

บทคัดย่อ

งานวิจัยนี้นำเสนอขั้นตอนวิธีเพื่อการจำแนกชนิดของผู้ป่วยโรคกล้ามเนื้ออ่อนแรง และโรคมะเร็งเต้านม ซึ่งเป็นโรคที่ยากต่อการวินิจฉัยโดยแพทย์ ดังนั้นการนำความรู้ในด้านการทำเหมืองข้อมูลมาใช้เพื่อสร้างตัวจำแนกประกอบกับขั้นตอนวิธีการรวมตัวจำแนกจึงเป็นวิธีการที่น่าจะทำให้การจำแนกผู้ป่วยมีประสิทธิภาพที่ดีขึ้นผู้วิจัยจึงได้นำเสนอขั้นตอนวิธีการสกัดลักษณะเด่นจากข้อมูลผู้ป่วยซึ่งประกอบด้วย ข้อมูลอาการป่วยและข้อมูลผลการตรวจเลือด โดยใช้ขั้นตอนวิธีต้นไม้ตัดสินใจเป็นเครื่องมือในการประยุกต์ใช้ หลังจากนั้นผู้วิจัยได้คิดค้นขั้นตอนวิธีที่สามารถสร้างฟังก์ชันความเป็นสมาชิกได้อย่างอัตโนมัติโดยไม่ต้องใช้มนุษย์ในการสร้างอีกต่อไป นอกจากนี้ผู้วิจัยยังได้คิดค้นขั้นตอนวิธีการรวมผลจำแนก และทำการทดลองโดยใช้มาตรวัดสามตัวได้แก่ Sensitivity, Specificity และ Accuracy โดยทำการเฉลี่ยประสิทธิภาพของการจำแนกแบบ 10 Fold Cross Validation ซึ่งผลการทดลองพบว่าขั้นตอนวิธีที่ผู้วิจัยเสนอสามารถให้ประสิทธิภาพในการจำแนกประเภทผู้ป่วยได้ดีกว่าวิธีดั้งเดิม

นอกจากนี้ผู้วิจัยได้ศึกษาผลกระทบที่อาจเกิดขึ้นจากการสร้างกราฟฟังก์ชันความเป็นสมาชิกแบบอัตโนมัติ ซึ่งได้แก่ ตำแหน่งจุดตัดกราฟ รูปกราฟที่แตกต่างกัน ได้แก่ รูปสามเหลี่ยม รูปประฆังคว่ำ และแบบเกาส์เซียน

Abstract

This research project proposes a new algorithm for medical diagnosis in two domains which are Chronic Fatigue Syndrome and Breast Cancer. These diseases are complex and make difficulty for physician to make diagnosis. Therefore, Data Mining technique is needed in order to build a set of classifiers with the fusion method in order to provide efficient classification result.

We propose a new feature extraction process with make use of decision tree as a tool to extract a set of attributes that are suitable for classification. We create two classifiers from different feature sets which are blood test from laboratory and symptom test. Our next contribution is a new algorithm that can automatically generate a set of membership functions used for fuzzy logic. We also propose a new fusion algorithm in order to integrate the classification results obtained from different classifiers. The performance measurements used in the experiments are Sensitivity, Specificity and Accuracy, respectively. The classification performances are average through 10 Fold Cross Validation. The experimental results show that fusion technique give the best performance compared to each single classifier.

Furthermore, we study the effect of membership functions which are automatically generated by our algorithm. The graph intersection and different forms of membership function (triangular, bell-shape, Gaussian) are studied as well.