RULE-DISCOVERY BASED COMPUTER AIDED DETECTION FOR BREAST CANCER DIAGNOSIS

ADCHARA CHAROENSUP 4936795 EGTI / M

M.Sc. (TECHNOLOGY OF INFORMATION SYSTEM MANAGEMENT)

THESIS ADVISORY COMMITTEE: SONGPOL ONGWATTANAKUL, Ph.D., CHOLATIP WIRATKEPUN, M.D., WARAKORN CHAROENSUK, Ph.D.

ABSTRACT

The current trend in breast cancer incidence is on the rise. At present, there are several tools for breast cancer screening and diagnosis. Mammography has been an effective method for diagnosis which has been capable of reducing breast cancer mortality rate by up to 35%. In general, the radiologist diagnoses breast cancer by screening for abnormalities, tumors, and calcifications, derived from the interpretation of the mammogram. The recent availability of the digital mammogram has revealed the opportunity to conduct Knowledge Discovery in Database (KDD) research in this field. The main objective of this research was to use data mining tools such as the decision tree to generate rules governed by features extracted from mammogram images. The decision tree automatically converts the feature based data into rules that can be interpreted as pieces of knowledge. The feature extraction for this research was performed by a human expert. Two methods of feature selection were tested, the ReliefF attribute evaluation and Consistency Subset Evaluation (CNS). The C4.5 decision tree algorithm was used to generate the rules. Two methods of pruning were implemented, Error Based Pruning (EBP) and Reduced Error Pruning (REP). From the results it was shown that all models provide sufficient accuracy and good performance, except the CNS with REP method.

KEY WORDS: RULE-DISCOVERY / COMPUTER AIDED DETECTION / BREAST CANCER DIAGDNOSIS

124 pages

การตรวจจับด้วยคอมพิวเตอร์โดยใช้การค้นพบกฎสำหรับการวินิจฉัยโรคมะเร็งเต้านม
RULE-DISCOVERY BASED COMPUTER AIDED DETECTION FOR BREAST CANCER DIAGNOSIS

อัจฉรา เจริญทรัพย์ 4936795 EGTI / M

วท.ม.(เทคโนโลยีการจัดการระบบสารสนเทศ)

คณะกรรมการที่ปรึกษาวิทยานิพนธ์: ทรงพล องค์วัฒนกุล, Ph.D., ชลทิพย์ วิรัตกพันธ์, Ph.D., วรากร เจริญสุข, Ph.D.,

บทคัดย่อ

โรคมะเร็งเต้านมเป็นโรคมะเร็งที่พบในหญิงส่วนใหญ่ทั่วโลกและมีแนวโน้มเพิ่มขึ้นปัจจุบันมีเครื่องมือในการตรวจหามะเร็งหลากหลายวิธี แต่วิธีแมมโมแกรมพิสูจน์แล้วว่าสามารถลดอัตราการเสียชีวิตด้วยโรคมะเร็งเต้านมได้ 20-35% ดังนั้นในปัจจุบันนี้การวินิจฉัยมะเร็งเต้านมเบื้องต้นใช้วิธีแมมโมแกรมอย่างแพร่หลาย โดยทั่วไปแพทย์ทำการวินิจฉัยโรคมะเร็งเต้านมจากการพิจารณาก้อนเนื้อและการก่อตัวของแคลเซียมโดยแปลผลจากภาพแมมโมแกรมภายในเต้านม ด้วยเหตุนี้จึงมีการศึกษาการจำแนกของผู้ป่วยที่สงสัยว่าเป็นมะเร็งในเต้านม และผู้ที่ปรกติที่แปลผลมาจากภาพแมมโมแกรมเพื่อสนับสนุนการวินิจฉัยโรคของแพทย์ โดยใช้การค้นพบกฎด้วยวิธีการตัดสินใจแบบต้นไม้(Decision tree) ซึ่งเป็นหนึ่งเทคนิคของการทำเหมืองข้อมูล(Data Mining) การตัดสินใจแบบต้นไม้นี้จะทำการเปลี่ยนข้อมูลให้กลายเป็นกฎ ซึ่งกฎทั้งหลายที่เกิดจากข้อมูลสามารถตีความเป็นองค์ความรู้ได้ งานวิจัยนี้ได้ทำการแปลงภาพแมมโมแกรมไปเป็นชุดของฟีเจอร์ต่างๆ โดยใช้ตามนุษย์ทำการคัดเลือกฟีเจอร์สองวิธีคือ ReliefF Attribute Evaluation (ReliefF) และ Consistency Subset Evaluation (CNS) สร้างการตัดสินใจแบบต้นไม้โดยใช้อัลกอรึธึม C4.5 สุดท้ายทำการตัดกิ่งต้นไม้สองวิธีคือ Error Based Pruning (EBP) และ Reduced Error Pruning (REP) ผลการทดลองแสดงว่าทุกโมเดลให้ประสิทธิภาพดี ยกเว้นโมเดลที่ใช้วิธี CNS และ REP

124 หน้า