enumerated in **Table 1** and associated network shown in **Figure 1**.

Table 1 Definition of BBN model's nodes (variables) and their states.

Category	**Node (variables)**	**States**
Diagnosis	Breast Cancer	present, absent
Patient History	Age (years)	<20, 20-30, 31-40, 41-50, 51-60, >60
	Age at Menarche (years)	<12, 12-13, >13
	Age at First Live Birth (years)	<20, 20-24, 25-29, >29
	Number of First-Degree Relatives with Breast Cancer	0, 1, 2
	Age at Menopause (years)	<40, 40-44, 45-40, >49
	Number of Pregnancy	<2, 2-3, >3
Physical Findings	Pain	present, absent

	Nipple Discharge Axilla Inflame	present, absent present, absent present, absent
Indirect Mammographic Findings	Breast Composition	present, absent
Direct Mammographic Findings	Mass Mass Present Mass Lympnode Mass Size (cm.) Calcification Calcification Present	malignant, benign, none yes, no present, absent 1,2,3,4 malignant, benign, none yes, no

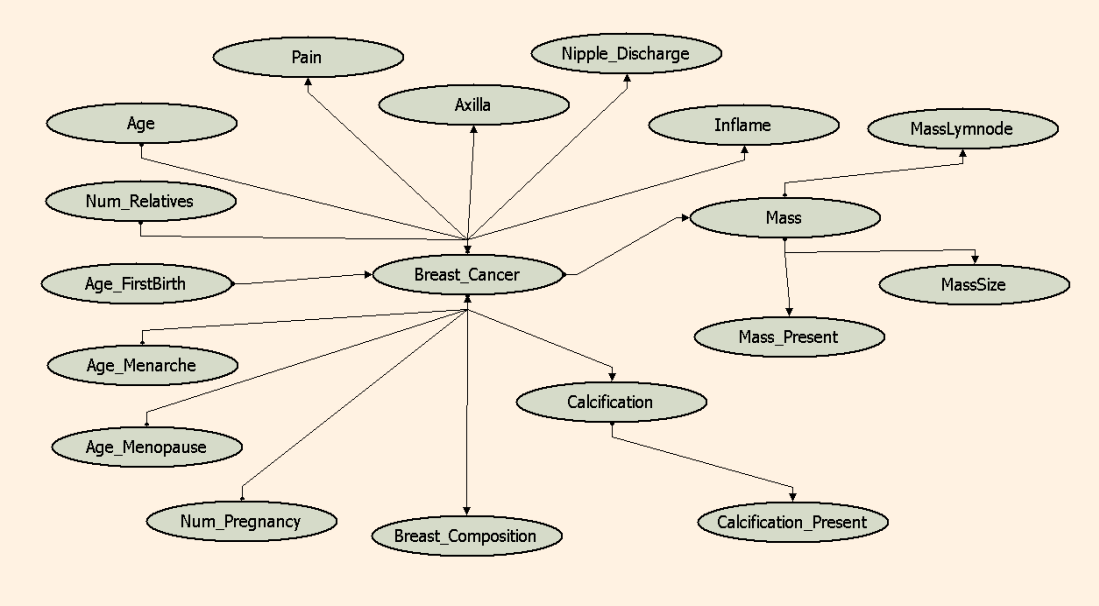
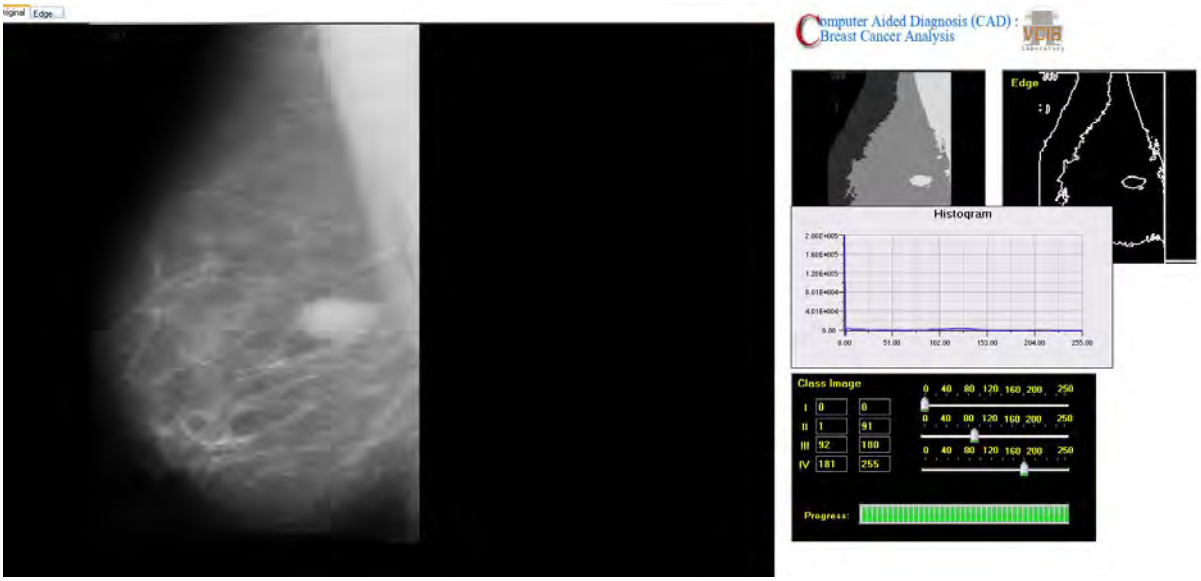


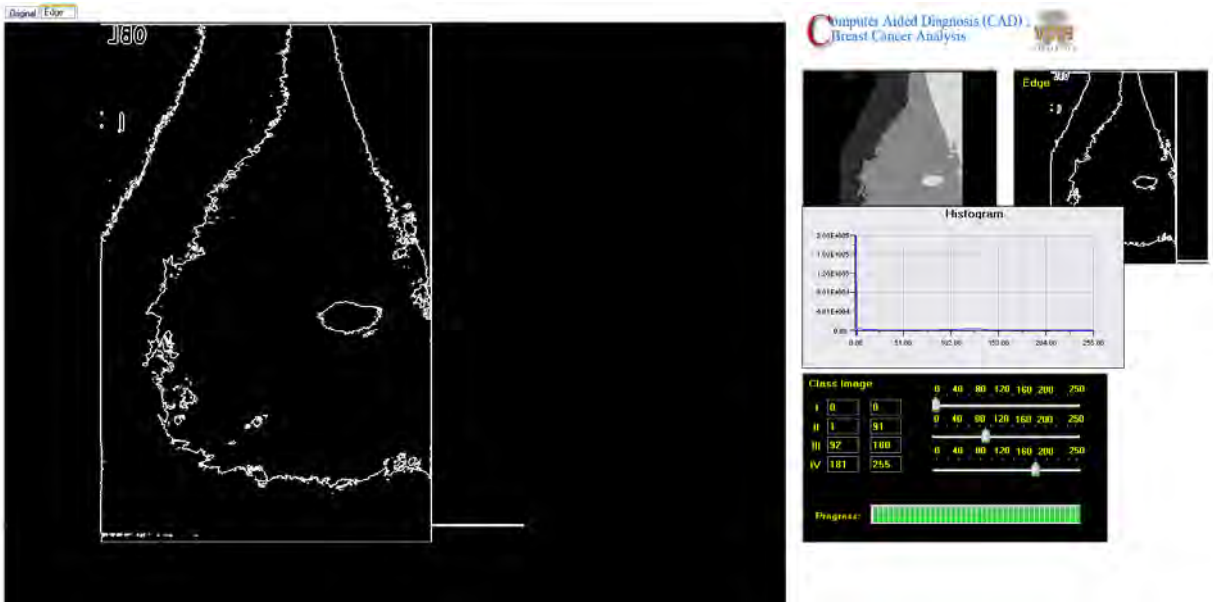
Fig. 1. Proposed BBN Model.

A. Direct Mammographic Findings

Mammographic image processing was carried out using cellular automata model (CA) [2],[4] to determine mass, calcification, and their features. **Figure 2** given below shows a process of mass detection in Computer Aided Diagnosis (CAD) software due to the research project.



(a) Before processing.



(b) After processing.

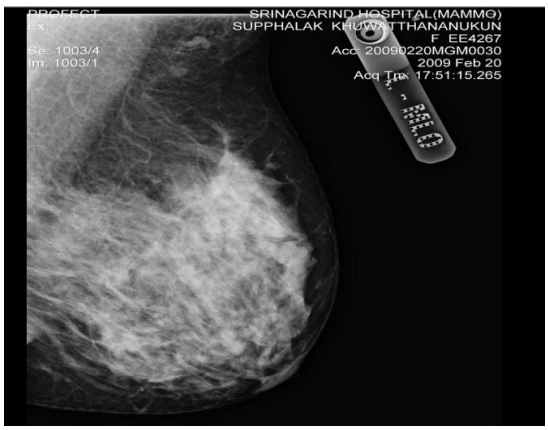
Fig. 2. Mammogram Image Processing Using Cellular Automata Model.

B. Data Acquisition and Inference Software

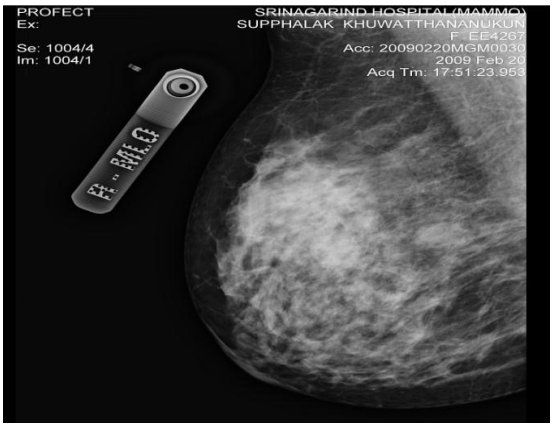
The proposed BBN's knowledge base was constructed from medical literature, health statistics reports and 450 patient's records collected at Srinakarind Hospital, Khon Kaen University. An example of patient's data was shown in Table 2. Figure 3 shows mammograms of a benign patient.

Table 2 An example of partial dataset utilized for training the BBN model

1 2	HN	Age	No. of Relative	First-Live Birth	Age-Menarch	Age-Menopause	No. of Pregnancy	Pain	Nipple Discharge	Axilla	Inframe	Tumor
3	AZ6182	50	0	29	12	43	2	0	0	0	0	2
4	GX3468	48	0	34	14	48	2	0	0	0	0	2
5	ER3619	55	0	24	13	50	1	0	0	0	0	2
6	HF3701	56	0	30	17	49	4	0	0	0	0	2
7	DL4407	46	0	0	12	37	0	0	0	0	0	2
8	GL1813	30	0	20	17	0	1	0	0	0	0	2
9	FP8570	54	0	36	15	47	2	0	0	0	0	2
10	GZ7630	55	3	20	15	40	4	0	0	0	0	2
11	EW7370	52	0	25	12	42	4	2	0	0	0	2
12	EH1260	41	0	23	14	40	3	0	0	0	0	2
13	HE5501	46	0	22	16	0	1	0	0	0	0	2
14	GU4996	57	3	37	16	48	3	0	0	0	0	2
15	FR1732	56	0	23	17	50	2	0	0	0	0	2
16	GM8244	64	0	24	14	55	4	0	0	0	0	2
17	GX6578	55	0	35	17	54	2	0	0	0	0	2
18	DG3696	64	0	18	15	54	3	0	0	0	0	2
19	HF1133	54	0	19	17	48	4	0	0	0	0	2
20	FX5857	74	0	19	12	50	3	2	0	0	0	2
21	HC4192	65	0	0	15	45	0	1	0	0	0	2
22	FT8211	61	2	29	13	52	4	0	0	0	0	2
23	FW4894	45	0	20	13	0	2	0	0	0	0	2
24	EP5357	35	2	0	13	0	0	0	0	0	0	1
25	DW9048	48	0	34	11	47	2	0	0	0	0	1
26	HD4666	39	1	0	14	0	2	1	0	1	0	1
27	GO9307	61	0	32	14	45	4	0	0	2	0	1



Left

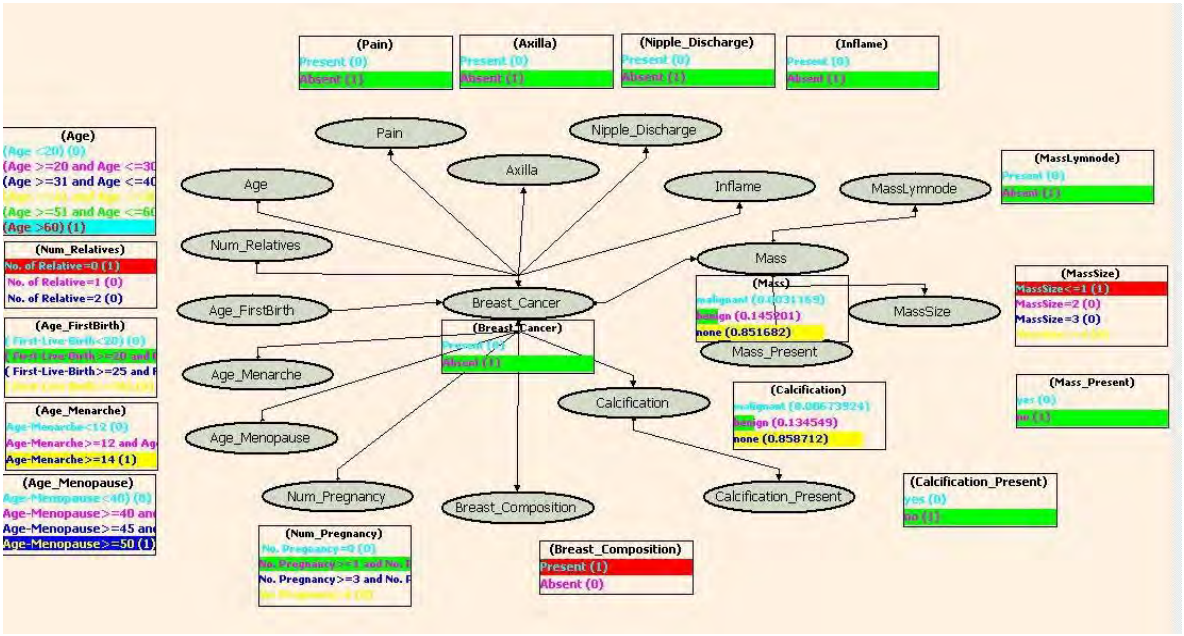


Right

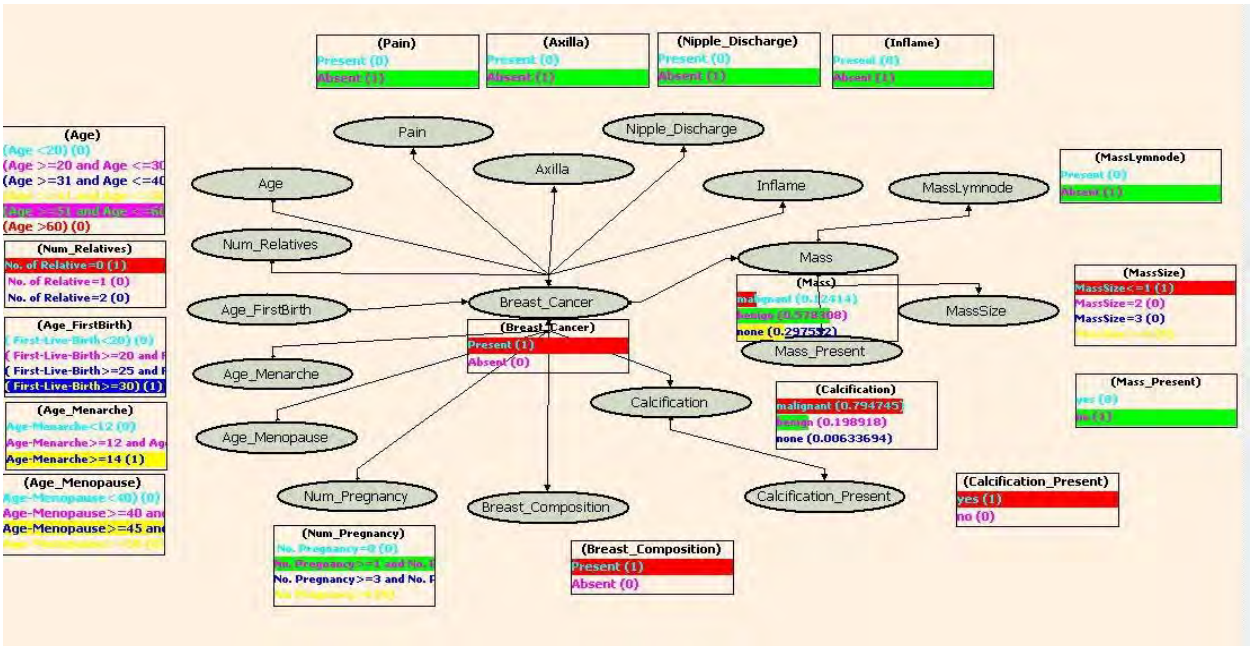
Fig. 3. Mammograms for a benign patient.

C. Modeling Software and Experimental Results

Microsoft Bayesian Networks (MSBNx) [8] was utilized as a software tool to model the breast cancer diagnosis problem. For performance evaluation, 100 patients comprised of 50 normal, 25 benign, and 25 malignant were tested on the proposed model. In this regard, it provides the promising results by reporting the percentage of 96.5 of accuracy in the diagnoses. **Figure 5** shows the results obtained from the system as implemented in three cases of patients.



a) Benign patient



b) Malignant patient

Fig. 4. An example of BBN's results for diagnosing the test patients: a) benign, b) malignant.

7.2 Results

A proposed Bayesian Belief Network (BBN) was investigated in modeling the medical breast cancer diagnosis by taking four types of datasets, namely, historic biodata, physical findings, indirect and direct mammographic findings into consideration. The structure of BBN was obtained by a suggestion of domain experts. 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 as mentioned in the previous sections. 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 regard, a dataset consisting of 500 cases is used for the performance evaluation of the model. In this respect, an 80% of data was used for training the model, while the rest of 20% was utilized for testing. The trained BBN model is tested on 100 patients consisting of 50, 25 and 25 for normal, benign and malignant patients, respectively. In addition, 5-fold and 10-fold cross-validation are implemented, the proposed BBN also reports the promising results. It provides 96.2 and 97.4 percentages of accuracy, respectively.

8. Conclusions and Discussion

Bayesian networks represent a promising technique for clinical decision support and provide a number of powerful capabilities for representing uncertain knowledge. They provide a flexible representation that allows one to specify dependence and independence of variables in a natural way through the network topology. Because Bayesian networks represent uncertainty using standard probability, one can collect the necessary data for the domain model by drawing directly on published statistical studies or elicited from the experts. The proposed BBN model is competitive to radiologists with varying levels of mammographic expertise. In addition, we are considering the addition of variables to improve the model performance, such as demographic features such as race and geographic location, and patient-history features such as diet, body habitus, history of hormone therapy, and previous cancers.

Acknowledgments

We thank The Thailand Research Fund (TRF) and The Office of Higher Education Commission for the financial support of this research project through the RMU5080010 contact.

9. References

[1] S. Wongthanavasu and C. Lursinap, "A 3-D CA-based Edge Operator for 3-D Images," The Proceeding of The 11th IEEE Int. Conference on Image Processing (IEEE-ICIP 2004), ISBN: 0-7803-8555-1, IEEE Catalog Number: 04CH37580C, pp. 235-238, 2004.

[2] S. Wongthanavasu and V. Tangvoraphonkchai, "Cellular Automata-based Algorithm and Its Application in Medical Images Processing, " The Proceeding of the 14th IEEE Int. Conference on Image Processing (IEEE-ICIP 2007), ISBN: 1-4244-1437-7, ISSN: 1522-4880, IEEE Catalog No.: 07CH37925C, pp. III-41 - 44, 2007.

[3] S. Wongthanavasu and R. Sadananda, "Pixel-level Edge Detection Using a Cellular Automata-Based Model," in Proc. Int. Conf. on Advances in Intelligent Systems: Theory and Applications. (59), IOS Press. The Netherlands, pp. 343-351, 2000.

[4] S. Wongthanavasu and V. Tanvoraphonkchai, "Cellular Automata-based Identification of The Pectoral Muscle in Mammograms," in Proc. 3rd Int. Conf. on Biomedical Engineering (ISBME 2008), November 11-12, 2008, Grand Mercure Fortune, Bangkok, Thailand, pp. 294-298, 2008.

[5] C.L. Huang, H.C. Liao, M.C. Chen, "Predication Model Building and Feature Selection with Support Vector Machines in Breast Cancer Diagnosis," Expert Systems with Applications, Elsevier, Vol. 34, pp. 578-587, 2008.

[6] H.D.Cheng, J. Shan, W. Ju, Y. Guo, L. Zhang, "Automated Breast Cancer Detection and Classification Using Ultrasound Images: A Survey," Pattern Recognition, Elsevier, vol. 43, pp. 299-317, 2010.

[7] S.M.Maskery, H. Hu, J. Hooke, C.D.Shriver, M.N. Liebman, "A Bayesian Derived Network of Breast Pathology Co-occurrence," Journal of Biomedical Informatics, Elsevier, vol. 41, pp. 242-250, 2008.

[8] John Locked, Microsoft Bayesian Network: Basics of Knowledge Engineering, Microsoft Technical Report, 1999.

[9] G. Hernandez and J.J. Herrmann, "Cellular Automata for Elementary Image Enhancement," Graphical Models and Image Processing (GMIP), Vol. 4, No. 58, pp. 82-89, 1996.

[10] P. Rosin, "Training Cellular Automata for Image Processing," IEEE Transactions on Image Processing, Vol. 15, No. 7, pp. 2076-2087, 2006.

[11] H.D. Cheng, X.J.Shi, R.Min, L.M. Hu, X.P.Cai, H.N.Du, "Approaches for Automated Dectection and Classification of Masses in Mammograms," Pattern Recognition, Vol. 39, pp. 646-668, 2006.

[12] B.Viher, A.Dobnikar, D.Zazula, “Cellular Automata and Follicle Recognition Problem and Possibilities of Using Cellular Automata for Image Recognition Purposes,” International Journal of Medical Informatics, Vol. 49, pp. 231-241, 1998.
[13] R.J. Ferrari, R.M Rangayyan, J.E.L. Desautels, R.A. Borges, A.F. Frere, “Automatic Identification of the Pectoral Muscle in Mammograms," IEEE Transactions on Medical Imaging, vol. 23, no. 2, pp. 232-245, 2004.
[14] N. Otsu, “A Threshold Selection Method from Gray-level Histograms”, IEEE Transactions on Systems Man Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.
[15] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, J. Savage, The mammographic images analysis society digital mammogram database, Exerpta Medica, International Congress Series 1069, pp. 375-378, 1994.
[16] F. Ma, J. Slavotinek, M.J. Bottema, “Two graph theory based methods for identifying the pectoral muscle in mammograms,” Pattern Recognition, vol. 40, pp. 2598-2598, 2007.

2. ผลงานของโครงการ

[1] S. Wongthanavasu and V. Tangvoraphongchai, "CA-based Algorithms and Its Application in Medical Image Processing," The Proceedings of the 14th IEEE Int. Conference on Image Processing (IEEE-ICIP 2007), IEEE Catalog No.: 07CH37925C, ISBN: 1-4244-1437-7, ISSN: 1522-4880, pp. III-41-III-44, 2007.

[2] S. Wongthanavasu and V. Tangvoraphongchai, "Cellular Automata-based Identification of The Pectoral Muscle in Mammograms," The Proceedings of the 3rd Int. Symposium on Biomedical Engineering (ISBME 2008), November 10-11, 2008, Bangkok, Thailand, pp. 294-298, 2008.

[3] S. Wongthanavasu, "Cellular Automata for Medical Image Processing," Book Chapter of the Book entitled "Cellular Autoamata – Innovative Modelling for Science and Engineering," ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

[4] S. Wongthanavasu, "A Bayesian Belief Network Model for Breast Cancer Diagnosis," Selected Papers of the 2010 Annual International Conference of German Operations Research Society. ISSN 0721-5924, ISBN 978-3-642-20008-3, e-ISBN 978-3-642-20009-0, Springer-Verlag Berlin Heidelberg, pp. 3-8, 2011.

2.1 ผลงานตามกิจกรรม

- Cellular Automata-based Image Processing
- Cellular Automata-based Mass Segmentation

[1] S. Wongthanavasu and V. Tangvoraphongchai, "CA-based Algorithms and Its Application in Medical Image Processing," The Proceedings of the 14th IEEE Int. Conference on Image Processing (IEEE-ICIP 2007), IEEE Catalog No.: 07CH37925C, ISBN: 1-4244-1437-7, ISSN: 1522-4880, pp. III-41-III-44, 2007.

[2] S. Wongthanavasu and V. Tangvoraphongchai, "Cellular Automata-based Identification of The Pectoral Muscle in Mammograms," The Proceedings of the 3rd Int. Symposium on Biomedical Engineering (ISBME 2008), November 10-11, 2008, Bangkok, Thailand, pp. 294-298, 2008.

[3] S. Wongthanavasu, "Cellular Automata for Medical Image Processing," Book Chapter of the Book entitled "Cellular Autoamata – Innovative Modelling for Science and Engineering," ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

2.2 ผลงานในกิจกรรม

- Cellular Automata-based Pectoral Muscle Removal

[1] S. Wongthanavasu and V. Tangvoraphongchai, "Cellular Automata-based Identification of The Pectoral Muscle in Mammograms," The Proceedings of the 3rd Int. Symposium on

Biomedical Engineering (ISBME 2008), November 10-11, 2008, Bangkok, Thailand, pp. 294-298, 2008.

[2] S. Wongthanavasu, "Cellular Automata for Medical Image Processing," Book Chapter of the Book entitled "Cellular Autoamata – Innovative Modelling for Science and Engineering," ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

2.3 ผลงานในกิจกรรม

- Breast Cancer Diagnosis Using Bayesian Belief Network

[1] S. Wongthanavasu, "Cellular Automata for Medical Image Processing," Book Chapter of the Book entitled "Cellular Autoamata – Innovative Modelling for Science and Engineering," ISBN 978-953-307-172-5, INTECH publisher, pp. 395-410, 2011.

[2] S. Wongthanavasu, "A Bayesian Belief Network Model for Breast Cancer Diagnosis," Selected Papers of the 2010 Annual International Conference of German Operations Research Society. ISSN 0721-5924, ISBN 978-3-642-20008-3, e-ISBN 978-3-642-20009-0, Springer-Verlag Berlin Heidelberg, pp. 3-8, 2011.

3. ความเห็นของผู้วิจัย

สำหรับการวิจัยในโครงการนี้ได้รับการสนับสนุนทุนจากโครงการเพิ่มขีดความสามารถด้านการวิจัยของอาจารย์รุ่นกลางในสถาบันอุดมศึกษา เป็นงานวิจัยที่นำเสนอต้นแบบการวิจัย Computer Aided Medical Diagnosis โดยใช้เทคนิค Artificial Intelligence และ Intelligent Systems สำหรับโรคมะเร็งเต้านม และสามารถนำไปเป็นต้นแบบในการพัฒนาระบบการวินิจฉัยโรค/กลุ่มอาการของโรคอื่นๆ ได้อย่างมีประสิทธิภาพ สำหรับความล่าช้าของโครงการส่วนหนึ่งมาจากการเก็บรวบรวมข้อมูลซึ่งมีจำนวน case ของผู่ป่วยที่เป็นมะเร็งเต้านมน้อยมากไม่เพียงพอต่อการวิเคราะห์ ต้องขยายเวลาโครงการออกไปเนื่องจาก ข้อมูลที่ปรากฏในระบบไม่มีความสมบูรณ์และไม่สามารถนำมาใช้กับระบบที่พัฒนาได้ ทำให้ต้องเก็บรวบรวมแบบ progressive สำหรับในการพัฒนาระบบ Computer Aided Medical Diagnosis นักวิจัยต้องคำนึงถึงปัจจัยดังกล่าวด้วย

4. เอกสารประกอบ

2007 IEEE International Conference on Image Processing

September 16-19, 2007 • San Antonio, Texas, U.S.A.

IEEE Catalog No.: 07CH37925C ISBN: 1-4244-1437-7 ISSN: 1522-4880

IEEE

IEEE Signal Processing Society

ICIP 2007

General Chair's Message

The ICIP 2007 Organizing Committee is excited to welcome this year's attendees to the 14th IEEE International Conference on Image Processing in San Antonio, Texas. ICIP 2007 is being held at the Hyatt Regency San Antonio, which is in the heart of the city and within easy walking distance to numerous sights. Just step outside the hotel and you will find San Antonio's famous River Walk. The 2.5 mile historic river district is lined on each side with cobblestone walkways, crossover bridges, restaurants, shops, and nightly entertainment, making it a must-see part of your stay.

San Antonio is probably best known for The Alamo. The Alamo was the site where defenders battled against the Republic of Mexico in a turning point of the Texas Revolution. To make sure that you do not miss this famous location, the ICIP 2007 Welcome Reception is scheduled to be held at Alamo Plaza, complete with food, beverages and entertainment, as well as tours of the Alamo. Whether you merely want to catch up with colleagues or soak in the historic past, this should be a fun event for everyone.

During your visit to San Antonio, you will soon learn that The Alamo is but one of the intriguing aspects of this city. You will be able to get a birds-eye view of the rest of San Antonio at the Conference Banquet. The Conference Banquet will be held at the 750-foot-tall Tower of the Americas, in the revolving restaurant that provides breathtaking, 360-degree views of the city. This tower was originally built in honor of the 1968 World's Fair. If you plan to attend the banquet, please be sure to register early because seating is limited.

You will also find that San Antonio offers numerous other interesting sights. For example, San Antonio has many cultural and historical museums where you can learn about the colorful history of Texas and its diverse people. If your trip to San Antonio includes a weekend, consider visiting the outstanding theme parks. Sea World, the world's largest marine life adventure park, offers a splashy line-up of more than 25 sensational shows, thrilling rides, animal attractions and educational experiences for all ages. Six Flags Fiesta Texas is another exciting theme park with shows, thrilling rides and games for all ages.

The focus of ICIP 2007 will be the technical program, which will feature three plenary sessions, six special sessions on timely and important topics, eight tutorials by experts in the field, and lecture presentations and poster sessions covering the gamut of image processing research. Topics will include image/video coding and transmission; image/video processing; image formation; image scanning, display, and printing; image/video storage, retrieval and authentication; and other applications.

I would like to express my deepest appreciation to the entire Organizing Committee and

Conference Management Services, whose tireless efforts have been essential in making ICIP'07 a reality. Finally, I would like to acknowledge the contributions of the Technical Program Committee members, the reviewers, the session chairs, the student volunteers and the corporate supporters, all of whom are essential elements of this large, team effort.

Jeff Rodriguez, General Chair

ICIP 2007

Technical Program Committee

John Apostolopoulos, HP Labs (Technical Program Co-Chair)
Bob Safranek, Benevue, Inc. (Technical Program Co-Chair)

Scott Acton, University of Virginia
Yucel Altunbasak, Georgia Institute of Technology
Ricardo De Queiroz, Universidade de Brasilia
Minh Do, University of Illinois at Urbana-Champaign
Jean-Luc Dugelay, Eurécom
James Fowler, Mississippi State University
Pascal Frossard, EPFL
Sheila Hemami, Cornell University
Jianying Hu, IBM T. J. Watson Research Center
Lina Karam, Arizona State University
William Karl, Boston University
Alex Kot, Nanyang Technological University
Benoit Macq, Université catholique de Louvain
B. S. Manjunath, University of California, Santa Barbara
Aleksandra Mojsilovic, IBM Research
Truong Nguyen, University of California, San Diego
Antonio Ortega, University of Southern California
Thrasyvoulos Pappas, Northwestern University
Fernando Pereira, IST-IT
Amy Reibman, AT&T Labs - Research
Kenneth Rose, University of California, Santa Barbara
Eli Saber, Rochester Institute of Technology
Amir Said, Hewlett-Packard Laboratories
Dan Schonfeld, University of Illinois at Chicago
Gaurav Sharma, University of Rochester
Eckehard Steinbach, Munich University of Technology
Qibin Sun, Institute for Infocomm Research, Singapore
H. Joel Trussell, North Carolina State University
Mihaela van der Schaar, University of California, Los Angeles
Min Wu, University of Maryland, College Park
Yehoshua Y. (Josh) Zeevi, Technion - Israel Institute of Technology

Reviewers

Til Aach, RWTH Aachen University
G. C. K. Abhayaratne, University of Sheffield

ICIP 2007

Technical Program Overview

It is our pleasure to welcome you to the 2007 IEEE International Conference on Image Processing in San Antonio, Texas, USA! We hope that you will find the conference both exciting and stimulating.

This year, we received 1708 properly completed paper submissions from authors from 58 countries. This is an increase of 7% over the 1596 submissions in ICIP 2006. Providing a thorough and fair review process for ICIP was made possible by the over 700 expert reviewers who generously provided over 5000 reviews. We also had a technical program committee of 33 people who personally assigned reviewers for each of the 1708 submissions based on their technical content, and also helped assess and recommend decisions for boundary papers, as well as performed their own reviews. There were many experts who generously provided their valuable feedback by reviewing up to 20 papers, and this included many current and past members of the IEEE Image and Multidimensional Signal Processing Technical Committee. On behalf of the entire ICIP community, we would like to thank all of the above whose valuable contributions have made ICIP 2007 possible!

Based on the feedback provided by the reviewers and technical program committee members, we accepted 843 papers that were organized into 36 lecture and 48 poster sessions, covering the whole gamut of image processing research. Similar to prior ICIP conferences, the most popular sessions, in terms of numbers of papers submitted, were in order: image and video segmentation, image and video coding, biomedical imaging, security, restoration and enhancement, stereoscopic and 3-D processing, and image & video storage and retrieval.

The ICIP 2007 technical program features three plenary lectures examining the timely and exciting research areas of compressive sensing, face recognition, and digital cinema. The three plenary lectures are:

- **Compressive Sensing**
  Presenter: Emmanuel J. Candes, California Institute of Technology
- **Survey of Automatic Face Recognition**
  Presenter: P. Jonathon Phillips, National Institute of Standards & Technology
- **An Overview of Digital Cinema**
  Presenter: Michael W. Marcellin, University of Arizona

The conference begins on Sunday, September 16, with eight tutorials organized by the tutorials chair Prof. Eli Saber. These tutorials were selected from among 18 proposals submitted in response to the Call for Tutorials by a team of experts under the guidance of

Prof. Saber. Attendees can get overviews of the state of the art in several key areas of image and video processing through the following tutorials:

- **Distributed Video Coding for Low Cost Video Encoding**
  Presenter: Anil Fernando, University of Surrey, UK
- **Moments and Moment Invariants in Image Analysis**
  Presenters: Jan Flusser, Barbara Zitova, and Tomas Suk, Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic
- **Perceptual Metrics for Image Quality Evaluation**
  Presenters: Thrasyvoulos Pappas, Northwestern University, and Sheila Hemami, Cornell University
- **Game Theoretic Approaches for Multi-user Multimedia Resource Allocation in Emerging Cognitive Radio Networks**
  Presenters: Mihaela van der Schaar, University of California, Los Angeles, and Sai Shankar N, Qualcomm Incorporated
- **Scalable Video Coding and Networking**
  Presenters: Jens-Rainer Ohm, RWTH Aachen University, Germany, and Thomas Wiegand, Fraunhofer HHI, Germany
- **Digital Image Forensics**
  Presenter: Yun Shi, New Jersey Institute of Technology
- **Image Processing using FPGAs**
  Presenter: Donald Bailey, Massey University, New Zealand
- **Content-Based Image and Video Retrieval**
  Presenters: Theo Gevers, Nicu Sebe, and Arnold Smeulders, University of Amsterdam

In addition to the tutorials, plenary addresses, and regular sessions, ICIP 2007 features six special sessions covering a broad spectrum of timely topics. We have selected the topics of these sessions by reviewing the 11 proposals submitted in response to the Call for Special Sessions. We would like to thank Prof. Scott Acton for his efforts as the special sessions chair. All special session submissions underwent a similar review process as regular submissions, and each received from three to six reviews. This year's special sessions and their organizers are:

- **Video Coding for Next Generation Displays**
  Organizer: Andrew Segall, Sharp Laboratories of America
- **Distributed Source Coding I: Low Complexity Video Coding**
  Organizers: Joern Ostermann. Leibniz Universität Hannover, Germany, and Da-ke He, IBM Research, USA, and Ashish Jagmohan, IBM Research, USA
- **Distributed Source Coding II: Distributed Image and Video Coding and Their Applications**
  Organizers: Da-ke He, IBM Research, USA, and Joern Ostermann, Leibniz Universität Hannover, Germany, and Ashish Jagmohan, IBM Research, USA
- **Challenges in Restoration for Media Production**
  Organizers: Anil Kokaram, Trinity College Dublin, Ireland, and Theodore Vlachos, University of Surrey, UK, and Bernard Besserer, University of La Rochelle, France, and Georg Thallinger, Joanneum Research, Austria

- **Image Processing and Analysis for Oncology**
  Organizer: Jinshan Tang, Alcorn State University, USA
- **Soft Computing in Image Processing: Recent Advances**
  Organizers: Mike Nachtegael, Ghent University, Belgium, and Etienne Kerre, Ghent University, Belgium, and Gerald Schaefer, Aston University, UK

ICIP 2007 includes several paper awards. For the IBM Student Paper Awards, the ICIP committee nominated ten papers from among the student paper submissions based on top review scores, and reviewer feedback on such topics as originality, potential impact, and presentation quality. The IBM award committee then selected the four winners. For the DoCoMo USA Labs Innovative Paper award, the ICIP committee nominated a second set of twenty-five best scoring papers, based on top review scores, and reviewer feedback on such topics as originality, potential impact, and presentation quality. An award committee coordinated by DoCoMo USA Labs then selected the two winners.

We would like to extend our deepest appreciation to the plenary speakers, tutorial presenters, special sessions chairs and presenters, and the ICIP authors and ICIP review committee for their contributions to ICIP 2007. We would also like to thank Conference Management Services, Inc. (CMS), and in particular Ms. Billene Mercer, for their expert guidance in organizing a conference of this size. We would also like to give a special thanks to Mr. Lance Cotton of CMS, whose tireless and friendly day-to-day help over the past year have been indispensable for the success of ICIP.

We would like to acknowledge and express our thanks to Hewlett-Packard for their support of ICIP, and to IBM and DoCoMo USA Labs for their support of the ICIP paper awards.
Finally, we offer our sincerest appreciation to the thousands of researchers all over the world for their excellent contributions, which resulted in an outstanding technical program. All of us on the program committee look forward to welcoming you to San Antonio, and offering you a rewarding, exciting, and memorable experience.

John Apostolopoulos and Robert Safranek
ICIP 2007 Technical Program Co-Chairs

Cellular Automata-Based Algorithm and its Application in Medical Image Processing

2 Author(s)

Wongthanavasu, S. ; Khon Kaen Univ., Khon Kaen ; Tangvoraphonkchai, V.

CELLULAR AUTOMATA-BASED ALGORITHM AND ITS APPLICATION IN MEDICAL IMAGE PROCESSING

S. Wongthanavasu

Department of Computer Science
Faculty of Science
Khon Kaen University
Khon Kaen 40002, Thailand
wongsar@kku.ac.th

V. Tangvoraphonkchai

Department of Radiology
Faculty of Medicine
Khon Kaen University
Khon Kaen 40002, Thailand
vorachai@kku.ac.th

ABSTRACT

This paper presents cellular automata algorithms for medical image processing. Mammogram images are comprehensively carried out to determine the hypothesis spots of breast cancer. In this respect, the main cellular automata algorithm and its variation are presented and studied to deal with binary and grayscale images. The results of the proposed algorithms are promising and helpful for physician and doctors in diagnosis of the breast cancer in further steps.

Index Terms— Cellular automata, mammogram image processing

1. INTRODUCTION

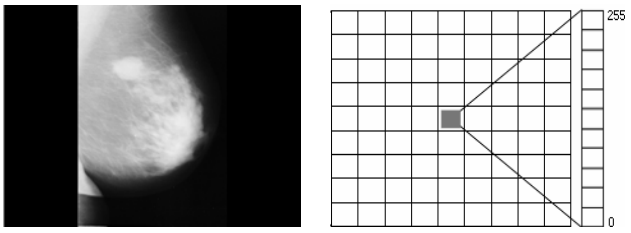
Cellular automata (CA) are a discrete spatiotemporal system whose behavior is specified in terms of local interactions. They appear as natural tools for image processing due to their local nature and simple parallel computation implementation. In this respect, there are a number of papers which generally discuss cellular automata for image processing. Hernandez et al. [1] presented CA for elementary 2-D image enhancement. Wongthanavasu et al. [2] presented 3-D CA for edge detection on binary and grayscale images and compared its performance evaluation to well-known edge operators. Rosin [3] presented algorithms for training cellular automata for image processing. Besides these, there are some papers discussed medical image processing. Cheng et al. [4] presented methods for mass detection and classification. Viher et al. [5] presented cellular automata algorithms for follicle image recognition.

This paper presents a cellular automata-based algorithm and its variation for white spot detection in mammogram. It starts by introducing cellular automata fundamentals necessary for understanding the proposed algorithms. Then, applications in noise removal, edge detection, and white spot detection in mammogram image for the breast cancer diagnosis are presented.

2. CELLULAR AUTOMATA

Let I denote the set of integer. A 2-D cellular space is a 4-tuple, (IxI, V, N, f), where IxI is a set of cartesian product of two integer sets, V is a set of cellular states, N is the type of neighborhood, and f is the local transition function from V^n into V. The relevant neighborhood function is a function from IxI into 2^{IxI} defined by $g(\alpha) = \{\alpha+\delta_1, \alpha+\delta_2,\ldots, \alpha+\delta_n\}$, for all $\alpha \in IxI$, where δ_i $(i = 1,2,\ldots,n) \in IxI$ is fixed. The neighborhood state function of a cell α at time t is defined by $h^t(\alpha) = (\nu^t(\alpha+\delta_1), \nu^t(\alpha+\delta_2), \ldots, \nu^t(\alpha+\delta_n))$. For 2-D von Neumann neighborhood, the neighborhood state function of the central cell (α) is defined by: $h^t(\alpha) = (\nu^t(\alpha+(0,0)), \nu^t(\alpha+(0,1)), \nu^t(\alpha+(1,0)), \nu^t(\alpha+(0,-1)), \nu^t(\alpha+(-1,0))$, where $\nu^t(\alpha+(0,0)$ is current state of the central cell, $\nu^t(\alpha+(0,1)$ $(\nu^t(\alpha+(0,-1))$ for the north (south) cells, $\nu^t(\alpha+(1,0)$ $(\nu^t(\alpha+(-1,0))$ for the east (west) cells.

Now we relate the neighborhood state of a cell α at time t to the cellular state of that cell at time $t+1$ by $f(h^t(\alpha)) = \nu^{t+1}(\alpha)$. The function f is referred to as the 2-D CA rule and is usually given in the form of a state table, specifying all possible pairs of the form $(h^t(\alpha), \nu^{t+1}(\alpha))$. Figure 1 shows 2-D digital image and 2-D CA.



(a) 2-D digital image. (b) 2-D cellular automata.

Figure 1. 2-D digital image vs. 2-D CA.

3. CA-BASED ALGORITHMS

As stated previously, cellular automata techniques appear as a natural tool for image processing due to their local nature and simple parallel computing implementation. In this section, we present one main algorithm and investigate its variation as methods for processing mammogram images. The methods will correspond to edge detection and noise removal for both binary and grayscale images, while the last one will correspond to spots detection for breast cancer diagnosis. Examples of the application of these cellular automata techniques to real mammogram images will be presented, which together with the results will show the performance characteristic.

The main cellular automata algorithm for k gray levels of digital images is on the basis of a bi-dimensional cellular automata (IxI, V, N, f) with $V = \{0, 1, 2, \ldots, k-1\}$, where k is a number of states, N is the type of neighborhood (e.g. n neighbors), while the local transition function f is from V^n into V. The proposed algorithm is shown in (1) following:

$$f((v^t(\alpha+\delta_1), v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_n)) = E(\alpha), \quad (1)$$
$$\text{if } \max_{j=0}^{k-1} N(C_j) = C_{target} \text{ and } sum(v^t(C_{target})) > k-1$$
$$= B(\alpha), \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values (states) in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, ..., m$ and $v^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is a number of neighbors of α which fall into class C_j.
$sum(v^t(C_{target}))$ is summation of $v^t(C_{target})$.
C_{target} is the majority class containing maximal number of neighbors.
$v^t(C_{target})$ denotes all of $v^t(\alpha+\delta_i) \in C_{target}$.
$E(\alpha)$ is the edge pixel value.
$B(\alpha)$ is the background pixel value.
k is a number of states.

For class arrangement, histogram distribution will be utilized to suggest the range of each class.

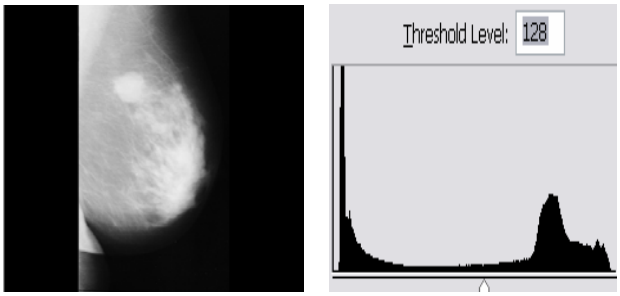
3.1. Gray level edge detection

The objective of the edge detection techniques is to enhance the magnitude of the local differences in gray level values between regions of the images. Over regions which are different, changes must be made to enhance the edges. The proposed variation of (1) which deals with this task is shown in formula (2) for Von Neumann's neighborhood, four classes and 256 gray levels as following:

$$f((v^t(\alpha+\delta_1), v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_5)) = 255, \quad (2)$$
$$\text{if } \max_{j=0}^{3} N(C_j) = C_{target} \text{ and } sum(v^t(C_{target})) > 255$$
$$= 0, \text{ otherwise}$$

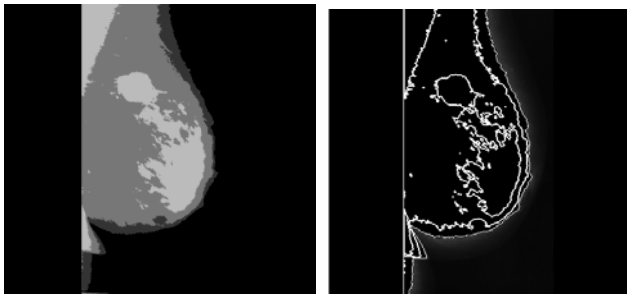
where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $v^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is number of neighbors of α which fall into class C_j.
C_{target} is the majority class containing maximal number of neighbors $(v^t(\alpha+\delta_i) \in C_{target})$.
$sum(v^t(C_{target}))$ is summation of $v^t(C_{target})$.
$v^t(C_{target})$ denotes all of $v^t(\alpha+\delta_i) \in C_{target}$.

In implementing (2) in an original mammogram image being supervised by the histogram information (Fig.2) for the class arrangement, the results were shown in Figure 3.



(a) Original mammogram. (b) Histogram.

Figure 2. Original mammogram and its histogram.



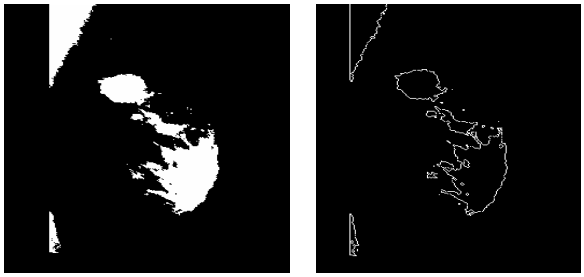
(a) Four class image. (b) Edge map result.

Figure 3. (a) Four classes of image, and (b) result of edge detection.

It is remarkable that the cellular automata algorithm simply provides the best edge maps shown in Figure 3(b).

3.2. Binary edge detection

In case of edge detection on binary image, the cellular automata algorithm in (2) is directly applied to binary image efficiently. It is no need to be changed. Figure 4 shows the promising result of the algorithm dealing with binary images. The edge result exhibits the superb quality with one pixel wide and edge has no break.



(a) Original binary image. (b) Edge result.

Figure 4. Edge result as dealing with binary image.

3.3. Noise filtering

The objective of the noise filtering is to reduce the local differences in gray level values between regions of the images. Over regions which are similar, no changes must be made, in order to avoid the destruction of the main characteristics of the image. The proposed variation of cellular automata algorithm given in (1) using von Neumann's neighborhood and two states dealing with this task is shown in formula (3) as follow:

$$f((v^t(\alpha+\delta_1),v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_5)) = \max(v^t(C_{target})), \quad \text{if } \max_{j=0}^{3} N(C_j) = C_{target} \qquad (3)$$

$$= 0, \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $v^t(\alpha+\delta_i) \in C_j$.

$N(C_j)$ is number of neighbors of α which fall into class C_j.

C_{target} is the majority class containing maximal number of neighbors $(v^t(\alpha+\delta_i) \in C_{target})$.

$v^t(C_{target})$ denotes all of $v^t(\alpha+\delta_i) \in C_{target}$.

$max(v^t(C_{target}))$ is the maximal state of $v^t(C_{target})$.

Figure 5 shows an original binary mammogram with salt and pepper noise at 2%. The noise filtering result using one iteration of implementing such an algorithm is shown in Figure 6. In this respect, the cellular automata algorithm provides the promising result.

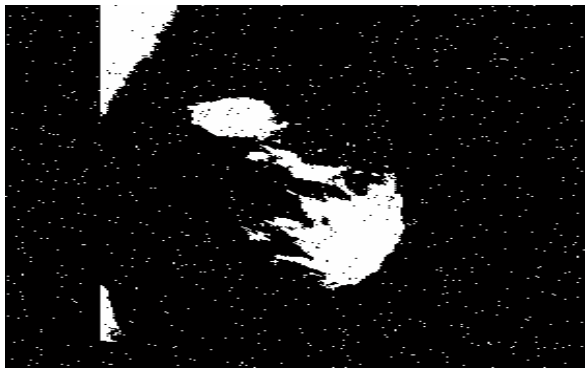


Figure 5. Binary image with salt and pepper noise (2%).

Figure 6. Binary image after noise removal.

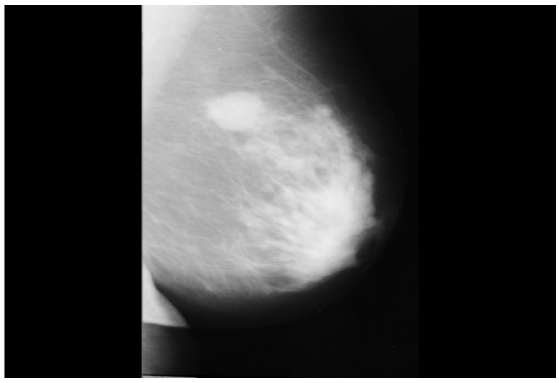
3.4. Spot detection

The objective of the spot detection is to assist the physicians and doctors in locating the hypothesis spots for breast cancer. The shape and spread region of the spots play a vital role for further steps of analysis and have to be comprehensively taken into account. In this regard, a set operator is presented to provide such an affect as following:

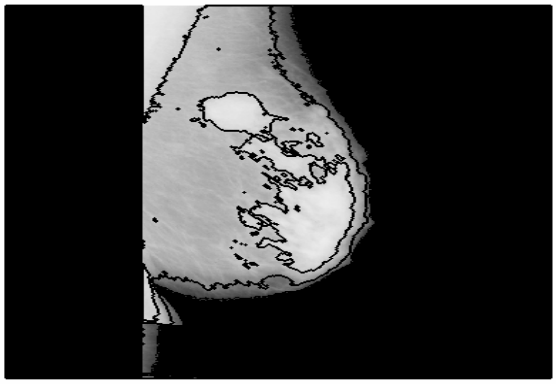
$$W = X - Y \qquad (4)$$

where X denotes an original image investigated,
Y denotes an edge map due to formula (2), and
W denotes the resulting image.

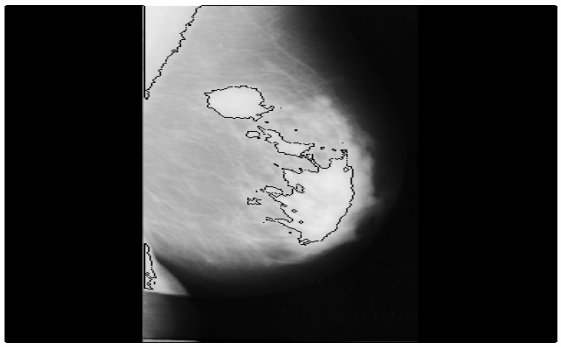
The difference (-) of two sets (images) is simply the subtraction of pixel values between two images (sets) in the same coordinate. By implementing such an operator in an original image (Fig. 7(a)) with respect to Y the results of formula (2) on grayscale and binary images, the resulting white spots detection were shown in Figure 7 (b) and (c), respectively.



(a) Malignant breast cancer.



(b) Spots detection due to (2) for grayscale.



(c) Spot detection due to (2) for binary image.

Figure 7. Mammogram image processing with white spot detection: (a) original malignant breast cancer, (b) cancer spot detection due to (4) for grayscale, and (c) due to (4) for binary image.

5. CONCLUSIONS AND DISCUSSIONS

The behavior of cellular automata is fascinating not only from a theoretical perspective but also from an experimental perspective. The uniformity of cell space offers the beauty and elegance of results. However in real life modeling non-uniformity of cells space may offer better in sights. In this work we have presented uniform cellular automata algorithms for elementary medical image processing. More specifically, cellular automata algorithms dealing with noise filtering, edge detection, and white spot detection for mammogram image in breast cancer diagnosis are presented and investigated. The results are promising, and quite encouraging in determining other tasks. In this regard, we have more investigations on application for mammogram image processing and hope report in the near future.

6. REFERENCES

[1] G. Hernandez and J.J. Herrmann, "Cellular Automata for Elementary Image Enhancement," *Graphical Models and Image Processing (GMIP)*, Vol. 4, No. 58, pp. 82-89, 1996.

[2] S. Wongthanavasu and C. Lursinsap, "A 3-D CA-based Edge Operator for 3-D Images," The Proceedings of the 11th IEEE Int. Conference on Image Processing (IEEE-ICIP 2004), IEEE press, pp. 235-238, 2004.

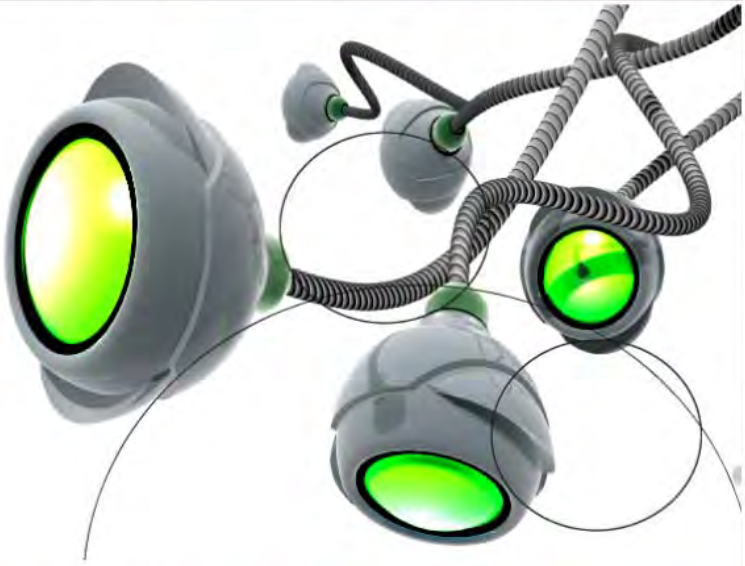
[3] P. Rosin, "Training Cellular Automata for Image Processing," *IEEE Transactions on Image Processing*, Vol. 15, No. 7, pp. 2076-2087, 2006.

[4] H.D.Cheng, X.J.Shi, R.Min, L.M. Hu, X.P.Cai, H.N.Du, "Approaches for Automated Dectection and Classification of Masses in Mammograms," *Pattern Recognition*, Vol. 39, pp. 646-668, 2006.

[5] B.Viher, A.Dobnikar, D.Zazula, "Cellular Automata and Follicle Recognition Problem and Possibilities of Using Cellular Automata for Image Recognition Purposes," *International Journal of Medical Informatics*, Vol. 49, pp. 231-241, 1998.

ISBME & BMEiCON 2008

NOVEMBER 10-11, 2008, BANGKOK, THAILAND



Honorary Conference Chairs Message

Dear Colleagues,

On behalf of the honorary conference chairs, it is our great pleasure to Bangkok and welcome you to the 3rd International Symposium on Biomedical Engineering – 2008 (ISBME-2008). This year the ISBME is organized in conjunction with the 2nd Symposium on Thai Biomedical Engineering (ThaiBME). The conference will be an excellent forum for engineers, physicians, and scientists to exchange the knowledge and discuss the advancement in biomedical engineering and its related fields.

The ISBME has been held biannually since 2004. In contrast, the ThaiBME has been initiated since last year. Both conferences have became major conferences in this field for this region of the world, providing an opportunity for researchers to get to know one another, exchange views and ideas, as well as their latest research findings, and strengthen existing friendship. Fortunately, this year we can combine both ISBME and ThaiBME into a single bigger event. We do hope that the participants will enjoy not only the technical aspects of the conference, but also the cultural and traditional of Thailand.

The “Loy Krathong”, a spectacular festival of lights, is held annually on the full moon of the 12th month of Thai lunar calendar in recognizing the goddess of the river. In addition to the conference, there are more opportunities for the speakers and participants to experience various social events.

Kitti Tirasesth

President, King Mongkut’s Institute of Technology Ladkrabang (KMITL)

Chiradet Ousawad

President, University of Thai Chamber of Commerce (UTCC)

General Chair Message

Colleagues, Distinguish Guests, Ladies and gentle men;

On behalf of the General Chair of Organizing Committee, it is my pleasure to welcome you to the 2nd Symposium on Thai Biomedical Engineering (ThaiBME 2008) which is jointly organized with the 3rd International Symposium on Biomedical Engineering – 2008 (ISBME-2008). This year we are fortunate to meet our colleagues in the beautiful city of Bangkok, Thailand. The conferences will take place during the "Loy Krathong Festival" a spectacular festival of lights annually held on the full moon of the 12th month of Thai lunar calendar. ThaiBME and ISBME promise to be a memorable event not to be missed.

The ISBME 2008 and ThaiBME 2008 feature a wide range of contributions, distributed across special sessions that represent the board range of interests in biomedical engineering. It is our hope that the conferences will provide a venue for exchange of scientific information, a meeting place for networking with colleagues, and an opportunity to present recent research.

Finally, I would like to say thanks to all honorable keynote speakers, invited speakers, authors, reviewers, session chairpersons, participants and other people for their supports, and wish all of you a successful and stimulating meeting and a pleasant stay in Bangkok, and also enjoy for Loy Krathong Festival.

Manas Sangworasil

King Mongkut's Institute of

Technology Ladkrabang (KMITL)

Thailand

President of the ThaiBME Association Message

Dear Colleagues:

On behalf of the President of the Thai Biomedical Engineering (ThaiBME) Association, it is my great pleasure and honor to welcome all the participants to the 3rd International Symposium on Biomedical Engineering, or ISBME-2008 which is held in conjunction with the 2nd Symposium on Thai Biomedical Engineering, or ThaiBME-2008. We have supported and organized these conferences as one of many activities conducted by our association. The ThaiBME association had been found since 2549 by the cooperation of engineers, scientists, medical doctors, dentists, veterinaries, pharmaceutical personnel, nurses, radiation technicians and other health personnel. The main objectives of the association are to promote the research and activities of biomedical engineering nationwide and to provide the forum for researchers and scientists to exchange and discuss matters regarding biomedical engineering and/or related fields.

The biomedical engineering has greatly expanded its roles by in integrating itself into and utilizing the knowledge and advantages of information technology. The BME conferences become more and more importance forum for researcher and scientists in such field.

I would like to express my sincere gratitude and deep appreciation to all of the conference committee members and staffs for their excellent work and to the authors from around the world for sharing their results by their presentations and discussions. I hope that the participants will enjoy not only the technical aspects of the conference, but also the banquet, food, and cultural of Thailand. Some inconveniences may occur during the arranging of the conference, but I do believe that the Organizing Committee Members will do there best to solve and to make sure that you participate in the conference with pleasure.

I hope that you will find a good opportunity to meet, to exchange ideas and to make research contacts and collaboration.

Somkiat Wattanasirichaigoon

President, ThaiBME

Technical Program Chairs Message

Colleagues, Distinguish Guests, Ladies and gentle men;

We, as a the technical program chairs of the ISBME-2008 and ThaiBME-2008, have a great pleasure to welcome all of you the odd event organized during the great time of the year, the marvelous “Loykrathong”. The ISBME this year in held in conjunction with the ThaiBME, these are good recipes for the all researchers in the area of BME, good opportunity for Thai participants and great chance for participants from abroad.

We have 130 papers submitted form researcher and scientists from 9 countries. Referees have enjoyed their hard times in reviewing the papers. Unfortunately, the conference can take 116 papers.

We would like to take this opportunity to express our sincere thanks to all the researchers who have submitted theirs good research works to be shared in the conferences. We, very much, appreciate all the referees who have spent theirs valuable times to review the manuscript. We also extend of our thanks to speakers, session chairpersons and all participants that have lively shared the session via the fruitful comment and discussion.

Due to the success of our annual national meeting “The ThaiBME”, the conference has extended its roles to an international one. From now on, it is known as “BMEiCON”. BMEiCON-2009 will be organized somewhere outside Bangkok, perhaps, Phuket. We are looking forward to have your great contributions as we do have this year, and we are looking forward to seeing you there.

Somsak Choomchuay
KMITL, Thailand

Kosin Chumnongthai
KMUTT, Thailand

Kazuhiko Hamamoto
Tokai Univ., Japan

Cellular Automata-based Identification of the Pectoral Muscle in Mammograms

CELLULAR AUTOMATA-BASED IDENTIFICATION OF THE PECTORAL MUSCLE IN MAMMOGRAMS

S. Wongthanavasu [1] and V. Tangvoraphonkchai[2]

[1] Department of Computer Science
Faculty of Science, Khon Kaen University
Khon Kaen 40002, Thailand
wongsar@kku.ac.th
[2] Department of Radiology
Faculty of Medicine, Khon Kaen University
Khon Kaen 40002, Thailand
vorachai@kku.ac.th

ABSTRACT

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms. Its inclusion can affect the results of intensity-based image processing methods. This paper presents a new method on the basis of cellular automata model for the identification of the pectoral muscle in MLO mammograms. A dataset of 84 MLO mammograms from the MIAS (Mammographic Image Analysis Society, London, U.K.) database was implemented throughout for evaluation. In this respect, the pectoral muscle edge detected in the mammograms was carried out based upon the percentage of false-positive (FP) and false-negative (FN) pixels determined by comparison between the numbers of pixels enclosed in the regions delimited by the edges identified by a radiologist and by the proposed method. The proposed CA-based method provides the promising results.

1. INTRODUCTION

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms. Its inclusion can affect the results of intensity-based image processing methods or bias procedures in the detection of breast cancer. An identification of the pectoral muscle in mammograms for the exclusion is one of preprocessing stages in computer-aided diagnosis for the breast cancer diagnosis.

Cellular automata (CA) are a discrete spatiotemporal system whose behavior is specified in terms of local interactions. They appear as natural tools for image processing due to their local nature and simple parallel computation implementation. In this respect, medical images are a popular domain of interest in present. Hernandez et al. [1] presented CA for elementary 2-D image enhancement. Wongthanavasu et al. [2] presented 3-D CA for edge detection on both binary and grayscale images and compared its performance evaluation to well-known edge operators. Rosin [3] presented algorithms for training cellular automata for image processing. Besides these, there are some papers discussed medical image processing for breast cancer. Cheng et al. [4] presented methods for mass detection and classification. Ferrari et al. [5] presented a method for identification of the pectoral muscle in mammograms.

This paper presents cellular automata-based algorithms for an identification of the pectoral muscle in mammograms. It starts by introducing cellular automata fundamentals necessary for understanding the proposed algorithms. Then, the steps of the proposed method and performance evaluation using FP (false-positive) and FN (false-negative) are comprehensively presented.

2. CELLULAR AUTOMATA

A 2-D cellular space is a 4-tuple, (IxI, V, N, f), where IxI is a set of cartesian product of two integer sets, V is a set of cellular states, N is the type of neighborhood, and f is the local transition function from V^n into V. The relevant neighborhood function is a function from IxI into 2^{IxI} defined by $g(\alpha) = \{\alpha+\delta_1, \alpha+\delta_2, \ldots, \alpha+\delta_n\}$, for all $\alpha \in IxI$, where δ_i $(i = 1,2,\ldots,n) \in IxI$ is fixed. The neighborhood state function of a cell α at time t is

defined by $h^t(\alpha) = (v^t(\alpha+\delta_1), v^t(\alpha+\delta_2), \ldots, v^t(\alpha+\delta_n))$. For 2-D von Neumann neighborhood, $h^t(\alpha) = (v^t(\alpha+(0,0)), v^t(\alpha+(0,1)), v^t(\alpha+(1,0)), v^t(\alpha+(0,-1)), v^t(\alpha+(-1,0))$, where $v^t(\alpha+(0,0)$ is current state of the central cell, $v^t(\alpha+(0,1)$ ($v^t(\alpha+(0,-1))$ for the north (south) cells, $v^t(\alpha+(1,0)$ ($v^t(\alpha+(-1,0))$ for the east (west) cells. We relate the neighborhood state of a cell α at time t to the cellular state of that cell at time $t+1$ by $f(h^t(\alpha)) = v^{t+1}(\alpha)$. The function f is referred to as the 2-D CA rule and is usually given in the form of a state table, specifying all possible pairs of the form $(h^t(\alpha), v^{t+1}(\alpha))$.

3. PROPOSED CA-BASED ALGORITHMS

In this section, the methods for identification of pectoral muscle in mammogram images were presented. The methods will correspond to edge detection for grayscale images, while the last method will correspond to image segmentation leading to the identification of the pectoral muscle. Examples of the application of the proposed method to real mammogram images will be presented, which together with the results will show the performance characteristic.

The proposed cellular automata algorithms for *256* gray levels of digital images are on the basis of a two-dimensional cellular automata *(IxI, V, N, f)* with *V = {0, 1, 2, ..., 255}*, *N* is von Neumann type of the neighborhood, while the local transition function *f* is from V^n into *V*.

3.1. Edge detection

The objective of the edge detection technique is to enhance the magnitude of the local differences in gray level values between regions of the images. Over regions which are different, changes must be made to enhance the edges. The cellular automata algorithm for this task was given in (1) as follow:

$$\text{if } ((v^t(\alpha+\delta_1), v^t(\alpha+\delta_2),\ldots,v^t(\alpha+\delta_5)) = 255, \quad (1)$$
$$\text{if } \max_{j=0}^{3} N(C_j) = C_{target} \text{ and } sum(v^t(C_{target})) > 255$$
$$= 0, \text{ otherwise}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $v^t(\alpha+\delta_i) \in C_j$.

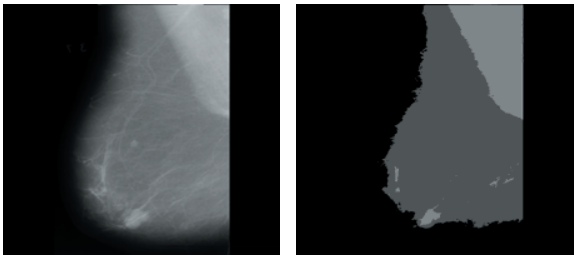
$N(C_j)$ is number of neighbors of α which fall into class C_j.

C_{target} is the majority class containing maximal number of neighbors $(v^t(\alpha+\delta_i) \in C_{target})$.

$sum(v^t(C_{target}))$ is summation of $v^t(C_{target})$.

$v^t(C_{target})$ denotes all of $v^t(\alpha+\delta_i) \in C_{target}$.

Prior to implementing the algorithm (1), an original mammogram image was classified into four classes (C_j) due to Otsu algorithm [6]. The results were shown in Figure 1.



(a) Original image mdb005. (b) Four class image.

(c) Edged image.

Fig. 1. Results obtained for the image mdb005. (a) Original image. (b) Its corresponding four-class image. (c) Edged image resulting from the implementing (1).

3.2. Pectoral muscle identification

The inclusion of pectoral muscle in mammogram can affect the results of intensity-based image processing methods or bias procedures in the detection of breast cancer. The objective of the pectoral muscle identification is to determine the pectoral muscle for the exclusion from the mammograms and being used for further steps of breast cancer diagnosis. In this respect, a number of mammogram processing steps presented earlier and the segmentation algorithm given below were used for carrying out this task.

Algorithm 1: Algorithm for removal of unqualified objects.

1. Implement the formula (1) in a mammogram image *P*, resulting in the image *Q*.
2. Filter out the objects consisting of possible concatenated white pixels (edge) in *Q* that are less than 2,500 pixels, resulting in the image R.
   { An object is defined by contiguous white pixels. }

Algorithm 2: Algorithm for identification of pectoral muscle.

1. FOR *i:= 1* TO *n* DO
   Find the area of $ROI_r(i)$.
   { $ROI_r(i)$ is a region *i* inside the image R obtained by algorithm 1 and it is defined by an area of contiguous black pixels surrounded by a closed boundary edge of white pixels. }

2. Pectoral muscle segmentation:
   2.1 MUSCLE := NIL {empty set}
   2.2 FOR *j:=1* TO *n* DO
   IF $ROI_r(j) > max\_size$ AND $ROI_r(j) \leq min\_size$
   THEN $ROI_r(j) :=$ background.
   { Set the $ROI_r(j)$ to background when it is not of interest regions. In practical uses in this paper, max_size = 95,200 pixles, and min_size = 2,500 pixles. }
   ELSE IF min (y, $ROI_r(j)$) > min_upper
   THEN $ROI_r(j) :=$ background.
   { min (y,$ROI_r(j)$) denotes minimal value of y for the coordinates (x,y) of all pixels in the region of $ROI_r(j)$. In the paper, min_upper = 60.}
   { Regions of interest locate at min(y, $ROI_r(j)$) at the topmost part of the image R. }
   ELSE IF mean (y, $ROI_r(j)$) > min_y_average
   THEN $ROI_r(j) :=$ background.
   { mean (y,$ROI_r(j)$) = average of y for the coordinates (x,y) of all pixels in the region of $ROI_r(j)$. average_y_average = 350. }
   ELSE IF mean (I, $ROI_r(j)$) > min_average_intensity
   THEN MUSCLE := MUSCLE $\cup$ $ROI_r(i)$
   { mean (I, $ROI_r(j)$) = average intensity of all pixels in the region of $ROI_r(j)$. In the paper, min_average_intensity = 157. }

The results of implementing the algorithm 1 on mdb005 and mdb011 and the results of implementing the algorithm 2 were shown in Fig. 2. Fig. 3 shows hand-drawn pectoral muscle edge by radiologist as applied to mdb005 (c), the result obtained by the proposed algorithm (b), and the pectoral muscle removal according to the proposed algorithm (d).

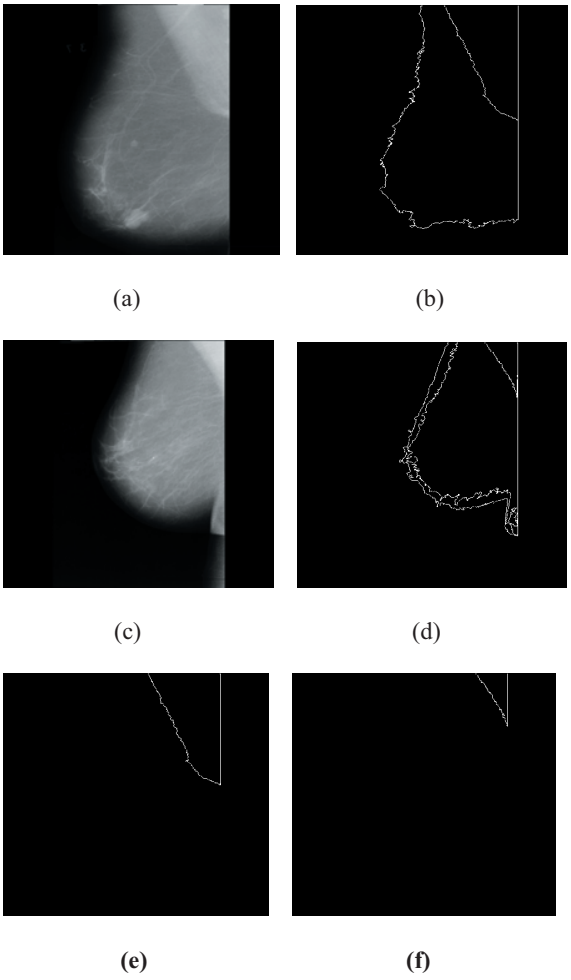
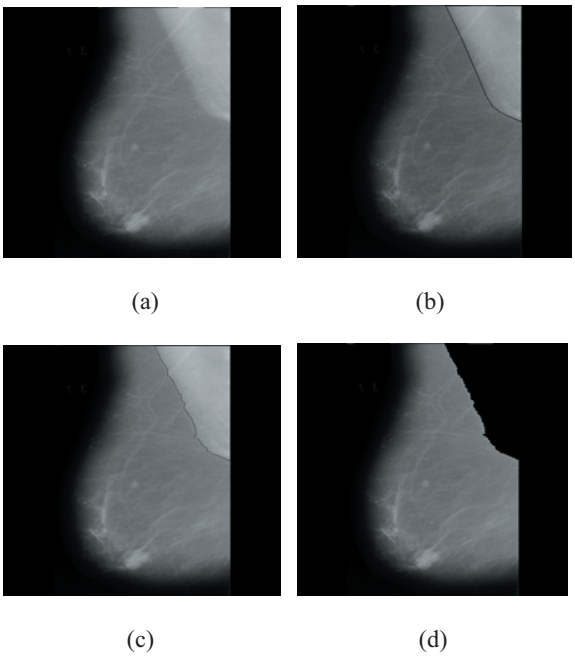


Fig. 2. Results obtained from algorithms 1 and 2. (a) and (c) are original images mdb005 and mdb011, respectively. (b) and (d) are results obtained from algorithm 1 as applied to (a) and (c), respectively. (e) and (f) are results obtained from algorithm 2 as applied to the edged images (b) and (d), respectively.



(a) (b)

(c) (d)

Fig. 3. Results obtained for the image mdb005. (a) Original image. (b) Hand-drawn pectoral muscle edge by radiologist superimposed on the original image. (c) Pectoral muscle edges detected by the CA-based method superimposed on the original image. (d) Pectoral muscle removal using the CA-based method.

4. RESULTS

The total of 84 images, randomly selected from the Mammographic Image Analysis Society, London, U.K. [7] were used in experimentation in this paper. The spatial resolution of these images is 200 μm and depth resolution in 8 bit. The images in the database are 1024x1024 pixels in size. The result obtained from the proposed method was evaluated in consultation with radiologists experienced in mammography. Then, the pectoral muscle edges were manually drawn by one of the authors under the supervision of radiologists, without referring to the results of detection by the proposed method. The segmentation results of the proposed method was evaluated based upon the number of false-positive (FP) and false-negative (FN) pixels in the regions demarcated by the manually drawn edges. An FP pixels was defined as a pixel outside the reference region that was included in the pectoral region segmented. An FN pixel was defined as a pixel in the reference region that was not present within the segmented region. Table 1 shows mean and standard deviation values of the FP and FN pixels for the result of the proposed method with 84 images.

Table 1. Mean and standard deviation values of the FP and FN pixels for the results of CA-based algorithm with 84 images.

CA-based algorithm	Statistics
Analysis of area enclosed	
FP ± σ	1.99±8.19%
FN ± σ	18.89±14.19%
# images with (FP and FN) < 5%	13
# images with 5% <(FP and FN) <10%	15
# images with (FP and FN) > 10%	56

From Table 1, it can be seen that the identification obtained by the CA-based algorithm provides the promising results with minimal FP and FN. In this regard, it is one of promising methods for being used in further steps in breast cancer diagnosis.

5. CONCLUSIONS AND DISCUSSIONS

The behavior of cellular automata is fascinating not only from a theoretical perspective but also from an experimental perspective. The uniformity of cell space offers the beauty and elegance of results. We have proposed a new algorithm for identification of the pectoral muscle in mammograms on the basis of cellular automata model. With reference to the number of pixels in manually demarcated pectoral muscle regions, the segmented regions provided by the proposed method resulted in average FP and FN rates of 1.99% and 18.89% over 84 images, respectively. According to the opinion of the radiologists involved in the study, the results are promising for application in the preprocessing stages of computer-aided diagnosis (CAD) systems.

REFERENCES

[1] G. Hernandez and J.J. Herrmann, "Cellular Automata for Elementary Image Enhancement," *Graphical Models and Image Processing (GMIP),* vol. 4, no. 58, pp. 82-89, 1996.

[2] S. Wongthanavasu and C. Lursinsap, "A 3-D CA-based Edge Operator for 3-D Images," in *Proc. IEEE Int. Conference on Image Processing (IEEE-ICIP 2004)*, ISBN: 0-7803-8555-1, IEEE Catalog Number: 04CH37580C, pp. 235-238, October 2004.

[3] P. Rosin, "Training Cellular Automata for Image Processing," *IEEE Transactions on Image Processing*, vol. 15, no. 7, pp. 2076-2087, 2006.

[4] H.D.Cheng, X.J.Shi, R.Min, L.M. Hu, X.P.Cai, and H.N.Du, "Approaches for Automated Detection and Classification of Masses in Mammograms," *Pattern Recognition,* vol. 39, pp. 646-668, 2006.

[5] R.J. Ferrari, R.M Rangayyan, J.E.L. Desautels, R.A. Borges, and A.F. Frere, "Automatic Identification of the Pectoral Muscle in Mammograms," *IEEE Transactions on Medical Imaging*, vol. 23, no. 2, pp. 232-245, 2004.

[6] N. Otsu, "A Threshold Selection Method from Gray-level Histograms", *IEEE Transactions on Systems Man Cybernetics,* vol. 9, no. 1, pp. 62-66, 1979.

[7] J. Suckling, J. Parker, D. Dance, S. Astley, I. Hutt, C. Boggis, I. Ricketts, E. Stamatakis, N. Cerneaz, S. Kok, P. Taylor, D. Betal, and J. Savage, *The mammographic images analysis society digital mammogram database*, Exerpta Medica, International Congress Series 1069, pp. 375-378, 1994.

[8] F. Ma, J. Slavotinek, M.J. Bottema, "Two graph theory based methods for identifying the pectoral muscle in mammograms," *Pattern Recognition,* Elsevier, vol. 40, pp. 2592-2602, 2007.

Published online: 11 April, 2011
Published in print edition: April, 2011

Computer and Information Science » Numerical Analysis and Scientific Computing

Cellular Automata - Innovative Modelling for Science and Engineering

Edited by Alejandro Salcido, ISBN 978-953-307-172-5, 426 pages, Publisher: InTech, Chapters published April 11, 2011 under CC BY-NC-SA 3.0 license

Modelling and simulation are disciplines of major importance for science and engineering. There is no science without models, and simulation has nowadays become a very useful tool, sometimes unavoidable, for development of both science and engineering. The main attractive feature of cellular automata is that, in spite of their conceptual simplicity which allows an easiness of implementation for computer simulation, as a detailed and complete mathematical analysis in principle, they are able to exhibit a wide variety of amazingly complex behaviour. This feature of cellular automata has attracted the researchers' attention from a wide variety of divergent fields of the exact disciplines of science and engineering, but

CELLULAR AUTOMATA - INNOVATIVE MODELLING FOR SCIENCE AND ENGINEERING

Edited by **Alejandro Salcido**

INTECHWEB.ORG

Cellular Automata - Innovative Modelling for Science and Engineering
Edited by Alejandro Salcido

Published by InTech
Janeza Trdine 9, 51000 Rijeka, Croatia

Publishing Process Manager Iva Lipovic
Technical Editor Teodora Smiljanic
Cover Designer Martina Sirotic
Image Copyright Radist, 2010. Used under license from Shutterstock.com

First published March, 2011
Printed in India

A free online edition of this book is available at www.intechopen.com
Additional hard copies can be obtained from orders@intechweb.org

Cellular Automata - Innovative Modelling for Science and Engineering,
Edited by Alejandro Salcido
p. cm.
ISBN 978-953-307-172-5

Contents

Preface

Modelling and simulation are disciplines of major importance for science and engineering. There is no science without models, and simulation has nowdays become a very useful tool, sometimes unavoidable, for development of both science and engineering. The numerical solution of differential equations has for many years been a paradigm of the computational approaches for simulation. Nevertheless, some conceptually different strategies for modelling and simulation of complex behaviour systems have been developed from the introduction of the innovative concept of cellular automata by Stanislaw Ulam and John Von Neumann in the early 1950s. Cellular automata are dynamical systems which consist of a finite-dimensional lattice, each site of which can have a finite number of states, and evolves in discrete time steps obeying a set of homogeneous local rules which define the system's dynamics. These rules are defined in such a way that the relevant laws of the phenomena of interest are fulfilled. Typically, only the nearest neighbours are involved in the updating of the lattice sites.

The main attractive feature of cellular automata is that, in spite of their conceptual simplicity which allows an easiness of implementation for computer simulation, such as a detailed and complete mathematical analysis in principle, they are able to exhibit a wide variety of amazingly complex behaviour. This feature of cellular automata has attracted the researchers' attention from a wide variety of divergent fields of the exact disciplines of science and engineering, but also of social sciences, and sometimes beyond. The collective complex behaviour of numerous systems, which emerge from the interaction of a multitude of simple individuals, is being conveniently modelled and simulated with cellular automata for very different purposes.

In this book, a number of innovative applications of cellular automata models in the fields of *Quantum Computing, Materials Science, Cryptography and Coding*, and *Robotics and Image Processing* are presented. Brief descriptions of these outstanding contributions are provided in the next paragraphs.

Quantum Computing. Chapter 1 presents an information-theoretic framework to investigate the relationship between stochastic behaviors and achievable reliable performance in quantum cellular automata technology. The central idea is that quantum cellular automata devices can be modelled as a network of unreliable information processing channels. In Chapter 2, cellular automata with graphane structured molecules and graphane nanoribbons to propagate and process digital information are proposed. The cells that make up the architecture of the automata correspond to the molecules and to sections of the nanoribbon. It is also intended to verify theoretically

that the proposed system is scalable, and binary information can be stored, propagated and processed at room temperature. Chapter 3 describes a magnetic quantum dot cellular automata approach for twisting computation and its technological implementation. The fundamental technological hypothesis (the snake-clock implementation) is explained, and an example of circuit description is given, followed by a specific architectural solution adopted with the low-level details. The contribution presented in Chapter 4 extends and implements several of the reversible and quantum computing concepts to the context of elementary cellular automata, and this includes a new method for modelling and processing via the reversibility property in the existence of noise. The main contribution is the creation of a new algorithm that can be used in noisy discrete systems modelling using conservative reversible elementary cellular automata and the corresponding quantum modelling of such discrete systems. This approach considers the important modelling and processing case which uses Swap-based operations to represent reversible elementary cellular automata even in the presence of noise. Chapter 5 reviews self-assembly of InAs quantum dot molecules with different features fabricated by the combination of conventional Stranski-Krastanow growth mode and modified molecular beam epitaxy technique using thin or partial capping as well as droplet epitaxy. InGaAs quantum rings with square shaped nanoholes are realized by droplet epitaxy, which are utilized as nano-templates for quadra quantum dot molecules where four InAs quantum dots are situated at the four corners of a square. This quadra quantum dot set is a basic quantum cellular automata cell for future quantum computation. Chapter 6 provides a brief overview of the basic properties of quantum information processing and analyzes the quantum versions of classical cellular automata models. Then it examines an application of quantum cellular automata, which uses quantum computing to realize real-life based truly random network organization. This abstract machine is called a quantum cellular machine, and it is designed for controlling a truly random biologically-inspired network, and to integrate quantum learning algorithms and quantum searching into a controlled, self-organizing system. Chapter 7 proposes a new and simple approach for designing high-performance molecular quantum cellular automata. It reviews two approaches for the theoretical study on the two-site molecular quantum cellular automata and discusses the influence of complex charge n on the signal transmission through molecular quantum cellular automata.

Materials Science. Chapter 8 of this section discusses a combined approach of a three-dimensional frontal cellular automata model with a finite element model which has been developed for modelling the macrostructure formation during the solidification in the continuous casting line. This joint has allowed improving accuracy of modelling. Calculated distribution of the temperature gives a basis for the simulation of macrostructure formation close to the real one. In Chapter 9, a novel point automata method is developed and applied to model the dendritic growth process. The main advantages of this method are: no need for mesh generation or polygonisation; the governing equations are solved with respect to the location of points (not polygons) on the computational domain; it allows rotating dendrites in any direction since it has a limited anisotropy of the node arrangements; it offers a simple and powerful approach of cellular automata type simulations; it offers straightforward node refinement possibility, and straightforward extension to 3D. Chapter 10 proposes a model based upon the coupling of a modified cellular automaton model with the growth calculation of a nucleus in a given nucleation cell, to simulate the evolution of the dendritic structure in

solidification of alloys. For a free dendritic growth in the undercooled melt, it is found that the proposed model can quantitatively describe the evolution of dendritic growth features, including the growing and coarsening of the primary trunks, the branching of the secondary and tertiary dendrite arms, as well as the solute segregation patterns. Moreover, the directional solidification with the columnar and equiaxed grains is simulated by the proposed method and the evolution from nucleation to impingement between grains is observed. Chapter 11 underlines that modelling at a mesoscopic scale using cellular automata appears as an interesting tool to understand general properties of corrosion processes. One part of the chapter focuses on the diffusion and the feedback effect of the layer on the corrosion rate, but no explicit reactions are taken into account. The model developed gives a simple description of a phenomenon of passivation. In other part the models are improved by introducing more explicitly some basic chemical and electrochemical processes. This was illustrated by considering a heterogeneous process in which a metal recovered by an insulating film is put in contact with an aggressive medium due to a local rupture in the film.

Cryptography and Coding. Chapter 12 describes the variable-length encryption method. It is a cryptographic model based on the backward interaction of cellular automata toggle rules. This method alternates during the ciphering process the employment of the original reverse algorithm with a variation inspired in Gutowitz's model, which adds extra bits when a preimage is calculated. Using the variable-length encryption method it is guaranteed that ciphering is possible even if an unexpected Garden-of-Eden state occurs. A short length ciphertext depends, however, on the secret key specification. The proper specification of the rules/key was deeply investigated in this chapter. In Chapter 13, as an application of quantum cellular automata technology, the Serpent block cipher (a finalist candidate of Advanced Encryption Standard) was implemented. This cryptographic algorithm has 32 rounds with a 128-bit block size and a 256-bit key size, and it consists of an initial permutation, 32 rounds, and a final permutation. Each round involves a key mixing operation, a pass through S-boxes, and a linear transformation. In the last round, the linear transformation is replaced by an additional key mixing operation. Simulation results were obtained from QCADesigner v2.0.3 software. Chapter 14 proposes a multi-dimensional and multi-rank pseudorandom generator for applied cryptography, which combines the 3D cellular automata algorithm and the linear feedback shift register architecture. The feasibility and efficiency of the algorithm were studied by using several tests. The final result showed they also can provide better pseudorandom key stream and pass the FIPS 140-1 standard tests. In Chapter 15, an efficient pseudorandom number generator based on hybrid between one-dimension and two-dimension cellular automata is proposed. The proposed algorithm is compared with previous works to check the quality of randomness. It could generate a good quality of randomness and pass by the ENT Walker and DIEHARD Marsaglia test suite. In Chapter 16, from a series of definitions, propositions and theorems, a solid foundation of variant logic framework has been constructed. Under selected sample images and operational matrices, a set of typical results is illustrated. This construction can be observed from different view-points under locally and globally symmetric considerations, in addition to detect emerging patterns from each recursively operations and a functional space viewpoint, further global transforming patterns can be identified. The mechanism can be developed further to establish foundations for logical construction of applications for computational models and structural optimisation requirements.

Robotics and Image Processing. Chapter 17 presents a problem of human and robot interaction called the Human State Problem and it shows how it can be partially analyzed and solved using deterministic hardware based approach using a Cellular Automaton as well as probabilistic Bayesian Network. This robotic processor is illustrated using cellular automata for adaptive resources selection and it is shown, in particular, how it can be used for machine learning of robot behavior by modifying the local state-transition function of the cellular automaton in real time. The aim of Chapter 18 is to apply a robust self-assembly strategy to the design of self-assembling robotics. It describes various models for morphogenesis and existing techniques for designing self-assembling robotics. Then, it introduces a cellular automata model for morphogenesis and determines the necessary conditions for its robust self-assembly and self-assembly to a pre-defined shape. Finally, it demonstrates the model coordinating the self-assembly of 55,000 cell virtual robot. Chapter 19 presents a number of cellular automata-based algorithms for medical image processing. It starts by introducing cellular automata fundamentals necessary for understanding the proposed algorithms. Then, a number of cellular automata algorithms for medical image edge detection, noise filtering, spot detection, pectoral muscle identification and segmentation were presented. In this regard, 2D mammogram images for the breast cancer diagnosis were investigated. Chapter 20 explores the possibility of using General Purpose Graphic Processing Unit in the context of integrative biology. The proposed approach is explained and presented with the implementation of two cellular automata algorithms to compare different memory usage. The performances showed significant speedup even when compared to the latest CPU processors. The results obtained were compared in particular with an Nvidia Tesla C1060 board to a sequential implementation on 2 kinds of CPU (Xeon Core 2 and Nehalem).

We hope that after reading the contributions presented in this book we will have succeeded in bringing across what engineers and scientists are doing about the application of the cellular automata techniques for modelling systems and processes in diverse disciplines, so as to produce innovative simulation tools and methods to support the development of science and engineering. We also hope that the readers will find this book interesting and useful.

Lastly, we would like to thank all the authors for their excellent contributions in different areas covered by this book.

Alejandro Salcido
Instituto de Investigaciones Eléctricas
Cuernavaca,
Mexico

19

Cellular Automata for Medical Image Processing

Sartra Wongthanavasu
Khon Kaen University
Thailand

1. Introduction

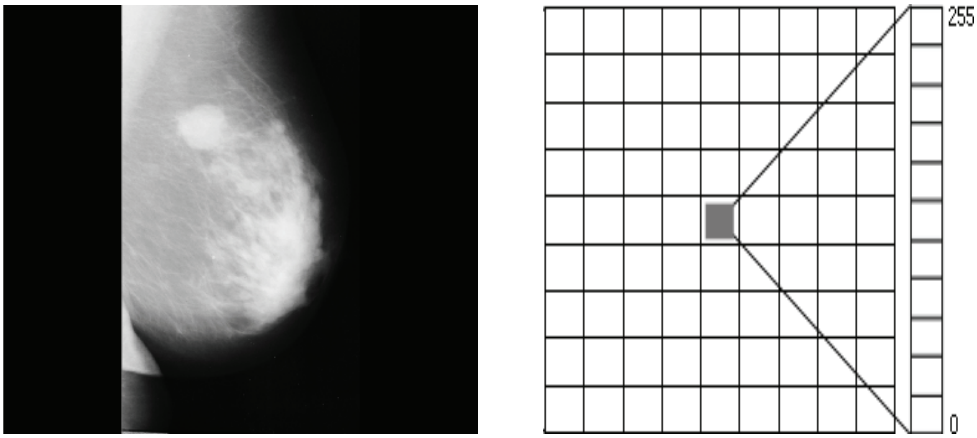
Cellular automata (CA) were introduced to provide a formal framework for investigating the behaviour of dynamic complex system in which time and space are discrete. They comprise an array of cells, where each cell can be in one of a finite number of possible states, which is updated synchronously in discrete time steps according to local transition rules (cell rules). A state of a cell at the next time step is determined by its neighboring cell's current state. A substantial numbers of CA activities occurred in the 1970s with the introduction of artificial life. There were a number of distinguished papers and books to date has investigated artificial life (Raghawan, 1993; Langton, 1986, 1992; Pesavento, 1995; Rietman, 1993). Interest in important features of physics was spawned largely by Tommaso Toffoli. Spephen Wolfram was responsible for capturing the wider interest of the physics community with a series of papers in the 1980s, while others were applying CAs to a variety of problems in other fields (Sarkar & Abbasi, 2006; Xiao et al., 2008; Bandini et al., 2001; Mizas et al., 2008). In present, CA are being studied from many widely different angles, and the relationship of these structures to existing problems in being constantly sought and discovered (Reynaga & Amthauer, 2003; Mitchell et al., 1994; Hecker et al., 1999). As the topology of CA, they appear as natural tools for image processing due to their local nature and simple parallel computation implementation. To date, there are a number of papers which generally discuss cellular automata for image processing (Hernandez & Herrmann, 1996; Rosin, 2006; Chen & Horng, 2010; Chen & Lai, 2007; Eslami et al., 2010; Tzionas et al., 1997; Umeo, 2001). In this regard, there were some papers discuss medical image processing using CA model (Cheng et al, 2006; Viher et al, 1998, Ferrari et al., 2004; Chen et al., 2008).
This paper presents a number of cellular automata-based algorithms for medical image processing. It starts by introducing cellular automata fundamentals necessary for understanding the proposed algorithms. Then, a number of cellular automata algorithms for medical image edge detection, noise filtering, spot detection, pectoral muscle identification and segmentation was presented. In this regard, 2-D mammogram images for the breast cancer diagnosis were investigated.

2. Cellular automata

Let I denote the set of integer. A 2-D cellular space is a 4-tuple, (IxI, V, N, f), where IxI is a set of cartesian product of two integer sets, V is a set of cellular states, N is the type of neighborhood, and f is the local transition function from V^n into V. The relevant neighborhood function is a function from IxI into 2^{IxI} defined by g (α) = {α+δ1, α+δ2,…,

$\alpha+\delta_n\}$, for all $\alpha \in IxI$, where δ_i $(i = 1,2,\ldots,n) \in IxI$ is fixed. The neighborhood state function of a cell α at time t is defined by $h^t(\alpha) = (v^t(\alpha+\delta_1), v^t(\alpha+\delta_2), \ldots, v^t(\alpha+\delta_n))$. For 2-D von Neumann neighborhood, the neighborhood state function of the central cell (α) is defined by: $h^t(\alpha) = (v^t(\alpha+(0,0)), v^t(\alpha+(0,1)), v^t(\alpha+(1,0)), v^t(\alpha+(0,-1)), v^t(\alpha+(-1,0))$, where $v^t(\alpha+(0,0)$ is current state of the central cell, $v^t(\alpha+(0,1)$ $(v^t(\alpha+(0,-1))$ for the north (south) cells, $v^t(\alpha+(1,0)$ $(v^t(\alpha+(-1,0))$ for the east (west) cells.

Now we relate the neighborhood state of a cell α at time t to the cellular state of that cell at time t+1 by $f(h^t(\alpha)) = v^{t+1}(\alpha)$. The function f is referred to as the 2-D CA rule and is usually given in the form of a state table, specifying all possible pairs of the form $(h^t(\alpha), v^{t+1}(\alpha))$. Figure 1 shows 2-D digital image and 2-D CA.



(a) 2-D digital image. (b) 2-D cellular automata with 256 states.

Fig. 1. A 2-D digital image vs A 2-D CA.

3. Cellular automata for medical image edge detection

As stated previously, cellular automata techniques appear as a natural tool for image processing due to their local nature and simple parallel computing implementation. This section presents one main algorithm and investigates its variation for processing mammogram images. The algorithms will cope with edge detection and noise removal for both binary and grayscale images, while the last one will correspond to hypothesized spots detection for breast cancer analysis. Examples of the application of these cellular automata techniques to real mammogram images will be presented, which together with the results will show the performance characteristics.

The main cellular automata algorithm for k gray levels of digital images is on the basis of a two dimensional cellular automata (IxI, V, N, f) with V = {0, 1, 2, …, k-1}, where k is a number of states, N is the type of neighborhood, while the local transition function f is from V^n into V. The proposed algorithm is shown in (1) below:

$$f((\nu^t(\alpha+\delta_1),\nu^t(\alpha+\delta_2),\ldots,\nu^t(\alpha+\delta_n)) = E(\alpha), \quad \text{if } \max_{j=0}^{k-1} N(C_j) = C_{target} \text{ and } sum(\nu^t(C_{target})) > k\text{-}1$$
$$= B(\alpha), \quad \text{otherwise} \tag{1}$$

where C_j is the j^{th} class of the pixel values (states) in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2,..,m$ and $\nu^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is a number of neighbors of α which fall into class C_j.
$sum(\nu^t(C_{target}))$ is summation of $\nu^t(C_{target})$.
C_{target} is the majority class containing maximal number of neighbors.
$\nu^t(C_{target})$ denotes all of $\nu^t(\alpha+\delta_i) \in C_{target}$.
$E(\alpha)$ is the edge pixel value.
$B(\alpha)$ is the background pixel value.
k is a number of states.
For class arrangement, histogram distribution will be utilized to supervise an identification of each class. Automatic class arrangement can be implemented using Otsu algorithm (Otsu, 1979).

3.1 Grayscale images

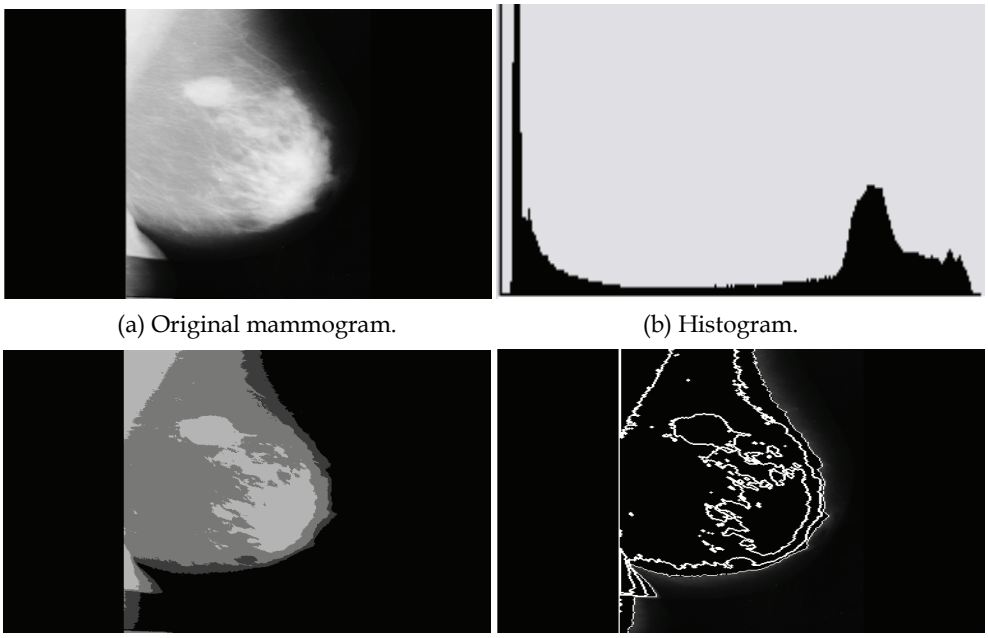
The objective of the edge detection techniques is to enhance the magnitude of the local differences in gray level values between regions of the images. Over regions which are different, changes must be made to enhance the edges. The proposed variation of (1) which deals with this task is shown in formula (2) for Von Neumann's neighborhood, four classes and 256 gray levels as following:

$$f((\nu^t(\alpha+\delta_1),\nu^t(\alpha+\delta_2),\ldots,\nu^t(\alpha+\delta_5)) = 255, \quad \text{if } \max_{j=0}^{3} N(C_j) = C_{target} \text{ and } sum(\nu^t(C_{target})) > 255$$
$$= 0, \quad \text{otherwise} \tag{2}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $\nu^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is number of neighbors of α which fall into class C_j.
C_{target} is the majority class containing maximal number of neighbors $(\nu^t(\alpha+\delta_i) \in C_{target})$.
$sum(\nu^t(C_{target}))$ is summation of $\nu^t(C_{target})$.
$\nu^t(C_{target})$ denotes all of $\nu^t(\alpha+\delta_i) \in C_{target}$.
Prior to implementing the formula (2) on a particular image, a class arrangement using histogram information was determined. Figure 2 shows an original mammogram image (a) and its histogram information (b). Results for four classes arrangement and the edge image were shown in Figure 2 (c) and (d), respectively.
It is explicitly shown that the cellular automata algorithm simply provides the promising edge maps shown in Fig. 3(d).

3.2 Binary images

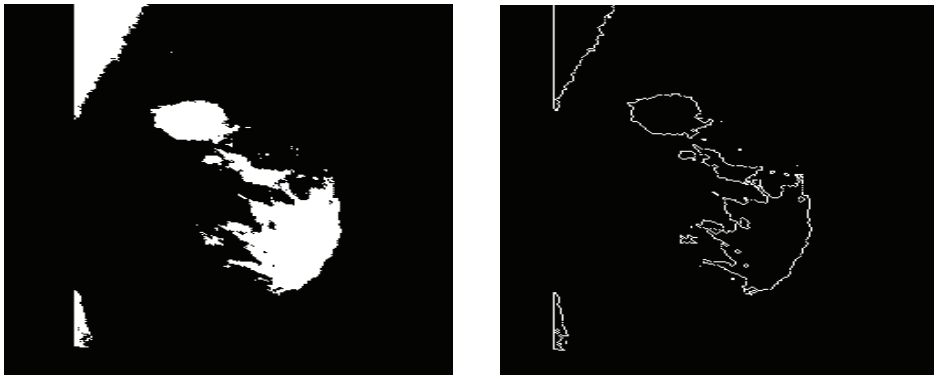
In case of edge detection on binary mammogram, the cellular automata algorithm given in the formula (2) directly enables carrying out to two states efficiently. There is no need to be changed. Figure 3 shows the promising result implementing the algorithm. An Edge result exhibits the superb quality with one pixel wide and edge has no break.



(a) Original mammogram. (b) Histogram.

(c) Four class image. (d) Edge map result.

Fig. 2. (a) Original mammogram, (b) histogram information, (c) four classes of image, and (d) result of edge detection.



(a) Original binary image. (b) Edge result.

Fig. 3. Edge result as dealing with binary image.

3.3 Noise filtering

The objective of the noise filtering is to reduce the local differences in gray level values between regions of the images. Over regions which are similar, no changes must be made, in order to avoid the destruction of the main characteristics of the image. The proposed

variation of cellular automata algorithm given in (1) using von Neumann's neighborhood and two states dealing with this task is shown in formula (3) as follow:

$$\begin{aligned} f((\nu^t(\alpha+\delta_1),\nu^t(\alpha+\delta_2),\ldots,\nu^t(\alpha+\delta_5)) &= \max(\nu^t(C_{target})), \quad \text{if } \max_{j=0}^{3} N(C_j) = C_{target} \\ &= 0, \qquad\qquad\qquad\quad \text{otherwise} \end{aligned} \tag{3}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $\nu^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is number of neighbors of α which fall into class C_j.
C_{target} is the majority class containing maximal number of neighbors $(\nu^t(\alpha+\delta_i) \in C_{target})$.
$\nu^t(C_{target})$ denotes all of $\nu^t(\alpha+\delta_i) \in C_{target}$.
$\max(\nu^t(C_{target}))$ is the maximal state of $\nu^t(C_{target})$.
Figure 4 shows an original binary mammogram with 2% salt and pepper noise. The noise filtering result obtained by implementing the algorithm (3) using one step of iteration was shown in Figure 4 (b). In this respect, the cellular automata algorithm (3) explicitly provides the promising result.

(a) Noisy binary image. (b) Noise removal.

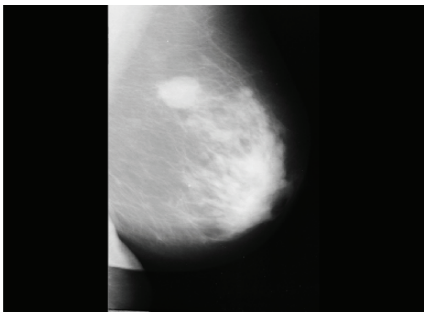
Fig. 4. Binary image with 2% salt and pepper noise (a) and the image after noise removal (b).

3.4 Spot detection

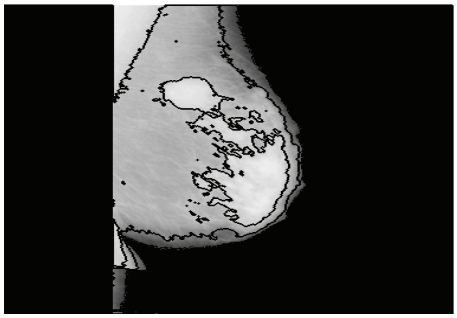
The objective of the spot detection is to assist the physicians and doctors in locating hypothesized spots for masses in breast cancer. The shape and spread region of spots play a vital role for further steps of analyses and have to be comprehensively taken into account. For this purpose, a set operator is implemented as follow:

$$W = X - Y \tag{4}$$

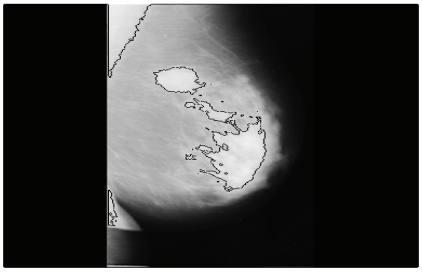
where X denotes an investigating original image,
Y denotes an edge map resulting from implementing the formula (2), and
W denotes the spot detection image.
The operator (-) represents the subtraction, meaning that it simply carries out the subtraction of pixel values in the same coordinate between two images (sets). By implementing such an operator in an original image (X) shown in Fig. 5(a) with respect to the edge image Y obtained by implementing the formula (2) on grayscale and binary images, the resulting spot detection was shown in Figure 5 (b) and (c), respectively.



(a) Malignant mammogram



(b) Spot detection using grayscale image.



(c) Spot detection using binary image.

Fig. 5. Spot detection due to (2) for binary and grayscale images.

Figure 5. shows mammogram processing with white spot detection: (a) original malignant mass in mammogram, (b) hypothesized mass detection using algorithm (2) for grayscale, and (c) hypothesized mass detection using algorithm (3) for binary image.

4. Cellular automata for identification of the pectorial muscle in mammograms

The pectoral muscle represents a predominant density region in most medio-lateral oblique (MLO) views of mammograms. Its inclusion can affect the results of intensity-based image processing methods. To date, there were some papers has investigated identification of the pectorial muscle in mammograms (Ferrari et al., 2004; Ma et al., 2007). This section presents a new method on the basis of cellular automata model for the identification of the pectoral muscle in MLO mammograms. A dataset of 84 MLO mammograms from the MIAS (Mammographic Image Analysis Society, London, U.K.) database (Suckling et al., 1994) was implemented throughout for evaluation. In this respect, the pectoral muscle edge detected in the mammograms was carried out based upon the percentage of false-positive (FP) and false-negative (FN) pixels determined by comparison between the numbers of pixels enclosed in the regions delimited by the edges identified by a radiologist and by the proposed method.

4.1 Proposed CA-based algorithm

In this section, the method for identification of pectoral muscle in mammogram images were presented. It starts by computing the edge detection by using the formula (2) stated earlier dealing with grayscale images. Then, the result will be segmented by using the rule-based

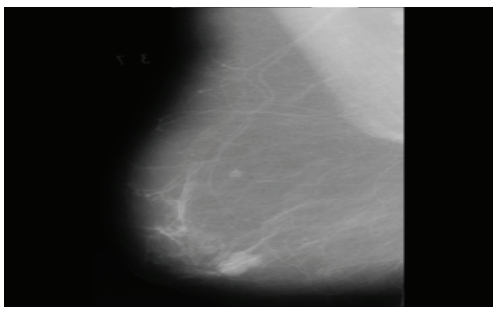
algorithm, leading to the identification of the pectoral muscle. Examples of the application of the proposed method to real mammogram images will be presented, which together with the results will show the performance characteristics.
The proposed cellular automata algorithm for 256 gray levels of digital images are on the basis of a two-dimensional cellular automata (IxI, V, N, f) with V = {0, 1, 2, …, 255}, N is von Neumann type of the neighborhood, while the local transition function f is from V^n into V.

4.1.1 Edge detection

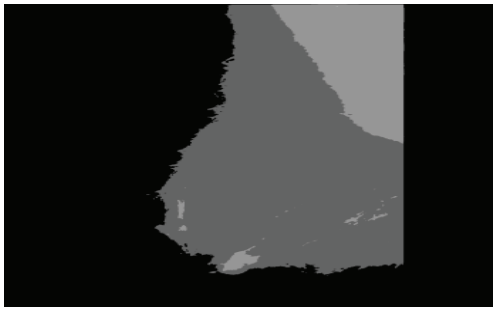
Referred to the formula (2) stated previously, the cellular automata algorithm for this task was revisited as follow:

$$\text{if } ((\nu^t(\alpha+\delta_1),\nu^t(\alpha+\delta_2),\ldots,\nu^t(\alpha+\delta_5)) = 255, \quad \text{if } \max_{j=0}^{3} N(C_j) = C_{target} \text{ and } \text{sum}(\nu^t(C_{target})) > 255$$
$$= 0, \quad \text{otherwise}$$

where C_j is the j^{th} class of the pixel values in its neighborhood $(h^t(\alpha))$ for $j = 0, 1, 2, 3$ and $\nu^t(\alpha+\delta_i) \in C_j$.
$N(C_j)$ is number of neighbors of α which fall into class C_j.
C_{target} is the majority class containing maximal number of neighbors $(\nu^t(\alpha+\delta_i) \in C_{target})$.
$sum(\nu^t(C_{target}))$ is summation of $\nu^t(C_{target})$.
$\nu^t(C_{target})$ denotes all of $\nu^t(\alpha+\delta_i) \in C_{target}$.



(a) Original image mdb005.



(b) Four class image.

(c) Edged image.

Fig. 6. Results obtained for the image mdb005: (a) Original image, (b) Its four class image, and (c) Edged image obtained by implementing (2).

Prior to implementing the algorithm (2), an original mammogram image of mdb005 (Suckling et al., 1994) was classified into four classes (Cj) due to Otsu algorithm (Otsu, 1979). The results were shown in Figure 6.

4.1.2 Pectoral muscle identification

As already mentioned, the inclusion of pectoral muscle in mammogram can affect the results of intensity-based image processing methods or bias procedures in the detection of breast cancer. The objective of the pectoral muscle identification is to determine the pectoral muscle for the exclusion from the mammograms prior to process in further steps for breast cancer diagnosis. In this regard, the edge resulting image was used in the segmentation algorithm given as follow:

Algorithm: 1 Pectoral muscle identification

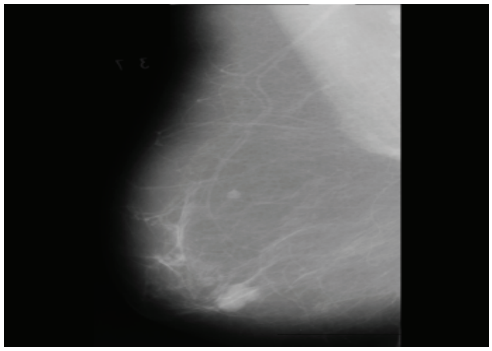
Step 1: Removal of unqualified objects.

1. Implement the formula (2) in a mammogram image P, resulting in the image Q.
2. Filter unqualified objects, objects consisting of contiguous edge less than 2,500 pixels, out of the Q resulting in the image R.

Step 2: Identification of pectoral muscle.

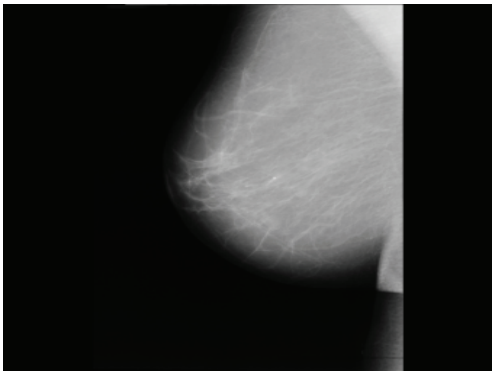
1. FOR i:= 1 TO n DO
   Find the area of $ROI_r(i)$.
   { $ROI_r(i)$ is a region of interest *i* inside the image R. ROI_r at region i is the i-th area surrounded by an edge boundary on image R.}
2. Pectoral muscle segmentation:
   2.1 MUSCLE := NIL {empty set}
   2.2 FOR j:=1 TO n DO
   IF $ROI_r(j) >$ max_size AND $ROI_r(j) \leq$ min_size
   THEN $ROI_r(j)$:= background
   { Set the $ROI_r(j)$ to background when it is not of interest regions. In practical uses in our implementation, max_size = 95,200 pixles, and min_size = 2,500 pixles.}
   ELSE IF min (y, $ROI_r(j)$) > min_upper
   THEN $ROI_r(j)$:= background
   { min (y,$ROI_r(j)$) denotes minimal value of y for a coordinate (x,y) of any pixels in the region of $ROI_r(j)$. In our implementation, min_upper = 60.}
   { Regions of interest will locate at min(y, $ROI_r(j)$) in which is at the topmost part of the image R. }
   ELSE IF mean (y, $ROI_r(j)$) > min_y_average
   THEN $ROI_r(j)$:= background
   { mean (y,$ROI_r(j)$) is an average of y for coordinates (x,y) inside the region of $ROI_r(j)$ In our implementation, min_y_average = 350. }
   ELSE IF mean (I, $ROI_r(j)$) > min_average_intensity
   THEN MUSCLE := MUSCLE $\cup$ $ROI_r(i)$
   { mean (I, $ROI_r(j)$) is an average intensity of all pixels inside the region of $ROI_r(j)$. In our implementation, min_average_intensity = 157. }

The results in implementing the proposed segmentation algorithm on mdb005 and mdb011 were shown in Fig. 7. Fig. 8 shows hand-drawn pectoral muscle edge by radiologist as applied to mdb005, the result obtained by the proposed algorithm, and the pectoral muscle removal according to the proposed algorithm.



(a) Original image mdb005.

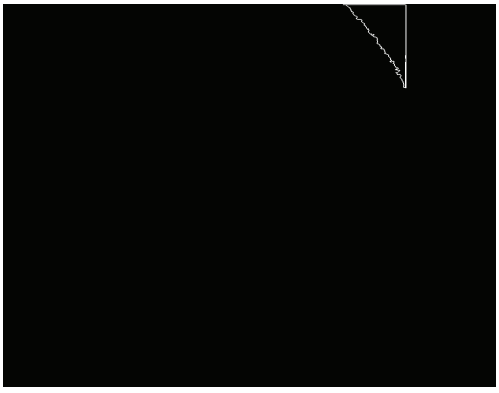
(b) Result obtained by the algorithm 2.



(c) Original image mdb011.

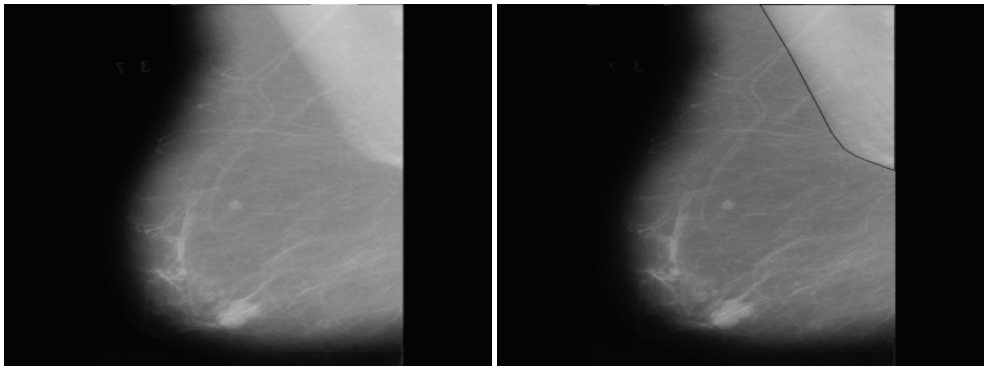
(d) Result obtained by the algorithm 2.

(e) Pectoral muscle identification of mdb005.

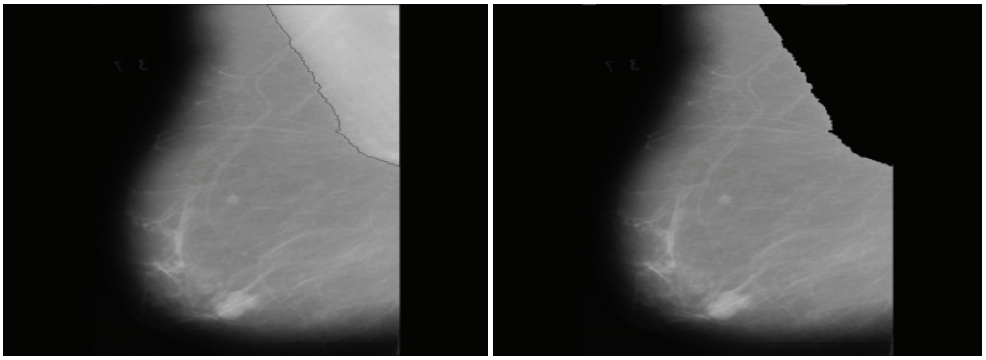


(f) Pectoral muscle identification of mdb011.

Fig. 7. Results obtained from proposed algorithm: (a) and (c) are original images mdb005 and mdb011, respectively, (b) and (d) are edge results due to original images (a) and (c), respectively, and (e) and (f) are results obtained from the pectoral muscle identification algorithm due to original images (a) and (c), respectively.



(a) Original image (b) Hand-drawn pectoral muscle edge by radiologist



(c) Pectoral muscle edges detected by the algorithm. (d) Pectoral muscle removal using the algorithm.

Fig. 8. Results obtained for the image mdb005: (a) original image, (b) hand-drawn pectoral muscle edge by radiologist superimposed on the original image, (c) pectoral muscle edges detected by the CA-based method superimposed on the original image, and (d) pectoral muscle removal using the CA-based method.

4.2 Performance evaluation and results

For performance evaluation purpose, the total of 84 images, randomly selected from the Mammographic Image Analysis Society, London, U.K. (Suckling et al., 1994), were used in experimentation. The spatial resolution of these images is 200 μm and depth resolution in 8 bit. The images in the database are 1024x1024 pixels in size. The result obtained from the proposed method was evaluated in consultation with radiologists experienced in mammography. Then, the pectoral muscle edges were manually drawn by the author under the supervision of radiologists, without referring to the results of detection by the proposed method. The segmentation results of the proposed method was evaluated based upon the number of false-positive (FP) and false-negative (FN) pixels in the regions demarcated by the manually drawn edges. An FP pixels was defined as a pixel outside the reference region that was included in the pectoral region segmented. An FN pixel was defined as a pixel in the reference region that was not present within the segmented region. Table 1 shows mean

and standard deviation values of the FP and FN pixels for the result of the proposed method with 84 images.

CA-based algorithm	Statistics
Analysis of area enclosed	
FP ± σ	1.99±8.19%
FN ± σ	18.89±14.19%
# images with (FP and FN) < 5%	13
# images with 5% <(FP and FN) <10%	15
# images with (FP and FN) > 10%	56

Table 1. Mean and standard deviation values of the FP and FN pixels for the results of CA-based algorithm with 84 images.

Table 1 shows an interesting statistical results that the identification obtained by the CA-based algorithm provides the promising results with minimal FP and FN. In this regard, the proposed CA-based algorithm is one of promising methods in pectoral muscle identification for mammogram processing.

5. Cellular automata for mass segmentation in mammograms

This section investigates algorithms on the basis of cellular automata model for coping with the segmentation of hypothesized masses in mammograms. The 256-states cellular automata algorithms were developed to deal with 256 grayscale mammogram images for determining masses. The proposed cellular automata algorithms investigated here are on the basis of two dimensional CAs (IxI, V, N, f) with V = {0, 1, 2, …, 255}, Von Neumann's neighborhood N, while the local transition function f is from Vn into V. The results will be used for determining mass features in breast cancer diagnosis. An empirical experimentation shows that the proposed cellular automata algorithms provide the superior results.

The proposed segmentation algorithm utilizes a concept of two-types bacteria propagation. The first type confiscates a mass seed, which is an object image, while the second one confiscates the background which is declared by a circle surrounding the mass seed. Both types of bacteria propagate to their neighbors in parallel at each time step depending on its and their neighbor's strengths. A circle surrounding a mass seed shown in Fig. 4 (a) represents the background seed being seized by the other type of bacteria. Both types of bacteria continue propagating in parallel in discrete time steps as stated earlier. In order to reduce the computation time, the proposed algorithm can be stopped anytime earlier so far as the propagation of bacteria taken up the mass seed is consistent. The CA-based segmentation algorithm is given as follow:

Algorithm: 2 CA-based algorithm for mass segmentation

// Identify hypothesized mass seed from a mammogram image P; mark the mass seed
// Identify object seed from a mammogram image P; mark the object seed
// For each cell (pixel) in P

for $\forall p \in P$

// copy previous state

$l_p^{t+1} = l_p^t$;

$\theta_p^{t+1} = \theta_p^t$;
// neighbors try to attack current cell
for $\forall q \in N(p)$
if $g(||C_p\text{-}C_q||_2).\theta_q^t > \theta_p^t$ then
$l_p^{t+1} = l_q^t$;
$\theta_p^{t+1} = g(||C_p\text{-}C_q||_2).\theta_q^t$
end if
end for
end for

where P denotes an image, p denotes a pixel.
θ_q^t denotes the strength of a cell $q \in Q$ at time t.
θ_p^t denotes the strength of a p's neighboring cells at time t.
l_p^t, l_p^{t+1} denote the centering cell state of p at time t and $t+1$, respectively.
l_q^t, l_q^{t+1} denote the neighboring cell state of q at time t and $t+1$, respectively.
$N(p)$ denotes neighboring cells of p.
$||C_p\text{-}C_q||_2$ denotes the Euclidian distance between a cell q and a neighboring cell p.
$g(x)$ represents a monotonous decreasing function defined by (2) as following

$$g(x) = 1 - \frac{x}{\max\|C\|_2} \tag{5}$$

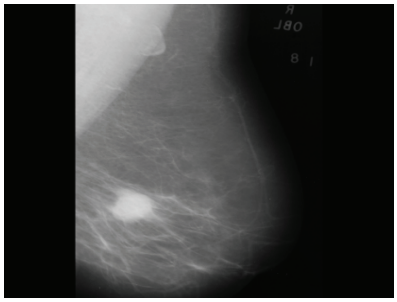
In implementing the algorithm 2, the termination condition was satisfied if the configuration of mass seed has not been changed, meaning that their states were steady. This abundantly reduces time complexity in practical implementation. Figure 9 shows the results of implementing the proposed mass segmentation algorithm on the mammogram mdb028. The hypothesized mass seeds, depicted by red shadow, and background seeds, depicted by blue circle, are represented two types of bacteria as shown in in Fig. 8 (a) and (b), respectively. Fig. 8 (c) and (d) show a series of propagating results on the 40 evolution steps and the final result at iteration 72. Fig. 10 shows the results of mass segmentation of mdb005 and mdb092 superimposed on the original mammograms, respectively.

6. Conclusions and discussions

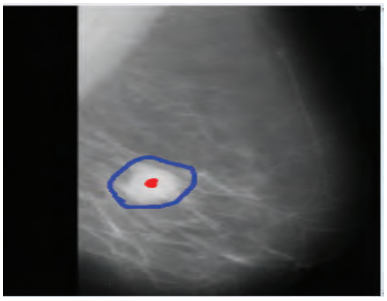
The behavior of cellular automata is fascinating not only from theory but also from applications. They offer the beauty and elegance of results in medical image processing as reported in literature. In this work we have presented a number of uniform cellular automata algorithms for mammogram image processing. Based on empirical experimentation, the proposed algorithms give the promising results when tested on MIAS mammograms. This quite encourages in determining other tasks for the research. In this regard, we have more investigations on applications for mammogram and other medical image processing and hope report in the near future.

7. Acknowledgement

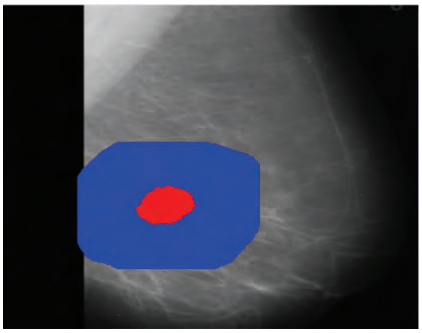
We deeply thank to The Thailand Research Funds (TRF) for financial support of the research project due to RMU5080010.



(a) Original image mdb028.



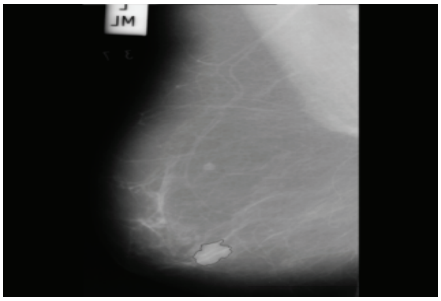
(b) Seeds of object and background.



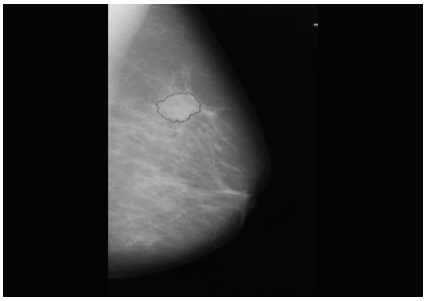
(c) Evolution result at 40 time steps

(d) Result at final time steps (72)

Fig. 9. Mass segmentation results obtained from the image mdb028: (a) original image, (b) initial seeds, (c) result after 40 time steps of evolution, and (d) final result at iteration 72.



(a) Mass segmentation result of mdb005.



(b) Mass segmentation result of mdb092.

Fig. 10. Mass segmentation obtained from the image mdb005 and mdb092: (a) mass segmentation result of mdb005 superimposed on the original mammogram, and (b) mass segmentation result of mdb092 superimposed on the original mammogram.

8. References

Hernandez, G. & Herrmann, J. (1996). Cellular Automata for Elementary Image Enhancement. *Graphical Models and Image Processing (GMIP),* Vol. 4, No. 58, pp. 82-89.

Wongthanavasu, S. & Lursinsap, C. (2004). A 3-D CA-based Edge Operator for 3-D Images, *Proceedings of ICIP 2004 11th International Conference on Image Processing,* pp. 235-238, ISBN 0-7803-8555-1, Singapore, October 24-27, 2004.

Rosin, P. (2006). Training Cellular Automata for Image Processing. *IEEE Transactions on Image Processing,* Vol. 15, No. 7, pp. 2076-2087.

Cheng, H. D., Shi, X.J., Min, R., Hu, L.M., Cai, X.P. & Du, H.N. (2006). Approaches for Automated Dectection and Classification of Masses in Mammograms. *Pattern Recognition,* Vol. 39, pp. 646-668.

Viher, B., Dobnikar, A. & Zazula, D. (1998). Cellular Automata and Follicle Recognition Problem and Possibilities of Using Cellular Automata for Image Recognition Purposes. *International Journal of Medical Informatics, Vol. 49, pp. 231-241.*

Ferrari, R.J., Rangayyan, R.M., Desautels, J.E., Borges, R.A. & Frere, A.F. (2004). Automatic Identification of the Pectoral Muscle in Mammograms. *IEEE Transactions on Medical Imaging,* Vol. 23, No. 2, pp. 232-245.

Otsu, N. (1979). A Threshold Selection Method from Gray-level Histograms. *IEEE Transactions on Systems Man Cybernetics,* Vol. 9, No. 1, pp. 62-66.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I, Boggis, C., Ricketts, I., Stamatakis, E., Cerneaz, N., Kok, S., Taylor, P., Betal, D. & Savage, J. (1994). The mammographic images analysis society digital mammogram database, Exerpta Medica. *International Congress Series 1069,* pp. 375-378

Ma, F., Slavotinek, J. & Bottema, M.J. (2007). Two graph theory based methods for identifying the pectoral muscle in mammograms. *Pattern Recognition,* Vol. 40, pp. 2598-2598

Wongthanavasu, S. & Tangvoraphongchai, V. (2007). CA-based Algorithms and Its Application in Medical Image Processing. *Proceedings of ICIP 2007 14th International Conference on Image Processing,* pp. III-41-44, ISBN: 1-4244-1437-7, ISSN: 1522-4880, San Antonio, Texas, U.S.A. September 16-19, 2007.

Wongthanavasu, S. & Tangvoraphongchai, V. (2008). Cellular Automata-based Identification of The Pectoral Muscle in Mammograms. *Proceedings of ISBME 2008 3rd International Symposium on Biomedical Engineering,* pp. 294-298, November 10-11, 2008, Bangkok, Thailand, November 10-11, 2008.

Chen, R.J. & Lai, L.L. (2007). Image security system using recursive cellular automata substitution. *Pattern Recognition,* Vol. 40, No. 5, May 2007, pp. 1621-1631

Chen, R.J. & Horng, S.J. (2010). Novel SCAN-CA-based image security system using SCAN and 2-D von Neumann cellular automata. *Signal Processing: Image Communication,* Vol. 25, Issue 6, July 2010, pp. 413-426

Reynaga, R. & Amthauer, E. (2003). Two-dimensional cellular automata of radius one for density classification task, *Pattern Recognition Letters,* Vol. 24, Issue 15, November 2003, pp. 2849-2856

Mitchell, M., Crutchfield, J.P & Hraber, P.T. (1994). Evolving cellular automata to perform computations: mechanisms and impediments. *Physica D: Nonlinear Phenomena,* Vol. 75, Issues 1-3, 1 August 1994, *pp.* 361-391

Wongthanavasu, S. & Sadananda, R. (2003). A CA-based edge operator and its performance evaluation. *Journal of Visual Communication and Image Representation,* Vol. 14, Issue 2, June 2003, pp. 83-96

Rogowska, J. (1999). Overview and Fundamentals of Medical Image Segmentation. *Handbook of Medical Image Processing and Analysis (Second Edition),* pp. 73-90

Hecker, C., Roytenberg, D., Sack, J.R. & Wang, Z. (1999). System development for parallel cellular automata and its applications. *Future Generation Computer Systems,* Vol. 16, Issues 2-3, December 1999, pp. 235-247

Karafyllidis, I. (1999). Acceleration of cellular automata algorithms using genetic algorithms. *Advances in Engineering Software,* Vol. 30, Issue 6, June 1999, pp. 419-437

Sarkar, C. & Abbasi, S.A. (2006). Cellular automata-based forecasting of the impact of accidental fire and toxic dispersion in process industries. *Journal of Hazardous Materials,* Vol. 137, Issue 1, 1 September 2006, pp. 8-30

Xiao, X., Wang, P. & Chou, K.-C. (2008). Predicting protein structural classes with pseudo amino acid composition: An approach using geometric moments of cellular automaton image. *Journal of Theoretical Biology,* Vol. 254, Issue 3, 7 October 2008, pp. 691-696

Eslami, Z., Razzaghi, S.H. & Ahmadabadi, J.Z. (2010). Secret image sharing based on cellular automata and steganography. *Pattern Recognition,* Vol. 43, Issue 1, January 2010, *pp.* 397-404

Tzionas, P., Thanailakis, A. & Tsalides, P. (1997). An efficient algorithm for the largest empty figure problem based on a 2D cellular automaton architecture. *Image and Vision Computing,* Vol. 15, Issue 1, January 1997, pp. 35-45

Umeo, H. (2001). Linear-time recognition of connectivity of binary images on 1-bit inter-cell communication cellular automaton. *Parallel Computing,* Vol. 27, Issue 5, April 2001, pp. 587-599

Bandini, S., Mauri, G. & Serra, R. (2001). Cellular automata: From a theoretical parallel computational model to its application to complex systems. *Parallel Computing,* Vol. 27, Issue 5, April 2001, pp. 539-553

Chen, J.-C., Yeh, C.-M. & Tzeng, J.-E. (2008). Pattern differentiation of glandular cancerous cells and normal cells with cellular automata and evolutionary learning. *Expert Systems with Applications,* Vol. 34, Issue 1, January 2008, pp. 337-346

Mizas, C., Sirakoulis, G.C., Mardiris, V., Karafyllidis, I., Glykos, N. & , Sandaltzopoulos R. (2008). Reconstruction of DNA sequences using genetic algorithms and cellular automata: Towards mutation prediction?. *Biosystems,* Vol. 92, Issue 1, April 2008, pp. 61-68

Richards, F.C., Meyer, T.P. & Packard, N.H. (1990). Extracting cellular automaton rules directly from experimental data. Physica D: Nonlinear Phenomena, Vol. 45, Issues 1-3, 2 September 1990, pp. 189-202

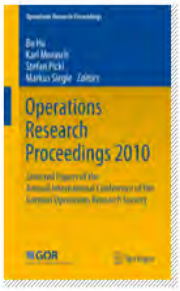
Raghavan, R. (1993). Cellular automata in pattern recognition. *Information Sciences,* Vol. 70, Issues 1-2, May 1993, pp. 145-177

Langton, C.G. (1986). Studying artificial life with cellular automata. *Physica D,* Vol. 22, pp. 120-140

Langton, C.G. (1992). Life at the edge of chaos. *Artificial life, Vol. 2, No. 4, pp* 41-91

Pesavento, U. (1995). An implementation of von Neumann's self-reproducing machine. *Artificial Life,* Vol. 2, No. 4, pp. 337-354

Rietman, E. (1993). *Creating Artificial Life : Self-organizatin.* Pensylvania : Windcrest/McGraw-Hill.



Operations Research Proceedings 2010

Selected Papers of the Annual International Conference of the German Operations Research Society
Series: » Operations Research Proceedings, Vol.

Hu, B.; Morasch, K.; Pickl, S.; Siegle, M. (Eds.)

2011, XVI, 667p. 131 illus..

Free Preview

Available Formats:

eBook
(gross) price for Thailand **166,59 €**
Add to cart

- ISBN 978-3-642-20009-0
- digitally watermarked, no DRM
- Included Format: PDF
- download immediately after purchase

» learn more about Springer eBooks
» add to marked items

Softcover **199,99 €**

SERVICES FOR THIS BOOK

Access an Online Book Review Copy
Download Product Flyer
Download High-Resolution Cover

NEW BOOK ALERT

Get alerted on new Springer publications in the subject area of Operation Research / Decision Theory.

Your E-Mail Address

SUBMIT

RECOMMENDED BOOKS

Handbook of Metaheuristics

Editor» Gendreau, Michel (et al.)

Price from **178,49 €**

Latest findings from research in OR Best papers from a leading OR conference Solutions for a globalized economy

This book contains selected papers from the symposium "Operations Research 2010" which was held from September 1-3, 2010 at the "Universität der Bundeswehr München", Germany. The international conference, which also serves as the annual meeting of the German Operations Research Society (GOR), attracted more than 600 participants from more than thirty countries. The general theme "Mastering Complexity" focusses on a natural component of the globalization process. Financial markets, traffic systems, network topologies and, last but not least, energy resource management, all contain complex behaviour and economic interdependencies which necessitate a scientific solution. Operations Research is one of the key instruments to model, simulate and analyze such systems. In the process of developing optimal solutions, suitable heuristics and efficient procedures are some of the challenges which are discussed in this volume.

Content Level » Research

Keywords » Decision Support - Multicriteria Analysis - Optimization - Scheduling - Simulation

Related subjects » Applications - Computational Intelligence and Complexity - Mathematics - Operations Research & Decision Theory

TABLE OF CONTENTS

POPULAR CONTENT WITHIN THIS PUBLICATION

Operations Research Proceedings 2010, Springer, pp. 3-8

A Bayesian Belief Network Model for Breast Cancer Diagnosis

S. Wongthanavasu

Abstract A statistical influence diagram, called Bayesian Belief Network (BBN), is investigated in modeling the medical breast cancer diagnosis. The proposed BBN is constructed under supervision by medical experts. Four types of datasets, namely, historic biodata, physical findings, indirect and direct mammographic findings are 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 are information obtained by mammogram image processing using the proposed cellular automata algorithms. A dataset is collected in real case of the breast cancer patients who come to get serviced at Srinakarind Hospital, Khon Kaen University, Thailand. A dataset of 500 cases is used throughout for model's performance evaluation.

In this regard, an 80 % of data is used for training the model, while the rest of 20 % is utilized for testing. The trained BBN model is tested on 100 patients consisting of 50, 25 and 25 for normal, benign and malignant patients, respectively. The proposed BBN provides the promising results reporting the 96.5 % of accuracy in the diagnosis. In addition, 5-fold and 10-fold cross-validation approach are implemented, the proposed BBN reports the promising results. It provides 96.2 and 97.4 percentages of accuracy, respectively.

1 Problem Overview

Mammography is an important tool in early detection of breast cancer. Unfortunately, many mammography findings cannot be classified easily as malignant or benign.

Wongthanavasu, S.
Department of Computer Science, Faculty of Science
Khon Kaen University, Khon Kaen, Thailand
e-mail: wongsar@kku.ac.th

B. Hu et al. (eds.), *Operations Research Proceedings 2010,* Operations Researrch Proceedings,
DOI 10.1007/978-3-642-20009-0_1,

Successful diagnosis depends on the ability of a physician to detect mammographic abnormalities and to integrate clinical information such as risk factors and physical findings to determine the likelihood of breast cancer. In this regard, machine learning are capable to successfully assist physicians in detecting the early detection of breast cancer. Bayesian network is one of the promising candidates. There are a number of papers that reported applications of machine learning techniques in breast cancer diagnosis [1, 2, 9]. In this respect, some papers to date investigates breast cancer diagnosis using Bayesian networks [4, 5, 6]. Bayesian Belief Network (BBN) is a promising machine learning technique superior to well-known techniques such as support vector machines and neural networks in several reasons. Firstly, BBN is capable of predicting and diagnosing disease by using incomplete input data while the two previously stated methods can not. Secondly, support vector machines and neural networks can not perform diagnosis besides prediction. That is they do not accept incomplete data for the test. Thirdly, there are many medical diagnosis models in literature report to date that BBN is superior in medical diagnosis problems. In this regard, there are several papers reports the investigation of breast cancer using the BBN. Consequently, the modeling is the key successful issue.

This paper proposes a breast cancer diagnosis model using Bayesian belief network. The major issues of the model are an integration of mammographic with clinical information.

2 Model Structure

2.1 Bayesian Belief Networks

A Bayesian network, also called a belief network or causal probabilistic network, is a graphical representation of probabilistic information: it is a directed, acyclic graph in which nodes represent random (stochastic) variables, and links between nodes represent direct probabilistic influences between the variables [5]. In this formalism, propositions are given numerical probability values signifying the degree of belief accorded them, and the values are combined and manipulated according to the rules of probability theory. Each node represents a variable and has two or more possible states. For example, the variable "Breast Cancer" has two states: "present" and "absent". Each state is associated with a probability value; for each node, these probability values sum to 1.

For implementing the proposed BBN, nodes and their states are enumerated in Table 1. Four types of datasets, namely, patient history, physical findings, indirect mammographic findings, and direct mammographic findings were used in the BBNs modeling. In this regard, the associated network was shown in Fig. 1.

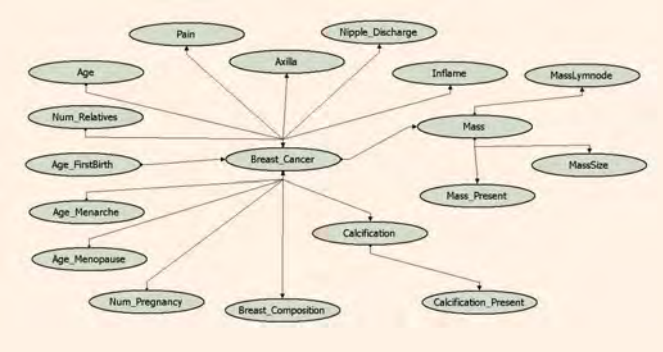


Fig. 1 Proposed BBN Model.

Table 1 Definition of BBN model's nodes (variables) and their states

Category	Node (variables)	States
Diagnosis	Age (years)	<20, 20-30, 31-40, 41-50, 51-60, >60
Patient History	Breast Cancer	present, absent
	Age at Menarche (years)	<12, 12-13, >13
	Age at First Live Birth (years)	<20, 20-24, 25-29, >29
	Number of First-Degree Relatives with Breast Cancer	0, 1, 2
	Age at First Live Birth (years)	<20, 20-24, 25-29, >29
	Age at Menopause (years)	<40, 40-44, 45-40, >49
	Number of Pregnancy	<2, 2-3, >3
Physical Findings	Pain	present, absent
	Nipple Discharge	present, absent
	Axilla	present, absent
	Inflame	present, absent
Indirect Mammographic Findings	Breast Composition	present, absent
Direct Mammographic Findings	Mass	malignant, benign, none
	Mass Present	yes, no
	Mass Lympnode	present, absent
	Mass Size (cm.)	1,2,3,4
	Calcification	malignant, benign, none
	Calcification Present	yes, no

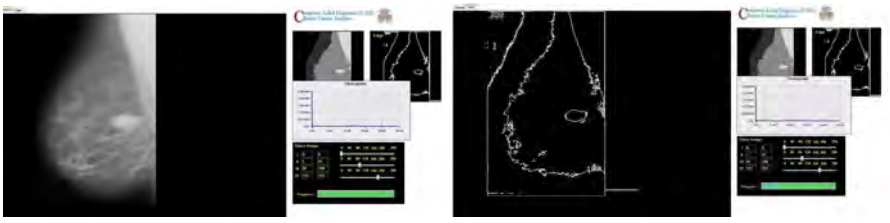
2.2 Direct Mammographic Findings

Mammographic image processing was carried out using cellular automata model (CA) [7, 8, 9, 10] to determine mass, calcification, and their features. Fig. 2 given below shows a process of mass detection in Computer Aided Diagnosis (CAD) software due to the research project.

2.3 Data Acquisition and Inference Software

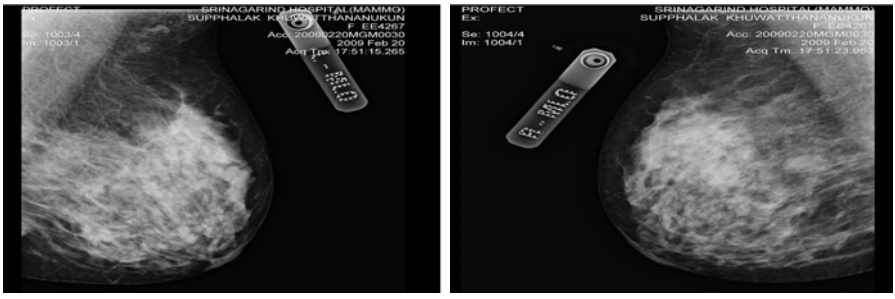
The proposed BBN's knowledge base was constructed from medical literature, health statistics reports and 500 patient's records collected at Srinakarind Hospital, Khon Kaen University. Reasoning in MSBNx inference software [3] uses conditional independence. It means that each variable (node) is independent with other variables which are not its successor given direct predecessors. This condition makes statistical reasoning possible by reducing combinatorial explosion of joint probability distribution. The patient dataset is calculated for conditional probabilities using conditional independence for each variable in the proposed BBN in Fig. 2. Fig. 2 shows mammograms of a benign patient.

Microsoft Bayesian Networks (MSBNx) [3] was utilized as a software tool to model the breast cancer diagnosis problem. MSBNx is a component-based Windows application for creating, assessing, and evaluating Bayesian Networks developed by Microsoft company. It is firstly developed as tools for assisting helpdesk in diagnosis of user's Microsoft office problems. MSBNx can deal with categorical data only by using conditional independence in reasoning. In implementation the proposed BBN model using MSBNx, 100 patients comprised of 50 normal, 25 benign, and 25 malignant were tested on the proposed model. In this regard, it provides the promising results by reporting the percentage of 96.5 of accuracy in the diagnoses.



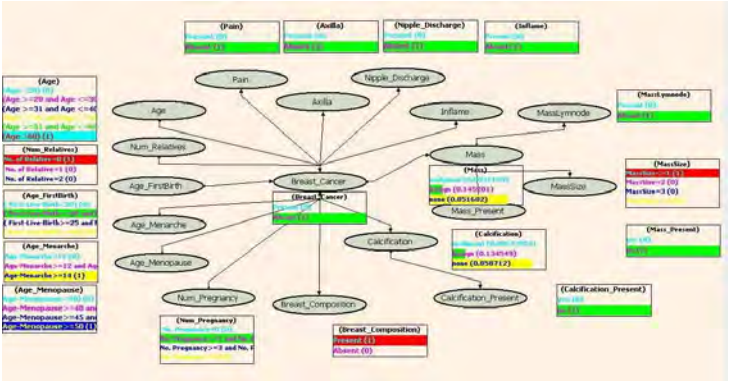
(a) Before processing (b) After processing

Fig. 2 Mammogram Image Processing Using Cellular Automata Model

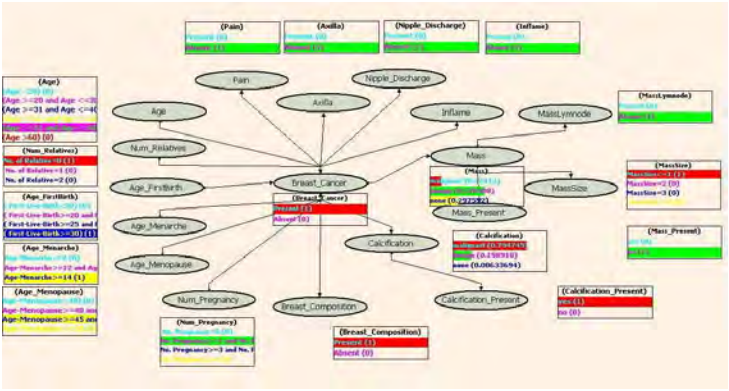


(a) Left (b) Right

Fig. 3 Mammograms for a benign patient



(a) Benign patient



(b) Malignant patient

Fig. 4 An example of BBN's results for diagnosing the test patients: a) benign, and b) malignant

Figure 1 shows the results obtained from the system as implemented in three cases of patients.

In addition, 5-fold and 10-fold cross-validation approach are implemented on such a dataset of 500 cases. On average, the proposed model reports 96.2 and 97.4 percentage of accuracy for 5-fold and 10-fold cross-validation, respectively.

3 Conclusions and Discussions

Bayesian networks represent a promising technique for clinical decision support and provide a number of powerful capabilities for representing uncertain knowledge. They provide a flexible representation that allows one to specify dependence and independence of variables in a natural way through the network topology.

Because Bayesian networks represent uncertainty using standard probability, one can collect the necessary data for the domain model by drawing directly on published statistical studies or elicited from the experts. The proposed BBN model is undergoing preclinical testing to compare its performance to that of radiologists with varying levels of mammographic expertise. In addition, we are considering the addition of variables to improve the model performance, such as demographic features such as race and geographic location, and patient-history features such as diet, body habitus, history of hormone therapy, and previous cancers. We hope to report the investigation soon.

Acknowledgements We thank The Thailand Research Fund (TRF) and The Office of Higher Education Commission for the financial support of the research through RMU5080010.

References

1. Cheng HD, Shan J, Ju W, Guo Y, Zhang L (2010) Automated Breast Cancer Detection and Classification Using Ultrasound Images: A Survey. In: Pattern Recognition, vol 43, pp 299–317
2. Huang CL, Liao HC, Chen MC (2008) Predication Model Building and Feature Selection with Support Vector Machines in Breast Cancer Diagnosis. Expert Systems with Applications. vol 34, pp 578–587
3. Locked J (1999) Microsoft Bayesian Network: Basics of Knowledge Engineering, Microsoft Technical Report
4. Maskery SM, Hu H, Hooke J, Shriver CD, Liebman MN (2008) A Bayesian Derived Network of Breast Pathology Co-occurrence. In: Journal of Biomedical Informatics, vol 41, pp 242–250
5. Nicandro CR, Héctor GAM, Humberto CC, Luis ANF, Rocío EBM (2007) Diagnosis of breast cancer using Bayesian networks: A case study. Computers in Biology and Medicine. vol 37, pp 1553–1564
6. Wang XH, Zheng B, Good WF, King JL, Chang YH (1999) Computer-assisted diagnosis of breast cancer using a data-driven Bayesian belief network. In: International Journal of Medical Informatics, vol 54(2), pp 115–126
7. Wongthanavasu S, Lursinap C (2004) A 3-D CA-based Edge Operator for 3-D Images. In: Proceedings of the 11th IEEE Int. Conf. Image Processing (IEEEICIP 2004), pp 235–238
8. Wongthanavasu S, Sadananda R (2000) Pixel-level Edge Detection Using a Cellular Automata-Based Model. In: Proceedings of Int. Conf. Advances in Intelligent Systems: Theory and Applications (59), IOS Press, The Netherlands, pp 343–351
9. Wongthanavasu S, Tangvoraphonkchai V (2007) Cellular Automata-based Algorithm and Its Application in Medical Images Processing. In: Proceedings of the 14th IEEE Int. Conf. Image Processing (IEEE-ICIP 2007), pp III–41–44
10. Wongthanavasu S, Tanvoraphonkchai V (2008) Cellular Automata-based Identification of The Pectoral Muscle in Mammograms. In: Proceedings of the 3rd Int. Conf. Biomedical Engineering (ISBME 2008), pp 294–298