

รายงานการวิจัยฉบับสมบูรณ์

การจำแนกประเภทข้อมูลโรคมะเร็งจากข้อมูลไมโครอาร์เรย์ของดีเอ็นเอ  
ด้วยเทคนิค Deep Belief Network

Classification of Cancer using DNA Microarray Data  
with Deep Belief Network Technique

คณะผู้วิจัย

ผู้ช่วยศาสตราจารย์ ดร.ภัทร อัยรักษ์

นางสาววรรณิภา แซ่ลิ้ม

โครงการวิจัยนี้ได้รับทุนสนับสนุนจากเงินรายได้มหาวิทยาลัยสงขลานครินทร์  
ประจำปีงบประมาณ 2555 รหัสโครงการ SCI550109S-0

## บทคัดย่อ

ที่ผ่านมากการวินิจฉัยและจำแนกประเภทข้อมูลของโรคมะเร็งโดยใช้ไมโครอาร์เรย์เป็นปัญหาสำคัญในการศึกษาทางด้านเทคนิคปัญญาประดิษฐ์ ความท้าทายของการวิเคราะห์ไมโครอาร์เรย์คือการเลือกจำนวนของยีนที่เกี่ยวข้องที่สุดที่ตอบสนองต่อการเปลี่ยนแปลงของเซลล์ โดยวัตถุประสงค์ของการศึกษานี้ คือ ศึกษาการคัดเลือกยีนและการจำแนกประเภทของข้อมูล ด้วยเทคนิค Deep Belief Network (DBN) ที่เรียกว่า Quadratic Deep Belief Network (QDBN) Network (DBN) จากการศึกษาพบว่าการใช้ QDBN ให้ความแม่นยำในการจำแนกประเภทข้อมูลเซลล์มะเร็งของโรคลูคีเมียกว่าวิธีการอื่นๆ เช่น DBN, k-nearest neighbor (KNN), Naive Bayes (NB) และ support vector machine (SVM) นอกจากนี้ วิธีการ QDBN, KNN และ SVM ยังให้ความแม่นยำสูงสุดเท่าๆ กันในการจำแนกประเภทข้อมูลเซลล์มะเร็งต่อมน้ำเหลือง (lymphoma) ในขณะที่วิธีการ KNN ให้ความแม่นยำสูงสุดในข้อมูลประเภทมะเร็งต่อมลูกหมาก

## Abstract

In recent years, cancer classification based on microarray data is well known to be difficult in the field of computational intelligence techniques. One of a challenging issue is to select a reasonable number of the most relevant gene. This work aims to investigate the combination of feature selection and classification methods. We propose a novel classifier based on Deep Belief Network (DBN) called Quadratic Deep Belief Network (QDBN). QDBN applies a novel deep architecture to combine the general capability of DBN with quadratic discriminant. The experimental results showed that the QDBN method outperforms the existing methods such as DBN, k-nearest neighbor (KNN), Naive Bayes (NB) and support vector machine (SVM), in terms of the prediction accuracy on leukemia dataset. QDBN, KNN and SVM reach the same highest accuracy on lymphoma dataset but KNN provides the best performance on prostate cancer dataset.