

การแบ่งกลุ่มย่อยตามลักษณะทางคลินิกรูปแบบใหม่โดยอาศัยปัญญาประดิษฐ์ ร่วมกับการวิเคราะห์โอมิกส์แบบบูรณาการหลายระดับ เพื่อระบุกลุ่มตัวแทนสารบ่งชี้ทางชีวภาพสำหรับภาวะอหิวาต์ซิมสเปกตรัม



วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต
สาขาวิชาชีวเคมีคลินิกและอนุทางการแพทย์ ภาควิชาเคมีคลินิก
คณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2563
ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Novel Artificial Intelligence-Assisted Subgrouping of Clinical Presentation and
Integrated Multi-Omics Analyses in Identification of Biomarker Candidates for Autism
Spectrum Disorder



A Thesis Submitted in Partial Fulfillment of the Requirements
for the Degree of Master of Science in Clinical Biochemistry and Molecular Medicine

Department of Clinical Chemistry
FACULTY OF ALLIED HEALTH SCIENCES
Chulalongkorn University

Academic Year 2020

Copyright of Chulalongkorn University

| | |
|---------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| หัวข้อวิทยานิพนธ์ | การแบ่งกลุ่มย่อยตามลักษณะทางคลินิกรูปแบบใหม่โดยอาศัยปัญญาประดิษฐ์ ร่วมกับการวิเคราะห์โอมิกส์แบบบูรณาการหลายระดับ เพื่อระบุกลุ่มตัวแทนสารบ่งชี้ทางชีวภาพสำหรับภาวะอหิวาต์ซิมสเปกตรัม |
| โดย | น.ส.กัลยภัตสรร์ หัสเสมอ |
| สาขาวิชา | ชีวเคมีคลินิกและอนุทางการแพทย์ |
| อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก | ผู้ช่วยศาสตราจารย์ ดร.เทวฤทธิ์ สระระชนะ |

คณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย อนุมัติให้หัวข้อวิทยานิพนธ์ฉบับนี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรมหาบัณฑิต

..... คณะบดีคณะสหเวชศาสตร์
(รองศาสตราจารย์ ดร.ปาลณี อัมรานนท์)

คณะกรรมการสอบวิทยานิพนธ์

..... ประธานกรรมการ
(ผู้ช่วยศาสตราจารย์ ดร.เทวิน เทนคำเนา)

..... อาจารย์ที่ปรึกษาวิทยานิพนธ์หลัก
(ผู้ช่วยศาสตราจารย์ ดร.เทวฤทธิ์ สระระชนะ)

..... กรรมการ
(อาจารย์ ดร.ศิริพร แสงสุธรรม)

..... กรรมการ
(อาจารย์ ดร.ณัฐชยา แหวนวงศ์)

..... กรรมการภายนอกมหาวิทยาลัย
(อาจารย์ ดร.มลฤดี สุขประสารทรัพย์)

6176652537 : MAJOR CLINICAL BIOCHEMISTRY AND MOLECULAR MEDICINE

KEYWORD: Autism spectrum disorder, Artificial Intelligence, Proteome, Transcriptome, Health Informatics,
Machine learning, Biomarkers

Kanlayaphat Hussem : Novel Artificial Intelligence-Assisted Subgrouping of Clinical Presentation and Integrated Multi-Omics Analyses in Identification of Biomarker Candidates for Autism Spectrum Disorder.

Advisor: Asst. Prof. TEWARIT SARACHANA, Ph.D.

Autism spectrum disorder (ASD) is an early onset neurodevelopmental disorder with a high degree of heterogeneity. Nowadays, there is still no biomarker for the diagnosis and subtyping for ASD, but there is evidence that early diagnosis and treatment are important for helping people with ASD in gaining necessary functions and behaviors in life. Current international standard ASD diagnosis is based on complicated and time-consuming questionnaires. Artificial intelligence (AI) analysis has been utilized to analyze various types of complicated big data. However, the research to demonstrate its applications on analyzing clinical presentations associated with ASD is still limited. In this study, we, therefore, sought to utilize AI tools to subtype 85 ASD individuals into subgroups based on a total of 123 items from ASD-related clinical phenotypic data. After getting ASD subgroups and the lists of subgroup members, the subgrouping models were compared to the effectiveness of different AI analyses. Moreover, we identified biomarker candidates for each ASD clinical subtype by using multi-omics approaches by performed transcriptome and proteome profiling analysis of lymphoblastoid cell lines (LCLs) derived from ASD patients. We conducted the proteomic analysis using 2D gel electrophoresis (2D-GE), followed by mass spectrometry, to identify proteins that are significantly disrupted in each ASD subgroup and then integrated the transcriptome profile and the proteome profiles to determine the correlation and association between the disrupted genes and proteins. Furthermore, we validated the gene expression by real-time qPCR and protein expression by Western blot analysis. Interestingly, we found that 'K-means clustering' which divided the ASD patients into 4 subgroups provided the best efficiency. And when we created the model for subtyping based on K-means clustering, the results show that 'Random forest' with 10 trees couple with 'forward selection' analyses gave the most accurate prediction result (100%) and using only 5 ADI-R items to divided ASD patients into subgroups. In addition, from the proteome profiling analysis of LCLs in ASD and ASD subgroups using 2D-GE, we selected 19 significant differentially expressed protein spots with at least two different intensity values to identify the protein by LC-MS/MS. Among those was HNRNPA2B1 which was increased in ASD with non-severe language impairments subgroup. While in ASD with severe language impairments subgroup found decreased expression of HNRNPA2B1 MDH2 DLD and MSN. In conclusion, this study suggests that the use of AI tools to subtyping ASD patients based on clinical manifestations could reduce the diversity of ASD symptoms in transcriptome and proteome profiling analysis, as well as more effective investigation of specific proteins in ASD subgroups which would serve as candidate genes or proteins for biomarker development research for ASD subtyping in the future.

Field of Study: Clinical Biochemistry and Molecular Student's Signature

Medicine

Academic Year: 2020 Advisor's Signature

กิตติกรรมประกาศ

วิทยานิพนธ์ฉบับนี้ดำเนินการจนสำเร็จลุล่วงไปได้ด้วยดี เนื่องด้วยความเมตตาและกรุณาของ ผู้ช่วยศาสตราจารย์ ดร. เทวฤทธิ์ สาระชนะ อาจารย์ผู้ซึ่งเป็นที่ปรึกษาวิทยานิพนธ์หลัก ที่ได้มอบความรู้ คำปรึกษา แนะนำแนวทางในการดำเนินงานวิจัย แนะนำแนวทางในการแก้ไขปัญหาและข้อบกพร่องที่เป็นประโยชน์ต่อการพัฒนางานวิจัย ตลอดจนแนะนำโอกาสทางการศึกษาที่เป็นประโยชน์ต่อนิสิต นอกจากนี้อาจารย์ยังติดตามความก้าวหน้าในการดำเนินงานวิจัยของนิสิตอย่างสม่ำเสมอ รวมถึงให้การสนับสนุนในด้านต่าง ๆ งานวิจัยนี้สำเร็จลุล่วงไปได้ด้วยดี จึงขอกราบขอบพระคุณอาจารย์เป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. เหวิน เทนคำเนาวัลย์ ประธานกรรมการสอบป้องกันวิทยานิพนธ์ อาจารย์ ดร. ศิริพร แสงสุธรรม อาจารย์ ดร. ณัฐชยา แหวนวงศ์ และอาจารย์ ดร. มลฤดี สุขประสารทรัพย์ ที่กรุณาสละเวลามาเป็นคณะกรรมการในการสอบป้องกันวิทยานิพนธ์ในครั้งนี้

ขอกราบขอบพระคุณ ศาสตราจารย์ ดร. Valerie W. Hu จากมหาวิทยาลัยจอร์จ วอชิงตัน กรุงวอชิงตัน ดีซี ประเทศสหรัฐอเมริกา ที่ให้ความอนุเคราะห์เซลล์ไลน์ลิ้มโพลาสต์ (LCLs) และโปรแกรมชีวสารสนเทศ IPA ในการศึกษา

ขอกราบขอบพระคุณ The Autism Genetic Resource Exchange (AGRE) Repository ที่ให้ความอนุเคราะห์ข้อมูลลักษณะอาการทางคลินิกของเซลล์ไลน์ลิ้มโพลาสต์ (LCLs) ตลอดจนขอกราบขอบพระคุณผู้ป่วยและผู้ปกครองของผู้ป่วยทุกท่านในการศึกษาครั้งนี้

ขอกราบขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร. เหวิน เทนคำเนาวัลย์ ที่สนับสนุนงานทางด้านสารเคมีที่ใช้ในการทำงานด้านโปรตีโอมิกส์ อาจารย์ ดร.ศิริพร แสงสุธรรม ที่กรุณาให้คำแนะนำและให้ความรู้เกี่ยวกับเทคนิคการทำงานด้านโปรตีโอมิกส์ รวมทั้งคุณอาจารย์ในภาควิชาเคมีคลินิกและในคณะสหเวชศาสตร์ที่อบรมสั่งสอน ให้ความรู้ คำปรึกษา คำแนะนำ และความช่วยเหลือที่เป็นประโยชน์อย่างยิ่งในการศึกษาวิจัย ตลอดจนขอขอบพระคุณบุคลากรทุกท่านในคณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่กรุณาอำนวยความสะดวกในด้านสถานที่ และอุปกรณ์บางส่วนที่จำเป็นต่อการศึกษานี้

ขอกราบขอบขอบคุณทุนอุดหนุนการศึกษาระดับบัณฑิตศึกษาจากบัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย เพื่อเฉลิมฉลองวโรกาสที่พระบาทสมเด็จพระเจ้าอยู่หัวทรงเจริญพระชนมายุครบ 72 พรรษา และ ทุน 90 ปี จุฬาลงกรณ์มหาวิทยาลัย กองทุนรัชดาภิเษกสมโภช

ขอขอบคุณพี่ ๆ เพื่อน ๆ และน้อง ๆ ในหน่วยปฏิบัติการวิจัยประสาทวิทยาศาสตร์เชิงระบบของอิมูโนและภาวะทางจิตเวช SYstems Neuroscience of Autism and PSychiatric disorders (SYNAPS Research Unit) และในหลักสูตรบัณฑิตศึกษา ตลอดจนหลักสูตรวิทยาศาสตร์บัณฑิต คณะสหเวชศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ที่ได้ให้ความช่วยเหลือ คำแนะนำ และคำปรึกษาที่เป็นประโยชน์ต่องานวิจัย

และสุดท้ายนี้ ขอกราบขอบพระคุณ คุณพ่อ คุณแม่ และครอบครัว ที่คอยส่งเสริมและให้การสนับสนุนในทุกด้าน และเป็นกำลังใจให้การทำงานวิจัยมาโดยตลอด ความดีของการศึกษาและคุณค่าของวิทยานิพนธ์นี้ ผู้วิจัยขออุทิศแด่บูรพาจารย์และผู้มีพระคุณทุกท่าน

กัลยภัสสรร์ ทัศสม

สารบัญ

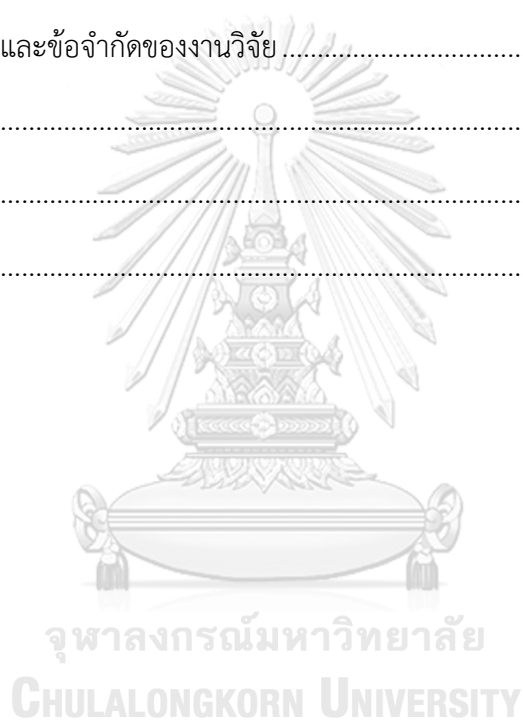
| | หน้า |
|---------------------------------------------------------------|------|
| บทคัดย่อภาษาไทย..... | ค |
| บทคัดย่อภาษาอังกฤษ..... | ง |
| กิตติกรรมประกาศ..... | จ |
| สารบัญ..... | ฉ |
| สารบัญตาราง..... | ฉ |
| สารบัญรูปภาพ..... | ช |
| บทที่ 1 | 1 |
| บทนำ..... | 1 |
| 1.1 ความสำคัญและที่มาของปัญหาการวิจัย..... | 1 |
| 1.2 คำถามงานวิจัย..... | 7 |
| 1.3 สมมติฐานงานวิจัย | 7 |
| 1.4 วัตถุประสงค์ของงานวิจัย | 7 |
| 1.5 กรอบแนวคิดงานวิจัย..... | 9 |
| 1.6 ระเบียบวิธีการวิจัย..... | 10 |
| 1.7 ขอบเขตและข้อจำกัดของงานวิจัย..... | 11 |
| 1.8 ประโยชน์ของงานวิจัย..... | 11 |
| บทที่ 2 | 13 |
| เอกสารและงานวิจัยที่เกี่ยวข้อง | 13 |
| 2.1 โรคออทิซึมสเปกตรัม (Autism Spectrum Disorder; ASD)..... | 13 |
| 2.2 แบบทดสอบ ADI-R (Autism Diagnostic Interview-Revised)..... | 17 |
| 2.3 เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence)..... | 20 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| 2.4 อัลกอริทึมที่ใช้ในโปรแกรม RapidMiner และการประเมินประสิทธิภาพ..... | 26 |
| 2.5 การศึกษาทรานสคริปโตมและโปรตีโอมในผู้ป่วยออทิซึมสเปกตรัม (Transcriptomic and Proteomic in ASD)..... | 41 |
| 2.6 เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines) | 43 |
| 2.7 ข้อมูลไมโครอาร์เรย์ (Microarray data)..... | 45 |
| 2.8 แมสสเปกโทรเมตรี (Mass spectrometry)..... | 50 |
| บทที่ 3 | 53 |
| วิธีดำเนินการวิจัย | 53 |
| <u>ตอนที่ 1</u> การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของคนไข้ออทิซึมสเปกตรัม โดยอาศัยการเรียนรู้ของเครื่องมือ (Machine learning) ด้วยโปรแกรม RapidMiner..... | 53 |
| <u>ตอนที่ 1.1</u> การรวบรวมข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัมจากฐานข้อมูล..... | 53 |
| <u>ตอนที่ 1.2</u> การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม RapidMiner..... | 54 |
| <u>ตอนที่ 1.3</u> การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner..... | 57 |
| <u>ตอนที่ 2</u> การวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs)..... | 61 |
| <u>ตอนที่ 2.1</u> การรวบรวมข้อมูลทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากฐานข้อมูล..... | 62 |
| <u>ตอนที่ 2.2</u> การวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ได้จากฐานข้อมูล | 62 |
| <u>ตอนที่ 3</u> การวิเคราะห์รูปแบบโปรตีโอมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid cell lines; LCLs) | 63 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| <u>ตอนที่ 3.1</u> การคัดเลือกสิ่งตัวอย่าง..... | 63 |
| <u>ตอนที่ 3.2</u> การเพาะเลี้ยงและการเก็บเซลล์..... | 64 |
| <u>ตอนที่ 3.3</u> การสกัดโปรตีนจากเซลล์ วัดปริมาณโปรตีน และตรวจสอบคุณภาพโปรตีน..... | 66 |
| <u>ตอนที่ 3.4</u> การแยกโปรตีนด้วยเทคนิค 2D-GE และการวิเคราะห์ความแตกต่างของจุดโปรตีน..... | 69 |
| <u>ตอนที่ 3.5</u> การย่อยแผ่นเจล สกัดเปปไทด์ และวิเคราะห์เปปไทด์ด้วย LC-MS/MS..... | 72 |
| <u>ตอนที่ 4</u> การศึกษาความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงลิมโฟบลาสต์จากผู้ป่วยโรคอหิวาต์ซีมสเปกตรัม และการยืนยันผล | 74 |
| <u>ตอนที่ 4.1</u> การวิเคราะห์ข้อมูลทางชีวสารสนเทศจากรายชื่อโปรตีนที่พบว่ามีแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยอหิวาต์ซีมสเปกตรัมและผู้ป่วยอหิวาต์ซีมสเปกตรัมกลุ่มย่อย | 75 |
| <u>ตอนที่ 4.2</u> การวิเคราะห์ความสัมพันธ์ของกลุ่มโปรตีนที่มีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยอหิวาต์ซีมสเปกตรัมและผู้ป่วยอหิวาต์ซีมสเปกตรัมกลุ่มย่อย | 76 |
| <u>ตอนที่ 4.3</u> การยืนยันผลการแสดงออกของยีนและโปรตีนที่สนใจโดยใช้วิธี Real-time PCR และ Western blot..... | 81 |
| <u>ตอนที่ 4.4</u> การวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงออกของโปรตีนกับคะแนนจากแบบทดสอบ ADI-R..... | 89 |
| บทที่ 4 | 90 |
| ผลการศึกษาวิจัย..... | 90 |
| 4.1 ผลการวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยอหิวาต์ซีมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม RapidMiner..... | 90 |
| 4.1.1 ผลการหาจำนวนกลุ่มย่อยที่เหมาะสมด้วยวิธี Elbow method | 90 |
| 4.1.2 ผลการเปรียบเทียบประสิทธิภาพในการแบ่งกลุ่มระหว่าง Unsupervised machine learning tools วิธี K-means และ K-medoids | 91 |

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.1.3 ผลการจัดกลุ่มผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อย และผลเปรียบเทียบลักษณะ อาการทางคลินิกของผู้ป่วย | 92 |
| 4.2 ผลการวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ ของเครื่องมือแบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner95 | |
| 4.2.1 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Artificial Neural Network (ANN) | 95 |
| 4.2.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม k-Nearest Neighbor (k- NN)..... | 101 |
| 4.2.3 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Naiive Bayes | 107 |
| 4.2.4 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Decision tree..... | 111 |
| 4.2.5 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Random forest..... | 115 |
| 4.2.6 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM)..... | 121 |
| 4.2.7 ผลการคัดเลือกโมเดลจากการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised Machine Learning) ที่ดีที่สุด | 126 |
| 4.3 ผลการวิเคราะห์รูปแบบทรานسكريปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วย ออทิ ซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย..... | 131 |
| 4.4 ผลการวิเคราะห์รูปแบบโปรตีโอมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึม สเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย | 132 |
| 4.5 ผลการวิเคราะห์ข้อมูลทางชีวสารสนเทศจากรายชื่อโปรตีนที่พบว่ามี การแสดงออกผิดปกติใน เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่ม ย่อย | 140 |
| 4.6 ผลการศึกษาความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติใน เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคออทิซึมสเปกตรัม..... | 143 |
| 4.7 ผลการยืนยันระดับการแสดงออกของยีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟ บลาสต์จากผู้ป่วยออทิซึมสเปกตรัม | 145 |

| | |
|----------------------------------------------------------------------------------------------------------------------------|-----|
| 4.8 ผลการยืนยันระดับการแสดงผลของโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยอหิวาต์ชิเมสเปกตรัม | 147 |
| 4.9 ผลการวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงผลของโปรตีนกับคะแนนจากแบบทดสอบ ADI-R | 149 |
| บทที่ 5 | 155 |
| อภิปรายผล | 155 |
| บทที่ 6 | 164 |
| สรุปผล ข้อเสนอแนะ และข้อจำกัดของงานวิจัย | 164 |
| บรรณานุกรม | 167 |
| ภาคผนวก | 178 |
| ประวัติผู้เขียน | 238 |



สารบัญตาราง

หน้า

| | |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| <u>ตารางที่ 1</u> รายการการศึกษารูปแบบทรานสคริปโตมาจากเซลล์เม็ดเลือดของผู้ป่วยอหิวาต์ชิมสเปกตรัม | 77 |
| <u>ตารางที่ 2</u> รายการการศึกษารูปแบบโปรตีนในเลือดของผู้ป่วยอหิวาต์ชิมสเปกตรัม | 79 |
| <u>ตารางที่ 3</u> ร้อยละความเข้มข้นของเจลที่เหมาะสมสำหรับการแยกมวลโมเลกุลของโปรตีน | 85 |
| <u>ตารางที่ 4</u> ขนาดมวลโมเลกุลและชนิดของสิ่งตัวอย่างสำหรับ Internal control ที่นิยมใช้ในการทำ Western blot | 88 |
| <u>ตารางที่ 5</u> ผลลัพธ์ประสิทธิภาพในการแบ่งกลุ่มด้วยวิธี K-means และ K-medoids เมื่อ k=4 | 92 |
| <u>ตารางที่ 6</u> ผลการแบ่งกลุ่มผู้ป่วยอหิวาต์ชิมสเปกตรัมออกเป็น 4 กลุ่มย่อยด้วยวิธี K-means | 93 |
| <u>ตารางที่ 7</u> ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 100 150 200 500 และ 1000 ในกระบวนการวิเคราะห์แบบ Split validation | 96 |
| <u>ตารางที่ 8</u> ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 75 100 และ 200 ในกระบวนการวิเคราะห์แบบ Backward elimination | 97 |
| <u>ตารางที่ 9</u> ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 100 500 1000 และ 2000 ในกระบวนการวิเคราะห์แบบ Forward selection | 98 |
| <u>ตารางที่ 10</u> เปรียบเทียบประสิทธิภาพของโมเดล Artificial Neural Network (ANN) จากทั้ง 3 กระบวนการวิเคราะห์ | 100 |
| <u>ตารางที่ 11</u> ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Split validation | 102 |
| <u>ตารางที่ 12</u> ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Backward elimination | 103 |
| <u>ตารางที่ 13</u> ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Forward selection | 104 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| ตารางที่ 14 เปรียบเทียบประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) จากทั้ง 3 กระบวนการวิเคราะห์..... | 106 |
| ตารางที่ 15 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการ วิเคราะห์แบบ Split validation | 107 |
| ตารางที่ 16 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการ วิเคราะห์แบบ Backward elimination..... | 108 |
| ตารางที่ 17 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการ วิเคราะห์แบบ Forward selection..... | 109 |
| ตารางที่ 18 เปรียบเทียบประสิทธิภาพของโมเดล Naïve Bayes หรือ Naïve Bayes kernel จากทั้ง 3 กระบวนการวิเคราะห์ | 110 |
| ตารางที่ 19 ประสิทธิภาพของโมเดล Decision tree ในกระบวนการวิเคราะห์แบบ Split validation Backward elimination และ Forward selection..... | 112 |
| ตารางที่ 20 เปรียบเทียบประสิทธิภาพของโมเดล Decision tree จากทั้ง 3 กระบวนการวิเคราะห์ | 114 |
| ตารางที่ 21 ประสิทธิภาพของโมเดล Random forest ที่ 1 5 10 25 50 และ 100 ต้น ใน กระบวนการวิเคราะห์แบบ Split validation..... | 116 |
| ตารางที่ 22 ประสิทธิภาพของโมเดล Random forest ที่ 5 10 25 50 และ 100 ต้น ในกระบวนการ วิเคราะห์แบบ Backward elimination..... | 117 |
| ตารางที่ 23 ประสิทธิภาพของโมเดล Random forest ที่ 1 5 10 25 50 และ 100 ต้น ใน กระบวนการวิเคราะห์แบบ Forward selection | 118 |
| ตารางที่ 24 เปรียบเทียบประสิทธิภาพของโมเดล Random forest จากทั้ง 3 กระบวนการวิเคราะห์ | 120 |
| ตารางที่ 25 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Split validation | 121 |
| ตารางที่ 26 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Backward elimination . | 122 |

| | |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| ตารางที่ 27 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Forward selection..... | 123 |
| ตารางที่ 28 เปรียบเทียบประสิทธิภาพของโมเดล Support Vector Machine (SVM) จากทั้ง 3 กระบวนการวิเคราะห์..... | 124 |
| ตารางที่ 29 เปรียบเทียบประสิทธิภาพของโมเดล A Library of Support Vector Machine (LibSVM) จากทั้ง 3 กระบวนการวิเคราะห์..... | 125 |
| ตารางที่ 30 เปรียบเทียบประสิทธิภาพของโมเดลที่ดีที่สุดจากแต่ละอัลกอริทึม | 127 |
| ตารางที่ 31 จำนวน Transcripts ที่มีความแตกต่างอย่างมีนัยสำคัญ และจำนวนโปรตีนที่ผลิตโดย Transcripts ในแต่ละกลุ่มย่อยเมื่อเปรียบเทียบกับกลุ่มควบคุม | 132 |
| ตารางที่ 32 รายชื่อจุดโปรตีนที่มีการแสดงออกผิดปกติอย่างมีนัยสำคัญในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย | 134 |
| ตารางที่ 33 หน้าที่ทางชีวภาพและลักษณะอาการที่เกี่ยวข้องกับโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัม..... | 137 |
| ตารางที่ 34 ผลการทำนายโรคและหน้าที่ทางชีววิทยาของกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย..... | 141 |
| ตารางที่ 35 ผลการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อยกับกลุ่มยีนจากกลุ่มตัวอย่างเดียวกัน และกลุ่มยีนและโปรตีนจากการศึกษาอื่น | 144 |
| ตารางที่ 36 ผลการวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงออกของโปรตีน MDH2 DLD ROA2 MSN และ ANXA5 กับคะแนนจากแบบทดสอบADI-R..... | 150 |
| ตารางที่ 37 คะแนน ADI-R เฉลี่ยของผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่ใช้ในการวิเคราะห์สหสัมพันธ์เทียบกับระดับการแสดงออกของโปรตีน..... | 236 |

สารบัญรูปภาพ

| | หน้า |
|-----------------------------------------------------------------------------|------|
| รูปที่ 1 Unsupervised machine learning | 22 |
| รูปที่ 2 Supervised machine learning | 23 |
| รูปที่ 3 Reinforcement Machine Learning..... | 24 |
| รูปที่ 4 K-means clustering | 27 |
| รูปที่ 5 K-medoids and K-means clustering | 28 |
| รูปที่ 6 Determining the optimal number of clusters using elbow method..... | 29 |
| รูปที่ 7 K-Nearest Neighbor (k-NN)..... | 30 |
| รูปที่ 8 Support Vector Mahine (SVM)..... | 31 |
| รูปที่ 9 Random forest..... | 34 |
| รูปที่ 10 Artificial Neural Network (ANN) | 37 |
| รูปที่ 11 Lymphoblastoid Cell Lines (LCLs)..... | 44 |
| รูปที่ 12 cDNA microarray | 45 |
| รูปที่ 13 TOF Mass spectrometer | 51 |
| รูปที่ 14 Column formatting in RapidMiner..... | 55 |
| รูปที่ 15 Confusion Matrix..... | 60 |
| รูปที่ 16 DNA RNA and protein isolation by GENEzol reagent..... | 67 |
| รูปที่ 17 ตัวอย่าง MASCOT search results | 74 |
| รูปที่ 18 Hypergeometric distribution calculator | 80 |
| รูปที่ 19 มวลโมเลกุลของโปรตีน Beta-actin และ GAPDH..... | 89 |
| รูปที่ 20 The elbow method for determining number of cluster | 90 |
| รูปที่ 21 Sample size calculation and comparison | 91 |

| | |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| รูปที่ 22 Heatmaps of ADI-R scores..... | 94 |
| รูปที่ 23 Random Forest (FS)_tree1 | 128 |
| รูปที่ 24 Random Forest (FS)_tree2 | 128 |
| รูปที่ 25 Random Forest (FS)_tree3 | 128 |
| รูปที่ 26 Random Forest (FS)_tree4 | 129 |
| รูปที่ 27 Random Forest (FS)_tree5 | 129 |
| รูปที่ 28 Random Forest (FS)_tree6 | 129 |
| รูปที่ 29 Random Forest (FS)_tree7 | 130 |
| รูปที่ 30 Random Forest (FS)_tree8 | 130 |
| รูปที่ 31 Random Forest (FS)_tree9 | 130 |
| รูปที่ 32 Random Forest (FS)_tree10..... | 131 |
| รูปที่ 33 จุดโปรตีนที่แสดงออกผิดปกติอย่างมีนัยสำคัญในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย | 136 |
| รูปที่ 34 ผลการทำนายเครือข่ายทางชีวภาพของโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์..... | 142 |
| รูปที่ 35 Box plot แสดงผลการแสดงออกของยีนจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับกลุ่มควบคุม | 146 |
| รูปที่ 36 Bar graph แสดงระดับการแสดงออกของโปรตีนจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับกลุ่มควบคุม | 148 |
| รูปที่ 37 Heatmaps แสดงความสัมพันธ์ของระดับการแสดงออกของโปรตีนจากเซลล์เพาะเลี้ยง ชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัมกับคะแนนจากแบบทดสอบ ADI-R | 154 |
| รูปที่ 38 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่มควบคุม..... | 232 |
| รูปที่ 39 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ ของกลุ่ม ASD (M+S subgroup)..... | 233 |

รูปที่ 40 โพรตีนที่ถูกวินิจฉัยด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ ของกลุ่ม M subgroup..... 234

รูปที่ 41 โพรตีนที่ถูกวินิจฉัยด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่ม S subgroup..... 235



บทที่ 1

บทนำ

1.1 ความสำคัญและที่มาของปัญหาการวิจัย

โรคออทิซึมสเปกตรัม (Autism Spectrum Disorder; ASD) เป็นโรคที่จัดอยู่ในกลุ่มโรคที่เกี่ยวข้องเนื่องกับความผิดปกติทางด้านพัฒนาการของระบบประสาท (Neurodevelopmental disorder) ซึ่งปัจจุบันยังไม่ทราบสาเหตุอย่างแน่ชัด แต่มีผลงานวิจัยจำนวนมากที่เชื่อกันว่ามีสาเหตุจากความผิดปกติในหลายปัจจัยด้วยกัน ทั้งปัจจัยทางด้านพันธุกรรม เช่น เกิดความผิดปกติที่บริเวณโครโมโซม 15q11-q13 (Nurmi et al., 2003) 8p11.23 และ 10q24.32 (Autism Spectrum Disorders Working Group of The Psychiatric Genomics, 2017) และปัจจัยทางด้านสิ่งแวดล้อม เช่น การขาดวิตามินดีในขณะตั้งครรภ์ (Vuillermot, Luan, Meyer, & Eyles, 2017) ผลของ Cytokine interleukin-6 ในขณะตั้งครรภ์ (Parker-Athill & Tan, 2010) และผลจากการได้รับสารบิสฟีนอลเอ (BPA) ขณะตั้งครรภ์ (Thongkorn et al., 2019) ผู้ป่วยโรคออทิซึมสเปกตรัมจะมีลักษณะความผิดปกติทางด้านพฤติกรรม โดยมีความบกพร่องด้านการสื่อสารและการเข้าสังคม (Impairments in communication and social interaction) หรือมีพฤติกรรมย้ำคิดย้ำทำ กระทำพฤติกรรมบางอย่างซ้ำ ๆ และมีความสนใจจำกัดเฉพาะเรื่อง (Repetitive behaviors and restricted interests) ซึ่งคำจำกัดความนี้เป็นไปตามนิยามของคู่มือการวินิจฉัยและสถิติสำหรับความผิดปกติทางจิตฉบับล่าสุดในปี ค.ศ.2013 (Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition; DSM-5) โดยสมาคมจิตแพทย์อเมริกัน (American Psychiatric Association; APA)

จากการศึกษาทางระบาดวิทยาโดยศูนย์ควบคุมและป้องกันโรคของประเทศสหรัฐอเมริกา (Center for Disease Control and Prevention; CDC) พบว่าอัตราความชุกของโรคออทิซึมสเปกตรัมสูงถึง 1 ใน 54 ราย ในปี ค.ศ.2020 (Center for Disease Control and Prevention, 2020) และในประเทศไทยจากการสำรวจความชุกของโรคออทิซึมสเปกตรัม พบว่ามีผู้ป่วยมากถึง 6 คนต่อจำนวนประชากร 1,000 คน ในปี พ.ศ.2555 (Chantayanon, 2016) จากความชุกของผู้ป่วยออทิซึมสเปกตรัมในประเทศไทยที่กล่าวมาข้างต้น คาดว่าจำนวนผู้ป่วยที่พบนั้นต่ำกว่าความเป็นจริงเนื่องจากผู้ป่วยโรคออทิซึมสเปกตรัมจำนวนมากในประเทศไทย ยังขาดโอกาสที่จะได้รับการวินิจฉัยที่ถูกต้องและเหมาะสม นอกจากนี้จากการศึกษาความชุกของโรคออทิซึมสเปกตรัมในประเทศ

สหรัฐอเมริกา ในปี ค.ศ.2020 ยังพบว่าอุบัติการณ์ของโรคอหิวาต์ชิเมสเปกตรัม มักพบในเด็กผู้ชายมากกว่าเด็กผู้หญิง โดยในเด็กผู้ชายจะมีความชุกของโรค 1 ใน 34 คน ขณะที่ความชุกในเด็กผู้หญิงมี 1 ใน 144 คน (Center for Disease Control and Prevention, 2020) แต่ปัจจุบันยังไม่สามารถสรุปทฤษฎีที่สามารถอธิบายความชุกที่เกิดขึ้นได้ แต่มีหลายงานวิจัยด้วยกันที่รายงานถึงปัจจัยทางพันธุกรรมที่เกี่ยวข้องกับความผิดปกติที่เกิดขึ้นบริเวณโครโมโซม X (Gong et al., 2008; Marco & Skuse, 2006) ซึ่งอาจมีความเกี่ยวข้องกับสาเหตุที่พบโรคนี้ได้ของเด็กผู้ชายมากกว่าเด็กผู้หญิง

ผู้ป่วยโรคอหิวาต์ชิเมสเปกตรัม นั้น มีการแสดงออกของอาการและความรุนแรงที่มีความหลากหลายเป็นอย่างมาก ซึ่งความหลากหลายและซับซ้อนของโรคนี้อาจมีสาเหตุมาจากปัจจัยร่วมกันระหว่างปัจจัยทางด้านพันธุกรรมและปัจจัยทางด้านสิ่งแวดล้อมดังที่ได้กล่าวไปในข้างต้น อย่างไรก็ตาม แม้จะยังไม่ทราบสาเหตุของโรคอย่างแน่ชัด แต่มีหลักฐานที่สนับสนุนว่าการได้รับการวินิจฉัยและการรักษาที่รวดเร็วเป็นสิ่งสำคัญที่จะช่วยให้ผู้ป่วยอหิวาต์ชิเมสเปกตรัมสามารถได้รับการเรียนรู้และพัฒนาพฤติกรรมที่จำเป็นในการดำรงชีวิตได้ (Barbaresi, Katusic, & Voigt, 2006) ในปัจจุบันมีนักวิจัยหลายท่านที่ได้ทำการศึกษาหาพื้นที่เกี่ยวข้องกับโรคอหิวาต์ชิเมสเปกตรัม ซึ่งทำให้มีการค้นพบรายชื่อยีนจำนวนมากคาดว่าเกี่ยวข้องกับโรคอหิวาต์ชิเมสเปกตรัม แต่ยังคงมีข้อจำกัดเนื่องจากไม่ทราบแน่ชัดว่าในทุกยีนที่ค้นพบนี้จะมีการแสดงออกที่ส่งผลให้เกิดอาการของโรคหรือไม่ ดังนั้นจึงเป็นเหตุให้นักวิจัยบางท่านเริ่มศึกษาการแสดงออกของยีนและโปรตีนที่สัมพันธ์กับโรคอหิวาต์ชิเมสเปกตรัม ซึ่งจะสะท้อนถึงการทำงานของยีนและอาจจะนำไปสู่การค้นพบสารชีวโมเลกุลที่มีความสัมพันธ์กับโรค

จากงานวิจัยที่ผ่านมาได้มีการศึกษาการแสดงออกของยีนและโปรตีนในผู้ป่วยโรคอหิวาต์ชิเมสเปกตรัมเป็นจำนวนมาก โดยจากการศึกษาการแสดงออกของยีนในผู้ป่วยอหิวาต์ชิเมสเปกตรัม พบว่าผู้ป่วยมีการแสดงออกของยีนบางยีนที่ความแตกต่างอย่างมีนัยสำคัญจากคนปกติ แต่อย่างไรก็ตามกลับมีข้อจำกัดในการศึกษา คือการแสดงออกของยีนที่พบนี้มีความแตกต่างกันในผู้ป่วยแต่ละราย (Gregg et al., 2008) และยังไม่ทราบแน่ชัดว่า RNA ที่พบจะถูกนำไปเป็นแม่แบบในการสังเคราะห์โปรตีนหรือไม่ เนื่องจากกระบวนการควบคุมการแสดงออกของยีน (Post-transcriptional regulation) เช่น กระบวนการ miRNA เป็นต้น รวมถึงยีนที่ค้นพบโดยส่วนใหญ่แล้วนั้นมักจะเป็นยีนที่มีความเกี่ยวข้องกับกระบวนการและวิถีหลักทางชีววิทยาของมนุษย์ เช่น การควบคุมการแสดงออกของยีน การพัฒนาการของสมอง การสื่อสารระหว่างเซลล์ประสาท และการสร้างและการทำงานของเซลล์ประสาท เป็นต้น ซึ่งอาจมีความเกี่ยวข้องกับพยาธิสภาพที่พบในผู้ป่วย (Duda, Haber,

Daniels, & Wall, 2017) ดังนั้นจึงเป็นเหตุให้มีการศึกษาโปรตีน เพื่อใช้ในการเชื่อมโยงข้อมูลที่ได้จากศึกษาการแสดงออกของของยีน เพื่อที่จะได้ทราบอย่างแน่ชัดถึงยีนที่เกี่ยวข้องกับโรคอย่างแท้จริง รวมถึงโปรตีนที่เกี่ยวข้องกับโรค เช่น ผลงานของ Dr. Jiang Xie ที่ได้ทำการศึกษาระดับของไซโตไคน์ (Cytokine) (Xie et al., 2017) แต่การค้นพบนี้อาจจะยังไม่ครอบคลุมที่จะนำมาเชื่อมโยงกับอาการของโรคได้อย่างเพียงพอ ทำให้มีการศึกษาโปรตีโอมในช่วงเวลาต่อมา เนื่องจากการศึกษาโปรตีโอมไม่ใช่เป็นเพียงการศึกษาหาความผิดปกติของโปรตีนตัวใดตัวหนึ่งแต่เป็นเครื่องมือที่ใช้ในการศึกษาชีววิทยาเชิงระบบที่มีประสิทธิภาพเป็นอย่างมาก ซึ่งจะช่วยให้เห็นภาพรวมของโปรตีนที่อาจเชื่อมโยงกับอาการที่แสดงออกของโรคและอาจมีประโยชน์ในการค้นหาสารบ่งชี้ทางชีวภาพเพื่อใช้ในการคัดกรองและวินิจฉัยผู้ป่วยในทางคลินิกได้ สำหรับผลงานวิจัยทางด้านโปรตีโอม มีผลงานวิจัยที่ศึกษาโปรตีโอมในซีรัมของผู้ป่วยโรคอหิวาต์แอฟริกาโดยอาศัยเทคนิคแมสสเปกโตรเมตรี เพราะเชื่อว่าจะพบความแตกต่างระหว่างโปรตีนของผู้ป่วยกับคนปกติ โดยพบว่ามีโปรตีนบางตัวที่แตกต่างจากคนปกติอย่างมีนัยสำคัญ (Corbett et al., 2007) แต่โปรตีนที่ค้นพบเหล่านี้คงยังมีข้อจำกัดเนื่องจากไม่ใช่ผู้ป่วยทุกคนที่พบว่ามีปริมาณของโปรตีนเหล่านี้ที่ผิดปกติไปอีกเช่นเดียวกัน ซึ่งจากงานวิจัยของ Dr.Valerie Hu และ Dr.Mara Steinberg ที่ได้นำข้อมูลผลจากการวินิจฉัยผู้ป่วยโรคอหิวาต์แอฟริกาตามลักษณะอาการที่แสดงออกด้วยแบบทดสอบ ADI-R มาทำการจัดกลุ่มย่อยโดยอาศัยเครื่องมือทางสถิติ ได้แก่ K-means เข้ามาช่วย ทำให้ได้กลุ่มย่อยที่มีลักษณะการแสดงออกของโรคใกล้เคียงกัน (Hu & Steinberg, 2009) การแยกกลุ่มตามลักษณะที่แสดงออกนี้อาจจะช่วยลดข้อจำกัดของการแสดงออกของยีนและโปรตีนที่พบในผู้ป่วยโรคอหิวาต์แอฟริกาที่หลากหลาย ซึ่งเป็นผลที่ทำให้ยีนหรือโปรตีนที่ค้นพบไม่อาจใช้บ่งบอกว่าผิดปกติในผู้ป่วยทุกคน แต่หากทำการจัดกลุ่มย่อย ยีนหรือโปรตีนที่ค้นพบก็จะสามารถบ่งบอกถึงความผิดปกติที่สัมพันธ์กับลักษณะการแสดงออกในเฉพาะกลุ่มนั้นได้ และจากงานวิจัยที่ได้มีการหาการแสดงออกของยีนในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs) ของผู้ป่วยโรคอหิวาต์แอฟริกาจากการแบ่งกลุ่มย่อยตามที่ได้กล่าวไปข้างต้น พบความแตกต่างอย่างมีนัยสำคัญเมื่อเทียบกับคนปกติ นอกจากนี้ยังพบการแสดงออกของยีนบางยีนที่แตกต่างกันในระหว่างกลุ่มย่อย (Hu et al., 2009) โดยจากงานวิจัยที่ได้กล่าวไปนี้ จึงเป็นข้อสนับสนุนที่ทำให้การจัดกลุ่มย่อยเพื่อลดความหลากหลายตามลักษณะอาการที่แสดงออกของโรคอหิวาต์แอฟริกาอาจเป็นแนวทางที่จะสามารถนำไปสู่การวิเคราะห์ทรานสคริปโตมและโปรตีโอม รวมถึงยีนที่จำเพาะและมีความสัมพันธ์กับอาการที่แสดงออก

ของโรคออทิซึมสเปกตรัมได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งจะเป็นประโยชน์ในการระบุสารบ่งชี้ทางชีวภาพและการศึกษากลไกในระดับโมเลกุลของแต่ละโรคออทิซึมสเปกตรัมกลุ่มย่อยได้

ในปัจจุบันยังไม่มีทั้งสารบ่งชี้สำหรับการวินิจฉัยโรคออทิซึมสเปกตรัมและสำหรับการจัดผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อยเพื่อลดความหลากหลายของอาการของผู้ป่วย การวินิจฉัยโรคออทิซึมสเปกตรัม จะอาศัยเพียงการสังเกตพฤติกรรมและทำการซักประวัติผู้ป่วยที่มีพัฒนาการผิดปกติเท่านั้น โดยจะทำได้ก็ต่อเมื่อผู้ป่วยมีอายุ 3 ปี เป็นอย่างน้อย เนื่องจากมักเป็นช่วงอายุที่อาการของโรคแสดงออกให้เห็นอย่างชัดเจน ซึ่งนี่เป็นหนึ่งในข้อจำกัดที่ทำให้ผู้ป่วยได้รับการวินิจฉัยล่าช้าและไม่ได้รับการรักษาอย่างทันทั่วถึง หรือบางครั้งหากไม่ใช้ผู้เชี่ยวชาญเป็นผู้วินิจฉัยก็จะทำให้การวินิจฉัยผิดพลาดได้ โดยหนึ่งในวิธีการวินิจฉัยโรคออทิซึมสเปกตรัมในประเทศสหรัฐอเมริกา คือการใช้แบบทดสอบ ADI-R (The Autism Diagnostic Interview-Revised) ของมหาวิทยาลัย University of Michigan แบบทดสอบ ADI-R ยังใช้เป็นวิธี Gold standard ในการวินิจฉัยโรคออทิซึมสเปกตรัม ซึ่งใช้แพร่หลายทั่วโลก โดยแบบทดสอบ ADI-R ประกอบด้วยรูปแบบการทดสอบทั้งหมด 123 หัวข้อ เหมาะสำหรับเด็กในช่วงอายุ 12-47 เดือนขึ้นไป (Kim & Lord, 2012) ซึ่งคำถามในการสัมภาษณ์จะเกี่ยวกับประวัติของผู้ป่วย พฤติกรรม พัฒนาการทางด้านภาษา พัฒนาการทางด้านสังคม และอาการทางคลินิกต่าง ๆ ที่เกี่ยวข้อง เช่น การทำร้ายตัวเอง ความก้าวร้าว เป็นต้น (Rutter, LeCouteur, & Lord, 2003) แบบทดสอบ ADI-R สามารถแบ่งออกได้เป็น 4 ส่วนการทดสอบ ได้แก่ ทดสอบการสื่อสาร (Communication) ทดสอบการเข้าสังคม (Social interaction) ทดสอบการกระทำพฤติกรรมบางอย่างซ้ำ ๆ (Repetitive behaviors) และทดสอบทักษะพิเศษ (Savant skills) ซึ่งมีเป้าหมายเพื่อประเมินบุคคลที่สงสัยว่าเป็นโรคออทิซึมสเปกตรัม และอาจนำไปสู่การแบ่งกลุ่มย่อยของโรคออทิซึมสเปกตรัมได้ อย่างไรก็ตาม แบบทดสอบ ADI-R มีข้อเสียคือ จำเป็นจะต้องใช้ผู้เชี่ยวชาญที่ผ่านการฝึกอบรมในการวินิจฉัย และยังมีจำนวนข้อในการสัมภาษณ์ที่มีปริมาณมาก ทำให้ต้องใช้เวลามากถึง 90-150 นาที ต่อผู้ป่วย 1 ราย โดยคะแนนจะถูกกำหนดให้อยู่ในระดับ 0-3 ซึ่งตัวเลขที่เพิ่มสูงขึ้นนั้นหมายถึงความรุนแรงของอาการ (Charman & Gotham, 2013) อีกทั้งในปัจจุบันยังไม่มีการรักษาที่มีประสิทธิภาพและจำเพาะต่อโรคโดยตรง

เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence technology) เป็นเทคโนโลยีที่ทั่วโลกให้ความสนใจมากขึ้นเรื่อย ๆ ในปัจจุบัน เทคโนโลยีปัญญาประดิษฐ์ หมายความว่าทุกอย่างที่ทำให้คอมพิวเตอร์สามารถทำตัวเสมือนมนุษย์ได้ โดยที่มีการแทรกแซงจากมนุษย์น้อยที่สุด (Mierswa,

2017) โดยมีการนำไปประยุกต์ใช้กับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big data analysis) ในหลากหลายสาขาด้วยกัน อันประกอบไปด้วย การประมวลภาษาธรรมชาติ (Natural language processing) การรู้จำคำพูด (Speech recognition) การเรียนรู้ของเครื่องมือ (Machine learning) การประมวลผลภาพ (Image processing) วิทยาการหุ่นยนต์ (Robotics) เครือข่ายเชิง-ความหมาย (Semantic web) เทคโนโลยีการศึกษา (Intelligent computer-aided instruction) ระบบผู้เชี่ยวชาญ (Expert systems) และสาขาปัญญาประดิษฐ์อื่น ๆ ซึ่งการเรียนรู้ของเครื่องมือ (Machine learning) คือการที่ระบบสามารถเรียนรู้แบบอัตโนมัติและปรับปรุงการจัดการระบบได้จากประสบการณ์ โดยไม่ต้องตั้งโปรแกรมไว้อย่างชัดเจน แต่มีการมุ่งเน้นการพัฒนาโปรแกรมคอมพิวเตอร์ที่สามารถเข้าถึงข้อมูลและใช้ข้อมูลนั้น ๆ ได้เอง โดยการเรียนรู้ของเครื่องมือสามารถแบ่งออกได้เป็น 2 ประเภทหลัก คือ การเรียนรู้แบบมีผู้สอน (Supervised machine learning) และการเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) ซึ่งมีหลายงานวิจัยด้วยกันที่นำการเรียนรู้ของเครื่องมือมาใช้ในการศึกษาต่าง ๆ รวมถึงการศึกษาที่เกี่ยวข้องกับโรคอหิวาต์ซิมสเปกตรัม เช่น ใช้การเรียนรู้ของเครื่องมือศึกษาเกี่ยวกับการเคลื่อนไหวของผู้ป่วยอหิวาต์ซิมสเปกตรัม (Crippa et al., 2015) ใช้การเรียนรู้ของเครื่องมือในการจำแนกโรคอหิวาต์ซิมสเปกตรัมกับโรคสมาธิสั้น (Duda et al., 2017) และใช้การเรียนรู้ของเครื่องมือในการลดจำนวนแบบทดสอบในการวินิจฉัยผู้ป่วย (Kosmicki, Sochat, Duda, & Wall, 2015) ดังนั้นเทคโนโลยีปัญญาประดิษฐ์ โดยเฉพาะอย่างยิ่งการเรียนรู้ของเครื่องมืออาจช่วยในการวิเคราะห์รูปแบบของข้อมูลและช่วยในการตัดสินใจที่จะเป็นประโยชน์ในการแบ่งผู้ป่วยอหิวาต์ซิมสเปกตรัมออกเป็นกลุ่มย่อยเพื่อลดความหลากหลายจากลักษณะอาการของผู้ป่วยได้ อย่างไรก็ตามการศึกษาเกี่ยวกับการใช้เทคโนโลยีปัญญาประดิษฐ์ในการแบ่งผู้ป่วยอหิวาต์ซิมสเปกตรัมออกเป็นกลุ่มย่อยยังคงมีข้อจำกัดอยู่

ดังนั้นในการศึกษานี้คณะผู้วิจัยต้องการใช้เทคโนโลยีปัญญาประดิษฐ์ในการแบ่งผู้ป่วยอหิวาต์ซิมสเปกตรัมออกเป็นกลุ่มย่อยตามลักษณะอาการทางคลินิกโดยอาศัยข้อมูลจากการวินิจฉัยผู้ป่วยด้วยแบบทดสอบ ADI-R ก่อนทำการศึกษารูปแบบการแสดงออกของทรานสคริปโตม (Transcriptome) และโปรตีโอม (Proteome) ในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs) ที่พัฒนามาจากเม็ดเลือดขาวของผู้ป่วยอหิวาต์ซิมสเปกตรัม เพื่อนำมาทำการเปรียบเทียบหาความแตกต่างจากคนปกติอย่างมีนัยสำคัญ ศึกษาปฏิสัมพันธ์และความสัมพันธ์ของยีนและโปรตีนที่พบว่ามีผลผิดปกติในผู้ป่วยอหิวาต์ซิมสเปกตรัมและผู้ป่วยอหิวาต์

สเปกตรัมกลุ่มย่อย โดยข้อมูลที่นำมาใช้ในการศึกษาทรานสคริปโตมเป็นข้อมูลซึ่งได้จากผลการวิเคราะห์ทรานสคริปโตมจาก Microarray ของ Dr.Valerie Hu และคณะ โดยคณะผู้วิจัยได้นำเครื่องมือทางสถิติและชีวสารสนเทศเข้ามาใช้ในการเปรียบเทียบหาความแตกต่างจากคนปกติ และการศึกษาโปรตีโอมศาสตร์ เทคนิค Two Dimensional Gel Electrophoresis (2D-GE) และแมสสเปกโตรเมตรี (Mass spectrometry) เข้ามาช่วยในการศึกษา โดย 2D-GE นั้นเป็นเทคนิคที่แยกโปรตีนโดยอาศัยทั้งมวลโมเลกุลและประจุของโปรตีน ซึ่งพบว่ามีความจำเพาะมากกว่าการแยกด้วยเทคนิคที่แยกโปรตีนเพียงมิติเดียวอย่าง SDS-PAGE พบได้ในผลงานวิจัยของ Dr. Armand G. Ngounou Wetie และคณะ ที่ได้ใช้เทคนิค 2D-GE และ LC-MS/MS ในการศึกษาโปรตีนในน้ำลายของผู้ป่วยโรคอหิวาต์แอฟริกา โดยเทคนิคดังกล่าวช่วยเพิ่มความจำเพาะและความไวในการวิเคราะห์มากขึ้น (A. G. Ngounou Wetie et al., 2015) แต่ยังคงมีข้อจำกัดเนื่องจากน้ำลายเป็นสารคัดหลั่งที่มีการปนเปื้อนได้ง่ายทำให้ผลที่ได้อาจจะเกิดข้อผิดพลาดได้ นอกจากนี้รูปแบบการแสดงออกของโปรตีนที่พบอาจจะไม่ครอบคลุมเท่ากับการใช้สิ่งตรวจอย่างเช่นโลหิต จากการศึกษาในครั้งนี้ คณะผู้วิจัยพบความแตกต่างอย่างมีนัยสำคัญของรูปแบบการแสดงออกของทรานสคริปโตม (Transcriptome) และโปรตีโอม (Proteome) ในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs) ที่พัฒนามาจากเม็ดเลือดขาวของผู้ป่วยอหิวาต์แอฟริกาและผู้ป่วยอหิวาต์แอฟริกาเล็กน้อยเมื่อเทียบกับคนปกติ นอกจากนี้ยังพบปฏิสัมพันธ์และความสัมพันธ์ของโปรตีนที่พบว่ามีคามผิดปกติในผู้ป่วยอหิวาต์แอฟริกาและผู้ป่วยอหิวาต์แอฟริกาเล็กน้อย ซึ่งนำไปสู่การค้นพบสารบ่งชี้ทางชีวภาพที่อาจสามารถใช้ในการตรวจวินิจฉัยโรคอหิวาต์แอฟริกาได้อย่างจำเพาะและมีประสิทธิภาพได้ในอนาคต

1.2 คำถามงานวิจัย

1. ปัญญาประดิษฐ์สามารถช่วยแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยตามอาการทางคลินิกได้หรือไม่ และแต่ละรูปแบบให้ผลแตกต่างกันอย่างไร
2. ผู้ป่วยแต่ละกลุ่มย่อยที่ถูกแบ่งโดยปัญญาประดิษฐ์มีรูปแบบทรานสคริปโตมและโปรตีโอมแตกต่างจากคนปกติหรือไม่
3. กลุ่มยีนและโปรตีนที่มีระดับการแสดงออกผิดปกติในผู้ป่วยโรคออทิสซึมสเปกตรัม มีปฏิสัมพันธ์ระหว่างกันและมีความเกี่ยวข้องกับชีวพยาธิสภาพที่พบในผู้ป่วยโรคออทิสซึมสเปกตรัมหรือไม่

1.3 สมมติฐานงานวิจัย

1. ปัญญาประดิษฐ์สามารถช่วยแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยตามอาการทางคลินิกโดยใช้ข้อมูลจากแบบวินิจฉัย ADI-R
2. ผู้ป่วยแต่ละกลุ่มย่อยที่ถูกแบ่งโดยปัญญาประดิษฐ์มีรูปแบบทรานสคริปโตมและโปรตีโอมแตกต่างจากคนปกติ และอาจจะมีลักษณะจำเพาะในแต่ละกลุ่มย่อย
3. กลุ่มยีนและโปรตีนที่มีระดับการแสดงออกผิดปกติในผู้ป่วยโรคออทิสซึมสเปกตรัม มีปฏิสัมพันธ์ระหว่างกันและมีความเกี่ยวข้องกับชีวพยาธิสภาพที่พบในผู้ป่วยโรคออทิสซึมสเปกตรัม

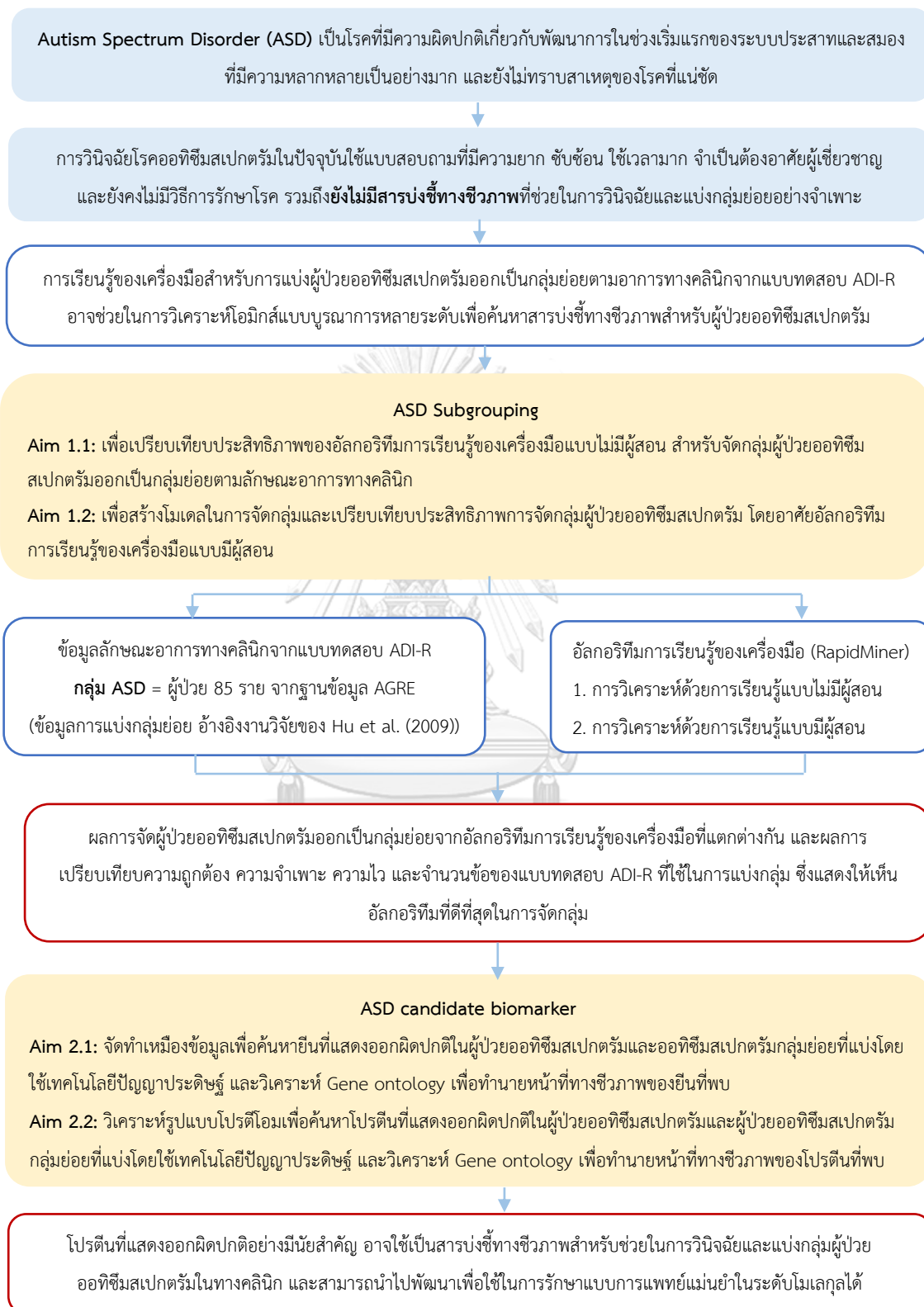
1.4 วัตถุประสงค์ของงานวิจัย

1. เพื่อทดลองแบ่งกลุ่มผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้ของเครื่องมือ (Machine learning) จากโปรแกรม RapidMiner โดยจะ
 - 1.1 เปรียบเทียบความสามารถหรือประสิทธิภาพในการแบ่งกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) ของเครื่องมือจากโปรแกรม RapidMiner
 - 1.2 สร้างโมเดลและเปรียบเทียบความสามารถหรือประสิทธิภาพของการแบ่งกลุ่มย่อยผู้ป่วยจากลักษณะอาการทางคลินิก โดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้แบบมีผู้สอน (Supervised machine learning) ของเครื่องมือจากโปรแกรม RapidMiner

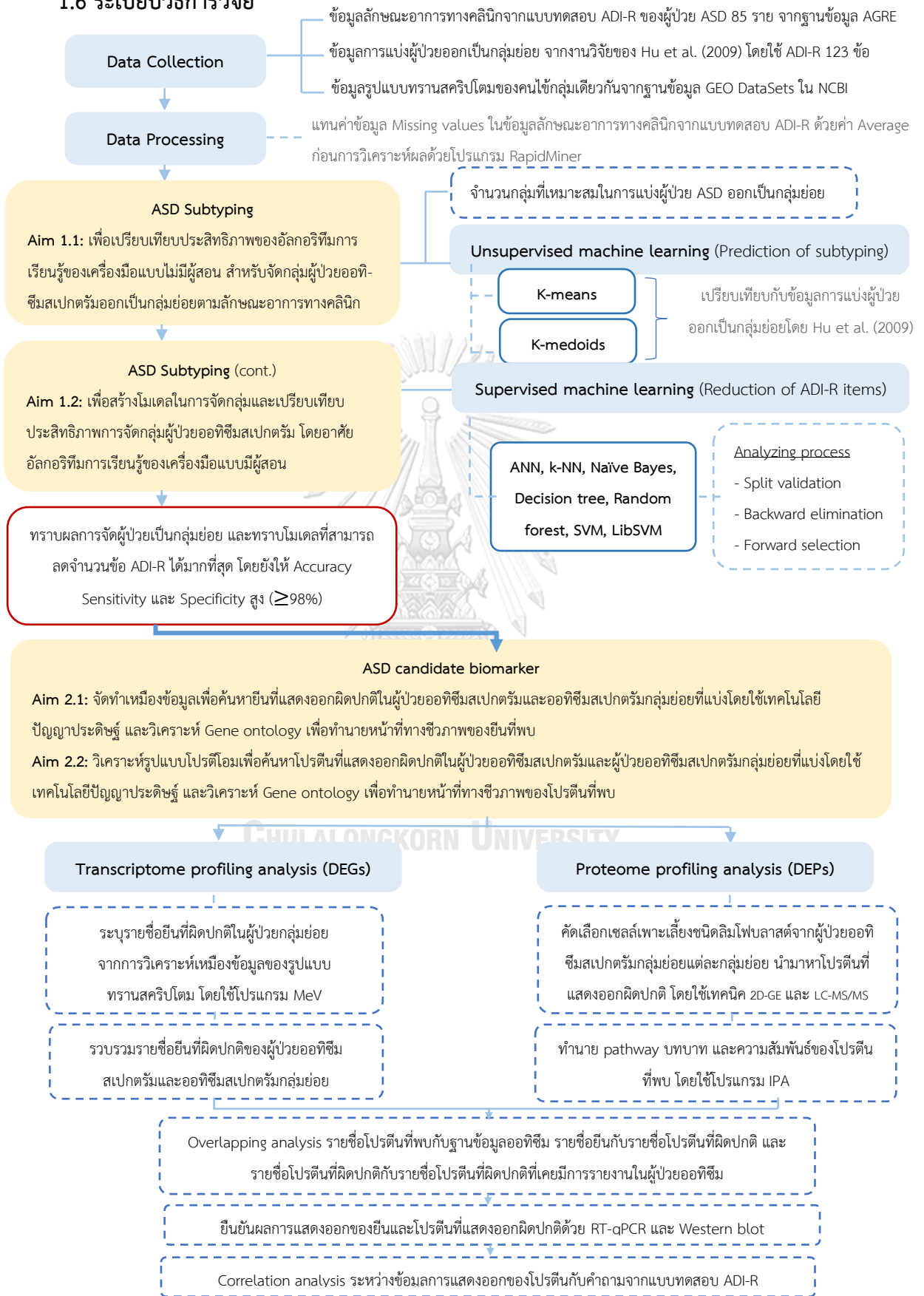
2. เพื่อศึกษารูปแบบทรานสคริปโตมและโปรตีโอมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิสซึมสเปกตรัมเปรียบเทียบกับรูปแบบการแสดงออกในเซลล์เพาะเลี้ยงที่มาจากคนปกติ และระบุกลุ่มยีนและโปรตีนที่มีระดับการแสดงออกผิดปกติในกลุ่มย่อยของโรคออทิสซึมสเปกตรัมที่แบ่งตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์
3. เพื่อศึกษาปฏิสัมพันธ์ (Interactome) ของกลุ่มยีนและโปรตีนที่มีระดับการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคออทิสซึมสเปกตรัม ทำนายหน้าที่ทางชีวภาพ และความสัมพันธ์กับชีวพยาธิสภาพที่พบในผู้ป่วยโรคออทิสซึมสเปกตรัม



1.5 กรอบแนวคิดงานวิจัย



1.6 ระเบียบวิธีการวิจัย



1.7 ขอบเขตและข้อจำกัดของงานวิจัย

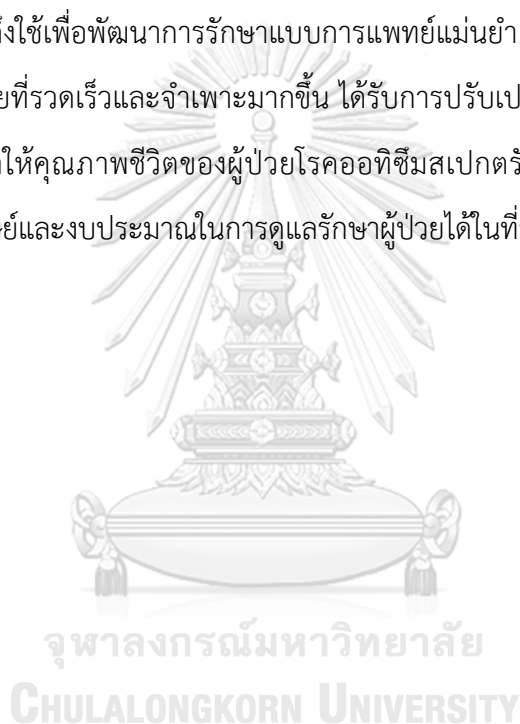
การศึกษานี้เป็นการวิจัยที่มีการออกแบบเป็น Case-control analytical study โดยคณะผู้วิจัยดำเนินการศึกษารูปแบบการแสดงของโปรตีนโพรตีโอมและอินเตอร์แอคโตมในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ Lymphoblastoid Cell Lines (LCLs) ที่พัฒนามาจากเซลล์จากเลือดของผู้ป่วยโรคออทิซึมสเปกตรัมและคนปกติเพศชาย การผลิตเซลล์เหล่านี้ดำเนินการโดย Autism Genetic Resource Exchange (AGRE) repository ซึ่งเป็นธนาคารชีวภาพที่รวบรวมสิ่งตัวอย่างและเซลล์สำหรับการศึกษาวิจัยเกี่ยวกับโรคออทิซึมสเปกตรัม ประเทศสหรัฐอเมริกา ซึ่งเซลล์เพาะเลี้ยงเหล่านี้ คณะผู้วิจัยได้รับความอนุเคราะห์จาก Dr.Valerie Hu มหาวิทยาลัยจอร์จทาวน์ กรุงวอชิงตัน ดีซี โดยเซลล์เพาะเลี้ยงจากผู้ป่วยจะถูกจำแนกออกเป็นกลุ่มย่อย (Subgroups) จำนวน 4 กลุ่ม และคัดเลือกเหลือเพียงกลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) จำนวน 9 ราย และกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroups; S subgroup) จำนวน 13 ราย โดยใช้เทคโนโลยีปัญญาประดิษฐ์ อาศัยการเรียนรู้ของคอมพิวเตอร์จากโปรแกรม RapidMiner มาแบ่งกลุ่มคนไข่ออกเป็นกลุ่มย่อย พิจารณาตามเกณฑ์ ADI-R และลักษณะอาการทางคลินิก ทั้งนี้คณะผู้วิจัยจะทำการศึกษาการแสดงออกของโปรตีนในช่วง pH 3-10 ระหว่างกลุ่มทดสอบและกลุ่มควบคุม โดยกำหนดจุดโปรตีนที่สนใจจะต้องมีค่า P-value < 0.05 จากนั้นทำการวิเคราะห์มวลของตัวอย่างโปรตีนที่สนใจด้วยเทคนิค LC-MS/MS นำผลที่ได้มาทำการวิเคราะห์เปรียบเทียบกับฐานข้อมูลและหาความสัมพันธ์ของโปรตีนที่สนใจ

ทั้งนี้งานวิจัยนี้ยังคงมีข้อจำกัดในเรื่องของกลุ่มตัวอย่างที่ไม่ใช่ประชากรคนไทย และยังมีข้อมูลที่ใช้ในการจัดกลุ่มด้วยเทคโนโลยีปัญญาประดิษฐ์ไม่มากนัก ส่งผลให้โมเดลการแบ่งกลุ่มที่ได้ อาจไม่ใช่ตัวแทนในการแบ่งกลุ่มของผู้ป่วยออทิซึมสเปกตรัมทั้งหมด แต่สามารถใช้เป็นแนวทางในการแบ่งกลุ่มผู้ป่วยออกเป็นกลุ่มย่อยในเบื้องต้นได้ ซึ่งจะเป็นประโยชน์ในการศึกษาเกี่ยวกับผู้ป่วยออทิซึมสเปกตรัมได้ต่อไปในอนาคต

1.8 ประโยชน์ของงานวิจัย

จากการที่โรคออทิซึมสเปกตรัมเป็นโรคที่ยังไม่ทราบสาเหตุของการเกิดโรคที่แน่ชัด อีกทั้งยังไม่มีสารบ่งชี้ หรือการรักษาที่จำเพาะต่อโรค ดังนั้นทางสาธารณสุขจึงมุ่งเน้นไปที่กระบวนการวินิจฉัย

ผู้ป่วย เพื่อดูแลโดยการปรับเปลี่ยนพฤติกรรม หรือสอนให้มีการเรียนรู้ทักษะที่จำเป็นและเป็นประโยชน์ต่อการใช้ชีวิตแก่ผู้ป่วย การศึกษาค้นคว้าครั้งนี้ทำให้เราสามารถแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยและได้โมเดลจากเทคโนโลยีปัญญาประดิษฐ์มาช่วยในการวินิจฉัยแยกผู้ป่วยออกเป็นกลุ่มย่อย และทำให้ทราบยีนและโปรตีนที่ผิดปกติในเซลล์ที่พัฒนามาจากเม็ดเลือดขาวของผู้ป่วยโรคออทิสซึมสเปกตรัม และเครือข่ายความสัมพันธ์ทางชีวภาพ (Interactome) ของยีนและโปรตีนเหล่านั้น ซึ่งอาจใช้เป็นตัวแทนสำหรับการศึกษาเชิงลึกในประชากรผู้ป่วย เพื่อพัฒนาเป็นสารบ่งชี้ทางชีวภาพ (Biomarker) หรือศึกษากลไกพยาธิสภาพในระดับชีวโมเลกุลของโรคออทิสซึมสเปกตรัมได้ต่อไปในอนาคต รวมถึงใช้เพื่อพัฒนาการรักษาแบบการแพทย์แม่นยำ (Targeted therapy) ส่งผลให้ผู้ป่วยได้รับการวินิจฉัยที่รวดเร็วและจำเพาะมากขึ้น ได้รับการปรับเปลี่ยนพฤติกรรม หรือการเรียนรู้ทักษะที่จำเป็น ส่งผลให้คุณภาพชีวิตของผู้ป่วยโรคออทิสซึมสเปกตรัมดีขึ้น และนำไปสู่การลดการสูญเสียทรัพยากรมนุษย์และงบประมาณในการดูแลรักษาผู้ป่วยได้ในที่สุด



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

2.1 โรคออทิซึมสเปกตรัม (Autism Spectrum Disorder; ASD)

“Autism” มีรากศัพท์มาจากภาษากรีก มาจากคำว่า Auto แปลว่า Self คือ การที่แยกตัวออกมาอยู่ตามลำพังในโลกของตัวเอง ซึ่งโรคออทิซึมสเปกตรัมนั้น เป็นโรคที่ความผิดปกติในด้านการพัฒนาการ ส่งผลให้ผู้ป่วยมีความผิดปกติทั้งด้านสังคม การสื่อสารและทางด้านพฤติกรรม ผู้ป่วยที่เป็นโรคออทิซึมสเปกตรัมนี้ โดยทั่วไปจะมีลักษณะที่ไม่แตกต่างจากคนปกติ เว้นเพียงแต่ผู้ป่วยโรคออทิซึมจะมีลักษณะการสื่อสาร การตอบสนอง พฤติกรรมและการเรียนรู้ที่แตกต่างจากคนทั่วไป โดยการเรียนรู้ ความคิดและการแก้ไขปัญหาของผู้ป่วยโรคนี้นี้จะมีตั้งแต่มีพรสวรรค์ที่จะสามารถเรียนรู้สิ่งต่างๆ ไปจนถึงมีความผิดปกติที่รุนแรงที่เรียนรู้สิ่งต่าง ๆ ได้น้อย ซึ่งทำให้ผู้ป่วยบางรายต้องการความช่วยเหลือเป็นอย่างมากในการดำเนินชีวิตประจำวัน (Center for Disease Control and Prevention, 2016)

ลักษณะอาการ

ผู้ป่วยโรคออทิซึมสเปกตรัมนั้นจะมีความบกพร่องของทักษะในด้านสังคม การแสดงออกของอารมณ์และการสื่อสาร ซึ่งพวกเขาจะมีแบบแผนพฤติกรรมเดิม ๆ ไม่ชอบการเปลี่ยนแปลง ผู้ป่วยออทิซึมสเปกตรัมในแต่ละรายก็จะมีการเรียนรู้ ความสนใจและการตอบสนองต่อสิ่งต่าง ๆ ที่แตกต่างกันออกไป ซึ่งลักษณะอาการของผู้ป่วยโรคนี้นี้จะแสดงออกตั้งแต่เด็กจนตลอดชีวิตของพวกเขา (Center for Disease Control and Prevention, 2016) โดยลักษณะอาการที่อาจจะบ่งชี้ได้แก่

1. ไม่ตอบสนองต่อการเรียกชื่อจนอายุ 12 เดือน
2. ไม่ให้ความสนใจกับสิ่งที่น่าสนใจ เช่น การไม่มองเครื่องบินที่บินอยู่เหนือตัวเอง
3. ไม่สนใจสิ่งที่คนอื่นให้ความสนใจ
4. มีปัญหาในการสร้างสัมพันธ์กับผู้อื่นหรือไม่สนใจที่จะมีปฏิสัมพันธ์กับผู้อื่น
5. ไม่สบสายตากับผู้อื่นและต้องการอยู่ลำพัง
6. มีปัญหาในการเข้าใจว่าคนอื่นรู้สึกอย่างไร หรือไม่สามารถบอกว่าคุณเองรู้สึกอย่างไร
7. ไม่ชอบที่จะกอดหรืออาจจะกอดเมื่อตัวผู้ป่วยต้องการเท่านั้น
8. เมื่อมีคนสื่อสารด้วยจะแสดงอาการไม่รู้ตัวแต่จะตอบสนองต่อเสียงอื่นแทน
9. มีความสนใจคนอื่นอย่างมากและไม่รู้วิธีที่จะพูดคุย เล่นหรือสร้างสัมพันธ์

10. มักจะพูดคำหรือประโยคตามคนที่มาพูดด้วยซ้ำ ๆ หรือจะพูดคำหรือประโยคซ้ำ ๆ แทนการสื่อสารแบบปกติ
11. มีปัญหาในการแสดงออกถึงความต้องการตัวเองโดยการใช้คำพูดหรือการแสดงอารมณ์แบบปกติ
12. ไม่สามารถเล่นเกมที่มีการสมมติบทบาทได้ เช่น การให้อาหารแก่ตุ๊กตา
13. ทำพฤติกรรมเดิมซ้ำ ๆ
14. มีปัญหาในการปรับตัวเมื่อชีวิตประจำวันเปลี่ยนแปลงไป
15. มีปฏิกริยาที่ผิดแปลกไปกับสิ่งของที่มีกลิ่น รสชาติ ความรู้สึกหรือเสียง
16. สูญเสียทักษะบางอย่างที่เคยมี เช่น การหยุดพูดคำที่เคยใช้

สาเหตุ

ในปัจจุบันยังไม่เป็นที่แน่ชัดในเรื่องของสาเหตุของการเกิดโรคออทิซึมสเปกตรัม แต่จากการศึกษานั้นคาดว่าน่าจะมีหลายสาเหตุสำหรับการเกิดโรคออทิซึมสเปกตรัม โดยแบ่งเป็น 2 กลุ่มใหญ่ ๆ (Center for Disease Control and Prevention, 2016) ได้แก่ สาเหตุทางด้านสิ่งแวดล้อม และสาเหตุทางด้านพันธุกรรมยกตัวอย่าง เช่น

1. สาเหตุด้านพันธุกรรม

มีนักวิทยาศาสตร์มากมายที่ทำการวิจัยจนเป็นที่ยอมรับว่าความผิดปกติของยีนมีผลต่อการเกิดโรคออทิซึมสเปกตรัม เช่น

- การพบว่าเด็กที่มีพี่น้องเป็นผู้ป่วยโรคออทิซึมสเปกตรัมจะมีความเสี่ยงสูงที่เป็นโรคออทิซึมสเปกตรัมมากกว่าเด็กที่ไม่มี (Hallmayer et al., 2011)
- โรคออทิซึมสเปกตรัมมักจะเกิดขึ้นกับคนที่มียีนหรือโครโมโซมที่ผิดปกติ เช่น โครโมโซมเอกซ์เปราะ (Fragile X syndrome) ทูเบอร์รัส สเคลอโรซิส (Tuberous sclerosis) และความผิดปกติของโครโมโซม 15q11-q13 (Nurmi et al., 2003)

ถึงแม้จะพบว่ามีความผิดปกติของยีนเกิดขึ้น แต่ตามความผิดปกติของยีนเพียงยีนเดียวไม่สามารถก่อให้เกิดโรคออทิซึมสเปกตรัมได้ ซึ่งจากงานวิจัยพบว่าส่วนใหญ่โรคออทิซึมสเปกตรัมนั้น จะเกิดเมื่อผู้ป่วยมีความผิดปกติในหลายปัจจัยหรือหลายยีนร่วมกัน จนพัฒนาไปเป็นโรคออทิซึมสเปกตรัม และความผิดปกติของยีนหลายยีนที่แสดงออกนี้ ยังสามารถใช้อธิบายลักษณะอาการที่แตกต่างกันในผู้ป่วยแต่ละคนได้อีกด้วย

2. สาเหตุด้านสิ่งแวดล้อม

นอกจากสาเหตุทางพันธุกรรมแล้วในปัจจุบันยังพบว่าโรคออทิซึมสเปกตรัมยังมีสาเหตุมาจากทางด้านสิ่งแวดล้อมด้วย เช่น

- การได้รับยาหรือสารบางอย่างขณะตั้งครรภ์ เช่น กรดวาลโพรอิก (Valproic acid) (Christensen et al., 2013) และ ธาลิโดไมด์ (Thalidomide) (Stromland, Nordin, Miller, Akerstrom, & Gillberg, 1994) เป็นต้น จะทำให้เด็กในครรภ์มีความเสี่ยงสูงที่จะเป็นโรคออทิซึมสเปกตรัม
- เด็กที่เกิดจากพ่อแม่ที่มีอายุมาก (Durkin et al., 2008)
- การขาดวิตามินดีในขณะตั้งครรภ์ (Vuillermot et al., 2017)
- ผลของ Cytokine interleukin-6 ในขณะตั้งครรภ์ (Parker-Athill & Tan, 2010)

ระบาดวิทยาของโรคออทิซึมสเปกตรัม

จากการศึกษาทางระบาดวิทยาโดยศูนย์ควบคุมและป้องกันโรคของประเทศสหรัฐอเมริกา (Center for Disease Control and Prevention; CDC) พบว่าอัตราความชุกของโรคออทิซึมสเปกตรัมสูงถึง 1 ใน 54 ราย ในปี ค.ศ.2020 (Center for Disease Control and Prevention, 2020) และในประเทศไทยจากการสำรวจความชุกของโรคออทิซึมสเปกตรัม พบว่ามีผู้ป่วยมากถึง 6 คนต่อจำนวนประชากร 1,000 คน ในปี พ.ศ.2555 (Chantayanon, 2016) จากความชุกของผู้ป่วยออทิซึมสเปกตรัมในประเทศไทยที่กล่าวมาข้างต้น คาดว่าจำนวนผู้ป่วยที่พบนั้น ต่ำกว่าความเป็นจริง เนื่องจากเด็กที่ป่วยเป็นโรคออทิซึมสเปกตรัมจำนวนมาก ยังขาดโอกาสที่จะได้รับการวินิจฉัยที่ถูกต้องและเหมาะสม นอกจากนี้จากการศึกษาความชุกของโรคออทิซึมสเปกตรัมในประเทศสหรัฐอเมริกา ในปี ค.ศ.2020 ยังพบว่าอุบัติการณ์ของโรคออทิซึมสเปกตรัม มักเกิดในเด็กผู้ชายมากกว่าเด็กผู้หญิง โดยในเด็กผู้ชายจะมีความชุกของโรค 1 ใน 34 คน ขณะที่ความชุกในเด็กผู้หญิงมี 1 ใน 144 คน (Center for Disease Control and Prevention, 2020)

การวินิจฉัยผู้ป่วยโรคออทิซึมสเปกตรัม

โรคออทิซึมสเปกตรัมยากที่จะทำการวินิจฉัย เนื่องจากโรคนี้ยังไม่มี การตรวจเลือด ที่จะใช้ในการตรวจเพื่อประกอบการวินิจฉัย ดังนั้นแพทย์จะทำการติดตามเฝ้าดูพฤติกรรม และการพัฒนาการของเด็กเพื่อใช้ในการวินิจฉัยแทน ด้วยวิธีนี้ในบางครั้งจะสามารถวินิจฉัยผู้ป่วยตั้งแต่อายุเพียง 18 เดือนหรือน้อยกว่านั้น แต่โดยปกติผู้เชี่ยวชาญจะสามารถวินิจฉัยผู้ป่วยได้อย่าง

แม่นยำมากที่สุดเมื่อผู้ป่วยมีอายุ 2 ปีขึ้นไป แต่อย่างไรก็ตาม มีเด็กที่ป่วยเป็นโรคออทิซึมจำนวนมากที่ไม่ได้รับการวินิจฉัยจนโต ซึ่งทำให้ผู้ป่วยเหล่านี้ไม่ได้รับการช่วยเหลืออย่างที่ควรจะเป็น โดยการวินิจฉัยโรคออทิซึมสเปกตรัมจะประกอบด้วย 2 ขั้นตอนหลัก ๆ (Center for Disease Control and Prevention, 2020) คือ

1. Developmental Screening

การทดสอบนี้จะเป็นการทดสอบสั้น ๆ ที่จะบอกว่าเด็กมีการเรียนรู้ทักษะพื้นฐานที่ควรจะมี หรือ มีการเรียนรู้แต่ช้ากว่าปกติหรือไม่ ในระหว่างการทดสอบแพทย์อาจจะถามผู้ปกครองของเด็ก หรือ พูดคุยและเล่นกับเด็กเพื่อสังเกตการณ์เรียนรู้ การพูด พฤติกรรมและการเคลื่อนไหว ซึ่งการพฤติกรรมเหล่านี้หากพบว่ามีผิดปกติไป อาจจะเป็นสัญญาณบ่งบอกการป่วยเป็นออทิซึมสเปกตรัมได้ ซึ่งการทดสอบการพัฒนาการช้าและทูลพลภาพจะทำเมื่ออายุ

- 9 เดือน
- 18 เดือน
- 24 หรือ 30 เดือน

ทั้งนี้การวินิจฉัยอาจจะมีการตรวจเพิ่มเติมหากเด็กมีความเสี่ยงสูงที่จะมีปัญหาด้านการพัฒนาการเนื่องจากการคลอดก่อนกำหนด มีน้ำหนักแรกเกิดน้อยกว่าปกติ ในกรณีที่เด็กมีญาติที่ป่วยเป็นโรคออทิซึมสเปกตรัมหรือในกรณีที่มีการแสดงออกของพฤติกรรมที่เกี่ยวข้องออกมา

2. Comprehensive Diagnostic Evaluation

การทดสอบนี้เป็น การทดสอบขั้นที่สอง โดยจะทำการติดตามพฤติกรรม การพัฒนาการ และ สัมภาษณ์ผู้ปกครองของเด็ก ซึ่งอาจจะทำการทดสอบการได้ยินและการมองเห็น การตรวจทางพันธุกรรม และอื่น ๆ ร่วมด้วย ซึ่งในขั้นตอนนี้จะต้องการผู้เชี่ยวชาญในการทำการทดสอบ เช่น Developmental pediatricians, Child neurologists, Child psychologists or Psychiatrists

เครื่องมือที่ใช้ในการวินิจฉัย

ในปัจจุบันมีเครื่องมือมากมายที่ช่วยในการวินิจฉัยโรคออทิซึมสเปกตรัมในเด็ก เครื่องมือที่ใช้ในการวินิจฉัยนี้จะอาศัยข้อมูลจากการสัมภาษณ์ผู้ปกครองของเด็กและการสังเกตพฤติกรรมของเด็ก ตัวอย่างของเครื่องมือที่ใช้ในการวินิจฉัย (Center for Disease Control and Prevention, 2016) เช่น

- Autism Diagnosis Interview-Revised (ADI-R)

แบบทดสอบ ADI-R (The Autism Diagnostic Interview-Revised) ของมหาวิทยาลัย University of Michigan เป็นแบบทดสอบที่ใช้ทั้งในเด็กและผู้ใหญ่ แบบทดสอบ ADI-R เป็นวิธี Gold Standard ในการวินิจฉัยโรคออทิซึมสเปกตรัม และใช้แพร่หลายในประเทศสหรัฐอเมริกา โดยแบบทดสอบ ADI-R ประกอบด้วยรูปแบบการทดสอบทั้งหมด 123 หัวข้อ แบ่งเป็น 4 ส่วนการทดสอบ ได้แก่ ทดสอบการสื่อสาร (Communication) ทดสอบการเข้าสังคม (Social interaction) ทดสอบการกระทำพฤติกรรมบางอย่างซ้ำ ๆ (Repetitive behaviors) และทดสอบทักษะอัจฉริยภาพพิเศษ (Savant skills) ซึ่งแบบทดสอบนี้สามารถใช้ได้กับผู้ป่วยที่มีอายุตั้งแต่ 18 เดือนขึ้นไปได้

- Autism Diagnostic Observation Schedule-Generc (ADOS-G)

ADOS-G เป็นแบบทดสอบที่ประเมินโดยอาศัยข้อมูลการปฏิสัมพันธ์ การสื่อสาร การเล่น และการใช้จินตนาการ ของผู้ที่สงสัยว่าจะป่วยเป็นโรคออทิซึมสเปกตรัม

- Childhood Autism Rating Scale (CARS)

แบบการทดสอบนี้เหมาะสำหรับเด็กที่มีอายุ 2 ปีขึ้นไป โดย CARS ประกอบด้วยส่วนที่ครอบคลุมการวินิจฉัยจากพฤติกรรม ลักษณะเฉพาะและความสามารถของผู้ที่สงสัยว่าป่วยเป็นโรคออทิซึมสเปกตรัม

- Gilliam Autism Rating Scale-Second Edition (GARS-2)

เป็นเครื่องมือที่ครู ผู้ปกครองและแพทย์ ใช้ในการวินิจฉัยเด็กที่มีอายุ 3 ปีถึง 22 ปี นอกจากนี้แบบทดสอบนี้ยังสามารถประเมินความรุนแรงของโรคได้อีกด้วย

2.2 แบบทดสอบ ADI-R (Autism Diagnostic Interview-Revised)

ในปัจจุบัน การวินิจฉัยผู้ป่วยออทิซึมสเปกตรัมที่เป็นมาตรฐานนั้น มีพื้นฐานมาจากการสังเกตพฤติกรรมของผู้ป่วยและการสอบถามผู้ปกครองเป็นหลัก (Murphy et al., 2016) กล่าวโดยย่อคือเด็กอายุประมาณ 3 ปี ที่สงสัยว่าเป็นโรคออทิซึมสเปกตรัมหรือไม่นั้น จะถูกวินิจฉัยจากการสังเกตพฤติกรรมในระหว่างที่แพทย์สอบถามคำถามจำนวนมากกับผู้ปกครอง ซึ่งเป็นสาเหตุหลักของข้อจำกัดในการวินิจฉัยผู้ป่วยออทิซึมสเปกตรัมที่ทำให้ผู้ป่วยได้รับการวินิจฉัยที่ล่าช้าและอาจไม่ได้รับการรักษาที่เหมาะสม

แบบทดสอบ ADI-R หรือ Autism Diagnostic Interview-Revised เป็นวิธีการวินิจฉัยโรคออทิซึมสเปกตรัมที่เป็นมาตรฐานและใช้แพร่หลายทั่วโลก (Howes et al., 2018) แบบทดสอบ

ADI-R เป็นเครื่องมือที่ถูกใช้มาเป็นเวลานาน และมีการพัฒนามาหลายเวอร์ชัน ซึ่งเวอร์ชันที่ใช้ในการศึกษานี้ ได้แก่ เวอร์ชันของปี 1995 และ 2003 มีจำนวนข้อที่ใช้ในการสัมภาษณ์ 111 และ 93 ข้อตามลำดับ โดย 2 เวอร์ชันนี้มีความแตกต่างกันตรงที่เวอร์ชัน 2003 มีการลดจำนวนข้อลงเพื่อให้ง่ายต่อการวินิจฉัยทั้งในทางคลินิกและในงานวิจัย รวมทั้งมีข้อที่เกี่ยวข้องกับพฤติกรรมของกลุ่มผู้ป่วยที่มีอายุน้อย (ต่ำกว่า 3 ปี) ที่ไม่มีในแบบทดสอบเวอร์ชัน 1995 ด้วย ซึ่งเมื่อทำการรวมข้อมูลทั้ง 2 เวอร์ชันแล้ว พบว่ามีจำนวนข้อที่สามารถใช้ในการศึกษาวิจัยทั้งหมด 123 ข้อ ใช้เวลาในประมาณ 90 นาที เพื่อประเมินพฤติกรรมทั้งหมด 3 ด้าน คือ ด้านการเข้าสังคม ด้านภาษาและการสื่อสาร และด้านพฤติกรรม เหมาะสำหรับเด็กอายุ 12-47 เดือน (Kim & Lord, 2012) โดยผู้ที่ใช้แบบทดสอบนี้ในการประเมินจะต้องเป็นแพทย์หรือผู้เชี่ยวชาญที่ผ่านการอบรมการใช้แบบทดสอบนี้มาโดยเฉพาะ โดยคำถามในการสัมภาษณ์จะเกี่ยวกับประวัติของผู้ป่วย พฤติกรรม พัฒนาการทางภาษา พัฒนาการด้านสังคม และอาการทางคลินิกต่าง ๆ ที่เกี่ยวข้อง เช่น ความก้าวร้าว การทำร้ายตัวเอง เป็นต้น (Rutter et al., 2003) แบบทดสอบ ADI-R สามารถแบ่งคำถามออกเป็น 4 หมวด ได้แก่ หมวดพัฒนาการด้านการสื่อสาร หมวดพัฒนาการทางด้านสังคม หมวดพฤติกรรมย้ำคิดย้ำทำ และหมวดทักษะอัจฉริยภาพพิเศษ การให้คะแนนจะให้ตามระดับความรุนแรงเป็นตัวเลข 0-3 (Charman & Gotham, 2013) ซึ่งขั้นตอนการใช้แบบสัมภาษณ์มีดังนี้ แพทย์ทำการสัมภาษณ์ผู้ปกครองเด็กที่สงสัยว่าจะเป็นโรคออทิซึมสเปกตรัม ร่วมกับการสังเกตพฤติกรรมของเด็ก จากนั้นแพทย์จะลงคะแนนตามอาการที่เกี่ยวข้องในแต่ละข้อคำถาม โดยระดับคะแนนที่มากขึ้นจะบ่งบอกอาการของโรคที่รุนแรงขึ้น คะแนนที่ได้จากแต่ละข้อคำถามของแต่ละด้านจะนำมารวมกันเพื่อดูว่าเกินค่าจุดตัด (Cut-off) ของผู้ที่มีอาการออทิซึมหรือไม่ ซึ่งการให้คะแนนของแบบสัมภาษณ์มีดังนี้

- 0 = ไม่แสดงพฤติกรรมของความผิดปกติใด ๆ ที่เกี่ยวข้องกับโรคออทิซึมสเปกตรัม
- 1 = เริ่มแสดงพฤติกรรมผิดปกติ แต่ยังไม่มากพอที่จะวินิจฉัยได้ว่ามีอาการของโรคออทิซึมสเปกตรัม
- 2 = แสดงพฤติกรรมผิดปกติของอาการของโรคออทิซึมสเปกตรัมอย่างชัดเจน
- 3 = แสดงพฤติกรรมผิดปกติของอาการของโรคออทิซึมสเปกตรัมอย่างรุนแรง
- 7 = แสดงพฤติกรรมผิดปกติอย่างชัดเจน แต่ไม่ใช่อาการของโรคออทิซึมสเปกตรัม
- 8 = ไม่สามารถประเมินพฤติกรรมออกมาได้
- 9 = ไม่ทราบข้อมูล หรือไม่ได้ทำการสัมภาษณ์ในข้อนั้น

คะแนนที่ได้จะขึ้นกับการประเมินของการโต้ตอบของผู้ปกครองเป็นหลัก ซึ่งเกณฑ์ในการวินิจฉัยว่าเป็นโรคออทิซึมสเปกตรัมหรือไม่นั้น ผู้ป่วยโรคออทิซึมสเปกตรัมจะต้องมีคะแนนทั้ง 3 ด้าน เกินค่า cut-off ดังนี้ *คะแนนในหมวดพัฒนาการด้านการสื่อสารจะต้องมากกว่าหรือเท่ากับ 8 คะแนน (กรณีที่เป็นวจนภาษา) หรือ 7 คะแนน (กรณีที่เป็นอวจนภาษา) และคะแนนในหมวดพัฒนาการทางด้านสังคมจะต้องมากกว่าหรือเท่ากับ 10 คะแนน และคะแนนในหมวดพฤติกรรมย่ำคิดย่ำทำ จะต้องมากกว่าหรือเท่ากับ 3 คะแนน* (Lord, Rutter, & Le Couteur, 1994) ซึ่งจุดประสงค์ของแบบทดสอบ ADI-R นั้น คือเพื่อประเมินบุคคลที่สงสัยว่าเป็นโรคออทิซึมสเปกตรัม และสามารถช่วยในการแบ่งกลุ่มผู้ป่วยโรคออทิซึมสเปกตรัมออกเป็นกลุ่มย่อยเพื่อลดความหลากหลายของอาการของผู้ป่วยโรคออทิซึมสเปกตรัมได้ (Hu et al., 2009) โดยข้อมูลจาก Spectrumnews ในปี 2011 ได้มีการระบุไว้ว่าแบบทดสอบ ADI-R ให้ความถูกต้องในการวินิจฉัยคนไข้ได้มากถึง 98% (Constantino et al., 2007; Fernandopulle, 2011; Warren et al., 2012) และข้อมูล Autism diagnosis in children and young people ใน Evidence update ปี 2013 จาก National Institute for Health and Care Excellence (NICE) (National Institute for Health and Care Excellence, 2013) ได้มีการกล่าวถึงการศึกษาความไว (Sensitivity) และความจำเพาะ (Specificity) ของแบบทดสอบ ADI-R โดย Kim & Lord ในปี 2012 ว่าแบบทดสอบ ADI-R นั้นสามารถวินิจฉัยกลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Non-verbal แยกจากกลุ่ม Non-spectrum ด้วยความไว 97% และมีความจำเพาะ 43% กลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Single word แยกจากกลุ่ม Non-spectrum ด้วยความไว 91% และมีความจำเพาะ 82% และกลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Phrase-speech แยกจากกลุ่ม Non-spectrum ด้วยความไว 70% และความจำเพาะ 68% นอกจากนี้ยังได้มีการลองทดสอบวิธีการวินิจฉัยด้วยแบบทดสอบ ADI-R แบบวิธีใหม่ คือตัดจำนวนข้อที่ซ้ำกันในเวอร์ชันปี 1995 และปี 2003 ออก พบว่า สามารถวินิจฉัยกลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Non-verbal แยกจากกลุ่ม Non-spectrum ด้วยความไว 85% และมีความจำเพาะเพิ่มขึ้นเป็น 70%, กลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Single word แยกจากกลุ่ม Non-spectrum ด้วยความไวเพิ่มขึ้นเป็น 94% และมีความจำเพาะ 81%, และกลุ่มคนไข้ออทิซึมสเปกตรัมแบบ Phrase-speech แยกจากกลุ่ม Non-spectrum ด้วยความไวเพิ่มขึ้นเป็น 80% และความจำเพาะเพิ่มขึ้นเป็น 70% (Kim & Lord, 2012) อย่างไรก็ตาม แบบทดสอบ ADI-R ยังคงมีข้อจำกัดคือ

1. ใช้เวลาในการทำแบบทดสอบนาน เนื่องจากจำนวนข้อของแบบทดสอบมี 123 ข้อ แต่ละข้อใช้เวลาในการตอบหรือให้คะแนน 2-3 นาที ทำให้เวลาที่ใช้ในการทำแบบทดสอบต่อการวินิจฉัย 1 ครั้ง จะใช้เวลาอย่างน้อย 90 นาที จึงไม่เหมาะกับการวินิจฉัยผู้ป่วยในสถานพยาบาลที่มีผู้เข้ารับการรักษาเป็นจำนวนมาก
2. ขาดแคลนผู้เชี่ยวชาญเฉพาะทางในการวิเคราะห์แบบทดสอบ ประกอบกับผู้ที่ทำการวินิจฉัยอาจมีการลงคะแนนผิด เนื่องจากไม่ชำนาญในการใช้แบบทดสอบ อาจทำให้ผู้คัดกรองลงคะแนนรวมผิด ส่งผลต่อการวางแผนการรักษาและฟื้นฟูได้
3. อายุของผู้ที่จะได้รับการวินิจฉัย พบว่าแบบทดสอบ ADI-R ใช้ได้สำหรับผู้ที่อายุ 12-18 เดือนขึ้นไป ทำให้อายุในการรักษาครั้งแรกลำบาก จากผลการศึกษาพบว่าอายุในการบำบัดส่งผลต่อการฟื้นฟูพัฒนาการในอนาคต จึงทำให้การลดข้อจำกัดด้านอายุของผู้ที่ถูกวินิจฉัยครั้งแรกเป็นสิ่งจำเป็น

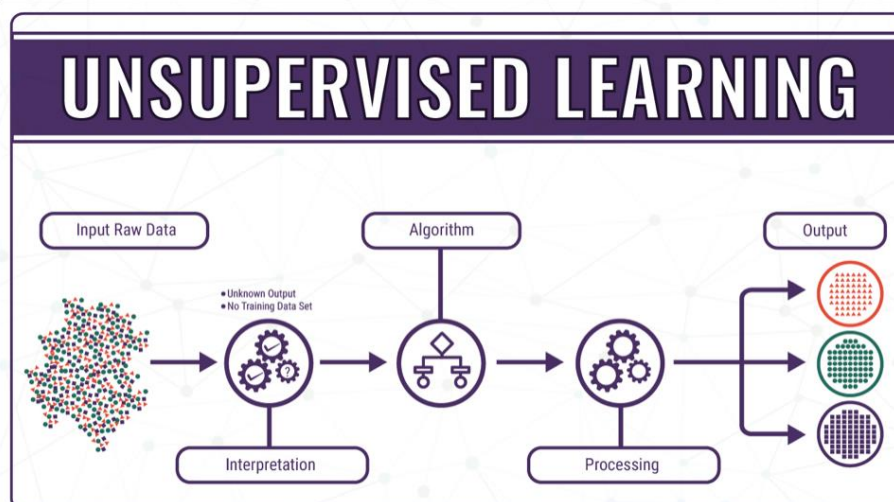
ดังนั้นเพื่อลดข้อจำกัดของแบบทดสอบ ADI-R การประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์อาจช่วยในการวินิจฉัยและแบ่งกลุ่มย่อยผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อยได้ จากลักษณะอาการทางคลินิกที่ได้จากคะแนนของแบบทดสอบ ADI-R

2.3 เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence)

เทคโนโลยีปัญญาประดิษฐ์ (Artificial Intelligence) คือความสามารถของคอมพิวเตอร์หรือเครื่องมือในแบบที่เป็นดิจิทัลในการแปลผลของข้อมูลที่นำเข้าระบบ การเรียนรู้จากข้อมูลที่ใส่เข้าไป หรือการใช้การเรียนรู้นั้นในการบรรลุเป้าหมายผ่านการปรับเปลี่ยนและพัฒนาระบบ เทคโนโลยีปัญญาประดิษฐ์มักใช้ในการอธิบายความสามารถของคอมพิวเตอร์ที่ดำเนินงานที่เกี่ยวข้องกับความชาญฉลาด หรือมักประยุกต์ในโครงการเกี่ยวกับการพัฒนาระบบตามกระบวนการความฉลาดของมนุษย์ เช่น ความสามารถในการมีเหตุมีผล การค้นหาความหมาย หรือการเรียนรู้จากประสบการณ์ (Copeland, 2019) เทคโนโลยีปัญญาประดิษฐ์สามารถนำไปประยุกต์ใช้กับการวิเคราะห์ข้อมูลขนาดใหญ่ (Big data analysis) ในหลากหลายสาขาด้วยกัน ประกอบไปด้วย การประมวลผลภาษาธรรมชาติ (Natural language processing) การรู้จำคำพูด (Speech recognition) การเรียนรู้ของเครื่อง (Machine Learning) การประมวลผลภาพ (Image processing) วิทยาการหุ่นยนต์ (Robotics) เครือข่ายเชิงความหมาย (Semantic web) เทคโนโลยีการศึกษา (Intelligent computer-aided instruction) ระบบผู้เชี่ยวชาญ (Expert systems) และสาขาปัญญาประดิษฐ์อื่น ๆ

การเรียนรู้ของเครื่องมือ (Machine learning) ซึ่งเป็นหนึ่งในสาขาของเทคโนโลยีปัญญาประดิษฐ์ คือกระบวนการวิเคราะห์ข้อมูลจากการสร้างรูปแบบผ่านโปรแกรมอัตโนมัติ เป็นการที่ระบบสามารถเรียนรู้ได้เองแบบอัตโนมัติและปรับปรุงการจัดการระบบได้จากประสบการณ์ โดยไม่ต้องตั้งโปรแกรมไว้อย่างชัดเจน แต่มีการมุ่งเน้นการพัฒนาโปรแกรมคอมพิวเตอร์ที่สามารถเข้าถึงข้อมูลและใช้ข้อมูลนั้น ๆ ได้ เพื่อให้การจัดการข้อมูลเป็นไปอย่างมีประสิทธิภาพ สามารถเรียนรู้ข้อมูลรูปแบบของข้อมูล และตัดสินใจแทนมนุษย์ได้ การเรียนรู้ของเครื่องมือสามารถใช้งานได้หลายรูปแบบด้วยกัน โดยอาศัยกลไกที่มีลักษณะเป็นระบบหรือเรียกว่าอัลกอริทึม (Algorithm) ซึ่งมีอยู่ด้วยกันหลายรูปแบบ ออกแบบโดยนักวิทยาศาสตร์ข้อมูล (Data scientist) ซึ่งหนึ่งในอัลกอริทึมที่ได้รับความนิยมสูงคือ การเรียนรู้เชิงลึก (Deep learning) ซึ่งถูกออกแบบมาให้ใช้งานได้ง่ายและประยุกต์ใช้ได้กับหลายลักษณะงาน อย่างไรก็ตามในการทำงานจริง นักวิทยาศาสตร์ข้อมูลจำเป็นจะต้องออกแบบตัวแปรต่าง ๆ ทั้งในการเรียนรู้เชิงลึกและหาอัลกอริทึมอื่น ๆ มาเป็นคู่เปรียบเทียบกับเพื่อมองหาอัลกอริทึมที่เหมาะสมที่ในการใช้งานจริง โดยปัจจุบันการเรียนรู้ของเครื่องมือมีอยู่ 4 ประเภทหลัก ดังนี้

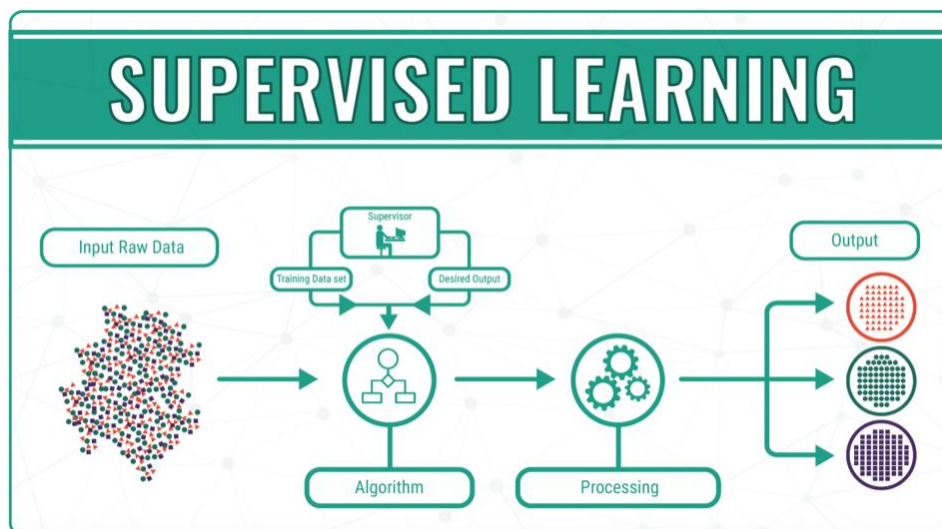
1. การเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) ใช้ในกรณีที่ข้อมูลที่ใช้ในการสอนไม่มีการจัดประเภทหรือติดฉลาก โดยจะเรียนรู้ว่าระบบสามารถอนุมานฟังก์ชันเพื่ออธิบายโครงสร้างที่ซ่อนอยู่จากข้อมูลที่ไม่ติดฉลากได้อย่างไร โดยระบบไม่ได้หาผลลัพธ์ที่ต้องการ แต่จะหาและสรุปข้อมูลจากชุดข้อมูลทั้งหมดเพื่ออธิบายโครงสร้างที่ซ่อนอยู่จากข้อมูลที่ไม่ได้ติดฉลากนั้น ซึ่งเราสามารถแบ่งการเรียนรู้แบบไม่มีผู้สอนออกเป็นการจัดกลุ่มของข้อมูล (Clustering) และการหาความสัมพันธ์ (Association) ตัวอย่างอัลกอริทึมการเรียนรู้แบบไม่มีผู้สอนที่นิยม เช่น K-means clustering, K-medoids clustering เป็นต้น



รูปที่ 1 Unsupervised machine learning

แสดงกระบวนการทำงานและวิเคราะห์ข้อมูลของการเรียนรู้แบบไม่มีผู้สอน ซึ่งไม่ต้องมีชุดข้อมูลที่เป็นชุดการสอน (Training dataset) เพื่อให้เครื่องมือเรียนรู้และวิเคราะห์ความสัมพันธ์ภายในชุดข้อมูลเอง แล้วทำนายผลลัพธ์ออกมา (Loon, 2018)

2. การเรียนรู้แบบมีผู้สอน (Supervised machine learning) เป็นการเรียนรู้ในแบบที่ระบบจะสามารถประยุกต์สิ่งที่ได้เรียนรู้จากในอดีตเพื่อใช้ทำนายข้อมูลใหม่ในอนาคตได้ มีเป้าหมายเพื่อการทำนายและจำแนกข้อมูล โดยใช้การติดฉลากตัวอย่างเพื่อทำนายเหตุการณ์ในอนาคต เริ่มจากการวิเคราะห์ชุดการสอน (Training dataset) อัลกอริทึมการเรียนรู้จะสร้างฟังก์ชันอนุमानเพื่อการทำนายผลลัพธ์ ซึ่งการเรียนรู้จะหยุดเมื่ออัลกอริทึมนั้น สามารถบรรลุเป้าหมายที่ยอมรับได้ นอกจากนี้อัลกอริทึมการเรียนรู้แบบมีผู้สอนนี้ยังสามารถเปรียบเทียบผลลัพธ์ที่ได้กับข้อมูลที่ถูกต้องเพื่อตัดแปลงแก้ไขโมเดลในการวิเคราะห์ข้อมูลได้ โดยเราสามารถแบ่งการเรียนรู้แบบมีผู้สอนออกเป็นการถดถอย (Regression) และการจำแนกกลุ่ม (Classification) สำหรับการถดถอยจะใช้เมื่อตัวแปรของผลลัพธ์เป็นค่าที่แท้จริง ส่วนการจำแนกกลุ่มจะใช้เมื่อตัวแปรของผลลัพธ์เป็นลักษณะของหมวดหมู่ ตัวอย่างของอัลกอริทึมการเรียนรู้แบบมีผู้สอนที่นิยม เช่น การถดถอยเชิงเส้น (Linear regression) ต้นไม้ตัดสินใจ (Decision tree) โครงข่ายประสาทเทียม (Artificial Neural Network) เป็นต้น



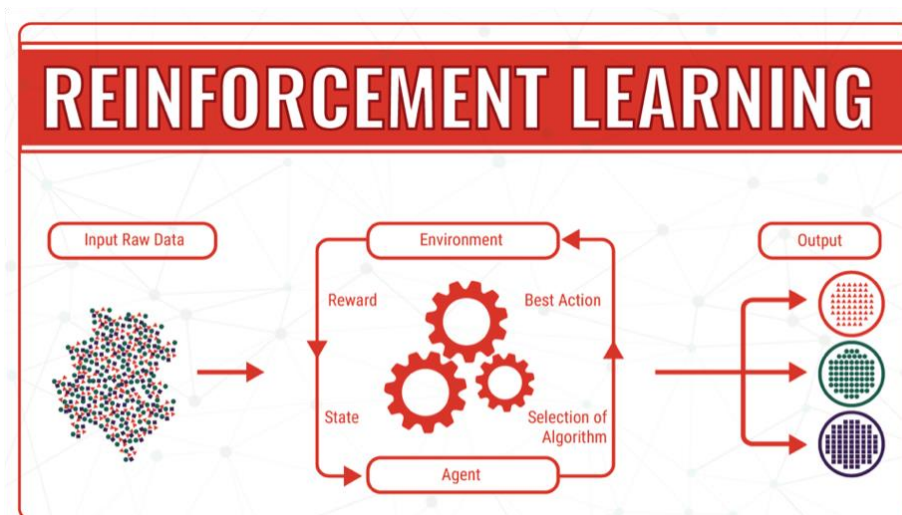
รูปที่ 2 Supervised machine learning

แสดงกระบวนการทำงานและวิเคราะห์ข้อมูลของการเรียนรู้แบบมีผู้สอน ซึ่งจำเป็นจะต้องมีชุดข้อมูลที่เป็นชุดการสอน (Training dataset) เพื่อให้เครื่องมือเรียนรู้แล้วเกิดเป็นโมเดลสำหรับการวิเคราะห์ข้อมูลชุดใหม่ในอนาคต (Loon, 2018)

3. การเรียนรู้แบบกึ่งมีผู้สอน (Semi-supervised machine learning) จากการเรียนรู้ของเครื่องมือทั้ง 2 ประเภทก่อนหน้า แม้ว่าจะมีหรือไม่มี การติดฉลากข้อมูลในชุดข้อมูลทั้งหมด การเรียนรู้แบบกึ่งมีผู้สอนจะสามารถใช้กับข้อมูลทั้ง 2 แบบได้ เนื่องจากในหลายสถานการณ์ ต้นทุนของการติดฉลากข้อมูลนั้นค่อนข้างสูงเนื่องจากจำเป็นต้องใช้ผู้เชี่ยวชาญที่เกี่ยวข้องกับข้อมูลในการติดฉลาก ดังนั้นแม้ว่าจะมีการติดฉลากของข้อมูลบางส่วนแค่เพียงเล็กน้อย การเรียนรู้แบบกึ่งมีผู้สอนก็จะสามารถสร้างโมเดลในการวิเคราะห์ข้อมูลได้ บนพื้นฐานมาจากความคิดที่ว่าถึงแม้สมาชิกในชุดข้อมูลส่วนมากที่ไม่ได้ติดฉลากนั้นเราไม่รู้ แต่ตัวข้อมูลในชุดข้อมูลนั้นก็จะมีโครงสร้างบางอย่างที่สำคัญซ่อนอยู่ในชุดข้อมูลซึ่งเราสามารถหาความสัมพันธ์มาอธิบายข้อมูลนั้นได้

4. การเรียนรู้แบบโต้ตอบ (Reinforcement Machine Learning) เป็นการเรียนรู้ของเครื่องมือด้วยการตอบโต้กับสิ่งแวดล้อมที่ได้รับข้อมูล โดยเครื่องมือจะยังไม่ทราบว่าข้อมูลที่ได้รับนั้นมีความสัมพันธ์กันอย่างไร แต่จะหาความเป็นไปได้ของความสัมพันธ์ในชุดข้อมูลนั้น ๆ แล้วโต้ตอบออกมา ซึ่งผลลัพธ์ของการแสดงออกจะมีการให้รางวัลเมื่อโต้ตอบได้ถูกต้อง และทำโทษเมื่อโต้ตอบผิด ดังนั้นเมื่อเครื่องมือเกิดการเรียนรู้ซ้ำ ๆ ก็จะทำให้การโต้ตอบกับชุดข้อมูลนั้น ควรมีแบบแผนหรือความสัมพันธ์อย่างไร เพื่อให้ทำได้ถูกต้องมากขึ้น อัลกอริทึมของการเรียนรู้แบบโต้ตอบ เช่น Q-learning, Temporal difference, Deep adversarial networks เป็นต้น โดยปัจจุบันจะนิยมใช้

การเรียนรู้แบบโต้ตอบกับการทำงานแบบอัตโนมัติของเครื่องมือ เช่น เกมส์คอมพิวเตอร์อย่างหมากรุก Robotic hands รถไร้คนขับ เป็นต้น (David, 2017)



รูปที่ 3 Reinforcement Machine Learning

แสดงกระบวนการทำงานและวิเคราะห์ข้อมูลของการเรียนรู้แบบโต้ตอบ ซึ่งเครื่องมือ จะทำการวิเคราะห์หาความสัมพันธ์ภายในชุดข้อมูล แล้วโต้ตอบกับสิ่งแวดล้อมที่ได้รับข้อมูล (Loon, 2018)

จากความสามารถในการเรียนรู้ของระบบคอมพิวเตอร์ มีหลายงานวิจัยด้วยกันที่มีการนำเทคโนโลยีปัญญาประดิษฐ์มาใช้ในการทำนายการวินิจฉัยในหลายโรคด้วยกันรวมถึงโรคออทิซึม สเปกตรัมด้วย ยกตัวอย่างเช่น การใช้การเรียนรู้ของเครื่องมือแบบมีผู้สอนในการจัดกลุ่มผู้ป่วยเด็กที่เป็นออทิซึมสเปกตรัมแบบ Low-function และมีความผิดปกติของการเคลื่อนไหวของร่างกาย (Simple-upper limb movement) (Crippa et al., 2015) หรือการใช้การเรียนรู้ของเครื่องมือแบบมีผู้สอนชนิด Logistic regression ในการจำแนกผู้ป่วยออทิซึมสเปกตรัมโดยใช้แบบทดสอบ Autism Spectrum Questions (AQ) (Habib, Harris, Pollick, & Melville, 2019) หรือการสร้างโมเดลจากอัลกอริทึมของการจัดกลุ่มสำหรับการจำแนกระหว่างโรคออทิซึมสเปกตรัม และโรคสมาธิสั้น (Attention Deficit Hyperactivity Disorder; ADHD) (Duda et al., 2017) หรือการใช้การเรียนรู้ของเครื่องมือแบบมีผู้สอนชนิด Support Vector Machine (SVM) ในการแยกผู้ป่วยโรคออทิซึมสเปกตรัม (Verbal individual with ASD) จากผู้ป่วยที่ไม่ได้เป็นโรคออทิซึมสเปกตรัม (Verbal individual with non-ASD) (Bone et al., 2016) และยังมี การใช้การเรียนรู้ของเครื่องมือในการลดจำนวนแบบทดสอบ ADOS (Autistic Diagnostic Observation Schedule) ที่

ใช้ในการคัดกรองผู้ป่วยโรคออทิซึมสเปกตรัม (Kosmicki et al., 2015), (Wall,2012) เป็นต้น และยังมีการศึกษาโดย Pichitpunpong และคณะ ในปี 2019 (Pichitpunpong et al., 2019) ที่ศึกษารูปแบบโปรตีนโอมและทรานสคริปโตมในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยแบบบูรณาการหลายระดับหลังจากที่ได้แบ่งผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อยโดยใช้เครื่องมือทางสถิติที่เป็นการเรียนรู้แบบไม่มีผู้สอน Hierarchical Clustering (HCL) และ Principle Component Analysis (PCA) โดยได้เลือกผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติทางด้านภาษารุนแรงมาศึกษารูปแบบทรานสคริปโตมและโปรตีนโอมในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ได้จากคนไข้ต่อ ซึ่งพบว่าจากโปรตีนที่มีการแสดงออกผิดปกติ นั้น มีโปรตีน DBI (Diazepam-binding inhibitor) ที่แสดงออกลดลงจริงในผู้ป่วย และมีความเกี่ยวข้องกับชีวพยาธิสภาพทางระบบประสาทและการอักเสบ จากงานวิจัยนี้ชี้ให้เห็นว่าการแบ่งกลุ่มคนไข่ออกเป็นกลุ่มย่อยและศึกษารูปแบบโปรตีนโอมเฉพาะในผู้ป่วยกลุ่มย่อยนั้น จะสามารถนำไปสู่การค้นพบสารบ่งชี้ทางชีวภาพที่สามารถช่วยในการวินิจฉัยและเกี่ยวข้องกับชีวพยาธิสภาพของโรคแบบจำเพาะได้ นอกจากนี้จากการศึกษาการใช้การเรียนรู้ของเครื่องมือเพื่อช่วยลดจำนวนแบบสอบถามที่ผ่านมามีการใช้เทคนิคการเรียนรู้ของเครื่องมือ 15 อัลกอริทึม จากโปรแกรม Waikato Environment for Knowledge Analysis หรือ WEKA เพื่อวิเคราะห์ข้อมูลจากแบบทดสอบ ADI-R ของผู้ป่วยออทิซึมสเปกตรัม ซึ่งเมื่อเปรียบเทียบความไวและความจำเพาะในการจำแนกผู้ป่วยออทิซึมสเปกตรัมทั้งหมด 891 คนจากกลุ่มควบคุมหรือคนที่ไม่มีอาการตามเกณฑ์ในการวินิจฉัยโรคออทิซึม 75 คน พบว่าต้นไม้ตัดสินใจแบบหมุนเวียน (Alternating decision tree; ADTree) สามารถจำแนกผู้ป่วยได้โดยอาศัยจำนวนข้อของแบบทดสอบแค่เพียง 7 ข้อ จาก 93 ข้อ ด้วยความแม่นยำถึง 99.9% ประกอบด้วย compsl5 (การเข้าใจภาษาง่าย ๆ) conver5 (การแลกเปลี่ยนบทสนทนา) play5 (การใช้จินตนาการในการเล่น) peerpl5 (การใช้จินตนาการในการเล่นกับเพื่อน) gaze5 (การจ้องมองโดยตรง) grplay5 (การเล่นเป็นกลุ่มกับเพื่อน) และ ageabn (อายุเมื่อมีความผิดปกติครั้งแรก) และเมื่อนำโมเดลนี้ไปศึกษาต่อโดยนำข้อมูลจาก The Simons Foundation และ The Boston Autism Consortium ของผู้ป่วยออทิซึมสเปกตรัมจำนวน 1,654 คน และ 322 คนตามลำดับ พบว่าสามารถจำแนกได้โดยมีความแม่นยำ 93.8% ถึง 99% แต่ยังมีข้อจำกัดในเรื่องของจำนวนกลุ่มตัวอย่างที่เป็นกลุ่มควบคุมยังคงไม่มากพอเมื่อเทียบกับกลุ่มตัวอย่างที่เป็นโรค (Wall, Dally, Luyster, Jung, & Deluca, 2012) ดังนั้นจะเห็นได้ว่าเทคโนโลยีปัญญาประดิษฐ์ โดยเฉพาะอย่างยิ่งการเรียนรู้ของเครื่องมืออาจจะช่วยลดจำนวนข้อของแบบทดสอบในการวินิจฉัยผู้ป่วยได้ จาก

การค้นหาคความน่าจะเป็นในการประมวลผลแบบต่าง ๆ เพื่อเลือกจำนวนข้อในการแบ่งผู้ป่วยออกเป็นกลุ่มย่อยให้น้อยที่สุด ซึ่งจะช่วยลดเวลาในการประมวลผลลงได้

2.4 อัลกอริทึมที่ใช้ในโปรแกรม RapidMiner และการประเมินประสิทธิภาพ

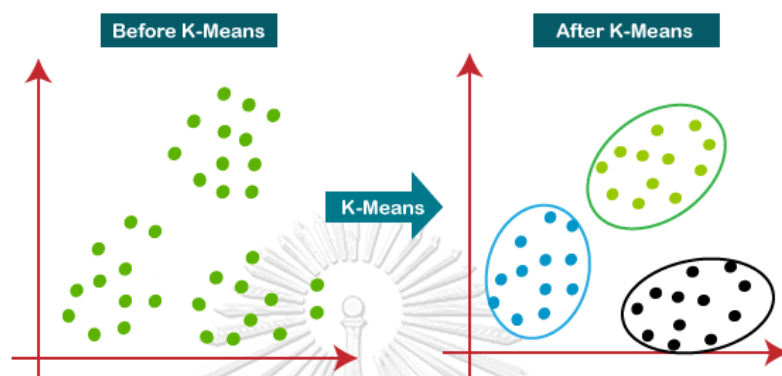
โปรแกรม RapidMiner studio เป็นโปรแกรมที่ถูกผลิตขึ้นโดย Ingo Mierswa และ Ralf Kilenberg มีการพัฒนาแรกเริ่มจากบริษัทที่ชื่อ Rapid-I ในประเทศเยอรมนี และเมื่อช่วงปลายปี 2013 ได้รับเงินลงทุนจากประเทศสหรัฐอเมริกา จึงเปลี่ยนชื่อบริษัทจาก Rapid-I เป็น RapidMiner แทน และย้ายสำนักงานใหญ่มาอยู่ที่ประเทศสหรัฐอเมริกา โปรแกรม RapidMiner ใช้ในงานทางด้านวิทยาศาสตร์ข้อมูล (Data sciences) ไม่ว่าจะเป็นการเรียนรู้ของเครื่องมือ (Machine learning) การทำเหมืองข้อมูล (Data mining) และการพยากรณ์ผลจากการวิเคราะห์ข้อมูล โดยฟังก์ชันหรืออัลกอริทึมที่นิยมใช้ในการจัดการข้อมูลในโปรแกรมนี้ได้แก่ ต้นไม้ตัดสินใจ (Decision tree) เป็นการสร้างกฎในรูปแบบ ถ้า/แล้ว จากคุณลักษณะของชุดข้อมูล โดยไม่มีสมการมากำกับความสัมพันธ์ระหว่างข้อมูลที่น่าเข้ากับผลลัพธ์ที่ได้ การจัดกลุ่มแบบ K-means เป็นหนึ่งในวิธีการจัดกลุ่มในรูปแบบการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน โดยอาศัยพื้นฐานทางสถิติในการวิเคราะห์ข้อมูล และโครงข่ายประสาทเทียม (Artificial Neural Network) เป็นหนึ่งในวิธีการวิเคราะห์ข้อมูลในรูปแบบการเรียนรู้ของเครื่องมือแบบมีผู้สอนซึ่งจำลองมาจากการทำงานของสมองมนุษย์ ที่มีส่วนการเรียนรู้ที่เล็กที่สุดคือ เซลล์ประสาท (Neuron) โดยจะให้ทำหน้าที่ในการคำนวณข้อมูลที่ได้รับเข้าระบบ แล้วหาความน่าจะเป็นของทุกเหตุการณ์ที่เป็นไปได้ เพื่อให้ได้ผลลัพธ์ที่มีความถูกต้องมากที่สุด โปรแกรม RapidMiner สามารถประยุกต์ใช้กับงานได้หลากหลายสาขามาก ไม่ว่าจะเป็นงานทางด้านเศรษฐศาสตร์ บัญชี วิทยาศาสตร์ หรืองานทางด้านทางการแพทย์ อย่างเช่นการพยากรณ์และป้องกันโรครวมถึงการแพทย์ส่วนบุคคล เนื่องจากสามารถจัดการกับข้อมูลจำนวนมากได้

1. อัลกอริทึมของการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning)

1.1 K-means clustering

เป็นหนึ่งในอัลกอริทึมของการเรียนรู้แบบไม่มีผู้สอนที่ง่ายที่สุด โดยจะทำการตัดแบ่ง หรือ Partitioning ข้อมูลออกเป็น k กลุ่ม จากนั้นจะทำการคำนวณค่าเฉลี่ยของข้อมูลในแต่ละกลุ่ม (Means) เพื่อใช้เป็นจุดศูนย์กลาง (Centroid) หรือ ค่าอ้างอิงในการวัดระยะห่างระหว่างข้อมูลใน

กลุ่มเดียวกัน ข้อมูลจะถูกจัดเข้ากลุ่มที่อยู่ใกล้ที่สุดจนครบทั้งหมด และจะมีการคำนวณจุดศูนย์กลางใหม่โดยการหาค่าเฉลี่ยของข้อมูลภายในกลุ่ม หากจุดศูนย์กลางในแต่ละกลุ่มถูกเปลี่ยนตำแหน่ง จะได้ข้อมูลที่มีความสัมพันธ์กับกลุ่มใหม่และใกล้กับจุดศูนย์กลางใหม่ โดยกระบวนการจัดนี้จะดำเนินไปจนกว่าข้อมูลทั้งหมดจะไม่มี การเปลี่ยนกลุ่มอีกต่อไป (กิตติศักดิ์ เกิดประสพ, 2552)

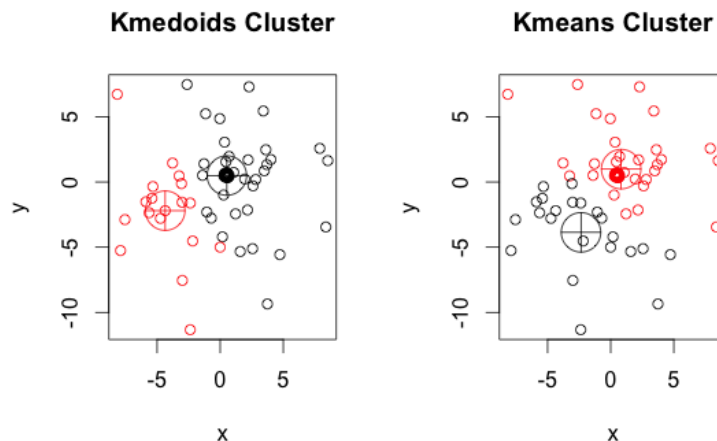


รูปที่ 4 K-means clustering

แสดงภาพก่อนและหลังการวิเคราะห์เพื่อจัดกลุ่มโดยใช้วิธี K-means

1.2 K-medoids clustering

เป็นอัลกอริทึมที่มีความแตกต่างกับ K-means clustering ตรงที่จะใช้ค่าเฉลี่ยระหว่างกลุ่มมาใช้เป็นจุดศูนย์กลาง (Centroid) วิธี K-medoids clustering นี้ จะแทนที่จุดศูนย์กลางด้วยค่าจริงของข้อมูลที่อยู่ในกลุ่มนั้น โดยมีข้อดีกว่าวิธี K-means คือ วิธี K-means ใช้ค่าจุดศูนย์กลางจากค่าเฉลี่ยที่เกิดจากการคำนวณ ไม่ใช่ตำแหน่งที่มีอยู่จริงของข้อมูล และในกรณีที่มี Outliers วิธี K-medoids จะให้ผลลัพธ์ที่ดีกว่าเนื่องจากจุดศูนย์กลางของกลุ่มจะยังคงใกล้เคียงกับข้อมูลส่วนใหญ่ของกลุ่ม แต่วิธี K-means ค่า Outliers จะทำให้จุดศูนย์กลางของกลุ่มไกลจากข้อมูลอื่น ๆ ภายในกลุ่มได้ แต่มีข้อเสียคือวิธี K-medoids จำเป็นต้องใช้ข้อมูลจริงในกลุ่มแทนจุดศูนย์กลางของกลุ่ม ทำให้ไม่เคร่งครัดเรื่องคุณภาพของตำแหน่งกลางเท่าวิธี K-means และเมื่อมีการเปลี่ยนแปลงการกระจายตัวของข้อมูลเพียงเล็กน้อย จะทำให้จุดศูนย์กลางของกลุ่มเปลี่ยนไปแบบไม่ต่อเนื่อง โดยกระโดดจากข้อมูลหนึ่งไปอีกข้อมูลหนึ่งในกลุ่ม (Kaufman & Rousseeuw, 2009)



รูปที่ 5 K-medoids and K-means clustering

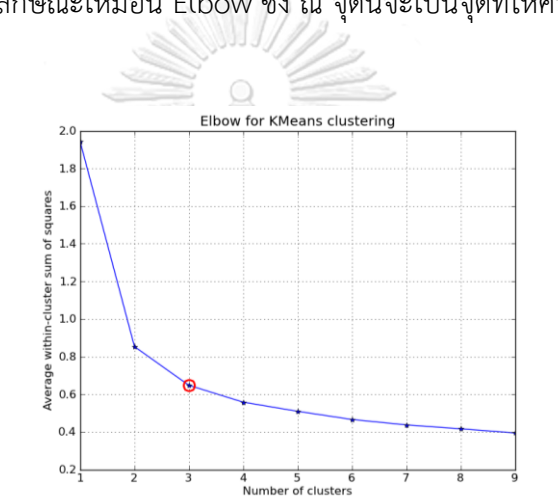
แสดงการเปรียบเทียบการวิเคราะห์เพื่อจัดกลุ่มระหว่างวิธี K-medoids กับวิธี K-means clustering

พารามิเตอร์ของการจัดกลุ่มแบบ K-means และ K-medoids

1. **Add cluster attribute** หากต้องการสร้าง Attribute ใหม่ในชุดข้อมูลผลลัพธ์ที่ได้ โดยจะเป็น Attribute ที่จัดกลุ่มข้อมูลเป็น Cluster ต่าง ๆ โดยสามารถนำไปใช้เป็น Label ในการวิเคราะห์อื่น ๆ ต่อไปได้ โดยใช้คำสั่ง Apply model กับการวิเคราะห์อื่น
2. **Add as label** ถ้าผลลัพธ์ที่ได้เป็นจริงก็จะสร้าง Attribute ใหม่ขึ้นและมีบทบาทพิเศษเป็น Label แต่ถ้าพารามิเตอร์ Add cluster attribute ไม่เป็นความจริง ก็จะไม่มีการสร้าง attribute ใหม่ขึ้น
3. **Remove unlabeled** ถ้าการจัดกลุ่มนี้เป็นจริง ตัวอย่างที่ไม่สามารถจัดกลุ่มได้จะถูกนำออกไปจากข้อมูลผลลัพธ์
4. **k** จำนวนกลุ่มที่เราต้องการให้จัด
5. **Max runs** จำนวนรอบในการวิเคราะห์ตั้งแต่เกิดการสุ่มเริ่มขึ้นที่จุดตั้งต้น
6. **Determine good start values** ถ้าการจัดกลุ่มนี้เป็นจริง จุดของค่า k จะถูกคำนวณจาก K-means++
7. **Max optimization steps** ระบุจำนวนที่มากที่สุดที่จะมีการทำซ้ำในหนึ่งรอบของการวิเคราะห์
8. **Measure types** การเลือกวิธีการวัดผลตามประเภทของข้อมูล มีดังนี้ MixedMeasures NominalMeasures NumericalMeasures และ BregmannDivergences

การวัดและเปรียบเทียบประสิทธิภาพของการจัดกลุ่มด้วยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน

ก่อนที่จะทำการวัดประสิทธิภาพการแบ่งกลุ่มเพื่อค้นหาอัลกอริทึมในการจัดกลุ่มด้วยการเรียนรู้แบบไม่มีผู้สอนที่ดีที่สุดนั้น จะต้องมีการหาค่า k หรือจำนวนกลุ่มที่ต้องการจะแบ่งที่ดีที่สุดก่อน ซึ่งสามารถทำได้โดยใช้วิธี Elbow method ใช้ในการหาจำนวนกลุ่มที่เหมาะสมสำหรับวิธี K-means โดยเฉพาะ เป็นหนึ่งในวิธีวัดข้อผิดพลาดของผลรวมระยะห่างระหว่างข้อมูลกับจุดศูนย์กลาง โดยแกน x เป็นจำนวนกลุ่มที่เรากำหนด ส่วนแกน y เป็นค่า Average within cluster sum of squares ที่คำนวณได้จากโปรแกรม โดยเมื่อความผิดพลาดลดน้อยลง เส้นโค้งที่มีความชันจะเริ่มโค้งและราบเรียบจนเกิดเป็นมุมลักษณะเหมือน Elbow ซึ่ง ณ จุดนี้จะเป็นจุดที่ให้ค่าจำนวนกลุ่มของ Cluster ที่ดีที่สุด



รูปที่ 6 Determining the optimal number of clusters using elbow method

แสดงวิธีการหาค่า k หรือจำนวนกลุ่มที่เหมาะสมในการจัดกลุ่มด้วยวิธี Elbow method

จากภาพจะเห็นว่าจำนวนกลุ่มเท่ากับ 3 จะเป็นจำนวนที่เหมาะสมสำหรับการจัดกลุ่มมากที่สุด

และเนื่องจากการจัดกลุ่มด้วยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน เป็นวิธีการจัดกลุ่มโดยแบ่งข้อมูลออกเป็น Cluster ทางโปรแกรม RapidMiner จะมีวิธีในการวัดประสิทธิภาพของการจัดกลุ่มแบบนี้ โดยเลือกใช้ Operator ที่มีชื่อว่า “Cluster distance performance” หรือการวัดระยะห่างของกลุ่มที่ถูกจัดขึ้น โดยจะมีค่าที่ปรากฏในหน้าต่างแสดงผลดังต่อไปนี้

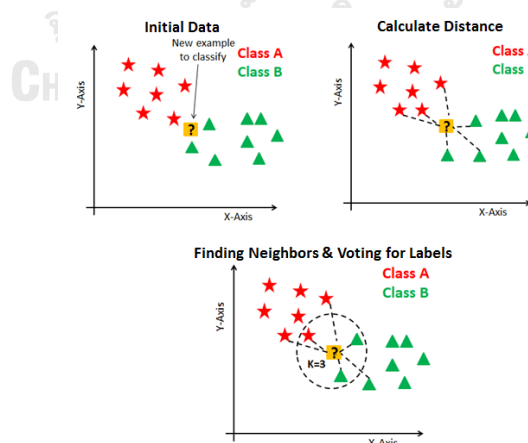
- **Average within centroid distance** หมายถึงค่าเฉลี่ยของระยะทางของแต่ละข้อมูลจากจุดศูนย์กลาง ซึ่งยิ่งค่าใกล้ 0 จะยิ่งดี แสดงให้เห็นว่าระยะห่างของข้อมูลภายในกลุ่มที่จัดขึ้นน้อย
- **Average within cluster distance** หมายถึงค่าเฉลี่ยของระยะทางหรือการกระจายตัวภายในกลุ่ม ยิ่งค่าน้อย ยิ่งมีการกระจายตัวของข้อมูลน้อยหรือมีการเกาะกลุ่มของข้อมูลมาก

- **Davies Bouldin** หรือ Davies-Bouldin index เป็นเกณฑ์การวัดคุณภาพของการจัดกลุ่มแบบ Clustering ที่ไม่ขึ้นกับจำนวนกลุ่มที่จะวิเคราะห์และไม่ขึ้นกับวิธีที่ใช้ในการแบ่งกลุ่ม จึงนิยมนำมาใช้ในการเปรียบเทียบประสิทธิภาพของอัลกอริทึมที่แตกต่างกันในการวิเคราะห์แบ่งกลุ่มข้อมูล โดยในการคำนวณจะเป็นการหาอัตราส่วนระหว่างผลรวมของการกระจายตัวของข้อมูลในแต่ละกลุ่ม และระยะห่างระหว่างกลุ่ม ซึ่งการแบ่งกลุ่มที่มีประสิทธิภาพที่ดีนั้น การกระจายตัวในกลุ่มจะต้องน้อย และจะต้องมีระยะห่างระหว่างกลุ่มมาก ดังนั้นค่าของ Davies Bouldin ที่น้อยที่สุด จะทำให้ได้การแบ่งกลุ่มที่ดีที่สุด (วีระศักดิ์ ช่องงูเหลือม, 2555)

2. อัลกอริทึมของการเรียนรู้ของเครื่องแบบมีผู้สอน (Supervised machine learning)

2.1 K-Nearest Neighbour (k-NN)

เป็นหนึ่งในอัลกอริทึมการจำแนกแบบ Lazy Model หรือการสร้างโมเดลแบบง่าย มีหลักการคือการดึงข้อมูลที่ใกล้เคียงตัวเองที่สุดมา k ตัว แล้วมานับจำนวนว่าข้อมูลเพื่อนบ้านใกล้เคียงทั้งหมด k ตัวนั้น เป็นกลุ่มใดมากที่สุด จำนวน k ที่น้อยเกินไปเป็นผลให้ความแม่นยำในการทำนายลดลงได้ แต่ถ้าหากจำนวนมากเกินไปก็จะไม่ใช่เรื่องดี เนื่องจากยังมีโอกาสครอบคลุมพื้นที่ของข้อมูลที่ไม่เกี่ยวข้องกันหรืออยู่คนละกลุ่มมากขึ้นด้วย และการเลือก k ที่เป็นเลขคู่ จะสามารถนำไปสู่ข้อสรุปที่ไม่ลงตัวได้เช่นกัน ทั้งนี้ขึ้นกับจำนวน Attributes ของข้อมูลที่น่ามาวิเคราะห์ด้วย

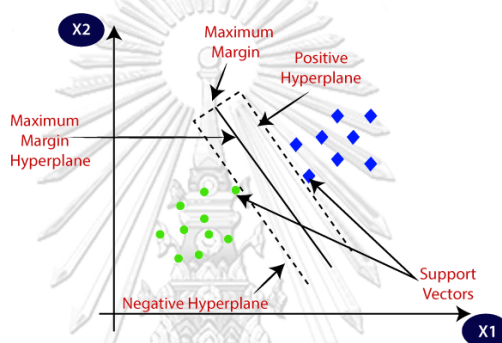


รูปที่ 7 K-Nearest Neighbor (k-NN)

แสดงกระบวนการวิเคราะห์ข้อมูลของอัลกอริทึม k-NN ที่จะมีการหาข้อมูลที่ใกล้ตัวเองมากที่สุด แล้วมานับจำนวนว่าข้อมูลที่ใกล้ตัวเองเหล่านั้นเป็นข้อมูลประเภทใดมากกว่า เพื่อจะทำการจำแนกประเภทของตัวข้อมูลนั้น

2.2 Support Vector Machine (SVM)

เป็นตัวจำแนกเชิงเส้น (Linear classification) ที่มีความแม่นยำสูงและมีประสิทธิภาพในการจำแนกข้อมูลที่มีมิติจำนวนมากได้ อาศัยหลักการของการหาสัมประสิทธิ์ของสมการเส้นตรงเพื่อสร้างเส้นในการแบ่งข้อมูลที่ถูกป้อนเข้าสู่เครื่องมือในกระบวนการสอนให้ระบบเรียนรู้ โดยจะทำการหาเส้นที่สามารถแบ่งแยกกลุ่มข้อมูลได้ดีที่สุด แนวความคิดของ SVM เกิดจากการนำค่าของกลุ่มข้อมูลมาวางในปริภูมิคุณลักษณะ (Feature space) จากนั้นจะทำการหาเส้นที่ใช้แบ่งข้อมูลทั้งสองออกจากกันโดยสร้างเส้นแบ่ง (Hyperplane) ที่เป็นเส้นตรงขึ้น และเพื่อให้ทราบว่าเส้นตรงที่แบ่งแยกกลุ่มออกจากกันนั้นเส้นตรงใดเป็นเส้นตรงที่ดีที่สุด



รูปที่ 8 Support Vector Machine (SVM)

แสดงกระบวนการวิเคราะห์ข้อมูลของอัลกอริทึม SVM ที่จะมีการขีดเส้นแบ่งข้อมูลเพื่อจำแนกกลุ่ม

หากข้อมูลที่นำมาใช้ในการวิเคราะห์ไม่ใช่ข้อมูลเชิงเส้น จะสามารถแก้ปัญหาได้โดยการนำ Kernel function มาใช้ ช่วยให้อัลกอริทึมสามารถ Optimize ค่าตัวแปรแบบ Polynomial ได้ โดยไม่ต้องเปลี่ยนรูปแบบและความสัมพันธ์ของคุณลักษณะของข้อมูลตั้งแต่ต้น สามารถเปลี่ยนข้อมูลที่มีมิติต่ำกว่าให้มีมิติสูงขึ้นเพื่อให้สามารถแบ่งข้อมูลแบบ Linear model ได้ การจะเลือกโมเดลของอัลกอริทึม SVM ควรจะเลือก Model ที่ไม่ Overfitting หรือไม่จำรูปแบบของข้อมูลที่ใช้สอน (Training dataset) มากจนเกินไป จนทำให้ไม่สามารถทำนายข้อมูลที่เป็น Unseen Data ได้

พารามิเตอร์ของ Support Vector Machine และ LibSVM

1. **Kernel type** เป็นการกำหนดว่าจะต้องมีการตรวจสอบความถูกต้องเพื่อป้องกันความน่าจะเป็นที่จะไม่มีโอกาสเกิดเหตุการณ์ขึ้นหรือไม่

2. **Estimation mode** ใช้สำหรับกรณีที่ต้องการหา Kernel density สามารถเลือกได้ 2 แบบ คือ Full จะสามารถกำหนด Bandwidth ได้ และ Greedy จะต้องกำหนด Bandwidth ขั้นต่ำและจำนวน Kernel
3. **Bandwidth selection** ใช้ในกรณีที่เลือก Mode full เพื่อตั้งค่า Kernel bandwidth สามารถกำหนดได้ว่าจะเป็น Heuristic หรือ Fix bandwidth
4. **Bandwidth** จะใช้ในกรณีที่ Bandwidth selection เลือกเป็น Fix ใช้กำหนดค่าจำเพาะของ Kernel bandwidth
5. **Minimum bandwidth** ใช้ในกรณีที่เลือก Mode greedy เพื่อกำหนดค่า Bandwidth ขั้นต่ำ
6. **Number of kernels** ใช้ในกรณีที่เลือก Mode greedy เพื่อกำหนดจำนวน Kernel
7. **Use application grid** เป็นการกำหนดว่าจำเป็นต้องใช้ Grid เพื่อวัด Kernel density ของโมเดลหรือไม่ เพื่อเพิ่มความเร็วและความแม่นยำในการทำงานของโมเดลที่สร้างขึ้น
8. **Application grid size** จะใช้ในกรณีที่เลือก Use application grid เพื่อกำหนดขนาดของ Grid ที่จะใช้ในการวัด

2.3 Naïve Bayes (NB)

เป็นการจำแนกประเภทรูปแบบหนึ่งที่อาศัยหลักการของความน่าจะเป็น (Probability) ตามทฤษฎีของเบย์ (Bayes' theorem) ถูกพัฒนาโดย Thomas Bayes ที่ว่าด้วยโอกาสที่จะเกิดเหตุการณ์ต่าง ๆ เมื่อเหตุการณ์หนึ่งได้เกิดขึ้นไปแล้ว เป็นเทคนิคที่ใช้ความน่าจะเป็นแบบมีเงื่อนไขเพื่อหาว่าสมมติฐานใดน่าจะถูกต้องที่สุด โดยใช้ความรู้ก่อนหน้า (Prior knowledge)

พารามิเตอร์ของ Naïve Bayes และ Naïve Bayes kernel

1. **Kernel type** กำหนดประเภทของ Kernel function ที่จะใช้ โดยมีประเภทของ Kernel ที่สามารถเลือกใช้ได้แก่ Dot, Radial, Polynomial, Neural, Anova, Epachnenikov, Gaussian combination, and Multiquadric
2. **C** เป็นค่า Complexity constant ที่ให้กำหนดความทนทานในกรณีที่ไม่สามารถจำแนกข้อมูลได้ โดยค่าสูงใช้สำหรับ Softer boundaries และค่าต่ำใช้สำหรับ Harder boundaries

แต่ถ้าค่านี้มากเกินไปจะทำให้เกิดการ Overfitting ได้ แต่ถ้าน้อยเกินไปจะเกิด Overgeneralization ได้

3. Convergence epsilon ใช้ในการ Optimize พารามิเตอร์เพื่อจำเพาะความแม่นยำเมื่ออยู่ในสถานะ KKT

4. Max iterations ใช้ในการ Optimize พารามิเตอร์เพื่อให้หยุดการทำซ้ำหลังจากที่ได้มีการกำหนดจำนวนรอบของการทำซ้ำ

5. Scale ใช้ในการตรวจสอบค่าของข้อมูลตัวอย่างหากเลือกพารามิเตอร์นี้ โดยจะถูกเก็บไว้ใช้สำหรับชุดทดสอบ

6. L pos เป็นปัจจัยของค่า C สำหรับข้อมูลตัวอย่างที่เป็น Positive

7. L neg เป็นปัจจัยของค่า C สำหรับข้อมูลตัวอย่างที่เป็น Negative

8. Epsilon ใช้กำหนดความจำเพาะของค่าคงที่ที่ไม่ไวพอ จะไม่มีการสูญหายของข้อมูลถ้าตั้งค่านี้นี้ใกล้เคียงกับค่าจริงในชุดข้อมูล

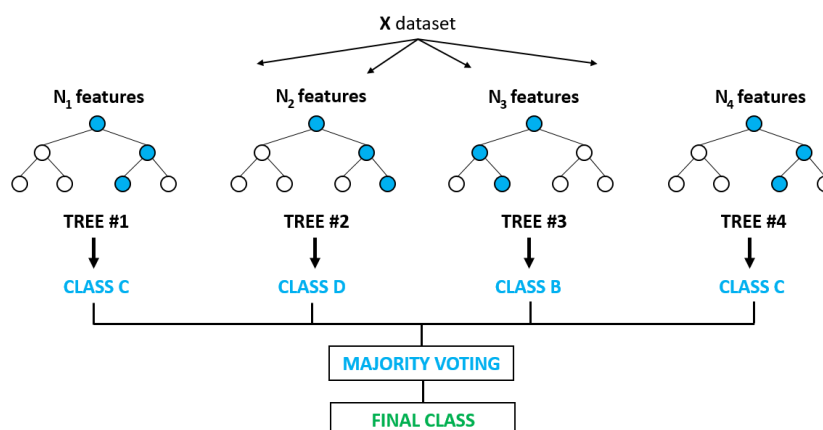
9. Balance cost เมื่อเลือกพารามิเตอร์นี้ จะมีการปรับค่า C pos และ C neg ให้สัมพันธ์กับขนาดของการจำแนกกลุ่ม

2.4 Decision tree

หรือต้นไม้ตัดสินใจ เป็นอัลกอริทึมที่ได้รับความนิยมเป็นอย่างมาก ใช้ในการสร้างโมเดลแบบ Rule-based คือการสร้างกฎ ถ้า/แล้ว จากคุณลักษณะของชุดข้อมูล โดยไม่มีสมการมากำกับความสัมพันธ์ระหว่างข้อมูลที่นำเข้ากับผลลัพธ์ที่ได้ มีลักษณะของโมเดลแบบต้นไม้กลับหัว โดยบนสุดคือราก และส่วนล่างที่ไม่มีการแตกแขนงต่อคือใบ การพิจารณาโมเดลนี้จะเริ่มจากจุดบนสุดที่เรียกว่า Root node ถ้าข้อมูลที่พบเป็นไปตามเงื่อนไขของโมเดลนั้น การตัดสินใจจะวิ่งไปทางซ้ายของ Root node ไปยังจุด Child node ซึ่งถ้าหากข้อมูลที่มาตามเส้นทางนี้ ตรงกับเงื่อนไขของ Child node จะถือว่าสิ้นสุด โดย Node สิ้นสุดนี้จะถูกเรียกว่า Leaf node ซึ่งปัจจัยสำคัญในการสร้างโมเดลของต้นไม้ตัดสินใจคือ “ความลึกของต้นไม้” แม้ว่าจำนวนชั้นของต้นไม้ นั้น ยิ่งลึกจะยิ่งละเอียดมากแค่ไหน แต่ก็ก็จะยิ่ง Overfitting มากขึ้น ในขณะเดียวกันหากจำนวนชั้นน้อยไป ก็จะไม่แม่นยำพอจะนำไปใช้งาน

2.5 Random forest

หรือป่าสุ่ม เป็นอัลกอริทึมประเภท Classification tree เช่นเดียวกับกับ Decision tree แต่เป็นโมเดลแบบ Ensemble คือการสร้างโมเดลในการจำแนกกลุ่มโดยนำ Decision tree หลาย ๆ ต้น มา Train ร่วมกัน เป็นหลักการที่มาจากกลไกการรวมการตัดสินใจของผู้ตัดสินใจจำนวนมากเข้าด้วยกัน มักให้ผลการตัดสินใจที่แม่นยำมากกว่าการพึ่งพาการตัดสินใจจากแหล่งเดียว ในทางสังคมเรียกวิธีการนี้ว่า “ปัญญาของฝูงชน” (Wisdom of crowd) โดยโมเดลที่ใช้ทำนายแต่ละตัวจะต้องเรียนรู้อย่างเป็นอิสระต่อกันมากที่สุดเหมือนเงื่อนไขในการตัดสินใจของฝูงชนว่าแต่ละคนจะต้องตัดสินใจด้วยตนเองให้มากที่สุดโดยไม่ได้นำเอาข้อมูลจากคนอื่นมาเป็นส่วนในการตัดสินใจ โดยแต่ละ Tree จะได้รับคุณลักษณะและข้อมูลทั้งหมดแบบสุ่ม และในขั้นตอนการทำนายผลลัพธ์ก็จะให้แต่ละ Decision tree ในป่าสุ่มนั้นทำนายผลลัพธ์ของตัวเอง และเลือกผลการทำนายผลลัพธ์สุดท้ายจากผลลัพธ์ที่ได้รับการโหวตมากที่สุด เรียกเทคนิคนี้ว่า Bagging หรือ Bootstrapping



รูปที่ 9 Random forest

แสดงแนวคิดของโมเดลแบบสุ่มป่าที่เกิดจากนำต้นไม้ตัดสินใจหลายต้นมาประกอบกัน โดยสุดท้ายจะพิจารณาผลลัพธ์สุดท้ายของการจำแนกกลุ่มจากผลลัพธ์ที่ได้รับการโหวตสูงสุด

พารามิเตอร์ของต้นไม้ตัดสินใจ (Decision tree) และป่าสุ่ม (Random forest)

- 1. Criterion** เป็นการเลือกเกณฑ์ว่า Attribute ไหนจะถูกเลือกในการแยกกลุ่ม ซึ่งในแต่ละเกณฑ์ที่มีให้เลือกนั้น ค่าที่ใช้ในการแยกกลุ่มจะถูกเลือกตามความเหมาะสมตามเกณฑ์นั้น ๆ ซึ่งสามารถเลือกเกณฑ์ได้ดังนี้

- **Information gain** ความไม่เป็นระเบียบของทุก Attribute จะถูกคำนวณ และจะเลือกเฉพาะ Attribute ที่มีความไม่เป็นระเบียบน้อยที่สุดมาใช้ในการแยกกลุ่ม ซึ่งวิธีนี้จะมี Bias ได้จากการเลือก Attribute ที่มีค่าเป็นตัวเลขขนาดใหญ่
- **Gain ratio** เป็นความหลากหลายของข้อมูลที่ได้รับที่ถูกนำมาใช้กับข้อมูลในแต่ละ Attribute เพื่อทำให้เกิดความกว้างและความเป็นรูปแบบเดียวกันของข้อมูลใน Attribute
- **Gini index** เป็นการวัดความไม่เท่ากันระหว่างการกระจายของตัวแปรที่ติดฉลาก โดยในการแยกกลุ่มจาก ผลของ Attribute ที่ถูกเลือกนั้น จะเกิดการลดลงของค่าเฉลี่ย Gini Index ของผลของข้อมูลย่อยในชุดข้อมูล
- **Accuracy** คือ Attribute ที่ถูกเลือกในการแยกกลุ่มที่ให้ค่าความถูกต้องมากที่สุดของทั้งต้นไม้ทั้งหมด
- **Least square** คือ Attribute ที่ถูกเลือกในการแยกกลุ่มที่ให้ค่าระยะห่างกำลังสองที่น้อยที่สุดระหว่างค่าเฉลี่ยของแต่ละ Node กับค่าจริง

2. **Maximal depth** คือความลึกของต้นไม้ ซึ่งขึ้นกับขนาดและคุณลักษณะของชุดข้อมูล ตัวอย่าง โดยพารามิเตอร์นี้ใช้จำกัดความลึกของต้นไม้ตัดสินใจ ถ้าค่าถูกตั้งเป็น -1 ค่าความลึกที่มากที่สุดของต้นไม้ก็จะไม่มีขอบเขต ในกรณีนี้ต้นไม้ก็จะถูกสร้างขึ้นจนกว่าจะเป็นไปตามเกณฑ์ แต่ถ้าค่าถูกตั้งเป็น 1 ต้นไม้ก็จะถูกสร้างโดยมีแค่ Node เดียว

3. **Apply pruning** โมเดลของต้นไม้ตัดสินใจจะถูกตัดแต่งหลังจากสร้างเสร็จ ตรวจสอบโดยบางกิ่งจะถูกแทนที่ด้วยใบตามพารามิเตอร์ Confidence

4. **Confidence** พารามิเตอร์นี้จำเพาะกับระดับความเชื่อมั่นที่ใช้ในการคำนวณค่าความผิดพลาดที่อาจเกิดขึ้นได้จากขั้นตอนการตัดแต่ง

5. **Apply prepruning** พารามิเตอร์นี้จะจำเพาะในกรณีที่มีเกณฑ์ในการหยุดมากกว่า Maximal Depth เหมาะสำหรับระหว่างการสร้างโมเดลต้นไม้ตัดสินใจ ตรวจสอบโดยใช้พารามิเตอร์ Minimal gain, Minimal leaf size, Minimal size of split, and Number of prepruning alternatives เป็นเกณฑ์ในการหยุด

6. Minimal gain จำนวนของ Node จะถูกคำนวณก่อนจะเกิดการแยก ซึ่ง Node จะถูกแยกถ้าจำนวนนั้นมากกว่า Minimal gain ถ้าค่า Minimal gain สูงจะมีการแยกกิ่งน้อย และต้นไม้จะมีขนาดเล็ก ดังนั้นถ้าค่าสูงไปจะไม่เกิดการแยกกิ่ง ต้นไม้ที่ได้ก็จะมีแค่ Node เดียว

7. Minimal leaf size ขนาดของใบก็คือจำนวนตัวอย่างที่อยู่ในกลุ่มนั้น โดยต้นไม้จะถูกสร้างขึ้นในแบบที่ทุกใบจะต้องมีจำนวนตัวอย่างอย่างน้อยตามพารามิเตอร์นี้

8. Minimal size for split ขนาดของ Node เป็นจำนวนของตัวอย่างที่อยู่ในกลุ่มนั้น ถ้าขนาดของ node มากกว่าหรือเท่ากับพารามิเตอร์นี้ เฉพาะ Node นั้นจะถูกแยกออกมา

9. Number of prepruning alternatives เมื่อการแยกถูกป้องกันโดยกระบวนการก่อนตัดแต่งใน Node ใด Node หนึ่ง พารามิเตอร์นี้จะให้จำนวน Node ทางเลือกอื่นในการแยก ซึ่งจะเกิดขึ้นขณะกระบวนการ Prepruning ตอนที่สร้างต้นไม้ เมื่อเกิดการป้องกันการแยกที่ Node นั้น ขณะที่มีการแยกเกิดขึ้นก็ไม่สามารถสร้างต้นไม้ได้ ในกรณีนี้ Node ทางเลือกก็จะพยายามแยกออกมา

10. Random splits หากเลือกพารามิเตอร์นี้จะทำให้เกิดการแบ่งจำนวน Attributes ให้ถูกเลือกแบบสุ่มแทนที่การ Optimize ในการเลือกแบบสุ่มนั้นจะมีรูปแบบในการสุ่มระหว่างค่าต่ำสุดและค่าสูงสุดของ Attribute ที่ทำการวิเคราะห์ หากใช้งานพารามิเตอร์นี้ในขณะที่ยกเลิกการ Pruning ป่าสุ่มจะทำให้ได้ป่าสุ่มโดยสมบูรณ์ (Extra-tree) และช่วยเพิ่มความเร็วให้กับโมเดลในกระบวนการวิเคราะห์

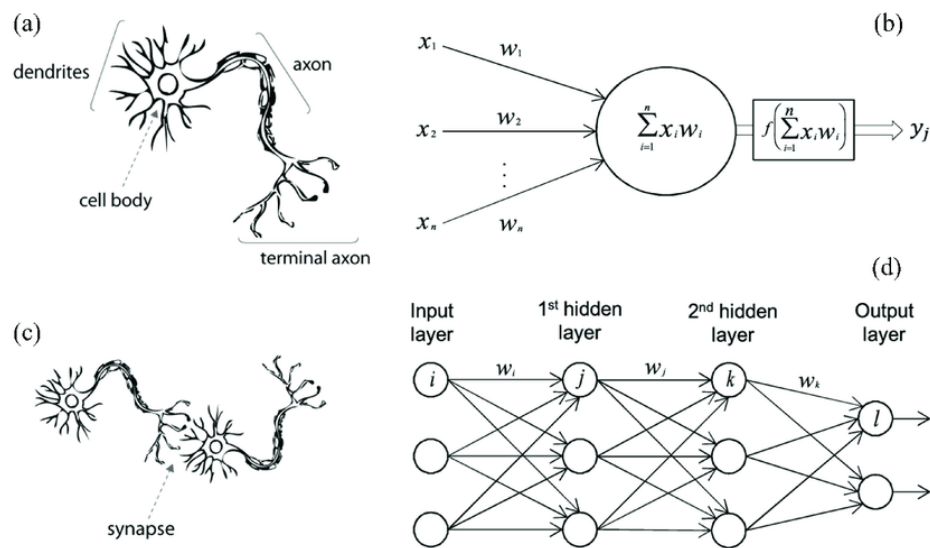
11. Voting strategy เลือกวิธีในการโหวต มี 2 แบบ ได้แก่ Confidence vote กับ Majority vote

12. Use local random seed ใช้เลือกเวลาที่ต้องการสุ่มข้อมูล เพื่อทำให้ได้ข้อมูลที่สุ่มเหมือนกันในทุกครั้ง

2.6 Artificial Neural Network (ANN)

หรือโครงข่ายประสาทเทียม เป็นอัลกอริทึมการเรียนรู้ที่ถูกพัฒนาขึ้นโดยอาศัยหลักการทำงานแบบเดียวกับระบบประสาทของมนุษย์ โดยใช้อัลกอริทึมหลายตัวทำงานพร้อมกันเป็นเครือข่าย อาศัยคุณสมบัติการทำงานของสมองมนุษย์ที่มีลักษณะคือ การที่เห็นภาพบางส่วนก็สามารถเข้าใจได้แล้วว่าสิ่งนั้นคืออะไร จากการเชื่อมโยงข้อมูลต่าง ๆ ในสมอง สามารถทำความเข้าใจและพัฒนาทักษะใน

หลาย ๆ ด้านที่หลากหลายและไม่เกี่ยวข้องกันได้ เข้าใจตรรกะและกติกาซึ่งเป็นเรื่องนามธรรมแม้ไม่เคยได้เห็นหรือสัมผัส และการเชื่อมโยงประมวลผลเพื่อแก้ไขปัญหาที่ซับซ้อน ทั้งหมดที่กล่าวมานี้ แม้จะไม่ได้ช่วยให้อัลกอริทึม ANN มีความคิดแบบมนุษย์ แต่ก็ช่วยให้สามารถเข้าใจชุดข้อมูลโดยไม่ต้องอาศัยข้อมูลจำนวนมาก และเห็นรูปแบบที่ซ่อนอยู่ในชุดข้อมูล ทำให้การคาดการณ์ข้อมูลบางประเภทเป็นไปได้เป็นอย่างดี



รูปที่ 10 Artificial Neural Network (ANN)

แสดงตัวอย่างแนวคิดในการสร้างโมเดลจากอัลกอริทึมโครงข่ายประสาทเทียม

ซึ่งเลียนแบบลักษณะการทำงานของเซลล์ประสาทของมนุษย์ (Meng, Hu, & Ancey, 2020)

สมองของมนุษย์ประกอบด้วยหน่วยประมวลผลพื้นฐานที่เรียกว่าเซลล์ประสาท (Neuron) และมีจุดต่อจำนวนโครงข่ายประสาทที่ประกอบด้วยส่วนสำคัญทั้งหมด 3 ส่วน คือ โยประสาท (Dendrite) ตัวเซลล์ (Soma) และแกนประสาท (Axon) และแต่ละโครงข่ายประสาทจะเชื่อมต่อกันโดยจุดประสานประสาท (Synapse) ซึ่งใน ANN การประมวลผลต่าง ๆ เกิดขึ้นในหน่วยประมวลผลย่อยที่เรียกว่า Node เป็นการจำลองลักษณะการทำงานของเซลล์ที่มีการส่งสัญญาณระหว่าง Node ที่เชื่อมต่อกัน โดย ANN จะประกอบด้วย Input Output ค่าน้ำหนัก ฟังก์ชันผลรวม และฟังก์ชันการแปลง (Irwin, NN application in control 1995) จะมีการจำลองให้ Input แต่ละอันมีค่าน้ำหนักที่เป็นตัวกำหนดของ Input ซึ่งแต่ละ Node จะมีค่า Threshold เป็นตัวกำหนดน้ำหนักรวมของ Input ว่าจะมีขนาดไหนจึงจะสามารถส่ง Output ไปยัง Node อื่นได้ เมื่อนำ Node มาต่อกันให้ทำงานร่วมกันแล้วก็จะเหมือนกับปฏิกิริยาเคมีที่เกิดขึ้นในสมองมนุษย์

พารามิเตอร์ของโครงข่ายประสาทเทียม (Artificial Neural Network)

1. Hidden layers พารามิเตอร์นี้ใช้ในการอธิบายโครงสร้างของ Neural Network ได้ จะเห็นเมื่อมีการสร้างโมเดลเท่านั้น ซึ่งทุกครั้งจำนวนของ Node จะมากกว่าค่าที่ตั้งไว้ค่าหนึ่งเสมอ เพราะจะมี Node ที่เพิ่มเติมเข้ามาในแต่ละชั้น แต่ Node ที่เพิ่มขึ้นนี้ ไม่ได้เชื่อมกับชั้นก่อนหน้า ถ้าหากตั้งค่าเป็น -1 ขนาดของชั้นนั้นจะถูกคำนวณจากจำนวนของ attributes ของชุดข้อมูลที่ใส่เข้าไป ในกรณีนี้ขนาดของชั้นจะถูกตั้งโดย (จำนวน Attributes + จำนวนของ Class)/2+1 แต่ถ้าหากไม่ตั้งค่าใด ๆ ก็จะเป็นค่าของประเภทข้อมูลและขนาดที่เท่ากับสมการข้างต้น ซึ่งถ้ามีแค่ชั้นเดียวโดยไม่มี Node ใด ๆ Input node ก็จะเชื่อมกับ Output node เลย โดยที่ไม่มี Hidden layer

2. Training cycles ใช้ระบุจำนวนรอบในการฝึก Neural Network ในการถ่ายทอดก่อนหน้า ค่าผลลัพธ์ที่แสดงออกมาจะถูกเปรียบเทียบกับค่าที่ถูกต้องเพื่อหาค่าความผิดพลาด โดยค่าความผิดพลาดนั้นจะส่งกลับไปยัง Network ซึ่งจะใช้ข้อมูลนี้ในการถ่วงน้ำหนักของการเชื่อมโยงข้อมูลเพื่อที่จะลดค่าความผิดพลาดที่เกิดขึ้นน้อยลง

3. Learning rate ใช้เพื่อบอกว่าเราจะเปลี่ยนการถ่วงน้ำหนักในแต่ละขั้นตอนเท่าไร ซึ่งไม่ควรมีค่าเท่ากับ 0

4. Momentum เต็มสัดส่วนของการถ่วงน้ำหนักก่อนหน้ามายังอันปัจจุบัน เพื่อป้องกันให้โครงข่ายที่ได้เป็นไปในทิศทางที่เหมาะสม

5. Decay เป็นพารามิเตอร์ที่จะใช้ในกรณีที่ Learning rate ควรลดค่าลงระหว่างที่เกิดการเรียนรู้ขึ้น

6. Shuffle เป็นพารามิเตอร์ที่จะใช้ในกรณีที่ข้อมูลที่นำเข้าควรมีการสับเปลี่ยนก่อนการเรียนรู้ แม้ว่า จะเพิ่มการใช้ความจำ แต่ก็แนะนำให้ใช้ในกรณีที่ข้อมูลถูกจัดเรียงไว้ก่อนแล้ว

7. Normalize เนื่องจากอัลกอริทึม Neural Network มีช่วงค่าของแต่ละ Attribute อยู่ที่ -1 ถึง +1 ดังนั้นจึงต้องมีการ Normalize ข้อมูลก่อนที่จะมีการเรียนรู้แม้ว่าจะเพิ่มระยะเวลาในการประมวลผลก็ตามแต่เป็นขั้นตอนที่จำเป็นอย่างมากในหลายกรณีด้วยกัน

8. Error epsilon การหารูปแบบที่เหมาะสมจะหยุดลงเมื่อค่าความผิดพลาดของการฝึกที่ได้ต่ำกว่าค่า Epsilon ที่กำหนด

9. Use local random seed ใช้เลือกเวลาที่ต้องการสุ่มข้อมูล เพื่อให้ได้ข้อมูลที่สุ่มเหมือนกันในทุกครั้ง

ในโปรแกรม RapidMiner studio มี Operator ที่ชื่อว่า Performance (Classification) ใช้สำหรับการประเมินประสิทธิภาพของการจำแนกกลุ่มแบบการเรียนรู้โดยมีผู้สอน อย่างการใช้ต้นไม้ตัดสินใจ หรือโครงข่ายประสาทเทียม ซึ่งจะทำการกำหนดประเภทของการเรียนรู้โดยอัตโนมัติและคำนวณเกณฑ์ที่พบบ่อยที่สุดสำหรับประเภทนั้น สำหรับการประเมินประสิทธิภาพเชิงสถิติของโมเดลการจำแนกกลุ่มแบบมีผู้สอนนั้น ชุดของข้อมูลควรมีการติดฉลาก (Label) ซึ่งตัวแปรที่ติดฉลากนี้จะเก็บค่าที่สังเกตได้จริง ในขณะที่ตัวแปรทำนายจะเก็บค่าของการทำนายโดยโมเดลการจำแนกกลุ่มนั้นแล้วสามารถนำมาใช้เปรียบเทียบเพื่อประเมินประสิทธิภาพได้

พารามิเตอร์ของ Performance (Classification)

1. **Accuracy** คือ ความถูกต้องของวิธีการวิเคราะห์ที่วัดได้ค่าใกล้เคียงกับค่าที่แท้จริงมากที่สุด (จำนวนสัมพัทธ์ของตัวอย่างที่จำแนกอย่างถูกต้องหรือร้อยละของการทำนายที่ถูกต้อง)
2. **Classification error** คือ ประเภทของความผิดพลาดในการจำแนกข้อมูล (จำนวนสัมพัทธ์ของ ตัวอย่างที่ไม่ได้ถูกนำมาจัดจำแนกหรือร้อยละของการทำนายที่ไม่ถูกต้อง)
3. **Kappa** คือ ใช้วัดความสอดคล้องของข้อมูลที่เป็นส่วนเดียวกันหรือคนเดียวกัน โดยผู้วัด 2 คนหรือ เครื่องมือวัด 2 ชนิด โดยข้อมูลที่ใช้ในการวิเคราะห์เป็นข้อมูลที่เป็นจำนวนนับ (ซึ่งเป็นสถิติที่ใช้ในการจัดจำแนกโดยทั่วไป มักใช้เป็นตัวชี้วัดที่มีประสิทธิภาพมากกว่าการคำนวณเปอร์เซ็นต์การคาดคะเนความถูกต้องแบบง่าย เนื่องจากเป็นการทำนายความถูกต้องที่เกิดขึ้นโดยบังเอิญ)
4. **Weighted mean recall** คือ ค่าเฉลี่ยถ่วงน้ำหนักของอัตราส่วนระหว่างการค้นพบข้อมูลที่ถูกต้องจากจำนวนข้อมูลที่ทำการค้นคืนมา
5. **Weighted mean precision** คือ ค่าเฉลี่ยถ่วงน้ำหนักของอัตราส่วนระหว่างการค้นพบข้อมูลที่ถูกต้องกับจำนวนข้อมูลที่ถูกต้องทั้งหมด
6. **Spearman Rho** คือ ความสัมพันธ์ระหว่างค่าจริงกับการทำนาย โดยใช้ Spearman Rho ซึ่งเป็นการวัดความสัมพันธ์เชิงเส้นระหว่างสองตัวแปรและกำหนดให้ตัวแปรสองตัว ในกรณีนี้คือตัวแปรติดฉลากและตัวแปรที่ทำนาย โดยปกติจะมีค่าอยู่ระหว่าง -1.00 ถึง 1.00

- ถ้ามีค่าติดลบหมายความว่า ตัวแปร 2 ตัวมีความสัมพันธ์ในทิศทางตรงกันข้าม

- ถ้ามีค่าเป็นบวกหมายความว่าตัวแปร 2 ตัวมีความสัมพันธ์ในทิศทางเดียวกัน
- ถ้ามีค่าเป็น 0 หมายความว่าตัวแปร 2 ตัวไม่มีความสัมพันธ์กัน

7. Kendall Tau คือ ความสัมพันธ์ระหว่างค่าจริงกับการทำนาย โดยใช้ Kendall Tau ซึ่งเป็นการวัดความสัมพันธ์และวัดความแข็งแกร่งของความสัมพันธ์ระหว่างตัวแปรสองตัว และกำหนดให้ตัวแปรสองตัวในกรณีนี้คือ ตัวแปรติดฉลากและตัวแปรที่ทำนาย โดยปกติจะมีค่าอยู่ระหว่าง -1.00 ถึง 1.00

- ถ้ามีค่าติดลบหมายความว่าตัวแปร 2 ตัวมีความสัมพันธ์ในทิศทางตรงกันข้าม
- ถ้ามีค่าเป็นบวกหมายความว่าตัวแปร 2 ตัวมีความสัมพันธ์ในทิศทางเดียวกัน
- ถ้ามีค่าเป็น 0 หมายความว่าตัวแปร 2 ตัวไม่มีความสัมพันธ์กัน

8. Absolute error คือ ค่าความคลาดเคลื่อนสัมบูรณ์เฉลี่ยของการทำนายจากค่าจริง โดยที่ค่าของตัวแปรติดฉลากคือค่าจริง

9. Relative error คือ ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการทำนายจากค่าจริงหารด้วยค่าจริง โดยที่ค่าของตัวแปรติดฉลากคือค่าจริง

10. Relative error lenient คือ ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการทำนายจากค่าจริงหารด้วยจำนวนสูงสุดของค่าจริงและการทำนาย โดยที่ค่าของตัวแปรติดฉลากคือค่าจริง

11. Relative error strict คือ ค่าเฉลี่ยของความคลาดเคลื่อนสัมบูรณ์ของการทำนายจากค่าจริงหารด้วยค่าต่ำสุดของค่าจริงและการทำนาย โดยที่ค่าของตัวแปรติดฉลากคือค่าจริง

12. Normalized absolute error คือ ค่าความคลาดเคลื่อนสัมบูรณ์หารด้วยค่าความคลาดเคลื่อนที่เกิดขึ้นจากการทำนายโดยเฉลี่ย

13. Root Mean Squared Error (RMSE) คือ รากของค่าคลาดเคลื่อนกำลังสองเฉลี่ย ซึ่งเป็นเครื่องมือในการเปรียบเทียบโมเดล โดยโมเดลใดที่มีค่า RMSE ต่ำสุดจะเป็นตัวแบบที่ดีที่สุด

14. Root relative squared error คือ รากของค่าคลาดเคลื่อนกำลังสองสัมพัทธ์

15. Squared Error คือ ค่าความคลาดเคลื่อนกำลังสองเฉลี่ย

16. Correlation คือ สถิติที่ใช้หาความสัมพันธ์ตัวแปรระหว่างตัวแปรติดฉลากและตัวแปรที่ทำนาย ซึ่งค่าสหสัมพันธ์ที่คำนวณได้ เรียกว่า ค่าสัมประสิทธิ์สหสัมพันธ์ (Correlation coefficient)

17. Squared correlation คือ ค่าสัมประสิทธิ์สหสัมพันธ์กำลังสองของตัวแปรติดฉลากและตัวแปรที่ทำนาย หากโมเดลใดมีค่า R-squared สูง (เข้าใกล้ 1) แสดงว่าโมเดลนั้นสามารถอธิบายค่าที่เกิดขึ้นได้มาก

18. Cross-entropy คือ ผลรวมของลอการิทึมของความเชื่อมั่นของตัวแปรติดฉลากที่แท้จริงหารด้วยจำนวนตัวอย่าง

2.5 การศึกษาทรานสคริปโตมและโปรตีโอมในผู้ป่วยออทิซึมสเปกตรัม (Transcriptomic and Proteomic in ASD)

ทรานสคริปโตมิกส์ เป็นการศึกษาการแสดงออกของยีนที่เป็น RNA transcript ทั้งหมดภายในเซลล์ ได้แก่ mRNA ที่เกิดจากการแสดงออกของยีนต่าง ๆ ในจีโนม ซึ่งแสดงถึงรูปแบบการแสดงออกของยีนในเซลล์ที่สภาวะใดสภาวะหนึ่งของเซลล์ ในการศึกษาทรานสคริปโตมจะใช้วิธีการที่เป็น High throughput อย่างเทคนิคไมโครอาร์เรย์ (Microarray analysis) และ Next-generation sequencing ในความเป็นจริงหลังจากการศึกษาจีโนมมนุษย์ พบว่าผลผลิตของยีนภายในเซลล์มีความหลากหลายมากกว่าจำนวนยีนที่ประมาณไว้มาก ซึ่งความแตกต่างของการแสดงออกและกระบวนการการดัดแปลงของ RNA transcript (RNA processing) ในเซลล์ต่างชนิดกันและในสภาวะต่าง ๆ กัน เป็นปัจจัยที่สำคัญอันหนึ่งที่ทำให้เกิดความหลากหลายของผลผลิตของยีนในมนุษย์ ดังกล่าว ดังนั้น การศึกษาเกี่ยวกับการแสดงออกของ RNA transcript หรือ ทรานสคริปโตมิกส์ (Transcriptomics) จึงมีประโยชน์อย่างมากในการที่จะทำให้เกิดความเข้าใจเกี่ยวกับชีววิทยาของมนุษย์ รวมทั้งกลไกและสาเหตุของโรคต่าง ๆ ที่อาจจะเกิดจากเปลี่ยนแปลงของทรานสคริปโตม

ส่วนโปรตีโอมิกส์ เป็นศาสตร์การศึกษาที่เกิดขึ้นภายหลังการศึกษาหาลำดับยีนทั้งหมดในสิ่งมีชีวิต (Autism Spectrum Disorders Working Group of The Psychiatric Genomics) ซึ่ง Mark Wilkins นักชีวสารสนเทศศาสตร์ ชาวออสเตรเลีย ได้ให้คำจำกัดความไว้ว่า เป็นการศึกษาโปรตีนทุกชนิดที่มีอยู่ในสิ่งมีชีวิตชนิดใดชนิดหนึ่ง ซึ่งโปรตีนทุกชนิดนั้นเป็นผลผลิตของยีนทั้งหมด โดยศึกษาโครงสร้าง การทำงาน บริเวณและภาวะที่พบโปรตีนเหล่านั้น รวมถึงความสัมพันธ์ระหว่างโปรตีนชนิดต่าง ๆ ในวิถีการสร้างและสลายสารเคมีภายในเซลล์ โดยโปรตีโอมิกส์สามารถลงลึกได้มากถึงระดับการศึกษาโปรตีนทุกชนิดที่พบได้ในเซลล์ (Whole proteome) ที่เป็นผลผลิตของยีนทั้งหมด (Whole genome) ซึ่งใช้เทคนิคจำเพาะต่อการแยกโปรตีนชนิดต่าง ๆ ออกจากกัน เช่น เทคนิคทาง

Liquid chromatography เพื่อแยกโปรตีนตามลักษณะต่าง ๆ เช่น ความเป็นขั้ว จากนั้นใช้แมสสเปกโตรเมตรี (Mass spectrometry) เพื่อวิเคราะห์ชนิดและปริมาณของโปรตีน โดยเทียบกับฐานข้อมูลของโปรตีน (Protein database) ซึ่งจะทำให้ทราบชนิดของโปรตีนที่สร้างขึ้นภายในเซลล์ เทคนิคทางโปรตีโอมิกส์นี้สามารถตรวจพบโปรตีนได้นับหมื่นชนิดในการตรวจวัดเพียงครั้งเดียว ทำให้เห็นภาพรวมได้ว่าโปรตีนทั้งเซลล์ทำงานเชื่อมโยงกันอย่างไร และเกิดความผิดปกติกับโปรตีนอะไรบ้างที่ก่อให้เกิดโรค เช่น โรคมะเร็ง และโปรตีนที่มีความผิดปกติส่งผลกับโปรตีนอื่น และสร้างความเสียหายให้กับเซลล์และร่างกายอย่างไร เป้าหมายหลักอีกอย่างหนึ่งของการศึกษาทางด้านโปรตีโอมิกส์ในปัจจุบัน คือการค้นหาสารบ่งชี้ทางชีวภาพ (Biomarkers) เพื่อใช้วินิจฉัยโรคต่าง ๆ ให้มีความถูกต้องรวดเร็ว และแม่นยำมากขึ้น รวมทั้งเพื่อใช้พยากรณ์ผลของการรักษา ทำให้การรักษาโรคมมีประสิทธิภาพมากขึ้น

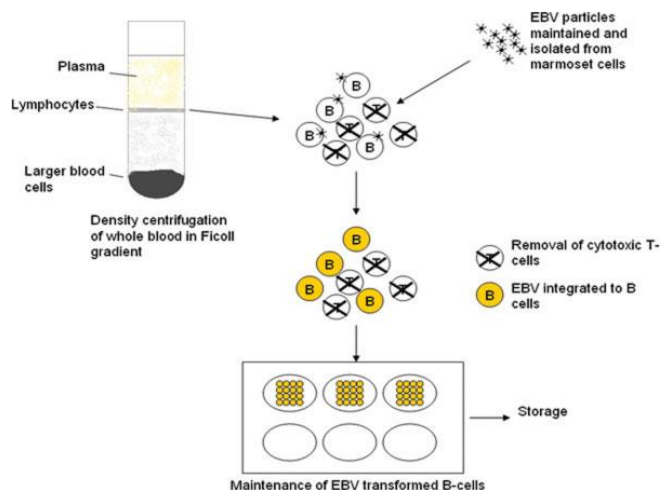
จากงานวิจัยที่ผ่านมาได้มีการศึกษาการแสดงออกของยีนและโปรตีนในผู้ป่วยโรคออทิซึมสเปกตรัมเป็นจำนวนมาก และมีการค้นพบว่ามียีนหรือโปรตีนที่มีการแสดงออกผิดปกติอย่างมีนัยสำคัญในผู้ป่วยออทิซึมสเปกตรัมเมื่อเทียบกับคนปกติ น่าสนใจเป็นอย่างยิ่งว่ารูปแบบทรานสคริปโตมจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยและคนปกติ พบการแสดงออกของยีนที่แตกต่างกันและมีความเกี่ยวข้องกับหน้าที่ทางระบบประสาทที่ส่งผลต่อโรคออทิซึมสเปกตรัม (Hu et al., 2009) นอกจากนี้มีหลายงานวิจัยด้วยกันที่ทำการศึกษารูปแบบโปรตีโอมของผู้ป่วยออทิซึมสเปกตรัมจากตัวอย่างต่าง ๆ ไม่ว่าจะเป็นเนื้อเยื่อสมองส่วนท้าย (Postmortem brain tissues) (Broek et al., 2014) ซีรัม (Corbett et al., 2007; Ramsey et al., 2013; Wetie et al., 2014) พลาสมา (Cortelazzo et al., 2016; Feng et al., 2017; L. Shen et al., 2018) เซลล์เม็ดเลือดขาวชนิดลิมโฟไซต์ (C. Shen et al., 2011) เลือดจากเด็กแรกเกิด (Neonatal blood) (Krakowiak et al., 2017; Zerbo et al., 2014) ปัสสาวะ (Yang & Rudser, 2016) และน้ำลาย (Castagnola et al., 2008; A. G. Ngounou Wetie et al., 2015; Armand G. Ngounou Wetie et al., 2015) ในการศึกษาโดย Broek et al. (2014) แสดงให้เห็นถึงโปรตีนจำนวนหนึ่งที่มีการแสดงออกแตกต่างในผู้ป่วยออทิซึมสเปกตรัมเทียบกับคนปกติ และมีความเกี่ยวข้องกับกระบวนการ Myelination กระบวนการควบคุม Synaptic vesicle และกระบวนการ Energy metabolism (Broek et al., 2014) นอกจากนี้จากการวิเคราะห์รูปแบบ Multiplex immunoassay ของตัวอย่างซีรัมจากผู้ป่วยออทิซึมสเปกตรัมจำนวน 37 ราย เทียบกับพี่น้องที่ไม่เป็นโรค ค้นพบโปรตีนที่แตกต่าง 12 ตัว โดยขึ้นกับอายุของคนไข้ ซึ่งมีความ

เกี่ยวข้องกับกระบวนการอักเสบ การเจริญเติบโต และการส่งสัญญาณของฮอริโมน แสดงให้เห็นถึงความสำคัญในการจัดผู้ป่วยออกเป็นกลุ่มย่อยและการวิเคราะห์ตัวอย่างโดยขึ้นกับอายุสำหรับการศึกษารูปแบบโปรตีโอมในคนไข้ (Ramsey et al., 2013) และในการการศึกษาพลาสมาโปรตีนจากผู้ป่วยออทิซึมสเปกตรัม 30 ราย เทียบกับคนปกติที่มีการพัฒนาตามปกติ 30 ราย โดย Cortelazzo et al. (2016) พบว่ารูปแบบการแสดงออกของไซโตไคน์ในเลือดของเด็กแรกเกิดที่เป็นออทิซึมสเปกตรัมอย่างมีนัยสำคัญกับความรุนแรงของโรค ซึ่งให้เห็นว่าการแบ่งผู้ป่วยออกเป็นกลุ่มย่อยตามความรุนแรงของลักษณะอาการของคนไข้อาจเป็นประโยชน์ต่อการค้นหาสารบ่งชี้ในเลือดสำหรับการวินิจฉัยและจัดกลุ่มผู้ป่วยออทิซึมสเปกตรัมได้ (Cortelazzo et al., 2016) ดังนั้นแม้ว่าจะมีงานวิจัยมากมายที่ศึกษาเกี่ยวกับรูปแบบโปรตีโอม แต่ความท้าทายในการศึกษารูปแบบโปรตีโอมคือโปรตีนที่ค้นพบเหล่านั้นมีคุณสมบัติในการทวนซ้ำต่ำ (Reproducibility) ไม่ได้พบในผู้ป่วยออทิซึมสเปกตรัมทุกรายเสมอไป ซึ่งปัญหานี้น่าจะมีผลมาจากความหลากหลายของลักษณะอาการของผู้ป่วยออทิซึมสเปกตรัม โดยงานวิจัยของ Pichitpunpong et al. ในปี 2019 (Pichitpunpong et al., 2019) ได้มีการศึกษารูปแบบโปรตีโอมและทรานสคริปโตมในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยแบบบูรณาการหลายระดับ หลังจากที่ได้แบ่งผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อย โดยได้เลือกผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติทางด้านภาษารุนแรงมาศึกษารูปแบบทรานสคริปโตมและโปรตีโอมในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ได้จากคนไข้ต่อ และทำการบูรณาการผลรูปแบบโปรตีโอมที่ได้กับรายชื่อโปรตีนที่แสดงออกผิดปกติในผู้ป่วยออทิซึมสเปกตรัมจากหลายสิ่งตัวอย่างที่เคยมีการตีพิมพ์ในฐานข้อมูล NCBI ซึ่งพบว่าจากโปรตีนที่มีการแสดงออกผิดปกติ นั้น มีโปรตีน DBI (Diazepam-binding inhibitor) ที่แสดงออกลดลงจริงในผู้ป่วย และมีความเกี่ยวข้องกับชีวพยาธิสภาพทางระบบประสาทและการอักเสบ จากงานวิจัยนี้ชี้ให้เห็นว่าการแบ่งกลุ่มคนไข้ออกเป็นกลุ่มย่อยและศึกษารูปแบบโปรตีโอมเฉพาะในผู้ป่วยกลุ่มย่อยนั้น จะสามารถนำไปสู่การค้นพบสารบ่งชี้ทางชีวภาพที่สามารถช่วยในการวินิจฉัยและเกี่ยวข้องกับชีวพยาธิสภาพของโรคแบบจำเพาะได้

2.6 เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines)

เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines) เป็นเซลล์ที่เกิดจากการนำเซลล์ Mature B lymphocyte มา Transformation โดยนิยมใช้ Epstein-Barr Virus (EBV) ใส่เข้าไปในเซลล์ ซึ่ง Lymphoblastoid Cell Lines ที่ได้นั้นจะมีคุณสมบัติที่จะเพิ่มจำนวนได้อย่าง

รวดเร็วกว่าและไม่จำกัด ซึ่งด้วยเหตุนี้จึงเป็นแหล่งของสารชีวโมเลกุล เช่น DAN RNA และโปรตีน ดังนั้นจึงเป็นที่นิยมนำมาใช้ในการศึกษาวิจัยในงานที่เกี่ยวข้อง เพราะวิธีนี้จะลดการเก็บตัวอย่างซ้ำจากผู้ป่วยได้ แต่อย่างไรก็ตามก็พบว่าเทคนิคนี้มีข้อจำกัดเนื่องจากการทำให้เกิด Lymphoblastoid Cell Lines อาศัยการติดเชื้อไวรัส ดังนั้นจะส่งผลทำให้เกิดยีนบางส่วนกลายพันธุ์ไปจากเดิม (Amoli, Carthy, Platt, & Ollier, 2008)



รูปที่ 11 Lymphoblastoid Cell Lines (LCLs)

แสดงขั้นตอนการ Transformation เซลล์เม็ดเลือดขาวชนิด Mature B-lymphocytes ด้วย Epstein-Barr Virus (EBV) (Sie, Loong, & Tan, 2009)

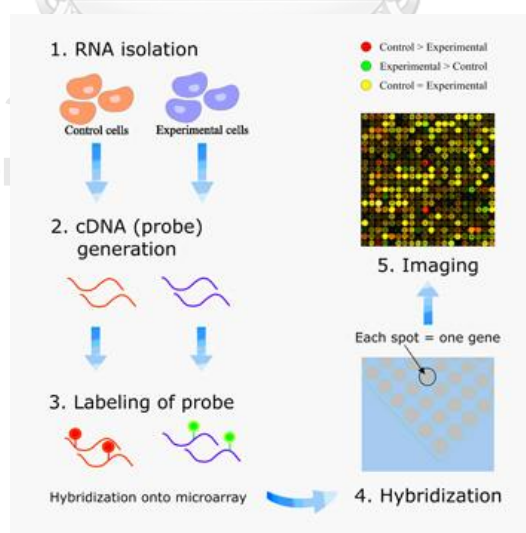
จากรูปเป็นการเปลี่ยน Mature B lymphocyte เป็น Lymphoblastoid Cell Lines โดยเริ่มจากการเก็บตัวอย่างเลือดจากผู้ป่วยและอาศัยเทคนิค Density centrifugation ในการแยกเซลล์ชนิดต่าง ๆ ซึ่งจากขั้นตอนนี้จะทำให้ได้ Mononuclear cell ออกมา โดยจะมีทั้ง T lymphocyte และ B lymphocyte จากนั้นจึงใส่ Epstein-Barr Virus ลงไป ตัวไวรัสจะทำการแทรกสารพันธุกรรมเข้าไปใน Mature B lymphocyte แบบ Nonrandom integration ซึ่งระหว่างนี้ก็จะทำการแยกเอา T lymphocyte ออกมาด้วย โดยอาศัยเทคนิคต่าง ๆ เช่น การทำให้เกิด Rosette formation ของ T lymphocyte โดยใช้เม็ดเลือดแดงของแกะ การเติมสาร T-cell mitogens เช่น Phytohemagglutinin (PHA) เพื่อกระตุ้นให้เกิด Blastogenesis และ Cell death ก่อนที่เซลล์จะเปลี่ยนไปเป็น Cytotoxic T cell การเติมสารพวก T cell immunosuppressant เช่น Cyclosporin A เพื่อไม่ให้ T lymphocyte เกิดการแบ่งตัว นอกจากนี้ยังมีการใช้ Magnetic Bead ที่ติด Anti-CD19 antibody ซึ่งจำเพาะกับ B Lymphocyte ในการแยกเอาเฉพาะ B lymphocyte (Amoli et al., 2008) เป็นต้น เมื่อทำการเปลี่ยน Mature B lymphocyte เป็น Lymphoblastoid

Cell Lines แล้วก็สามารถนำมาเพิ่มจำนวนและใช้ในการศึกษาได้ หรือสามารถนำว่าเก็บใน Growth medium ก่อนจะนำไปแช่แข็งเพื่อไว้ใช้ต่อไปได้

การใช้ lymphoblastoid Cell Lines นั้นมีข้อจำกัดเนื่องจากบางยีนถูกทำให้กลายจากการที่ EBV เข้าไปทำให้เซลล์กลายเป็น ซึ่งมีการค้นพบแล้ว 11 ยีน (Bernacki et al., 2005) เช่น ยีนของ Huntington's disease ดังนั้นการจะนำ Lymphoblastoid Cell Lines ไปใช้จะต้องทำการศึกษายีนที่ผิดปกติที่เกิดจากการ EBV หรือไม่

อย่างไรก็ตามจากการศึกษาก่อนหน้าที่ได้นำ Lymphoblastoid Cell Lines มาใช้ก็แสดงให้เห็นว่า Lymphoblastoid Cell Lines นั้นก็ยังคงเป็นทางเลือกที่ดีในการศึกษาสารชีวโมเลกุลดังเช่นงานวิจัยที่ใช้ Lymphoblastoid Cell Lines จากเด็กออทิสซึมสเปกตรัมและจากญาติที่เป็นคนปกติมาทำการศึกษาการแสดงออกของยีน ซึ่งในการศึกษาคั้งนี้ยังยืนยันอีกว่าการใช้ Lymphoblastoid Cell Lines นั้นสามารถใช้ในการศึกษาการแสดงออกของยีนในโรคออทิสซึมสเปกตรัมได้ (Baron, Liu, Hicks, & Gregg, 2006) และงานวิจัยของ Dr.Valerie Hu และคณะ ในปี ค.ศ. 2009 ที่ได้นำ Lymphoblastoid Cell Lines สกัด RNA เพื่อศึกษาการแสดงออกของยีนที่สัมพันธ์กับอาการทางคลินิกในผู้ป่วยโรคออทิสซึม โดยใช้ DNA microarray (Hu et al., 2009)

2.7 ข้อมูลไมโครอาร์เรย์ (Microarray data)



รูปที่ 12 cDNA microarray

แสดงขั้นตอนของกระบวนการทำ cDNA microarray จาก RNA ต้นแบบ (Staudt & Brown, 2000)

ข้อมูลไมโครอาร์เรย์ (Microarray data) มาจากการใช้เทคนิคไมโครอาร์เรย์ (Microarray) ที่ถูกพัฒนาขึ้นในช่วงทศวรรษที่ผ่านมาและถูกใช้ในการตรวจหาความผิดปกติของโครโมโซมอย่างแพร่หลายในปัจจุบัน โดยเทคนิคนี้ได้รวมเอาข้อดี ของเทคนิค G-banding และเทคนิค FISH เข้าไว้ด้วยกันซึ่งทำให้เทคนิคไมโครอาร์เรย์สามารถตรวจหาความผิดปกติของโครโมโซมทุกแห่งได้ด้วยความละเอียด (Resolution) สูงกว่าเทคนิค G-banding อย่างน้อย 10 เท่า เทคนิคไมโครอาร์เรย์อาศัยหลักการไฮบริดเซชัน (Hybridization) ระหว่าง DNA กับโพรบ (Probe) ที่จำเพาะกับแต่ละตำแหน่งบนโครโมโซมที่นำมาเรียงบนวัสดุ เช่น แผ่นสไลด์แก้ว หรือแผ่นชิป (Chip) การตรวจ โครโมโซมด้วยเทคนิคไมโครอาร์เรย์จึงเทียบเท่ากับการตรวจด้วยเทคนิค FISH หลาย ๆ ครั้งในคราวเดียวกัน เทคนิคไมโครอาร์เรย์สามารถตรวจหาความไม่สมดุลของจีโนม (Genome) ซึ่งเกิดจากความแปรผันของจำนวนซ้ำ (Copy Number Variants, CNVs) ของบริเวณต่าง ๆ ในจีโนม ได้แก่ การลดลงของจำนวนซ้ำ (Copy number loss) เช่นที่เกิดจาก deletion และการเพิ่มขึ้นของจำนวนซ้ำ (Copy number gain) ดังเช่นที่เกิดจากการ Duplication

การพัฒนาเทคโนโลยีไมโครอาร์เรย์เป็นไปอย่างรวดเร็วโดยเฉพาะอย่างยิ่งหลังจาก Human Genome Project เสร็จสมบูรณ์ ทำให้การออกแบบโพรบทำได้ง่ายขึ้นและครอบคลุมทั่วทั้งจีโนม (Tiling array) หรือครอบคลุมเฉพาะตำแหน่งที่สนใจ (Targeted array) เช่น ตำแหน่งที่ทราบแน่ชัดว่าทำให้เกิด Microdeletion และ Microduplication syndromes และส่วน Subtelomere ของโครโมโซม โดยปัจจุบันเทคนิคไมโครอาร์เรย์ที่ใช้ในการตรวจโครโมโซมมีอยู่ 2 ชนิดใหญ่ คือ

1) Array Comparative Genomic Hybridization (Array CGH) ซึ่งโพรบที่ใช้ทำมาจาก Bacterial Artificial Chromosome (BAC) หรือ Oligonucleotide ที่สังเคราะห์ขึ้น เทคนิค Array CGH มีหลักการเช่นเดียวกับเทคนิค CGH แบบดั้งเดิมโดยการนำดีเอ็นเอของผู้ป่วยและดีเอ็นเอปกติในปริมาณที่เท่ากันมา ติดฉลากด้วยสารฟลูออเรสเซนต์ที่สีต่างกันและให้ไฮบริดซ์กับโพรบที่อยู่บนไมโครอาร์เรย์ (แทนที่จะไฮบริดซ์กับโครโมโซมในระยะเมทาเฟสดังเช่นในเทคนิค CGH แบบดั้งเดิม) ข้อมูลที่ได้จะเป็นความเข้มของสารฟลูออเรสเซนต์ โดยเป็นค่า Log ratio หรือ Fold change เทียบกับดีเอ็นเอปกติ การวิเคราะห์ข้อมูลด้วยซอฟต์แวร์ที่จำเพาะกับไมโครอาร์เรย์จะมีการควบคุมคุณภาพของข้อมูลเพื่อลดผลบวกปลอมก่อนที่จะประมวลผลเป็นค่าจำนวนซ้ำ

2) Single Nucleotide Polymorphism (SNP) array สำหรับ SNP array ข้อดีของ SNP Array นอกจากใช้ตรวจการเปลี่ยนแปลงของ จำนวนซ้ำได้เช่นเดียวกับ Array CGH แล้ว ยังใช้

ตรวจหาบริเวณของจีโนมที่เรียกว่า Long Contiguous Stretches of Homozygosity (LCSHs) ซึ่งอาจเกิดได้จาก

- 1) การแต่งงานในเครือญาติที่ไม่ทราบจากการชักประวัติผู้ป่วย
- 2) Uniparental Disomy (UPD) แบบ Isodisomic UPD SNP array จะตรวจ Heterodisomic UPD ได้ก็ต่อเมื่อมี Isodisomic block ปะปนอยู่ด้วย
- 3) Loss of Heterozygosity (LOH) แบบ Copy number neutral ซึ่งมีจำนวนซ้ำ ปกติ ภาวะ LOH นี้พบได้บ่อยมากในโรคมะเร็ง

ข้อดีของเทคนิคไมโครอาร์เรย์ ได้แก่

- 1) เป็นการตรวจ DNA ดังนั้นจึงไม่จำเป็นต้องเพาะเลี้ยงเซลล์ และ ใช้กับสิ่งส่งตรวจที่ไม่มีเซลล์ที่มีชีวิตได้ เช่น ชิ้นเนื้อในพาราฟิน ตัวอ่อนที่เริ่มเน่าแล้ว (Macerated fetus) หรือ สิ่งส่งตรวจที่ไม่สามารถเพาะเลี้ยงเพื่อให้ได้โครโมโซมในระยะ Metaphase เช่น น้ำลาย
- 2) สามารถตรวจความผิดปกติและความแปรผันของจีโนมหลาย ๆ ตำแหน่งได้ในเวลาเดียวกัน
- 3) สามารถตรวจหาความผิดปกติของโครโมโซมขนาดเล็ก ๆ ที่มองไม่เห็นด้วยกล้องจุลทรรศน์

ทำให้เทคนิคไมโครอาร์เรย์เพิ่มประสิทธิภาพในการวินิจฉัยโรคทางพันธุกรรมกว่าวิธีดั้งเดิม โดยพบว่าในกลุ่มผู้ป่วยที่มีพัฒนาการช้า ผู้ป่วยที่มีความบกพร่องด้านสติปัญญาที่ไม่ทราบสาเหตุ ผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยที่มีความพิการแต่กำเนิดหลายอย่างร่วมกัน เมื่อตรวจด้วยเทคนิค G-banding จะพบความผิดปกติของโครโมโซมประมาณ 3-4% เมื่อตรวจด้วย เทคนิค G-banding ร่วมกับเทคนิค FISH ที่จำเพาะกับ Subtelomere ของแต่ละโครโมโซม จะพบความผิดปกติของโครโมโซมเพิ่มขึ้นเป็นประมาณ 7-9% และเมื่อตรวจด้วยไมโครอาร์เรย์อย่างเดียวจะพบความผิดปกติของโครโมโซมเพิ่มขึ้นเป็นประมาณ 15-20% ขึ้นอยู่กับชนิดและความละเอียดของเทคนิคไมโครอาร์เรย์ที่ใช้ นอกจากนี้เมื่อเริ่มมีการตรวจโครโมโซมด้วยเทคนิคไมโครอาร์เรย์เพิ่มขึ้น มีรายงานการค้นพบโรคและกลุ่มอาการใหม่ ๆ ที่เกิดจากความผิดปกติของโครโมโซมเพิ่มขึ้นตามมา เช่น 1q21.1 microdeletion syndrome 17q21.31 microdeletion syndrome ซึ่งความผิดปกติเหล่านี้มักมีขนาดเล็กมากหรือมีภาวะโมเสอซิซึม นอกจากนี้เทคนิคไมโครอาร์เรย์ยังช่วยให้นักวิจัยทราบข้อมูลเพิ่มเติมเกี่ยวกับยีนหรือค้นพบยีนที่เกี่ยวข้องกับโรคหลายโรค เช่น การขาดหายไปของยีน CHD7 ใน

CHARGE syndrome และโรคในกลุ่ม Complex disorder เช่น ออทิสซึมสเปกตรัม, ลมชัก, อัลไซเมอร์ ช่วยในการ Characterization และให้ข้อมูลเพิ่มเติมเกี่ยวกับ Genotype-phenotype correlation

ข้อจำกัดของเทคนิคไมโครอาร์เรย์ คือ

1) ไม่สามารถตรวจความผิดปกติแบบสมดุลที่ไม่มี การขาดหรือเกินของโครโมโซม เช่น Balanced Translocation และ Inversion

2) ไม่สามารถตรวจภาวะ Polyploidy ยกเว้นการใช้ SNP Array Genotype Data

3) ไม่สามารถตรวจภาวะโมเสอซิซึมที่มีโครโมโซมผิดปกติปริมาณน้อย ๆ (Low level mosaicism) เช่น น้อยกว่า 10-20% ทั้งนี้ขึ้นกับชนิดของไมโครอาร์เรย์ ซอฟต์แวร์ และชนิดของตัวอย่าง ภาวะโมเสอซิซึมของโครโมโซมที่พบแต่กำเนิดพบได้ไม่บ่อยและลักษณะทางคลินิกของผู้ป่วยมักจะแปรผันตรงกับปริมาณของเซลล์ที่ผิดปกติ อย่างไรก็ตามภาวะโมเสอซิซึมหรือ Cellular admixture มีความสำคัญมากในการตรวจวินิจฉัยโรคมะเร็ง เนื่องจากโรคมะเร็งมีลักษณะความผิดปกติที่หลากหลาย (Heterogeneity) และมักพบเซลล์ที่มีการกลายพันธุ์ต่างชนิดกันปนอยู่ด้วยกันหรือปนอยู่กับเซลล์ที่ปกติ ซึ่งการตรวจพบเซลล์ที่ผิดปกติแม้เพียงไม่มาก แต่อาจมีผลต่อการวินิจฉัย การพยากรณ์โรค การรักษาและติดตามการรักษา ดังนั้นก่อนการนำเทคนิคไมโครอาร์เรย์มาใช้ในห้องปฏิบัติการ จึงควรทำการทดสอบเพื่อหาระดับต่ำสุดของโมเสอซิซึมที่เทคนิคไมโครอาร์เรย์นั้นสามารถตรวจพบได้

ยิ่งเทคนิคไมโครอาร์เรย์มีความละเอียดมาก ก็จะมีโอกาสตรวจพบ CNVs มากขึ้น โดย CNVs ที่พบอาจเป็น

1) CNVs ที่ก่อให้เกิดโรค (Pathogenic CNVs) เช่น CNVs ที่พบในฐานข้อมูลของคนที่ เป็นโรค หรือ CNVs ที่มีขนาดใหญ่ (1-2 Kb) และอยู่ในตำแหน่งของยีน หรือ CNVs ที่เปลี่ยนแปลงจำนวนของยีน

2) CNVs ที่ไม่น่าจะก่อให้เกิดโรค (Likely benign CNVs) เช่น CNVs ที่พบเหมือนกับในพ่อหรือแม่ที่ปกติและยังไม่เคยมีรายงานมาก่อนในผู้ป่วยรายอื่น ๆ

3) CNVs ที่ไม่ทราบความสำคัญทางคลินิก (Variants of uncertain clinical significance) ซึ่งอันสุดท้ายนี้อาจก่อให้เกิดความกังวลแก่ผู้ป่วยและแพทย์ การตรวจยืนยัน CNVs ที่พบในพ่อและแม่

การใช้ Bioinformatics เพื่อพิจารณาความสำคัญของ CNVs การตรวจสอบกับฐานข้อมูล CNVs เช่น Database of Genomic Variants (DGV) และรายงานการค้นพบ CNVs ในคนที่เป็โรค และคนปกติ จะช่วยในการพิจารณาว่า CNVs นั้นเป็น Pathogenic หรือ Benign CNVs

ขณะนี้ มีแนวทางปฏิบัติในการใช้เทคนิคไมโครอาร์เรย์สำหรับการตรวจวินิจฉัยหลังคลอด ออกมาจากหลายแห่ง เช่น International Standard Cytogenomic Array (ISCA) ประเทศ สหรัฐอเมริกา Human Genetics Society of Australia Italian Society of Human Genetics (SIGU) เป็นต้น ISCA แนะนำให้ใช้เทคนิคไมโครอาร์เรย์เป็นการตรวจอันดับแรก เพื่อหาความผิดปกติ ของโครโมโซมในผู้ป่วยที่มีพัฒนาการช้า ผู้ป่วยที่มีความบกพร่องด้านสติปัญญาที่ไม่ทราบสาเหตุ ผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยที่มีความพิการแต่กำเนิดหลายอย่างร่วมกัน สำหรับข้อแนะนำใน การใช้เทคนิคไมโครอาร์เรย์เพื่อตรวจวินิจฉัยก่อนคลอดนั้นมีการศึกษาเพิ่มมากขึ้น SIGU แนะนำให้ใช้ เทคนิคไมโครอาร์เรย์เป็นการตรวจอันดับสองหลังจากการตรวจแครีโอไทป์ด้วยเทคนิค Banding โดยเฉพาะเมื่อผลตรวจอัลตราซาวด์พบว่าทารกในครรภ์มีความพิการแต่กำเนิด การศึกษาล่าสุดพบว่า การตรวจด้วยเทคนิคไมโครอาร์เรย์ช่วยให้พบความผิดปกติของโครโมโซมเพิ่มจากการตรวจด้วย เทคนิค G-banding ประมาณ 5.3- 6.4% สำหรับในผู้หญิงตั้งครรภ์อายุมาก หรือมีผลการตรวจคัด กรองจากเลือดมารดาเป็นบวก การตรวจด้วยเทคนิคไมโครอาร์เรย์ช่วยให้พบความผิดปกติของ โครโมโซมเพิ่มเติมประมาณ 1.7% นอกจากนี้การให้ข้อมูลกับผู้ป่วยควรต้องแจ้งล่วงหน้าเกี่ยวกับ โอกาสที่จะพบ CNVs ที่ไม่ทราบความสำคัญทางคลินิก ซึ่งพบได้ ประมาณ 0-4% เมื่อใช้ Targeted microarray และ 9-12% เมื่อใช้ SNP microarray ที่มีความละเอียดสูง นอกจากนี้เทคนิคไมโคร อาร์เรย์ยังได้ถูกนำไปใช้ในการตรวจคัดกรองหรือวินิจฉัยก่อนการนำตัวอ่อนกลับเข้าไปฝังตัวในมดลูก

ปัจจุบันเริ่มมีการใช้เทคนิคไมโครอาร์เรย์แพร่หลายมากขึ้นในการตรวจหาความผิดปกติของ โครโมโซม แม้ว่ายังไม่ได้ใช้ในทุกห้องปฏิบัติการโดยเฉพาะในประเทศไทยเนื่องจากมีราคาแพง ใน อนาคตคาดว่าเมื่อเทคนิคไมโครอาร์เรย์มีราคาต่ำลง และเมื่อมีข้อมูลความสำคัญทางคลินิกของ CNVs มากขึ้น เทคนิคไมโครอาร์เรย์น่าจะเป็นเทคนิคที่มีบทบาทมากขึ้น เนื่องจากมีความคุ้มทุนในด้านการ วินิจฉัย การรักษา และการวิจัย ทางการแพทย์

2.8 แมสสเปกโตรเมตรี (Mass spectrometry)

แมสสเปกโตรเมตรี (Mass spectrometry) เป็นเทคนิคการวิเคราะห์สารชนิดหนึ่งซึ่งอาศัยหลักการที่ตรวจวัดเชิงมวลของสารในรูปของไอออน โดยเทคนิคนี้มีถูกนำมาใช้ประโยชน์อย่างกว้างขวางและแพร่หลายอย่างมากในด้านต่าง ๆ ซึ่งสามารถให้ข้อมูลได้ทั้งในการวิเคราะห์เชิงคุณภาพ (Qualitative analysis) เช่น การทดสอบหาโครงสร้างทางเคมี (Chemical structure) การทดสอบหาชนิดหรือน้ำหนักโมเลกุลของสาร และการวิเคราะห์เชิงปริมาณ (Quantitative analysis) เช่น การทดสอบหาปริมาณของสารแต่ละชนิดที่ตรวจวัด

หลักการของพื้นฐานของเทคนิคแมสสเปกโตรเมตรีนี้ ประกอบไปด้วยการนำสารตัวอย่างเข้าสู่ส่วนผลิตไอออน (Ionization source หรือ Ion source) ของเครื่องแมสสเปกโตรมิเตอร์ โมเลกุลของสารตัวอย่างจะได้รับพลังงานที่สูงมากพอทำให้เกิดการไอออไนเซชันกลายเป็นไอออนของโมเลกุล (Molecular ion) หรือไอออนแม่ (Parent ion) และไอออนแม่ที่เกิดขึ้นนี้ ถ้ามีพลังงานสะสมสูงเกินกว่าพลังงานพันธะก็จะเกิดการแตกออกของพันธะภายในไอออนโมเลกุลนั้นกลายเป็นไอออนย่อย (Fragment ion) หรือ ไอออนลูก (Daughter ion) ที่มีขนาดเล็กกว่าเดิม โดยหากไอออนลูกนี้ยังคงมีพลังงานสูงอยู่ก็สามารถเกิดการแตกตัวไปเป็นไอออนย่อยได้อีก จากนั้นไอออนเหล่านี้จะถูกส่งผ่านไปยังส่วนวิเคราะห์มวล (Mass analyzer) ซึ่งจะทำการแยกไอออนทั้งหลายที่เกิดขึ้นตามค่ามวลต่อประจุ (Mass-to-charge ratio; m/z ratio) ของไอออนเหล่านั้น และไอออนที่ผ่านการแยกมวลแล้วจะถูกส่งไปสู่ส่วนตรวจวัด (Detector) ซึ่งจะทำการบันทึกปริมาณไอออนแต่ละชนิดที่ผ่านเข้ามา และประมวลผลข้อมูลที่ได้จากการตรวจวัดนี้ในรูปของแมสสเปกตรัม (Mass spectrum) ของสารตัวอย่าง

แม้ในการวิเคราะห์สารตัวอย่างด้วยเทคนิคนี้จะทำให้ตัวอย่างสูญเสียไปหลังผ่านกระบวนการเปลี่ยนให้เป็นไอออน แต่อย่างไรก็ตามเทคนิคการวิเคราะห์นี้ก็มีควมไวในการตรวจวัด (Sensitivity) สูงมาก นั้นหมายถึงสามารถตรวจสารที่มีปริมาณต่ำมากได้ นอกจากนี้ความสามารถในการแยกแยะไอออนตามค่ามวลต่อประจุของเครื่องแมสสเปกโตรเมตรีบางประเภท ยังสามารถทำได้โดยมีควมละเอียดของการวิเคราะห์สูงมากอีกด้วย ดังนั้นในปัจจุบันจึงได้มีการนำเทคนิคนี้มาประยุกต์ใช้ในการศึกษามากมายหลายด้าน ซึ่งนั้นรวมถึงการนำไปใช้ในงานวิจัยทางอณูชีววิทยาเช่นกัน (ศุภคร วินิชเวชารุ่งเรือง, 2553)

ประเภทของแมสสเปกโตรเมตรี

ประเภทของแมสสเปกโตรเมตรีนั้นสามารถจัดตามเทคนิคของการวิเคราะห์มวล (ศุภสร วิ
นิชเวชารุ่งเรือง, 2553)

ส่วนวิเคราะห์มวล (Mass analyzer)

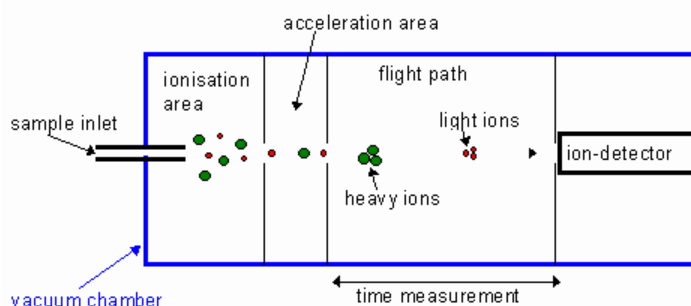
ส่วนวิเคราะห์มวลของแมสสเปกโตรเมตรีมีหลายชนิดด้วยกัน แต่จะขอยกมาแค่เฉพาะชนิดที่
จะใช้ในการวิเคราะห์นี้เท่านั้น ดังต่อไปนี้

1. Quadrupole mass analyzer

ส่วนวิเคราะห์มวลแบบ quadrupole ประกอบด้วยแท่งโลหะกลม 4 แท่งที่เชื่อมต่อกันทาง
ไฟฟ้า จัดวางตัวขนานกัน หลักการทำงาน คือ ไอออนจากแหล่งผลิตไอออนจะถูกโฟกัสด้วย Ion
Optic ให้เคลื่อนที่ผ่านเข้าไปในบริเวณช่องว่างตรงกลางระหว่างแท่งโลหะทั้งสี่ ผ่านกลไกการแยกแยะ
ไอออน ออกสู่ส่วนตรวจวัดหรือ Detector การแยกแยะไอออนทำโดยปรับค่าสนามไฟฟ้าภายใน
บริเวณช่องกลางของแท่งโลหะเหล่านี้ ทำให้ไอออนที่มี m/z เฉพาะค่าหนึ่ง ๆ เท่านั้น ที่สามารถ
เคลื่อนที่ในรูปแบบการส่ายหรือสั่นผ่านออกไปยัง Detector ได้ ในขณะที่ไอออนที่มีค่า m/z อื่น ๆ
ซึ่งไม่สอดคล้องกับสนามไฟฟ้าในขณะนั้น จะมีเส้นทางการเกิด Oscillation ไปชนผนังและไปไม่ถึง
Detector การสแกน m/z สำหรับการวิเคราะห์แต่ละครั้งทำได้โดยปรับค่าสนามไฟฟ้าไปเรื่อย ๆ โดย
ให้อยู่ในช่วงที่ครอบคลุมการวิ่งผ่านของค่า m/z ที่ต้องการบันทึก

2. Time of Flight (TOF) mass analyzer

TOF เป็นเครื่องวิเคราะห์มวลที่ทำการแยกแยะไอออนโดยอาศัยความแตกต่างของความเร็ว
ในการเคลื่อนที่ของไอออนที่มีมวลต่างกัน เมื่อถูกเร่งด้วยสนามไฟฟ้าดังรูป



รูปที่ 13 TOF Mass spectrometer

แสดงส่วนวิเคราะห์มวลของเครื่องแมสสเปกโตรเมตรีแบบ Time of Flight (TOF)
โดยจะเห็นว่าไอออนที่มีมวลเบากว่า จะเคลื่อนที่ไปยัง Detector ได้ไวกว่า (Adam-Day, 2018)

ในการศึกษานี้จะใช้เทคนิคแมสสเปกโตรเมตรีในการศึกษารูปแบบโปรตีนหรือการศึกษาโปรตีนทั้งหมดในเซลล์สิ่งมีชีวิต เพื่อให้เห็นภาพรวมกลไกการทำงานของเซลล์ในระดับโปรตีน ซึ่งการศึกษารูปแบบโปรตีนประกอบด้วยขั้นตอนย่อย ๆ ได้แก่ การสกัดโปรตีน การวัดปริมาณโปรตีน การแยกโปรตีนด้วยกระแสไฟฟ้า 2 มิติ (Two Dimensional Gel Electrophoresis) วิเคราะห์การแสดงออกที่แตกต่างกันของโปรตีนบนแผ่นเจล ระบุโปรตีนโดยใช้เทคนิค LC-MS/MS โดยในขั้นตอนการระบุโปรตีนด้วยเทคนิค LC-MS/MS ข้อมูล Chromatogram ที่ได้จากเครื่องจะถูกแปลงเป็นลำดับกรดอะมิโน ซึ่งจะถูกลำดับไปเปรียบเทียบกับข้อมูลโปรตีนในระบบฐานข้อมูลโดยใช้โปรแกรมทางชีวสารสนเทศศาสตร์ เพื่อระบุชนิดของโปรตีนดังกล่าว



บทที่ 3

วิธีดำเนินการวิจัย

ขั้นตอนและวิธีการทำงานวิจัย

สำหรับขั้นตอนและวิธีการทำงานวิจัย ผู้วิจัยได้แบ่งออกเป็น 4 ส่วนหลัก ๆ ดังต่อไปนี้

ตอนที่ 1 การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของคนไข้ออทิซึมสเปกตรัม โดยอาศัยการเรียนรู้ของเครื่องมือ (Machine learning) ด้วยโปรแกรม RapidMiner

ตอนที่ 2 การวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs)

ตอนที่ 3 การวิเคราะห์รูปแบบโปรตีโอมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs)

ตอนที่ 4 การศึกษาความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคออทิซึมสเปกตรัม และการยืนยันผล

ตอนที่ 1 การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของคนไข้ออทิซึมสเปกตรัม โดยอาศัยการเรียนรู้ของเครื่องมือ (Machine learning) ด้วยโปรแกรม RapidMiner

วัตถุประสงค์: เพื่อทำการเปรียบเทียบความสามารถหรือประสิทธิภาพในการแบ่งกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) และแบบมีผู้สอน (Supervised machine learning) ของเครื่องมือจากโปรแกรม RapidMiner และสร้างโมเดลในการแบ่งกลุ่มผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อย

ตอนที่ 1.1 การรวบรวมข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัมจาก

ฐานข้อมูล

วัตถุประสงค์: เพื่อกำหนดเกณฑ์ในการคัดเลือกข้อมูลลักษณะอาการทางคลินิกจากผู้ป่วยที่นำมาใช้ศึกษา

วิธีการวิจัย

รวบรวมข้อมูลของแบบทดสอบ Autism Diagnostic Interview-Revised (ADI-R) เพื่อการวินิจฉัยโรคออทิซึมสเปกตรัม จากฐานข้อมูล Autism Genetic Resource Exchange (AGRE) repository ซึ่งเป็นแหล่งเก็บรวบรวมข้อมูลทั้ง Genotype และ Phenotype ของผู้ป่วยออทิซึม

สำหรับการศึกษาวิจัย ซึ่งข้อมูลที่ใช้ในงานวิจัยนี้ ประกอบด้วยข้อมูลแบบทดสอบ ADI-R จำนวน 123 ข้อ ของผู้ป่วยออทิซึมสเปกตรัม จำนวน 85 ราย ที่ผ่านวิธีคัดเลือกตามเกณฑ์ของ Dr. Valerie Hu และคณะ ซึ่งทำการคัดเลือกผู้ป่วยชายที่เป็นโรคออทิซึมแบบโปรแบนด์ และคัดแยกผู้ป่วยที่มี Cognitive impairment (Raven's scores <70) ผู้ป่วยที่มีความผิดปกติทางพันธุกรรมชัดเจน (Fragile X, Rett syndrome, Tuberous sclerosis, Chromosome 15q11-q13 duplication) ผู้ป่วยที่คลอดก่อนกำหนด (<35 สัปดาห์) และผู้ป่วยที่ได้รับการวินิจฉัยด้วยโรคจิตประสาทที่เป็น Comorbid Disorders ของออทิซึมสเปกตรัม ได้แก่ Bipolar disorder, Obsessive compulsive disorder, and Severe anxiety ออกไป เหลือเพียงผู้ป่วยที่เป็นโรคออทิซึมสเปกตรัมแบบไม่ทราบสาเหตุ (Idiopathic) รวมทั้งสิ้น 1,954 ราย และได้ทำการสุ่มเลือกผู้ป่วยที่มีอาการทุกรูปแบบออกมา จำนวน 85 ราย

ทั้งนี้ผู้วิจัยได้ใช้ข้อมูลในการแบ่งกลุ่มย่อยของ Dr. Hu และ Dr. Steinberg จากการศึกษาในปี 2009 (Hu & Steinberg, 2009) มาเปรียบเทียบผลการแบ่งผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อย โดยการศึกษาของ Dr. Valerie Hu นั้น ได้แบ่งผู้ป่วยทั้ง 85 รายนี้ จากการใช้ข้อมูลลักษณะอาการทางคลินิกจากแบบทดสอบ ADI-R 123 ข้อ ออกเป็น 4 กลุ่มย่อย ได้แก่ กลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup) กลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup) กลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroups) และกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษร่วมกับความผิดปกติด้านภาษารุนแรง (Savant skills with severe language impairments subgroup)

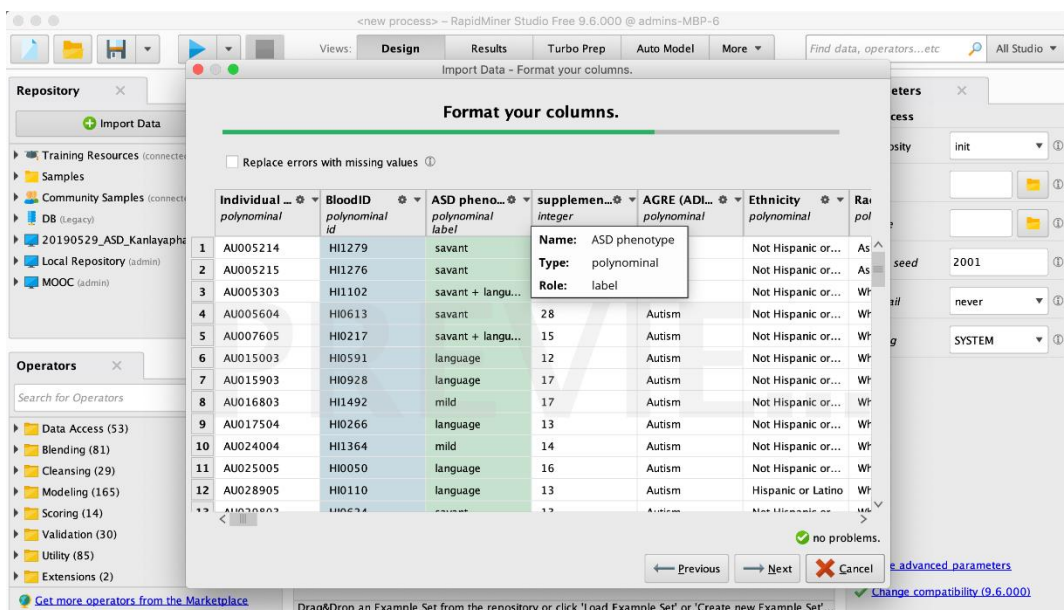
ตอนที่ 1.2 การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม RapidMiner

วัตถุประสงค์: เพื่อเปรียบเทียบความสามารถหรือประสิทธิภาพในการแบ่งกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) ของเครื่องมือจากโปรแกรม RapidMiner

วิธีการวิจัย

เมื่อรวบรวมข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมมาเรียบร้อยแล้ว จากนั้นนำข้อมูลเข้าสู่โปรแกรม RapidMiner โดยมีขั้นตอนดังนี้

1. เปิดโปรแกรม RapidMiner Studio version 9.6.000 จากนั้นคลิกปุ่ม Import Data ในหน้าต่าง Repository เลือก My Computer เพื่อค้นหาไฟล์ตารางของข้อมูลที่ต้องการจะนำมาวิเคราะห์
2. ทำการตรวจสอบตาราง และตั้งค่าชนิดของตาราง โดยเลือก Change Role ของ Attribute (Column) ที่มีชื่อว่า Blood ID เป็น ID และ Attribute ที่มีชื่อว่า ASD Phenotype เป็น Label จากนั้นกด Next



รูปที่ 14 Column formatting in RapidMiner

แสดงหน้าต่างการตั้งค่าชนิดตารางในโปรแกรม RapidMiner

โดยสามารถ Change role ของ Attribute ที่เราต้องการติดฉลากเป็นชนิด Label

3. ตั้งชื่อตารางสำหรับการ Save ลงโปรแกรม เมื่อตั้งชื่อเสร็จกด Finish ก็จะได้ข้อมูลของตาราง ซึ่งจะอยู่ในส่วน Repository ทางด้านซ้ายมือของโปรแกรม
4. ลากไฟล์ตารางจาก Repository เข้ามาในหน้าต่าง Process จากนั้นทำการเลือก Attribute ที่เป็นแบบทดสอบ ADI-R โดยการนำ Operator ที่มีชื่อว่า Select attribute มาต่อกับข้อมูลที่ต้องการวิเคราะห์ โดยทำการปรับ Parameters ดังนี้ เลือก Attribute filter type เป็น Subset คลิกปุ่ม Select attribute เลือกอันที่ไม่ใช่ ADI-R Items มาอยู่ทางฝั่ง Select เมื่อเสร็จแล้วคลิก Apply แต่จะต้องมีการคลิก Invert selection เพื่อเลือกเฉพาะข้อมูล ADI-R Items มาวิเคราะห์เท่านั้น
5. ลาก Operator ที่มีชื่อว่า Replace missing values ต่อจาก Select attribute แล้วตั้งค่า Parameters ให้แทนค่าที่หายไปด้วย Average

6. จากนั้นเลือก Tool ของ Unsupervised machine learning ซึ่งจะอยู่ในหมวด Segmentation เป็นการวิเคราะห์แบบ Clustering ทั้งหมด
- 6.1 Tools ชนิด Centroid cluster model ได้แก่ K-means K-medoids และ x-means จะสามารถวัดผลได้ทั้ง Cluster distance performance และ Cluster density performance
- 6.2 Tools ชนิด Cluster model ได้แก่ K-means (Kernel) DBSCAN Expectation maximization clustering Support Vector clustering และ Random clustering จะสามารถวัดผลได้เฉพาะ Cluster density performance
- 6.3 Tools ชนิด Hierarchical clustering model ได้แก่ Agglomerative clustering Top Down clustering และ Flatten clustering ไม่สามารถวัดผลได้ เนื่องจากผลลัพธ์ไม่ได้อยู่ในรูปแบบ Cluster แต่จะอยู่ในรูปแบบตารางแทน
- ตั้งค่า Parameters โดยเลือก Add cluster attribute ตั้งค่า k หรือจำนวนกลุ่มตามต้องการ จากนั้นตั้งค่าชนิดของข้อมูลที่วัดเป็น Numerical Measures แบบ Euclidean Distance วิเคราะห์จำนวน 100 ครั้ง และใช้ Local random seed 1992 ในทุก Tools
7. วัดผลการแบ่งกลุ่มโดยใช้ Operator ที่ชื่อว่า Cluster distance performance โดยลากข้อมูลตารางและผลลัพธ์การจัดกลุ่มเข้า Input Port และลาก Output port ไปสู่การแสดงผลการประเมินทางด้านขวามือ
8. สำหรับการวัดผล Cluster Density Performance จำเป็นจะต้องใช้ Operator ที่มีชื่อว่า Data to similarity มาวิเคราะห์ จากนั้นลากเข้า Input port ของ Cluster density performance เช่นเดียวกับข้อมูลตารางและผลลัพธ์การจัดกลุ่ม และลาก Output port ไปสู่การแสดงผลการประเมินทางด้านขวามือ
9. ในการเปรียบเทียบหาค่า k ที่เหมาะสมสำหรับแต่ละอัลกอริทึม จะต้องใช้ Operator ที่มีชื่อว่า Item distribution performance โดย Input port จะต้องนำผลลัพธ์การจัด Cluster และ Performance ของ Cluster distance หรือ Cluster density performance มาต่อ และลาก Output Port ไปสู่การแสดงผลการประเมินทางด้านขวามือเช่นกัน
10. ลาก Output Port ของผลลัพธ์การจัดกลุ่ม จาก Clustering tool ที่ใช้ด้วย แล้วกดวิเคราะห์ โปรแกรมจะแสดงผลลัพธ์การจัดกลุ่ม และผลการประเมินให้ในมุมมอง Results

11. นำผลลัพธ์ที่ได้มาทำการเปรียบเทียบค่า k หรือจำนวนกลุ่มที่เหมาะสมด้วยวิธี Elbow method โดยค่าที่ได้จาก Operator Item distribution performance จะเป็นค่าที่เรียกว่า Average within cluster sum of squares มาสร้างกราฟ และเปรียบเทียบประสิทธิภาพการแบ่งกลุ่มระหว่างวิธี K-means และ K-medoids ที่จำนวนกลุ่มเท่ากันจากค่าที่ได้จาก Operator Cluster distance performance และ Cluster density performance โดยจะมีค่า Average within centroid distance บอกค่าเฉลี่ยของแต่ละตัวอย่างจากจุดกึ่งกลางของกลุ่ม Average within cluster distance บอกค่าเฉลี่ยการกระจายตัวภายในกลุ่มหรือการเกาะกลุ่ม และค่า Davies Bouldin บอกการกระจายตัวภายในกลุ่ม และระยะห่างระหว่างกลุ่ม มาใช้ในการเปรียบเทียบประสิทธิภาพ

ตอนที่ 1.3 การวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยอหิวาต์ซิมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner

วัตถุประสงค์: เพื่อสร้างโมเดลและเปรียบเทียบความสามารถหรือประสิทธิภาพในการแบ่งกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ โดยอาศัยการเรียนรู้แบบมีผู้สอน (Supervised machine learning) ของเครื่องมือจากโปรแกรม RapidMiner

วิธีการวิจัย

หลังจากที่ได้ Unsupervised machine learning tool ที่มีประสิทธิภาพมากที่สุดและสามารถแบ่งผู้ป่วยอหิวาต์ซิมสเปกตรัมออกเป็นกลุ่มย่อยจากอาการทางคลินิกได้แล้ว ในขั้นตอนนี้ ผู้วิจัยจะสร้างโมเดลที่ใช้ในการแบ่งกลุ่มคนไข้ตามที่ได้จัดไว้จากวิธีข้างต้น

โดยผู้วิจัยจะใช้ Supervised machine learning tools ซึ่งในงานวิจัยนี้จะทำการเปรียบเทียบ 7 อัลกอริทึม ได้แก่ Artificial Neural Network (ANN) k-Nearest Neighbor (k-NN) Naive Bayes Decision tree Random forest Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) โดยในแต่ละอัลกอริทึมจะใช้กระบวนการในการวิเคราะห์ข้อมูล 3 แบบ ได้แก่ Split validation (กระบวนการนี้จะมีการแบ่งข้อมูลเพื่อสอนคอมพิวเตอร์ 70% และทดสอบโมเดล 30%) Cross validation/Backward elimination (กระบวนการนี้คอมพิวเตอร์จะลองตัดตัวแปรออกไปก่อนจะทำการวิเคราะห์ หากตัวแปรใดถูกนำออก

แล้วลดความถูกต้องของโมเดลลง จะยังคงตัวแปรนั้นไว้ แต่ถ้าตัดออกแล้วไม่ทำให้ค่าความถูกต้องเปลี่ยนไป หรือเพิ่มค่าความถูกต้องมากขึ้น ตัวแปรนั้นก็จะถูกนำออกจากโมเดล) และ Cross validation/Forward selection (กระบวนการนี้คอมพิวเตอร์จะลองเพิ่มตัวแปรที่เข้ามาใช้วิเคราะห์ทีละตัว หากตัวแปรที่เพิ่มเข้ามาทำให้ค่าความถูกต้องของโมเดลมีค่าเพิ่มมากขึ้นก็จะเก็บตัวแปรนั้นไว้ แต่ถ้าหากเพิ่มเข้ามาแล้วไม่ทำให้ค่าความถูกต้องเปลี่ยนไป หรือลดค่าความถูกต้องลง ตัวแปรนั้นจะถูกตัดออกจากโมเดล) ตามวิธีการดังนี้

1. จากการ Clustering ด้วย Unsupervised machine learning tool ที่เลือก เราจะสร้าง Attribute ใหม่ที่มีการจัดผู้ป่วยออกที่ซิมสเปกตรัมออกเป็นกลุ่มย่อย โดยตั้งชื่อว่า Cluster และ Change role ของ Attribute นี้เป็น Label เพื่อให้ Supervised machine learning algorithms ใช้ในการวิเคราะห์ข้อมูลเพื่อสร้างโมเดล

2. หลังจาก Save ตารางใหม่นี้แล้ว ลากไฟล์ตารางจาก Repository เข้ามาในหน้าต่าง Process จากนั้นทำการเลือก Attribute ที่เป็นแบบทดสอบ ADI-R โดยการนำ Operator ที่มีชื่อว่า Select Attribute มาต่อกับข้อมูลที่ต้องการวิเคราะห์และปรับค่า Parameters เช่นเดียวกับขั้น Unsupervised machine learning

3. ลาก Operator ที่มีชื่อว่า Replace missing values มาต่อจาก Select attribute แล้วตั้งค่า Parameters ให้แทนค่าที่หายไปด้วย Average

4. ลาก Operator ที่มีชื่อว่า Split validation มาต่อจาก Replace missing values เพื่อทำการแบ่งข้อมูลเข้าไปใช้ในการสร้างโมเดลเพื่อเป็นการสอนเครื่องมือ (Training dataset) 70% และไว้สำหรับทดสอบโมเดล (Testing dataset) 30% เมื่อตั้งค่า Parameters เสร็จแล้ว ให้ Double click เข้าไปในกล่อง Operator Split validation เพื่อใส่อัลกอริทึมที่จะใช้ในการวิเคราะห์ ในกรณีที่ใช้ Operator Backward elimination และ Forward selection ให้ Double click เข้าไปใน Operator และทำการใส่ Operator Cross validation เข้าไป เพื่อให้สามารถใส่อัลกอริทึมได้

5. จากนั้นเลือก Tool ของ Supervised machine learning ซึ่งจะอยู่ในหมวด Modeling > Predictive โดยมีการสร้างโมเดลอยู่หลายรูปแบบมาก ได้แก่ Artificial Neural Network (ANN), k-Nearest Neighbor (k-NN), Naive Bayes, Decision tree, Random forest, Support Vector Machine (SVM), A Library of Support Vector Machine (LibSVM) มาใส่ทางฝั่ง Training

6. ปรับค่า Parameter ของแต่ละ Model ตามความเหมาะสม เช่น เลือก Use local random seed เพื่อที่จะได้ทำให้ทุก Algorithms ที่มีการสุ่มข้อมูลมาใช้สร้างโมเดลนั้น สุ่มข้อมูลชุดเดียวกันในทุก ๆ ครั้ง เป็นต้น

7. ทางฝั่ง Testing ให้เลือก Operator ที่มีชื่อว่า Apply model และ Performance (Classification) มาต่อกัน โดยตรง Parameters ของ Performance (Classification) จะสามารถเลือกดูข้อมูลการประเมินผลได้หลากหลาย เช่น Accuracy, Classification error, Kappa, Weighted mean recall, Weighted mean precision, Spearman Rho, Kendall Tua, Absolute error, Relative error, Relative error lenient, Relative error strict, Normalized absolute error, Root mean squared error, Root relative error, Squared error, Correlation, Squared correlation, Cross-entropy, Margin, Soft margin loss, Logistic loss, Skip undefined labels, Use sample weights

8. กลับไปที่หน้า Process แล้วลากเส้นจาก Output port ของ Operator Split validation ไปยัง result ทางด้านขวามือ เพื่อให้แสดงผล

9. ในการตรวจสอบการสร้างโมเดลแบบใช้ Testing dataset 100% ให้นำเข้าตาราง ADI-R scores ที่ไม่มีข้อมูลใน Attribute cluster โดยปล่อยเป็นช่องว่าง แล้ว Change role เป็น Prediction ลากเข้าหน้าต่างต่าง Process เดียวกัน แล้วเลือก Operator Select attribute กับ Replace missing values มาต่อเช่นเดิม

10. ลาก Operator Multiply ต่อกับ Output port ที่เป็น Model ของ Operator Split validation ในข้อ 8 โดยจะแบ่ง Model อันนี้ไปต่อกับ Result และอีกอันหนึ่งมาใช้เป็นโมเดลเพื่อทำนายการจัดกลุ่ม

11. ลาก Operator Apply model และ Performance (Classification) มาต่อกัน โดยทาง Input port จะนำข้อมูลตารางจากข้อ 9 และ Model จากข้อ 10 มาต่อ ลาก Output Port ไปทางขวามือที่เป็น Result

12. กวดิเคราะห์ โปรแกรมจะแสดงผลลัพธ์การจัดกลุ่ม และผลการประเมินให้ในมุมมอง Results การเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึมของการเรียนรู้ของเครื่องมือ

นอกจากนี้ในการประเมินประสิทธิภาพในการทำนายของโมเดลจากผลการวิเคราะห์ของ

โปรแกรม RapidMiner แล้ว เรายังสามารถคำนวณความไว (Sensitivity) ความจำเพาะ (Specificity) และ F-score ได้จากการใช้เครื่องมือ Confusion matrix โดยแต่ละค่าสามารถคำนวณได้ดังต่อไปนี้

| | | True Class | | |
|------------|---------|------------|---------|---------|
| | | Class 2 | Class 1 | Class 0 |
| Prediction | Class 2 | | | |
| | Class 1 | | | |
| | Class 0 | | | |

① Class 2's Precision

| | | True Class | | |
|------------|---------|--------------------|---------------------|---------------------|
| | | Class 2 | Class 1 | Class 0 |
| Prediction | Class 2 | True Positive (TP) | False Positive (FP) | False Positive (FP) |

$$Precision = \frac{TP}{TP + FP}$$

② Class 2's Recall

| | | True Class |
|------------|---------|---------------------|
| | | Class 2 |
| Prediction | Class 2 | True Positive (TP) |
| | Class 1 | False Negative (FN) |
| | Class 0 | False Negative (FN) |

$$Recall = \frac{TP}{TP + FN}$$

③ Class 2's F-Value

$$F - Value = \frac{2 * Recall * Precision}{Recall + Precision}$$

รูปที่ 15 Confusion Matrix

แสดงการคำนวณค่า Precision Recall และ F-score จากผลลัพธ์การทำนายที่ได้จากโปรแกรม RapidMiner

- Accuracy** ค่าความถูกต้องในการทำนายของโมเดล เป็นประสิทธิภาพโดยรวมของทั้งโมเดล ซึ่งสามารถคำนวณได้จาก

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

- Precision** หรืออีกชื่อหนึ่งคือ Positive Predictive Value (PPV) เป็นค่าความแม่นยำที่สนใจเฉพาะส่วนที่ทำนายถูกต้อง

$$Precision = \frac{TP}{TP + FN}$$

- Sensitivity** หรือ Recall มีอีกชื่อหนึ่งคือ True Positive Rate (TTR) บอกถึงความสามารถการทำนายค่าที่ Positive ของโมเดลได้ถูกต้องมากแค่ไหน บอกความไวในการทำนายของโมเดล

$$Sensitivity = \frac{TP}{TP + FN}$$

4. **Specificity** มีอีกชื่อหนึ่งคือ True Negative Rate (TNR) บอถึงความสามารถในการทำนายค่าที่ Negative ของโมเดลได้ถูกต้องมากแค่ไหน เป็นการบอกความจำเพาะของโมเดล

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}$$

5. **F-score** เป็นค่าเฉลี่ยของ Precision และ Recall

$$\text{F-score} = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

เมื่อได้ผลลัพธ์ของประสิทธิภาพของแต่ละโมเดลออกมาแล้ว ผู้วิจัยจะทำการเปรียบเทียบประสิทธิภาพ เพื่อหา Supervised machine learning ที่ดีที่สุด โดยมีเกณฑ์ดังนี้

1. ในแต่ละอัลกอริทึมจะมีกระบวนการในการวิเคราะห์ที่แตกต่างกันออกไปทั้งหมด 3 แบบ คือ Split validation, Backward elimination, and Forward selection รวมถึงการปรับพารามิเตอร์ที่แตกต่างกันออกไป ในแต่ละกระบวนการวิเคราะห์จะมีการเลือกการปรับค่าพารามิเตอร์ที่ดีที่สุดออกมา (Accuracy สูงที่สุด) และจะถูกนำมาเปรียบเทียบกับระหว่างกระบวนการวิเคราะห์ของแต่ละอัลกอริทึมต่อไป
2. จากนั้นในแต่ละอัลกอริทึม จะพิจารณากระบวนการวิเคราะห์ที่ให้ค่า Accuracy ที่มากที่สุด ($\geq 98\%$) ที่สามารถลดจำนวนข้อลง โดยที่ใช้เวลาในการวิเคราะห์ไม่เกิน 1/2 ชม.
3. นำโมเดลที่ดีที่สุดของแต่ละ Algorithm มาเปรียบเทียบกับกัน เพื่อหาโมเดลที่ให้ค่า Accuracy สูงที่สุดและสามารถลดจำนวนข้อของแบบสอบถามได้มากที่สุด

ตอนที่ 2 การวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์

(Lymphoblastoid Cell Lines; LCLs)

วัตถุประสงค์: เพื่อศึกษารูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมเปรียบเทียบกับรูปแบบการแสดงออกในเซลล์เพาะเลี้ยงที่มาจากคนปกติ ด้วยเทคนิคไมโครอาร์เรย์ และระบุยีนที่มีระดับการแสดงออกผิดปกติทั้งในผู้ป่วยออทิซึมสเปกตรัมและในกลุ่มย่อยที่แบ่งตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์

ตอนที่ 2.1 การรวบรวมข้อมูลทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จาก

ฐานข้อมูล

วัตถุประสงค์: เพื่อกำหนดเกณฑ์ในการคัดเลือกข้อมูลทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากการศึกษาด้วยเทคนิคไมโครอาร์เรย์

วิธีการวิจัย

ทำการคัดเลือกข้อมูลรูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยชุดเดียวกันกับการแบ่งผู้ป่วยออกเป็นกลุ่มย่อยตามลักษณะอาการทางคลินิกโดยใช้เทคโนโลยีปัญญาประดิษฐ์ และผู้ป่วยชุดเดียวกันกับที่จะใช้ในการศึกษารูปแบบโปรตีโอมจำนวน 85 เซลล์ จากการศึกษาของ Dr. Hu และคณะ ในปี 2009 (Hu et al., 2009) ซึ่งเป็นการศึกษารูปแบบการแสดงออกของยีนโดยใช้เทคนิค cDNA microarrays จำนวน 39,936 ยีน (TIGR 40K Human array, GPL3427) และมี Probe อยู่ทั้งหมด 42,000 Probe สำหรับการศึกษาการแสดงออกของ RNA ทั้งหมด 42,000 ทรานสคริป

โดยเราจะทำการดึงข้อมูลจากฐานข้อมูล GSE15402 จาก NCBI ที่อยู่ในรูปแบบของข้อมูลตารางที่มีการเก็บรูปแบบการแสดงออกของยีนที่ได้ทำการ Normalize แล้ว จากโปรแกรมในชุด TM4: Microarray Software Suite ซึ่งได้แก่ โปรแกรม Microarray Data Analysis System (MIDAS) และ SpotFinder ดังที่ได้มีการศึกษาของ Dr. Hu และคณะ ในปี 2009 จากจำนวนผู้ป่วยทั้งหมด 85 ราย และกลุ่มควบคุมซึ่งเป็นคนปกติจำนวน 29 ราย โดยผู้วิจัยจะคัดเลือกข้อมูลจากกลุ่มผู้ป่วยออทิสซึมสเปกตรัมกลุ่มย่อย C จำนวน 9 ราย กลุ่มย่อย D จำนวน 13 ราย และกลุ่มควบคุมจำนวน 7 ราย เพื่อให้สอดคล้องกับการศึกษารูปแบบโปรตีโอมของผู้ป่วยออทิสซึมสเปกตรัมจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในการศึกษาครั้งนี้ (Hu et al., 2009)

ตอนที่ 2.2 การวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ได้จาก

ฐานข้อมูล

วัตถุประสงค์: เพื่อทำการเปรียบเทียบรูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิสซึมสเปกตรัมและผู้ป่วยออทิสซึมสเปกตรัมกลุ่มย่อยกับคนปกติ ตลอดจนเกณฑ์ในการคัดเลือกรายชื่อยีนที่มีการแสดงออกผิดปกติไป

วิธีการวิจัย

เมื่อได้ข้อมูลที่ผ่านการคัดเลือกมาแล้ว ผู้วิจัยได้ทำการ Normalize ข้อมูลอีกครั้งโดยการใช้โปรแกรม Multi-Experimental Viewer (MeV) และใช้ 70% Data filter เพื่อทำการกำจัดข้อมูลการแสดงออกของยีนที่มีการสูญเสียค่าความเข้มมากกว่าร้อยละ 30 ของจำนวนเซลล์เพาะเลี้ยงทั้งหมด จากนั้นจึงนำข้อมูลการแสดงออกของยีนที่เหลือมาหา yin ที่แสดงออกแตกต่างกันอย่างมีนัยสำคัญด้วยวิธีการทางสถิติ T-test แบบ Adjusted Bonferroni correction ($P < 0.05$) โดยใช้ข้อมูลของกลุ่มผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับข้อมูลของกลุ่มควบคุม ซึ่งรายชื่อยีนที่พบว่าการแสดงออกผิดปกติในผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยจะถูกเก็บไว้สำหรับการศึกษาในขั้นตอนหลังจากนี้ต่อไป

ตอนที่ 3 การวิเคราะห์รูปแบบโปรตีนของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid cell lines; LCLs)

วัตถุประสงค์: เพื่อศึกษารูปแบบการแสดงออกของโปรตีนในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมเปรียบเทียบกับรูปแบบการแสดงออกในเซลล์เพาะเลี้ยงที่มาจากคนปกติ และระบุโปรตีนที่มีระดับการแสดงออกผิดปกติในกลุ่มย่อยของโรคออทิซึมสเปกตรัมที่แบ่งตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์

ตอนที่ 3.1 การคัดเลือกสิ่งตัวอย่าง

วัตถุประสงค์: เพื่อกำหนดเกณฑ์ในการคัดเลือกผู้ป่วยที่เหมาะสมในแต่ละกลุ่มตัวอย่าง

วิธีการวิจัย

ทำการศึกษาในเซลล์เพาะเลี้ยงชนิด Lymphoblastoid Cell Lines (LCLs) ที่พัฒนามาจาก Mononuclear Cells ในเลือดของผู้ป่วยโรคออทิซึมสเปกตรัมและคนปกติ โดย Autism Genetic Resource Exchange (AGRE) repository ซึ่งคณะผู้วิจัยได้รับความอนุเคราะห์เซลล์เพาะเลี้ยงเหล่านี้จาก Dr. Valerie Hu มหาวิทยาลัยจอร์จวอชิงตัน กรุงวอชิงตัน ดีซี ประเทศสหรัฐอเมริกา เซลล์เหล่านี้ได้ผ่านการคัดเลือกตามหลักเกณฑ์ของ Dr. Valerie Hu และคณะ ตลอดจนใช้ในการวิเคราะห์การแสดงออกของทรานสคริปโตมิกส์โดยใช้ cDNA microarray analysis ที่ได้ตีพิมพ์ไว้แล้ว (GSE15402) กล่าวโดยย่อคือ ทำการคัดเลือกผู้ป่วยชายที่เป็นโรคออทิซึมแบบโปรแบนด์ (Autistic probands) โดยใช้ข้อมูลจากพฤติกรรมของผู้ป่วย 123 ประการในชุดแบบสอบถามเพื่อการ

วินิจฉัยโรคออทิซึมสเปกตรัมที่เรียกว่า Autism Diagnostic Interview-Revised (ADI-R) คัดแยกผู้ป่วยที่มี Cognitive impairment (Raven's scores <70) ผู้ป่วยที่มีความผิดปกติทางพันธุกรรมชัดเจน (Fragile X, Rett syndrome, Tuberous sclerosis, Chromosome 15 q11-q13 duplication) ผู้ป่วยที่คลอดก่อนกำหนด (<35 สัปดาห์) และผู้ป่วยที่ได้รับการวินิจฉัยด้วยโรคจิตประสาทที่เป็น Comorbid disorders ของออทิซึมสเปกตรัม ได้แก่ Bipolar disorder, Obsessive Compulsive disorder, Severe anxiety ออกไป เหลือเพียงผู้ป่วยที่เป็นโรคออทิซึมสเปกตรัมแบบ Idiopathic รวมทั้งสิ้น 1,954 ราย และได้ทำการสุ่มเลือกผู้ป่วยที่มีอาการทุกรูปแบบออกมาจำนวน 85 ราย จากนั้นทำการจัดกลุ่มเป็น 4 กลุ่มย่อย (Subgroups) ตามลักษณะอาการทางคลินิก (Clinical phenotypes) ด้วยวิธีปัญญาประดิษฐ์

โดยสำหรับการศึกษานี้ จะทำการศึกษารูปแบบโปรตีโอมในผู้ป่วยออทิซึมสเปกตรัมเปรียบเทียบกับคนปกติที่มีอายุเท่ากับผู้ป่วย (Age-matched control) จะทำการแบ่งผู้ป่วยออทิซึมสเปกตรัมออกเป็น 4 กลุ่มย่อย โดยใช้ลักษณะอาการทางคลินิกและวิธีปัญญาประดิษฐ์ในการแบ่งกลุ่มดังที่ได้กล่าวไปในขั้นตอนที่ 1 และทำการคัดเลือกกลุ่มของผู้ป่วยกลุ่มย่อย C จำนวน 9 ราย กลุ่มย่อย D จำนวน 13 ราย และกลุ่มควบคุมจำนวน 7 ราย รวมทั้งสิ้น 29 ราย

ตอนที่ 3.2 การเพาะเลี้ยงและการเก็บเซลล์

วัตถุประสงค์: เพื่อให้ได้เซลล์สำหรับการสกัดโปรตีนไว้ใช้ในการศึกษารูปแบบโปรตีโอมจากผู้ป่วยออทิซึมสเปกตรัมและกลุ่มควบคุม

วิธีการวิจัย

ขั้นตอนการเพาะเลี้ยงเซลล์

เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid cell lines; LCLs) จะถูกเก็บไว้ในไนโตรเจนเหลวจนกว่าจะนำมาเพาะเลี้ยงตามระเบียบวิธีการของ Rutgers University Cell and DNA repository ผู้ผลิตและดูแลรักษาเซลล์ใน Autism Genetic Research Exchange (AGRE) collection กล่าวโดยย่อคือ เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จะถูกนำมาเพาะเลี้ยงในอาหารเลี้ยงเซลล์ RPMI-1640 medium (GE Healthcare, Chicago, IL, USA) ที่มีส่วนผสมของ Fetal Bovine Serum 15% (Sigma-Aldrich, St. Louis, MO, USA) และ Penicillin/Streptomycin 1% (Mediatech, Manassas, VA, USA) หลังจาก

ดำเนินการเพาะเลี้ยง เซลล์ที่ได้จะถูกเก็บรักษาไว้ใน RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) ที่อุณหภูมิ -20 องศาเซลเซียส ซึ่งจะช่วยคงสภาพสารพันธุกรรมจนกว่าจะทำการวิเคราะห์ในกระบวนการต่อ ๆ ไป

1. ขั้นตอนการเพาะเลี้ยงเซลล์จะต้องทำใน Biological Safety Cabinet class II โดยเริ่มจากเตรียมอาหารเลี้ยงเซลล์ที่ประกอบด้วย RPMI-1640 Fetal Bovine Serum (FBS) และ Penicillin/Streptomycin ลงใน Culture flask ขนาด 25 cm² (T-25) แล้วนำไป Incubate ในตู้บเลี้ยงเซลล์เพื่อปรับอุณหภูมิ

2. นำเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ถูกเก็บไว้ในไนโตรเจนเหลว มาละลายอย่างรวดเร็วที่อุณหภูมิ 37 องศาเซลเซียสในอ่างควบคุมอุณหภูมิ ปิเปตเซลล์ใส่ลงใน Culture flask ที่เตรียมไว้ในข้อแรกอย่างช้า ๆ เขย่าเบา ๆ ให้เซลล์กับอาหารเลี้ยงเซลล์ผสมให้เข้ากัน เก็บไว้ในตู้เลี้ยงเซลล์ที่อุณหภูมิ 37 องศาเซลเซียส ที่มีปริมาณก๊าซคาร์บอนไดออกไซด์ร้อยละ 5 เป็นเวลา 24 ชั่วโมง

3. เมื่อครบ 24 ชั่วโมงต้องทำการเปลี่ยน Culture flask เป็นขนาด 75 cm² โดยปิเปตเซลล์ 8 ml ลงใน Culture flask แล้วเติมอาหารเลี้ยงเซลล์อีก 12 ml แล้วนำเข้าตู้เลี้ยงเซลล์ต่อ

4. ซึ่งเมื่ออาหารเลี้ยงเซลล์เปลี่ยนเป็นสีเหลือง จะต้องทำการเปลี่ยนอาหารเลี้ยงเซลล์ โดยตั้ง Culture flask ดังกล่าวไว้เพื่อให้เซลล์ตกตะกอน จากนั้นปิเปตอาหารเลี้ยงเซลล์ทิ้งโดยปิเปตออกมาให้มากที่สุด และใส่อาหารเลี้ยงเซลล์ใหม่ลงไปมากกว่าปริมาตรที่ปิเปตออกมา 1 ml

ทั้งนี้ ขั้นตอนทั้งหมดต้องใช้เวลาในการเพาะเลี้ยงเซลล์เป็นเวลา 3 สัปดาห์ หรือประมาณ 1 เดือน เพื่อให้ได้ปริมาณเซลล์มากกว่า 4×10^8 cell/cm³ ซึ่งเป็นจำนวนเซลล์ที่ได้เพียงพอต่อการวิจัยจึงค่อยทำการเก็บเซลล์

ขั้นตอนการเก็บเซลล์

เมื่อได้จำนวนเซลล์ที่เพียงพอแล้ว ให้เก็บเซลล์ใน RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) ที่อุณหภูมิ - 20 องศาเซลเซียส จนกว่าจะนำมาสกัดโปรตีนต่อไป ส่วนการเก็บเซลล์ไว้ในไนโตรเจนเหลวสำหรับการทดลองอื่น ๆ จะใช้ Freezing Medium ที่ประกอบด้วย Fetal Bovine Serum และ Dimethyl sulfoxide (DMSO) ใน

อัตราส่วน 90 ml ต่อ 10 ml ตามลำดับ ทั้งนี้ ขั้นตอนการเก็บเซลล์ทุกขั้นตอนจะต้องรักษาอุณหภูมิของเซลล์ไว้ที่ 4 องศาเซลเซียส กล่าวโดยย่อคือ

1. ปิเปตต์เซลล์จาก Culture flask ใส่ลงใน Microcentrifuge tube ปริมาณ 1 ml
2. นำเซลล์ไปปั่นแยก ดูดอาหารเลี้ยงเซลล์ออกให้หมดให้คงเหลือไว้แต่เซลล์ที่ตกตะกอนใส่ RNeasy Lysis Buffer จากนั้นปิเปตขึ้นลงให้เป็นเนื้อเดียวกัน แล้วเก็บที่ -80 องศาเซลเซียส จนกว่าจะนำมาสกัด DNA RNA และโปรตีน

ตอนที่ 3.3 การสกัดโปรตีนจากเซลล์ วัดปริมาณโปรตีน และตรวจสอบคุณภาพโปรตีน

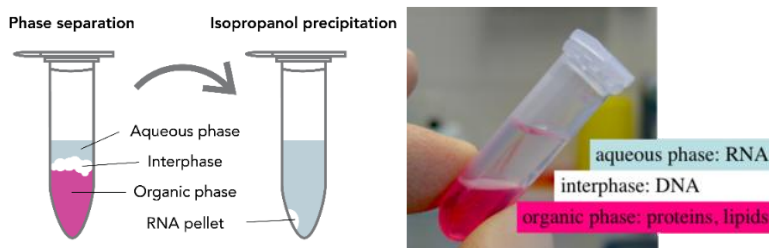
วัตถุประสงค์: เพื่อสกัดโปรตีน ทำให้ปริมาณโปรตีนบริสุทธิ์ขึ้น ลดปริมาณเกลือส่วนเกินของโปรตีน ทำการวัดปริมาณโปรตีนที่คงเหลือ และตรวจสอบคุณภาพโปรตีนที่นำมาทำการศึกษา เพื่อให้ได้ปริมาณโปรตีนที่เหมาะสมในการศึกษาขั้นต่อไป

วิธีการวิจัย

ขั้นตอนการสกัดโปรตีน

โปรตีนจากเซลล์ที่เก็บไว้ใน RNeasy Lysis Buffer (Thermo Fisher Scientific, Waltham, MA, USA) จะถูกสกัดโดยใช้ RNeasy lysis reagent (Geneaid Biotech Ltd., New Taipei City, Taiwan) ด้วยเทคนิค Phenol-chloroform extraction ตามคำแนะนำจากบริษัทผู้ผลิตนี้มา กล่าวโดยย่อคือ

1. นำเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ใน RNeasy Lysis Buffer ที่ -80 องศาเซลเซียส มาละลายที่อุณหภูมิห้อง ทำการปั่นเหวี่ยงและปิเปตต์ RNeasy Lysis Buffer ออก จากนั้นเติม RNeasy lysis reagent ลงไป โดยให้มีปริมาตรเป็น 2 เท่าของปริมาณเซลล์ ทำการปิเปตต์ขึ้นลงเพื่อทำให้เซลล์แตก แล้วตั้งทิ้งไว้ที่อุณหภูมิห้อง 5 นาที
2. เติม Chloroform 0.2 ml ลงไป ปิดฝาหลอดให้แน่น จากนั้นเขย่าหลอดแรง ๆ ด้วยมือ เป็นเวลา 15 วินาที และตั้งไว้ที่อุณหภูมิห้อง 2-3 นาที แล้วนำไปปั่นแยกที่ 12,000 g อุณหภูมิ 4 องศาเซลเซียส 15 นาที เพื่อให้สิ่งตัวอย่างแยกชั้น ซึ่งจะแยกออกเป็น 3 ชั้น โดยชั้นล่างสุดหรือ Organic phase จะเป็นชั้นที่มีสีแดงของ Phenol-chloroform คือชั้นของโปรตีน ส่วนชั้นกลางหรือ Interphase จะเป็นชั้นของ DNA และชั้นบนสุดจะเป็นชั้นสารละลายไม่มีสีหรือ Aqueous phase ซึ่งเป็นชั้นของ RNA



รูปที่ 16 DNA RNA and protein isolation by GENEzol reagent

แสดงการแยกชั้นของ DNA RNA และโปรตีน หลังจากตกตะกอนด้วย Chloroform ซึ่งจะแยกออกเป็น 3 ชั้น โดยชั้นบนสุดซึ่งเป็นชั้น Aqueous phase จะเป็นชั้นของ RNA ส่วนชั้นกลาง Interphase เป็นชั้นของ DNA และชั้นล่างสุดซึ่งเป็น Organic phase จะเป็นชั้นของโปรตีน (Suckale, 2009)

3. ปิเปตชั้นสารละลายใสและชั้นกลางออกไปใส่หลอดใหม่ให้ได้มากที่สุด เพื่อนำไปใช้ในขั้นตอนการสกัด RNA และ DNA ตามลำดับ

4. นำสารละลายในชั้น Phenol-chloroform จากข้อ 2 มาทำการสกัดโปรตีนต่อ โดยเติม Isopropanol ลงไป 1.5 ml ตั้งไว้ที่อุณหภูมิห้องเป็นเวลา 10 นาที แล้วนำไปปั่นที่ 12,000 g อุณหภูมิ 4 องศาเซลเซียส เป็นเวลา 10 นาที แล้วดูดส่วนน้ำใสออก เพื่อให้เหลือแต่ตะกอนโปรตีน (Protein pellet)

5. เติม Wash solution ได้แก่ Ice-cold acetone (Acetone ใน -20 องศาเซลเซียส) ลงไปในปริมาณ 2 เท่า ของ GENEzol reagent ที่ใช้ในขั้นตอนที่ 1 นำไป Vortex เป็นเวลา 1 นาที จึงนำไปปั่นที่ 16,000 g อุณหภูมิ 4 องศาเซลเซียส 10 นาที

6. เท Wash solution ทั้ง ทำขั้นตอนที่ 5 ซ้ำอีก 2 รอบ รอบสุดท้ายให้ใช้ Pipette ดูด Wash solution ออกให้ได้มากที่สุด โดยไม่ให้โดนตะกอน จากนั้นตากตะกอนให้แห้ง

7. ละลายตะกอนกลับด้วย Lysis buffer ที่ประกอบด้วย 7M Urea, 2M Thiourea, 4% CHAPS, 100mM DTT และ 1% Protease inhibitor ทำการผสมให้เข้ากัน แล้วใช้ที่บดเซลล์ (Grinder) ช่วยบดให้โปรตีนผสมเป็นเนื้อเดียวกันกับ Lysis buffer

8. หากตะกอนยังไม่ละลาย สามารถนำไป Sonicate ที่ 4 องศาเซลเซียส เป็นเวลา 1 ชั่วโมง เพื่อให้ตะกอนละลายได้ดีขึ้น แต่ถ้ายังไม่ละลายให้เติม Lysis buffer เพิ่มเข้าไป และ Sonicate ต่ออีก 1 ชั่วโมง

9. เมื่อตะกอนโปรตีนละลายดีแล้ว ให้เก็บโปรตีนไว้ที่ -20 องศาเซลเซียส ได้จนกว่าจะทำการวัดปริมาณโปรตีนและตรวจสอบคุณภาพโปรตีนต่อไป

ขั้นตอนการวัดปริมาณโปรตีน

ทำการหาปริมาณโปรตีนด้วยวิธี Bradford's assay ซึ่งเป็นหลักการ Dye-binding principle ดังนี้

1. เตรียมสารละลายโปรตีนมาตรฐาน โดยใช้น้ำ Milli-Q และ 10% BSA ให้ได้ 5 ความเข้มข้น ได้แก่ 1, 0.5, 0.25, 0.125, 0.0625 $\mu\text{g}/\mu\text{l}$ ตามลำดับ
2. ปิเปตสารละลายโปรตีนมาตรฐานทั้ง 5 ความเข้มข้น Lysis buffer และโปรตีนที่สกัดได้จากขั้นตอนการสกัดโปรตีน (Dilute 10 เท่า โดยใช้น้ำ Milli-Q 31.5 μl และโปรตีน 3.5 μl) ปริมาณ 10 μl ลงใน 96 wells plate โดยในแต่ละตัวอย่างจะต้องมีการทำซ้ำเป็นจำนวนทั้งสิ้นสามหลุม (Triplicate)
3. เติม Bradford solution ลงไปใน 96 wells plate ที่มีการเติมสารละลายโปรตีนมาตรฐานทั้ง 5 ความเข้มข้น Blank และโปรตีนที่เราต้องการวัดปริมาณ จากนั้นใช้ฝ่ามือเคาะขอบทั้ง 4 ด้านของ 96 wells plate เบา ๆ เพื่อให้สารละลายทั้งหมดเข้ากันเป็นเนื้อเดียวและไม่เกิดฟองอากาศ ตั้งไว้ที่อุณหภูมิห้อง เป็นเวลา 10 นาที
4. นำไปวัดค่าการดูดกลืนแสงที่ 595 nm ด้วยเครื่อง Microplate reader และนำค่าที่ได้มาเปรียบเทียบกับกราฟมาตรฐานของโปรตีนที่ทราบความเข้มข้น เพื่อเปรียบเทียบหาความเข้มข้นของโปรตีนตัวอย่าง โดยใช้โปรแกรม Microsoft Excel ในการคำนวณ

ขั้นตอนการตรวจสอบคุณภาพโปรตีน

ทำการตรวจสอบคุณภาพโปรตีนด้วยวิธี SDS-PAGE และ Coomassie staining กล่าวโดยย่อคือ

1. เตรียมเจลโดยใช้ 10% Separating gel (30% Acrylamide/Bis Solution, 1.5M Tris pH 8.8, 10% SDS, 10% APS, 0.1% TEMED) และ 6% Stacking gel (30% Acrylamide/Bis solution, 0.5M Tris pH 6.8, 10% SDS, 10% APS, 0.1% TEMED)
2. เซตระบบในการวิเคราะห์เจล โดยวิเคราะห์ระบบ Gel electrophoresis ด้วยการเท Running Buffer ที่มีส่วนประกอบของ Tris, SDS, และ Glycine จำนวน 1 L ลงในระบบ จากนั้นหยอดตัวอย่างโปรตีนความเข้มข้น 5 μl ผสมกับ Laemli buffer + Protease inhibitor (Tris pH 6.8, SDS, Glycerol, Bromophenol blue, DTT) ปริมาณ 10 μl

และ Broad Range Protein ปริมาณ 4 μ l ทำการต่อระบบเข้ากับกระแสไฟฟ้าโดยแบ่งเป็น 2 ช่วงการทำงาน

ช่วงแรกกำหนด $V = 70$ volt เป็นเวลา 30 นาที

ช่วงที่สอง กำหนด $V = 120$ volt เป็นเวลา 60 นาที

3. แผ่นเจลที่ได้ จะถูกตัดส่วน Stacking gel ออก แชล่งใน Fixative solution (Acetic acid, Methanol, Water) เป็นเวลา 1 ชั่วโมง จากนั้นล้างเจลด้วย Milli-Q Water 3 ครั้ง ครั้งละ 10 นาที แล้วจึงนำแผ่นเจลไปย้อมสีด้วย Colloidal Coomassie Brilliant Blue G-250 Solution (Ammonium sulfate, 85% Ortho-phosphoric acid, Coomassie Brilliant Blue G-250, Methanol) เป็นเวลา 16-18 ชม. แล้วล้างด้วยน้ำ Milli-Q ที่อุณหภูมิ 45-60 องศาเซลเซียส อีก 3 ครั้ง ครั้งละ 10 นาที

4. บันทึกภาพเจลที่ด้วยเครื่อง Gel doc โดยสามารถเก็บ Gel ที่ย้อมสำเร็จไว้ที่อุณหภูมิ 2-8 องศาเซลเซียส

ตอนที่ 3.4 การแยกโปรตีนด้วยเทคนิค 2D-GE และการวิเคราะห์ความแตกต่างของจุดโปรตีน

วัตถุประสงค์: เพื่อแยกโปรตีนตามประจุ Isoelectric point (pI) และมวลโมเลกุลของโปรตีน

วิธีการวิจัย

ขั้นตอนการแยกโปรตีนด้วยเทคนิค Two Dimensional Gel Electrophoresis (2D-GE)

การแยกโปรตีนในมิติที่ 1 (First dimension) เป็นการแยกโปรตีนโดยอาศัยค่า pI (Isoelectric point) ของโปรตีน โดยจะใช้เทคนิค Isoelectric Focusing (IEF) โดยอาศัยระบบ Ettan IPGphor3 Isoelectric Focusing ซึ่งมีขั้นตอนดังต่อไปนี้

1. เจือจางตัวอย่างโปรตีนให้มีความเข้มข้นสุดท้าย 220-350 μ g ด้วย Rehydration buffer ซึ่งมีส่วนประกอบดังนี้
 - 8 M Urea จะทำหน้าที่ช่วยให้โปรตีนละลายได้ดีและทำหน้าที่รักษาสภาพโปรตีน
 - 4% CHAPS เป็นสาร Detergent ช่วยละลายโปรตีนในส่วนที่ไม่ชอบน้ำและลดการเกาะกันของโปรตีน
 - 3mM DTT ช่วยคงสภาพหมู่ Thiol ของโปรตีน ทำให้แถบโปรตีนที่ได้จากการแยกโปรตีนชัดเจนยิ่งขึ้น

- 2% IPG Buffer หรือ Pharmalyte ซึ่งมี pH เท่ากับ IPG Strip ได้แก่ Immobiline DryStrip pH 3-10, 13 cm (Linear) ที่ใช้ ประกอบไปด้วยตัวพาที่มีคุณสมบัติเป็นแอมโฟไลท์ (Ampholyte) มีส่วนช่วยในการเร่งปฏิกิริยา โดยสารเหล่านี้หากเป็นแอมโฟไลท์ที่มีค่า pI สูง จะมีประจุลบจะพาโปรตีนที่มีประจุลบไปยังขั้วแอโนด และหากเป็นแอมโฟไลท์ที่มีค่า pI ต่ำจะมีประจุบวกจะพาโปรตีนที่มีประจุบวกไปยังขั้วแคโทด
 - 0.001% Bromophenol blue เป็นตัวที่ใช้ในการติดตามปฏิกิริยาที่เกิดขึ้น โดยหากมีปฏิกิริยาเกิดขึ้นสารจะเคลื่อนที่ไปยังขั้วแอโนด
2. ทำความสะอาดอุปกรณ์ด้วยกระดาษ Kimwipe ที่ชุบด้วย Methanol จากนั้นใช้ปิเปต ดูด Sample ใส่ลงไปที่กึ่งกลางของ Strip holder อย่างรวดเร็วแต่ระวังอย่าให้เกิดฟอง
 3. ใช้ Forceps คีบ IPG strip ได้แก่ Immobiline DryStrip pH 3-10, 13 cm, Linear (GE Healthcare, Chicago, IL, USA) ออกจากห่อ โดยพยายามคีบบริเวณด้านข้างของ Strip เริ่มจากด้านแอโนดของ Strip แล้วคว่ำด้านเจลลงใน Strip holder ทำอย่างรวดเร็วจนไม่ให้เกิดฟอง ขั้นตอนนี้จะทำการ Passive rehydration เจล หรือเป็นการที่ทำให้ Sample เข้าไปในเจล จากนั้นใช้ปิเปต ดูด Cover fluid เคลือบทับ Strip ใน Strip holder จนเต็ม แล้วปิดด้วย Cover ของ Strip holder
 4. นำไปวางในเครื่องโดยให้ด้านท้ายของ Strip อยู่ด้าน Anode ปิดด้วย Lid adapter เปิดเครื่อง และตั้งค่าโดยแบ่งกระแสไฟฟ้าวออกเป็น 5 ช่วง
 - ช่วงแรก: ช่วง Rehydrate strip ตั้งเวลา 12 ชั่วโมง ที่อุณหภูมิ 20 องศาเซลเซียส
 - ช่วงที่สอง: 500 Volts ไปถึง 500 Volt-h
 - ช่วงที่สาม: 1,000 Volts ไปถึง 800 Volt-h
 - ช่วงที่สี่: 10,000 Volts ไปถึง 27,000 Volt-h

โดยกำหนดปริมาณ A_{max} อยู่ที่ 75 μA ต่อ Strip ที่อุณหภูมิ 20 °C

 - ช่วงสุดท้าย: ช่วง Hold strip 10 ชั่วโมง (ใช้ A แค่ 1 μA)
 5. เตรียม 100mM DTT ใน Equilibration buffer 15 ml และ 150mM Iodoacetamide ใน Equilibration buffer 15 ml โดย Equilibration buffer ประกอบไปด้วย
 - 50mM Tris-HCl pH 8.8 เป็น Equilibration buffer ช่วยรักษา pH ของ strip ให้เหมาะสมในการทำ Electrophoresis

- 6M Urea และ 30% Glycerol จะช่วยเพิ่มความหนืดของ Buffer เพื่อไปลดผลจากการเกิดปรากฏการณ์ Electroendosmosis ซึ่งจะไปรบกวนการเคลื่อนของโปรตีนจาก Strip ไปยังเจลในชั้นมิติที่สอง
 - 2% SDS จะไปทำให้โปรตีนเสียสภาพและเกิด Protein-SDS complex ซึ่งจะมีประจุเป็นลบ
 - 0.002% Bromophenol blue เป็นสารที่ใช้ติดตามกระบวนการ Electrophoresis
6. หลังจากนั้นนำ Strip ออกจาก Strip holder ใส่ลงใน Tube อย่าให้คว่ำหน้าลง ชั้นตอนนี้คือการทำ Equilibration ซึ่งควรทำหลังจาก IEF ทันที
 7. ใส่ 100mM DTT ใน Equilibration buffer ลงไปใน Tube ให้ท่วม Strip โดยสาร DTT จะไปช่วยรักษาสภาพรีดิวซ์ของโปรตีน แช่ไว้ 15 นาที โดยใส่ลงในกล่องโฟมกันเคลื่อน แล้วเขย่าเบา ๆ ตลอด เมื่อครบเวลาให้เทสารออก ล้าง Strip ด้วย น้ำ Mili-Q ในกระบอกตวง และซับน้ำด้วยกระดาษ Kimwipe
 8. ใส่ 150mM Iodoacetamide ใน Equilibration buffer ลงไปใน tube ให้ท่วม strip สาร Iodoacetamide จะไปช่วยป้องกันการเกิด Protein re-oxidation ซึ่งเป็นผลทำให้แถบที่ได้ไม่ได้คุณภาพ แช่ไว้ 15 นาทีโดยใส่ลงในกล่องโฟมกันเคลื่อน เขย่าเบา ๆ ตลอด เมื่อครบเวลาให้เทสารออก ล้าง Strip ด้วย น้ำ Mili-Q ในกระบอกตวง และซับน้ำด้วยกระดาษ Kimwipe
 9. จากนั้นเก็บโดยปิดเกลียวที่อุณหภูมิต่ำกว่า -60 องศาเซลเซียส หากยังไม่วิเคราะห์เจล

การแยกโปรตีนในมิติที่สอง (Second dimension) จะเป็นการแยกโปรตีนโดยอาศัยมวลโมเลกุล (Molecular weight) ของโปรตีน ใช้วิธีที่เรียกว่า Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis หรือ SDS-PAGE ด้วยเครื่อง Protein® II xi Cell (Bio-Rad Laboratories, Hercules, CA, USA) ซึ่งวิธีนี้จะแยกโปรตีนโดยมวลโมเลกุลเพียงอย่างเดียว ประจุของโปรตีนจะไม่มีผลในการแยกด้วยวิธีนี้ เนื่องจากสาร SDS เป็นสารจำพวก Anionic detergent คือเมื่อมีการละลายในน้ำจะแตกตัวเป็นโมเลกุลที่มีประจุลบ ไปล้อมรอบโปรตีนโดยหันเอาส่วนที่ชอบน้ำออกเกิดเป็น SDS-protein complex ซึ่งการเกิดปฏิกิริยาเช่นนี้จึงทำให้ประจุของโปรตีนไม่มีผลต่อการแยก นอกจากนี้ในขั้น Equilibration มีสาร DTT และ Iodoacetamide ที่จะไปทำลาย Disulfide bond ทำให้โปรตีนที่ทำปฏิกิริยากับสารทั้งสองตัวนี้ถูกแยกตามมวลโมเลกุลเพียงอย่างเดียว ซึ่งในขั้นตอนนี้จะใช้ 12.5% Separating gel (สำหรับ 14-100 kDa) ที่ประกอบไป

ด้วย 30% Acrylamide/Bis solution, 1.5M Tris pH 8.8, 10% SDS, 10% Ammonium persulfate (APS) และ 0.1% TEMED ซึ่งเราจะนำ strip ที่ได้จากการแยกในมิติแรกมาใส่ที่ด้านบนของเจล จากนั้นจึงทำการทำต่อระบบเข้ากับกระแสไฟฟ้า โดยแบ่งเป็น 2 ช่วงการทำงาน

ช่วงแรกกำหนด $V = V_{\max}$ กำหนด mA ให้คงที่ = 32mA/gel

ช่วงที่สอง กำหนด $V_{\max} = 300$ V, กำหนดปริมาณ mA ให้คงที่ = 48 mA/gel

หลังจากนั้นจะนำเจลที่วิเคราะห์แยกโปรตีนเสร็จไปย้อมสี โดยเริ่มจากการนำเจลไปแช่ Fixation solution ซึ่งประกอบไปด้วย 10% Acetic acid และ 40% Methanol เป็นเวลา 1 ชั่วโมง จากนั้นล้างเจลด้วยน้ำ Milli-Q 3 ครั้ง ครั้งละ 10 นาที แล้วจึงนำแผ่นเจลไปย้อมสีด้วย Colloidal Coomassie Brilliant Blue G-250 solution ที่ประกอบไปด้วย Ammonium sulfate, 85% Ortho-phosphoric acid, Coomassie Brilliant Blue G-250, and Methanol เป็นเวลา 16-18 ชม. แล้วล้างด้วยน้ำ Milli-Q ที่อุณหภูมิ 45-60 องศาเซลเซียสอีก 3 ครั้ง ครั้งละ 10 นาที แล้วจึงนำไปบันทึกภาพด้วยเครื่อง Image scanner III (GE Healthcare, Chicago, IL, USA) และวิเคราะห์ความแตกต่างของการแสดงออกของโปรตีนด้วยโปรแกรม Image Master 2D Platinum version 7.0 (GE Healthcare, Chicago, IL, USA) โดยที่เจลของแต่ละกลุ่มที่ได้ดำเนินการซ้ำ 3 ครั้ง จะต้องมีความ Consistency โดยซ้อนแผ่นเจลและกำหนดค่า Vector ในโปรแกรมให้แผ่นเจลทั้ง 3 ซ้ำนั้น มีตำแหน่งตรงกัน เพื่อทำการเก็บจุดโปรตีนที่มีความเข้มแตกต่างกันอย่างมีนัยสำคัญ ($P < 0.05$) และความเข้มนั้นจะต้องต่างกันอย่างน้อย 2 เท่า ระหว่างกลุ่มทดสอบและกลุ่มควบคุม ก่อนนำไปวิเคราะห์ต่อยด้วยเทคนิค Mass spectrometry เพื่อระบุชนิดของโปรตีนต่อไป ซึ่งค่าความเข้มที่ได้จะเป็นค่าความเข้มแบบ Relative (Relative spot intensities) โดยเป็นการเทียบค่าของสัญญาณความเข้มของจุดโปรตีนที่วัดได้จากสัญญาณความเข้มของจุดโปรตีนทั้งหมด ทั้งนี้หากโปรตีนมีค่าความเข้มระหว่างแผ่นเจลใน 3 ซ้ำ แตกต่างกันมาก จะต้องมีการตัดค่าที่ต่างกว่าค่าอื่นออกไปเพื่อคำนวณค่าเฉลี่ยความเข้มของจุดโปรตีนใหม่

ตอนที่ 3.5 การย่อยแผ่นเจล สกัดเปปไทด์ และวิเคราะห์เปปไทด์ด้วย LC-MS/MS

วัตถุประสงค์: เพื่อสกัดเปปไทด์จากจุดโปรตีนที่มีความแตกต่างอย่างมีนัยสำคัญหลังจากขั้นตอนการแยกด้วยเทคนิค 2D-Gel และทำให้เปปไทด์บริสุทธิ์ก่อนนำไปวิเคราะห์หาชนิดของโปรตีน

วิธีการวิจัย

ผู้วิจัยจะส่งแผ่นเจลที่ได้ทำการแยกโปรตีนด้วยเทคนิค 2D-GE แล้ว ไปยังคณะเทคนิคการแพทย์ มหาวิทยาลัยมหิดล เพื่อทำการวิเคราะห์หาจุดโปรตีนซ้ำอีกครั้งและวิเคราะห์เปปไทด์ด้วย LC-MS/MS โดยหลังจากวิเคราะห์จุดโปรตีนแล้ว จะทำการย่อยแผ่นเจล สกัดเปปไทด์ และวิเคราะห์เปปไทด์ ดังต่อไปนี้

1. จุดโปรตีนที่มีความเข้มแตกต่างกันอย่างมีนัยสำคัญ ($P < 0.05$) จะถูกตัดออกจากเจลและนำมาล้างสีออกด้วย De-stain solution (50% ACN ใน 100 μ L ของ 25mM ammonium bicarbonate) และย่อยด้วยเอนไซม์ Trypsin ข้ามคืนที่ 37 องศาเซลเซียส ใน 25mM Ammonium bicarbonate เปปไทด์จะถูกนำมาสกัดใน 50 mL of 5% Formic acid/50% ACN แล้วนำไปใส่ใน Ultrasonic bath เป็นเวลา 15 นาที และทำให้แห้งด้วย Speed-vacuum
2. นำตัวอย่างโปรตีนไปละลายด้วย 98% H₂O, 2% ACN and 0.1% formic acid และปั่นแยกที่อุณหภูมิ 4 องศา 15,000 g เป็นเวลา 5 นาที ก่อนจะนำเข้าสู่ระบบการวิเคราะห์ LC-MS/MS ด้วยเครื่อง UltiMate 3000 RSLCnano System (Ultimate 3000, Dionex USA) ต่อกับเครื่อง micrOTOF-Q II ESI-Qq-TOF mass spectrometer (Bruker Daltonics, Germany) ซึ่งเป็นระบบ Q-TOF Mass spectrometry เป็นเทคนิคที่นำหลักการสองหลักการมารวมกัน ประกอบด้วย Quadrupole analyzer และ Time of flight mass analyzer (TOF) ซึ่งมีหลักการดังนี้

เริ่มแรก sample จะถูกเปลี่ยนเป็นไอออนโดยเทคนิค Electrospray ionization โดยเทคนิคนี้จะมีเข็มที่ให้ความต่างศักย์ที่สูง พ่นสิ่งตัวอย่างที่อยู่ในลักษณะของเหลว ทำให้เกิดละอองเกิดขึ้น ละอองที่พ่นออกมาจะระเหย ให้ความหนาแน่นของประจุเพิ่มขึ้นจึงคายไอออนออกมา ไอออนบวกจะถูกส่งเข้าตัว Mass analyzer ซึ่งในขั้นแรกจะเป็น Quadrupole analyzer ที่มี Cylindrical rod สี่อันเรียงขนานกัน ซึ่งตรงนี้ Ion ที่มีมวลต่อประจุที่ต้องการเท่านั้นจึงผ่านไปได้ก่อนจะเข้าสู่ขั้นที่สองนั่นก็คือ TOF analyzer โดยตรงนี้จะมีท่อขนาด 1 เมตรที่ให้ไอออนเคลื่อนที่ผ่าน ซึ่งจะมีการวัดเวลาในการเคลื่อนที่ผ่านซึ่งเวลาจะแปรผกผันกับมวลก่อนจะส่งเข้า Detector

3. ข้อมูลที่ได้จะถูกนำไปวิเคราะห์ระบุหาโปรตีนโดยใช้เครื่องมือ MASCOT search engine และ Swiss-Prot database ซึ่งจะมีการตั้งค่า Fixed modifications เป็น Carbamidomethy (C)

ซึ่งเป็นการเติมกรดอะมิโนเพิ่ม 57 kDa เนื่องจากเรามีการใช้ DTT และ IAA เพื่อให้โปรตีนไม่จับเป็น Disulfide bond จึงต้องตั้งค่าเพื่อให้โปรแกรมคำนวณได้ถูกต้อง ส่วน Variable modifications ตั้งค่าเป็น Oxidation ตามความน่าจะเป็นที่อาจเกิดขึ้นได้ โดยข้อมูล MASCOT score (Matrix Science, London, UK) สำหรับเปปไทด์ควรมีคะแนน MASCOT scores มากกว่า 75 คะแนนขึ้นไป ส่วนโปรตีนควรมีคะแนน MASCOT scores มากกว่าเปปไทด์ แต่ไม่ได้มีเกณฑ์ระบุไว้ชัดเจน และควรดูความ Unique ของลำดับกรดอะมิโนหรือเปปไทด์ที่ตรงกับฐานข้อมูลของโปรตีนนั้น ๆ ซึ่งรายชื่อโปรตีนจะถูกรายงานหลังจากมีการตรวจสอบความถูกต้องของรูปแบบโปรตีน

MASCOT Search Results

Protein View: ROA2_HUMAN
Heterogeneous nuclear ribonucleoproteins A2/B1 OS=Homo sapiens OX=9606 GN=HNRNPA2B1 PE=1 SV=2

Search parameters

- Type of search: MS/MS Ion Search
- Enzyme: Trypsin
- Fixed modifications: Carbamidomethyl (C)
- Variable modifications: Oxidation (M)
- Mass values: Monoisotopic
- Protein mass: Unrestricted
- Peptide mass tolerance: ± 1.2 Da
- Fragment mass tolerance: ± 0.6 Da
- Max missed cleavages: 1
- Instrument type: ESI-QUAD-TOF
- Number of queries: 6,170

Peptide score distribution

There are 253 peptide matches above identity threshold and 342 matches above homology threshold for 6,170 queries. Histogram score range is (0, 164). On average, individual ions scores > 32 (beyond green shading) indicate identity or extensive homology (p<0.05).

Matched peptides shown in bold red.

| Query Start - End | Observed | M(expt) | M(calc) | Delta M Score | Expect Rank | U | Peptide |
|-------------------|----------|-----------|-----------|---------------|-------------|---|----------------------------------|
| #1114 39 - 46 | 544.2723 | 1086.5300 | 1086.4770 | 0.0530 | 0.29 | U | R.NYFEQWGR.L |
| #1115 39 - 46 | 544.2723 | 1086.5309 | 1086.4770 | 0.0539 | 0.30 | U | R.NYFEQWGR.L |
| #1116 39 - 46 | 544.2731 | 1086.5317 | 1086.4770 | 0.0547 | 0.29 | U | R.NYFEQWGR.L |
| #1117 39 - 46 | 544.2735 | 1086.5325 | 1086.4770 | 0.0554 | 0.29 | U | R.NYFEQWGR.L |
| #1118 39 - 46 | 544.2743 | 1086.5341 | 1086.4770 | 0.0571 | 0.29 | U | R.NYFEQWGR.L |
| #1119 39 - 46 | 544.2753 | 1086.5361 | 1086.4770 | 0.0590 | 0.5 | U | R.NYFEQWGR.L |
| #1361 47 - 54 | 505.2703 | 1008.5261 | 1008.4793 | 0.0529 | 0.54 | U | K.LTDCVYMR.D + Oxidation (M) |
| #5891 47 - 59 | 754.2986 | 1506.7962 | 1506.7171 | 0.0852 | 1.14 | U | K.LTDCVYMRPASK.R + Oxidation (M) |
| #5893 47 - 59 | 503.2724 | 1006.7959 | 1006.7171 | 0.0788 | 1.18 | U | K.LTDCVYMRPASK.R + Oxidation (M) |
| #5894 47 - 59 | 503.2726 | 1006.7959 | 1006.7171 | 0.0788 | 1.25 | U | K.LTDCVYMRPASK.R + Oxidation (M) |
| #5895 47 - 59 | 503.2727 | 1006.7962 | 1006.7171 | 0.0791 | 1.19 | U | K.LTDCVYMRPASK.R + Oxidation (M) |
| #5896 47 - 59 | 503.2733 | 1006.7980 | 1006.7171 | 0.0809 | 1.23 | U | K.LTDCVYMRPASK.R + Oxidation (M) |

CHULALONGKORN UNIVERSITY

รูปที่ 17 ตัวอย่าง MASCOT search results

แสดงภาพตัวอย่างผล MASCOT search results โดยจะมีการตั้งค่า Fixed modifications และ Variable modifications รวมถึงส่วนของลำดับเปปไทด์ของโปรตีนที่ตรงกัน และมีลำดับที่ Unique ตรงกับโปรตีนในฐานข้อมูล

ตอนที่ 4 การศึกษาความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่ค้นพบว่าแสดงออก

ผิดปกติในเซลล์เพาะเลี้ยงลิมโฟบลาสต์จากผู้ป่วยโรคอหิวาต์ชิมสเปกตรัม และการยืนยันผล

วัตถุประสงค์: เพื่อเปรียบเทียบหาความสอดคล้องและความสัมพันธ์ระหว่างกลุ่มยีนและกลุ่มโปรตีนที่มีระดับการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยอหิวาต์ชิมสเปกตรัมใน

การศึกษาครั้งนี้ ทั้งในแง่ของชนิดและหน้าที่ทางชีวภาพของโปรตีน และเพื่อยืนยันผลการแสดงออกของยีนและโปรตีนที่พบ

ตอนที่ 4.1 การวิเคราะห์ข้อมูลทางชีวสารสนเทศจากรายชื่อโปรตีนที่พบว่ามี การแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

วัตถุประสงค์: เพื่อกันหากกลุ่มโรคทางระบบประสาทหรือพยาธิสภาพที่เกี่ยวข้องกับโรคออทิซึมสเปกตรัม และหน้าที่ทางชีวภาพ ตลอดจนสร้างเครือข่ายทางชีวภาพที่แสดงความเกี่ยวข้องกับกลุ่มโปรตีนดังกล่าว

วิธีการวิจัย

รายชื่อของโปรตีนที่พบว่ามี การแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยจะถูกนำไปทำ Gene ontology analysis เพื่อทำนายหน้าที่ทางชีวภาพของโปรตีนเหล่านั้น ซึ่งรายชื่อโปรตีนจะถูกนำเข้าโปรแกรม Ingenuity Pathway Analysis (IPA) ซึ่งเป็นโปรแกรมชีวสารสนเทศที่ใช้ในการวิเคราะห์หรือทำนายหน้าที่ทางชีวภาพของสารชีวโมเลกุลต่าง ๆ และยังเป็นเครื่องมือในการค้นหาข้อมูลสำหรับงานด้านโอมิกส์ โดยโปรแกรม IPA จะเก็บรวบรวมข้อมูลต่าง ๆ จากผลงานวิจัยที่ได้รับการตีพิมพ์เกี่ยวกับโรค Canonical pathways หน้าที่ทางชีวภาพ ปฏิสัมพันธ์ และตัวควบคุมระหว่างยีนกับโปรตีนในแง่มุมต่าง ๆ ไว้ในฐานะข้อมูลของระบบ เนื่องจากยีนหรือโปรตีนหนึ่งตัว อาจเคยถูกรายงานว่ามีความเกี่ยวข้องได้หลายบทบาทตามข้อมูลที่เคยมีการตีพิมพ์เอาไว้ เพื่อที่จะทำนายความเกี่ยวข้องกับโรคออทิซึมสเปกตรัม โปรแกรม IPA จะทำการ Overlap รายชื่อของยีนหรือโปรตีนกับฐานข้อมูลเพื่อวิเคราะห์ผล ซึ่งจะมีการคำนวณค่าความน่าเชื่อถือหรือนัยสำคัญทางสถิติโดยใช้ Fisher's Exact test ($P < 0.05$) ทั้งนี้โปรแกรมดังกล่าวคณะผู้วิจัยได้รับความอนุเคราะห์จาก Dr. Valerie Hu มหาวิทยาลัยจอร์จวอชิงตัน กรุงวอชิงตัน ดีซี ประเทศสหรัฐอเมริกา

ตอนที่ 4.2 การวิเคราะห์ความสัมพันธ์ของกลุ่มโปรตีนที่มีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

วัตถุประสงค์: เพื่อทำการกำหนดเกณฑ์ในการคัดเลือกการศึกษาที่นำมาวิเคราะห์ความสอดคล้องตลอดจนวิเคราะห์ความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่มีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

วิธีการวิจัย

รายชื่อของโปรตีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยจะถูกนำไปทำ Overlapping analysis ร่วมกับยีนและโปรตีนที่แสดงออกผิดปกติในผู้ป่วยออทิซึมสเปกตรัม โดยแบ่งการ Overlap ได้ออกเป็น 3 ส่วน

ส่วนที่ 1 เป็นการ Overlap ระหว่างรายชื่อโปรตีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยกับรายชื่อยีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยจาก GSE15402 ซึ่งมีความสอดคล้องกับตัวอย่างจากการศึกษารูปแบบโปรตีโอมของผู้ป่วยจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ ซึ่งผู้วิจัยได้ดำเนินการตามขั้นตอนที่ 2

ส่วนที่ 2 เป็นการ Overlap ระหว่างรายชื่อโปรตีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมกับรายชื่อยีนที่แสดงออกผิดปกติจากเซลล์เม็ดเลือดของผู้ป่วยออทิซึมสเปกตรัมที่ได้จากฐานข้อมูล GEO DataSers ใน NCBI จำนวนทั้งสิ้น 6 การศึกษา (Alter et al., 2011; Gazestani et al., 2019; Gregg et al., 2008; Kong et al., 2012; Kuwano et al., 2011; Pramparo et al., 2015) มาวิเคราะห์ด้วยโปรแกรม Multi-Experimental Viewer (MEV) เพื่อหาพื้นที่แสดงออกแตกต่างกันอย่างมีนัยสำคัญด้วยวิธีการทางสถิติ T-test แบบ Adjusted Bonferroni correction ($P < 0.05$) โดยมีเกณฑ์ในการคัดเลือกดังนี้ (1) เป็นการศึกษาแบบทรานสคริปโตมในผู้ป่วยออทิซึมสเปกตรัมเทียบกับกลุ่มควบคุมที่ไม่ได้เป็นโรคออทิซึมสเปกตรัม (2) เป็นการศึกษาที่ใช้สิ่งตัวอย่างจากเซลล์เม็ดเลือดของกลุ่มตัวอย่าง (3) มีจำนวนกลุ่มตัวอย่างไม่น้อยกว่า 3 รายต่อกลุ่ม โดยมีการศึกษาที่ผู้วิจัยได้คัดเลือกมาแสดงดังตารางที่ 1

ตารางที่ 1 รายการการศึกษาในรูปแบบทรานสคริปโตมจากเซลล์เม็ดเลือดของผู้ป่วยออทิซึมสเปกตรัม

| No. | GEO DataSets | Author | Study title | Sample information | | | No. of transcripts | |
|-----|--------------|------------------------|---------------------------------------------------------------------------------------------------------------------------------|------------------------------|------------------------------|-------------------------|--------------------|-------------------|
| | | | | Type | Controls | Sample size | #DEGs | #protein products |
| 1 | GSE6575 | Gregg et al., 2008 | Gene expression in blood of children with autism spectrum disorder | Whole blood | Age and sex-matched controls | 35 ASD 12 controls | 302 | 266 |
| 2 | GSE25507 | Alter et al., 2011 | Autism and increased paternal age-related changes in global levels of gene expression regulation | Peripheral blood lymphocytes | Age and sex-matched controls | 82 ASD 64 controls | 2369 | 1712 |
| 3 | GSE26415 | Kuwano et al., 2011 | Autism-associated gene expression signatures in peripheral blood leukocytes | Leukocytes | Age-matched controls | 21 ASD 42 controls | 5210 | 3420 |
| 4 | GSE18123 | Kong et al., 2012 | Blood gene expression signatures distinguish autism spectrum disorders from controls | Whole blood | Age and sex-matched controls | 170 ASD 115 controls | 3130 | 2123 |
| 5 | GSE42133 | Pramparo et al., 2015 | Disrupted functional networks in autism underlie early brain maldevelopment and provide accurate classification | Leukocytes | Age and sex-matched controls | 91 ASD 56 controls | 2628 | 2082 |
| 6 | GSE111175 | Gazestani et al., 2019 | Perturbations of PI3K/AKT, RAS/ERK, WNT/B-catenin networks in leukocytes are linked to ASD genetics and fetal origins of autism | Leukocytes | Age and sex-matched controls | 119 ASD 107 controls | 2835 | 2313 |

ส่วนที่ 3 เป็นการ Overlap ระหว่างรายชื่อโปรตีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยกับรายชื่อโปรตีนที่เคยมีรายงานว่าพบการแสดงออกผิดปกติในเลือดของผู้ป่วยออทิซึมสเปกตรัมในฐานข้อมูลชีวสารสนเทศ (L. Shen et al., 2020) จำนวนทั้งสิ้น 7 การศึกษา ซึ่งจะถูกรวบรวมขึ้นโดยมีเกณฑ์ในการคัดเลือกดังนี้ (1) เป็นการศึกษาในรูปแบบโปรตีโอมในผู้ป่วยออทิซึมสเปกตรัมเทียบกับกลุ่มควบคุมที่ไม่ได้เป็นโรคออทิซึมสเปกตรัม (2) เป็นการศึกษาที่ใช้สิ่งตัวอย่างจากเลือดของกลุ่มตัวอย่าง (3) มีจำนวนกลุ่มตัวอย่างไม่น้อยกว่า 3 รายต่อกลุ่ม โดยมีการศึกษาที่ผู้วิจัยได้คัดเลือกมาแสดงดังตารางที่ 2

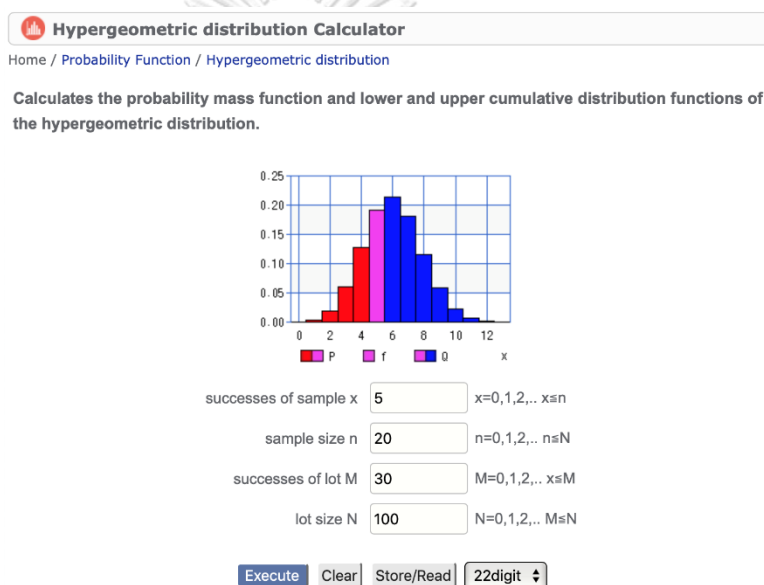


ตารางที่ 2 รายการการศึกษาแบบโปรตีโอมในเลือดของผู้ป่วยออทิซึมสเปกตรัม

| No. | Author | Study title | Sample information | | Related proteins | |
|-----|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------|---------------------------------|------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| | | | Type | Size | # DEPs | Protein lists |
| 1 | Corbett et al., 2007 | A proteomic study of serum from children with autism showing differentially expression of apolipoproteins and complement proteins | Serum | 69 ASD 35 controls | 4 | Increased: APOB, C1Q, FHR1, FN1 |
| 2 | Ngounou Wetie et al., 2014 | A pilot proteomic study of protein markers in autism spectrum disorder | Serum | 7 ASD 7 controls | 3 | Increased: APOA1, APOA4, PON1 |
| 3 | Cortelazzo et al., 2016 | Expression and oxidative modifications of plasma proteins in autism spectrum disorders: Interplay between inflammatory response and lipid peroxidation | Plasma | 30 ASD 30 controls | 12 | Increased: A2M, AAT, HP, FGB, FGG, IGH1, IGHG Decreased: APOA1, APOA4, APOJ, PALB, TIBC |
| 4 | Feng et al., 2017 | Redox proteomic identification of carbonylated proteins in autism plasma: insight into oxidative stress and its related biomarkers in autism | Plasma | 15 ASD 15 controls | 2 | Increased: C8A, IGKC |
| 5 | Shen et al., 2018 | iTRAQ-based proteomic analysis reveals protein profile in plasma from children with autism | Plasma | 30 ASD 30 controls | 23 | Increased: AAT, VTN, AGT, APOE, C3, C5, EHD3, FN1, FBLN1, IGFALS Decreased: ACTG1, ACTN1, ENO1, PARVB, CALM, CALR, FERMT3, ITGA2B, MAPRE2, TLN1 THBS1, VCP, VCL |
| 6 | Shen et al., 2019 | Proteomics study of peripheral blood mononuclear cells (PBMCs) in autistic children | PBMCs | 30 ASD 30 controls | 48 | Increased: C3, CALR, SERPINA1, ATP5A1, ATP5B, MDH2, IFNG, IL1B, IL6, IL1, TNF, IDH2, HYOU1, VIM, HSP90B1, ATP5F1B, HSPD1, APOA1, CBX3, LCP1, GLUD1, SAMHD1, DLD, AHNK, ACO2, ME2, ETFA, HSPA5, P4HB, GOT2, HSPA9, IQGAP1, LMNB2, ATP5F1A, ALB, SERPIND1, ITIH4, LBP, LMNB1 Decreased: FLOT1, UQCRC2, VDACC3, USP9X, FLOT2, PLEC, GPD2, STOM, ANXA11 |
| 7 | Pichitpunpong et al., 2019 | Phenotypic subgrouping and multi-omics analyses reveal reduced diazepam-binding inhibitor (DBI) protein levels in autism spectrum disorder with severe language impairment | LCLs | 6 ASD in subgroup 6 controls | 18 | Increased: COX5A, LGALS1, GSTP1, HNRNP1A1, ERH, CALM1 Decreased: TUBB, DLD, ANXA5, H3F3C, PGAM1, ENO1, TPT1, CLTA, CCT5, GSTP1, AHSB, IDH2, ERH |

หลังจากที่ได้รายชื่อยีนหรือรายชื่อโปรตีนที่แสดงออกผิดปกติแล้ว ผู้วิจัยจะทำการ Overlap ข้อมูลเพื่อหารายชื่อโปรตีนที่พบร่วมกันกับข้อมูลในแต่ละส่วน ด้วยเครื่องมือ Venn diagram ในโปรแกรม Venny 2.1.0 (Oliveros, 2007-2015) แล้วนำจำนวนโปรตีนที่ปรากฏร่วมกันมาวิเคราะห์หาความสัมพันธ์ทางสถิติของความสัมพันธ์โดยใช้ Hypergeometric distribution analysis จากเครื่องมือ Keisan online calculator เป็นการคำนวณหาความสัมพันธ์ทางสถิติ ในกรณีที่ตัวอย่างถูกสุ่มแบบ Without replacement selected มีขั้นตอนในการวิเคราะห์ดังนี้

1. เข้าเว็บไซต์ <https://keisan.casio.com/exec/system/1180573201> เพื่อไปยังเครื่องมือ Hypergeometric distribution calculator ดังภาพ



รูปที่ 18 Hypergeometric distribution calculator

แสดงภาพเครื่องมือในการคำนวณค่าความสัมพันธ์ทางสถิติจาก Keisan online calculator

2. เติมจำนวนโปรตีนที่ได้จากการทำ Overlapping analysis ลงในช่องว่าง โดยแต่ละช่องมีความหมายดังนี้

Success of sample x = จำนวนโปรตีนที่พบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์และพบในข้อมูลอีกส่วนที่เราทำการ Overlap

Sample size n = จำนวนโปรตีนทั้งหมดที่พบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์

Successes of lot M = จำนวนยีนหรือโปรตีนในข้อมูลอีกส่วนที่เราทำการ Overlap

Lot size N = จำนวน Predicted protein encoded จากจีโนมของมนุษย์ซึ่งได้จากฐานข้อมูล HUPO (Human Proteome Organization) โดยมีจำนวนทั้งสิ้น 19,773 โปรตีน

3. กดปุ่ม Execute และดูผลลัพธ์ของค่านัยสำคัญทางสถิติในช่อง Upper cumulative distribution ($P < 0.05$)

ตอนที่ 4.3 การยืนยันผลการแสดงออกของยีนและโปรตีนที่สนใจโดยใช้วิธี Real-time PCR และ Western blot

วัตถุประสงค์: เพื่อเป็นการยืนยันผลจากการวิเคราะห์การแสดงออกของยีนจากการศึกษารูปแบบทรานสคริปโตม และการแสดงออกของโปรตีนจากการศึกษารูปแบบโปรตีโอม กลุ่มของยีนหรือโปรตีนที่แสดงออกผิดปกติ และมีความเกี่ยวข้องอย่างมีนัยสำคัญต่อเครือข่ายและหน้าที่ทางชีวภาพที่เกี่ยวข้องกับโรคออทิซึมสเปกตรัมจะถูกคัดเลือกมาทำการวิเคราะห์ด้วยวิธี Real-time PCR และ Western blot

วิธีการวิจัย

เกณฑ์ในการเลือกยีนหรือโปรตีนมาทำการยืนยันผล

หลังจากที่ผู้วิจัยได้ทำการศึกษารูปแบบทรานสคริปโตมและโปรตีโอม และได้มีการศึกษาความสอดคล้องและความสัมพันธ์กับการศึกษาในผู้ป่วยออทิซึมสเปกตรัมจากการศึกษาอื่นแล้ว ผู้วิจัยจะทำการคัดเลือกยีนและโปรตีนมายืนยันผลของระดับการแสดงออก ด้วยวิธี Real-time qPCR และ Western blot ตามลำดับ โดยมีเกณฑ์ในการคัดเลือกดังต่อไปนี้

- มีค่านัยสำคัญทางสถิติ P-value ต่ำ จากการศึกษารูปแบบโปรตีโอม
- มีระดับการเปลี่ยนแปลงมาก จากค่า Log 2 ratio
- เคยมีการรายงานว่าพบในผู้ป่วยออทิซึมสเปกตรัมมาก่อน (ASD candidate genes/proteins)
- มีทิศทางการเปลี่ยนแปลงที่มีความสอดคล้องกับที่เคยรายงานในผู้ป่วยออทิซึมสเปกตรัม
- มีค่า MASCOT score และ %Coverage สูง
- มีหน้าที่ที่มีความเกี่ยวข้องกับพยาธิสภาพที่พบในผู้ป่วยออทิซึมสเปกตรัม

ขั้นตอนการทำ Real-time qPCR

ยีนที่มีการแสดงออกผิดปกติจากการศึกษารูปแบบทรานสคริปโตมและผ่านการทำการวิเคราะห์ความสัมพันธ์ระหว่างยีนกับโปรตีนแล้ว จะถูกเลือกมาวิเคราะห์เชิงปริมาณด้วยวิธี RT-PCR โดย RNA ที่สกัดจากเซลล์เพาะเลี้ยงชนิดของผู้ป่วยอหิวาต์ซึ่งอยู่ใน RNAlater (Thermo Fisher Scientific, Waltham, MA, USA) จะถูกนำมาสกัดต่อตามกระบวนการตามคำแนะนำของบริษัทผู้ผลิตน้ำยา ซึ่ง RNA ที่ได้นั้นจะถูกเปลี่ยนเป็น cDNA หลังจากนั้น cDNA ที่ได้จะถูกเป็นต้นแบบสำหรับปฏิกิริยา Real-time PCR ต่อไป โดยใช้น้ำยา AccuPower® RT PreMix (Bioneer corporation, Daejeon, Korea) และ RealMOD™ Green W2 2X qPCR mix (iNtRON Biotechnology, Korea) ตามลำดับ

สำหรับการออกแบบ primer ที่จำเพาะต่อยีนที่เลือกมา จะใช้โปรแกรมชีวสารสนเทศช่วยในการออกแบบ Primer โดยใช้โปรแกรม UCSC genome browser primer3 และ Ensembl ซึ่งจะต้องเลือก Exon สำหรับให้ Forward primer และ Reverse primer มาจับ ซึ่งต้องห่างกันเพื่อป้องกันการ Contamination จาก Genomic DNA

ขั้นตอนการทำ Reverse transcription โดยใช้น้ำยา AccuPower® RT PreMix นั้น RNA จำนวน 0.5-1 µg จะถูกนำมาผสมกับ OligoT₁₈ primer ปริมาณ 0.5 µg แล้วนำไปปมที่อุณหภูมิ 70 องศาเซลเซียสเป็นเวลา 5 นาที ปิดเตา DEPC-water เพื่อปรับปริมาตรของปฏิกิริยาให้เท่ากับ 20 µl ลงใน AccuPower® RT PreMix โดยภายในจะมีเอนไซม์ RTnase สำหรับปฏิกิริยา Reverse transcription โดย cDNA ที่ได้จะถูกใช้เป็นต้นแบบสำหรับปฏิกิริยา Real-time PCR ในขั้นตอนต่อไป

ขั้นตอนการทำ Real-time PCR โดยใช้ชุดน้ำยา RealMOD™ Green W2 2X qPCR mix โดย cDNA ที่ได้จากขั้นตอน Reverse transcription จำนวน 2-5 µl นำมาเปิดลงใน Real-time PCR tube หลังจากนั้นเติม qPCR master-mix ซึ่งประกอบไปด้วย 2X Greenstar จำนวน 10 µl Specific primers ทั้ง Reverse primer และ Forward primer อย่างละ 1 µl จากนั้นใช้น้ำ PCR grade water ปรับจนได้ปริมาตร 20 µl แล้วจะนำมาทำปฏิกิริยา Pre-denaturation denaturation และ Annealing/Extension โดยจะใช้การคำนวณการแสดงออกของยีนด้วยวิธี Delta-delta CT value และใช้ Human RNA 18s เป็น Endogenous control

ขั้นตอนการทำ Western blot

Western blot เป็นวิธีการที่จะใช้ยืนยันโปรตีนที่ได้จากการวิเคราะห์ด้วย Mass spectrometry และโปรแกรมชีวสารสนเทศ โดยจะเลือกกลุ่มโปรตีนที่แสดงออกผิดปกติ และมีความเกี่ยวข้องอย่างมีนัยสำคัญต่อเครือข่ายและหน้าที่ทางชีวภาพที่เกี่ยวข้องกับโรคอหิวาต์ที่ซึมสเปกตรัมมาทำการวิเคราะห์ยืนยันผล

การทำ Western Blot หรือเรียกอีกชื่อหนึ่งว่า Protein immunoblot เนื่องจากการใช้แอนติบอดีที่จำเพาะในการตรวจวัดแอนติเจน ซึ่งเป็นวิธีที่เป็นที่ยอมรับสำหรับการวิเคราะห์โปรตีนที่จำเพาะในสิ่งตัวอย่าง โดยการใช้ SDS-Polyacrylamide Gel Electrophoresis (SDS-PAGE) เพื่อแยกโปรตีนที่มีความหลากหลายออกจากกัน ไม่ว่าจะเป็นในด้านโครงสร้างสามมิติ หรือความยาวของ Polypeptide โดยย้ายโปรตีนที่ต้องการแยกลงบนพื้นผิว Nitrocellulose membrane ซึ่งโปรตีนจะถูกย้อมด้วยแอนติบอดี (เช่นเดียวกับการใช้ Probe) ที่จำเพาะกับโปรตีน โปรตีนเป้าหมายจะถูกวิเคราะห์การเคลื่อนที่ในสนามไฟฟ้าจากความเข้มข้น (Intensity) และตำแหน่ง (Location) เนื่องจากวิธีนี้มีประสิทธิภาพสูง จำเพาะ และแม่นยำ เพราะสามารถแยกโปรตีนเป้าหมายที่มีขนาดเล็กถึง 1 ng ได้ จึงนิยมใช้ในเทคนิคทางอณูชีววิทยา ชีวเคมี และงานทางด้านพันธุศาสตร์ภูมิคุ้มกัน

Western Blot จะประกอบด้วย 2 ขั้นตอนหลัก คือ SDS-PAGE และ Protein blotting and detection

1. SDS-polyacrylamide gel electrophoresis (SDS-PAGE)

การแยกในขั้นแรกนี้ เป็นการแยกโปรตีนตามมวลโมเลกุลเป็นหลัก โดยจะมีการแยกโปรตีนตามอนุภาคประจุที่เคลื่อนที่ตรงข้ามกับขั้ว Electrode ภายใต้สนามไฟฟ้า โดยขึ้นกับประจุ ขนาดโมเลกุล และโครงสร้างของโปรตีน

ในส่วนของ SDS ที่ได้กล่าวไป จะพบว่า การเคลื่อนที่ในสนามไฟฟ้าจะขึ้นกับน้ำหนักโมเลกุล แขนงประจุและขนาดของโปรตีน ซึ่ง SDS คือ Anionic detergent ที่สามารถทำลายพันธะไฮโดรเจน ทั้งในและระหว่างโมเลกุลเพื่อที่จะคลายโปรตีน และทำลายโครงสร้างทุติยภูมิและตติยภูมิได้ เช่นเดียวกับ Denaturing agent และ Hydrotropy agent โดยในโปรตีนส่วนใหญ่ การจับของ SDS ที่สาย Polypeptide จะทำให้เกิดการกระจายของประจุต่อมวล ด้วยวิธีดังกล่าว ประจุที่แท้จริงของ Polypeptide จะไม่สำคัญอีกต่อไปเมื่อเทียบกับประจุลบที่เกิดจาก SDS เนื่องจากมีแรงผลักทางไฟฟ้าเกิดขึ้น จึงทำให้โปรตีนคลายตัวเป็นรูปแท่งและสามารถแยกความต่างในเจลได้ ดังนั้นสิ่ง

ตัวอย่างที่ทำการตรวจวิเคราะห์จะผสมอยู่กับ SDS และทำปฏิกิริยาตามกระบวนการที่เกี่ยวข้องอย่างเป็นลำดับขั้นตอน โดยให้ความร้อนแก่สิ่งตัวอย่าง อย่างน้อย 60 องศาเซลเซียส เพื่อให้โปรตีนถูกทำลายและเกิดการ Depolymerization รวมทั้งยังช่วยให้ SDS จับและคลายโปรตีนเป็นรูปแท่งและยึดกันด้วยประจุลบ อาจเติมสี Bromophenol blue เพื่อให้เห็นการดำเนินของกระบวนการในระหว่างการวิเคราะห์ในกระแสไฟฟ้า และเติม Glycerol ในปริมาณที่เหมาะสมเพื่อเพิ่มความหนาแน่นและเร่งการเคลื่อนที่ของสิ่งตัวอย่าง

โดยในกระบวนการ Gel electrophoresis จะใช้ระบบบัฟเฟอร์ที่มีค่าของ pH ที่แตกต่างกัน คือ Tris-glycine หรือระบบ “Laemmli” ที่เริ่มจาก pH 6.8 และโปรตีนจะเกิดการแยกที่ pH ประมาณ 8.3-9.0 แต่ระบบนี้มีข้อเสียคือที่ pH นี้ อาจทำให้เกิดพันธะ Disulfide ระหว่าง Cysteine ในโปรตีน (เนื่องจากค่า pKa ของ Cysteine อยู่ในช่วง 8-9) และ Reducing agent ในบัฟเฟอร์ที่ใส่ไปไม่เคลื่อนที่ไปกับโปรตีน จึงมีการแก้ไขปัญหาโดยการแยกโปรตีนที่ pH ที่ต่ำกว่า pKa ของ Cysteine เช่นการใช้ Bis-tris, pH 6.5 ซึ่งทำให้เจลมีความเสถียรมากขึ้นด้วยจาก pH ที่ต่ำลง ทำให้สามารถเก็บเจลไว้ได้นานมากขึ้นก่อนการใช้งาน รวมถึงการใช้ Reducing agent เช่น Sodium bisulfite ที่เคลื่อนที่นำหน้าโปรตีนเพื่อรักษาสถานะของปฏิกิริยา

พื้นที่ที่กระแสไฟฟ้าเข้ามา ประจุลบจะเคลื่อนเข้าหาขั้วบวก นำโดย Cl^- ตามด้วย Glycinate (Gly^-) ส่วนอนุภาค SDS-protein จะอยู่ระหว่าง Cl^- ของเจลบัฟเฟอร์ และ Gly^- ของ Cathode บัฟเฟอร์ ที่ตรงขอบ เนื่องจากกระแสไฟฟ้าที่ลดลงโปรตีนจึงเกิดการแยกในชั้นเจล โดยโปรตีนที่มีความเป็นประจุลบมากจะเคลื่อนที่ไว และโปรตีนที่มีน้ำหนักโมเลกุลน้อยก็เช่นกัน ทำให้เกิดการแยกของโปรตีนในตำแหน่งต่าง ๆ ของเจล

Polyacrylamide Gel Electrophoresis (PAGE) ใช้ในการแยกโปรตีนที่มีขนาดตั้งแต่ 5-2000 kDa ขึ้นอยู่กับรูของเจลโดยควบคุมที่ความเข้มข้นของ Acrylamide ตอนเตรียมเจล โดย Stacking gel (6%) จะถูกเทอยู่ข้างบน Resolving gel และใส่ Comb เพื่อเกิด Wells ขึ้น ซึ่งเปอร์เซ็นต์ความเข้มข้นที่เลือกจะขึ้นอยู่กับขนาดของโปรตีนที่ใช้ในสิ่งตัวอย่าง หากโปรตีนที่ทราบขนาดนั้นมีขนาดเล็กจะต้องใช้เปอร์เซ็นต์สูง และการเปลี่ยนระบบบัฟเฟอร์จะช่วยให้สามารถแยกโปรตีนที่มีขนาดเล็กมาก ๆ ได้ดียิ่งขึ้น

ตารางที่ 3 ร้อยละความเข้มข้นของเจลที่เหมาะสมสำหรับการแยกมวลโมเลกุลของโปรตีน

| Range of molecular weight (KDa) | Concentration of gel (%) |
|---------------------------------|--------------------------|
| <10 | 15 |
| 10-30 | 12 |
| 30-100 | 10 |
| 100-500 | 8 |
| >500 | 5 |

2. Western blotting and detection

ประกอบด้วย 5 ขั้นตอน ได้แก่ Transfer, Blocking, Primary antibody incubation, Secondary antibody incubation, Protein detection and analysis

2.1 Transfer

โปรตีนจะถูกย้ายจากในเจลมาอยู่บน Nitrocellulose (NC) หรือ Poly-vinylidene difluoride (PVDF) membrane โปรตีนจะจับกับ NC membrane ด้วยแรง Hydrophobic และ membrane จะสามารถผลิตสีย้อมที่ไม่จำเพาะออกมา วิธีนี้จึงเป็นวิธีที่ต้นทุนต่ำและง่ายต่อการใช้งาน แต่อย่างไรก็ตามโปรตีนขนาดเล็กก็อาจสูญหายได้ขณะที่ทำการล้าง ส่วน PVDF membrane นั้น มีประสิทธิภาพสูงกว่า แต่จำเป็นต้องแช่ใน Methanol ก่อนใช้งาน เพื่อให้เกิดประจุบวกบน Membrane ช่วยในการจับกับประจุลบของโปรตีน

NC membrane จะถูกนำมาใช้อย่างจำเพาะ เนื่องจากมีขนาดของรูที่แตกต่างกันไปตามน้ำหนักโมเลกุลของโปรตีน ซึ่งขนาดที่นิยมใช้มากที่สุดคือขนาด 0.45 และ 0.2 μm โดยขนาด 0.45 μm จะใช้กับโปรตีนที่มีน้ำหนักโมเลกุลมากกว่า 20 kDa ส่วนขนาด 0.2 μm จะใช้กับโปรตีนที่มีน้ำหนักโมเลกุลน้อยกว่า 20 kDa

PVDF membrane จะใช้ในการตรวจหาโปรตีนที่มีน้ำหนักโมเลกุลขนาดเล็กได้ดีที่สุด เนื่องจากมีความไวและความละเอียดสูงกว่า Membrane ทั่วไป

เทคนิค Transfer ที่นิยมใช้มากที่สุดคือเทคนิค Semi-dry transfer และ Wet transfer โดยเทคนิค Semi-dry transfer จะใช้หลักการ Membrane-filter sandwich อยู่ระหว่าง Filter กับ บัฟเฟอร์ ซึ่งกระบวนการ Transfer จะขึ้นอยู่กับการนำของกระแสไฟฟ้าที่ผลิตจากบัฟเฟอร์ เทคนิคนี้ จะใช้เวลาเพียงเล็กน้อยเนื่องจากกระแสไฟฟ้าจะผ่านชั้นเจลและ Membrane โดยตรง ในขณะที่

เทคนิค Wet transfer ตัว Membrane-filter sandwich จะอยู่ในถัง Transfer ในแนวตั้ง ดังนั้น โปรตีนจะถูก Transfer ภายใต้การควบคุมของสนามไฟฟ้าที่ผลิตจากแผ่นขั้วไฟฟ้าที่ขนานกับ Sandwich จึงใช้เวลามากกว่า โปรตีนที่ถูกย้ายจะมีประสิทธิภาพดีกว่า และสามารถย้ายโปรตีนได้จากหลายเจล

2.2 Blocking

เป็นการยับยั้งบริเวณทำปฏิกิริยาบน Membrane เพื่อลดจำนวนการจับของโปรตีนที่ไม่จำเพาะในแต่ละขั้นตอนของการทำงานที่มีการใช้ Inert protein หรือ Nonionic detergent โดย Blocking buffer จะสามารถยับยั้งได้ทุกบริเวณ ไม่วางโปรตีนเป้าหมายลงบน Membrane ซ้ำ ไม่จับกับ Epitope ของโปรตีนเป้าหมาย และไม่เกิด Cross-reaction กับแอนติบอดีหรือ Detection reagent ซึ่งตัวยับยั้งที่ใช้โดยทั่วไป ได้แก่ BSA, Nonfat dry milk, Casein, Gelatin, and Nonionic detergent Tween-20 ส่วนบัฟเฟอร์ที่ใช้มากที่สุดคือ TBS และ PBS

2.3 Primary antibody incubation

หลังจากขั้นตอนการยับยั้ง Primary antibody ที่จำเพาะกับโปรตีนเป้าหมายจะถูก Incubate ไปกับ Membrane เพื่อให้เกิดการจับกับโปรตีน โดย Primary antibody ควรจะมีการตรวจสอบก่อนใช้งาน และการเลือกใช้จะขึ้นอยู่กับแอนติเจนที่ใช้ในตรวจวัด ซึ่งสามารถใช้ได้ทั้ง Polyclonal และ Monoclonal antibody โดยที่ Monoclonal antibody จะจับกับ Single specific antigenic epitope จึงมีความจำเพาะสูงกว่า แต่หาก Epitope เป้าหมายถูกทำลายจะมีผลได้ การใช้ Polyclonal จึงให้ความสามารถในการจับมากกว่า และเสถียรกว่าแม้ว่า Epitope เป้าหมายจะถูกทำลาย

2.4 Secondary antibody incubation

หลังจากล้าง Primary antibody ที่ไม่ได้จับออกไปแล้ว Specific enzyme conjugated secondary antibody จะเข้าจับกับ Primary antibody ที่ทำปฏิกิริยากับโปรตีนเป้าหมายบน Membrane โดยแอนติบอดีที่นิยมใช้ที่สุดคือ Anti-mouse และ Anti-rabbit immunoglobulin เนื่องจาก Host ของ Primary antibody คือหนูและกระต่าย จึงใช้แพะในการเลี้ยง Anti-mouse และ Anti-rabbit polyclonal antibodies ดังนั้นจึงถือได้ว่านิยมใช้ Goat anti-mouse และ Goat anti-rabbit immunoglobulin มากที่สุด โดยการเลือกใช้จะขึ้นกับชนิดของสัตว์ที่เลี้ยง Primary antibodies เช่น ถ้าเกิด Primary antibody เป็น Mouse monoclonal antibody ดังนั้น

Secondary antibody จะต้องเป็น Anti-mouse antibody

2.5 Protein detection and analysis

Protein detection (Color development)

สารตั้งต้นจะทำปฏิกิริยากับแอนติบอดีที่ติดอยู่กับ Secondary antibody เพื่อให้เกิดสี ระดับโปรตีนเป้าหมายในเซลล์หรือเนื้อเยื่อจะถูกประเมินผ่านเครื่อง Densitometry และประเมินจากตำแหน่งของแถบโปรตีนที่เกิดขึ้น

Alkaline phosphatase (AP) และ Horseradish peroxidase (HRP) คือแอนติบอดีที่นิยมใช้กันอย่างแพร่หลาย สำหรับหน้าที่ของ AP ในการเร่งปฏิกิริยานั้น BCIP ซึ่งเป็นสารตั้งต้นที่ไม่มีสีจะถูกเปลี่ยนเป็นผลิตภัณฑ์ที่มีสีน้ำเงิน โดย H_2O_2 , 3-amino-9-ethyl carbazole และ 4-chlorine naphthol จะถูกออกซิไดส์เป็นสีน้ำตาลของสารตั้งต้น จากนั้นจะเกิดสีน้ำเงินของผลิตภัณฑ์โดยการเร่งปฏิกิริยาของ HRP การเพิ่มสี Chemiluminescence เป็นอีกวิธีหนึ่งที่ใช้ในการตรวจวัด HRP ในการใช้ HRP สารเรืองแสงต้นตัม Luminol จะถูกออกซิไดส์โดย H_2O_2 และเกิดการเรืองแสงขึ้น นอกเหนือจากนี้ตัวเพิ่มการทำงานจะยังช่วยให้แสงมีความเข้มมากขึ้น 1000 เท่า HRP จะสามารถตรวจวัดได้เมื่อจุดโปรตีนถูกกระตุ้นบน Photographic film

Analysis

a. Control design

เพื่อรับรองความถูกต้องและความแม่นยำของผลการทดสอบ ทำให้สามารถระบุปัญหาที่เกิดขึ้นได้อย่างรวดเร็วและตรงประเด็น

ชนิดของ Control

Positive control: ย่อยเซลล์เพาะเลี้ยงชนิดหรือเนื้อเยื่อตัวอย่างที่ทราบการแสดงออกของโปรตีนที่เราต้องการตรวจวัด ใช้เพื่อยืนยันว่าแอนติบอดีที่ใช้นั้นทำงานได้อย่างมีประสิทธิภาพ

Negative control: ย่อยเซลล์เพาะเลี้ยงชนิดหรือเนื้อเยื่อตัวอย่างที่ทราบว่าไม่มีการแสดงออกของโปรตีนที่เราต้องการตรวจวัด ใช้เพื่อตรวจสอบว่าแอนติบอดีที่ใช้มีความจำเพาะ เป็นตัวบอกถึงการจับอย่างไม่จำเพาะ และผล False positive ที่เกิดขึ้นได้ในปฏิกิริยา

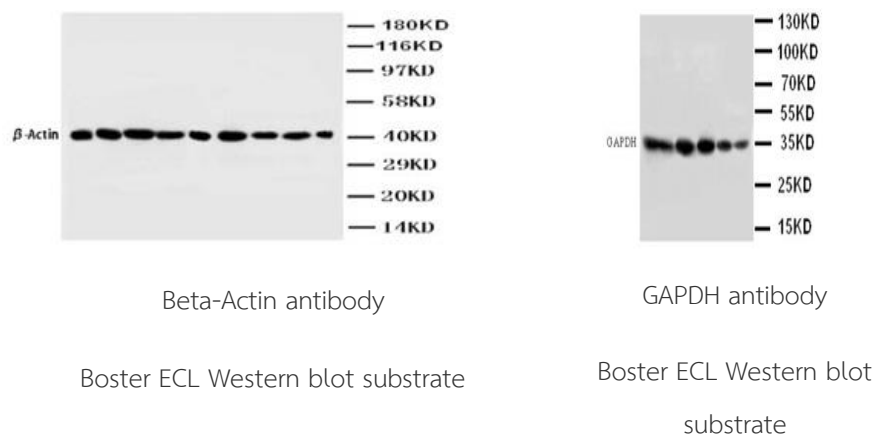
Secondary antibody control: เต็มแค่ Secondary antibody ลงไป ไม่ได้มีการเติม Primary antibody ดังนั้นใช้เพื่อตรวจสอบว่า Secondary antibody ที่ใช้มีความจำเพาะ เป็นตัวบอกถึงการจับอย่างไม่จำเพาะ และผล False positive ที่เกิดขึ้นได้ในปฏิกิริยา

Blank control: ไม่มีการใส่แอนติบอดีทั้งสองลงไป ใน Membrane เพื่อตรวจสอบตัว Membrane เองและตรวจสอบผลการยับยั้ง

Internal control: มีความจำเป็นอย่างยิ่งในการใช้ตรวจสอบคุณภาพสิ่งตัวอย่างและระบบการทำงานของ Secondary antibody โดยเป็นแอนติบอดีต่อ “House-keeping proteins” หรือโปรตีนที่แสดงออกเท่ากันในทุกเซลล์และเนื้อเยื่อ ใช้เพื่อตรวจสอบว่าได้มีการใส่ตัวอย่างลงไป ใน แผลวหรือไม่ โดยเฉพาะอย่างยิ่งในการทดสอบเพื่อเปรียบเทียบการแสดงออกของระดับโปรตีนในสิ่ง ตัวอย่างที่แตกต่างกัน

ตารางที่ 4 ขนาดมวลโมเลกุลและชนิดของสิ่งตัวอย่างสำหรับ Internal control ที่นิยมใช้ในการ ทำ Western blot

| Internal control | Molecular weight (kDa) | Sample type |
|------------------|------------------------|---------------------------------------------------------------------------|
| Beta-Actin | 43 | Whole Cell/cytoplasmic |
| GAPDH | 30-40 | Whole Cell/cytoplasmic |
| Tubulin | 55 | Whole Cell/cytoplasmic |
| VCDA1/Porin | 31 | Mitochondrial |
| COXIV | 16 | Mitochondrial |
| Lamin B1 | 16 | Nuclear (Not suitable for samples where the nuclear envelope is removed.) |
| TBP | 38 | Nuclear (Not suitable for samples where DNA is removed) |



รูปที่ 19 มวลโมเลกุลของโปรตีน Beta-actin และ GAPDH

แสดงขนาดมวลโมเลกุลของโปรตีน Beta-actin และ GAPDH ซึ่งเป็น
Internal control ที่นิยมในการทำ Western blot (www.thermofisher.com)

ตอนที่ 4.4 การวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงออกของโปรตีนกับคะแนนจาก

แบบทดสอบ ADI-R

วัตถุประสงค์: เพื่อวิเคราะห์สหสัมพันธ์ (Correlation analysis) ระหว่างระดับการแสดงออกของโปรตีนกับคะแนนจากแบบทดสอบ ADI-R

วิธีการวิจัย

ในการวิเคราะห์สหสัมพันธ์ (Correlation analysis) ผู้วิจัยจะใช้วิธี Pearson correlation analysis ($P < 0.05$) ในการวิเคราะห์เปรียบเทียบระดับการแสดงออกของโปรตีนกับคะแนนจากแบบทดสอบ ADI-R จำนวน 123 ข้อ ของผู้ป่วยออทิซึมสเปกตรัม โดยใช้โปรแกรม IBM SPSS Statistics version 22 และสร้าง Correlation heatmaps ด้วยโปรแกรม MeV และ Galaxy version 20.09.rc1

บทที่ 4

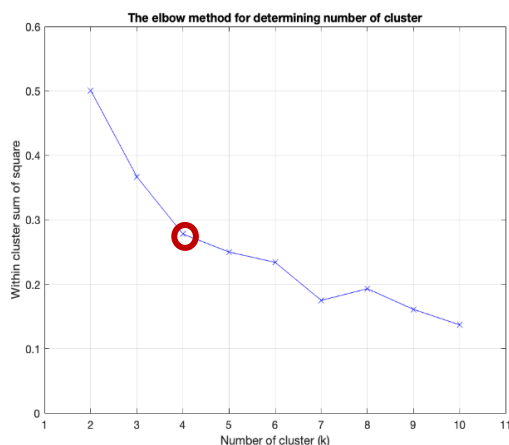
ผลการศึกษาวิจัย

4.1 ผลการวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม

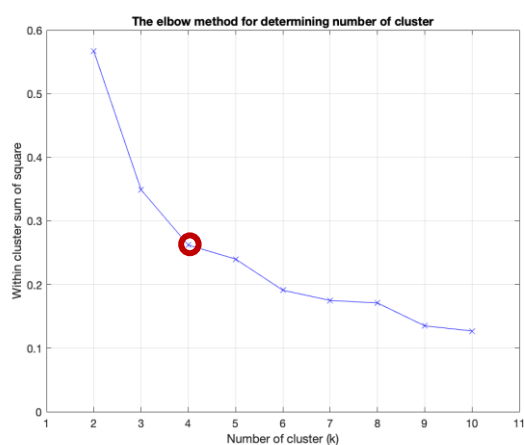
RapidMiner

4.1.1 ผลการหาจำนวนกลุ่มย่อยที่เหมาะสมด้วยวิธี Elbow method

ข้อมูลลักษณะอาการทางคลินิกจากแบบทดสอบ ADI-R จำนวน 123 ข้อ ของผู้ป่วยออทิซึมสเปกตรัมจำนวน 85 ราย จากฐานข้อมูล Autism Genetic Resource Exchange (AGRE) repository ที่ผ่านวิธีคัดเลือกตามเกณฑ์ของ Dr. Valerie Hu และคณะ ได้ถูกนำมาวิเคราะห์โดยใช้การเรียนรู้ของเครื่องมือแบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม RapidMiner โดยใช้อัลกอริทึม K-means และ K-medoids ซึ่งผู้วิจัยได้ทำการเปรียบเทียบหาจำนวนกลุ่มที่เหมาะสมในการแบ่งกลุ่มย่อย (Clustering) ของแต่ละอัลกอริทึมด้วยวิธี Elbow method พบว่าจากกราฟ The elbow method for determining number of cluster นั้น ทั้งวิธี K-means และ K-medoids นั้น มีจำนวนกลุ่มย่อยที่เหมาะสม คือ 4 กลุ่มย่อย ดังรูปที่ 19



K-means (Within cluster sum of square)

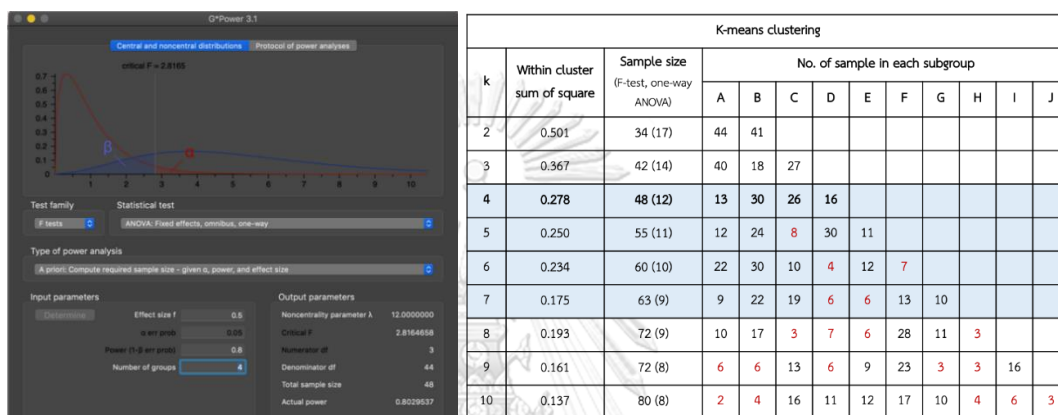


K-medoids (Within cluster sum of square)

รูปที่ 20 The elbow method for determining number of cluster

แสดงภาพวิธี Elbow method ในการหาจำนวนกลุ่มที่เหมาะสมในการ Clustering ของอัลกอริทึม K-means และ K-medoids โดยแกน x คือจำนวนกลุ่ม และแกน y คือค่า Within cluster sum of square

และจากการคำนวณจำนวนกลุ่มตัวอย่างที่เหมาะสมดังรูปที่ 20 ด้วยโปรแกรม G*Power 3.1 พบว่า เมื่อใช้สถิติในการทดสอบแบบ F-test, One-way ANOVA ที่ $P\text{-value} < 0.05$ จำนวนกลุ่มตัวอย่างสำหรับ 4 กลุ่มย่อย ที่สามารถใช้คำนวณค่าทางสถิติในการวิเคราะห์ได้จะอยู่ที่จำนวนรวมอย่างน้อย 48 ราย เฉลี่ยคิดเป็นกลุ่มย่อยละ 12 รายเป็นอย่างน้อย ซึ่งเมื่อเทียบกับจำนวนตัวอย่างที่ถูกแบ่งโดยวิธี K-means หรือ K-medoids แล้ว การแบ่งกลุ่มย่อยออกเป็น 4 กลุ่ม ยังคงให้จำนวนกลุ่มตัวอย่างที่มากพอในการคำนวณค่าทางสถิติได้



Total sample size = 48 (n=4, at least 12 samples/subgroup)

รูปที่ 21 Sample size calculation and comparison

แสดงภาพการคำนวณขนาดกลุ่มตัวอย่างด้วยโปรแกรม G*power 3.1 (ซ้าย) และภาพของตารางการเปรียบเทียบขนาดกลุ่มตัวอย่างเมื่อมีจำนวนกลุ่มแตกต่างกัน

4.1.2 ผลการเปรียบเทียบประสิทธิภาพในการแบ่งกลุ่มระหว่าง Unsupervised machine learning tools วิธี K-means และ K-medoids

หลังจากได้จำนวนกลุ่มย่อยที่เหมาะสมในการแบ่งผู้ป่วยแล้ว ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพในการแบ่งกลุ่ม (Clustering) ที่จำนวนเท่ากัน โดยใช้ผลลัพธ์จากการประเมินประสิทธิภาพในการแบ่งกลุ่มที่ได้จากโปรแกรม RapidMiner ได้แก่ Average within centroid distance, Average within cluster distance, and Davies Bouldin พบว่าจากตารางที่ 5 วิธี K-means clustering ให้ประสิทธิภาพในการจัดกลุ่มที่ดีกว่าวิธี K-medoids คือมีค่า Average within centroid distance มากกว่า บอกถึงค่าเฉลี่ยของแต่ละตัวอย่างจากจุดศูนย์กลางของกลุ่ม แสดงถึงความหนาแน่นภายในกลุ่มที่มากกว่า มีค่า Average within cluster distance โดยรวมที่น้อยกว่า บอกถึงการกระจายตัวภายในกลุ่มที่น้อยกว่า มีการกระจุกตัวของข้อมูลภายในกลุ่มที่มากกว่า และมี

ค่า Davies Bouldin ที่น้อยกว่า บอถึงการกระจายตัวภายในกลุ่มน้อยและมีระยะห่างระหว่างกลุ่มมากของวิธี K-means เมื่อเทียบกับวิธี K-medoids

ตารางที่ 5 ผลลัพธ์ประสิทธิภาพในการแบ่งกลุ่มด้วยวิธี K-means และ K-medoids เมื่อ k=4

| K-means clustering (k=4) | | | K-medoids clustering (k=4) | | |
|--------------------------|-------------------------------|------------------------------|----------------------------|-------------------------------|------------------------------|
| Performance | Avg. within centroid distance | Avg. within cluster distance | Performance | Avg. within centroid distance | Avg. within cluster distance |
| all | -80.087 | -291.174 | all | -151.965 | -278.198 |
| Cluster_0 (gr.A) | -61.621 | -136.271 | Cluster_0 (gr.A) | -144.450 | -256.822 |
| Cluster_1 (gr.B) | -76.308 | -372.662 | Cluster_1 (gr.B) | -186.789 | -258.704 |
| Cluster_2 (gr.C) | -94.208 | -332.762 | Cluster_2 (gr.C) | -141.897 | -353.507 |
| Cluster_3 (gr.D) | -80.348 | -194.170 | Cluster_3 (gr.D) | -139.059 | -196.666 |
| Davies Bouldin | -1.960 | | Davies Bouldin | -1.664 | |

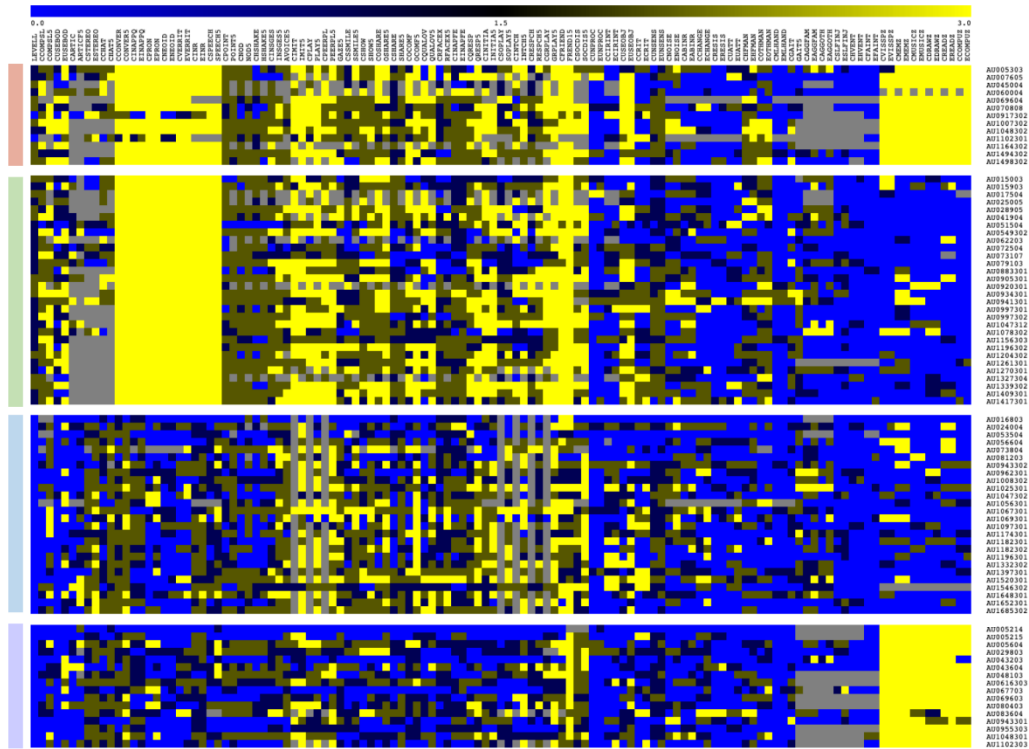
4.1.3 ผลการจัดกลุ่มผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อย และผลเปรียบเทียบลักษณะอาการทางคลินิกของผู้ป่วย

เมื่อได้อัลกอริทึมและจำนวนกลุ่มย่อยที่เหมาะสม ได้แก่ วิธี K-means clustering จำนวน 4 กลุ่มย่อยในการแบ่งผู้ป่วยออทิซึมสเปกตรัม จำนวน 85 ราย ดังตารางที่ 6 โดยกลุ่มย่อย A มีจำนวนตัวอย่าง 13 ราย กลุ่มย่อย B มีจำนวนตัวอย่าง 30 ราย กลุ่มย่อย C มีจำนวนตัวอย่าง 26 ราย และกลุ่มย่อย D จำนวน 16 ราย

ตารางที่ 6 ผลการแบ่งกลุ่มผู้ป่วยอหิวาต์ชิเมสเปกตรัมออกเป็น 4 กลุ่มย่อยด้วยวิธี K-means

| K-means clustering (k=4) | | | |
|--------------------------|----------------|----------------|----------------|
| Group A (n=13) | Group B (n=30) | Group C (n=26) | Group D (n=16) |
| HI1102 | HI0591 | HI1492 | HI1279 |
| HI0217 | HI0928 | HI1364 | HI1276 |
| HI0570 | HI0266 | HI1495 | HI0613 |
| HI0927 | HI0050 | HI1234 | HI0624 |
| HI1428 | HI0110 | HI2251 | HI0652 |
| HI1911 | HI0649 | HI2028 | HI0700 |
| HI2530 | HI0792 | HI2576 | HI1555 |
| HI2706 | HI2437 | HI2609 | HI2597 |
| HI2677 | HI1943 | HI4461 | HI2008 |
| HI2815 | HI1657 | HI3379 | HI1429 |
| HI2883 | HI1367 | HI2680 | HI2039 |
| HI4196 | HI1861 | HI4420 | HI2163 |
| HI4185 | HI3323 | HI3424 | HI2575 |
| | HI2427 | HI2828 | HI2791 |
| | HI2402 | HI2905 | HI2679 |
| | HI2536 | HI2853 | HI2824 |
| | HI2532 | HI2835 | |
| | HI2781 | HI2846 | |
| | HI2779 | HI4870 | |
| | HI2683 | HI3470 | |
| | HI2687 | HI3660 | |
| | HI4004 | HI4002 | |
| | HI3522 | HI4341 | |
| | HI2983 | HI4751 | |
| | HI3118 | HI4820 | |
| | HI3290 | HI4838 | |
| | HI3303 | | |
| | HI3369 | | |
| | HI3599 | | |
| | HI3673 | | |

ซึ่งผู้วิจัยได้ทำการเปรียบเทียบลักษณะอาการทางคลินิกของแต่ละกลุ่มย่อย พบว่าผู้ป่วยในแต่ละกลุ่มย่อยมีลักษณะอาการทางคลินิกที่มีรูปแบบคล้ายคลึงกันภายในกลุ่ม แต่แตกต่างกันระหว่างกลุ่ม ดังรูปที่ 21



รูปที่ 22 Heatmaps of ADI-R scores

แสดงภาพลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัมทั้ง 85 ราย โดย Column คือ ADI-R items และ Row คือคะแนนของผู้ป่วยแต่ละราย แถบสีส้มคือกลุ่มย่อย A (n=13) แถบสีเขียวคือกลุ่มย่อย B (n=30) แถบสีฟ้าคือกลุ่มย่อย C (n=26) และแถบสีม่วงคือกลุ่มย่อย D (n=16)

นอกจากนี้ผู้วิจัยได้เปรียบเทียบผลการแบ่งกลุ่มย่อยในข้างต้นกับข้อมูลในการแบ่งกลุ่ม ของ Dr. Hu และ Dr. Steinberg จากการศึกษาในปี 2009 (Hu & Steinberg, 2009) ซึ่งใช้วิธี K-means clustering, Principle Components Analysis, and Hierarchical clustering ร่วมกัน พบว่ามีผลลัพธ์ในการแบ่งกลุ่มที่ตรงกันทั้งจำนวนกลุ่มย่อย และรายชื่อกลุ่มตัวอย่างภายในกลุ่ม โดยกลุ่มย่อย A (n=13) เทียบได้กับกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษร่วมกับความผิดปกติด้านภาษารุนแรง (Savant skills with severe language impairments subgroup; SL subgroup) กลุ่มย่อย B (n=30) เทียบได้กับกลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup; L subgroup) กลุ่มย่อย C (n=26) เทียบได้กับกลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษา

รุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) และกลุ่มย่อย D (n=16) เทียบได้กับกลุ่มย่อยที่มีทักษะอัจฉริยะภาพพิเศษ (Savant skills subgroup; S subgroup)

4.2 ผลการวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner

หลังจากที่ผู้วิจัยได้ทำการแบ่งผู้ป่วยออทิซึมสเปกตรัมทั้ง 85 ราย ออกเป็นกลุ่มย่อย 4 กลุ่ม ด้วยวิธี K-means แล้ว เพื่อเปรียบเทียบประสิทธิภาพในการวิเคราะห์ข้อมูลลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม โดยการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner ผู้วิจัยได้ทำการติดฉลากคนไข้ตามที่แบ่งกลุ่มย่อยดังกล่าว

4.1.3 และใช้กระบวนการวิเคราะห์ 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection ในแต่ละอัลกอริทึมเพื่อสร้างโมเดลในการแบ่งคนไข้ออกเป็นกลุ่มย่อย โดยมีผลลัพธ์ของแต่ละอัลกอริทึมดังต่อไปนี้

4.2.1 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Artificial Neural Network (ANN)

สำหรับอัลกอริทึม Artificial Neural Network (ANN) ผู้วิจัยได้ทำการเปรียบเทียบการปรับค่าพารามิเตอร์ของ Training cycles ในกระบวนการวิเคราะห์แบบ Split validation ที่ 25 50 100 150 200 500 และ 1000 Training cycles ดังตารางที่ 7 ส่วนกระบวนการวิเคราะห์แบบ Backward elimination ได้ทำการเปรียบเทียบการปรับค่าพารามิเตอร์ของ Training cycles ที่ 25 50 75 100 และ 200 Training cycles ดังตารางที่ 8 และได้ทำการเปรียบเทียบการปรับค่าพารามิเตอร์ของ Training cycles ในกระบวนการวิเคราะห์แบบ Forward selection ที่ 25 50 100 500 1000 และ 2000 Training cycles ดังตารางที่ 9

ตารางที่ 7 ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 100 150 200 500 และ 1000 ในกระบวนการวิเคราะห์แบบ Split validation

| | ANN25 | apply_ANN25 | ANN50 | apply_ANN50 | ANN100 | apply_ANN100 | ANN150 | apply_ANN150 | ANN200 | apply_ANN200 | ANN500 | apply_ANN500 | ANN1000 | apply_ANN1000 |
|-----------------------------|---------------------|------------------|---------------------|------------------|---------------------|------------------|---------------------|------------------|---------------------|------------------|----------------------|------------------|----------------------|------------------|
| accuracy | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% |
| classification error | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% |
| kappa | 0.893 | 1 | 0.893 | 1 | 0.893 | 1 | 0.893 | 1 | 0.893 | 1 | 0.893 | 1 | 0.893 | 1 |
| weighted mean recall | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% |
| weighted mean precision | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% |
| spearman rho | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 |
| kendall tau | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 |
| absolute error | 0.127 ± 0.198 | 0.037 ± 0.016 | 0.110 ± 0.209 | 0.024 ± 0.011 | 0.100 ± 0.218 | 0.016 ± 0.007 | 0.096 ± 0.222 | 0.013 ± 0.006 | 0.093 ± 0.224 | 0.011 ± 0.005 | 0.087 ± 0.231 | 0.006 ± 0.003 | 0.084 ± 0.235 | 0.005 ± 0.002 |
| relative error | 12.65% ± 19.78% | 3.69% ± 1.62% | 11.04% ± 20.94% | 2.39% ± 1.06% | 10.01% ± 21.77% | 1.59% ± 0.71% | 9.56% ± 22.16% | 1.26% ± 0.56% | 9.30% ± 22.41% | 1.07% ± 0.48% | 8.68% ± 23.09% | 0.65% ± 0.29% | 8.37% ± 23.51% | 0.50% ± 0.22% |
| relative error strict | 43.59% ± 143.76% | 3.86% ± 1.80% | 54.73% ± 204.34% | 2.46% ± 1.13% | 69.05% ± 276.41% | 1.62% ± 0.74% | 79.22% ± 326.16% | 1.28% ± 0.58% | 87.39% ± 365.70% | 1.09% ± 0.49% | 119.87% ± 521.42% | 0.66% ± 0.30% | 152.77% ± 678.42% | 0.50% ± 0.23% |
| normalized absolute error | 0.164 | 0.047 | 0.143 | 0.047 | 0.13 | 0.02 | 0.124 | 0.016 | 0.121 | 0.014 | 0.113 | 0.008 | 0.109 | 0.006 |
| root mean squared error | 0.235 ± 0.000 | 0.040 ± 0.000 | 0.237 ± 0.000 | 0.040 ± 0.000 | 0.240 ± 0.000 | 0.017 ± 0.000 | 0.241 ± 0.000 | 0.014 ± 0.000 | 0.243 ± 0.000 | 0.012 ± 0.000 | 0.247 ± 0.000 | 0.007 ± 0.000 | 0.250 ± 0.000 | 0.005 ± 0.000 |
| root relative squared error | 0.305 | 0.052 | 0.308 | 0.034 | 0.312 | 0.022 | 0.314 | 0.018 | 0.315 | 0.015 | 0.321 | 0.009 | 0.324 | 0.007 |
| squared error | 0.055 ± 0.166 | 0.002 ± 0.002 | 0.055 ± 0.166 | 0.001 ± 0.001 | 0.057 ± 0.192 | 0.000 ± 0.000 | 0.058 ± 0.197 | 0.000 ± 0.000 | 0.059 ± 0.200 | 0.000 ± 0.000 | 0.061 ± 0.209 | 0.000 ± 0.000 | 0.062 ± 0.216 | 0.000 ± 0.000 |
| correlation | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 |
| squared correlation | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 |
| cross-entropy | 0.271 | 0.054 | 0.266 | 0.035 | 0.268 | 0.023 | 0.271 | 0.018 | 0.274 | 0.015 | 0.288 | 0.009 | 0.301 | 0.007 |

ตารางที่ 8 ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 75 100 และ 200 ในกระบวนการวิเคราะห์

แบบ Backward elimination

| | ANN25 | apply_ANN25 | ANN50 | apply_ANN50 | ANN75 | apply_ANN75 | ANN100 | apply_ANN100 | ANN200 | apply_ANN200 |
|-----------------------------|-------------------|------------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|-------------------|-----------------|
| weight attribute | 122 | 56.42 m | 122 | | 122 | | 122 | 1.10.50 hr | 122 | 2.00.03 |
| accuracy | 98.82% | 100% | 98.82% | 100% | 98.82% | 100% | 98.82% | 100% | 98.82% | 0% |
| classification error | 1.18% | 0% | 1.18% | 0% | 1.18% | 0% | 1.18% | 0% | 1.18% | 1 |
| kappa | 0.984 | 1 | 0.984 | 1 | 0.984 | 1 | 0.984 | 1 | 0.984 | 100% |
| weighted mean recall | 98.08% | 100% | 98.08% | 100% | 98.08% | 100% | 98.08% | 100% | 98.08% | 100% |
| weighted mean precision | 98.53% | 100% | 98.53% | 100% | 98.53% | 100% | 98.53% | 100% | 98.53% | 1 |
| spearman rho | 9.981 | 1 | 9.981 | 1 | 9.981 | 1 | 9.981 | 1 | 9.981 | 1 |
| kendall tau | 9.956 | 1 | 9.956 | 1 | 9.956 | 1 | 9.956 | 1 | 9.956 | 0.023 +/- 0.010 |
| absolute error | 0.163 +/- 0.108 | 0.110 +/- 0.045 | 0.115 +/- 0.111 | 0.061 +/- 0.026 | 0.098 +/- 0.112 | 0.045 +/- 0.019 | 0.088 +/- 0.113 | 0.037 +/- 0.016 | 0.072 +/- 0.115 | 2.34% +/- 1.03% |
| relative error | 16.31% +/- 10.84% | 11.05% +/- 4.54% | 11.46% +/- 11.09% | 6.10% +/- 2.59% | 9.76% +/- 11.24% | 4.49% +/- 1.94% | 8.84% +/- 11.34% | 3.67% +/- 1.60% | 7.23% +/- 11.52% | 2.34% +/- 1.03% |
| relative error strict | 24.03% +/- 40.33% | 12.74% +/- 6.32% | 17.58% +/- 42.53% | 6.58% +/- 3.08% | 15.75% +/- 45.60% | 4.75% +/- 2.20% | 14.89% +/- 48.24% | 3.84% +/- 1.77% | 13.74% +/- 55.55% | 0.03 |
| normalized absolute error | 0.21 | 0.142 | 0.148 | 0.079 | 0.126 | 0.058 | 0.114 | 0.047 | 0.093 | 0.026 +/- 0.000 |
| root mean squared error | 0.196 +/- 0.000 | 0.119 +/- 0.000 | 0.159 +/- 0.000 | 0.066 +/- 0.000 | 0.149 +/- 0.000 | 0.049 +/- 0.000 | 0.144 +/- 0.000 | 0.040 +/- 0.000 | 0.144 +/- 0.000 | 0.033 |
| root relative squared error | 0.244 | 0.154 | 0.199 | 0.085 | 0.186 | 0.063 | 0.179 | 0.052 | 0.169 | 0.001 +/- 0.001 |
| squared error | 0.038 +/- 0.075 | 0.014 +/- 0.014 | 0.025 +/- 0.075 | 0.004 +/- 0.004 | 0.022 +/- 0.076 | 0.002 +/- 0.002 | 0.021 +/- 0.077 | 0.002 +/- 0.002 | 0.018 +/- 0.080 | 1 |
| correlation | 0.995 | 1 | 0.995 | 1 | 0.995 | 1 | 0.995 | 1 | 0.995 | 1 |
| squared correlation | 0.99 | 1 | 0.99 | 1 | 0.99 | 1 | 0.99 | 1 | 0.99 | 0.054 |
| cross-entropy | 0.247 | 0.169 | 0.174 | 0.09 | 0.15 | 0.066 | 0.138 | 0.054 | 0.116 | 0% |

ตารางที่ ๒ ประสิทธิภาพของโมเดล Artificial Neural Network (ANN) ที่ Training cycles เท่ากับ 25 50 100 500 1000 และ 2000 ในกระบวนการวิเคราะห์แบบ Forward selection

| | ANN25 | apply ANN25 | ANN50 | apply ANN50 | ANN100 | apply ANN100 | ANN500 | apply ANN500 | ANN1000 | apply ANN1000 |
|-----------------------------|--------------------|---------------------|-------------------|-------------------|-------------------|-------------------|------------------|------------------|---------------------|--------------------|
| weight attribute | 2 | | 4 | | 4 | | 3 | | 6 | 19.43 m |
| accuracy | 66.11% + 7.36% | 65.56% | 96.25% ± 6.04% | 96.67% | 97.50% +/- 5.27% | 97.78% | 96.25% ± 6.04% | 96.67% | 98.75% +/- 3.95% | 98.82% |
| classification error | 33.89% + 7.36% | 34.44% | 3.75% ± 6.04% | 3.33% | 2.50% +/- 5.27% | 2.22% | 3.75% ± 6.04% | 3.33% | 1.25% +/- 3.95% | 1.18% |
| kappa | 0.449 + 0.082 | 0.488 | 0.945 ± 0.089 | 0.954 | 0.962 +/- 0.080 | 0.969 | 0.945 ± 0.089 | 0.954 | 0.980 +/- 0.062 | 0.984 |
| weighted mean recall | 50.00% +0.00% | 50.00% | 88.75% ± 14.97% | 94.46% | 91.25% +/- 14.49% | 96.74% | 88.75% ± 14.97% | 94.46% | 93.75% +/- 13.50% | 98.44% |
| weighted mean precision | 34.38% +4.54% | 32.92% | 87.29% ± 16.68% | 96.36% | 90.62% +/- 15.38% | 97.67% | 87.29% ± 16.68% | 96.36% | 94.38% +/- 11.95% | 99.07% |
| spearman rho | 0.198 +/- 0.329 | 0.184 | 0.957 ± 0.126 | 0.950% | 0.959 +/- 0.126 | 0.957 | 0.957 ± 0.126 | 0.950% | 0.960 +/- 0.126 | 0.956 |
| kendall tau | 0.185 +/- 0.308 | 0.169 | 0.953 ± 0.125 | 0.950% | 0.958 +/- 0.126 | 0.957 | 0.953 ± 0.125 | 0.950% | 0.960 +/- 0.126 | 0.959 |
| absolute error | 0.590 +/- 0.016 | 0.568 +/- 0.152 | 0.332 +/- 0.018 | 0.298 +/- 0.112 | 0.188 +/- 0.030 | 0.165 +/- 0.096 | 0.165 ± 0.039 | 0.147 ± 0.149 | 0.073 +/- 0.041 | 0.049 +/- 0.102 |
| relative error | 58.98% +/- 1.58% | 56.79% +/- 15.21% | 33.25% +/- 1.83% | 29.76% +/- 11.20% | 18.75% +/- 3.03% | 16.45% +/- 9.56% | 16.54% ± 3.87% | 14.67% ± 14.93% | 7.32% +/- 4.08% | 4.90% +/- 10.22% |
| relative error strict | 199.53% +/- 25.09% | 179.15% +/- 149.54% | 60.15% +/- 16.50% | 49.86% +/- 51.53% | 33.98% +/- 28.00% | 23.31% +/- 33.59% | 82.37% ± 133.90% | 52.13% ± 229.26% | 214.19% +/- 651.43% | 50.26% +/- 424.20% |
| normalized absolute error | 0.817 +/- 0.260 | 0.73 | 0.459 +/- 0.139 | 0.383 | 0.256 +/- 0.071 | 0.212 | 0.227 ± 0.078 | 0.189 | 0.096 +/- 0.044 | 0.063 |
| root mean squared error | 0.609 +/- 0.018 | 0.588 +/- 0.000 | 0.353 +/- 0.030 | 0.318 +/- 0.000 | 0.215 +/- 0.062 | 0.190 +/- 0.000 | 0.216 ± 0.094 | 0.209 ± 0.000 | 0.118 +/- 0.098 | 0.113 +/- 0.000 |
| root relative squared error | 0.844 +/- 0.270 | 0.756 | 0.486 +/- 0.147 | 0.409 | 0.289 +/- 0.087 | 0.245 | 0.293 ± 0.131 | 0.269 | 0.153 +/- 0.112 | 0.146 |
| squared error | 0.371 +/- 0.022 | 0.346 +/- 0.185 | 0.125 +/- 0.022 | 0.101 +/- 0.098 | 0.050 +/- 0.031 | 0.036 +/- 0.071 | 0.055 ± 0.047 | 0.044 ± 0.146 | 0.023 +/- 0.038 | 0.013 +/- 0.102 |
| correlation | 0.154 +/- 0.284 | 0.05 | 0.951 ± 0.131 | 0.949 | 0.955 +/- 0.132 | 0.953 | 0.951 ± 0.131 | 0.949 | 0.958 +/- 0.132 | 0.955 |
| squared correlation | 0.097 +/- 0.235 | 0.003 | 0.920 ± 0.207 | 0.9 | 0.927 +/- 0.208 | 0.908 | 0.920 ± 0.207 | 0.9 | 0.934 +/- 0.209 | 0.911 |
| cross-entropy | 1.265 +/- 0.077 | 1.313 | 0.553 +/- 0.064 | 0.531 | 0.297 +/- 0.097 | 0.272 | 0.318 ± 0.190 | 0.295 | 0.169 +/- 0.245 | 0.117 |

จากตารางทั้ง 3 ตารางในข้างต้น ผู้วิจัยได้ใช้ค่า Accuracy เป็นหลัก ในการพิจารณาเลือกโมเดลที่มีประสิทธิภาพดีที่สุดในกระบวนการวิเคราะห์ที่ต่างกันทั้ง 3 แบบ พบว่ากระบวนการวิเคราะห์แบบ Split validation ที่ 25 Training cycles ให้ค่า Accuracy ของโมเดลเท่ากับ 92.31% (จาก Testing data 30%) โดยใช้ข้อความ 123 ข้อ, กระบวนการวิเคราะห์แบบ Backward elimination ที่ 25 Training cycles ให้ค่า Accuracy ของโมเดลเท่ากับ 98.82-100% โดยตัดข้อความ LEVELL ออก เหลือข้อความที่ใช้วิเคราะห์ 122 ข้อ, และกระบวนการวิเคราะห์แบบ Forward selection ที่ 1000 Training cycles ให้ค่า Accuracy ของโมเดลเท่ากับ 98.82% ใช้ข้อความเพียง 6 ข้อ (CCONVER, CINR, CONVER5, CSPEECH, ECOMPUZ, and EINAPPO) โดยที่ กระบวนการวิเคราะห์แบบ Backward elimination ใช้เวลาในการวิเคราะห์นานที่สุดดังแสดงในตารางที่ 10



ตารางที่ 10 เปรียบเทียบประสิทธิภาพของโมเดล Artificial Neural Network (ANN) จากทั้ง 3 กระบวนการวิเคราะห์

| Subgroups | Parameter | Algorithms: ANN | | | | |
|-------------|--------------------|---------------------------|------|-------------|------|---------------|
| | | Split validation (25 TCs) | | BE (25 TCs) | | FS (1000 TCs) |
| | No. of ADI-R items | 123 | | 122 | | 6 |
| | time | 1 s | | 56.42 m | | 19.43 m |
| | accuracy | 92.31% | 100% | 98.82% | 100% | 98.82% |
| A (n=13) | Precision | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 50% | 100% | 92.3% | 100% | 100% |
| | Specificity | 100% | 100% | 98.61% | 100% | 100% |
| | Recall | 50% | 100% | 92.31% | 100% | 100% |
| | F-score | 66.67% | 100% | 96% | 100% | 100% |
| B (n=30) | Precision | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 100% | 100% | 100% |
| C (n=26) | Precision | 100% | 100% | 100% | 100% | 96.3% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 98.3 |
| | Recall | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 100% | 100% | 98.12% |
| D (n=16) | Precision | 71.43% | 100% | 94.12% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 93.75% |
| | Specificity | 90.48% | 100% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 93.75% |
| | F-score | 83.33% | 100% | 96.97% | 100% | 96.77% |

4.2.2 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม k-Nearest Neighbor (k-NN)

สำหรับอัลกอริทึม k-Nearest Neighbor (k-NN) ผู้วิจัยได้ทำการเปรียบเทียบการปรับค่าพารามิเตอร์ของ ค่า k ในกระบวนการวิเคราะห์ทั้ง 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection ที่ 3 5 9 15 25 45 และ 85 ดังตารางที่ **11-13** ตามลำดับ



ตารางที่ 11 ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Split validation

| | k-NN(k=3) | apply_k-NN(k=3) | k-NN(k=5) | apply_k-NN(k=5) | k-NN(k=9) | apply_k-NN(k=9) | k-NN(k=15) | apply_k-NN(k=15) | k-NN(k=25) | apply_k-NN(k=25) | k-NN(k=45) | apply_k-NN(k=45) | k-NN(k=85) | apply_k-NN(k=85) |
|-----------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|------------------|-----------------|------------------|-------------------|------------------|-------------------|-------------------|
| accuracy | 96.15% | 100% | 100% | 98% | 96.15% | 98% | 92.31% | 97.65% | 88.46% | 95.29% | 50.00% | 67.06% | 34.62% | 35.29% |
| classification error | 3.85% | 0% | 0% | 2.35% | 3.85% | 2% | 7.69% | 2.35% | 11.54% | 4.71% | 50.00% | 32.94% | 65.38% | 64.71% |
| kappa | 0.947 | 1 | 1 | 0.967 | 0.947 | 0.967 | 0.894 | 0.967 | 0.838 | 0.935 | 0.251 | 0.511 | 0 | 0 |
| weighted mean recall | 93.75% | 100% | 100% | 98.08% | 95.00% | 98.08% | 90.00% | 98.08% | 85.62% | 94.59% | 37.50% | 52.76% | 25.00% | 25.00% |
| weighted mean precision | 95.83% | 100% | 100% | 98% | 95.00% | 98% | 91.67% | 98.44% | 92.33% | 95.80% | 24.35% | 57.83% | 8.65% | 8.82% |
| spearman rho | 0.993 | 1 | 1 | 0.984 | 0.993 | 0.984 | 0.988 | 0.984 | 0.805 | 0.976 | -0.185 | 0.285 | 0 | 0 |
| kendall tau | 0.984 | 1 | 1 | 0.979 | 0.984 | 0.979 | 0.971 | 0.979 | 0.799 | 0.966 | -0.17 | 0.273 | 0 | 0 |
| absolute error | 0.088 ± 0.172 | 0.029 ± 0.104 | 0.100 ± 0.149 | 0.063 ± 0.142 | 0.128 ± 0.162 | 0.091 ± 0.147 | 0.258 ± 0.191 | 0.163 ± 0.157 | 0.434 ± 0.149 | 0.304 ± 0.172 | 0.644 ± 0.103 | 0.519 ± 0.135 | 0.809 ± 0.055 | 0.722 ± 0.078 |
| relative error | 8.82% ± 17.18% | 2.91% ± 10.37% | 9.95% ± 14.86% | 6.34% ± 14.19% | 12.78% ± 16.17% | 9.10% ± 14.68% | 25.81% ± 19.11% | 16.32% ± 15.71% | 43.44% ± 14.92% | 30.41% ± 17.18% | 64.37% ± 10.28% | 51.93% ± 13.53% | 80.87% ± 5.51% | 72.24% ± 7.84% |
| relative error strict | 16.80% ± 40.25% | 5.04% ± 19.39% | 14.95% ± 23.82% | 12.11% ± 38.78% | 20.20% ± 30.77% | 15.11% ± 33.63% | 47.29% ± 50.96% | 25.57% ± 34.33% | 90.89% ± 54.45% | 54.04% ± 43.14% | 210.55% ± 105.68% | 128.70% ± 75.16% | 480.72% ± 204.53% | 298.97% ± 139.10% |
| normalized absolute error | 0.115 | 0.037 | 0.129 | 0.082 | 0.166 | 0.117 | 0.335 | 0.21 | 0.565 | 0.392 | 0.837 | 0.669 | 1.051 | 0.93 |
| root mean squared error | 0.193 ± 0.000 | 0.108 ± 0.000 | 0.179 ± 0.000 | 0.155 ± 0.000 | 0.206 ± 0.000 | 0.173 ± 0.000 | 0.321 ± 0.000 | 0.227 ± 0.000 | 0.459 ± 0.000 | 0.349 ± 0.000 | 0.652 ± 0.000 | 0.537 ± 0.000 | 0.811 ± 0.000 | 0.727 ± 0.000 |
| root relative squared error | 0.251 | 0.139 | 0.233 | 0.2 | 0.268 | 0.223 | 0.417 | 0.292 | 0.597 | 0.450 | 0.847 | 0.691 | 1.054 | 0.936 |
| squared error | 0.037 ± 0.090 | 0.012 ± 0.047 | 0.032 ± 0.056 | 0.024 ± 0.080 | 0.042 ± 0.076 | 0.030 ± 0.077 | 0.103 ± 0.124 | 0.051 ± 0.082 | 0.211 ± 0.132 | 0.122 ± 0.112 | 0.425 ± 0.136 | 0.288 ± 0.146 | 0.657 ± 0.091 | 0.528 ± 0.117 |
| correlation | 0.986 | 1 | 1 | 0.99 | 0.985 | 0.99 | 0.971 | 0.99 | 0.816 | 0.98 | 0 | 0.264 | 0 | 0 |
| squared correlation | 0.972 | 1 | 1 | 0.98 | 0.969 | 0.98 | 0.943 | 0.98 | 0.666 | 0.96 | 0 | 0.07 | 0 | 0 |
| cross-entropy | 0.163 | 0.054 | 0.168 | 0.119 | 0.219 | 0.162 | 0.47 | 0.285 | 0.844 | 0.564 | 1.5 | 1.109 | 2.366 | 1.895 |

ตารางที่ 12 ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Backward elimination

| | k-NN(k=3) | apply_k-NN(k=3) | k-NN(k=5) | apply_k-NN(k=5) | k-NN(k=9) | apply_k-NN(k=9) | k-NN(k=15) | apply_k-NN(k=15) | k-NN(k=25) | apply_k-NN(k=25) | k-NN(k=45) | apply_k-NN(k=45) | k-NN(k=85) | apply_k-NN(k=85) |
|-----------------------------|-----------|-----------------|------------|-----------------|------------|-----------------|------------|------------------|------------|------------------|-------------|------------------|-------------|------------------|
| accuracy | 122 | 33 s | 122 | 97.65% | 122 | 97.65% | 122 | 97.65% | 122 | 95.29% | 122 | 64.71% | 122 | 35.29% |
| classification error | 97.65% | 98.82% | 97.65% | 2.35% | 97.65% | 2.35% | 97.65% | 2.35% | 94.12% | 4.71% | 62.35% | 35.29% | 64.71% | 64.71% |
| kappa | 0.967 | 0.984 | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 | 0.967 | 0.918 | 0.935 | 0.437 | 0.474 | 0 | 0 |
| weighted mean recall | 98.33% | 99.04% | 98.08% | 98.08% | 98.08% | 98.08% | 98.08% | 98.08% | 92.67% | 94.59% | 47.12% | 49.64% | 25.00% | 25.00% |
| weighted mean precision | 98.75% | 99.19% | 98.44% | 98.44% | 98.44% | 98.44% | 98.44% | 98.44% | 94.98% | 95.80% | 31.04% | 57.05% | 8.82% | 8.82% |
| spearman rho | 0.998 | 0.992 | 0.981 | 0.984 | 0.981 | 0.981 | 0.984 | 0.981 | 0.976 | 0.976 | 1.455 | 0.186 | 0 | 0 |
| kendall tau | 0.995 | 0.989 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.979 | 0.957 | 0.966 | 1.368 | 0.175 | 0 | 0 |
| absolute error | 0.058 +/- | 0.029 +/- | 0.073 +/- | 0.059 +/- | 0.112 +/- | 0.094 +/- | 0.173 +/- | 0.212 +/- | 0.350 +/- | 0.300 +/- | 0.563 +/- | 0.521 +/- | 0.752 +/- | 0.722 +/- |
| relative error | 0.170 | 0.104 | 0.153 | 0.141 | 0.158 | 0.148 | 0.160 | 0.178 | 0.169 | 0.171 | 0.130 | 0.136 | 0.072 | 0.078 |
| relative error | 5.85% +/- | 2.91% +/- | 7.26% +/- | 5.91% +/- | 11.20% +/- | 9.36% +/- | 17.26% +/- | 21.18% +/- | 35.02% +/- | 30.03% +/- | 56.29% +/- | 52.09% +/- | 75.23% +/- | 72.24% +/- |
| relative error strict | 17.03% | 10.37% | 15.26% | 14.10% | 15.82% | 14.81% | 15.99% | 17.79% | 16.92% | 17.11% | 13.00% | 13.61% | 7.24% | 7.84% |
| relative error strict | ∞% | 5.04% +/- | 14.98% +/- | 11.58% +/- | 19.46% +/- | 15.50% +/- | 26.89% +/- | 35.91% +/- | 66.28% +/- | 53.03% +/- | 154.34% +/- | 129.96% +/- | 351.11% +/- | 298.97% +/- |
| normalized absolute error | 49.78% | 19.39% | 49.78% | 38.79% | 44.17% | 33.76% | 32.64% | 43.74% | 49.45% | 42.66% | 87.91% | 76.65% | 164.61% | 139.10% |
| normalized absolute error | 0.075 | 0.037 | 0.094 | 0.076 | 0.144 | 0.121 | 0.222 | 0.273 | 0.451 | 0.387 | 0.725 | 0.671 | 0.969 | 0.93 |
| root mean squared error | 0.180 +/- | 0.108 +/- | 0.169 +/- | 0.153 +/- | 0.194 +/- | 0.175 +/- | 0.235 +/- | 0.277 +/- | 0.389 +/- | 0.346 +/- | 0.578 +/- | 0.538 +/- | 0.756 +/- | 0.727 +/- |
| root mean squared error | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| root relative squared error | 0.224 | 0.139 | 0.211 | 0.197 | 0.242 | 0.226 | 0.303 | 0.345 | 0.485 | 0.445 | 0.72 | 0.693 | 0.942 | 0.936 |
| squared error | 0.032 +/- | 0.012 +/- | 0.029 +/- | 0.023 +/- | 0.038 +/- | 0.031 +/- | 0.055 +/- | 0.077 +/- | 0.151 +/- | 0.119 +/- | 0.334 +/- | 0.290 +/- | 0.571 +/- | 0.528 +/- |
| squared error | 0.128 | 0.047 | 0.091 | 0.080 | 0.087 | 0.077 | 0.080 | 0.100 | 0.124 | 0.111 | 0.151 | 0.148 | 0.112 | 0.117 |
| correlation | 0.99 | 0.995 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.99 | 0.974 | 0.98 | 0 | 0.106 | 0 | 0 |
| squared correlation | 0.98 | 0.99 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.98 | 0.95 | 0.96 | 0 | 0.011 | 0 | 0 |
| cross-entropy | ∞ | 0.054 | 0.126 | 0.112 | 0.183 | 0.167 | 0.302 | 0.347 | 0.604 | 0.556 | 1.133 | 1.115 | 1.868 | 1.895 |

ตารางที่ 13 ประสิทธิภาพของโมเดล k-Nearest Neighbor (k-NN) ที่ k เท่ากับ 3 5 9 15 25 45 และ 85 ในกระบวนการวิเคราะห์แบบ Forward selection

| | k-NN(k=3) | apply_k- NN(k=3) | k-NN(k=5) | apply_k- NN(k=5) | k-NN(k=9) | apply_k- NN(k=9) | k-NN(k=15) | apply_k- NN(k=15) | k-NN(k=25) | apply_k- NN(k=25) | k-NN(k=45) | apply_k- NN(k=45) | k-NN(k=85) | apply_k- NN(k=85) |
|-----------------------------|-----------|---------------------|-----------|---------------------|-----------|---------------------|------------|----------------------|------------|----------------------|------------|----------------------|------------|----------------------|
| | 5 | | 7 | | 7 | | 8 | 43 s | 6 | | 3 | | 1 | |
| accuracy | 98.82% | 98.89% | 98.75% | 98.89% | 98.75% | 97.78% | 100% | 100% | 100% | 100% | 79.17% | 77.78% | 66.11% | 65.56% |
| classification error | 1.18% | 1.11% | 1.25% | 1.11% | 1.25% | 2.22% | 0% | 0% | 0% | 0% | 20.83% | 22.22% | 33.89% | 34.44% |
| kappa | 0.984 | 0.985 | 0.982 | 0.985 | 0.982 | 0.969 | 1 | 1 | 1 | 1 | 0.711 | 0.692 | 0.499 | 0.488 |
| weighted mean recall | 98.33% | 98.21% | 92.50% | 98.21% | 92.50% | 96.43% | 95% | 100% | 95% | 100% | 66.25% | 68.75% | 50.00% | 50.00% |
| weighted mean precision | 98.75% | 98.61% | 91.67% | 98.61% | 91.67% | 97.37% | 95% | 100% | 95% | 100% | 59.38% | 59.13% | 34.38% | 32.92% |
| spearman rho | 0.998 | 0.998 | 0.998 | 0.998 | 0.998 | 0.996 | 1 | 1 | 1 | 1 | 0.845 | 0.84 | 0.198 | 0.184 |
| kendall tau | 0.995 | 0.995 | 0.996 | 0.995 | 0.996 | 0.989 | 1 | 1 | 1 | 1 | 0.83 | 0.816 | 0.185 | 0.169 |
| absolute error | 0.029 +/- | 0.023 +/- | 0.076 +/- | 0.046 +/- | 0.079 +/- | 0.063 +/- | 0.114 +/- | 0.094 +/- | 0.258 +/- | 0.205 +/- | 0.576 +/- | 0.550 +/- | 0.681 +/- | 0.650 +/- |
| | 0.034 | 0.090 | 0.056 | 0.116 | 0.048 | 0.140 | 0.032 | 0.131 | 0.042 | 0.198 | 0.021 | 0.133 | 0.011 | 0.103 |
| relative error | 2.90% +/- | 2.28% +/- | 7.64% +/- | 4.65% +/- | 7.88% +/- | 6.26% +/- | 11.38% +/- | 9.39% +/- | 25.81% +/- | 20.55% +/- | 57.62% +/- | 54.98% +/- | 68.14% +/- | 64.99% +/- |
| | 3.36% | 9.00% | 5.61% | 11.57% | 4.85% | 14.03% | 3.18% | 13.07% | 4.24% | 19.83% | 2.06% | 13.32% | 1.10% | 10.26% |
| relative error strict | ∞% | 16.26% | ∞% | 33.43% | 13.85% | 39.63% | 4.56% | 21.46% | 9.31% | 37.78% | ∞% | 487.44% | 19.18% | 183.56% |
| | 0.038 | 0.029 | 0.101 | 0.06 | 0.122 | 0.081 | 0.163 | 0.121 | 0.367 | 0.264 | 0.799 | 0.707 | 0.944 | 0.836 |
| normalized absolute error | | | | | | | | | | | | | | |
| root mean squared error | 0.096 +/- | 0.093 +/- | 0.153 +/- | 0.125 +/- | 0.144 +/- | 0.154 +/- | 0.175 +/- | 0.161 +/- | 0.328 +/- | 0.286 +/- | 0.590 +/- | 0.566 +/- | 0.688 +/- | 0.658 +/- |
| | 0.107 | 0.000 | 0.097 | 0.000 | 0.085 | 0.000 | 0.029 | 0.000 | 0.036 | 0.000 | 0.022 | 0.000 | 0.012 | 0.000 |
| root relative squared error | 0.125 | 0.119 | 0.216 | 0.16 | 0.22 | 0.198 | 0.246 | 0.207 | 0.461 | 0.367 | 0.817 | 0.727 | 0.954 | 0.846 |
| squared error | 0.018 +/- | 0.009 +/- | 0.032 +/- | 0.016 +/- | 0.027 +/- | 0.024 +/- | 0.031 +/- | 0.026 +/- | 0.108 +/- | 0.082 +/- | 0.348 +/- | 0.320 +/- | 0.473 +/- | 0.433 +/- |
| | 0.027 | 0.039 | 0.037 | 0.064 | 0.025 | 0.078 | 0.010 | 0.051 | 0.023 | 0.095 | 0.026 | 0.163 | 0.016 | 0.142 |
| correlation | 0.996 | 0.995 | 0.996 | 0.995 | 0.996 | 0.991 | 1 | 1 | 1 | 1 | 0.816 | 0.807 | 0.154 | 0.05 |
| squared correlation | 0.992 | 0.991 | 0.993 | 0.991 | 0.993 | 0.981 | 1 | 1 | 1 | 1 | 0.714 | 0.65 | 0.097 | 0.003 |
| cross-entropy | ∞ | 0.041 | ∞ | 0.085 | 0.13 | 0.117 | 0.173 | 0.158 | 0.439 | 0.377 | ∞ | 1.248 | 1.55 | 1.582 |

จากตารางทั้ง 3 ตารางในข้างต้น ผู้วิจัยได้ใช้ค่า Accuracy เป็นหลัก ในการพิจารณาเลือกโมเดลที่มีประสิทธิภาพดีที่สุดในจำนวนการวิเคราะห์ที่ต่างกันทั้ง 3 แบบ พบว่ากระบวนการวิเคราะห์แบบ Split validation ที่ $k=5$ ให้ค่า Accuracy ของโมเดลเท่ากับ 100% (จาก Testing dataset 30%) โดยใช้ข้อความ 123 ข้อ กระบวนการวิเคราะห์แบบ Backward elimination ที่ $k=3$ ให้ค่า Accuracy ของโมเดลเท่ากับ 97.65-98.82% โดยตัดข้อความ LEVELL ออก เหลือข้อความที่ใช้วิเคราะห์ 122 ข้อ และกระบวนการวิเคราะห์แบบ Forward selection ที่ $k=15$ ให้ค่า Accuracy ของโมเดลเท่ากับ 100% ใช้ข้อความ 8 ข้อ (CIMIT, CMUSICZ, CSPEECH, ECOMPUZ, EINAPPQ, LEVELL, QUALOV5, and RFACEX5) โดยที่กระบวนการวิเคราะห์ทั้ง 3 แบบใช้เวลาไม่เกิน 1 นาที ดังแสดงในตาราง 14



4.2.3 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Naïve Bayes

สำหรับอัลกอริทึม Naïve Bayes ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของ Naïve Bayes และ Naïve Bayes kernel ในกระบวนการวิเคราะห์ทั้ง 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection แบบ Naïve Bayes และ Naïve Bayes kernel ดังตารางที่ 15-17 ตามลำดับ

ตารางที่ 15 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการวิเคราะห์แบบ Split validation

| | Naïve bayes | apply_Naïve bayes | Naïve bayes_kernel (10) | apply_Naïve bayes-kernel (10) |
|-----------------------------|-------------------|-------------------|-------------------------|-------------------------------|
| accuracy | 76.92% | 100% | 86.36% | 100% |
| classification error | 23.08% | 0% | 13.64% | 0% |
| kappa | 0.672 | 1 | 0.808 | 1 |
| weighted mean recall | 67.22% | 100% | 75.00% | 100% |
| weighted mean precision | 56.67% | 100% | 64.73% | 100% |
| spearman rho | 0.802 | 1 | 0.972 | 1 |
| kendall tau | 0.775 | 1 | 0.945 | 1 |
| absolute error | 0.231 +/- 0.421 | 0.000 +/- 0.000 | 0.136 +/- 0.343 | 0.000 +/- 0.000 |
| relative error | 23.08% +/- 42.13% | 0.00% +/- 0.00% | 13.64% +/- 34.32% | 0.00% +/- 0.00% |
| relative error strict | ∞% | 0.00% +/- 0.00% | ∞% | 0.00% +/- 0.00% |
| normalized absolute error | 0.3 | 0 | 0.177 | 0 |
| root mean squared error | 0.480 +/- 0.000 | 0.000 +/- 0.000 | 0.369 +/- 0.000 | 0.000 +/- 0.000 |
| root relative squared error | 0.624 | 0 | 0.48 | 0 |
| squared error | 0.231 +/- 0.421 | 0.000 +/- 0.000 | 0.136 +/- 0.343 | 0.000 +/- 0.000 |
| correlation | 0.793 | 1 | 0.954 | 1 |
| squared correlation | 0.629 | 1 | 0.91 | 1 |
| cross-entropy | ∞ | 0 | unknown | 0 |

ตารางที่ 16 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการ
วิเคราะห์แบบ Backward elimination

| | Naïve bayes | apply_Naïve bayes | Naïve bayes_kernel (10) | apply_Naïve bayes-kernel (10) |
|-----------------------------|------------------|-------------------|-------------------------|-------------------------------|
| | 119 | 31 s | 120 | 52 s |
| accuracy | 90.59% | 100% | 86.90% | 100% |
| classification error | 9.41% | 0% | 13.10% | 0% |
| kappa | 0.867 | 1 | 0.814 | 1 |
| weighted mean recall | 85.70% | 100% | 79.93% | 100% |
| weighted mean precision | 92.69% | 100% | 90.60% | 100% |
| spearman rho | 8.516 | 1 | 8.066 | 1 |
| kendall tau | 8.451 | 1 | 7.903 | 1 |
| absolute error | 0.094 +/- 0.292 | 0.000 +/- 0.000 | 0.131 +/- 0.337 | 0.000 +/- 0.000 |
| relative error | 9.43% +/- 29.20% | 0.00% +/- 0.00% | 13.10% +/- 33.73% | 0.00% +/- 0.00% |
| relative error strict | ∞% | 0.00% +/- 0.00% | ∞% | 0.00% +/- 0.00% |
| normalized absolute error | 0.121 | 0 | 0.169 | 0 |
| root mean squared error | 0.307 +/- 0.000 | 0.000 +/- 0.000 | 0.362 +/- 0.000 | 0.000 +/- 0.000 |
| root relative squared error | 0.382 | 0 | 0.452 | 0 |
| squared error | 0.094 +/- 0.292 | 0.000 +/- 0.000 | 0.131 +/- 0.337 | 0.000 +/- 0.000 |
| correlation | 0.842 | 1 | 0.814 | 1 |
| squared correlation | 0.709 | 1 | 0.662 | 1 |
| cross-entropy | ∞ | 0 | unknown | 0 |



ตารางที่ 17 ประสิทธิภาพของโมเดล Naïve Bayes และ Naïve Bayes kernel ในกระบวนการวิเคราะห์แบบ Forward selection

| | Naïve bayes | apply_Naïve bayes | Naïve bayes_kernel (10) | apply_Naïve bayes-kernel (10) |
|-----------------------------|-----------------|-------------------|-------------------------|-------------------------------|
| | | | 5 | 28 s |
| accuracy | 93.89% | 92.94% | 97.50% | 98.82% |
| classification error | 6.11% | 7.06% | 2.50% | 1.18% |
| kappa | 0.911 | 0.902 | 0.963 | 0.984 |
| weighted mean recall | 88.33% | 90.50% | 90.00% | 98.08% |
| weighted mean precision | 88.75% | 91.05% | 88.54% | 98.53% |
| spearman rho | 0.91 | 0.901 | 0.95 | 0.998 |
| kendall tau | 0.904 | 0.894 | 0.949 | 0.995 |
| absolute error | 0.078 +/- 0.094 | 0.062 +/- 0.218 | 0.034 +/- 0.049 | 0.023 +/- 0.110 |
| relative error | 7.78% +/- 9.41% | 6.19% +/- 21.80% | 3.37% +/- 4.90% | 2.33% +/- 11.03% |
| relative error strict | ∞% | ∞% | ∞% | 32.53% +/- 284.63% |
| normalized absolute error | 0.085 | 0.08 | 0.044 | 0.03 |
| root mean squared error | 0.180 +/- 0.189 | 0.227 +/- 0.000 | 0.087 +/- 0.141 | 0.113 +/- 0.000 |
| root relative squared error | 0.199 | 0.292 | 0.114 | 0.145 |
| squared error | 0.065 +/- 0.091 | 0.051 +/- 0.203 | 0.025 +/- 0.052 | 0.013 +/- 0.101 |
| correlation | 0.907 | 0.887 | 0.932 | 0.955 |
| squared correlation | 0.847 | 0.787 | 0.906 | 0.99 |
| cross-entropy | ∞ | 0.527 | unknown | 0.074 |

จากตารางทั้ง 3 ตารางในข้างต้น ผู้วิจัยได้ใช้ค่า Accuracy เป็นหลัก ในการพิจารณาเลือกโมเดลที่มีประสิทธิภาพดีที่สุดในกระบวนการวิเคราะห์ที่ต่างกันทั้ง 3 แบบ พบว่ากระบวนการวิเคราะห์แบบ Split validation และ Forward Selection ให้ค่า Accuracy ของโมเดลแบบ Naïve Bayes kernel เท่ากับ 86.36-100% (จาก Testing dataset 30%) โดยใช้ข้อความ 123 ข้อ และ 97.50-98.82% โดยใช้ข้อความ 5 ข้อ (CCONVER, CMEMZ, CMUSICZ, CSPEECH, and EDRAWZ) ตามลำดับ ส่วนกระบวนการวิเคราะห์แบบ Backward elimination ให้ค่า Accuracy ของโมเดลแบบ Normal Naïve Bayes เท่ากับ 90.59-100% โดยตัดข้อความ CHAT5 NOD5 AVOICE5 และ EMLHAND ออก เหลือข้อความที่ใช้วิเคราะห์ 119 ข้อ โดยที่กระบวนการวิเคราะห์ทั้ง 3 แบบใช้เวลาไม่เกิน 1 นาที ดังแสดงในตารางที่ 18

ตารางที่ 18 เปรียบเทียบประสิทธิภาพของโมเดล Naïve Bayes หรือ Naïve Bayes kernel จากทั้ง 3 กระบวนการวิเคราะห์

| Subgroups | Parameter | Algorithms: Naïve Bayes and Naïve Bayes-kernel | | | | | |
|-------------|--------------------|------------------------------------------------|------|-------------|------|-------------|--------|
| | | Split validation (kernel) | | BE (normal) | | FS (kernel) | |
| | No. of ADI-R items | 123 | | 119 | | 5 | |
| | time | 0 s | | 31 s | | 28 s | |
| | accuracy | 86.36% | 100% | 90.59% | 100% | 97.5% | 98.82% |
| A (n=13) | Precision | 0% | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 0% | 100% | 61.54% | 100% | 92.31% | 92.31% |
| | Specificity | 100% | 100% | 100% | 100% | 100% | 98.63% |
| | Recall | 0% | 100% | 61.54% | 100% | 100% | 92.31% |
| | F-score | 0% | 100% | 76.19% | 100% | 96% | 96% |
| B (n=30) | Precision | 87.5% | 100% | 88.24% | 100% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 92.31% | 100% | 92.16% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 93.33% | 100% | 93.75% | 100% | 100% | 100% |
| C (n=26) | Precision | 100% | 100% | 89.66% | 100% | 96.15% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 94.44% | 100% | 98.28% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 94.55% | 100% | 98.04% | 100% |
| D (n=16) | Precision | 71.43% | 100% | 92.86% | 100% | 93.75% | 94.12% |
| | Sensitivity | 100% | 100% | 81.25% | 100% | 93.75% | 100% |
| | Specificity | 87.5% | 100% | 98.46% | 100% | 98.53% | 100% |
| | Recall | 100% | 100% | 81.25% | 100% | 93.75% | 100% |
| | F-score | 83.33% | 100% | 86.67% | 100% | 93.75% | 96.97% |

4.2.4 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Decision tree

สำหรับอัลกอริทึม Decision tree ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของโมเดล kernel ในกระบวนการวิเคราะห์ทั้ง 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection ดังตารางที่ 19



ตารางที่ 19 ประสิทธิภาพของโมเดล Decision tree ในกระบวนการวิเคราะห์แบบ Split validation Backward elimination และ Forward selection

| | 1 tree | apply_1 tree | 1 tree_BE | apply_1 tree_BE | 1 tree_FS | apply_1 tree_FS | |
|-----------------------------|-------------------|--------------------|-------------------|--------------------|-----------------|--------------------|---------|
| | 4 | 0 s | 122 >> 3 | 20 s | 4 | 45 s | |
| accuracy | 84.62% | 97.65% | 89.41% | 97.65% | 95.14% | 95.29% | CVERRIT |
| classification error | 15.38% | 2% | 10.59% | 2.35% | 4.86% | 4.71% | EDRAWZ |
| kappa | 0.789 | 0.967 | 0.853 | 0.967 | 0.93 | 0.935 | ECOMPUZ |
| weighted mean recall | 81.25% | 96.88% | 86.66% | 96.88% | 87.92% | 94.59% | ESTEREO |
| weighted mean precision | 88.89% | 98.21% | 87.90% | 98.21% | 86.46% | 94.33% | QUALOV5 |
| spearman rho | 0.736 | 0.913 | 7.487 | 0.913 | 0.944 | 0.944 | |
| kendall tau | 0.735 | 0.918 | 7.507 | 0.918 | 0.942 | 0.938 | |
| absolute error | 0.154 +/- 0.361 | 0.044 +/- 0.141 | 0.120 +/- 0.291 | 0.044 +/- 0.141 | 0.093 +/- 0.059 | 0.082 +/- 0.185 | |
| relative error | 15.38% +/- 36.08% | 4.37% +/- 14.12% | 12.01% +/- 29.13% | 4.37% +/- 14.12% | 9.29% +/- 5.89% | 8.20% +/- 18.52% | |
| relative error strict | ∞% | 32.94% +/- 196.72% | ∞% | 32.94% +/- 196.72% | ∞% | 71.76% +/- 371.63% | |
| normalized absolute error | 0.2 | 0.056 | 0.155 | 0.056 | 0.13 | 0.106 | |
| root mean squared error | 0.392 +/- 0.000 | 0.148 +/- 0.000 | 0.315 +/- 0.000 | 0.148 +/- 0.000 | 0.179 +/- 0.155 | 0.203 +/- 0.000 | |
| root relative squared error | 0.51 | 0.19 | 0.393 | 0.19 | 0.247 | 0.261 | |
| squared error | 0.154 +/- 0.361 | 0.022 +/- 0.130 | 0.099 +/- 0.285 | 0.022 +/- 0.130 | 0.054 +/- 0.066 | 0.041 +/- 0.169 | |
| correlation | 0.741 | 0.909 | 0.744 | 0.909 | 0.947 | 0.94 | |
| squared correlation | 0.55 | 0.827 | 0.553 | 0.827 | 0.911 | 0.884 | |
| cross-entropy | ∞ | 0.121 | ∞ | 0.121 | ∞ | 0.226 | |

จากตารางในข้างต้น พบว่ากระบวนการวิเคราะห์แบบ Split validation ให้ค่า Accuracy ของโมเดลเท่ากับ 84.62-97.65% (จาก Testing dataset 30%) โดยใช้ข้อคำถามเพียง 4 ข้อ (LEVELL, ECOMPUZ, CCOMPUZ, and ESTEREO) กระบวนการวิเคราะห์แบบ Backward elimination ให้ค่า Accuracy ของโมเดลเท่ากับ 89.41-97.65% โดยใช้ข้อคำถามเพียง 3 ข้อ (LEVELL, ECOMPUZ, and ESTEREO) และกระบวนการวิเคราะห์แบบ Forward selection ให้ค่า Accuracy ของโมเดลเท่ากับ 95.14-95.29% ใช้ข้อคำถาม 4 ข้อ (CVERRIT, EDRAWZ, ECOMPUZ, and QUALOV5) โดยที่กระบวนการวิเคราะห์ทั้ง 3 แบบใช้เวลาไม่เกิน 1 นาที ดังแสดงในตารางที่

20



ตารางที่ 20 เปรียบเทียบประสิทธิภาพของโมเดล Decision tree จากทั้ง 3 กระบวนการ
วิเคราะห์

| Subgroups | Parameter | Algorithms: Decision tree | | | | | |
|-------------|--------------------|---------------------------|--------|--------|--------|--------|--------|
| | | Split validation | | BE | | FS | |
| | No. of ADI-R items | 123>>4 | | 122>>3 | | 4 | |
| | time | 0 s | | 20 s | | 45 s | |
| | accuracy | 84.62% | 97.65% | 89.41 | 97.65% | 95.14% | 95.29% |
| A (n=13) | Precision | 100% | 100% | 90.91% | 100% | 92.31% | 92.31% |
| | Sensitivity | 50% | 100% | 76.92% | 100% | 92.31% | 92.31% |
| | Specificity | 100% | 100% | 98.51% | 100% | 98.57% | 98.57% |
| | Recall | 50% | 100% | 76.92% | 100% | 92.31% | 92.31% |
| | F-score | 66.67% | 100% | 83.33% | 100% | 92.31% | 92.31% |
| B (n=30) | Precision | 100% | 100% | 100% | 100% | 96.77% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 98.08% | 98.08% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 100% | 100% | 98.36% | 98.36% |
| C (n=26) | Precision | 100% | 92.86% | 88.46% | 92.86% | 96.15% | 100% |
| | Sensitivity | 75% | 100% | 88.46% | 100% | 96.15% | 92.31% |
| | Specificity | 80.95% | 96.61% | 94.64% | 96.61% | 98.24% | 100% |
| | Recall | 75% | 100% | 88.46% | 100% | 96.15% | 92.31% |
| | F-score | 85.71% | 96.3% | 88.46% | 96.3% | 96.15% | 96% |
| D (n=16) | Precision | 55.56% | 100% | 72.22% | 100% | 93.33% | 88.24% |
| | Sensitivity | 100% | 87.5% | 81.25% | 87.5% | 87.5% | 93.75% |
| | Specificity | 80.95% | 100% | 92.65% | 100% | 98.53% | 97.05% |
| | Recall | 100% | 87.5% | 81.25% | 87.5% | 87.5% | 93.75% |
| | F-score | 71.43% | 93.33% | 76.47% | 93.33% | 90.32% | 90.91% |

4.2.5 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Random forest

สำหรับอัลกอริทึม Random forest ผู้วิจัยได้ทำการเปรียบเทียบการปรับค่าพารามิเตอร์ของจำนวน Tree ในกระบวนการวิเคราะห์ทั้ง 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection ที่ 1 5 10 25 50 และ 100 ต้น ดังตารางที่ 21-23 ตามลำดับ



ตารางที่ 21 ประสิทธิภาพของโมเดล Random forest ที่ 1 5 10 25 50 และ 100 ต้น ในกระบวนการวิเคราะห์แบบ Split validation

| | 1 tree | apply_1 tree | 5 trees | apply_5 trees | 10 trees | apply_10 trees | 25 trees | apply_25 trees | 50 trees | apply_50 trees | 100 trees | apply_100 trees |
|-----------------------------|------------|--------------|------------|---------------|------------|----------------|------------|----------------|------------|----------------|------------|-----------------|
| | 11 | | 35 | | 63 | | 63 | | 63 | | 63 | |
| accuracy | 84.62% | 94.12% | 80.77% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% | 92.31% | 100% |
| classification error | 15.38% | 5.88% | 19.23% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% | 7.69% | 0% |
| kappa | 0.786 | 0.919 | 0.732 | 1 | 0.893 | 1 | 0.893 | 1 | 0.893 | 100% | 0.893 | 100% |
| weighted mean recall | 80.62% | 94.36% | 74.38% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% | 87.50% | 100% |
| weighted mean precision | 81.88% | 93.30% | 76.70% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% | 92.86% | 100% |
| spearman rho | 0.688 | 0.9 | 0.807 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 | 0.988 | 1 |
| kendall tau | 0.689 | 0.897 | 0.788 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 | 0.971 | 1 |
| absolute error | 0.154 +/- | 0.059 +/- | 0.246 +/- | 0.061 +/- | 0.289 +/- | 0.163 +/- | 0.286 +/- | 0.154 +/- | 0.288 +/- | 0.141 +/- | 0.295 +/- | 0.140 +/- |
| | 0.361 | 0.235 | 0.256 | 0.106 | 0.210 | 0.110 | 0.176 | 0.096 | 0.176 | 0.092 | 0.161 | 0.088 |
| relative error | 15.38% +/- | 5.88% +/- | 24.62% +/- | 6.12% +/- | 28.85% +/- | 16.31% +/- | 28.63% +/- | 15.44% +/- | 28.82% +/- | 14.07% +/- | 29.51% +/- | 13.99% +/- |
| | 36.08% | 23.53% | 25.61% | 10.64% | 21.03% | 11.03% | 17.61% | 9.61% | 17.58% | 9.17% | 16.07% | 8.80% |
| relative error strict | ∞% | ∞% | 65.06% +/- | 8.24% +/- | 91.09% +/- | 21.94% +/- | 58.96% +/- | 19.97% +/- | 60.67% +/- | 18.00% +/- | 57.23% +/- | 17.78% +/- |
| | | | 107.34% | 15.37% | 197.78% | 18.82% | 86.59% | 15.29% | 94.48% | 15.51% | 75.61% | 15.05% |
| normalized absolute error | 0.2 | 0.076 | 0.32 | 0.079 | 0.375 | 0.21 | 0.372 | 0.199 | 0.375 | 0.181 | 0.384 | 0.18 |
| root mean squared error | 0.392 +/- | 0.243 +/- | 0.355 +/- | 0.123 +/- | 0.357 +/- | 0.197 +/- | 0.336 +/- | 0.182 +/- | 0.338 +/- | 0.168 +/- | 0.336 +/- | 0.165 +/- |
| | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| root relative squared error | 0.51 | 0.312 | 0.462 | 0.158 | 0.464 | 0.254 | 0.437 | 0.234 | 0.439 | 0.216 | 0.437 | 0.213 |
| squared error | 0.154 +/- | 0.098 +/- | 0.126 +/- | 0.015 +/- | 0.127 +/- | 0.039 +/- | 0.113 +/- | 0.033 +/- | 0.114 +/- | 0.028 +/- | 0.113 +/- | 0.027 +/- |
| | 0.361 | 0.244 | 0.187 | 0.032 | 0.198 | 0.048 | 0.147 | 0.038 | 0.149 | 0.039 | 0.135 | 0.038 |
| correlation | 0.659 | 0.235 | 0.774 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 | 0.975 | 1 |
| squared correlation | 0.434 | 0.901 | 0.599 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 | 0.95 | 1 |
| cross-entropy | ∞ | 0.811 | 0.512 | 0.101 | ∞ | 0.268 | 0.533 | 0.249 | 0.539 | 0.225 | 0.54 | 0.223 |

ตารางที่ 22 ประสิทธิภาพของโมเดล Random forest ที่ 5 10 25 50 และ 100 ต้น ในกระบวนการวิเคราะห์แบบ Backward elimination

| | 5 trees | apply_5 trees | 10 trees | apply_10 trees | 25 trees | apply_25 trees | 50 trees | apply_50 trees | 100 trees | apply_100 trees |
|-----------------------------|-------------------|------------------|-------------------|------------------|-------------------|------------------|-------------------|-----------------|-------------------|-----------------|
| | 122>>45 | | 122>>72 | | 122>> | | 122>> | | 122>> | |
| accuracy | 97.65% | 100% | 97.65% | 100% | 97.65% | 100% | 97.65% | 100% | 96.47% | 100% |
| classification error | 2.35% | 0% | 2.35% | 0% | 2.35% | 0% | 2.35% | 0% | 3.53% | 0% |
| kappa | 0.967 | 1 | 0.967 | 1 | 0.967 | 1 | 0.967 | 1 | 0.951 | 100% |
| weighted mean recall | 96.15% | 100% | 96.15% | 100% | 96.15% | 100% | 96.15% | 100% | 94.23% | 100% |
| weighted mean precision | 97.72% | 100% | 97.72% | 100% | 97.72% | 100% | 97.72% | 100% | 96.42% | 100% |
| spearman rho | 9.948 | 1 | 9.935 | 1 | 9.935 | 1 | 9.935 | 1 | 9.928 | 1 |
| kendall tau | 9.906 | 1 | 9.885 | 1 | 9.885 | 1 | 9.885 | 1 | 9.862 | 1 |
| absolute error | 0.165 +/- 0.187 | 0.056 +/- 0.095 | 0.188 +/- 0.175 | 0.073 +/- 0.092 | 0.178 +/- 0.163 | 0.071 +/- 0.079 | 0.193 +/- 0.155 | 0.073 +/- 0.068 | 0.189 +/- 0.156 | 0.073 +/- 0.064 |
| relative error | 16.47% +/- 18.71% | 5.65% +/- 9.51% | 18.83% +/- 17.47% | 7.28% +/- 9.23% | 17.84% +/- 16.29% | 7.06% +/- 7.94% | 19.35% +/- 15.50% | 7.28% +/- 6.75% | 18.88% +/- 15.61% | 7.26% +/- 6.40% |
| relative error strict | 30.20% +/- 53.99% | 7.25% +/- 12.71% | 30.74% +/- 36.41% | 9.16% +/- 13.33% | 34.71% +/- 84.54% | 8.49% +/- 10.48% | 31.63% +/- 42.96% | 8.50% +/- 8.95% | 31.80% +/- 48.76% | 8.41% +/- 8.38% |
| normalized absolute error | 0.212 | 0.073 | 0.242 | 0.094 | 0.23 | 0.091 | 0.249 | 0.094 | 0.243 | 0.094 |
| root mean squared error | 0.249 +/- 0.000 | 0.111 +/- 0.000 | 0.257 +/- 0.000 | 0.118 +/- 0.000 | 0.242 +/- 0.000 | 0.106 +/- 0.000 | 0.248 +/- 0.000 | 0.099 +/- 0.000 | 0.245 +/- 0.000 | 0.097 +/- 0.000 |
| root relative squared error | 0.311 | 0.142 | 0.32 | 0.151 | 0.301 | 0.137 | 0.309 | 0.128 | 0.305 | 0.125 |
| squared error | 0.062 +/- 0.107 | 0.012 +/- 0.024 | 0.066 +/- 0.091 | 0.014 +/- 0.029 | 0.058 +/- 0.115 | 0.011 +/- 0.020 | 0.061 +/- 0.098 | 0.010 +/- 0.017 | 0.060 +/- 0.104 | 0.009 +/- 0.016 |
| correlation | 0.99 | 1 | 0.99 | 1 | 0.99 | 1 | 0.99 | 1 | 0.985 | 1 |
| squared correlation | 0.98 | 1 | 0.98 | 1 | 0.98 | 1 | 0.98 | 1 | 0.971 | 1 |
| cross-entropy | 0.276 | 0.091 | 0.304 | 0.116 | 0.295 | 0.11 | 0.309 | 0.112 | 0.304 | 0.111 |

ตารางที่ 23 ประสิทธิภาพของโมเดล Random forest ที่ 1 5 10 25 50 และ 100 ต้น ในกระบวนการวิเคราะห์ที่แบบ Forward selection

| | 5 trees | apply_5 trees | 10 trees | apply_10 trees | 25 trees | apply_25 trees | 50 trees | apply_50 trees | 100 trees | apply_100 trees |
|-----------------------------|--------------------|-------------------|-----------------|------------------|-------------------|------------------|-------------------|-----------------|------------------|-------------------|
| | 3 | | 5 | 38s | 5 | | 5 | | 3 | |
| accuracy | 96.39% +/- 5.83% | 96.67% | 98.75% | 100% | 98.75% | 100% | 97.64% | 100% | 95.14% | 97.78% |
| classification error | 3.61% | 3.33% | 1.25% | 0% | 1.25% | 0% | 2.36% | 0% | 4.86% | 2.22% |
| kappa | 0.947 | 0.954 | 0.982 | 1 | 0.982 | 1 | 0.967 | 100% | 0.928 | 0.969 |
| weighted mean recall | 90.00% | 95.59% | 92.50% | 100% | 92.50% | 100% | 91.25% | 100% | 89.17% | 96.74% |
| weighted mean precision | 89.79% | 96.67% | 91.67% | 100% | 91.67% | 100% | 90.42% | 100% | 89.37% | 97.67% |
| spearman rho | 0.917 | 0.917 | 0.998 | 1 | 0.999 | 1 | 0.956 | 1 | 0.895 | 0.957 |
| kendall tau | 0.919 | 0.919 | 0.996 | 1 | 0.998 | 1 | 0.959 | 1 | 0.897 | 0.957 |
| absolute error | 0.134 +/- 0.071 | 0.077 +/- 0.132 | 0.075 +/- 0.053 | 0.037 +/- 0.096 | 0.079 +/- 0.065 | 0.032 +/- 0.076 | 0.078 +/- 0.063 | 0.035 +/- 0.069 | 0.137 +/- 0.072 | 0.099 +/- 0.125 |
| relative error | 13.38% +/- 7.08% | 7.70% +/- 13.20% | 7.53% +/- 5.35% | 3.71% +/- 9.58% | 7.93% +/- 6.54% | 3.18% +/- 7.59% | 7.82% +/- 6.30% | 3.47% +/- 6.86% | 13.67% +/- 7.24% | 9.90% +/- 12.55% |
| relative error strict | 77.33% +/- 127.63% | 12.56% +/- 31.43% | ∞% | 5.43% +/- 16.14% | 13.35% +/- 14.68% | 4.15% +/- 11.18% | 13.41% +/- 13.42% | 4.30% +/- 9.96% | ∞% | 15.04% +/- 30.99% |
| normalized absolute error | 0.185 | 0.099 | 0.105 | 0.048 | 0.102 | 0.041 | 0.105 | 0.045 | 0.182 | 0.127 |
| root mean squared error | 0.203 +/- 0.120 | 0.153 +/- 0.000 | 0.147 +/- 0.113 | 0.103 +/- 0.000 | 0.128 +/- 0.101 | 0.082 +/- 0.000 | 0.130 +/- 0.102 | 0.077 +/- 0.000 | 0.216 +/- 0.135 | 0.160 +/- 0.000 |
| root relative squared error | 0.278 | 0.196 | 0.205 | 0.132 | 0.166 | 0.106 | 0.175 | 0.099 | 0.294 | 0.206 |
| squared error | 0.054 +/- 0.058 | 0.023 +/- 0.072 | 0.033 +/- 0.041 | 0.011 +/- 0.038 | 0.026 +/- 0.032 | 0.007 +/- 0.024 | 0.026 +/- 0.030 | 0.006 +/- 0.021 | 0.063 +/- 0.067 | 0.026 +/- 0.071 |
| correlation | 0.912 | 0.91 | 0.996 | 1 | 0.996 | 1 | 0.954 | 1 | 0.902 | 0.953 |
| squared correlation | 0.86 | 0.829 | 0.993 | 1 | 0.992 | 1 | 0.926 | 1 | 0.85 | 0.908 |
| cross-entropy | 0.277 | 0.136 | ∞ | 0.063 | 0.128 | 0.051 | 0.127 | 0.055 | ∞ | 0.169 |

จากตารางทั้ง 3 ตารางในข้างต้น ผู้วิจัยได้ใช้ค่า Accuracy เป็นหลัก ในการพิจารณาเลือกโมเดลที่มีประสิทธิภาพดีที่สุดในกระบวนการวิเคราะห์ที่ต่างกันทั้ง 3 แบบ พบว่ากระบวนการวิเคราะห์แบบ Split validation ที่ใช้ต้นไม้ 10 ต้น ให้ค่า Accuracy ของโมเดลเท่ากับ 92.31-100% (จาก Testing dataset 30%) โดยใช้ข้อความทดลองเหลือ 63 ข้อ กระบวนการวิเคราะห์แบบ Backward elimination ที่ใช้ต้นไม้ 5 ต้น ให้ค่า Accuracy ของโมเดลเท่ากับ 90.59-100% ใช้ข้อความทดลองเหลือ 45 ข้อ และกระบวนการวิเคราะห์แบบ Forward selection ที่ใช้ต้นไม้ 10 ต้น ให้ค่า Accuracy ของโมเดลเท่ากับ 98.82-100% ใช้ข้อความเพียง 5 ข้อ (CPRON, CVERRIT, ECOMPUZ, LEVELL, and RFACEX5) โดยที่กระบวนการวิเคราะห์แบบ Backward elimination ใช้เวลาในการวิเคราะห์นานกว่า 1 นาที ดังแสดงในตารางที่ 24



ตารางที่ 24 เปรียบเทียบประสิทธิภาพของโมเดล Random forest จากทั้ง 3 กระบวนการ
วิเคราะห์

| Subgroups | Parameter | Algorithms: Random forest | | | | | |
|-------------|--------------------|-----------------------------|------|--------------|------|---------------|------|
| | | Split validation (10 trees) | | BE (5 trees) | | FS (10 trees) | |
| | No. of ADI-R items | 63 | | 122>>45 | | 5 | |
| | time | 0 s | | 1.03 m | | 38 s | |
| | accuracy | 92.31% | 100% | 90.59% | 100% | 98.82% | 100% |
| A (n=13) | Precision | 100% | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 50% | 100% | 84.62% | 100% | 92.31% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 98.55% | 100% |
| | Recall | 50% | 100% | 84.62% | 100% | 92.31% | 100% |
| | F-score | 66.67% | 100% | 91.67% | 100% | 96% | 100% |
| B (n=30) | Precision | 100% | 100% | 96.77% | 100% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 98.15% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 98.36% | 100% | 100% | 100% |
| C (n=26) | Precision | 100% | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 100% | 100% | 100% | 100% | 100% | 100% |
| D (n=16) | Precision | 71.43% | 100% | 94.12% | 100% | 94.12% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 90.48% | 100% | 98.53% | 100% | 98.55% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% | 100% |
| | F-score | 83.33% | 100% | 96.97% | 100% | 96.97% | 100% |

4.2.6 ผลการเปรียบเทียบประสิทธิภาพของโมเดลจากอัลกอริทึม Support Vector

Machine (SVM) และ A Library of Support Vector Machine (LibSVM)

ผู้วิจัยได้ทำการเปรียบเทียบประสิทธิภาพของ Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์ทั้ง 3 แบบ ได้แก่ Split validation Backward elimination และ Forward selection แบบ Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ดังตารางที่ 25-27 ตามลำดับ

ตารางที่ 25 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Split validation

| | SVM (Classification by regression) | apply_SVM | LibSVM | apply_LibSVM |
|-----------------------------|------------------------------------|-------------------|-------------------|-----------------|
| accuracy | 76.92% | 87.06% | 80.77% | 100% |
| classification error | 23.08% | 12.94% | 19.23% | 0% |
| kappa | 0.675 | 0.815 | 0.73 | 1 |
| weighted mean recall | 68.12% | 79.57% | 76.94% | 100% |
| weighted mean precision | 79.44% | 91.28% | 90.38% | 100% |
| spearman rho | 0.663 | 0.877 | 0.675 | 1 |
| kendall tau | 0.663 | 0.856 | 0.652 | 1 |
| absolute error | 0.231 +/- 0.421 | 0.129 +/- 0.336 | 0.192 +/- 0.394 | 0.000 +/- 0.000 |
| relative error | 23.08% +/- 42.13% | 12.94% +/- 33.57% | 19.23% +/- 39.41% | 0.00% +/- 0.00% |
| relative error strict | ∞% | ∞% | ∞% | 0.00% +/- 0.00% |
| normalized absolute error | 0.3 | 0.167 | 0.25 | 0.000 +/- 0.000 |
| root mean squared error | 0.480 +/- 0.000 | 0.360 +/- 0.000 | 0.439 +/- 0.000 | 0.000 +/- 0.000 |
| root relative squared error | 0.624 | 0.463 | 0.57 | 0.000 +/- 0.000 |
| squared error | 0.231 +/- 0.421 | 0.129 +/- 0.336 | 0.192 +/- 0.394 | 0.000 +/- 0.000 |
| correlation | 0.659 | 0.869 | 0.741 | 1 |
| squared correlation | 0.435 | 0.756 | 0.549 | 1 |
| cross-entropy | ∞ | ∞ | ∞ | 0 |

ตารางที่ 26 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Backward elimination

| | SVM (Classification by regression) | apply_SVM | LibSVM | apply_LibSVM |
|-----------------------------|------------------------------------|-------------------|------------------|-----------------|
| | 119 | 6.21 m | 122 | 38 s |
| accuracy | 83.53% | 87.06% | 97.65% | 100% |
| classification error | 16.47% | 12.94% | 2.35% | 0% |
| kappa | 0.764 | 0.815 | 0.967 | 1 |
| weighted mean recall | 74.16% | 79.57% | 96.51% | 100% |
| weighted mean precision | 88.31% | 91.28% | 97.51% | 100% |
| spearman rho | 8.284 | 0.877 | 9.502 | 1 |
| kendall tau | 8.095 | 0.856 | 9.494 | 1 |
| absolute error | 0.165 +/- 0.371 | 0.129 +/- 0.336 | 0.024 +/- 0.152 | 0.000 +/- 0.000 |
| relative error | 16.47% +/- 37.09% | 12.94% +/- 33.57% | 2.35% +/- 15.16% | 0.00% +/- 0.00% |
| relative error strict | ∞% | ∞% | ∞% | 0.00% +/- 0.00% |
| normalized absolute error | 0.212 | 0.167 | 0.03 | 0.000 +/- 0.000 |
| root mean squared error | 0.406 +/- 0.000 | 0.360 +/- 0.000 | 0.153 +/- 0.000 | 0.000 +/- 0.000 |
| root relative squared error | 0.506 | 0.463 | 0.191 | 0.000 +/- 0.000 |
| squared error | 0.165 +/- 0.371 | 0.129 +/- 0.336 | 0.024 +/- 0.152 | 0.000 +/- 0.000 |
| correlation | 0.816 | 0.869 | 0.95 | 1 |
| squared correlation | 0.666 | 0.756 | 0.902 | 1 |
| cross-entropy | ∞ | ∞ | ∞ | 0 |

ตารางที่ 27 ประสิทธิภาพของโมเดล Support Vector Machine (SVM) และ A Library of Support Vector Machine (LibSVM) ในกระบวนการวิเคราะห์แบบ Forward selection

| | SVM (classification by regression) | apply_SVM | LibSVM | apply_LibSVM |
|---------------------------|------------------------------------|-------------------|-------------------|-----------------|
| | 9 | 4.28 m | 6 | 1.03 m |
| accuracy | 82.50% +/- 5.64% | 82.35% | 100% | 100% |
| classification error | 17.50% +/- 5.64% | 17.65% | 0% | 0% |
| kappa | 0.743 +/- 0.085 | 0.745 | 1 | 1 |
| weighted mean recall | 70.42% +/- 8.44% | 71.88% | 95.00% +/- 10.54% | 100% |
| weighted mean precision | 63.08% +/- 8.57% | 64.04% | 95.00% +/- 10.54% | 100% |
| spearman rho | 0.855 +/- 0.152 | 0.841 | 1 | 1 |
| kendall tau | 0.824 +/- 0.146 | 0.81 | 1 | 1 |
| absolute error | 0.175 +/- 0.056 | 0.176 +/- 0.381 | 0.000 +/- 0.000 | 0.000 +/- 0.000 |
| relative error | 17.50% +/- 5.64% | 17.65% +/- 38.12% | 0.00% +/- 0.00% | 0.00% +/- 0.00% |
| relative error strict | ∞% | ∞% | 0.00% +/- 0.00% | 0.00% +/- 0.00% |
| normalized absolute error | 0.243 +/- 0.120 | 0.227 | 0.000 +/- 0.000 | 0.000 +/- 0.000 |
| root mean squared error | 0.413 +/- 0.068 | 0.420 +/- 0.000 | 0.000 +/- 0.000 | 0.000 +/- 0.000 |
| squared error | 0.175 +/- 0.056 | 0.176 +/- 0.381 | 0.000 +/- 0.000 | 0.000 +/- 0.000 |
| correlation | 0.855 +/- 0.135 | 0.84 | 1 | 1 |
| squared correlation | 0.747 +/- 0.209 | 0.706 | 1 | 1 |
| cross-entropy | ∞ | ∞ | 0 | 0 |

จากตารางทั้ง 3 ตารางในข้างต้น ผู้วิจัยได้ใช้ค่า Accuracy เป็นหลัก ในการพิจารณาเลือกโมเดลที่มีประสิทธิภาพดีที่สุดจากกระบวนการวิเคราะห์ที่ต่างกันทั้ง 3 แบบ พบว่า Support Vector Machine (SVM) กระบวนการวิเคราะห์แบบ Split validation ให้ค่า Accuracy ของโมเดลเท่ากับ 76.92-87.06% (จาก Testing dataset 30%) โดยใช้ข้อความ 123 ข้อ กระบวนการวิเคราะห์แบบ Backward elimination ให้ค่า Accuracy ของโมเดลเท่ากับ 83.53-87.06% โดยตัดข้อความ CARTIC CNEOID CPEERPL และ CSOPLAY ออก เหลือข้อความที่ใช้วิเคราะห์ 119 ข้อ และกระบวนการวิเคราะห์แบบ Forward selection ให้ค่า Accuracy ของโมเดลเท่ากับ 82.35% ใช้ข้อความ 9 ข้อ (CAGGFAM, CCOMPUZ, CHAT5, CINAPPQ, CNEOID, CVISSPZ, ESLFINJ, OCOMF5, and RFACEX5) ส่วน A Library of Support Vector Machine (LibSVM) กระบวนการวิเคราะห์แบบ Split validation ให้ค่า Accuracy ของโมเดลเท่ากับ 76.92-87.06% (จาก Testing dataset 30%) โดยใช้ข้อความ 123 ข้อ, กระบวนการวิเคราะห์แบบ Backward elimination ให้ค่า Accuracy ของโมเดลเท่ากับ 97.65-100% โดยตัดข้อความ CUATT ออก เหลือข้อความที่ใช้วิเคราะห์ 122 ข้อ และกระบวนการวิเคราะห์แบบ Forward selection ให้ค่า accuracy ของโมเดลเท่ากับ 100% ใช้ข้อความ 6 ข้อ (CCOMPUZ, CSPEECH, EDRAWZ, INITIA5, LEVELL, and RFACEX5) ดังแสดงในตารางที่ 28 และ 29

ตารางที่ 28 เปรียบเทียบประสิทธิภาพของโมเดล Support Vector Machine (SVM) จากทั้ง 3 กระบวนการวิเคราะห์

| Subgroups | Parameter | Algorithms: SVM | | | | |
|-------------|--------------------|------------------|--------|--------|--------|--------|
| | | Split validation | | BE | | FS |
| | No. of ADI-R items | 123 | | 119 | | 9 |
| | time | 0 s | | 6.21 m | | 4.28 m |
| | accuracy | 76.92% | 87.06% | 83.53% | 87.06% | 82.35% |
| A (n=13) | Precision | 100% | 100% | 100% | 100% | 0% |
| | Sensitivity | 25% | 30.77% | 15.38% | 30.77% | 0% |
| | Specificity | 100% | 100% | 100% | 100% | 100% |
| | Recall | 25% | 30.77% | 15.38% | 30.77% | 0% |
| | F-score | 40% | 47.06% | 26.67% | 47.06% | 0% |
| B (n=30) | Precision | 90% | 78.95% | 76.92% | 78.95% | 73.17% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 91.67% | 84.62% | 82% | 84.62% | 78.43% |
| | Recall | 100% | 100% | 100% | 100% | 100% |
| | F-score | 94.74% | 88.24% | 86.95% | 88.24% | 84.51% |
| C (n=26) | Precision | 77.78% | 92.86% | 89.66% | 92.86% | 89.66% |
| | Sensitivity | 87.5% | 100% | 100% | 100% | 100% |
| | Specificity | 86.67% | 96% | 93.75% | 96% | 93.62% |
| | Recall | 87.5% | 100% | 100% | 100% | 100% |
| | F-score | 82.35% | 92.3% | 94.55% | 96.3% | 94.55% |
| D (n=16) | Precision | 50% | 93.33% | 86.67% | 93.33% | 93.33% |
| | Sensitivity | 60% | 87.5% | 81.25% | 87.5% | 87.5% |
| | Specificity | 85% | 98.36% | 96.67% | 98.36% | 98.24% |
| | Recall | 60% | 87.5% | 81.25% | 87.5% | 87.5% |
| | F-score | 54.54% | 90.32% | 83.87% | 90.32% | 90.32% |

ตารางที่ 29 เปรียบเทียบประสิทธิภาพของโมเดล A Library of Support Vector Machine (LibSVM) จากทั้ง 3 กระบวนการวิเคราะห์

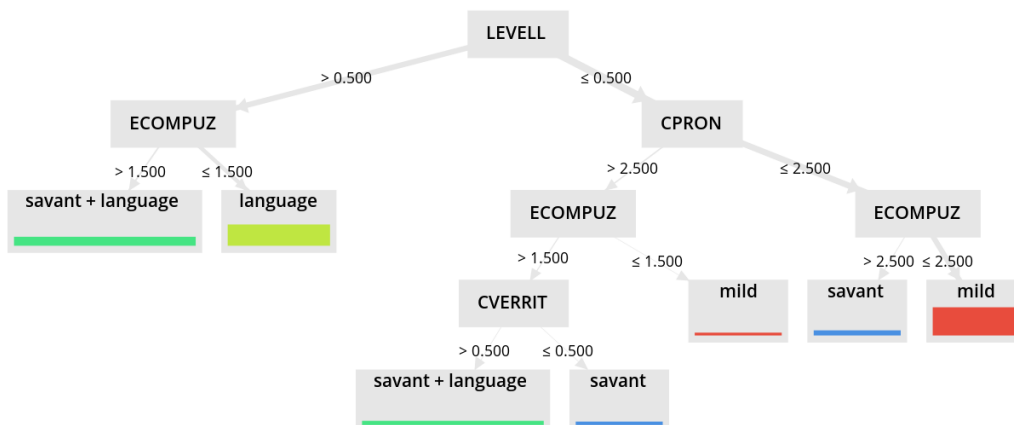
| Subgroups | Parameter | Algorithms: LibSVM | | | | |
|-------------|--------------------|--------------------|--------|--------|------|--------|
| | | Split validation | | BE | | FS |
| | No. of ADI-R items | 123 | | 122 | | 6 |
| | time | 0 s | | 38 s | | 1.03 m |
| | accuracy | 76.92% | 87.06% | 97.65% | 100% | 100% |
| A (n=13) | Precision | 100% | 100% | 100% | 100% | 100% |
| | Sensitivity | 25% | 30.77% | 92.31% | 100% | 100% |
| | Specificity | 100% | 100% | 100% | 100% | 100% |
| | Recall | 25% | 30.77% | 92.31% | 100% | 100% |
| | F-score | 40% | 47.06% | 96% | 100% | 100% |
| B (n=30) | Precision | 90% | 78.95% | 100% | 100% | 100% |
| | Sensitivity | 100% | 100% | 100% | 100% | 100% |
| | Specificity | 91.67% | 84.62% | 100% | 100% | 100% |
| | Recall | 100% | 100% | 100% | 100% | 100% |
| | F-score | 94.74% | 88.24% | 100% | 100% | 100% |
| C (n=26) | Precision | 77.78% | 92.86% | 96.3% | 100% | 100% |
| | Sensitivity | 87.5% | 100% | 100% | 100% | 100% |
| | Specificity | 86.67% | 96% | 98.28% | 100% | 100% |
| | Recall | 87.5% | 100% | 100% | 100% | 100% |
| | F-score | 82.35% | 92.3% | 98.12% | 100% | 100% |
| D (n=16) | Precision | 50% | 93.33% | 93.75% | 100% | 100% |
| | Sensitivity | 60% | 87.5% | 93.75% | 100% | 100% |
| | Specificity | 85% | 98.36% | 98.55% | 100% | 100% |
| | Recall | 60% | 87.5% | 93.75% | 100% | 100% |
| | F-score | 54.54% | 90.32% | 93.75% | 100% | 100% |

4.2.7 ผลการคัดเลือกโมเดลจากการเรียนรู้ของเครื่องมือแบบมีผู้สอน (Supervised Machine Learning) ที่ดีที่สุด

จากการเปรียบเทียบผลลัพธ์ของการประเมินประสิทธิภาพของโมเดลในแต่ละอัลกอริทึมจากกระบวนการวิเคราะห์ที่แตกต่างกันทั้งหมด 3 แบบ คือ Split validation Backward elimination และ Forward selection ทำให้ได้โมเดลที่ดีที่สุดจากแต่ละอัลกอริทึมดังแสดงในตารางที่ 30 ซึ่งเมื่อเปรียบเทียบระหว่างอัลกอริทึมทั้ง 7 แบบแล้ว พบว่าอัลกอริทึม **Random forest** ที่ใช้ต้นไม้จำนวน 10 ต้น ตามรูปที่ 22-31 (มีวิธีการอ่านแผนภาพโมเดลเช่นเดียวกับโมเดลของ Decision tree ยกตัวอย่างรูปที่ 22 คือให้เริ่มจากการอ่าน Root node ได้แก่ ข้อคำถาม LEVELL ถ้าข้อมูลที่พบเป็นไปตามเงื่อนไข คือคะแนนข้อ LEVELL < 0.5 การตัดสินใจจะวิ่งไปทางซ้ายของ Root node ไปยังจุด Child node ได้แก่ ECOMPUS ซึ่งถ้าหากข้อมูลที่มาตามเส้นทางนี้ ตรงกับเงื่อนไขของ Child node คือคะแนน ECOMPUS < 1.5 จะถือว่าสิ้นสุด โดย Node สิ้นสุดนี้จะถูกเรียกว่า Leaf node คือคนไข้จะอยู่ในกลุ่มย่อย SL subgroup โดยความหนาของกิ่งต้นไม้และความหนาของแถบสีบริเวณ Leaf node บ่งบอกถึงความชุกของตัวอย่างที่ผ่านการตัดสินใจตามกิ่งนั้นหรือมีการตัดสินใจมาตามทิศทางดังกล่าว) ให้ประสิทธิภาพของโมเดลที่ดีที่สุด คือ Accuracy 100% ในเวลาไม่เกิน 1 นาที โดยใช้ข้อคำถามแค่ 5 ข้อ ได้แก่ CPRON (Pronominal reversal; communication-verbal) CVERRIT (Verbal rituals; restricted, repetitive, and stereotyped pattern of behavior) ECOMPUS (Computer ability; savant skills) LEVELL (Overall level of language; communication-verbal) และ RFACEX5 (Range of facial expressions used to communicate; reciprocal social interaction)

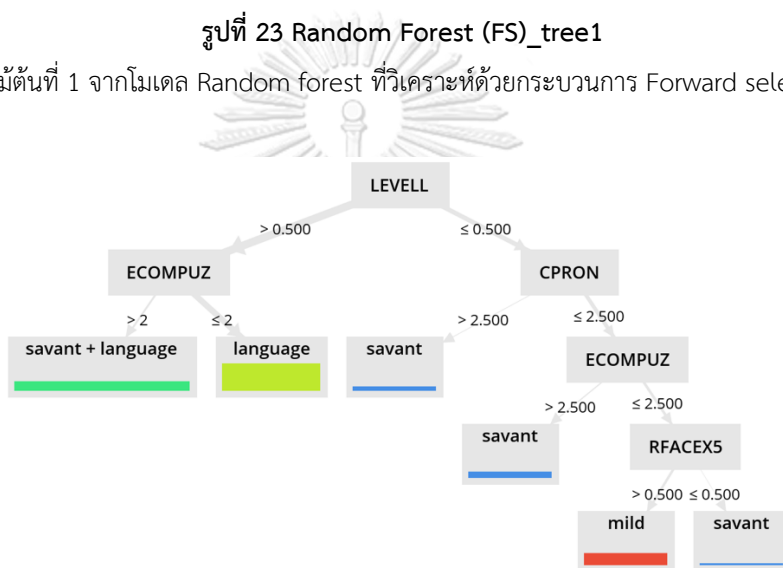
ตารางที่ 30 เปรียบเทียบประสิทธิภาพของโมเดลที่ดีที่สุดจากแต่ละอัลกอริทึม

| Algorithms | ANN | k-NN | Naive Bayes | Decision tree | Random forest | SVM | LibSVM |
|---------------------------|-------------------------------------------------------------|-----------------------------------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------------------------|--------------------------------------------------|--------------------------------------------------------|--------------------------------------------------------------|
| Analyzing process | FS | FS | FS | BE | FS | BE | FS |
| Special parameter applied | 1000 Training cycles | k=15 | Naive Bayes-kernel | - | 10 Trees | - | - |
| Accuracy | 98.82% | 100% | 98.82% | 97.65% | 100% | 87.06% | 100% |
| No. of ADI-R items | 6 | 8 | 5 | 122>>3 | 5 | 119 | 6 |
| Process analyzing time | 19.43 m | 43 s | 28 s | 20 s | 38 s | 4.28 m | 1.03 m |
| ADI-R items | CCONVER CINR CONVER5 CSPEECH ECOMPUZ EINAPPO | CIMIT CMUSICZ CSPEECH ECOMPUZ EINAPPO LEVELL QUALOV5 RFACEX5 | CCONVER CMEMZ CMUSICZ CSPEECH EDRAWZ | ตัดข้อ CCOMPUZ ออกไปก่อน และ เหลือข้อ LEVELL ECOMPUZ ESTEREO | CPRON CVERRIT ECOMPUZ LEVELL RFACEX5 | ทุกข้อยกเว้น CARTIC CNEOID CPEERPL CSOPLAY | CCOMPUZ CSPEECH EDRAWZ INITIA5 LEVELL RFACEX5 |



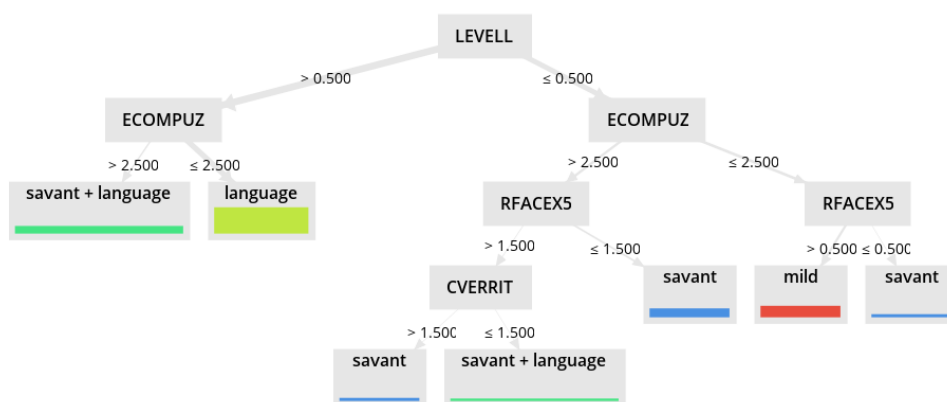
รูปที่ 23 Random Forest (FS)_tree1

ต้นไม้ต้นที่ 1 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



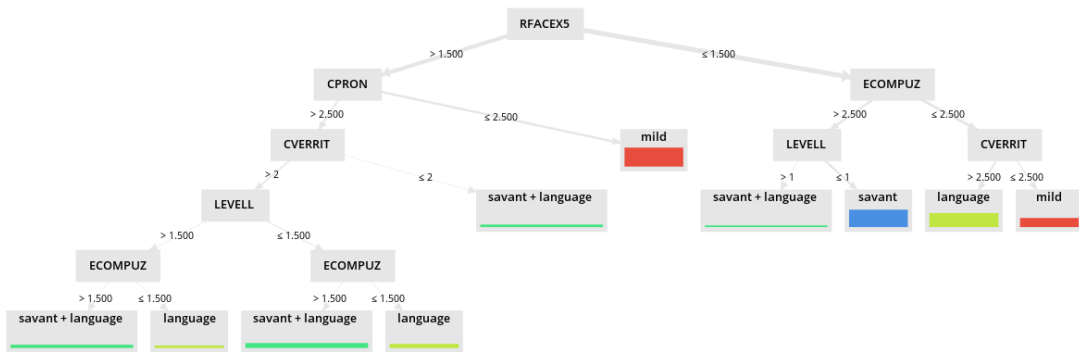
รูปที่ 24 Random Forest (FS)_tree2

ต้นไม้ต้นที่ 2 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



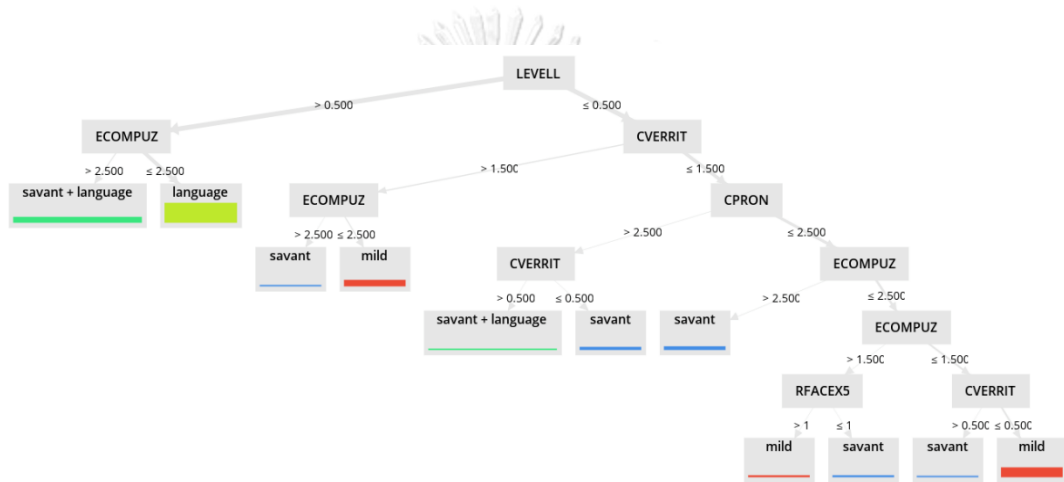
รูปที่ 25 Random Forest (FS)_tree3

ต้นไม้ต้นที่ 3 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



รูปที่ 26 Random Forest (FS)_tree4

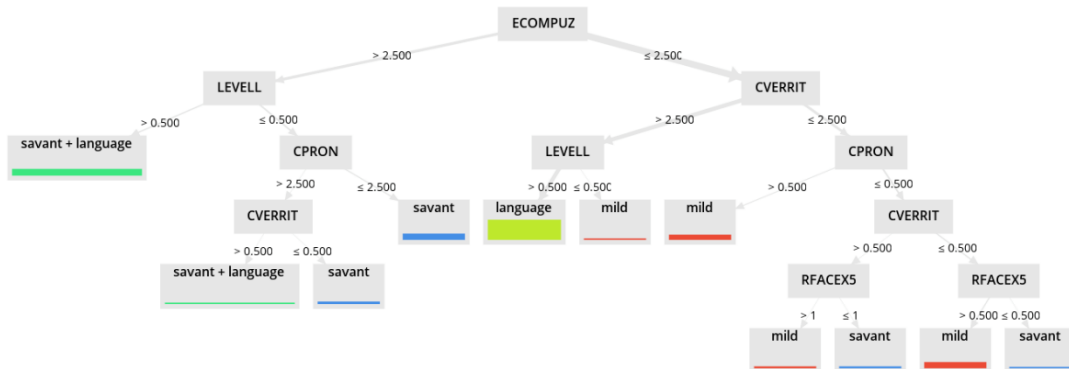
ต้นไม้ต้นที่ 4 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



รูปที่ 27 Random Forest (FS)_tree5

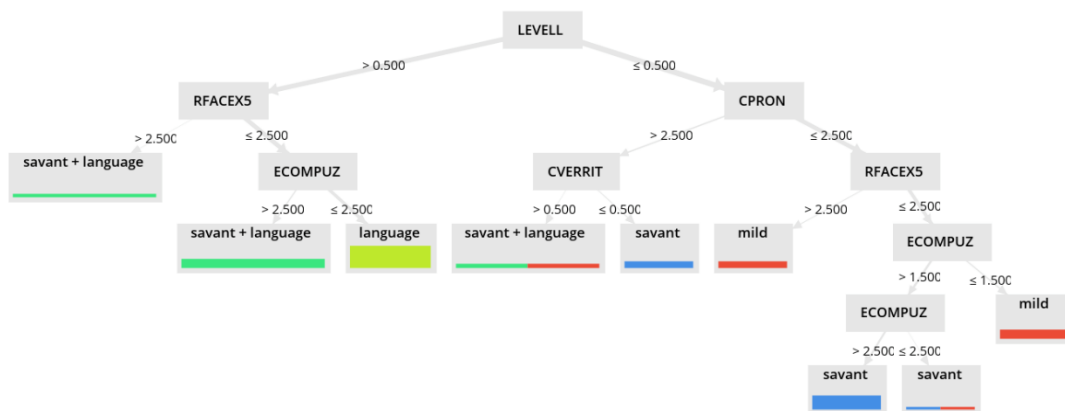
ต้นไม้ต้นที่ 5 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection

จุฬาลงกรณ์มหาวิทยาลัย



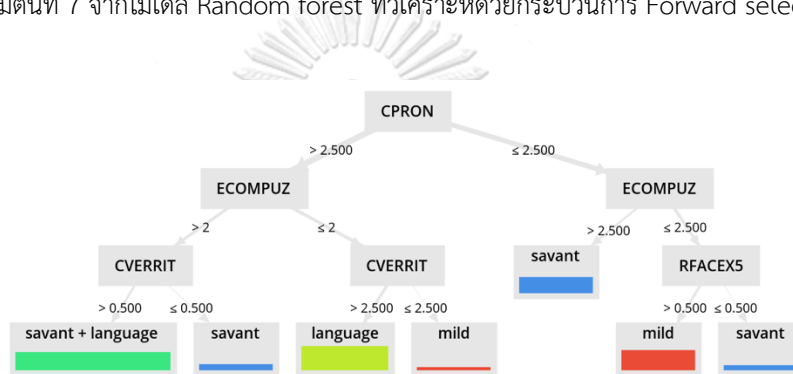
รูปที่ 28 Random Forest (FS)_tree6

ต้นไม้ต้นที่ 6 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



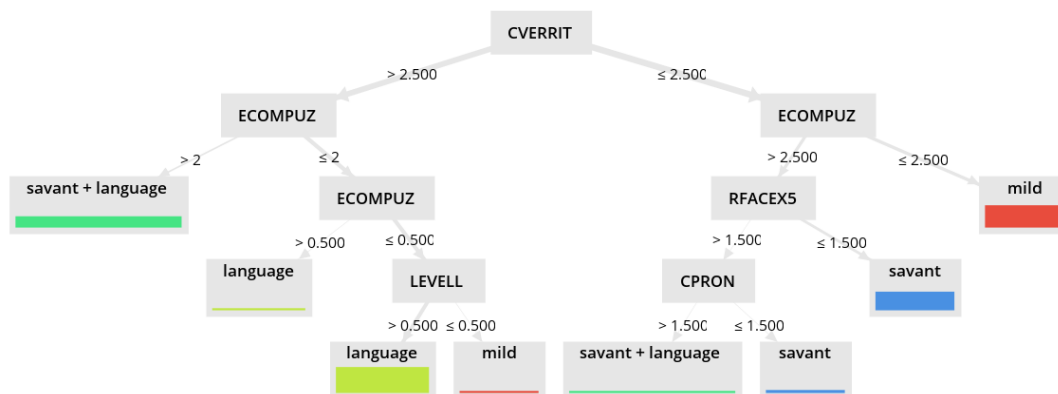
รูปที่ 29 Random Forest (FS)_tree7

ต้นไม้ต้นที่ 7 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



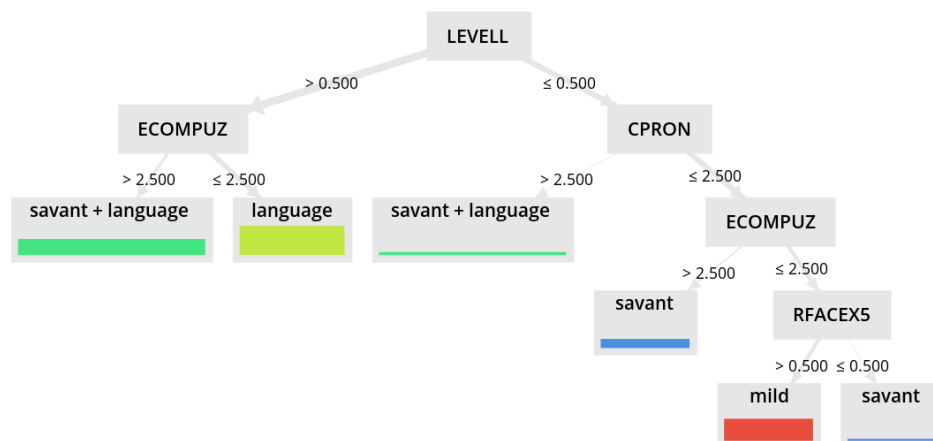
รูปที่ 30 Random Forest (FS)_tree8

ต้นไม้ต้นที่ 8 จากโมเดล Random Forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



รูปที่ 31 Random Forest (FS)_tree9

ต้นไม้ต้นที่ 9 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection



รูปที่ 32 Random Forest (FS)_tree10

ต้นไม้ต้นที่ 10 จากโมเดล Random forest ที่วิเคราะห์ด้วยกระบวนการ Forward selection

4.3 ผลการวิเคราะห์รูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

หลังจากที่ผู้วิจัยได้ทำการแบ่งผู้ป่วยออทิซึมสเปกตรัมทั้ง 85 ราย ออกเป็นกลุ่มย่อย 4 กลุ่ม ด้วยวิธี K-means และเปรียบเทียบผลลัพธ์การแบ่งกลุ่มกับการศึกษาของ Dr. Hu และ Dr. Steinberg ในปี 2009 (Hu & Steinberg, 2009) ตามหัวข้อ 4.1.3 แล้ว ผู้วิจัยได้ทำการศึกษารูปแบบทรานสคริปโตมของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ จากฐานข้อมูล GSE15402 ใน NCBI จากการศึกษาของ Dr. Hu ในปี 2009 (Hu et al., 2009) ซึ่งเป็นการศึกษารูปแบบการแสดงออกของยีนโดยใช้เทคนิค cDNA microarrays จำนวน 39,936 ยีน และมี Probe สำหรับการศึกษารูปแบบการแสดงออกของ RNA ทั้งหมด 42,000 Transcripts โดยคัดเลือกข้อมูลจากกลุ่มผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย C หรือ Mild subgroup (M subgroup) จำนวน 9 ราย กลุ่มย่อย D หรือ Savant subgroup (S subgroup) จำนวน 13 ราย และกลุ่มควบคุมจำนวน 7 ราย ที่มีเพศเดียวกัน และมีอายุใกล้เคียงกันกับกลุ่ม (Age- and sex-matched controls) ด้วยวิธีการทางสถิติ T-test แบบ Adjusted Bonferroni correction ($P < 0.05$) โดยจำนวน Transcripts ที่มีความแตกต่างอย่างมีนัยสำคัญในแต่ละกลุ่มเมื่อเปรียบเทียบกับกลุ่มควบคุม และจำนวนโปรตีนที่ผลิตโดย Transcripts เหล่านี้ ได้แสดงไว้ดังตารางที่ 31

ตารางที่ 31 จำนวน Transcripts ที่มีความแตกต่างอย่างมีนัยสำคัญ และจำนวนโปรตีนที่ผลิตโดย Transcripts ในแต่ละกลุ่มย่อยเมื่อเปรียบเทียบกับกลุ่มควบคุม

| Comparison | #Differentially expressed transcripts | #Corresponding protein products |
|----------------------------|---------------------------------------|---------------------------------|
| Control vs M+S subgroup | 266 | 147 |
| Control vs Mild subgroup | 544 | 351 |
| Control vs Savant subgroup | 276 | 158 |

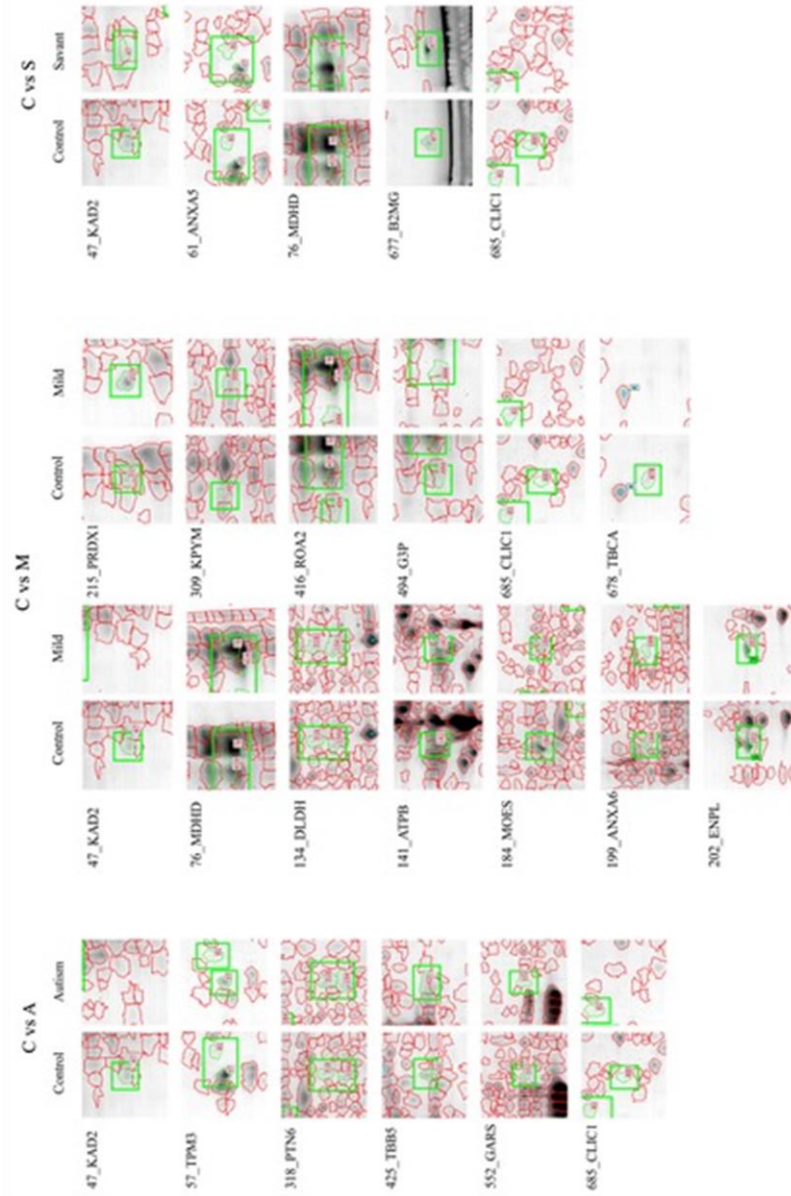
4.4 ผลการวิเคราะห์รูปแบบโปรตีนของเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

เพื่อศึกษารูปแบบการแสดงออกของโปรตีนในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมเปรียบเทียบกับรูปแบบการแสดงออกในเซลล์เพาะเลี้ยงชนิดที่มาจากคนปกติ และระบุโปรตีนที่มีระดับการแสดงออกผิดปกติในกลุ่มย่อยของโรคออทิซึมสเปกตรัม ผู้วิจัยได้ทำการศึกษารูปแบบโปรตีนในผู้ป่วยออทิซึมสเปกตรัมเปรียบเทียบกับคนปกติที่มีอายุใกล้เคียงกันกับผู้ป่วย (age-matched control) โดยคัดเลือกกลุ่มของผู้ป่วยกลุ่มย่อย C หรือกลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) จำนวน 9 ราย กลุ่มย่อย D หรือกลุ่มย่อยที่มีทักษะอัจฉริยะภาพพิเศษ (Savant skills subgroups; S subgroup) จำนวน 13 ราย และกลุ่มควบคุมจำนวน 7 ราย อาศัยเทคนิค Two Dimensional Gel Electrophoresis (2D-GE) เพื่อทำการวิเคราะห์โปรตีนในช่วง pH 3-10 โดยในแต่ละกลุ่มจะทำการทดสอบทั้งหมด 3 ซ้ำ (Triplicate) ซึ่งโปรตีนที่ใช้ทดสอบสกัดมาจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่ถูกเพาะเลี้ยงตามระเบียบวิธีการของ Rutgers University Cell and DNA repository ผู้ผลิตและดูแลรักษาเซลล์ใน Autism Genetic Research Exchange (AGRE) collection และได้ทำการวัดปริมาณและตรวจสอบคุณภาพโปรตีนก่อนทำการทดสอบ โดยเมื่อดำเนินการแล้วแผ่นเจลภายหลังการทดสอบได้ถูกนำไปวิเคราะห์ความแตกต่างของการแสดงออกของโปรตีนด้วยโปรแกรม Image Master 2D Platinum version 7.0

จากการศึกษาพบว่าปริมาณจุดโปรตีนทั้งสิ้น 686 จุดต่อแผ่นเจล จากแผ่นเจลในกลุ่ม Control กลุ่ม ASD (M+S) กลุ่ม M และกลุ่ม S ซึ่งเมื่อทำการวิเคราะห์ด้วยสถิติ F-test, ANOVA ที่ $P\text{-value} < 0.05$ พบว่ามีจุดโปรตีนที่มีค่าสัดส่วนความเข้มแตกต่างกันอย่างมีนัยสำคัญระหว่างทั้ง 4 กลุ่ม มีเพียง 20 จุดโปรตีนต่อกลุ่ม จึงได้ทำการวิเคราะห์สถิติ ANOVA ที่ $P\text{-value} < 0.05$ แยกแต่ละกลุ่มเทียบกับกลุ่มควบคุม พบว่าในกลุ่ม ASD (M+S subgroup) มีจุดโปรตีนที่มีค่าสัดส่วนความเข้มแตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับกลุ่ม control 10 จุดโปรตีน กลุ่ม M subgroup มีจุดโปรตีนที่มีค่าสัดส่วนความเข้มแตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับกลุ่ม control 18 จุดโปรตีน และกลุ่ม S subgroup มีจุดโปรตีนที่มีค่าสัดส่วนความเข้มแตกต่างกันอย่างมีนัยสำคัญเมื่อเทียบกับกลุ่ม control 9 จุดโปรตีน จากนั้นจึงทำการคัดเลือกจุดโปรตีนดังกล่าวอีกครั้ง โดยให้มีค่าสัดส่วนความเข้มแตกต่างกันลดลงหรือเพิ่มขึ้นอย่างน้อย 2 เท่า พบว่ามีจุดโปรตีนที่มีการแสดงออกผิดปกติอย่างมีนัยสำคัญในกลุ่ม ASD (M+S subgroup) 6 จุด กลุ่ม M subgroup 13 จุด และกลุ่ม S subgroup 5 จุด ดังแสดงในตารางที่ 32 และรูปที่ 32 โดยจากโปรตีนทั้งหมดพบว่ามี 14 โปรตีนที่เคยมีรายงานในฐานข้อมูลออร์ทิมสเปกตรัม AutismKB database ดังตารางที่ 33

ตารางที่ 32 รายชื่อจุดโปรตีนที่มีการแสดงออกผิดปกติอย่างมีนัยสำคัญในเซลล์เพาะเลี้ยงชนิดลิโมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัม
กลุ่มย่อย

| Group | Protein | Description | Fold-change (ASD gr./Control) | P-value (ANOVA; ASD gr. vs Control) | MASCOT scores | %Coverage | Mass | pI |
|--------------------------|---------|--------------------------------------------------|----------------------------------|----------------------------------------|------------------|-----------|-------|------|
| ASD (M+S subgroup) | AK2 | Adenylate kinase 2 | ↓0.039 | 4.13E-03 | 1032 | 33% | 26689 | 7.67 |
| | PTPN6 | Tyrosine-protein phosphatase non-receptor type 6 | ↑6.852 | 5.28E-03 | 664 | 23% | 67918 | 7.65 |
| | CLIC1 | Chloride intracellular channel protein 1 | ↓0.020 | 5.43E-03 | 1415 | 25% | 27248 | 5.09 |
| | GARS | Glycine-tRNA ligase | ↑7.499 | 1.22E-02 | 487 | 20% | 83854 | 6.61 |
| | TUBB | Tubulin beta chain | ↑9.721 | 2.77E-02 | 1584 | 29% | 50095 | 4.78 |
| | TPM3 | Tropomyosin alpha-3 | ↓0.457 | 3.78E-02 | 994 | 31% | 32987 | 4.68 |
| | TBCA | Tubulin-specific chaperone | ↓0.045 | 2.36E-03 | 348 | 36% | 12904 | 5.25 |
| | AK2 | Adenylate kinase 2 | ↓0.036 | 4.13E-03 | 1032 | 33% | 26689 | 7.67 |
| | CLIC1 | Chloride intracellular channel protein 1 | ↓0.021 | 5.43E-03 | 1415 | 25% | 27248 | 5.09 |
| | PRDX1 | Peroxiredoxin-1 | ↑2.174 | 6.89E-03 | 1158 | 56% | 22324 | 8.27 |
| M subgroup | ROA2 | Heterogeneous nuclear ribonucleoproteins A2/B1 | ↑6.562 | 9.62E-03 | 1330 | 48% | 37464 | 8.97 |
| | ENPL | Endoplasmic | ↓0.329 | 1.28E-02 | 1345 | 27% | 92696 | 4.76 |
| | MDH2 | Malate dehydrogenase, mitochondria | ↓0.399 | 2.23E-02 | 1446 | 52% | 35937 | 8.92 |
| | MOES | Moesin | ↓0.276 | 2.24E-02 | 521 | 32% | 67892 | 6.08 |



รูปที่ 33 จุดโปรตีนที่แสดงออกผิดปกติอย่างมีนัยสำคัญในเซลล์เพาะเลี้ยงของผู้ป่วยออทิซึมสติปดรีมและออทิซึมสติปดรีมกลุ่มย่อย

แสดงจุดโปรตีนที่แสดงออกผิดปกติจากการวิเคราะห์ด้วยเทคนิค 2D-GE ในเซลล์เพาะเลี้ยงของผู้ป่วยออทิซึมสติปดรีมและออทิซึมสติปดรีมกลุ่มย่อยเทียบกับคนปกติ

วงกลมสีแดงแสดงถึงจุดโปรตีนที่พบบนแผ่นเจล ในขณะที่วงกลมสีเขียวแสดงถึงจุดโปรตีนที่แสดงออกผิดปกติอย่างมีนัยสำคัญ (ANOVA; P-value<0.05)

ตารางที่ 33 หน้า that ทางชีวภาพและลักษณะอาการที่เกี่ยวข้องกับโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัม

| No. | Protein | Description | GenesCards functions/Phenotypes | AutismKB |
|-----|----------------|--------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----------|
| 1 | AK2 (KAD2) | Adenylate kinase 2 | Plays an important role in cellular energy homeostasis and in adenine nucleotide metabolism. Abnormality in this gene can cause hearing impairment. | ✓ |
| 2 | CLIC1 | Chloride intracellular channel protein 1 | Can insert into membranes and form chloride ion channels. Involved in regulation of the cell cycle. | ✓ |
| 3 | TPM3 | Tropomyosin alpha-3 | Plays a central role, in association with the troponin complex, in the calcium dependent regulation of vertebrate striated muscle contraction. Abnormality in this gene can cause the neurological abnormality, neurodevelopmental delay, or abnormal synaptic transmission. | ✓ |
| 4 | PTPN6 | Tyrosine-protein phosphatase non-receptor type 6 | Modulates signaling by tyrosine phosphorylated cell surface receptors such as KIT and the EGF receptor/EGFR. | ✓ |
| 5 | GARS | Glycine-tRNA ligase | Catalyzes the ATP-dependent ligation of glycine to the 3'-end of its cognate tRNA, via the formation of an aminoacyl(-adenylate intermediate (Gly-AMP). Mutation in this gene can cause an abnormality of the nervous system. | - |
| 6 | TUBB (TBB5) | Tubulin beta chain | Tubulin is the major constituent of microtubules. Abnormality in this gene can cause abnormality of neuronal migration, abnormality of brain morphology, or neurodevelopmental delay. | ✓ |
| 7 | MDH2 | Malate dehydrogenase, mitochondria | This protein is localized to the mitochondria and may play pivotal roles in the malate-aspartate shuttle that operates in the metabolic coordination between cytosol and mitochondria. Mutation in this gene can cause abnormality of the nervous system, neurological speech impairment, or neurodevelopmental delay. | ✓ |
| 8 | TBCA | Tubulin-specific chaperone | Involved in the early step of the tubulin folding pathway. | ✓ |
| 9 | MOES | Moesin | ERM proteins appear to function as cross-linkers between plasma membranes and actin-based cytoskeletons. Moesin is localized to filopodia and other membranous protrusions that are important for cell-cell recognition and signaling and for cell movement. | ✓ |

| | | | | |
|----|------------------|------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| 10 | ENPL (HS90B1) | Endoplasmic reticulum | This protein is a member of a family of adenosine triphosphate (ATP)-metabolizing molecular chaperones with roles in stabilizing and folding other proteins. The encoded protein is localized to melanosomes and the endoplasmic reticulum. | ✓ |
| 11 | ATPB | ATP synthase subunit beta | This protein is a subunit of mitochondrial ATP synthase. Mitochondrial ATP synthase catalyzes ATP synthesis, utilizing an electrochemical gradient of protons across the inner membrane during oxidative phosphorylation. | ✓ |
| 12 | PRDX1 | Peroxiredoxin-1 | This protein is a member of the peroxiredoxin family of antioxidant enzymes, which reduce hydrogen peroxide and alkyl hydroperoxides. Abnormality in this gene can cause abnormality of the nervous system, abnormality of movement, or neurodevelopmental delay. | - |
| 13 | ANXA6 | Annexin A6 | Annexin VI belongs to a family of calcium-dependent membrane and phospholipid binding proteins. | - |
| 14 | DLD | Dihydrolipoyl dehydrogenase, mitochondria | The encoded protein has been identified as a moonlighting protein. In homodimeric form, the encoded protein functions as a dehydrogenase and is found in several multi-enzyme complexes that regulate energy metabolism. However, as a monomer, this protein can function as a protease. Abnormality in this gene can cause microcephaly, visual impairment, or global developmental delay. | ✓ |
| 15 | G3P | Glyceraldehyde-3-phosphate dehydrogenase | Has both glyceraldehyde-3-phosphate dehydrogenase and nitrosylase activities, thereby playing a role in glycolysis and nuclear functions, respectively. | ✓ |
| 16 | ROA2 | Heterogeneous nuclear ribonucleoproteins A2/B1 | The hnRNPs are RNA binding proteins and they complex with heterogeneous nuclear RNA (hnRNA). These proteins are associated with pre-mRNAs in the nucleus and appear to influence pre-mRNA processing and other aspects of mRNA metabolism and transport. Mutation in this gene can cause abnormality of brain morphology, abnormality of the nervous system, or neurological speech impairment. | - |
| 17 | KPYM | Pyruvate kinase | Glycolytic enzyme that catalyzes the transfer of a phosphoryl group from phosphoenolpyruvate (PEP) to ADP, generating ATP. | ✓ |

| | | | | |
|----|-------|----------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---|
| 18 | ANXA5 | Annexin A5 | This protein is an anticoagulant protein that acts as an indirect inhibitor of the thromboplastin-specific complex, which is involved in the blood coagulation cascade. Mutation in this gene can cause an abnormality of the fetus or the birth of the fetus, excluding structural abnormalities. | - |
| 19 | B2M | Beta-2-microglobulin | Component of the class I major histocompatibility complex (MHC). Involved in the presentation of peptide antigens to the immune system. Abnormality in this gene can cause abnormality of the nervous system. | ✓ |



4.5 ผลการวิเคราะห์ข้อมูลทางชีวสารสนเทศจากรายชื่อโปรตีนที่พบว่ามีการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย

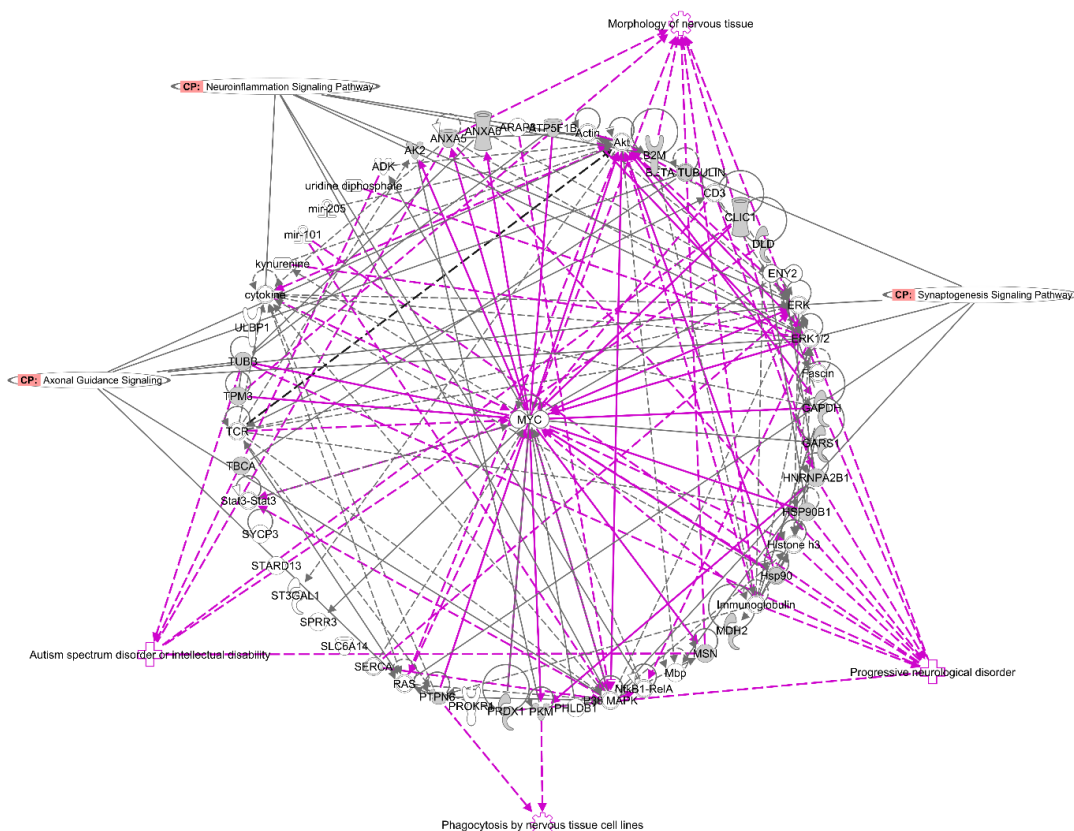
หลังจากที่ทำการศึกษารูปแบบการแสดงออกของโปรตีนในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยและระบุโปรตีนที่มีระดับการแสดงออกผิดปกติแล้ว เพื่อค้นหากลุ่มโรคทางระบบประสาทหรือพยาธิสภาพที่เกี่ยวข้องกับโรคออทิซึมสเปกตรัม และหน้าที่ทางชีวภาพ ตลอดจนสร้างเครือข่ายทางชีวภาพที่แสดงความเกี่ยวข้องกับกลุ่มโปรตีนดังกล่าว ผู้วิจัยได้นำรายชื่อโปรตีนที่พบทั้ง 19 ตัว ไปทำการวิเคราะห์ Gene ontology analysis ในโปรแกรม Ingenuity Pathway Analysis (IPA) เพื่อทำนายหน้าที่ทางชีวภาพของโปรตีนด้วยนัยสำคัญทางสถิติที่ $P\text{-value} < 0.05$ โดยใช้ Fisher's Exact test พบว่าโปรตีนทั้ง 19 ตัว มีความเกี่ยวข้องกับโรคพัฒนาการบกพร่อง (Developmental disorder) โดยเฉพาะอย่างยิ่งโรคออทิซึมสเปกตรัม (Autism spectrum disorder or intellectual disability) ที่ นัยสำคัญ $8.17E-04$ ถึง 5 โปรตีน และเกี่ยวข้องกับภาวะบกพร่องทางพัฒนาการอื่น ๆ อย่าง GARS-associated growth retardation and developmental delay และ Pervasive developmental disorder อย่างมีนัยสำคัญอีกด้วย นอกจากนี้โปรตีนที่แสดงออกผิดปกติยังมีความเกี่ยวข้องกับโรคทางระบบประสาท (Neurological disease) อย่างเช่น Neurological signs, Synaptic dysfunction of neuromuscular junctions, and Progressive neurological disorder อย่างมีนัยสำคัญ รวมถึงโปรตีนดังกล่าวยังมีความเกี่ยวข้องกับหน้าที่และพัฒนาการของระบบประสาท (Nervous system development and function) ไม่ว่าจะเป็น Morphology of neurons, Mitogenesis of astrocytes, and Size of presynaptic terminals อย่างมีนัยสำคัญดังแสดงในตารางที่ 34

ตารางที่ 34 ผลการทำนายโรคและหน้าที่ทางชีววิทยาของกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย

| Diseases and functions | P-value | #DEPS |
|------------------------------------------------------------|----------|-------|
| Developmental Disorder | | |
| Autism spectrum disorder or intellectual disability | 8.17E-04 | 5 |
| GARS-associated growth retardation and developmental delay | 8.36E-04 | 1 |
| Pervasive developmental disorder | 1.58E-03 | 3 |
| Neurological Disease | | |
| Neurological signs | 1.49E-03 | 5 |
| Synaptic dysfunction of neuromuscular junctions | 1.67E-03 | 1 |
| Progressive neurological disorder | 2.14E-03 | 6 |
| Nervous system development and function | | |
| Morphology of neurons | 8.82E-04 | 5 |
| Mitogenesis of astrocytes | 3.34E-03 | 1 |
| Size of presynaptic terminals | 4.17E-03 | 1 |

และในการศึกษาปฏิสัมพันธ์ของโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย ผู้วิจัยได้ทำการสร้างเครือข่ายทางชีวภาพของความสัมพันธ์ระหว่างโปรตีนต่อโปรตีน (Interactome) ของโปรตีนที่ผิดปกติทั้งหมด ดังรูปที่ 33 พบว่ามีความเกี่ยวข้องกับโรคออทิซึมสเปกตรัม (Autism spectrum disorder or intellectual disability) โรคทางระบบประสาทแบบก้าวหน้า (Progressive neurological disorder) การ Phagocytosis โดยเซลล์เพาะเลี้ยงชนิดเนื้อเยื่อประสาท (Phagocytosis by nervous tissue cell lines) และรูปร่างของเนื้อเยื่อประสาท (Morphology of

nervous tissue) รวมถึงสัมพันธ์กับ Canonical pathway ที่เกี่ยวข้องกับการทำงานของระบบประสาทและสมองที่สำคัญอย่าง Neuroinflammation signaling pathway, Axonal guidance signaling, and Synaptogenesis signaling pathway โดยมีโมเลกุลของ MYC เป็นโมเลกุลที่เชื่อมโยงโมเลกุลต่าง ๆ ในเครือข่าย



รูปที่ 34 ผลการทำนายเครือข่ายทางชีวภาพของโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย

แสดงปฏิสัมพันธ์ระหว่างโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย ต่อหน้าที่ที่เกี่ยวข้องกับระบบประสาทและสมอง จากการวิเคราะห์ด้วยโปรแกรม

IPA (Gray = Input molecules; ⚙️ = Function, ⚖️ = Phosphatase, 🚚 = Transporter, 🏠 = Transmembrane receptor, 🚪 = Ion channel, 🧪 = Enzyme, ☒ = Chemical/Toxicant, ○ = Complex/Group/Other; — = Direct interaction, - - - = Indirect interaction)

4.6 ผลการศึกษาความสอดคล้องและความสัมพันธ์ของกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคออทิซึมสเปกตรัม

จากการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย กับกลุ่มยีนที่แสดงออกผิดปกติในกลุ่มตัวอย่างเดียวกัน จาก GSE15402 หัวข้อ 4.3 โดยใช้ Venn diagram และ Hypergeometric distribution analysis พบโปรตีนที่แสดงออกผิดปกติจากทั้งสองรายชื่อเพียงโปรตีนเดียว คือ TPM3 ในกลุ่ม ASD (M+S subgroup) ที่ระดับนัยสำคัญ $4.37E-02$ แต่ไม่พบโปรตีนที่แสดงออกผิดปกติจากทั้งสองรายชื่อในกลุ่ม M subgroup และกลุ่ม S subgroup

นอกจากนี้ผู้วิจัยได้วิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิซึมสเปกตรัม กับรายชื่อยีนที่แสดงออกผิดปกติจากเซลล์เม็ดเลือดของผู้ป่วยออทิซึมสเปกตรัมที่ได้จากฐานข้อมูล GEO DataSets ใน NCBI จำนวนทั้งสิ้น 6 การศึกษา พบโปรตีนที่แสดงออกผิดปกติที่มีความสัมพันธ์กับยีนที่แสดงออกผิดปกติจากเซลล์เม็ดเลือดของผู้ป่วยออทิซึมสเปกตรัม ในการศึกษาจาก GSE18123 จำนวน 3 โปรตีน (HNRNPA2B1, MSN, and DLD) GSE26415 จำนวน 7 โปรตีน (B2M, CLIC1, HNRNPA2B1, PRDX1, AK2, TBCA, and DLD) GSE42133 จำนวน 1 โปรตีน (B2M) และ GSE111175 จำนวน 7 โปรตีน (TPM3, TUBB, HNRNPA2B1, PTPN6, PRDX1, TBCA, and DLD)

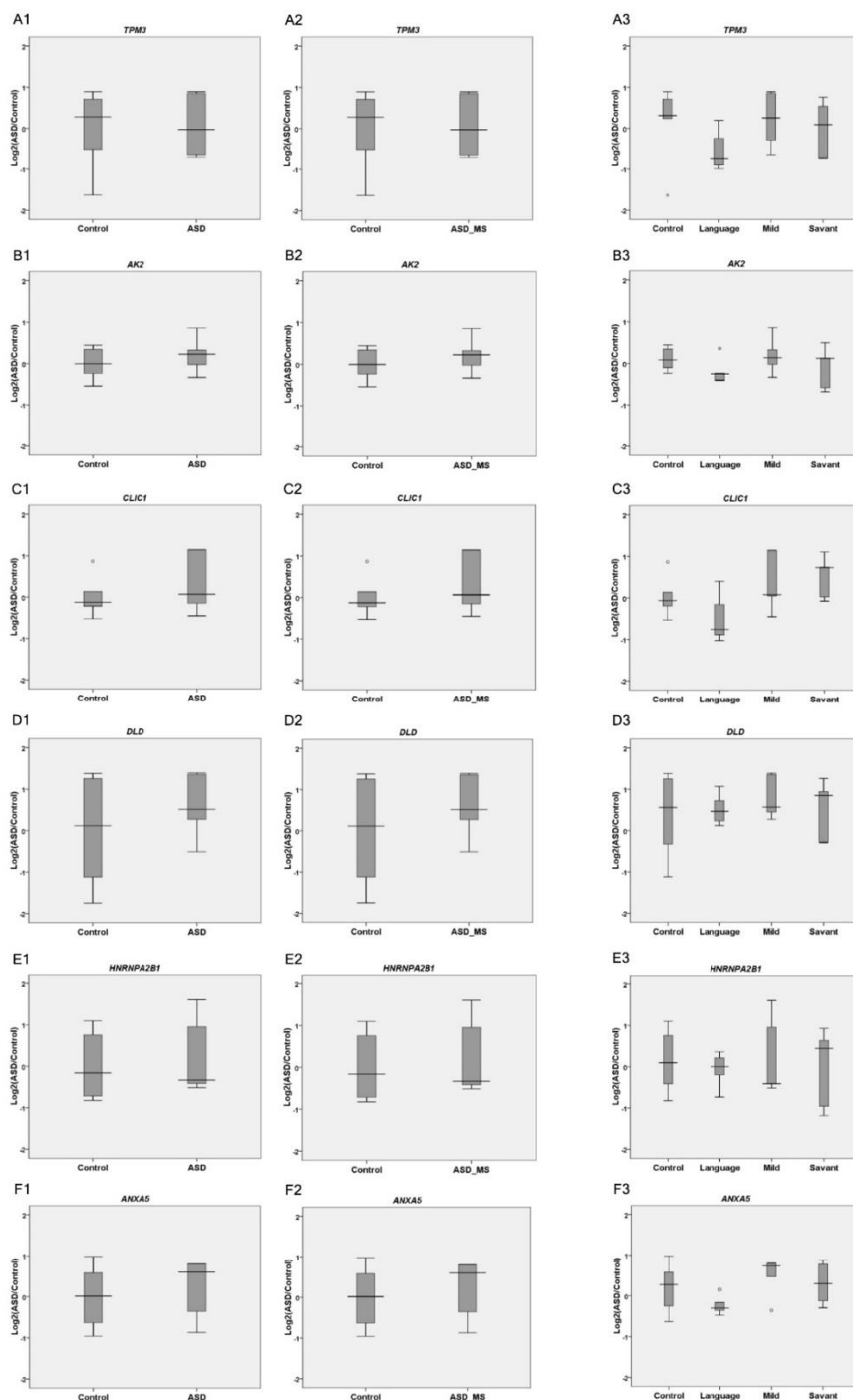
และจากการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย กับโปรตีนที่เคยมีรายงานว่าพบการแสดงออกผิดปกติในเลือดของผู้ป่วยออทิซึมสเปกตรัมในฐานข้อมูลชีวสารสนเทศ จำนวนทั้งสิ้น 7 การศึกษา พบโปรตีนที่แสดงออกผิดปกติที่มีความสัมพันธ์กับโปรตีนที่เคยมีรายงานว่าพบการแสดงออกผิดปกติในเลือดของผู้ป่วย ในกลุ่ม ASD (M+S subgroup) คือ TUBB กลุ่ม M subgroup คือ HSP90B1, MDH2, and DLD และกลุ่ม S subgroup คือ MDH2 และ ANXA5 ที่ระดับนัยสำคัญ $2.30E-02$ $3.46E-05$ และ $2.51E-04$ ตามลำดับ ดังแสดงในตารางที่ 35

ตารางที่ 35 ผลการวิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่ค้นพบว่าแสดงออกผิดปกติใน เซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยโรคอหิวาต์และอหิวาต์สเปกตรัมกลุ่มย่อย กับกลุ่มยีนจากกลุ่มตัวอย่างเดียวกัน และกลุ่มยีนและโปรตีนจากการศึกษาอื่น

| Overlap types | DEPs subgroup | Overlap with studies (No. of gene lists) | P-value from hypergeometric analysis | #Overlapping proteins | Protein symbols |
|---------------------------------------------------------|---------------|------------------------------------------|--------------------------------------|-----------------------|------------------------------------------------|
| ส่วนที่ 1 DEGs (GSE15402) vs DEPs | M+S subgroup | GSE15402 (M+S subgroup) | 4.37E-02 | 1 | TPM3 |
| | M subgroup | GSE15402 (M subgroup) | 1.00 | 0 | - |
| | S subgroup | GSE15402 (S subgroup) | 1.00 | 0 | - |
| ส่วนที่ 2 DEGs จาก การศึกษาก่อนหน้านี้ vs DEPs | All DEPs | GSE6575 (226) | 1.00 | 0 | - |
| | | GSE18123 (2123) | 0.334 | 3 | HNRNPA2B1 MSN DLD |
| | | GSE25507 (1712) | 1.00 | 0 | - |
| | | GSE26415 (3420) | 3.42E-02 | 7 | B2M CLIC1 HNRNPA2B1 PRDX1 AK2 TBCA DLD |
| | | GSE42133 (2082) | 0.879 | 1 | B2M |
| | | GSE111175 (2313) | 4.17E-03 | 7 | TPM3 TUBB HNRNPA2B1 PTPN6 PRDX1 TBCA DLD |
| ส่วนที่ 3 DEPs จาก การศึกษาก่อนหน้านี้ vs DEPs | M+S subgroup | Previous proteomic studies | 2.30E-02 | 1 | TUBB |
| | M subgroup | | 3.46E-05 | 3 | HSP90B1 MDH2 DLD |
| | S subgroup | | 2.51E-04 | 2 | MDH2 ANXA5 |

4.7 ผลการยืนยันระดับการแสดงออกของยีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟ บลาสต์จากผู้ป่วยออทิซึมสเปกตรัม

จากผลการศึกษาความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย กับกลุ่มยีนจากกลุ่มตัวอย่างเดียวกัน และกลุ่มยีนและโปรตีนในเซลล์เม็ดเลือดของผู้ป่วยจากการศึกษาอื่น ผู้วิจัยได้ทำการเลือกยีนจากบทบาทหน้าที่ทางชีววิทยา ระดับความเชื่อมั่น และทิศทางที่สัมพันธ์กันระหว่างรูปแบบทรานสคริปโตมและรูปแบบโปรตีโอม จำนวนทั้งสิ้น 6 ยีน ได้แก่ *TPM3*, *AK2*, *CLIC1*, *DLD*, *HNRNPA2B1*, and *ANXA5* โดยเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่มผู้ป่วยออทิซึมสเปกตรัมและกลุ่มควบคุม (กลุ่ม L subgroup 5 ราย กลุ่ม M subgroup 6 ราย กลุ่ม S subgroup 8 ราย และกลุ่มควบคุม 6 ราย) ได้ถูกนำมาสกัด RNA เพื่อยืนยันการแสดงออกด้วยเทคนิค RT-qPCR พบว่าเมื่อเปรียบเทียบระดับการแสดงออกของยีนทั้ง 6 ยีนข้างต้น ไม่พบการเปลี่ยนแปลงระดับแสดงออกของยีนอย่างมีนัยสำคัญทั้งในกลุ่มผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเมื่อเทียบกับกลุ่มควบคุม ดังแสดงในรูปที่ 34



รูปที่ 35 Box plot แสดงผลการแสดงออกของยีนจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึม สเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับกลุ่มควบคุม

ผลการแสดงออกของยีน (A) *TPM3* (B) *AK2* (C) *CLIC1* (D) *DLD* (E) *HNRNPA2B1* และ (F) *ANXA5*

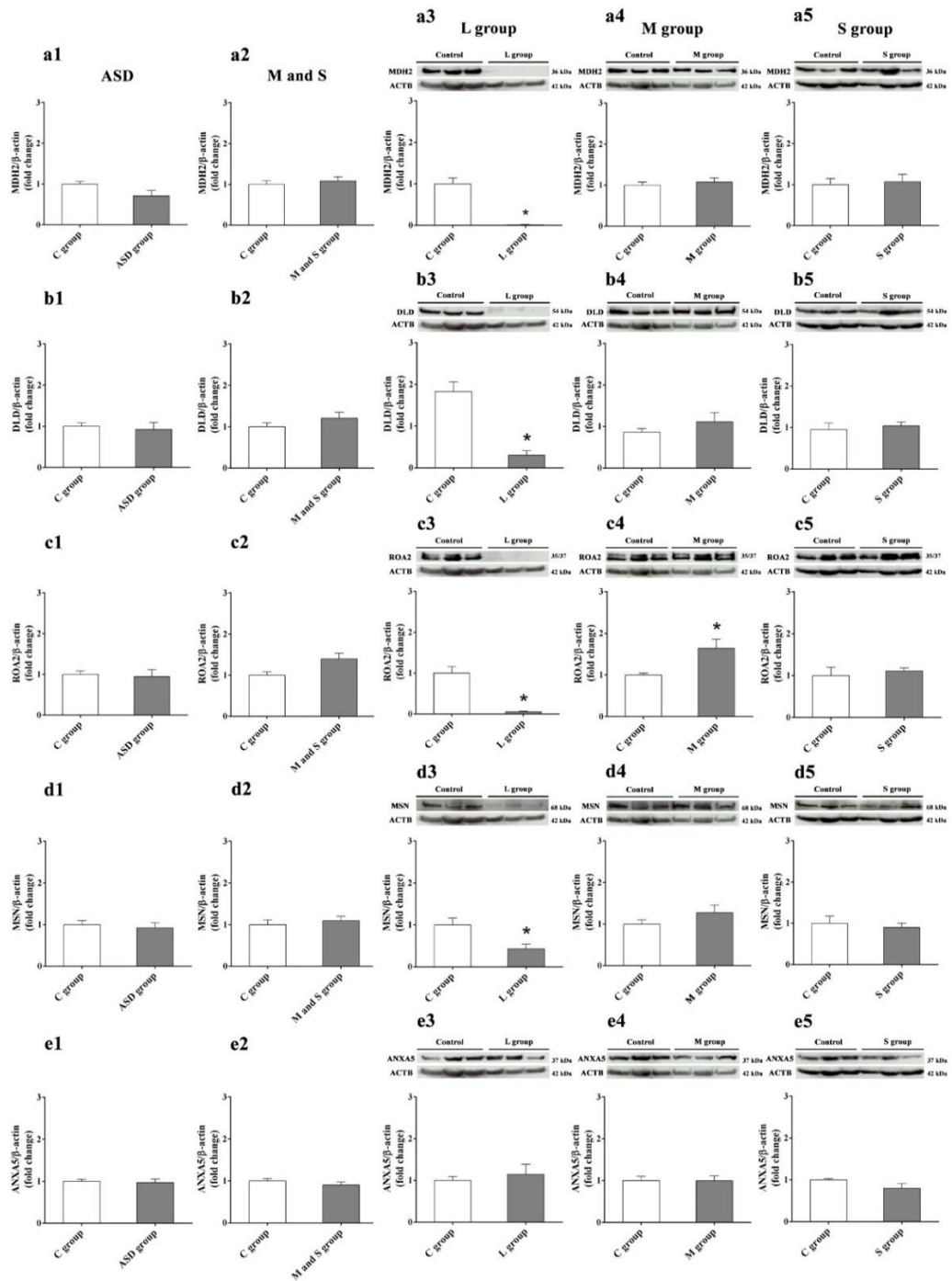
จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย (ASD n=19;

ASD_MS n=14; L subgroup n=5, M subgroup n=6, S subgroup n=8) เทียบกับกลุ่มควบคุม

(Control n=6) โดยใช้สถิติ two-tailed Student's T-test ที่ P-value<0.05

4.8 ผลการยืนยันระดับการแสดงออกของโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัม

เมื่อไม่พบความแตกต่างอย่างมีนัยสำคัญของการแสดงออกของยีนในกลุ่มผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเมื่อเทียบกับกลุ่มควบคุม ดังนั้นผู้วิจัยจึงมุ่งเน้นไปที่การยืนยันการแสดงออกของโปรตีนจากการศึกษารูปแบบโปรตีโอมที่แสดงออกผิดปกติแทน โดยได้คัดเลือกโปรตีนจากบทบาทหน้าที่ทางชีววิทยา ระดับความเชื่อมั่น และความสัมพันธ์กันระหว่างรูปแบบโปรตีโอมกับรูปแบบทรานสคริปโตม และกลุ่มยีนและโปรตีนจากการศึกษาก่อนหน้านี้ จำนวน 5 โปรตีน ได้แก่ MDH2 DLD HNRNPA2B1 (ROA2) MSN และ ANXA5 โดยผู้วิจัยได้นำโปรตีนจากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่มผู้ป่วยออทิซึมสเปกตรัมและกลุ่มควบคุม (กลุ่ม L subgroup 6 ราย กลุ่ม M subgroup 6 ราย กลุ่ม S subgroup 6 ราย และกลุ่มควบคุม 5 ราย) มาทำการศึกษาระดับการแสดงออกของโปรตีนด้วยเทคนิค Western blot พบว่ามีโปรตีน HNRNPA2B1 (ROA2) ที่มีระดับการแสดงออกเพิ่มขึ้นอย่างมีนัยสำคัญ ($P < 0.05$) ในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย M subgroup เมื่อเทียบกับกลุ่มควบคุม และพบโปรตีน 4 ตัว ได้แก่ MDH2 DLD HNRNPA2B1 (ROA2) และ MSN ที่มีระดับการแสดงออกลดลงอย่างมีนัยสำคัญ ($P < 0.05$) ในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย L subgroup แต่ไม่พบการเปลี่ยนแปลงของระดับการแสดงออกของโปรตีนทั้ง 5 ตัว ในผู้ป่วยออทิซึมสเปกตรัมแบบรวมกลุ่มย่อย และในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย S subgroup ดังแสดงในรูปที่ 35



รูปที่ 36 Bar graph แสดงระดับการแสดงออกของโปรตีนจากเซลล์เพาะเลี้ยงชนิดลิ้มโพลาสตีในผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับกลุ่มควบคุม

ระดับการแสดงออกของโปรตีน (a) MDH2 (b) DLD (c) ROA2 (d) MSN (e) ANXA5 จากเซลล์เพาะเลี้ยงชนิดลิ้มโพลาสตีในผู้ป่วยออทิซึมสเปกตรัม และผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย (ASD n=18; ASD_MS n=12;

L subgroup n=6, M subgroup n=6, S subgroup n=6) เทียบกับกลุ่มควบคุม (Control n=5)

โดยใช้สถิติ two-tailed Student's T-test ที่ P-value<0.05

4.9 ผลการวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงออกของโปรตีนกับคะแนนจากแบบทดสอบ ADI-R

เพื่อวิเคราะห์สหสัมพันธ์ (Correlation analysis) ระหว่างระดับการแสดงออกของโปรตีนทั้ง 5 ตัว ได้แก่ MDH2 DLD HNRNPA2B1 (ROA2) MSN และ ANXA5 กับลักษณะอาการทางคลินิกของผู้ป่วยออทิซึมสเปกตรัม ผู้วิจัยได้ทำการวิเคราะห์สหสัมพันธ์ระหว่างระดับการแสดงออกของโปรตีนข้างต้นกับคะแนนจากแบบทดสอบ ADI-R จำนวน 123 ข้อของผู้ป่วยแต่ละรายจากผู้ป่วยออทิซึมสเปกตรัม 3 กลุ่มย่อย (กลุ่ม L subgroup กลุ่ม M subgroup และกลุ่ม S subgroup) โดยใช้ Pearson correlation analysis ที่ระดับนัยสำคัญ $P\text{-value} < 0.05$ ด้วยโปรแกรม IBM SPSS Statistics version 22 พบว่าโปรตีนทั้ง 5 ตัว มีความสัมพันธ์อย่างมีนัยสำคัญกับคะแนนจากแบบทดสอบ ADI-R หลายข้อด้วยกัน ดังแสดงในตารางที่ 36 โดยโปรตีน MDH2 สัมพันธ์กับข้อความจากแบบทดสอบ ADI-R มากที่สุด รองลงมาคือโปรตีน ROA2 DLD MSN และ ANXA5 ตามลำดับ โดยส่วนมากข้อความที่มีความสัมพันธ์แบบ Negative correlation กับระดับการแสดงออกของโปรตีนทั้ง 5 ตัว จำนวนทั้งสิ้น 47 ข้อ จะเป็นข้อความเกี่ยวกับความบกพร่องทางด้านภาษาแบบ Verbal (แถบสีเขียวเข้ม) ความบกพร่องทางด้านพฤติกรรม (แถบสีชมพู) และความบกพร่องในการเข้าสังคม (แถบสีเหลือง) ดังแสดงในรูปที่ 36

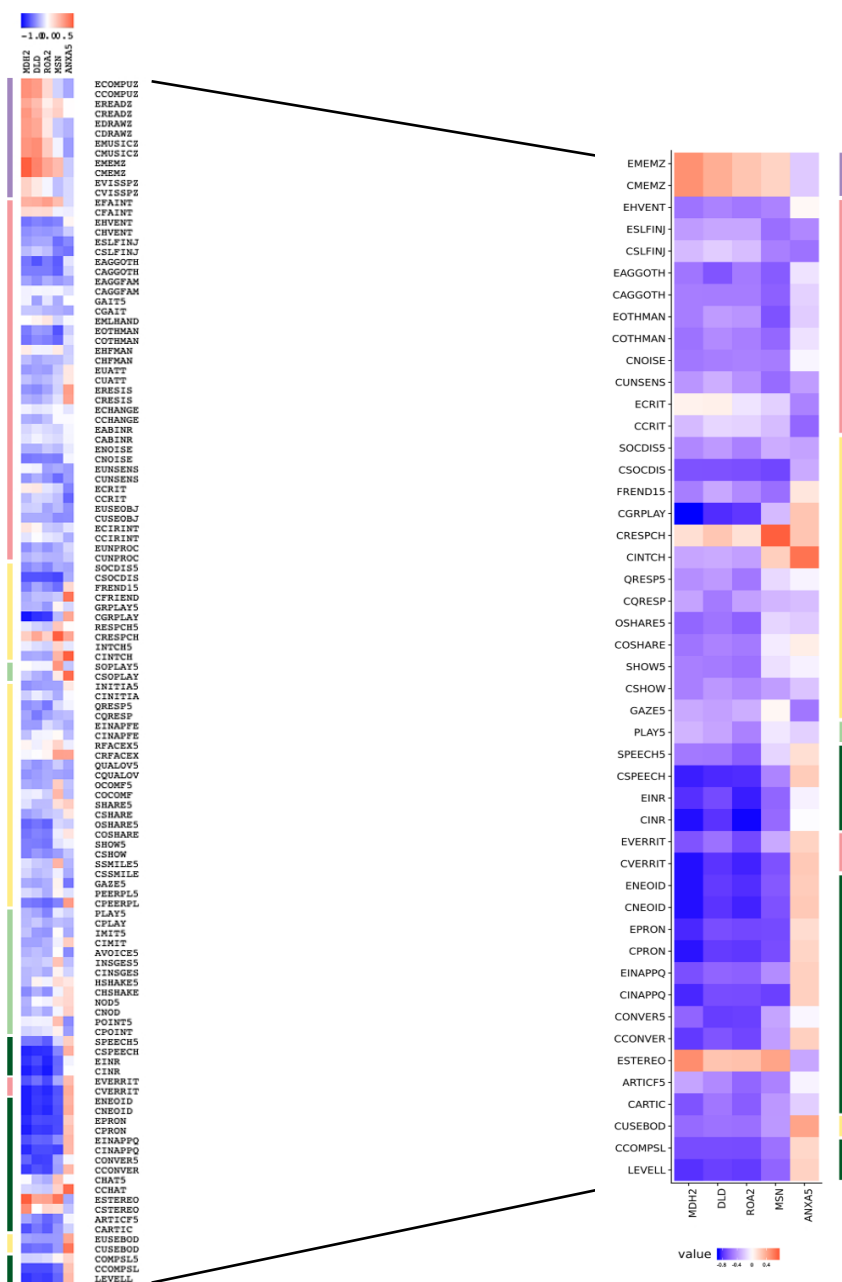
ตารางที่ 36 ผลการวิเคราะห์ความสัมพันธ์ระหว่างการแสดงออกของโปรตีน MDH2 DLD ROA2 MSN และ ANXA5 กับคะแนนจากแบบทดสอบADI-R

| MDH2 vs ADI-R scores | | | | | | | | | |
|----------------------|-------------|----------------------------------------------|---------------------|-----|-------------|--------------------------------------------------------|---------------------|--|--|
| No. | ADI-R items | ADI-R description | Pearson correlation | No. | ADI-R items | ADI-R description | Pearson correlation | | |
| 1 | LEVELL | Overall level of language | 5.44E-05 | 17 | EINR | Intonation/volume/rhythm/rate | 4.74E-05 | | |
| 2 | CCOMPSL | Comprehension of simple language | 3.25E-03 | 18 | CSPEECH | Current communicative speech | 4.14E-06 | | |
| 3 | CUSEBOD | Use of other's body to communicate | 9.82E-03 | 19 | SPEECH5 | Current communicative speech | 2.30E-02 | | |
| 4 | CARTIC | Articulation/pronunciation | 2.58E-03 | 20 | CSHOW | Showing and directing attention | 3.81E-02 | | |
| 5 | ESTEREO | Stereotyped utterances and delayed echolalia | 3.34E-02 | 21 | SHOW5 | Showing and directing attention | 3.92E-02 | | |
| 6 | CCONVER | Reciprocal conversation | 1.59E-04 | 22 | COSHARE | Offering to share | 1.80E-02 | | |
| 7 | CONVER5 | Reciprocal conversation | 6.83E-03 | 23 | OSHARE5 | Offering to share | 9.78E-03 | | |
| 8 | CINAPPO | Inappropriate questions or statements | 2.72E-05 | 24 | CGRPLAY | Group play with peers | 5.48E-04 | | |
| 9 | EINAPPO | Inappropriate questions or statements | 1.37E-03 | 25 | CSOCIDS | Social disinhibition | 3.43E-03 | | |
| 10 | CPRON | Pronominal reversal | 1.94E-06 | 26 | CNOISE | Undue general sensitivity to noise | 2.06E-02 | | |
| 11 | EPRON | Pronominal reversal | 2.98E-05 | 27 | COTHMAN | Other complex mannerisms or stereotyped body movements | 1.97E-02 | | |
| 12 | CNEOID | Neologisms/diosyncratic language | 1.24E-06 | 28 | EOTHMAN | Other complex mannerisms or stereotyped body movements | 3.42E-02 | | |
| 13 | ENEOID | Neologisms/diosyncratic language | 1.23E-06 | 29 | EHVENT | Hyperventilation | 3.97E-02 | | |
| 14 | CVERRIT | Verbal rituals | 1.24E-06 | 30 | CMEMZ | Memory skill | 2.78E-02 | | |
| 15 | EVERRIT | Verbal rituals | 1.91E-03 | 31 | EEMEMZ | Memory skill | 2.78E-02 | | |
| 16 | CINR | Intonation/volume/rhythm/rate | 5.00E-07 | | | | | | |

| DLD vs ADI-R scores | | | | | | | |
|---------------------|-------------|---------------------------------------|---------------------|-----|-------------|----------------------------------------------------|---------------------|
| No. | ADI-R items | ADI-R description | Pearson correlation | No. | ADI-R items | ADI-R description | Pearson correlation |
| 1 | LEVELL | Overall level of language | 3.20E-04 | 14 | EVERRIT | Verbal rituals | 1.48E-02 |
| 2 | CCOMPSL | Comprehension of simple language | 3.21E-03 | 15 | CINR | Intonation/volume/rhythm/rate | 6.93E-05 |
| 3 | CUSEBOD | Use of other's body to communicate | 1.61E-02 | 16 | EINR | Intonation/volume/rhythm/rate | 8.26E-04 |
| 4 | CARTIC | Articulation/pronunciation | 2.50E-02 | 17 | CSPEECH | Current communicative speech | 1.89E-05 |
| 5 | CCONVER | Reciprocal conversation | 1.99E-03 | 18 | SPEECH5 | Current communicative speech | 2.08E-02 |
| 6 | CONVER5 | Reciprocal conversation | 2.35E-04 | 19 | SHOW5 | Showing and directing attention | 3.21E-02 |
| 7 | CINAPPO | Inappropriate questions or statements | 1.54E-03 | 20 | COSHARE | Offering to share | 3.79E-02 |
| 8 | EINAPPO | Inappropriate questions or statements | 6.97E-03 | 21 | OSHARE5 | Offering to share | 2.17E-02 |
| 9 | CPRON | Pronominal reversal | 3.02E-04 | 22 | CQRESP | Appropriateness of social responses | 2.98E-02 |
| 10 | EPRON | Pronominal reversal | 1.61E-03 | 23 | CGRPLAY | Group play with peers | 6.92E-03 |
| 11 | CNEOID | Neologisms/diosyncratic language | 1.30E-04 | 24 | CSOCIDS | Social disinhibition | 3.33E-03 |
| 12 | ENEOID | Neologisms/diosyncratic language | 2.76E-04 | 25 | CNOISE | Undue general sensitivity to noise | 2.83E-02 |
| 13 | CVERRIT | Verbal rituals | 1.30E-04 | 26 | EAGGOTH | Aggression to non-caregivers or non-family members | 2.26E-02 |

| HNRNPA2B1 (ROAA2) vs ADI-R scores | | | | | | | | | |
|-----------------------------------|-------------|---------------------------------------|---------------------|-----|-------------|--------------------------------------------------------|---------------------|-------------------|---------------------|
| No. | ADI-R items | ADI-R description | Pearson correlation | No. | ADI-R items | ADI-R description | Pearson correlation | ADI-R description | Pearson correlation |
| 1 | LEVELL | Overall level of language | 1.71E-04 | 16 | CINR | Intonation/volume/rhythm/rate | 1.79E-07 | | |
| 2 | CCOMPSL | Comprehension of simple language | 2.90E-03 | 17 | EINR | Intonation/volume/rhythm/rate | 3.47E-06 | | |
| 3 | CUSEBOD | Use of other's body to communicate | 1.37E-02 | 18 | CSPEECH | Current communicative speech | 3.02E-05 | | |
| 4 | CARTIC | Articulation/pronunciation | 5.29E-03 | 19 | SPEECH5 | Current communicative speech | 4.38E-03 | | |
| 5 | ARTICF5 | Articulation/pronunciation | 2.80E-02 | 20 | PLAY5 | Imaginative play | 4.29E-02 | | |
| 6 | CCONVER | Reciprocal conversation | 5.52E-04 | 21 | SHOW5 | Showing and directing attention | 1.90E-02 | | |
| 7 | CONVER5 | Reciprocal conversation | 3.14E-04 | 22 | COSHARE | Offering to share | 2.42E-02 | | |
| 8 | CINAPPQ | Inappropriate questions or statements | 1.45E-03 | 23 | OSHARE5 | Offering to share | 7.42E-03 | | |
| 9 | EINAPPQ | Inappropriate questions or statements | 5.11E-03 | 24 | ORESP5 | Appropriateness of social responses | 2.57E-02 | | |
| 10 | CPRON | Pronominal reversal | 1.88E-04 | 25 | CGRPLAY | Group play with peers | 1.27E-02 | | |
| 11 | EPRON | Pronominal reversal | 1.02E-03 | 26 | CSOCIDS | Social disinhibition | 2.63E-03 | | |
| 12 | CNEOID | Neologisms/idiosyncratic language | 1.41E-05 | 27 | SOCDIS5 | Social disinhibition | 4.82E-02 | | |
| 13 | ENEOID | Neologisms/idiosyncratic language | 5.91E-05 | 28 | CNOISE | Undue general sensitivity to noise | 3.51E-02 | | |
| 14 | CVERRIT | Verbal rituals | 1.41E-05 | 29 | COTHMAN | Other complex mannerisms or stereotyped body movements | 3.67E-02 | | |
| 15 | EVERRIT | Verbal rituals | 7.44E-04 | 30 | EHVENT | Hyperventilation | 4.76E-02 | | |

| MSN vs ADI-R scores | | | | | | | |
|-----------------------|-------------|---------------------------------------|---------------------|-----|-------------|--------------------------------------------------------|---------------------|
| No. | ADI-R items | ADI-R description | Pearson correlation | No. | ADI-R items | ADI-R description | Pearson correlation |
| 1 | LEVELL | Overall level of language | 6.93E-03 | 12 | CRESPCH | Response to approaches of other children | 1.27E-02 |
| 2 | CCOMPSL | Comprehension of simple language | 2.91E-02 | 13 | FREND15 | Friendships | 1.28E-02 |
| 3 | CINAPPQ | Inappropriate questions or statements | 5.08E-04 | 14 | CSOCIDS | Social disinhibition | 1.32E-03 |
| 4 | CPRON | Pronominal reversal | 1.56E-03 | 15 | CUNSENS | Unusual sensory interests | 1.07E-02 |
| 5 | EPRON | Pronominal reversal | 1.22E-03 | 16 | CNOISE | Undue general sensitivity to noise | 2.78E-02 |
| 6 | CNEOID | Neologisms/diosyncratic language | 2.33E-03 | 17 | COTHMAN | Other complex mannerisms or stereotyped body movements | 1.08E-02 |
| 7 | ENEOID | Neologisms/diosyncratic language | 3.89E-03 | 18 | EOTHMAN | Other complex mannerisms or stereotyped body movements | 2.42E-03 |
| 8 | CVERRIT | Verbal rituals | 2.33E-03 | 19 | CAGGOTH | Aggression to non-caregivers or non-family members | 3.76E-02 |
| 9 | CINR | Intonation/volume/rhythm/rate | 9.72E-03 | 20 | EAGGOTH | Aggression to non-caregivers or non-family members | 3.03E-02 |
| 10 | EINR | Intonation/volume/rhythm/rate | 7.04E-03 | 21 | ESLFINJ | Self injury | 3.22E-02 |
| 11 | CSPEECH | Current communicative speech | 4.10E-02 | | | | |
| ANXA5 vs ADI-R scores | | | | | | | |
| No. | ADI-R items | ADI-R description | Pearson correlation | No. | ADI-R items | ADI-R description | Pearson correlation |
| 1 | GAZE5 | Direct gaze | 2.45E-02 | 4 | ECRIT | Compulsions/rituals | 3.74E-02 |
| 2 | CINTCH | Interest in children | 3.20E-02 | 5 | CSLFINJ | Self injury | 3.81E-02 |
| 3 | CCRIT | Compulsions/rituals | 7.96E-03 | | | | |



รูปที่ 37 Heatmaps แสดงความสัมพันธ์ของระดับการแสดงออกของโปรตีนจากเซลล์เพาะเลี้ยง

ชนิดลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัมกับคะแนนจากแบบทดสอบ ADI-R

ภาพทางซ้ายคือ Heatmap แสดงความสัมพันธ์ของระดับการแสดงออกของโปรตีนทั้ง 5 ตัว กับคะแนนจากแบบทดสอบ ADI-R จำนวน 123 ข้อ ส่วนภาพทางขวา คือ Heatmap แสดงความสัมพันธ์อย่างมีนัยสำคัญของระดับการแสดงออกของโปรตีนทั้ง 5 ตัว กับคะแนนจากแบบทดสอบ ADI-R จำนวน 47 ข้อ โดยแถบสีด้านข้างบ่งบอกข้อความเกี่ยวกับความบกพร่องทางด้านภาษาแบบ Verbal (แถบสีเขียวเข้ม) ความบกพร่องทางด้านภาษาแบบ Non-verbal (แถบสีเขียวอ่อน) ความบกพร่องทางด้านพฤติกรรม (แถบสีชมพู) ความบกพร่องในการเข้าสังคม (แถบสีเหลือง) และทักษะอัจฉริยภาพพิเศษ (สีม่วง)

บทที่ 5

อภิปรายผล

การศึกษาวินิจฉัยในครั้งนี ผู้วิจัยทำการศึกษาโดยใช้เทคโนโลยีปัญญาประดิษฐ์ในการแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยตามลักษณะอาการทางคลินิก โดยอาศัยข้อมูลจากการวินิจฉัยผู้ป่วยด้วยแบบทดสอบ ADI-R ก่อนทำการศึกษารูปแบบการแสดงออกของทรานสคริปโตม (Transcriptome) และโปรตีโอม (Proteome) ในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ (Lymphoblastoid Cell Lines; LCLs) ที่พัฒนามาจากเม็ดเลือดขาวของผู้ป่วยออทิสซึมสเปกตรัมด้วยเทคนิค 2D-GE และ LC-MS/MS เพื่อทำการเปรียบเทียบหาโปรตีนที่มีความแตกต่างจากคนปกติอย่างมีนัยสำคัญ และศึกษาปฏิสัมพันธ์และความสัมพันธ์ของโปรตีนที่พบว่ามีคามผิดปกติในผู้ป่วยออทิสซึมสเปกตรัมและผู้ป่วยออทิสซึมสเปกตรัมกลุ่มย่อยในกลุ่มที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) และกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroups; S subgroup) โดยคาดหวังว่าโปรตีนที่ได้จากการวิเคราะห์ในครั้งนี จะสามารถนำไปใช้เป็นตัวแทนสำหรับการศึกษาเชิงลึกในประชากรผู้ป่วย เพื่อพัฒนาเป็นสารบ่งชี้ทางชีวภาพ (Biomarker) หรือศึกษากลไกพยาธิสภาพในระดับชีวโมเลกุลของโรคออทิสซึมสเปกตรัม สืบเนื่องจากผู้ป่วยโรคออทิสซึมสเปกตรัมนั้น มีการแสดงออกของอาการและความรุนแรงที่มีความหลากหลายเป็นอย่างมาก ซึ่งความหลากหลายและซับซ้อนของโรคนี้อาจมีสาเหตุมาจากปัจจัยร่วมกันระหว่างปัจจัยทางด้านพันธุกรรม เช่น เกิดความผิดปกติที่บริเวณโครโมโซม 15q11-q13 (Nurmi et al., 2003) และปัจจัยทางด้านสิ่งแวดล้อม เช่น การขาดวิตามินดีในขณะตั้งครรภ์ (Vuillermot et al., 2017), และผลจากการได้รับสารบิสฟีนอลเอ (BPA) ขณะตั้งครรภ์ (Thongkorn et al., 2019) เป็นต้น อย่างไรก็ตามยังไม่มีทั้งสารบ่งชี้สำหรับการวินิจฉัยโรคออทิสซึมสเปกตรัมและสำหรับการจัดผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยเพื่อลดความหลากหลายของอาการของผู้ป่วย รวมถึงยังไม่มีการรักษาที่มีประสิทธิภาพและจำเพาะต่อโรคโดยตรง ซึ่งการวินิจฉัยในปัจจุบันอาศัยเพียงแบบสอบถามหรือแบบทดสอบที่วินิจฉัยจากการสังเกตพฤติกรรมและทำการซักประวัติผู้ป่วยที่มีพัฒนาการผิดปกติ โดยจำเป็นจะต้องแพทย์ผู้เชี่ยวชาญ ปริมาณข้อคำถามมาก และใช้เวลานาน แม้ว่าจะเคยมีการศึกษาที่นำการเรียนรู้ของเครื่องมาใช้ในการศึกษาต่าง ๆ รวมถึงการศึกษาที่เกี่ยวข้องกับโรคออทิสซึมสเปกตรัม เช่น ใช้การเรียนรู้ของเครื่องมือศึกษาเกี่ยวกับการ

เคลื่อนไหวของผู้ป่วยออทิสซึมสเปกตรัม (Crippa et al., 2015), ใช้การเรียนรู้ของเครื่องมือในการจำแนกโรคออทิสซึมสเปกตรัมกับโรคสมาธิสั้น (Duda et al., 2017), และใช้การเรียนรู้ของเครื่องมือในการลดจำนวนแบบทดสอบในการวินิจฉัยผู้ป่วย (Kosmicki et al., 2015) อย่างไรก็ตามการศึกษาเกี่ยวกับการใช้เทคโนโลยีปัญญาประดิษฐ์ในการแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยยังคงมีข้อจำกัดอยู่

เพื่อลดความหลากหลายตามลักษณะอาการที่แสดงออกของโรคออทิสซึมสเปกตรัม ผู้วิจัยได้ทำการแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยโดยใช้วิธีปัญญาประดิษฐ์ อาศัยการเรียนรู้แบบไม่มีผู้สอน (Unsupervised machine learning) ด้วยโปรแกรม RapidMiner พบว่าจำนวนกลุ่มย่อยที่เหมาะสมในการแบ่งจากข้อมูลลักษณะอาการทางคลินิกจากแบบทดสอบ ADI-R จำนวน 123 ข้อของผู้ป่วย 85 ราย คือ จำนวน 4 กลุ่ม และแบ่งโดยใช้วิธี K-means clustering จะให้ประสิทธิภาพในการแบ่งที่ดีที่สุด กล่าวคือกลุ่มย่อยที่ได้จะมีความเหมือนกันของตัวอย่างภายในกลุ่มและมีความแตกต่างระหว่างกลุ่มมากกว่าอีกวิธี ซึ่งผลการแบ่งกลุ่มนี้มีความสอดคล้องกับการแบ่งผู้ป่วยกลุ่มเดียวกันจากการศึกษาของ Hu and Steinberg ในปี 2009 (Hu & Steinberg, 2009) ซึ่งใช้วิธี K-means clustering Principle Components Analysis และ Hierarchical clustering ร่วมกัน จากโปรแกรม MeV พบว่ามีผลลัพธ์ในการแบ่งกลุ่มที่ตรงกันทั้งจำนวนกลุ่มย่อย และรายชื่อกลุ่มตัวอย่างภายในกลุ่ม ได้แก่ กลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษร่วมกับความผิดปกติด้านภาษารุนแรง (Savant skills with severe language impairments subgroup; SL subgroup) กลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup; L subgroup) กลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) และกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroups; S subgroup) แสดงให้เห็นว่าแม้ว่าจะใช้โปรแกรมในการวิเคราะห์ผลที่แตกต่างกัน แต่การแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็น 4 กลุ่มย่อยด้วยวิธี K-means clustering นั้น เป็นวิธีที่สามารถแบ่งผู้ป่วยทั้ง 85 รายนี้ได้ดีที่สุดและสามารถเชื่อถือได้ โดยจากผลการเปรียบเทียบลักษณะอาการทางคลินิกจากแบบทดสอบ ADI-R พบว่าผู้ป่วยในแต่ละกลุ่มย่อยมีลักษณะอาการทางคลินิกที่มีรูปแบบคล้ายคลึงกันภายในกลุ่ม แต่แตกต่างกันระหว่างกลุ่ม

นอกจากนี้จากการศึกษาการใช้เทคโนโลยีปัญญาประดิษฐ์เพื่อลดข้อจำกัดจากการวินิจฉัยโดยใช้แบบทดสอบ ADI-R โดยการลดจำนวนข้อคำถาม เคยมีการเปรียบเทียบประสิทธิภาพของ

โมเดลจากโปรแกรม Waikato Environment for Knowledge Analysis (WEKA) ในการวินิจฉัยผู้ป่วยออทิสซึมสเปกตรัมจากผู้ป่วย 891 ราย และคนที่ไม่มีอาการตามเกณฑ์ของโรคออทิสซึมสเปกตรัม 75 ราย โดยอัลกอริทึม Alternating decision tree ให้ความถูกต้องสูงถึง 99.9% โดยใช้ข้อความเพียง 7 ข้อ (Kosmicki et al., 2015) แต่อย่างไรก็ตามยังไม่เคยมีการใช้เทคโนโลยีปัญญาประดิษฐ์เพื่อลดจำนวนข้อในการแบ่งผู้ป่วยออทิสซึมสเปกตรัมออกเป็นกลุ่มย่อยมาก่อน ผู้วิจัยจึงได้เปรียบเทียบประสิทธิภาพในการแบ่งกลุ่มย่อยตามอาการทางคลินิกโดยใช้วิธีปัญญาประดิษฐ์ อาศัยการเรียนรู้แบบมีผู้สอน (Supervised machine learning) ด้วยโปรแกรม RapidMiner โดยการสร้างโมเดลสำหรับการแบ่งกลุ่มผู้ป่วยออทิสซึมสเปกตรัมออกเป็น 4 กลุ่มย่อยจากวิธี K-means clustering ในข้างต้นพบว่าจากการเปรียบเทียบอัลกอริทึม 7 แบบ ได้แก่ Artificial Neural Network (ANN), k-Nearest Neighbor (k-NN), Naive Bayes, Decision tree, Random forest, Support Vector Machine (SVM), and A Library of Support Vector Machine (LibSVM) โมเดลจากอัลกอริทึม Random forest จำนวน 10 ต้น ที่ใช้กระบวนการวิเคราะห์แบบ Forward selection ในการคัดเลือกข้อมูลให้ประสิทธิภาพในการจัดกลุ่มที่มี Accuracy ของโมเดลสูงถึง 100% มี Sensitivity และ Specificity เฉลี่ยในการจัดกลุ่ม 100% โดยใช้ข้อความจากแบบทดสอบ ADI-R เพียง 5 ข้อ จากทั้งหมด 123 ข้อ ได้แก่ CPRON (Pronominal reversal การเลือกใช้คำสรรพนามที่ไม่เหมาะสม) CVERRIT (Verbal rituals การมีคำติดปากที่ไม่ได้สื่อความหมาย) ECOMPUS (Computational ability ความสามารถในการคำนวณ) LEVELL (Overall level of language ความสามารถในการใช้ภาษาโดยรวม) และ RFACEX5 (Range of facial expression used to communicate การแสดงออกทางสีหน้าเวลาสื่อสาร) ซึ่งมีข้อความในหมวดความบกพร่องทางการใช้ภาษาแบบ Verbal 2 ข้อ คือ LEVELL และ CPRON คำถามในหมวดความบกพร่องทางพฤติกรรม 2 ข้อ คือ CVERRIT และ RFACEX5 และคำถามในหมวดทักษะพิเศษ คือ ECOMPUS

ในการศึกษารูปแบบโปรตีโอมของผู้ป่วยออทิสซึมสเปกตรัม ผู้วิจัยได้ใช้เซลล์เพาะเลี้ยงชนิดลิ้มโพลาสต์ซึ่งเป็นเซลล์ที่พัฒนามาจากเม็ดเลือดขาวของผู้ป่วย โดยคัดเลือกเฉพาะกลุ่มย่อยที่ไม่มี ความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup; M subgroup) และกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroups; S subgroup) มาเปรียบเทียบกับกลุ่มควบคุม ด้วยเทคนิค 2D-GE และ LC-MS/MS พบว่ามีจุดโปรตีนที่แสดงออกผิดปกติอย่างมีนัยสำคัญในแต่ละกลุ่มย่อยเทียบกับกลุ่มควบคุม ในกลุ่ม ASD (M+S

subgroup) 10 จุด กลุ่ม M subgroup 18 จุด และกลุ่ม S subgroup 9 จุด โดยได้คัดเลือกเฉพาะจุดที่มีค่าสัดส่วนความเข้มแตกต่างกันลดลงหรือเพิ่มขึ้นอย่างน้อย 2 เท่า จำนวนทั้งสิ้น 19 จุด ไปทำการวิเคราะห์เพื่อระบุโปรตีนต่อด้วย LC-MS/MS และพบว่า 14 โปรตีน จาก 19 โปรตีนข้างต้น เคยมีการรายงานว่าผิดปกติในผู้ป่วยออทิสซึมสเปกตรัมในฐานข้อมูล AutismKB database และเมื่อทำการวิเคราะห์เครือข่ายทางชีวภาพของโปรตีนทั้ง 19 ตัว ผู้วิจัยพบความเกี่ยวข้องกับหน้าที่หรือโรคทางระบบประสาท (Neurological disease and functions) และโรคพัฒนาการบกพร่อง (Developmental disorder) โดยเฉพาะอย่างยิ่งมีความเกี่ยวข้องกับโรคออทิสซึมสเปกตรัม (Autism spectrum disorder or intellectual disability) ซึ่งมีโมเลกุลของ MYC หรือ c-MYC เป็นโมเลกุลที่มีปฏิสัมพันธ์กับโปรตีนที่พบว่ามีผิดปกติในผู้ป่วยออทิสซึมสเปกตรัมและเป็นศูนย์กลางในการเชื่อมโยงโมเลกุลต่าง ๆ ไว้ด้วยกัน โดย MYC เป็น Proto-oncogene protein ที่ทำหน้าที่เป็น Transcription factor ควบคุมการแสดงออกของยีนในกลุ่มที่เกี่ยวข้องกับการเจริญของเซลล์ กระบวนการ Metabolism และกระบวนการที่เกี่ยวข้องกับ miRNA (Hutter, Bolin, Weishaupt, & Swartling, 2017; Wierstra & Alves, 2008) และเป็น Target gene ที่สำคัญใน WNT signaling pathway ที่เคยมีรายงานว่ามีความเกี่ยวข้องกับกระบวนการทางระบบประสาทและสมองอย่างเช่น กระบวนการ Synaptogenesis (Vallée & Vallée, 2018) และมีผลต่อกระบวนการพัฒนาการของสมอง (Wey & Knoepfler, 2010) นอกจากนี้ยังเคยพบความผิดปกติของการแสดงออกของยีนเพิ่มขึ้นในผู้ป่วยออทิสซึมสเปกตรัมอีกด้วย (Forés-Martos et al., 2019) อย่างไรก็ตามความผิดปกติของโมเลกุล MYC และกลไกการทำงานของ WNT signaling pathway กับพยาธิสภาพของโรคออทิสซึมสเปกตรัมจำเป็นต้องมีการศึกษาเพิ่มเติมต่อไป

จากการวิเคราะห์ความสอดคล้องและความสัมพันธ์ ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิสซึมสเปกตรัมและออทิสซึมสเปกตรัมกลุ่มย่อยกับกลุ่มยีนที่แสดงออกผิดปกติในกลุ่มตัวอย่างเดียวกันจากการศึกษารูปแบบทรานสคริปโตมของ GSE15402 พบโปรตีนที่แสดงออกผิดปกติจากทั้งสองรายชื่ออย่างมีนัยสำคัญเพียงโปรตีนเดียว คือ TPM3 ในกลุ่ม ASD (M+S subgroup) แต่ไม่พบโปรตีนที่แสดงออกผิดปกติจากทั้งสองรายชื่อในกลุ่ม M subgroup และกลุ่ม S subgroup และเมื่อได้วิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิสซึมสเปกตรัม กับรายชื่อยีนที่แสดงออกผิดปกติจากเซลล์เม็ดเลือดของผู้ป่วยออทิสซึมสเปกตรัมที่ได้จากฐานข้อมูล GEO DataSets

ใน NCBI จำนวนทั้งสิ้น 6 การศึกษา พบโปรตีนที่แสดงออกผิดปกติที่มีความสัมพันธ์กับยีนที่แสดงออกผิดปกติจากเซลล์เม็ดเลือดของผู้ป่วยออทิซึมสเปกตรัม จำนวนทั้งสิ้น 11 โปรตีน ได้แก่ HNRNPA2B1 MSN DLD B2M CLIC1 PRDX1 AK2 TBCA TPM3 TUBB และ PTPN6 แสดงให้เห็นว่ากลุ่มโปรตีนที่แสดงออกผิดปกติในผู้ป่วยโรคออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อยจากการศึกษาครั้งนี้ มีความสัมพันธ์กับกลุ่มยีนที่เคยรายงานว่าพบการแสดงออกผิดปกติในผู้ป่วยออทิซึมสเปกตรัม นอกจากนี้ผู้วิจัยได้วิเคราะห์ความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยโรคออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย กับโปรตีนที่เคยมีรายงานว่าพบการแสดงออกผิดปกติในเลือดของผู้ป่วยออทิซึมสเปกตรัมในฐานข้อมูลชีวสารสนเทศ จำนวนทั้งสิ้น 7 การศึกษา พบโปรตีนที่แสดงออกผิดปกติที่มีความสัมพันธ์กับโปรตีนที่เคยมีรายงานว่าพบการแสดงออกผิดปกติในเลือดของผู้ป่วย จำนวน 5 โปรตีน ได้แก่ คือ TUBB HSP90B1 MDH2 DLD และ ANXA5

จากผลการศึกษาความสัมพันธ์ระหว่างกลุ่มโปรตีนที่แสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์จากผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อยดังที่ได้กล่าวไปในข้างต้น ผู้วิจัยได้ทำการเลือกยีนจากบทบาหน้าที่ทางชีววิทยา ระดับความเชื่อมั่น และทิศทางที่สัมพันธ์กันระหว่างรูปแบบทรานสคริปโตมและรูปแบบโปรตีโอม จำนวนทั้งสิ้น 6 ยีน ได้แก่ *TPM3 AK2 CLIC1 DLD HNRNPA2B1* และ *ANXA5* มายืนยันการแสดงออกด้วยเทคนิค RT-qPCR พบว่าเมื่อเปรียบเทียบระดับการแสดงออกของยีนทั้ง 6 ยีนข้างต้น ไม่พบการเปลี่ยนแปลงระดับแสดงออกของยีนอย่างมีนัยสำคัญทั้งในกลุ่มผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยเมื่อเทียบกับกลุ่มควบคุม ซึ่งจากการศึกษาก่อนหน้านี้หลายการศึกษาด้วยกันแม้จะพบว่าผู้ป่วยออทิซึมสเปกตรัมมีการแสดงออกของยีนบางยีนที่ความแตกต่างอย่างมีนัยสำคัญจากคนปกติ แต่การแสดงออกของยีนที่พบนี้มีความแตกต่างกันในผู้ป่วยแต่ละราย (Gregg et al., 2008) อีกทั้งยีนส่วนมากที่เคยมีการรายงานความผิดปกติในผู้ป่วยออทิซึมสเปกตรัมมักจะไม่ค่อยมีความสัมพันธ์กับระดับการแสดงออกของโปรตีนที่ยีนเหล่านั้นผลิตออกมา ซึ่งสาเหตุที่ทำให้การแสดงออกของยีนและโปรตีนไม่สอดคล้องกันดังกล่าว อาจมีความเกี่ยวข้องโดยตรงกับกระบวนการหลังการถอดรหัส (Post-translational modifications) เช่น กระบวนการ Methylation กระบวนการ Phosphorylation กระบวนการ Glycosylation กระบวนการ Ubiquitination หรือ กระบวนการ Proteolysis เป็นต้น ที่ทำให้ระดับการแสดงออกของโปรตีนมีการเปลี่ยนแปลงไปและมีความแตกต่างกันไปในแต่ละบุคคล

และเมื่อไม่พบความแตกต่างอย่างมีนัยสำคัญของการแสดงออกของยีนในกลุ่มผู้ป่วยออทิซึม สเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย ดังนั้นผู้วิจัยจึงมุ่งเน้นไปที่การยืนยันการแสดงออกของโปรตีนจากการศึกษารูปแบบโปรตีโอมที่แสดงออกผิดปกติแทน โดยได้คัดเลือกโปรตีนจากบทบาทหน้าที่ทางชีววิทยา ระดับความเชื่อมั่น และความสัมพันธ์กันระหว่างรูปแบบโปรตีโอมกับรูปแบบทรานสคริปโตม และกลุ่มยีนและโปรตีนจากการศึกษาก่อนหน้า จำนวน 5 โปรตีน ได้แก่ MDH2 DLD HNRNPA2B1 (ROA2) MSN และ ANXA5 มาทำการศึกษาระดับการแสดงออกของโปรตีนด้วยเทคนิค Western blot พบว่ามีโปรตีน HNRNPA2B1 (ROA2) ที่มีระดับการแสดงออกเพิ่มขึ้นอย่างมีนัยสำคัญ ($P < 0.05$) ในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย M subgroup เมื่อเทียบกับกลุ่มควบคุม และพบโปรตีน 4 ตัว ได้แก่ MDH2 DLD HNRNPA2B1 (ROA2) และ MSN ที่มีระดับการแสดงออกลดลงอย่างมีนัยสำคัญ ($P < 0.05$) ในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย L subgroup แต่ไม่พบการเปลี่ยนแปลงของระดับการแสดงออกของโปรตีนทั้ง 5 ตัว ในผู้ป่วยออทิซึมสเปกตรัมแบบรวมกลุ่มย่อย และในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย S subgroup ซึ่งการที่ไม่พบว่าโปรตีนทั้ง 5 ตัวมีระดับการแสดงออกผิดปกติในผู้ป่วยออทิซึมสเปกตรัมแบบรวมกลุ่มย่อยนั้น เป็นข้อสนับสนุนสำคัญที่ทำให้การจัดกลุ่มย่อยเพื่อลดความหลากหลายตามลักษณะอาการที่แสดงออกของโรคออทิซึมสเปกตรัมนั้น เป็นแนวทางที่จะสามารถนำไปสู่การระบุสารบ่งชี้ทางชีวภาพและกลไกในระดับโมเลกุลของแต่ละโรคออทิซึมสเปกตรัมกลุ่มย่อยได้จริง แต่การที่ยังไม่สามารถยืนยันโปรตีนที่มีระดับการแสดงออกผิดปกติในกลุ่ม S subgroup หรือกลุ่มย่อยที่มีทักษะอัจฉริยภาพพิเศษ (Savant skills subgroup) ได้ ผู้วิจัยคาดว่าอาจเกิดจากสาเหตุที่ลักษณะอาการทางคลินิกของผู้ป่วยในกลุ่มย่อยนี้ มีความรุนแรงของความบกพร่องทางการสื่อสาร พฤติกรรม และการเข้าสังคมค่อนข้างน้อย เมื่อเทียบกับกลุ่มย่อยอื่น ๆ อีกทั้งยังมีการแสดงออกของยีนที่ผิดปกติในจำนวนที่น้อยกว่ากลุ่มอื่นและมีจำนวนใกล้เคียงกับจำนวนยีนที่แสดงออกผิดปกติในผู้ป่วยออทิซึมแบบรวมกลุ่มย่อย ดังนั้นการค้นหายีนหรือโปรตีนที่แสดงออกผิดปกติในผู้ป่วยกลุ่มนี้ยังคงต้องศึกษาเพิ่มเติมต่อไปในอนาคต

HNRNPA2B1 (Heterogenous nuclear ribonucleoprotein A2/B1 หรือ ROA2) ซึ่งพบระดับการแสดงออกเพิ่มขึ้นในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup) แต่พบระดับการแสดงออกลดลงในกลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup) เป็นโปรตีนที่มีความเกี่ยวข้องกับกระบวนการ Pre-mRNAs และเกี่ยวข้องกับการขนส่ง mRNA ที่จำเพาะ

ไปยัง Cytoplasm ของเซลล์ประสาทอีกด้วย ทั้งยังมีความเชื่อมโยงกับภาวะการเสื่อมของระบบประสาทในหลายโรคด้วยกัน (Hutten & Dormann, 2016) แต่บทบาทที่แท้จริงของโปรตีนตัวนี้ในระบบประสาทและสมองยังไม่เป็นที่ทราบแน่ชัด นอกจากนี้ยังมีความเกี่ยวข้องกับกระบวนการ Telomere extension และกระบวนการ Chromatin regulation แม้ว่าจะยังไม่เคยมีการรายงานความผิดปกติของยีนหรือโปรตีนตัวนี้ในฐานข้อมูลออทิซึมมาก่อน แต่เคยมีการศึกษาโดย Kong ในปี 2012 (Kong et al., 2012) พบว่า RNA ของโปรตีนตัวนี้แสดงออกเพิ่มขึ้นในผู้ป่วยออทิซึมซึ่งสอดคล้องกับในกลุ่มย่อย M subgroup ในขณะที่มีอีกการศึกษาหนึ่งที่ได้ทำการกลายพันธุ์ยีน *Chd8* ซึ่งส่งผลให้หนูมีพฤติกรรมคล้ายลักษณะอาการที่พบในผู้ป่วยออทิซึมสเปกตรัมกลับมีระดับการแสดงออกของยีนและโปรตีนตัวนี้ลดลง (Gompers et al., 2017) ซึ่งสอดคล้องกับกลุ่มย่อย L subgroup

DLD (Dihydrolipoamide dehydrogenase) เป็นโปรตีนองค์ประกอบของระบบ glycine cleavage โดยเป็น Subunit E3 ของ Mitochondrial multienzyme complexes 3 ชนิด ได้แก่ Pyruvate dehydrogenase (PDH) Alpha-ketoglutarate dehydrogenase (α -KGDH) และ Dehydrogenase 2-oxoacid branched chain (BCKDH) ซึ่ง DLD จัดเป็น Moonlighting protein เนื่องจากความสามารถในการทำงานที่มีกลไกแตกต่างกันได้ โดยเมื่ออยู่ในรูปแบบ Homodimer จะทำหน้าที่เป็นเอนไซม์ Dehydrogenase ที่เกี่ยวข้องกับกระบวนการควบคุมการสร้างและสลายพลังงาน แต่เมื่ออยู่ในรูป Monomer จะทำหน้าที่เป็นเอนไซม์ Protease แทน โดยโปรตีนตัวนี้มีความเกี่ยวข้องกับกระบวนการการสร้างและสลายพลังงานในสมอง และเคยมีการรายงานในโรคที่มีภาวะการเสื่อมของระบบประสาท (Yan, Thangthaeng, & Forster, 2008) และหากเกิดการขาดโปรตีนตัวนี้ในเด็กแรกเกิด จะทำให้เกิดภาวะการตึงตัวของกล้ามเนื้อต่ำจากภาวะเลือดเป็นกรด ซึ่งมักจะทำให้เด็กมีอายุไม่เกิน 2-3 ปีแรก หรือหากมีชีวิตรอดก็จะมีปัญหาในด้านการเจริญเติบโตและพบความบกพร่องทางระบบประสาท (Quinonez SC, 2014) โดยจากการศึกษารูปแบบโปรตีนของ Pichitpunpong ในปี 2019 (Pichitpunpong et al., 2019) พบความผิดปกติของโปรตีนนี้ในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup) ซึ่งสอดคล้องกับในการศึกษานี้ ที่มีการแสดงออกของโปรตีน DLD ลดลงในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย L subgroup

MDH2 (Malate dehydrogenase 2) เป็นเอนไซม์ที่ทำหน้าที่ในกระบวนการ Reverse oxidation ในการเปลี่ยน Malate เป็น Oxaloacetate ในกระบวนการ TCA cycle ไม่ได้มีหน้าที่หรือลักษณะอาการอื่นที่เกี่ยวข้องกับออทิซึมสเปกตรัมโดยตรง แต่ก่อให้เกิดโรค Epileptic encephalopathy ในเด็กหากพบความผิดปกติที่ยีน จะมีอาการ Neurodevelopmental impairment รวมถึง Cognitive และ Motor delays (Ait-El-Mkadem et al., 2017) และเคยมีการรายงานการพบโปรตีนตัวนี้ในสมองของผู้ป่วย Alzheimer's disease อีกด้วย (Shi & Gibson, 2011) นอกจากนี้ยังมีรายงานการเพิ่มขึ้นของโปรตีนตัวนี้จากการศึกษารูปแบบโปรตีโอมใน PBMCs ของผู้ป่วยออทิซึมสเปกตรัมด้วยเทคนิค iTRAQ labeling และ LC-MS/MS โดย Shen ในปี 2019 (L. Shen et al., 2018) ซึ่งทิศทางของการแสดงออกของโปรตีนตรงข้ามกับที่พบในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย L subgroup

MSN (Moesin) หรือ MOES อยู่ใน ERM family protein (Erizin-radixin-moesin) ทำหน้าที่เชื่อม Actin cytoskeleton ให้กับเยื่อหุ้มเซลล์ โดยควบคุมโครงสร้างและหน้าที่ของ Specific domains บริเวณผิวเซลล์ ส่งผลต่อรูปร่างเซลล์ การขนส่งผ่านเยื่อหุ้มเซลล์ การส่งสัญญาณระหว่างเซลล์ และมีหน้าที่สำคัญเกี่ยวกับภูมิคุ้มกันด้วย โดยเกี่ยวข้องกับการ Migration ของเซลล์ ภูมิคุ้มกัน รวมไปถึงบริเวณ Blood-brain barrier แต่ไม่ได้มีหน้าที่หรือลักษณะอาการอื่นที่เกี่ยวข้องกับออทิซึมสเปกตรัมโดยตรง ซึ่งจากการศึกษาระดับของโปรตีนตัวนี้ในเนื้อเยื่อสมองของผู้ป่วยออทิซึมสเปกตรัมพบว่าการแสดงออกลดลง และเมื่อศึกษากลไกการทำงานในระดับโมเลกุลพบว่าการลดลงของโปรตีนตัวนี้ เกิดเนื่องมาจาก Non-coding RNA MSNP1AS ที่เพิ่มขึ้นในผู้ป่วยออทิซึมสเปกตรัม (Kerin et al., 2012) ซึ่งสอดคล้องกับในการศึกษานี้ที่มีการแสดงออกของโปรตีน MSN ลดลงในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย L subgroup

ANXA5 (Annexin A5) อยู่ใน Annexin family เป็น Calcium-dependent phospholipid binding protein โดย ANXA5 ทำหน้าที่เป็นโปรตีน Anticoagulant ไปยับยั้งการทำงานของ Thromboplastin-specific complex ทางอ้อม และทำหน้าที่สำคัญในกระบวนการการส่งต่อสัญญาณในเซลล์ (Cellular signal transduction) กระบวนการอักเสบ (Inflammation) การเจริญเติบโตและการพัฒนาของเซลล์ (Growth and differentiation) โดยเคยมีรายงานจากการศึกษารูปแบบโปรตีโอมของ Pichitpunpong ในปี 2019 (Pichitpunpong et al., 2019) พบการแสดงออกของโปรตีนนี้ลดลงในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติด้านภาษา

รุนแรง (Severe language impairments subgroup) ซึ่งไม่สอดคล้องกับผลการศึกษาระดับการ
แสดงออกของโปรตีนในการศึกษานี้ที่ไม่พบการเปลี่ยนแปลงของระดับการแสดงออกของโปรตีน
ANXA5 ในผู้ป่วยออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อย

และจากการวิเคราะห์สหสัมพันธ์ (Correlation analysis) ระหว่างระดับการแสดงออกของ
โปรตีนทั้ง 5 ตัว ได้แก่ MDH2 DLD HNRNPA2B1 MSN และ ANXA5 กับลักษณะอาการทาง
คลินิกของผู้ป่วยออทิซึมสเปกตรัม พบว่าโปรตีนทั้ง 5 ตัว มีความสัมพันธ์อย่างมีนัยสำคัญกับคะแนน
จากแบบทดสอบ ADI-R หลายข้อด้วยกัน โดยโปรตีน MDH2 สัมพันธ์กับข้อคำถามจากแบบทดสอบ
ADI-R มากที่สุด รองลงมาคือโปรตีน ROA2 DLD MSN และ ANXA5 ตามลำดับ โดยส่วนมากข้อ
คำถามที่มีความสัมพันธ์แบบ Negative correlation กับระดับการแสดงออกของโปรตีนทั้ง 5 ตัว
จำนวนทั้งสิ้น 47 ข้อ โดยจะเป็นข้อคำถามเกี่ยวกับความบกพร่องทางด้านภาษาแบบ Verbal เช่น
LEVELL CPRON SPEECH5 CCONVER ข้อคำถามเกี่ยวกับความบกพร่องทางด้านพฤติกรรม เช่น
CVERRIT CNOISE ข้อคำถามเกี่ยวกับความบกพร่องในการเข้าสังคม เช่น GAZE CSHOW FREN15
และข้อคำถามทักษะพิเศษ EMEMZ และ CMEMZ ซึ่งความสัมพันธ์ระหว่างระดับการแสดงออกของ
โปรตีนกับลักษณะอาการทางคลินิกที่เป็นแบบ Negative correlation นี้ แสดงให้เห็นว่าหากระดับ
การแสดงออกของโปรตีนลดลง ลักษณะอาการทางคลินิกในข้อที่สัมพันธ์อย่างมีนัยสำคัญนั้นจะยังมี
ความรุนแรงมากขึ้น สังเกตได้จากในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติด้านภาษา
รุนแรง (Severe language impairments subgroup) ซึ่งมีลักษณะอาการทางคลินิกรุนแรงกว่า
ผู้ป่วยในกลุ่มย่อยอื่น จะมีระดับการแสดงออกของโปรตีน MDH2 DLD HNRNPA2B1 และ MSN
ที่ลดลงเมื่อเทียบกับคนปกติ

บทที่ 6

สรุปผล ข้อเสนอแนะ และข้อจำกัดของงานวิจัย

การศึกษาวินิจฉัยในครั้งนี้ได้แสดงให้เห็นว่าการใช้เทคโนโลยีปัญญาประดิษฐ์ในการแบ่งผู้ป่วยออทิซึมสเปกตรัมออกเป็นกลุ่มย่อยตามลักษณะอาการทางคลินิก โดยอาศัยข้อมูลจากการวินิจฉัยผู้ป่วยด้วยแบบทดสอบ ADI-R นั้น สามารถลดความหลากหลายตามลักษณะอาการที่แสดงออกของโรคออทิซึมสเปกตรัมในการวิเคราะห์รูปแบบทรานสคริปโตมและโปรตีโอม รวมถึงการค้นหาโปรตีนที่จำเพาะต่อผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อย และมีความสัมพันธ์กับอาการที่แสดงออกของโรคออทิซึมสเปกตรัมได้อย่างมีประสิทธิภาพมากขึ้น ซึ่งเป็นประโยชน์ในการระบุสารบ่งชี้ทางชีวภาพและการศึกษากลไกในระดับโมเลกุลของแต่ละโรคออทิซึมสเปกตรัมกลุ่มย่อยได้ กลุ่มโปรตีนที่พบการแสดงออกผิดปกติในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมอาจสะท้อนถึงกลไกการเปลี่ยนแปลงในระดับโมเลกุลที่เกิดขึ้นในผู้ป่วย รวมถึงเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์อาจใช้เป็นตัวแทนในการค้นหาสารบ่งชี้ทางชีวภาพของผู้ป่วยได้ และเนื่องจากระดับการแสดงออกของยีนที่ผิดปกติส่วนใหญ่ไม่สอดคล้องกันกับระดับการแสดงออกของโปรตีนในผู้ป่วยออทิซึมสเปกตรัม ทำให้จำเป็นต้องมีการศึกษารูปแบบโปรตีโอมในโรคออทิซึมสเปกตรัมเพื่อค้นหาสารบ่งชี้ทางชีวภาพและกลไกในระดับโมเลกุลที่เกิดขึ้นในผู้ป่วย ซึ่งจากการวิเคราะห์รูปแบบโปรตีโอมในเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วยออทิซึมสเปกตรัมและผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยในการศึกษานี้ ได้แสดงให้เห็นถึงโปรตีนที่แสดงออกผิดปกติหลายตัวด้วยกัน รวมถึงโปรตีน HNRNPA2B1 ที่มีระดับการแสดงออกเพิ่มขึ้นในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่ไม่มีความผิดปกติด้านภาษารุนแรง (Non-severe language impairments or Mild subgroup) แต่พบระดับการแสดงออกลดลงในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรง (Severe language impairments subgroup) เมื่อเทียบกับกลุ่มควบคุม โดยมีโปรตีน MDH2 DLD และ MSN ที่มีระดับการแสดงออกลดลงในผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่มีความผิดปกติด้านภาษารุนแรงอีกเช่นกัน ซึ่งโปรตีนเหล่านี้อาจใช้เป็นโปรตีนบ่งชี้ในการศึกษาเชิงลึกในประชากรผู้ป่วย เพื่อพัฒนาเป็นสารบ่งชี้ทางชีวภาพ (biomarker) หรือศึกษากลไกพยาธิสภาพในระดับชีวโมเลกุลของโรคออทิซึมสเปกตรัมได้ต่อไป เช่น การศึกษา Activity ของโปรตีน ROA2 รวมถึงใช้เพื่อพัฒนาการรักษาแบบการแพทย์แม่นยำ (Targeted therapy) ซึ่งจะช่วยให้ผู้ป่วยได้รับการวินิจฉัยที่รวดเร็วและจำเพาะมากขึ้น ได้รับ

การปรับเปลี่ยนพฤติกรรม หรือการเรียนรู้ทักษะที่จำเป็น ส่งผลให้คุณภาพชีวิตของผู้ป่วยโรคอหิวาต์ซึมสเปกตรัมดีขึ้น และนำไปสู่การลดการสูญเสียทรัพยากรมนุษย์และงบประมาณในการดูแลรักษาผู้ป่วยได้ในอนาคต

อย่างไรก็ตามงานวิจัยในครั้งนี้นี้ยังคงมีข้อจำกัดในเรื่องของจำนวนกลุ่มตัวอย่าง เนื่องจากการศึกษาครั้งนี้เป็นการศึกษาเบื้องต้นเท่านั้น ซึ่งในอนาคตควรมีการเพิ่มจำนวนของกลุ่มตัวอย่างทั้งในขั้นตอนการแบ่งกลุ่มย่อยและการสร้างโมเดลในการแบ่งกลุ่มโดยอาศัยเทคโนโลยีปัญญาประดิษฐ์ รวมถึงในการศึกษารูปแบบทรานสคริปโตมและโปรตีโอม และในเรื่องของสิ่งตัวอย่างที่ใช้ในการศึกษานี้คือเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ที่พัฒนามาจากเซลล์เม็ดเลือดขาวของผู้ป่วยอหิวาต์ซึมสเปกตรัมจากประเทศสหรัฐอเมริกา โดยในอนาคตหากต้องการค้นหาโปรตีนที่มีระดับการแสดงออกผิดปกติ เพื่อเป็นตัวแทนของสารบ่งชี้ทางชีวภาพในผู้ป่วยที่สามารถใช้ในการวินิจฉัยโรคทางคลินิกได้จริง อาจจะต้องมีการศึกษาเพิ่มเติมในสิ่งตัวอย่างที่เป็น Whole blood จากประชากรไทย นอกจากนี้จากข้อมูลของผู้ป่วย เราไม่ทราบว่ามีผู้ป่วยรายใดที่อยู่ระหว่างการรักษาอาการของโรคอหิวาต์ซึมสเปกตรัมและจำเป็นจะต้องมีการทานยาหรือไม่ ควรมีการซักถามข้อมูลเกี่ยวกับเรื่องนี้เพิ่มเติม เนื่องจากยาที่ใช้ อาจส่งผลต่อระดับการแสดงออกของยีนหรือโปรตีนของผู้ป่วยได้ และควรมีการศึกษาเพิ่มเติมในแง่ของความผิดปกติของระดับการแสดงออกของโปรตีนที่พบในเลือดสามารถสะท้อนถึงชีวพยาธิสภาพที่เกิดขึ้นในระบบประสาทและสมองของผู้ป่วยด้วยในอนาคต นอกจากนี้ในการศึกษานี้ผู้วิจัยได้ใช้ข้อมูลรูปแบบทรานสคริปโตมที่ได้มีการวิเคราะห์ด้วยเทคนิค Microarray ไว้ก่อนหน้า ซึ่งในขั้นตอนของการศึกษารูปแบบโปรตีโอมและการยืนยันผลได้มีการวิเคราะห์ในภายหลัง จึงอาจเป็นสาเหตุสำคัญของความแปรปรวนที่เกิดขึ้นและส่งผลให้ความผิดปกติของยีนและโปรตีนที่พบไม่สอดคล้องกันแม้จะเป็นสิ่งตัวอย่างจากผู้ป่วยรายเดียวกันก็ตาม จึงควรทำการวิเคราะห์รูปแบบทรานสคริปโตมและโปรตีโอมรวมถึงการยืนยันผลจากสิ่งตัวอย่างที่มีการเก็บหรือเพาะเลี้ยงในระยะเวลาเดียวกัน อีกทั้งเทคนิคที่ใช้ในการศึกษาแบบองค์รวม (Omics) ครั้งนี้ ได้แก่ เทคนิค Microarray และ 2D-GE ล้วนเป็นเทคนิคที่มี False discovery rate (FDR) เกิดขึ้นได้ จึงจำเป็นต้องมีการยืนยันผลต่อด้วย RT-qPCR และ Western blot นอกจากนี้ในการวิเคราะห์รูปแบบโปรตีโอมด้วยเทคนิค 2D-GE ทางผู้วิจัยได้คัดเลือกเฉพาะจุดโปรตีนที่มีค่าความเข้มแตกต่างกันสูง โดยในอนาคตอาจเลือกทุกจุดที่พบว่ามีค่าความเข้มแตกต่างกันอย่างมีนัยสำคัญเพื่อทำการระบุโปรตีน และ

อาจใช้เทคนิคที่มี Reproducibility สูงกว่าเทคนิค 2D-GE ในการศึกษารูปแบบโปรตีนในผู้ป่วย
อหิวสเปกตรัมและอหิวสเปกตรัมกลุ่มย่อยต่อไป



บรรณานุกรม



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

- Adam-Day, S. (2018). Mass spectrometry. Retrieved from <https://alevelnotes.com/notes/chemistry/elements-of-life/mass-spectrometry>
- Ait-El-Mkadem, S., Dayem-Quere, M., Gusic, M., Chaussonot, A., Bannwarth, S., François, B., . . . Paquis-Flucklinger, V. (2017). Mutations in MDH2, Encoding a Krebs Cycle Enzyme, Cause Early-Onset Severe Encephalopathy. *American journal of human genetics*, *100*(1), 151-159. doi:10.1016/j.ajhg.2016.11.014
- Alter, M. D., Kharkar, R., Ramsey, K. E., Craig, D. W., Melmed, R. D., Grebe, T. A., . . . Stephan, D. A. (2011). Autism and increased paternal age related changes in global levels of gene expression regulation. *PloS one*, *6*(2), e16715-e16715. doi:10.1371/journal.pone.0016715
- Amoli, M. M., Carthy, D., Platt, H., & Ollier, W. E. (2008). EBV Immortalization of human B lymphocytes separated from small volumes of cryo-preserved whole blood. *Int J Epidemiol*, *37 Suppl 1*, i41-45. doi:10.1093/ije/dym285
- Autism Spectrum Disorders Working Group of The Psychiatric Genomics, C. (2017). Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*, *8*, 21. doi:10.1186/s13229-017-0137-9
- Barbaresi, W. J., Katusic, S. F., & Voigt, R. G. (2006). Autism: a review of the state of the science for pediatric primary health care clinicians. (1072-4710 (Print)).
- Baron, C. A., Liu, S. Y., Hicks, C., & Gregg, J. P. (2006). Utilization of lymphoblastoid cell lines as a system for the molecular modeling of autism. *J Autism Dev Disord*, *36*(8), 973-982. doi:10.1007/s10803-006-0134-x
- Bernacki, S. H., Beck, J. C., Muralidharan, K., Schaefer, F. V., Shrimpton, A. E., Richie, K. L., . . . Stenzel, T. T. (2005). Characterization of publicly available lymphoblastoid cell lines for disease-associated mutations in 11 genes. *Clin Chem*, *51*(11), 2156-2159. doi:10.1373/clinchem.2005.048686
- Bone, D., Bishop, S. L., Black, M. P., Goodwin, M. S., Lord, C., & Narayanan, S. S. (2016). Use of machine learning to improve autism screening and diagnostic instruments: effectiveness, efficiency, and multi-instrument fusion. *J Child Psychol Psychiatry*, *57*(8), 927-937. doi:10.1111/jcpp.12559

- Broek, J. A. C., Brombacher, E., Stelzhammer, V., Guest, P. C., Rahmoune, H., & Bahn, S. (2014). The need for a comprehensive molecular characterization of autism spectrum disorders. *International Journal of Neuropsychopharmacology*, 17(4), 651-673. doi:10.1017/S146114571300117x
- Castagnola, M., Messina, I., Inzitari, R., Fanali, C., Cabras, T., Morelli, A., . . . Gurrieri, F. (2008). Hypo-Phosphorylation of Salivary Peptidome as a Clue to the Molecular Pathogenesis of Autism Spectrum Disorders. *Journal of Proteome Research*, 7(12), 5327-5332. doi:10.1021/pr8004088
- Center for Disease Control and Prevention. (2016). Autism Spectrum Disorder (ASD). Retrieved from <https://www.cdc.gov/ncbddd/autism/data.html>
- Center for Disease Control and Prevention. (2020). Autism Spectrum Disorder (ASD). Retrieved from <https://www.cdc.gov/ncbddd/autism/data.html>
- Chantayanon. (2016). More than 3,000 autistic children outside the system advise how parents care; 2012. Retrieved from http://www.autisticthai.com/attach/knowledge_1429275314_Roadmap.pdf
- Charman, T., & Gotham, K. (2013). Measurement Issues: Screening and diagnostic instruments for autism spectrum disorders - lessons from research and practice. (1475-357X (Print)).
- Christensen, J., Gronborg, T. K., Sorensen, M. J., Schendel, D., Parner, E. T., Pedersen, L. H., & Vestergaard, M. (2013). Prenatal valproate exposure and risk of autism spectrum disorders and childhood autism. *JAMA*, 309(16), 1696-1703. doi:10.1001/jama.2013.2270
- Constantino, J. N., Lavesser, P. D., Zhang, Y. I., Abbacchi, A. M., Gray, T., & Todd, R. D. (2007). Rapid Quantitative Assessment of Autistic Social Impairment by Classroom Teachers. *Journal of the American Academy of Child & Adolescent Psychiatry*, 46(12), 1668-1676. doi:<https://doi.org/10.1097/chi.0b013e318157cb23>
- Copeland, B. J. (2019). Artificial intelligence. Retrieved from <https://www.britannica.com/technology/artificial-intelligence>
- Corbett, B. A., Kantor, A. B., Schulman, H., Walker, W. L., Lit, L., Ashwood, P., . . . Sharp, F. R. (2007). A proteomic study of serum from children with autism

- showing differential expression of apolipoproteins and complement proteins. *Mol Psychiatry*, 12(3), 292-306. doi:10.1038/sj.mp.4001943
- Cortelazzo, A., De Felice, C., Guerranti, R., Signorini, C., Leoncini, S., Zollo, G., . . . Hayek, J. (2016). Expression and oxidative modifications of plasma proteins in autism spectrum disorders: Interplay between inflammatory response and lipid peroxidation. *Proteomics Clinical Applications*, 10(11), 1103-1112. doi:10.1002/prca.201500076
- Crippa, A., Salvatore, C., Perego, P., Forti, S., Nobile, M., Molteni, M., & Castiglioni, I. (2015). Use of Machine Learning to Identify Children with Autism and Their Motor Abnormalities. (1573-3432 (Electronic)).
- David, F. (2017). Types of Machine Learning Algorithms You Should Know: Toward Data Sciences. Retrieved from <https://towardsdatascience.com/types-of-machine-learning-algorithms-you-should-know-953a08248861>
- Duda, M., Haber, N., Daniels, J., & Wall, D. P. (2017). Crowdsourced validation of a machine-learning classification system for autism and ADHD. *Translational Psychiatry*, 7. doi:ARTN e113310.1038/tp.2017.86
- Durkin, M. S., Maenner, M. J., Newschaffer, C. J., Lee, L. C., Cunniff, C. M., Daniels, J. L., . . . Schieve, L. A. (2008). Advanced parental age and the risk of autism spectrum disorder. *Am J Epidemiol*, 168(11), 1268-1276. doi:10.1093/aje/kwn250
- Feng, C. Y., Chen, Y. J., Pan, J. T., Yang, A. C., Niu, L., Min, J., . . . Shen, L. M. (2017). Redox proteomic identification of carbonylated proteins in autism plasma: insight into oxidative stress and its related biomarkers in autism. *Clinical Proteomics*, 14. doi:ARTN 210.1186/s12014-017-9138-0
- Fernandopulle, N. (2011). Measurement of autism: a review of four screening measures. *Indian journal of psychological medicine*, 33(1), 5-10. doi:10.4103/0253-7176.85389
- Forés-Martos, J., Catalá-López, F., Sánchez-Valle, J., Ibáñez, K., Tejero, H., Palma-Gudiel, H., . . . Tabarés-Seisdedos, R. (2019). Transcriptomic metaanalyses of autistic brains reveals shared gene expression and biological pathway

abnormalities with cancer. *Molecular autism*, 10, 17-17. doi:10.1186/s13229-019-0262-8

- Gazestani, V. H., Pramparo, T., Nalabolu, S., Kellman, B. P., Murray, S., Lopez, L., . . . Lewis, N. E. (2019). A perturbed gene network containing PI3K-AKT, RAS-ERK and WNT- β -catenin pathways in leukocytes is linked to ASD genetics and symptom severity. *Nature neuroscience*, 22(10), 1624-1634. doi:10.1038/s41593-019-0489-x
- Gompers, A. L., Su-Feher, L., Ellegood, J., Copping, N. A., Riyadh, M. A., Stradleigh, T. W., . . . Nord, A. S. (2017). Germline Chd8 haploinsufficiency alters brain development in mouse. *Nature neuroscience*, 20(8), 1062-1073. doi:10.1038/nn.4592
- Gong, X., Bacchelli, E., Blasi, F., Toma, C., Betancur, C., Chaste, P., . . . Bourgeron, T. (2008). Analysis of X chromosome inactivation in autism spectrum disorders. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics*, 147B(6), 830-835. doi:10.1002/ajmg.b.30688
- Gregg, J. P., Lit, L., Baron, C. A., Hertz-Picciotto, I., Walker, W., Davis, R. A., . . . Sharp, F. R. (2008). Gene expression changes in children with autism. *Genomics*, 91(1), 22-29. doi:10.1016/j.ygeno.2007.09.003
- Habib, A., Harris, L., Pollick, F., & Melville, C. (2019). A meta-analysis of working memory in individuals with autism spectrum disorders. *PloS one*, 14(4), e0216198-e0216198. doi:10.1371/journal.pone.0216198
- Hallmayer, J., Cleveland, S., Torres, A., Phillips, J., Cohen, B., Torigoe, T., . . . Risch, N. (2011). Genetic heritability and shared environmental factors among twin pairs with autism. *Arch Gen Psychiatry*, 68(11), 1095-1102. doi:10.1001/archgenpsychiatry.2011.76
- Howes, O. D., Rogdaki, M., Findon, J. L., Wichers, R. H., Charman, T., King, B. H., . . . Murphy, D. G. (2018). Autism spectrum disorder: Consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. *J Psychopharmacol*, 32(1), 3-29. doi:10.1177/0269881117741766

- Hu, V. W., Sarachana, T., Kim, K. S., Nguyen, A., Kulkarni, S., Steinberg, M. E., . . . Lee, N. H. (2009). Gene expression profiling differentiates autism case-controls and phenotypic variants of autism spectrum disorders: evidence for circadian rhythm dysfunction in severe autism. *Autism Res, 2*(2), 78-97.
doi:10.1002/aur.73
- Hu, V. W., & Steinberg, M. E. (2009). Novel clustering of items from the Autism Diagnostic Interview-Revised to define phenotypes within autism spectrum disorders. *Autism Res, 2*(2), 67-77. doi:10.1002/aur.72
- Hutten, S., & Dormann, D. (2016). hnRNPA2/B1 Function in Neurodegeneration: It's a Gain, Not a Loss. *Neuron, 92*(4), 672-674.
doi:https://doi.org/10.1016/j.neuron.2016.11.014
- Hutter, S., Bolin, S., Weishaupt, H., & Swartling, F. J. (2017). Modeling and Targeting MYC Genes in Childhood Brain Tumors. *Genes, 8*(4), 107.
doi:10.3390/genes8040107
- Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: an introduction to cluster analysis* (Vol. 344): John Wiley & Sons.
- Kerin, T., Ramanathan, A., Rivas, K., Grepo, N., Coetzee, G. A., & Campbell, D. B. (2012). A Noncoding RNA Antisense to Moesin at 5p14.1 in Autism. *Science Translational Medicine, 4*(128), 128ra140. doi:10.1126/scitranslmed.3003479
- Kim, S. H., & Lord, C. (2012). New autism diagnostic interview-revised algorithms for toddlers and young preschoolers from 12 to 47 months of age. (1573-3432 (Electronic)).
- Kong, S. W., Collins, C. D., Shimizu-Motohashi, Y., Holm, I. A., Campbell, M. G., Lee, I.-H., . . . Kohane, I. S. (2012). Characteristics and predictive value of blood transcriptome signature in males with autism spectrum disorders. *PloS one, 7*(12), e49475-e49475. doi:10.1371/journal.pone.0049475
- Kosmicki, J. A., Sochat, V., Duda, M., & Wall, D. P. (2015). Searching for a minimal set of behaviors for autism detection through feature selection-based machine learning. *Translational Psychiatry, 5*. doi:ARTN e51410.1038/tp.2015.7
- Krakowiak, P., Goines, P. E., Tancredi, D. J., Ashwood, P., Hansen, R. L., Hertz-Picciotto, I., & Van de Water, J. (2017). Neonatal Cytokine Profiles Associated With

Autism Spectrum Disorder. *Biol Psychiatry*, 81(5), 442-451.

doi:10.1016/j.biopsych.2015.08.007

Kuwano, Y., Kamio, Y., Kawai, T., Katsuura, S., Inada, N., Takaki, A., & Rokutan, K. (2011). Autism-Associated Gene Expression in Peripheral Leucocytes Commonly Observed between Subjects with Autism and Healthy Women Having Autistic Children. *PloS one*, 6(9), e24723.

doi:10.1371/journal.pone.0024723

Loon, R. v. (2018). Machine learning explained: Understanding supervised, unsupervised, and reinforcement learning. Retrieved from <https://bigdata-madesimple.com/machine-learning-explained-understanding-supervised-unsupervised-and-reinforcement-learning/>

Lord, C., Rutter, M., & Le Couteur, A. (1994). Autism Diagnostic Interview-Revised: A revised version of a diagnostic interview for caregivers of individuals with possible pervasive developmental disorders. *Journal of Autism and Developmental Disorders*, 24(5), 659-685. doi:10.1007/BF02172145

Marco, E. J., & Skuse, D. H. (2006). Autism-lessons from the X chromosome. *Social cognitive and affective neuroscience*, 1(3), 183-193. doi:10.1093/scan/nsi028

Meng, Z., Hu, Y., & Ancey, C. (2020). Using a Data Driven Approach to Predict Waves Generated by Gravity Driven Mass Flows. . *Water*.

Mierswa, I. (2017). What is Artificial Intelligence, Machine Learning, and Deep Learning. Retrieved from <https://ingomierswa.com/2017/04/19/what-is-artificial-intelligence-machine-learning-and-deep-learning/>

Murphy, C. M., Wilson, C. E., Robertson, D. M., Ecker, C., Daly, E. M., Hammond, N., . . . McAlonan, G. M. (2016). Autism spectrum disorder in adults: diagnosis, management, and health services development. *Neuropsychiatr Dis Treat*, 12, 1669-1686. doi:10.2147/NDT.S65455

National Institute for Health and Care Excellence. (2013). Autism spectrum disorder in under 19s: support and management Retrieved from <https://www.nice.org.uk/guidance/cg170/resources/autism-spectrum-disorder-in-under-19s-support-and-management-pdf-35109745515205>

- Ngounou Wetie, A. G., Wormwood, K. L., Charette, L., Ryan, J. P., Woods, A. G., & Darie, C. C. (2015). Comparative two-dimensional polyacrylamide gel electrophoresis of the salivary proteome of children with autism spectrum disorder. *J Cell Mol Med*, *19*(11), 2664-2678. doi:10.1111/jcmm.12658
- Ngounou Wetie, A. G., Wormwood, K. L., Russell, S., Ryan, J. P., Darie, C. C., & Woods, A. G. (2015). A Pilot Proteomic Analysis of Salivary Biomarkers in Autism Spectrum Disorder. *Autism Research*, *8*(3), 338-350. doi:10.1002/aur.1450
- Nurmi, E. L., Dowd, M., Tadevosyan-leyfer, O., Haines, J. L., Folstein, S. E., & Sutcliffe, J. S. (2003). Exploratory Subsetting of Autism Families Based on Savant Skills Improves Evidence of Genetic Linkage to 15q11-q13. *Journal of the American Academy of Child & Adolescent Psychiatry*, *42*(7), 856-863. doi:http://dx.doi.org/10.1097/01.CHI.0000046868.56865.0F
- Oliveros, J. C. (2007-2015). Venny. An interactive tool for comparing lists with Venn's diagrams. Retrieved from <https://bioinfogp.cnb.csic.es/tools/venny/index.html>
- Parker-Athill, E. C., & Tan, J. (2010). Maternal immune activation and autism spectrum disorder: interleukin-6 signaling as a key mechanistic pathway. *Neurosignals*, *18*(2), 113-128. doi:10.1159/000319828
- Pichitpunpong, C., Thongkorn, S., Kantayaprasit, S., Yuwattana, W., Plaingam, W., Sangsuthum, S., . . . Sarachana, T. (2019). Phenotypic subgrouping and multi-omics analyses reveal reduced diazepam-binding inhibitor (DBI) protein levels in autism spectrum disorder with severe language impairment. *PloS one*, *14*(3). doi:ARTN e021419810.1371/journal.pone.0214198
- Pramparo, T., Pierce, K., Lombardo, M. V., Carter Barnes, C., Marinero, S., Ahrens-Barbeau, C., . . . Courchesne, E. (2015). Prediction of Autism by Translation and Immune/Inflammation Coexpressed Genes in Toddlers From Pediatric Community Practices. *JAMA Psychiatry*, *72*(4), 386-394. doi:10.1001/jamapsychiatry.2014.3008
- Quinonez SC, T. J. (2014). Dihydrolipoamide Dehydrogenase Deficiency. *GeneReviews*. Retrieved from <https://www.ncbi.nlm.nih.gov/books/NBK220444/>
- Ramsey, J. M., Guest, P. C., Broek, J. A. C., Glennon, J. C., Rommelse, N., Franke, B., . . . Bahn, S. (2013). Identification of an age-dependent biomarker signature in

- children and adolescents with autism spectrum disorders. *Molecular autism*, 4. doi:Artn 2710.1186/2040-2392-4-27
- Rutter, M., LeCouteur, A., & Lord, C. (2003). The Autism Diagnostic Interview-Revised. *Los Angeles, CA: Western Psychological Services*, 24, 659-685.
- Shen, C., Zhao, X. L., Ju, W. N., Zou, X. B., Huo, L. R., Yan, W., . . . Zhong, N. (2011). A Proteomic Investigation of B Lymphocytes in an Autistic Family: A Pilot Study of Exposure to Natural Rubber Latex (NRL) May Lead to Autism. *Journal of Molecular Neuroscience*, 43(3), 443-452. doi:10.1007/s12031-010-9463-5
- Shen, L., Liu, X., Zhang, H., Lin, J., Feng, C., & Iqbal, J. (2020). Biomarkers in autism spectrum disorders: Current progress. *Clinica Chimica Acta*, 502, 41-54. doi:https://doi.org/10.1016/j.cca.2019.12.009
- Shen, L., Zhang, K., Feng, C., Chen, Y., Li, S., Iqbal, J., . . . Zhai, J. (2018). iTRAQ-Based Proteomic Analysis Reveals Protein Profile in Plasma from Children with Autism. *PROTEOMICS – Clinical Applications*, 12(3), 1700085. doi:10.1002/prca.201700085
- Shi, Q., & Gibson, G. E. (2011). Up-regulation of the mitochondrial malate dehydrogenase by oxidative stress is mediated by miR-743a. *Journal of neurochemistry*, 118(3), 440-448. doi:10.1111/j.1471-4159.2011.07333.x
- Sie, L., Loong, S., & Tan, E. K. (2009). Utility of lymphoblastoid cell lines. *J Neurosci Res*, 87(9), 1953-1959. doi:10.1002/jnr.22000
- Staudt, L., & Brown, P. (2000). Genomic Views of the Immune System*. *Annual review of immunology*, 18, 829-859. doi:10.1146/annurev.immunol.18.1.829
- Stromland, K., Nordin, V., Miller, M., Akerstrom, B., & Gillberg, C. (1994). Autism in thalidomide embryopathy: a population study. *Dev Med Child Neurol*, 36(4), 351-356. Retrieved from <https://www.ncbi.nlm.nih.gov/pubmed/8157157>
- <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-8749.1994.tb11856.x?sid=nlm%3Apubmed>
- Suckale, J. (2009). RNA extraction using trizol/tri. Retrieved from https://openwetware.org/wiki/RNA_extraction_using_trizol/tri

- Thongkorn, S., Kanlayaprasit, S., Jindatip, D., Tencomnao, T., Hu, V. W., & Sarachana, T. (2019). Sex Differences in the Effects of Prenatal Bisphenol A Exposure on Genes Associated with Autism Spectrum Disorder in the Hippocampus. *Scientific Reports*, *9*. doi:ARTN 303810.1038/s41598-019-39386-w
- Vallée, A., & Vallée, J.-N. (2018). Warburg effect hypothesis in autism Spectrum disorders. *Molecular brain*, *11*(1), 1-1. doi:10.1186/s13041-017-0343-6
- Vuillermot, S., Luan, W., Meyer, U., & Eyles, D. (2017). Vitamin D treatment during pregnancy prevents autism-related phenotypes in a mouse model of maternal immune activation. *Molecular autism*, *8*, 9. doi:10.1186/s13229-017-0125-0
- Wall, D. P., Dally, R., Luyster, R., Jung, J. Y., & Deluca, T. F. (2012). Use of artificial intelligence to shorten the behavioral diagnosis of autism. *PLoS One*, *7*(8), e43855. doi:10.1371/journal.pone.0043855
- Warren, Z., Vehorn, A., Dohrmann, E., Nicholson, A., Sutcliffe, J. S., & Veenstra-VanderWeele, J. (2012). Accuracy of phenotyping children with autism based on parent report: what specifically do we gain phenotyping “rapidly”? *Autism Research*, *5*(1), 31-38. doi:10.1002/aur.230
- Wetie, A. G. N., Wormwood, K., Thome, J., Dudley, E., Taurines, R., Gerlach, M., . . . Darie, C. C. (2014). A pilot proteomic study of protein markers in autism spectrum disorder. *Electrophoresis*, *35*(14), 2046-2054. doi:10.1002/elps.201300370
- Wey, A., & Knoepfler, P. S. (2010). C-myc and N-myc in the developing brain. *Aging (Albany NY)*, *2*(5), 261-262. Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2898015/>
- Wierstra, I., & Alves, J. (2008). The c-myc Promoter: Still MysterY and Challenge. In *Advances in Cancer Research* (Vol. 99, pp. 113-333): Academic Press.
- Xie, J., Huang, L., Li, X., Li, H., Zhou, Y., Zhu, H., . . . Xu, W. (2017). Immunological cytokine profiling identifies TNF-alpha as a key molecule dysregulated in autistic children. *Oncotarget*. doi:10.18632/oncotarget.19326
- Yan, L.-J., Thangthaeng, N., & Forster, M. J. (2008). Changes in dihydrolipoamide dehydrogenase expression and activity during postnatal development and

aging in the rat brain. *Mechanisms of ageing and development*, 129(5), 282-290. doi:10.1016/j.mad.2008.01.006

Yang, L., & Rudser, K. (2016). Urine Protein Biomarker Candidates for Autism. *Journal of Proteomics & Bioinformatics*, s14. doi:10.4172/jpb.S14-004

Zerbo, O., Yoshida, C., Grether, J. K., Van de Water, J., Ashwood, P., Delorenze, G. N., . . . Croen, L. A. (2014). Neonatal cytokines and chemokines and risk of Autism Spectrum Disorder: the Early Markers for Autism (EMA) study: a case-control study. *Journal of Neuroinflammation*, 11. doi:Artn 11310.1186/1742-2094-11-113

กิตติศักดิ์ เกิดประสพ. (2552). การพัฒนาเหมืองข้อมูลแบบจัดกลุ่ม.

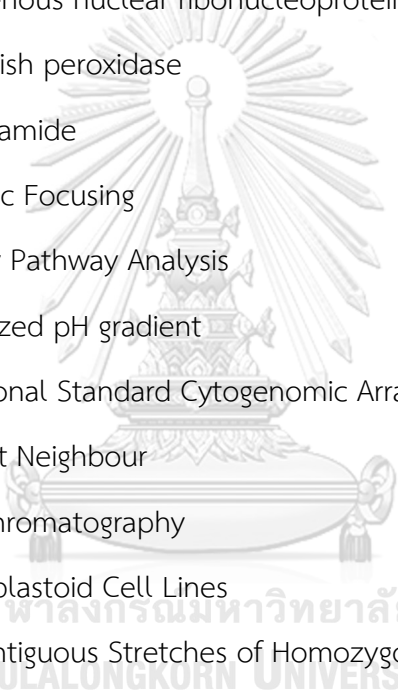
วีระศักดิ์ ช่องจูเหลื่อม. (2555). การจัดกลุ่มข้อมูลด้วยเทคนิคกราฟเคมิตอยส์แบบขนานบนหน่วยประมวลผลกลางแบบหลายแกนหลัก. มหาวิทยาลัยเทคโนโลยีสุรนารี

ศุภศร วินิขเวชารุ่งเรือง, ณ. ป. (2553). แมสสเปกโตรเมตรี (*Mass Spectrometry*).

ภาคผนวก

รายการตัวย่อ (Abbreviations)

| | |
|-------|----------------------------------------------------------------------|
| 2D-GE | Two-Dimensional Gel Electrophoresis |
| ADHD | Attention Deficit Hyperactivity Disorder |
| ADI-R | Autism Diagnostic Interview-Revised |
| ADOS | Autism Diagnostic Observation Schedule |
| AGRE | Autism Genetic Resource Exchange |
| AK2 | Adenylate kinase 2 |
| ANN | Artificial Neural Network |
| ANXA5 | Annexin A5 |
| AQ | Autism Spectrum Questions |
| AP | Alkaline phosphatase |
| APA | American Psychiatric Association |
| ASD | Autism Spectrum Disorder |
| BAC | Bacterial Artificial Chromosome |
| BPA | Bisphenol A |
| BSA | Bovine serum albumin |
| CARS | Childhood Autism Rating Scale |
| CDC | Center for Disease Control and Prevention |
| CGH | Comparative Genomic Hybridization |
| CLIC1 | Chloride intracellular channel protein 1 |
| CNVs | Copy Number Variants |
| DGV | Database of Genomic Variants |
| DBI | Diazepam-binding inhibitor |
| DLD | Dihydrolipoamide dehydrogenase |
| DSM-5 | Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition |
| DTT | Dithiothreitol |



| | |
|-----------|----------------------------------------------|
| EBV | Epstein-Barr Virus |
| FBS | Fetal Bovine Serum |
| FDR | False discovery rate |
| FISH | Fluorescence in situ hybridization |
| GARS | Glycine-tRNA ligase |
| GARS-2 | Gilliam Autism Rating Scale–Second Edition |
| HCL | Hierarchical Clustering |
| HNRNPA2B1 | Heterogenous nuclear ribonucleoprotein A2/B1 |
| HRP | Horseradish peroxidase |
| IAA | Iodoacetamide |
| IEF | Isoelectric Focusing |
| IPA | Ingenuity Pathway Analysis |
| IPG | Immobilized pH gradient |
| ISCA | International Standard Cytogenomic Array |
| k-NN | K-Nearest Neighbour |
| LC | Liquid Chromatography |
| LCLs | Lymphoblastoid Cell Lines |
| LCSHs | Long Contiguous Stretches of Homozygosity |
| LibSVM | A Library of Support Vector Machine |
| LOH | Loss of Heterozygosity |
| MDH2 | Malate dehydrogenase 2 |
| MeV | Multi-Experimental Viewer |
| MIDAS | Microarray Data Analysis System |
| MS | Mass Spectrometry |
| MSN | Moesin |
| MYC | Myc proto-oncogene |
| NC | Nitrocellulose |

| | |
|----------|-----------------------------------------------------------------------|
| NCBI | National Center for Biotechnology Information |
| NICE | National Institute for Health and Care Excellence |
| PBS | Phosphate-buffered saline |
| PCA | Principle Component Analysis |
| PCR | Polymerase Chain Reaction |
| PHA | Phytohemagglutinin |
| pI | Isoelectric point |
| PVDF | Poly-vinylidene difluoride |
| SDS-PAGE | Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis |
| SIGU | Human Genetics Society of Australia Italian Society of Human Genetics |
| SNP | Single Nucleotide Polymorphism |
| SVM | Support Vector Machine |
| TBS | Tris-buffered saline |
| TOF | Time of Flight |
| TPM3 | Tropomyosin alpha-3 |
| UCSC | University of California Santa Cruz Genome Institute |
| UPD | Uniparental Disomy |
| WEKA | Waikato Environment for Knowledge Analysis |

รายการข้อคำถาม ADI-R จำนวน 123 ข้อ และเกณฑ์การให้คะแนน

1. LEVELL -Over all level of language

0 = Functional use of spontaneous echoed or stereotyped language that, on a daily basis, involves phrases of three words or more that at least sometimes include a verb and are comprehensible to other people

1 = No functional use of three-word phrases in spontaneous, echoed, or stereotyped speech, but uses speech on a daily basis with at least five different words in the last month

2 = Fewer than five words total or speech not used on a daily basis

9 = N/K or not asked

2. CCOMPSSL -Comprehension of simple language (current)

0 = In response to a request can usually perform an unexpected action with an unexpected object; or could place an object, other than something to be used by self (such as boots or a toy), in an unexpected location in a different room ("Put the keys on the kitchen table")

1 = In response to a request can usually get an object, other than something for self or something highly contextualized, from another room ("Get the keys from the kitchen table"), but usually cannot carry out a new action on this object or put it in a "new" place

2 = Understands many words (more than 50), including "yes" and names of familiar people, toys, or foods, but does not meet criteria for "0" or "1"

3 = Understands fewer than 50 words, but some comprehension of "no" and names of a few favorite objects, foods, or people, or words within familiar routines

4 = Little or no comprehension of words, even in context

8 = N/A (e.g., deaf)

9 = N/K or not asked

3. COMPSSL5 -Comprehension of simple language (most abnormal)

0 = Functional use of spontaneous echoed or stereotyped language that, on a daily basis, involves phrases of three words or more that at least sometimes include a verb and are comprehensible to other people

1 = No functional use of three-word phrases in spontaneous, echoed, or stereotyped speech, but uses speech on a daily basis with at least five different words in the last month

2 = Fewer than five words total or speech not used on a daily basis

9 = N/K or not asked

4. CUSEBOD -Use of other's body to communicate (current)

0 = No use of other's body to communicate, except in situations where other strategies have not worked (e.g., when parent conversing with someone and subject cannot get her/his attention), or when taking someone's hand to lead them places

1 = Occasional placement of other's hand on objects or use of other's hand as a tool or to point, but some combination with other modes of communication

2 = Occasional placement of other's hand or use of other's hand as a tool or to demonstrate "for" the subject without integration with other modes of communication

3 = Regular use of other's hand as a tool or to gesture "for" the subject

8 = Little or no spontaneous communication

9 = N/K or not asked

5. EUSEBOD -Use of other's body to communicate (ever)

0 = No use of other's body to communicate, except in situations where other strategies have not worked (e.g., when parent conversing with someone and subject cannot get her/his attention), or when taking someone's hand to lead them places

1 = Occasional placement of other's hand on objects or use of other's hand as a tool or to point, but some combination with other modes of communication

2 = Occasional placement of other's hand or use of other's hand as a tool or to demonstrate "for" the subject without integration with other modes of communication

3 = Regular use of other's hand as a tool or to gesture "for" the subject

8 = Little or no spontaneous communication

9 = N/K or not asked

6. CARTIC -Articulation/pronunciation (current)

0 = Understood by anybody (i.e., clear enunciation of most sounds, but may include a few consonant omissions or substitutions)

1 = Understood better by family than others because of difficulty with some sounds, but mostly comprehensible to strangers at first encounter

2 = Definite articulation difficulties such that some words are very difficult for strangers to understand until they get to know her/him

3 = Strangers find speech almost impossible to understand or parents have significant difficulties understanding because of articulation

8 = N/A; no speech or subject below age 4.0 years

9 = N/K or not asked

7. ARTICF5 -Articulation/pronunciation (ever)

0 = Understood by anybody (i.e., clear enunciation of most sounds, but may include a few consonant omissions or substitutions)

1 = Understood better by family than others because of difficulty with some sounds, but mostly comprehensible to strangers at first encounter

2 = Definite articulation difficulties such that some words are very difficult for strangers to understand until they get to know her/him

3 = Strangers find speech almost impossible to understand or parents have significant difficulties understanding because of articulation

8 = N/A; no speech or subject below age 4.0 years

9 = N/K or not asked

8. CSTEREO -Stereotyped utterances and delayed echolalia (current)

0 = Rarely or never uses stereotyped phrases

1 = Speech tends/tended to be more repetitive than most subjects at the same level of complexity, but not stereotyped in an odd or unusual way; or occasional stereotyped utterances, but consistent productive language as well

2 = Often uses/used stereotyped utterances with productive language as well

3 = Phrases are almost exclusively stereotyped utterances

8 = N/A; no phrases (code "1" or "2" on Item 30)

9 = N/K or not asked

9. ESTEREO -Stereotyped utterances and delayed echolalia (ever)

0 = Rarely or never uses stereotyped phrases

1 = Speech tends/tended to be more repetitive than most subjects at the same level of complexity, but not stereotyped in an odd or unusual way; or occasional stereotyped utterances, but consistent productive language as well

2 = Often uses/used stereotyped utterances with productive language as well

3 = Phrases are almost exclusively stereotyped utterances

8 = N/A; no phrases (code "1" or "2" on Item 30)

9 = N/K or not asked

10. CCHAT -Social vocalization (current)

0 = Verbalizes or chats with clear social quality of talking to be friendly or to express interest, rather than to make needs known

1 = Some social use of speech in response to caregiver or to get attention with no other obvious motivation, but limited in frequency or range of contexts

2 = Uses some speech to alert caregiver to immediate needs or wants, but little or no purely social use of verbalization

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

11. CHAT5 -Social vocalization (ever)

0 = Verbalizes or chats with clear social quality of talking to be friendly or to express interest, rather than to make needs known

1 = Some social use of speech in response to caregiver or to get attention with no other obvious motivation, but limited in frequency or range of contexts

2 = Uses some speech to alert caregiver to immediate needs or wants, but little or no purely social use of verbalization

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

12. CCONVER -Reciprocal conversation, at whatever verbal level of complexity possible (current)

0 = Conversation flows, including both offering information and building on other person's response in a manner that leads to ongoing dialogue

1 = Occasional reciprocal conversation, but less frequent than normal or limited in flexibility or topics

2 = Little or no reciprocal conversation; others find it difficult to build a conversation even if there is apparent positive or social talk by subject; subject fails to follow anyone else's conversation topic; may ask or answer questions but not as part of a single interchange

3 = Very little spontaneous speech

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

13. CONVER5 -Reciprocal conversation, at whatever verbal level of complexity possible (ever)

0 = Conversation flows, including both offering information and building on other person's response in a manner that leads to ongoing dialogue

1 = Occasional reciprocal conversation, but less frequent than normal or limited in flexibility or topics

2 = Little or no reciprocal conversation; others find it difficult to build a conversation even if there is apparent positive or social talk by subject; subject fails to follow anyone else's conversation topic; may ask or answer questions but not as part of a single interchange

3 = Very little spontaneous speech

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

14. CINAPPQ -Inappropriate questions or statements (current)

0 = No or very rare use of questions/statements inappropriate to conversation or setting

1 = Use of some questions/statements regardless of situation; questions or statements are slightly inappropriate and may be repetitious, but are not usually very odd or highly embarrassing

2 = Frequent use of questions/statements that are odd or obviously inappropriate to the situation

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

15. EINAPPQ -Inappropriate questions or statements (ever)

0 = No or very rare use of questions/statements inappropriate to conversation or setting

1 = Use of some questions/statements regardless of situation; questions or statements are slightly inappropriate and may be repetitious, but are not usually very odd or highly embarrassing

2 = Frequent use of questions/statements that are odd or obviously inappropriate to the situation

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

16. CPRON -Pronominal reversal (current)

0 = No confusion between 1st and 2nd or 3rd person after phrase speech with pronoun use established

1 = Refers/referred to self by name instead of "I" after phrase speech established, but no persistent "you/s/he-I" confusion

2 = "You/I" or "s/he/I" confusion after phrase speech established, but "you" or "s/he" not used with intonation of a question

3 = Pronominal confusion when asking a question in which "you" or "s/he" is used for "I"

7 = Other types of pronominal confusion (other than "I/me"), such as "he/you"

8 = N/A (code "1" or "2" on Item 30 or no pronouns used)

9 = N/K or not asked

17. EPRON -Pronominal reversal (ever)

0 = No confusion between 1st and 2nd or 3rd person after phrase speech with pronoun use established

1 = Refers/referred to self by name instead of "I" after phrase speech established, but no persistent "you/s/he-I" confusion

2 = "You/I" or "s/he/I" confusion after phrase speech established, but "you" or "s/he" not used with intonation of a question

3 = Pronominal confusion when asking a question in which "you" or "s/he" is used for "I"

7 = Other types of pronominal confusion (other than "I/me"), such as "he/you"

8 = N/A (code "1" or "2" on Item 30 or no pronouns used)

9 = N/K or not asked

18. CNEOID -Neologisms/idiosyncratic language (current)

0 = No use of neologisms or idiosyncratic language

1 = Occasional use of neologisms and/or idiosyncratic words and phrases used consistently over a period of time

2 = Regular use of neologisms and/or idiosyncratic ways of saying things, including generalization of unusual term to reference beyond the example that may have fostered the initial idiosyncratic word or phrase

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

19. ENEOID -Neologisms/idiosyncratic language (ever)

0 = No use of neologisms or idiosyncratic language

1 = Occasional use of neologisms and/or idiosyncratic words and phrases used consistently over a period of time

2 = Regular use of neologisms and/or idiosyncratic ways of saying things, including generalization of unusual term to reference beyond the example that may have fostered the initial idiosyncratic word or phrase

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

20. CVERRIT -Verbal rituals (current)

0 = No verbal rituals

1 = Tendency to say things in ritualized way or to require others to do so, but no indication that this is compulsive, and readily stops if asked to do so

2 = Subject has to say one or more things in a special way; rituals may intrude on family life, may involve other family members and some distress at interruption, or may cause some disturbance or minor reorganization of family life that can be tolerated by most families

3 = As for "2," but with marked difficulty to control and marked intrusiveness on family life; family members involved to a degree that causes definite social impairment, disruption, or prevention of some family activities; serious distress at any attempted interruption

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

21. EVERRIT -Verbal rituals (ever)

0 = No verbal rituals

1 = Tendency to say things in ritualized way or to require others to do so, but no indication that this is compulsive, and readily stops if asked to do so

2 = Subject has to say one or more things in a special way; rituals may intrude on family life, may involve other family members and some distress at interruption, or may cause some disturbance or minor reorganization of family life that can be tolerated by most families

3 = As for "2," but with marked difficulty to control and marked intrusiveness on family life; family members involved to a degree that causes definite social impairment, disruption, or prevention of some family activities; serious distress at any attempted interruption

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

22. CINR -Intonation/volume/rhythm/rate (current)

0 = Normal, appropriately varying intonation, reasonable volume, and normal rate of speech, with regular rhythm coordinated with breathing

1 = Speech that shows one or other of the abnormalities listed under "2," but not obviously peculiar and no interference with intelligibility

2 = Speech that is clearly abnormal in terms of any or all of the following: (a) odd intonation or inappropriate pitch and stress; (b) markedly flat and toneless or mechanical speech; (c) consistently abnormal volume that lacks modulation; (d) inappropriate, poorly modulated rate or rhythm (either unusually slow or halting or unusually rapid or jerky) to a degree that creates some interference with intelligibility

3 = Speech that is frequently obviously peculiar or difficult to understand because of abnormalities of type specified under "2"

7 = Stutter or stammer; abnormal intonation due to severe hearing impairment; or intonation abnormality due to neurological disorder causing motor dysfunction

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

23. EINR -Intonation/volume/rhythm/rate (ever)

0 = Normal, appropriately varying intonation, reasonable volume, and normal rate of speech, with regular rhythm coordinated with breathing

1 = Speech that shows one or other of the abnormalities listed under "2," but not obviously peculiar and no interference with intelligibility

2 = Speech that is clearly abnormal in terms of any or all of the following: (a) odd intonation or inappropriate pitch and stress; (b) markedly flat and toneless or mechanical speech; (c) consistently abnormal volume that lacks modulation; (d) inappropriate, poorly modulated rate or rhythm (either unusually slow or halting or unusually rapid or jerky) to a degree that creates some interference with intelligibility

3 = Speech that is frequently obviously peculiar or difficult to understand because of abnormalities of type specified under "2"

7 = Stutter or stammer; abnormal intonation due to severe hearing impairment; or intonation abnormality due to neurological disorder causing motor dysfunction

8 = N/A (code "1" or "2" on Item 30)

9 = N/K or not asked

24. CSPEECH -Current communication speech (current)

0 = Speech, at whatever level attained, used frequently and communicatively in a variety of contexts, including some reference to events not present (Do not include requests here)

1 = Some communicative use of words (i.e., words used regularly to communicate, with or without an abnormal element), but somewhat restricted in frequency or contexts

2 = Some spontaneous words and/or echolalic language, but with limited communicative use

3 = Little or no communicative language (including exclusively non-communicative echolalia), though subject has some language

8 = N/A (code "1" or "2" on Item 30, or under 5 years of age for 5,0 year coding)

9 = N/K or not asked

25. SPEECH5 -Current communication speech (ever)

0 = Speech used frequently and communicatively in a variety of contexts

1 = Some communicative use of words

2 = Some spontaneous words and/or echolalic language

3 = Little or no communicative language

8 = N/A

9 = N/K or not asked

26. CPOINT -Pointing to express interest (current)

0 = Spontaneously points at objects at a distance with finger to express interest, using coordinated eye gaze in order to communicate

1 = Makes some limited attempt to express interest by pointing, but with limited flexibility and/or lack of coordination (e.g., uses arm or points with finger, but without consistent coordination with eye gaze)

2 = Little or no spontaneous unprompted attempts to point to express interest in objects at a distance

8 = N/A

9 = N/K or not asked

27. POINT5 -Pointing to express interest (ever)

0 = Spontaneously points at objects at a distance with finger to express interest, using coordinated eye gaze in order to communicate

1 = Makes some limited attempt to express interest by pointing, but with limited flexibility and/or lack of coordination (e.g., uses arm or points with finger, but without consistent coordination with eye gaze)

2 = Little or no spontaneous unprompted attempts to point to express interest in objects at a distance

8 = N/A

9 = N/K or not asked

28. CNOD -Nodding (current)

0 = Consistently nods spontaneously

1 = Sometimes nods spontaneously

2 = Never nods spontaneously

8 = N/A

9 = N/K or not asked

29. NOD5 -Nodding (most abnormal)

0 = Consistently nods spontaneously

1 = Sometimes nods spontaneously

2 = Never nods spontaneously

8 = N/A

9 = N/K or not asked

30. CHSHAKE -Head shaking (current)

0 = Consistently shakes head spontaneously

1 = Sometimes shakes head spontaneously

2 = Never shakes head spontaneously

8 = N/A

9 = N/K or not asked

31. HSHAKE5 -Head shaking (most abnormal)

0 = Consistently shakes head spontaneously

1 = Sometimes shakes head spontaneously

2 = Never shakes head spontaneously

8 = N/A



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

9 = N/K or not asked

32. CINSGES -Conventional/instrumental gestures (current)

0 = Appropriate and spontaneous use of a variety of conventional or instrumental gestures

1 = Spontaneous use of conventional or instrumental gestures, but limited in range or contexts

2 = Inconsistent spontaneous use, or use of elicited or well-rehearsed simple conventional or instrumental gestures only

3 = No use of conventional or instrumental gestures

8 = N/A

9 = N/K or not asked

33. INSGES5 -Conventional/instrumental gestures (most abnormal)

0 = Appropriate and spontaneous use of a variety of conventional or instrumental gestures

1 = Spontaneous use of conventional or instrumental gestures, but limited in range or contexts

2 = Inconsistent spontaneous use, or use of elicited or well-rehearsed simple conventional or instrumental gestures only

3 = No use of conventional or instrumental gestures

8 = N/A

9 = N/K or not asked

34. AVOICE5 -Attention to voice (most abnormal)

0 = Usually looks up and pays attention when spoken to in a positive manner in contexts other than to do something that s/he may not want to do

1 = Does not consistently appear to pay attention (e.g., might look up briefly, but little sustained attention), but sometimes responds to what is said or responds on occasion only to firm, loud voice

2 = Usually does not look up or pay attention when spoken to, and does not respond to what is said; or responds to her/his name only or when her/his attention is caught very deliberately

3 = Rarely responds, although hearing normal

8 = N/A

9 = N/K or not asked

35. CIMIT -Spontaneous imitation of actions (current)

0 = Spontaneous imitation of a varied range of non-taught actions, at least some of which are incorporated into play outside the context of the observed behavior of the imitated person

1 = Some indication of spontaneous imitation that goes beyond copying a frequent use of an object, but not of sufficient flexibility or number to meet the criteria for "0"

2 = Spontaneous imitation limited to a few familiar routines that are not incorporated into play; including frequent appropriate use of an object probably learned through imitation (e.g., mowing lawn with a toy mower)

3 = Very rare or no spontaneous imitation

8 = N/A

9 = N/K or not asked

36. IMIT5 -Spontaneous imitation of actions (ever)

0 = Spontaneous imitation of a varied range of non-taught actions, at least some of which are incorporated into play outside the context of the observed behavior of the imitated person

1 = Some indication of spontaneous imitation that goes beyond copying a frequent use of an object, but not of sufficient flexibility or number to meet the criteria for "0"

2 = Spontaneous imitation limited to a few familiar routines that are not incorporated into play; including frequent appropriate use of an object probably learned through imitation (e.g., mowing lawn with a toy mower)

3 = Very rare or no spontaneous imitation

8 = N/A

9 = N/K or not asked

37. CPLAY -Imaginative play (current)

0 = Variety of pretend play, including use of dolls/animals/toys as self-initiating agents

1 = Some pretend play, including actions directed to dolls or cars, etc., but limited in variety or frequency

2 = Occasional, spontaneous pretend actions or highly repetitive pretend play (which may be frequent), or only play that has been taught by others

3 = No pretend play

8 = N/A

9 = N/K or not asked

38. PLAY5 -Imaginative play (ever)

0 = Variety of pretend play, including use of dolls/animals/toys as self-initiating agents

1 = Some pretend play, including actions directed to dolls or cars, etc., but limited in variety or frequency

2 = Occasional, spontaneous pretend actions or highly repetitive pretend play (which may be frequent), or only play that has been taught by others

3 = No pretend play

8 = N/A

9 = N/K or not asked

39. CPEERPL -Imaginative play with peers (current)

0 = Imaginative, cooperative play with other children in which the subject both takes the lead and follows another child in spontaneous, pretend activities

1 = Some participation in pretend play with another child, but not truly reciprocal and/or pretending is very limited in variety

2 = Some play with other children, but little or no pretending

3 = No play with other children or no pretend play even on own

8 = N/A

9 = N/K or not asked

40. PEERPL5 -Imaginative play with peers (ever)

0 = Imaginative, cooperative play with other children in which the subject both takes the lead and follows another child in spontaneous, pretend activities

1 = Some participation in pretend play with another child, but not truly reciprocal and/or pretending is very limited in variety

2 = Some play with other children, but little or no pretending

3 = No play with other children or no pretend play even on own

8 = N/A

9 = N/K or not asked

41. GAZE5 -Direct gaze (most abnormal)

0 = Normal reciprocal direct gaze used to communicate across a range of situations and people

1 = Definite direct gaze, but only of brief duration or not consistent during social interactions

2 = Uncertain/occasional direct gaze, or gaze rarely used during social interactions

3 = Unusual or odd use of gaze

8 = N/A

9 = N/K or not asked

42. CSSMILE -Social smiling (current)

0 = Regularly predictable, reciprocal, social smiles in response to the smiles of a variety of people besides parent/caregiver

1 = Some evidence of reciprocal social smiling, but not sufficient to code "0"

2 = Some evidence of smiling while looking at people, but generally not reciprocal; code here if smiles only to parent/caregiver, smiles only upon request, or smiles in odd situations or odd ways

3 = Little or no smiling at people, though may smile at other things

8 = N/A

9 = N/K or not asked

43. SSMILE5 -Social smiling (most abnormal)

0 = Regularly predictable, reciprocal, social smiles in response to the smiles of a variety of people besides parent/caregiver

1 = Some evidence of reciprocal social smiling, but not sufficient to code "0"

2 = Some evidence of smiling while looking at people, but generally not reciprocal; code here if smiles only to parent/caregiver, smiles only upon request, or smiles in odd situations or odd ways

3 = Little or no smiling at people, though may smile at other things

8 = N/A

9 = N/K or not asked

44. CSHOW -Showing and directing attention (current)

0 = Regular showing of objects by bringing things to parent/caregiver and directing her/his attention, with no manifest motive other than sharing

1 = Possible showing as described above, but not sufficiently frequent or of purely communicative quality to meet criteria for "0"

2 = Some bringing things to parent/caregiver and/or showing, but associated with preoccupations, food, or need for help

3 = Rare or no social approaches of this type

8 = N/A

9 = N/K or not asked

45. SHOW5 -Showing and directing attention (most abnormal)

0 = Regular showing of objects by bringing things to parent/caregiver and directing her/his attention, with no manifest motive other than sharing

1 = Possible showing as described above, but not sufficiently frequent or of purely communicative quality to meet criteria for "0"

2 = Some bringing things to parent/caregiver and/or showing, but associated with preoccupations, food, or need for help

3 = Rare or no social approaches of this type

8 = N/A

9 = N/K or not asked

46. COSHARE -Offering to share (current)

0 = Frequent, spontaneous, and varied offers to share different sorts of objects (e.g., toys, comforters) with other people

1 = Some spontaneous offers to share, but limited in number of contexts or frequency (must be more than food)

2 = Will sometimes share if requested, but not spontaneously, or spontaneous sharing of food only

3 = No sharing

8 = N/A

9 = N/K or not asked

47. OSHARE5 -Offering to share (ever)

0 = Frequent, spontaneous, and varied offers to share different sorts of objects (e.g., toys, comforters) with other people

1 = Some spontaneous offers to share, but limited in number of contexts or frequency (must be more than food)

2 = Will sometimes share if requested, but not spontaneously, or spontaneous sharing of food only

3 = No sharing

8 = N/A

9 = N/K or not asked

48. CSHARE -Seeking to share his/her enjoyment with others (current)

0 = Frequent attempts across a variety of contexts to direct several other people's attention to things that s/he enjoys or has done well (must be with more than one parent)

1 = Some attempts to share enjoyment, but limited in number or variety or spontaneity, or lacking clear quality of shared pleasure

2 = Few or no attempts to share enjoyment

8 = N/A

9 = N/K or not asked

49. SHARE5 -Seeking to share his/her enjoyment with others (ever)

0 = Frequent attempts across a variety of contexts to direct several other people's attention to things that s/he enjoys or has done well (must be with more than one parent)

1 = Some attempts to share enjoyment, but limited in number or variety or spontaneity, or lacking clear quality of shared pleasure

2 = Few or no attempts to share enjoyment

8 = N/A

9 = N/K or not asked

50. COCOMF -Offers comfort (current)

0 = Flexibly and spontaneously offers comfort in a range of circumstances and ways, for example, by gesture, touching, or vocalization, or offers of objects (e.g., blanket); must include change in facial expression

1 = Has partial response (e.g., stands nearby and looks concerned) or indirect physical approach (e.g., comes to sit in lap, but with no clear attempt to comfort); or offers comfort in response to exaggerated expression (e.g., to pretend crying) or in one routine situation (e.g., to baby sister when she's hungry)

2 = Rarely offers comfort or only in odd ways

3 = Never offers comfort to others

8 = N/A

9 = N/K or not asked

51. OCOMF5 -Offers comfort (most abnormal)

0 = Flexibly and spontaneously offers comfort in a range of circumstances and ways, for example, by gesture, touching, or vocalization, or offers of objects (e.g., blanket); must include change in facial expression

1 = Has partial response (e.g., stands nearby and looks concerned) or indirect physical approach (e.g., comes to sit in lap, but with no clear attempt to comfort); or offers comfort in response to exaggerated expression (e.g., to pretend crying) or in one routine situation (e.g., to baby sister when she's hungry)

2 = Rarely offers comfort or only in odd ways

3 = Never offers comfort to others

8 = N/A

9 = N/K or not asked

52. CQUALOV -Quality of social overtures (current)

0 = Consistently uses coordinated eye gaze with accompanying vocalization in typical situations when motivated to communicate

1 = May use eye gaze or vocalization, but these are weakly integrated

2 = Rarely shows well-coordinated, focused social intentionality involving eye gaze or vocalization, or shows this in odd ways

3 = Shows no coordination of eye gaze and vocalization

8 = N/A

9 = N/K or not asked

53. QUALOV5 -Quality of social overtures (most abnormal)

0 = Consistently uses coordinated eye gaze with accompanying vocalization in typical situations when motivated to communicate

1 = May use eye gaze or vocalization, but these are weakly integrated

2 = Rarely shows well-coordinated, focused social intentionality involving eye gaze or vocalization, or shows this in odd ways

3 = Shows no coordination of eye gaze and vocalization

8 = N/A

9 = N/K or not asked

54. CRFACEX -Range of facial expressions used to communicate (current)

0 = Full range of facial expression

1 = Somewhat limited facial expression; may be rather stilted, exaggerated, in manner

2 = Markedly limited range of facial expressions or tendency to have just one facial expression (e.g., happy) for all circumstances

3 = Facial expression showing little or no indication of emotion of any kind

8 = N/A

9 = N/K or not asked

55. RFACEX5 -Range of facial expressions used to communicate (most abnormal)

0 = Full range of facial expression

1 = Somewhat limited facial expression; may be rather stilted, exaggerated, in manner

2 = Markedly limited range of facial expressions or tendency to have just one facial expression (e.g., happy) for all circumstances

3 = Facial expression showing little or no indication of emotion of any kind

8 = N/A

9 = N/K or not asked

56. CINAPFE -Inappropriate facial expressions (current)

0 = Facial expressions almost always appropriate to mood, situation, and context

1 = Facial expressions slightly or occasionally inappropriate or odd

2 = Facial expressions obviously inappropriate in several different situations

8 = N/A (almost no variation in facial expression, appropriate or inappropriate, as in coding of "3" in Item 57)

9 = N/K or not asked

57. EINAPFE -Inappropriate facial expressions (ever)

0 = Facial expressions almost always appropriate to mood, situation, and context

1 = Facial expressions slightly or occasionally inappropriate or odd

2 = Facial expressions obviously inappropriate in several different situations

8 = N/A (almost no variation in facial expression, appropriate or inappropriate, as in coding of "3" in Item 57)

9 = N/K or not asked

58. CQRESP -Appropriateness of social responses (current)

0 = Appropriate response to overtures by familiar and unfamiliar adults

1 = Some clear positive responses and interactions, but not consistent

2 = Responds to parents/caregiver and others in familiar settings, but responses are stereotyped, inappropriate, or very limited

3 = Little or no interest in or response to people, except parents/caregiver or very familiar significant others

8 = N/A

9 = N/K or not asked

59. QRESP5 -Appropriateness of social responses (most abnormal)

- 0 = Appropriate response to overtures by familiar and unfamiliar adults
- 1 = Some clear positive responses and interactions, but not consistent
- 2 = Responds to parents/caregiver and others in familiar settings, but responses are stereotyped, inappropriate, or very limited
- 3 = Little or no interest in or response to people, except parents/caregiver or very familiar significant others
- 8 = N/A
- 9 = N/K or not asked

60. CINITIA -Initiation of appropriate activities (current)

- 0 = Spontaneously takes up a range of appropriate play activities, without prompting or organization by other people
- 1 = Spontaneously initiates only a limited range of appropriate activities
- 2 = Engages in passive, but otherwise appropriate, activity, such as watching TV or listening to the radio
- 3 = Engages in no active or constructive play, or engages only in repetitive activities or motor stereotypes
- 8 = N/A
- 9 = N/K or not asked

61. INITIA5 -Initiation of appropriate activities (most abnormal)

- 0 = Spontaneously takes up a range of appropriate play activities, without prompting or organization by other people
- 1 = Spontaneously initiates only a limited range of appropriate activities
- 2 = Engages in passive, but otherwise appropriate, activity, such as watching TV or listening to the radio
- 3 = Engages in no active or constructive play, or engages only in repetitive activities or motor stereotypes
- 8 = N/A

9 = N/K or not asked

62. CSOPLAY -Imitative social play (current)

0 = Normal social play, including clear evidence that the child initiates and responds to simple infant social games and can take both parts

1 = Some to-and-fro (i.e., reciprocal) social play, but limited in amount, duration, or contexts in which shown (e.g., only plays peekaboo or pat-a-cake with parents/caregiver)

2 = Little to-and-fro social play (e.g., plays peekaboo or pat-a-cake in a limited way only, but not reciprocal)

3 = No evidence of to-and-fro social play

8 = N/A

9 = N/K or not asked

63. SOPLAY5 -Imitative social play (most abnormal)

0 = Normal social play, including clear evidence that the child initiates and responds to simple infant social games and can take both parts

1 = Some to-and-fro (i.e., reciprocal) social play, but limited in amount, duration, or contexts in which shown (e.g., only plays peekaboo or pat-a-cake with parents/caregiver)

2 = Little to-and-fro social play (e.g., plays peekaboo or pat-a-cake in a limited way only, but not reciprocal)

3 = No evidence of to-and-fro social play

8 = N/A

9 = N/K or not asked

64. CINTCH -Interest in children (current)

0 = Often watches other children and sometimes makes a clear effort to approach them or get their attention

1 = Usually watches other children or indicates interest in them to parent/caregiver in some way (e.g., by pointing, vocalizing, or trying to imitate what they are doing, but no attempt to seek them out); or approaches other children without trying to get their attention

2 = Occasionally watches other children, but almost never tries to approach them, or to direct parent's/caregiver's attention to them, or copy them

3 = Shows no, or almost no, interest in other children

8 = N/A

9 = N/K or not asked

65. INTCH5 -Interest in children (most abnormal)

0 = Often watches other children and sometimes makes a clear effort to approach them or get their attention

1 = Usually watches other children or indicates interest in them to parent/caregiver in some way (e.g., by pointing, vocalizing, or trying to imitate what they are doing, but no attempt to seek them out); or approaches other children without trying to get their attention

2 = Occasionally watches other children, but almost never tries to approach them, or to direct parent's/caregiver's attention to them, or copy them

3 = Shows no, or almost no, interest in other children

8 = N/A

9 = N/K or not asked

66. CRESPCH -Response to approaches of other children (current)

0 = Generally responsive to other children's approaches, although may be hesitant initially if other children are too rough or intrusive; sometimes makes a clear effort to keep an interaction going with a child other than a sibling by gesturing, vocalizing, offering an object, etc.

1 = Sometimes responsive to other children's approaches, but response is limited, somewhat unpredictable, or only to a sibling or a very familiar child

2 = Rarely or never responds to the approach of even a familiar child (although may show interest in non approaching children or babies)

3 = Consistently and persistently avoids approaches of other children

8 = N/A

9 = N/K or not asked

67. RESPCH5 -Response to approaches of other children (most abnormal)

0 = Generally responsive to other children's approaches, although may be hesitant initially if other children are too rough or intrusive; sometimes makes a clear effort to keep an interaction going with a child other than a sibling by gesturing, vocalizing, offering an object, etc.

1 = Sometimes responsive to other children's approaches, but response is limited, somewhat unpredictable, or only to a sibling or a very familiar child

2 = Rarely or never responds to the approach of even a familiar child (although may show interest in non approaching children or babies)

3 = Consistently and persistently avoids approaches of other children

8 = N/A

9 = N/K or not asked

68. CGRPLAY -Group play with peers (current)

0 = Actively seeks and plays cooperatively in several different groups (three or more subjects) in a variety of activities or situations

1 = Some cooperative play, but of insufficient initiative, flexibility, frequency, or variety to code "0"

2 = Enjoys "parallel" active play (such as jumping in turn on 3 trampoline or falling down with others during Ring A Ring O'Roses), but little or no cooperative play

3 = Seeks no play that involves participation in groups of other children, though may chase or play catch

8 = N/A

9 = N/K or not asked

69. GRPLAY5 -Group play with peers (most abnormal)

0 = Actively seeks and plays cooperatively in several different groups (three or more subjects) in a variety of activities or situations

1 = Some cooperative play, but of insufficient initiative, flexibility, frequency, or variety to code "0"

2 = Enjoys "parallel" active play (such as jumping in turn on 3 trampoline or falling down with others during Ring A Ring O'Roses), but little or no cooperative play

3 = Seeks no play that involves participation in groups of other children, though may chase or play catch

8 = N/A

9 = N/K or not asked

70. CFRIEND -Friendships (current)

0 = One or more relationships with person in approximately own age group with whom subject shares nonstereotyped activities of personal variety; whom subject sees outside prearranged group (such as club); and with whom there is definite reciprocity and mutual responsiveness

1 = One or more relationships that involve some personal shared activities outside a prearranged situation, with some initiative taken by subject, but limited in terms of restricted interests (e.g., model railways) or less than normal responsiveness/reciprocity

2 = People with whom subject has some kind of personal relationship involving seeking of contact, but only in group situation (such as club, church, etc.) or in school or at work

3 = No peer relationships that involve selectivity and sharing

8 = N/A (subject is outside the specified age range, or there has been an unusually serious lack of opportunity for peer contact)

9 = N/K or not asked

71. FREN15 -Friendships (most abnormal)

0 = One or more relationships with person in approximately own age group with whom subject shares nonstereotyped activities of personal variety; whom subject sees outside prearranged group (such as club); and with whom there is definite reciprocity and mutual responsiveness

1 = One or more relationships that involve some personal shared activities outside a prearranged situation, with some initiative taken by subject, but limited in terms of restricted interests (e.g., model railways) or less than normal responsiveness/reciprocity

2 = People with whom subject has some kind of personal relationship involving seeking of contact, but only in group situation (such as club, church, etc.) or in school or at work

3 = No peer relationships that involve selectivity and sharing

8 = N/A (subject is outside the specified age range, or there has been an unusually serious lack of opportunity for peer contact)

9 = N/K or not asked

72. CSOCDIS -Social disinhibition (current)

0 = Normal social inhibition

1 = Occasional cheekiness or disinhibition more than others at same developmental level, but not to the extent of embarrassment: somewhat socially naive or imperceptive for developmental level

2 = Definite lack of appreciation of social cues, contexts, or requirements: definitely lacks normal social inhibitions and sometimes behaves in socially embarrassing ways; fails to modulate behavior according to social context

3 = Marked social disinhibition: appears unaware of social cues and social requirements so that behavior frequently embarrassing or inappropriate

8 = N/A

9 = N/K or not asked

73. SOCDIS5 -Social disinhibition (most abnormal)

0 = Normal social inhibition

1 = Occasional cheekiness or disinhibition more than others at same developmental level, but not to the extent of embarrassment: somewhat socially naive or imperceptive for developmental level

2 = Definite lack of appreciation of social cues, contexts, or requirements: definitely lacks normal social inhibitions and sometimes behaves in socially embarrassing ways; fails to modulate behavior according to social context

3 = Marked social disinhibition: appears unaware of social cues and social requirements so that behavior frequently embarrassing or inappropriate

8 = N/A

9 = N/K or not asked

74. CUNPROC -Unusual preoccupations (current)

0 = No unusual preoccupations

1 = Unusual preoccupation that does not interfere with significant activities of family life or does not cause social impairment of the subject

2 = Definite, repetitive preoccupation that intrudes into family life, but does not disrupt it significantly; or definite, repetitive preoccupation that does not cause substantial interference with social functioning, but which does constrain or intrude upon subject's other activities

3 = Definite preoccupation that causes substantial interference or social impairment and severely limits the subject's other activities

9 = N/K or not asked

75. EUNPROC -Unusual preoccupations (ever)

0 = No unusual preoccupations

1 = Unusual preoccupation that does not interfere with significant activities of family life or does not cause social impairment of the subject

2 = Definite, repetitive preoccupation that intrudes into family life, but does not disrupt it significantly; or definite, repetitive preoccupation that does not cause substantial interference with social functioning, but which does constrain or intrude upon subject's other activities

3 = Definite preoccupation that causes substantial interference or social impairment and severely limits the subject's other activities

9 = N/K or not asked

76. CCIRINT -Circumscribed interests (current)

0 = No circumscribed interests

1 = Special interest of unusual degree, but not definitely intrusive into or constraining of the subject's or family's other activities

2 = Definite circumscribed interest that does not cause substantial interference with social functioning, but which does constrain or intrude upon subject's or family's other activities

3 = Definite circumscribed interest that causes clear-cut social impairment

8 = N/A (age below 3.0 years)

9 = N/K or not asked

77. ECIRINT -Circumscribed interests (ever)

0 = No circumscribed interests

1 = Special interest of unusual degree, but not definitely intrusive into or constraining of the subject's or family's other activities

2 = Definite circumscribed interest that does not cause substantial interference with social functioning, but which does constrain or intrude upon subject's or family's other activities

3 = Definite circumscribed interest that causes clear-cut social impairment

8 = N/A (age below 3.0 years)

9 = N/K or not asked

78. CUSEOBJ -Repetitive use of objects or interest in parts of objects (current)

0 = Little or no repetitive use of objects

1 = some repetitive use of objects (e.g., shaking strings or spinning things), or interest in parts (e.g., turning wheels and dials) or very specific types of objects (e.g., collecting bits of paper), in conjunction with several other activities but not causing social impairment

2 = Play linked to highly stereotypic use of objects or attention to specific parts or types of objects, but which does not constrain or intrude upon subject's other activities

3 = Play linked to highly stereotypic use of objects to an extent that prevents or seriously interferes with other activities

7 = An interest in "infant" toys, such as music boxes or rattles, but play is with a variety of objects and not in a highly stereotypic fashion

8 = N/A (no play with objects)

9 = N/K or not asked

79. EUSEOBJ -Repetitive use of objects or interest in parts of objects (ever)

0 = Little or no repetitive use of objects

1 = some repetitive use of objects (e.g., shaking strings or spinning things), or interest in parts (e.g., turning wheels and dials) or very specific types of objects (e.g.,

collecting bits of paper), in conjunction with several other activities but not causing social impairment

2 = Play linked to highly stereotypic use of objects or attention to specific parts or types of objects, but which does not constrain or intrude upon subject's other activities

3 = Play linked to highly stereotypic use of objects to an extent that prevents or seriously interferes with other activities

7 = An interest in "infant" toys, such as music boxes or rattles, but play is with a variety of objects and not in a highly stereotypic fashion

8 = N/A (no play with objects)

9 = N/K or not asked

80. CCRIT -Compulsions/rituals (current)

0 = No compulsions/rituals

1 = Some activities with unusually fixed sequences, but no activity that appears compulsive in quality

2 = One or more activities that subject has to perform in a special way; subject appears to be under pressure or becomes anxious if activity disrupted, or family goes to unusual lengths to avoid interrupting ritual or to make sure subject is forewarned if it is necessary to interrupt her/him; compulsive quality present, but little interference with family life or social impairment

3 = One or more activities that subject has to perform in a special way; subject appears to be under marked pressure or becomes extremely anxious or distressed if activity disrupted; degree of compulsive quality intrudes upon family life or causes definite social impairment to subject

9 = N/K or not asked

81. ECRIT -Compulsions/rituals (ever)

0 = No compulsions/rituals

1 = Some activities with unusually fixed sequences, but no activity that appears compulsive in quality

2 = One or more activities that subject has to perform in a special way; subject appears to be under pressure or becomes anxious if activity disrupted, or family goes to unusual lengths to avoid interrupting ritual or to make sure subject is forewarned if it is necessary to interrupt her/him; compulsive quality present, but little interference with family life or social impairment

3 = One or more activities that subject has to perform in a special way; subject appears to be under marked pressure or becomes extremely anxious or distressed if activity disrupted; degree of compulsive quality intrudes upon family life or causes definite social impairment to subject

9 = N/K or not asked

82. CUNSENS -Unusual sensory interests (current)

0 = Has no unusual sensory interests

1 = Shows one or two unusual interests regularly

2 = Has unusual sensory interest that takes up a major amount of time or prevents or limits alternative use of that material in its ordinary function

9 = N/K or not asked

83. EUNSENS -Unusual sensory interests (ever)

0 = Has no unusual sensory interests

1 = Shows one or two unusual interests regularly

2 = Has unusual sensory interest that takes up a major amount of time or prevents or limits alternative use of that material in its ordinary function

9 = N/K or not asked

84. CNOISE -Undue general sensitivity to noise (current)

0 = No general sensitivity to noise

1 = Slight sensitivity to noise: somewhat sensitive to loud sounds such as the vacuum cleaner, motorbikes, or other appliances

2 = Definite sensitivity to noises that are not distressing to most other people, the sensitivity being accompanied by a clear behavioral change (such as avoidance, hands over ears, or crying)

3 = Definite sensitivity to noises to the extent that subject's distress/disturbance in relation to certain noises interferes with family or household routines

9 = N/K or not asked

85. ENOISE -Undue general sensitivity to noise (ever)

0 = No general sensitivity to noise

1 = Slight sensitivity to noise: somewhat sensitive to loud sounds such as the vacuum cleaner, motorbikes, or other appliances

2 = Definite sensitivity to noises that are not distressing to most other people, the sensitivity being accompanied by a clear behavioral change (such as avoidance, hands over ears, or crying)

3 = Definite sensitivity to noises to the extent that subject's distress/disturbance in relation to certain noises interferes with family or household routines

9 = N/K or not asked

86. CABINR -Abnormal idiosyncratic negative response to specific sensory stimuli (current)

0 = No predictable, abnormal, idiosyncratic, negative response to specific sensory stimuli

1 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli, but reaction mild or controllable so does not give rise to avoidance or to any interference with ordinary life

2 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli; some intrusion into ordinary activities so there are occasional

tantrums/disturbances or attempts by family to avoid subject being exposed to specific stimulus; however, no substantial interference with general pattern of family life

3 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli that causes substantial interference with family life or totally, or almost totally, prevents some activity

9 = N/K or not asked

87. EABINR -Abnormal idiosyncratic negative response to specific sensory stimuli (ever)

0 = No predictable, abnormal, idiosyncratic, negative response to specific sensory stimuli

1 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli, but reaction mild or controllable so does not give rise to avoidance or to any interference with ordinary life

2 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli; some intrusion into ordinary activities so there are occasional tantrums/disturbances or attempts by family to avoid subject being exposed to specific stimulus; however, no substantial interference with general pattern of family life

3 = Predictable, abnormal, idiosyncratic, negative response to one or more specific stimuli that causes substantial interference with family life or totally, or almost totally, prevents some activity

9 = N/K or not asked

88. CCHANGE -Difficulties with minor changes in subject's own routines or personal environment (current)

0 = No difficulties with minor changes in own subject's own routines

1 = Unusually negative reaction to minor changes in subject's own routines, but with no serious distress and little or no interference in family life

2 = Definite, unusual reactions to minor changes in subject's own routines, causing resistance or distress/or family going to unusual lengths to avoid changing minor aspects of subject's routines or to prepare subject for minor changes; but no substantial interference in family life

3 = Definite, unusual, and marked resistance to minor changes in subject's own routines, with substantial interference with or impairment of family activities

9 = N/K or not asked

89. ECHANGE -Difficulties with minor changes in subject's own routines or personal environment (ever)

0 = No difficulties with minor changes in own subject's own routines

1 = Unusually negative reaction to minor changes in subject's own routines, but with no serious distress and little or no interference in family life

2 = Definite, unusual reactions to minor changes in subject's own routines, causing resistance or distress/or family going to unusual lengths to avoid changing minor aspects of subject's routines or to prepare subject for minor changes; but no substantial interference in family life

3 = Definite, unusual, and marked resistance to minor changes in subject's own routines, with substantial interference with or impairment of family activities

9 = N/K or not asked

90. CRESIS -Resistance of trivial changes in the environment, not directly affecting the subject (current)

0 = None

1 = Unusually negative reaction to trivial changes in the environment, but with no serious distress and little or no interference in family life

2 = Definite, unusual reactions to trivial changes in the environment, causing marked distress or causing family to go to unusual lengths to avoid trivial changes in the environment or to prepare subject for such trivial changes; but no substantial interference in family life

3 = Definite, unusual, and marked resistance to trivial changes in the environment, with substantial interference with or impairment of family activities

9 = N/K or not asked

91. ERESIS -Resistance of trivial changes in the environment, not directly affecting the subject (ever)

0 = None

1 = Unusually negative reaction to trivial changes in the environment, but with no serious distress and little or no interference in family life

2 = Definite, unusual reactions to trivial changes in the environment, causing marked distress or causing family to go to unusual lengths to avoid trivial changes in the environment or to prepare subject for such trivial changes; but no substantial interference in family life

3 = Definite, unusual, and marked resistance to trivial changes in the environment, with substantial interference with or impairment of family activities

9 = N/K or not asked

92. CUATT -Unusual attachment to objects (current)

0 = No attachment or attachment only to cuddly object used as comforter

1 = Some attachment to slightly unusual object, such as piece of paper or soft brush, or several similar interchangeable objects, but puts down if asked to do so and can tolerate separation from it; no interference with activities

2 = Attachment to an unusual object causes significant distress on separation or causes caregivers to try to ensure object always readily available for subject because of anticipated distress; occasional interference with activities

3 = Unusual attachment so intrusive that it prevents many everyday activities

6 = Attachment to soother/comforter or blanket or other usual object beyond age 5 or with such intensity that interferes with social functioning or activities (if has also had an unusual attachment, code that instead)

7 = Series of short-lasting (1-3 days) attachments to unusual objects or groups of objects, replaced by new attachment to different kind of unusual object also for short time

9 = N/K or not asked

93. EUATT -Unusual attachment to objects (ever)

0 = No attachment or attachment only to cuddly object used as comforter

1 = Some attachment to slightly unusual object, such as piece of paper or soft brush, or several similar interchangeable objects, but puts down if asked to do so and can tolerate separation from it; no interference with activities

2 = Attachment to an unusual object causes significant distress on separation or causes caregivers to try to ensure object always readily available for subject because of anticipated distress; occasional interference with activities

3 = Unusual attachment so intrusive that it prevents many everyday activities

6 = Attachment to soother/comforter or blanket or other usual object beyond age 5 or with such intensity that interferes with social functioning or activities (if has also had an unusual attachment, code that instead)

7 = Series of short-lasting (1-3 days) attachments to unusual objects or groups of objects, replaced by new attachment to different kind of unusual object also for short time

9 = N/K or not asked

94. CHFMAN -Hand and finger mannerisms (current)

0 = No hand and finger mannerisms

1 = Occasional hand and finger mannerisms only or type not as clearly specified as for rating of "2"; 2 = Definite, frequent hand mannerisms or finger flicking/twisting, but no interference with other activities or distress if interrupted

3 = Marked mannerisms of type specified; associated with social impairment or distress when interrupted or infrequent interruption due to others' concern about subject's reaction

8 = N/A (e.g., physically disabled)

9 = N/K or not asked

95. EHFMAN -Hand and finger mannerisms (ever)

0 = No hand and finger mannerisms

1 = Occasional hand and finger mannerisms only or type not as clearly specified as for rating of "2"; 2 = Definite, frequent hand mannerisms or finger flicking/twisting, but no interference with other activities or distress if interrupted

3 = Marked mannerisms of type specified; associated with social impairment or distress when interrupted or infrequent interruption due to others' concern about subject's reaction

8 = N/A (e.g., physically disabled)

9 = N/K or not asked

96. COTHMAN -Other complex mannerisms or stereotyped body movements, do not include isolated rocking (current)

0 = No hand and finger mannerisms

1 = Occasional hand and finger mannerisms only or type not as clearly specified as for rating of "2"; 2 = Definite, frequent hand mannerisms or finger flicking/twisting, but no interference with other activities or distress if interrupted

3 = Marked mannerisms of type specified; associated with social impairment or distress when interrupted or infrequent interruption due to others' concern about subject's reaction

8 = N/A (e.g., physically disabled)

9 = N/K or not asked

97. EOTHMAN -Other complex mannerisms or stereotyped body movements, do not include isolated rocking (ever)

0 = No complex mannerisms or stereotyped body movements

1 = Only occasional complex mannerisms or stereotyped body movements

2 = Definite, frequent other mannerisms or stereotypies, but will stop without distress if interrupted

3 = Marked mannerisms associated with social impairment

8 = N/A

9 = N/K or not asked

98. CMLHAND -Mid-line hand movements (current)

0 = No midline hand movements

1 = Only occasional midline hand movements or type not as clearly specified as for a rating of "2"

2 = Definite abnormal wringing hand movements mainly in the midline

8 = N/A

9 = N/K or not asked

99. EMLHAND -Mid-line hand movements (ever)

0 = No midline hand movements

1 = Only occasional midline hand movements or type not as clearly specified as for a rating of "2"

2 = Definite abnormal wringing hand movements mainly in the midline

8 = N/A

9 = N/K or not asked

100. CGAIT -Gait (current)

0 = Normal gait

- 1 = Somewhat unusual gait
- 2 = Definitely odd gait (e.g., toe-walking or abnormal bouncing)
- 3 = Gait sufficiently odd to be noticed by others outside family or teachers
- 7 = Gait apparently due to a neurological condition or to severe developmental motor impairment
- 8 = N/A
- 9 = N/K or not asked

101. GAIT5 -Gait (most abnormal)

- 0 = Normal gait
- 1 = Somewhat unusual gait
- 2 = Definitely odd gait (e.g., toe-walking or abnormal bouncing)
- 3 = Gait sufficiently odd to be noticed by others outside family or teachers
- 7 = Gait apparently due to a neurological condition or to severe developmental motor impairment
- 8 = N/A
- 9 = N/K or not asked

102. CAGGFAM -Aggression to caregivers or family members (current)

- 0 = No aggression or only rare episodes; rare aggressiveness, not a significant problem
- 1 = Mild aggressiveness only (threatening without physical contact; or behavior that might represent just unduly rough play or momentary, provoked lashing out)
- 2 = Definite physical aggression involving hitting or biting but no use of implements
- 3 = Violence that involves the use of implements
- 9 = N/K or not asked

103. EAGGFAM -Aggression to caregivers or family members (ever)

0 = No aggression or only rare episodes; rare aggressiveness, not a significant problem

1 = Mild aggressiveness only (threatening without physical contact; or behavior that might represent just unduly rough play or momentary, provoked lashing out)

2 = Definite physical aggression involving hitting or biting but no use of implements

3 = Violence that involves the use of implements

9 = N/K or not asked

104. CAGGOTH -Aggression to non-caregivers or non-family members (current)

0 = Aggression rare, not a significant problem

1 = Mild aggressiveness only (threatening without physical contact; or behavior that might represent just unduly rough play or momentary provoked lashing out)

2 = Definite physical aggression involving hitting or biting but no use of implements

3 = Violence that involves the use of implements

9 = N/K or not asked

105. EAGGOTH -Aggression to non-caregivers or non-family members (ever)

0 = Aggression rare, not a significant problem

1 = Mild aggressiveness only (threatening without physical contact; or behavior that might represent just unduly rough play or momentary provoked lashing out)

2 = Definite physical aggression involving hitting or biting but no use of implements

3 = Violence that involves the use of implements

9 = N/K or not asked

106. CSLFINJ -Self injury (current)

0 = No self-injury

1 = Slight self-injury (e.g., occasionally bites own hand/arm when annoyed, pulls hair, or slaps face); no substantial tissue damage

2 = Self-injury definitely present (e.g., actual bruises or calluses, repeated head banging, hair pulling, biting associated with definite tissue damage; do not count picking of spots)

3 = Definite self-injury with serious damage (e.g., skull fracture, eye injury, etc.)

9 = N/K or not asked

107. ESLFINJ -Self injury (ever)

0 = No self-injury

1 = Slight self-injury (e.g., occasionally bites own hand/arm when annoyed, pulls hair, or slaps face); no substantial tissue damage

2 = Self-injury definitely present (e.g., actual bruises or calluses, repeated head banging, hair pulling, biting associated with definite tissue damage; do not count picking of spots)

3 = Definite self-injury with serious damage (e.g., skull fracture, eye injury, etc.)

9 = N/K or not asked

108. CHVENT -Hyperventilation (current)

0 = No hyperventilating

1 = Occasional hyperventilating

2 = Frequent hyperventilating

9 = N/K or not asked

109. EHVENT -Hyperventilation (ever)

0 = No hyperventilating

1 = Occasional hyperventilating

2 = Frequent hyperventilating

9 = N/K or not asked

110. CFAINT -Faints/fits/blackouts (current)

0 = No attacks

1 = History of attacks that might be epileptic, but diagnosis not established

2 = Definite diagnosis of epilepsy

7 = Febrile convulsions only, with no continuing daily medication outside the period of fever

9 = N/K or not asked

111. EFAINT -Faints/fits/blackouts (ever)

0 = No attacks

1 = History of attacks that might be epileptic, but diagnosis not established

2 = Definite diagnosis of epilepsy

7 = Febrile convulsions only, with no continuing daily medication outside the period of fever

9 = N/K or not asked

112. CVISSPZ -Visuospatial ability, i.e. in puzzles, jigsaws, shapes, patterns, etc. (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or

participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

113. EVISSPZ -Visuospatial ability, i.e. in puzzles, jigsaws, shapes, patterns, etc.

(ever)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

114. CMEMZ -Memory skill, accurate memory for detail, as of dates or timetables (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

115. EMEMZ -Memory skill, accurate memory for detail, as of dates or timetables (ever)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

116. CMUSICZ -Musical ability, recognition, composition, absolute, pitch, or performance (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

117. EMUSICZ -Musical ability, recognition, composition, absolute, pitch, or performance (ever)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally

or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

118. CDRAWZ -Drawing skill, unusually skilled use of perspective or creative approach (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

119. EDRAWZ -Drawing skill, unusually skilled use of perspective or creative approach (ever)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

120. CREADZ -Reading ability e.g. early sight reading (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

121. EREADZ -Reading ability e.g. early sight reading (ever)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

122. CCOMPUZ -Computational ability e.g. mental arithmetic (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

123. ECOMPUZ -Computational ability e.g. mental arithmetic (current)

0 = No outstanding skills/knowledge in relation to overall level of ability, whether high or low

1 = Isolated skill/knowledge that is definitely out of keeping with subject's general level of ability, but not above general population norms

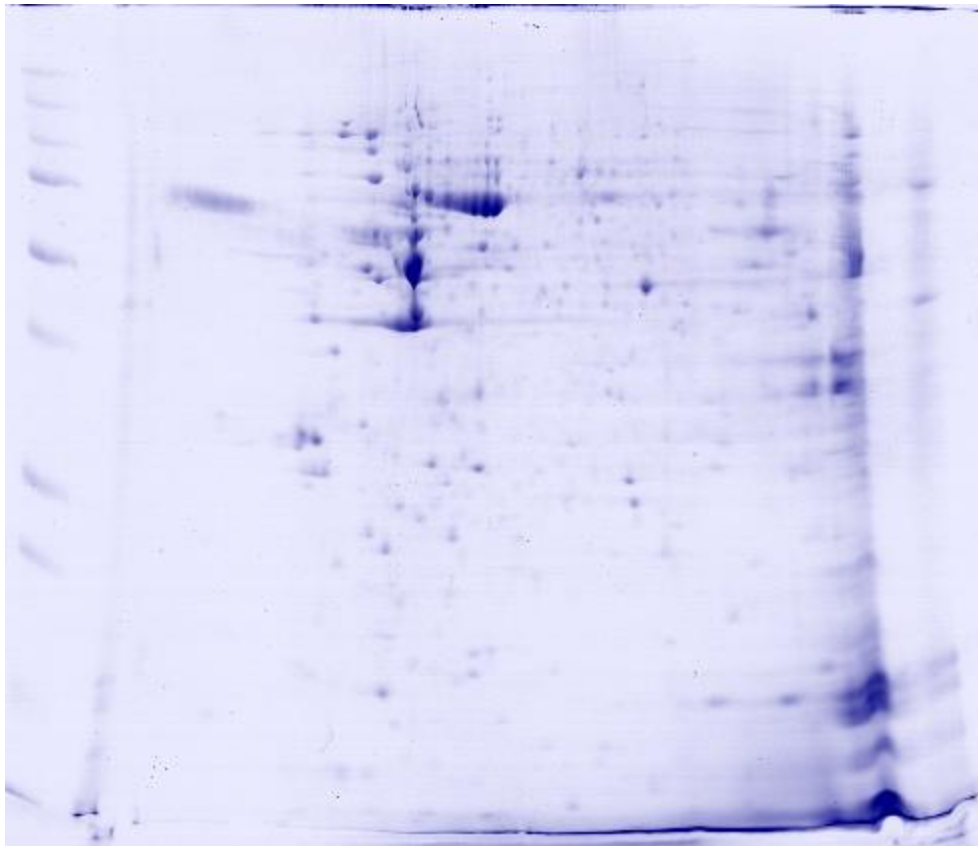
2 = Isolated skill/knowledge that is definitely above the subject's general level of ability and above the general population's normal level, but is not used functionally or meaningfully to any marked extent (e.g., a preschool child who can read without comprehension or a calendrical calculator would be coded here)

7 = Isolated skill/knowledge that is above the subject's general level and above the normal population's level of ability and is used meaningfully (i.e., genuine talent or ability used adaptively, such as performing music for others' enjoyment or participating in age-appropriate children's hobbies such as model-building or computer programming); subject is recognized by peers as having exceptional skill

8 = N/A (e.g., reading in a nonverbal subject)

9 = N/K or not asked

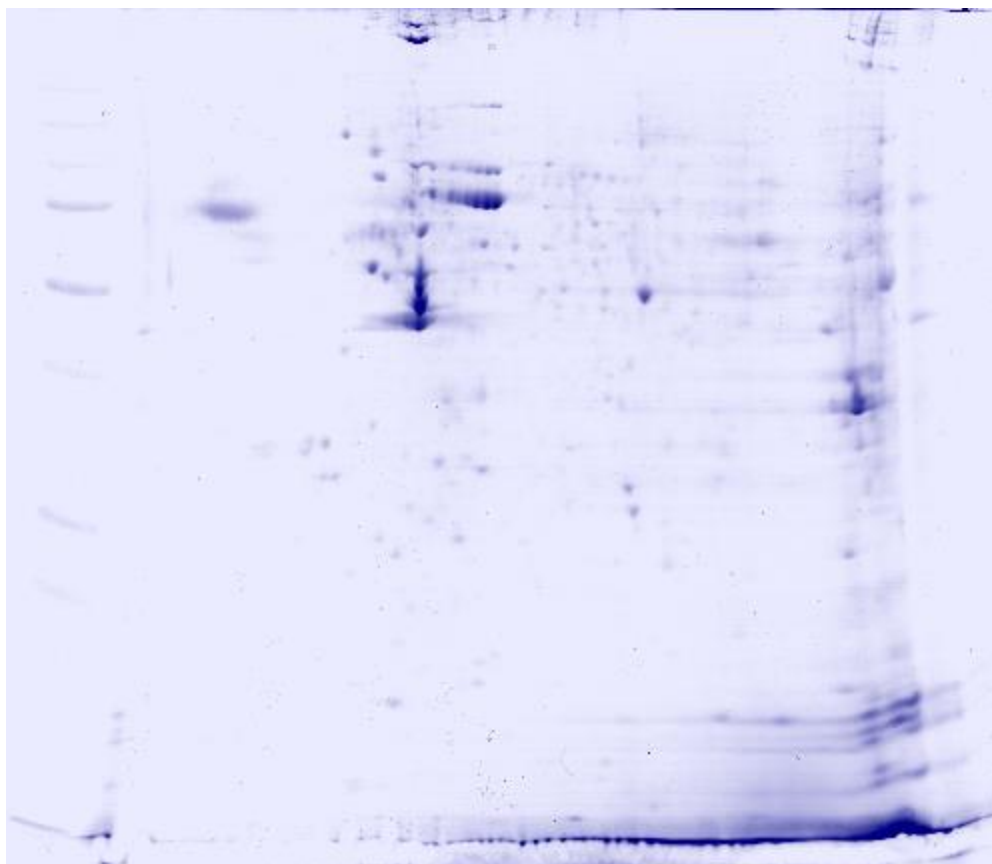
**ภาพโปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของผู้ป่วย
ออทิซึมสเปกตรัมและออทิซึมสเปกตรัมกลุ่มย่อยเทียบกับโปรตีนที่ได้จากกลุ่มควบคุม**



รูปที่ 38 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่มควบคุม

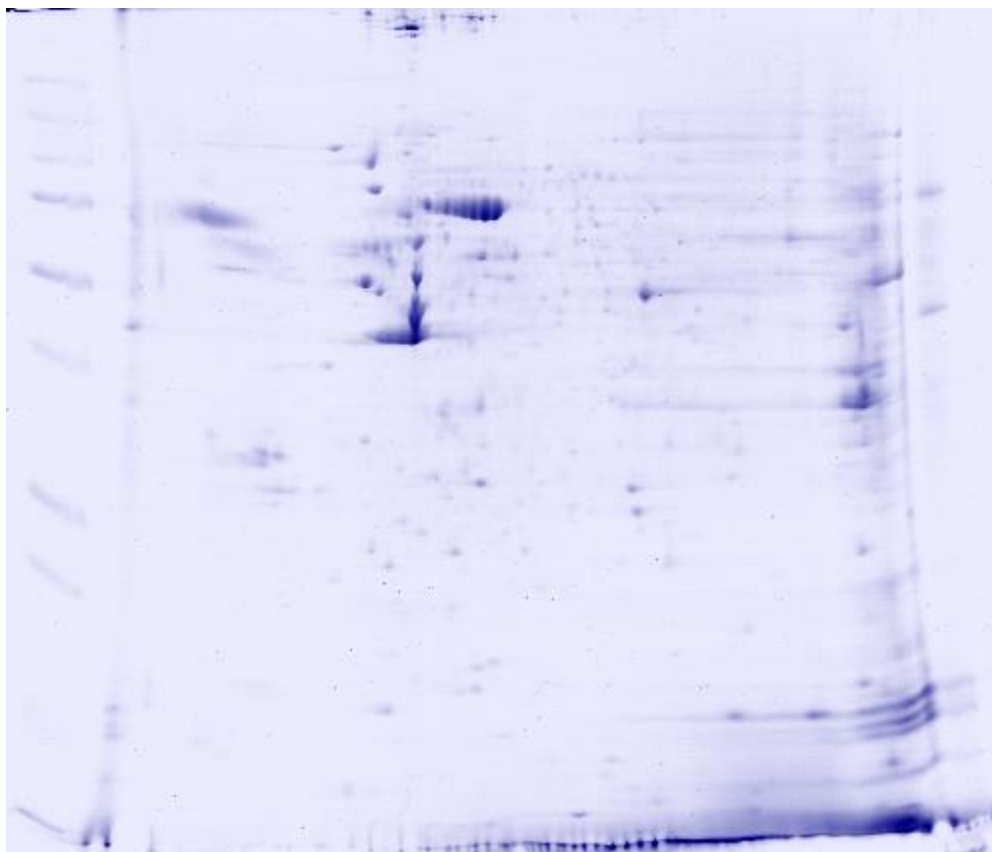
แสดงภาพโปรตีนบนแผ่นเจลที่ได้จากการวิเคราะห์ด้วยเทคนิค 2D-GE

จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในกลุ่มควบคุม (n=6)



รูปที่ 39 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์
ของกลุ่ม ASD (M+S subgroup)

แสดงภาพโปรตีนบนแผ่นเจลที่ได้จากการวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิด
ลิมโฟบลาสต์ในผู้ป่วยออทิซึมสเปกตรัมแบบรวมกลุ่มย่อย ASD M+S subgroup (n=22)

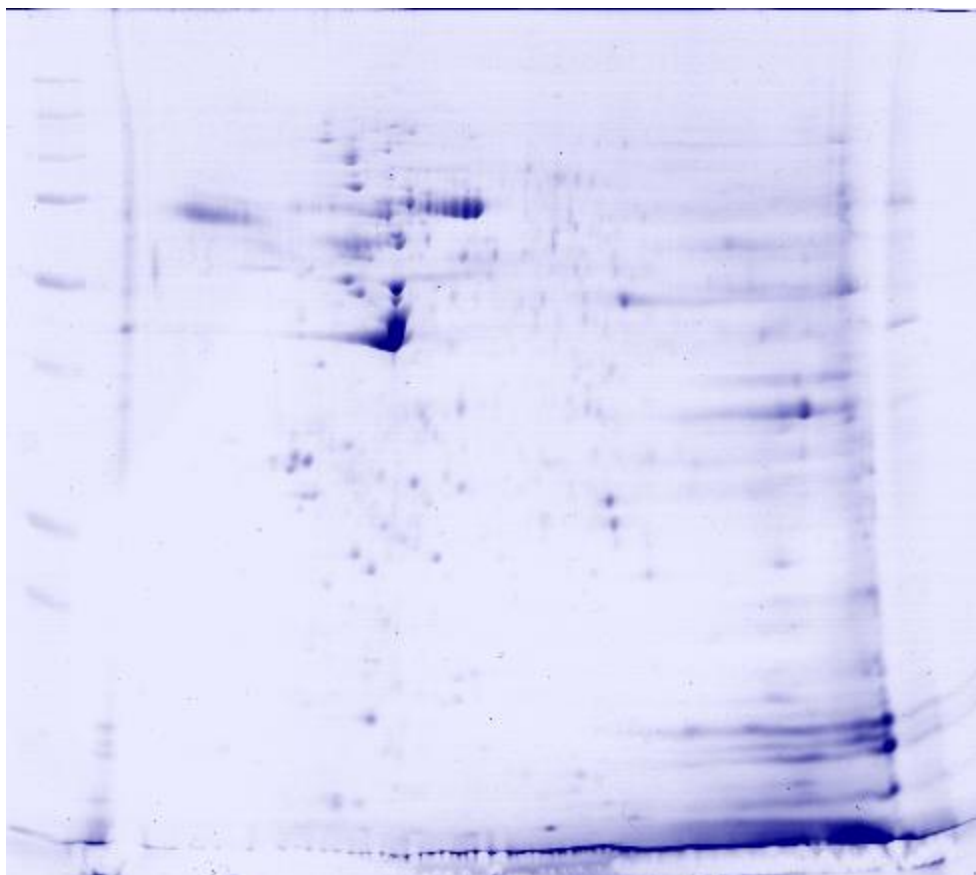


รูปที่ 40 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์
ของกลุ่ม M subgroup

แสดงภาพโปรตีนบนแผ่นเจลที่ได้จากการวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิด
ลิมโฟบลาสต์ในผู้ป่วยอหิวาต์ชิมสเปกตรัมกลุ่มย่อย M subgroup (n=9)

จุฬาลงกรณ์มหาวิทยาลัย

CHULALONGKORN UNIVERSITY



รูปที่ 41 โปรตีนที่ถูกวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ของกลุ่ม S subgroup แสดงภาพโปรตีนบนแผ่นเจลที่ได้จากการวิเคราะห์ด้วยเทคนิค 2D-GE จากเซลล์เพาะเลี้ยงชนิดลิมโฟบลาสต์ในผู้ป่วยออทิสซึมสเปกตรัมกลุ่มย่อย S subgroup (n=13)

**ตารางที่ 37 คะแนน ADI-R เฉลี่ยของผู้ป่วยออทิซึมสเปกตรัมกลุ่มย่อยที่ใช้ในการวิเคราะห์
สหสัมพันธ์เทียบกับระดับการแสดงออกของโปรตีน**

| ADI-R items | ADI-R scores (Mean ± SD) | | | ADI-R items | ADI-R scores (Mean ± SD) | | |
|-------------|--------------------------|-------------|-------------|-------------|--------------------------|-------------|-------------|
| | L subgroup | M subgroup | S subgroup | | L subgroup | M subgroup | S subgroup |
| LEVELL | 1.33 ± 0.52 | 0.00 ± 0.00 | 0.00 ± 0.00 | SOPLAY5 | 1.40 ± 1.14 | 1.50 ± 1.38 | 0.67 ± 0.52 |
| CCOMPSL | 1.67 ± 1.03 | 0.33 ± 0.58 | 0.17 ± 0.41 | CINTCH | 1.80 ± 1.30 | 2.00 ± 0.00 | 0.40 ± 0.55 |
| COMPSL5 | 2.20 ± 0.84 | 2.00 ± 0.00 | 1.00 ± 1.26 | INTCH5 | 1.80 ± 1.30 | 1.67 ± 1.37 | 1.00 ± 1.26 |
| CUSEBOD | 1.17 ± 1.17 | 0.00 ± 0.00 | 0.00 ± 0.00 | CRESPCH | 1.00 ± 0.71 | 3.00 ± 0.00 | 0.40 ± 0.55 |
| EUSEBOD | 1.67 ± 1.21 | 0.50 ± 0.84 | 0.83 ± 1.33 | RESPCH5 | 1.60 ± 1.14 | 1.17 ± 1.17 | 0.67 ± 0.82 |
| CARTIC | 1.00 ± 0.71 | 0.00 ± 0.00 | 0.17 ± 0.41 | CGRPLAY | 2.40 ± 0.55 | | 0.75 ± 0.50 |
| ARTICF5 | 2.00 ± 0.00 | 0.17 ± 0.41 | 0.40 ± 0.89 | GRPLAY5 | 2.60 ± 0.55 | 1.67 ± 1.37 | 1.20 ± 1.10 |
| CSTEREO | 0.25 ± 0.50 | 1.67 ± 1.03 | 1.67 ± 0.52 | CFRIEND | 2.67 ± 0.52 | 2.00 ± 1.26 | 1.50 ± 1.05 |
| ESTEREO | 0.25 ± 0.50 | 1.67 ± 1.03 | 1.67 ± 0.52 | FRIEND15 | 3.00 ± 0.00 | 1.83 ± 1.17 | 2.83 ± 0.41 |
| CCHAT | 1.67 ± 1.03 | 1.50 ± 0.55 | 1.00 ± 0.89 | CSOCDIS | 2.20 ± 0.45 | 0.60 ± 0.89 | 1.67 ± 0.52 |
| CHAT5 | 2.00 ± 0.00 | 1.83 ± 0.98 | 1.33 ± 1.03 | SOCDIS5 | 2.60 ± 0.55 | 1.20 ± 1.30 | 2.17 ± 0.75 |
| CCONVER | 3.00 ± 0.00 | 1.83 ± 0.98 | 1.83 ± 0.75 | CUNPROC | 0.50 ± 0.84 | 0.00 ± 0.00 | 0.33 ± 0.82 |
| CONVER5 | 3.00 ± 0.00 | 1.00 ± 0.89 | 1.33 ± 0.82 | EUNPROC | 0.83 ± 0.98 | 0.00 ± 0.00 | 0.33 ± 0.82 |
| CINAPPQ | 3.00 ± 0.00 | 0.40 ± 0.55 | 1.17 ± 0.98 | CCIRINT | 0.20 ± 0.45 | 0.00 ± 0.00 | 0.50 ± 0.84 |
| EINAPPQ | 3.00 ± 0.00 | 1.00 ± 0.89 | 1.33 ± 0.82 | ECIRINT | 0.60 ± 0.89 | 0.00 ± 0.00 | 1.00 ± 1.26 |
| CPRON | 3.00 ± 0.00 | 0.40 ± 0.89 | 0.67 ± 1.21 | CUSEOBJ | 1.00 ± 1.55 | 0.17 ± 0.41 | 0.67 ± 0.82 |
| EPRON | 3.00 ± 0.00 | 0.40 ± 0.89 | 1.17 ± 1.47 | EUSEOBJ | 1.67 ± 1.51 | 0.83 ± 1.33 | 1.67 ± 0.82 |
| CNEOID | 3.00 ± 0.00 | 0.00 ± 0.00 | 0.17 ± 0.41 | CCRIT | 0.50 ± 0.84 | 0.00 ± 0.00 | 0.83 ± 1.17 |
| ENEOID | 3.00 ± 0.00 | 0.20 ± 0.45 | 0.33 ± 0.52 | ECRIT | 0.67 ± 0.82 | 0.17 ± 0.41 | 1.50 ± 1.38 |
| CVERRIT | 3.00 ± 0.00 | 0.00 ± 0.00 | 0.17 ± 0.41 | CUNSENS | 0.67 ± 0.52 | 0.00 ± 0.00 | 0.83 ± 0.41 |
| EVERRIT | 3.00 ± 0.00 | 0.50 ± 1.22 | 0.83 ± 0.98 | EUNSENS | 0.83 ± 0.41 | 0.33 ± 0.52 | 1.17 ± 0.41 |
| CINR | 3.00 ± 0.00 | 0.83 ± 0.75 | 1.50 ± 0.55 | CNOISE | 1.33 ± 0.82 | 0.17 ± 0.41 | 0.67 ± 1.03 |
| EINR | 3.00 ± 0.00 | 1.00 ± 0.89 | 1.83 ± 0.41 | ENOISE | 1.33 ± 0.82 | 0.67 ± 1.03 | 0.83 ± 1.33 |
| CSPEECH | 3.00 ± 0.00 | 0.67 ± 0.52 | 0.33 ± 0.52 | CABINR | 0.67 ± 0.82 | 0.33 ± 0.82 | 0.67 ± 0.82 |
| SPEECH5 | 3.00 ± 0.00 | 1.33 ± 1.03 | 1.00 ± 1.26 | EABINR | 0.83 ± 0.75 | 0.33 ± 0.82 | 0.83 ± 1.17 |
| CPOINT | 1.00 ± 1.10 | 1.00 ± 1.10 | 0.83 ± 0.98 | CCHANGE | 1.00 ± 0.89 | 0.50 ± 0.84 | 0.67 ± 0.82 |
| POINT5 | 1.20 ± 1.10 | 1.40 ± 0.89 | 1.00 ± 0.89 | ECHANGE | 1.17 ± 0.98 | 0.83 ± 0.98 | 1.33 ± 0.82 |
| CNOD | 1.17 ± 0.98 | 0.50 ± 0.84 | 0.67 ± 1.03 | CREISIS | 0.50 ± 0.84 | 0.17 ± 0.41 | 0.17 ± 0.41 |
| NOD5 | 1.60 ± 0.89 | 1.40 ± 0.89 | 0.67 ± 1.03 | ERESIS | 0.67 ± 0.82 | 0.17 ± 0.41 | 0.17 ± 0.41 |
| CHSHAKE | 1.50 ± 0.84 | 0.67 ± 1.03 | 0.67 ± 1.03 | CUATT | 1.17 ± 2.86 | 0.17 ± 0.41 | 0.17 ± 0.41 |
| HSHAKE5 | 1.60 ± 0.89 | 1.33 ± 1.03 | 0.67 ± 1.03 | EUATT | 2.33 ± 3.61 | 0.17 ± 0.41 | 0.33 ± 0.52 |
| CINSGES | 1.50 ± 1.05 | 1.00 ± 1.26 | 0.67 ± 0.82 | CHFMAN | 0.83 ± 1.17 | 0.50 ± 0.84 | 0.17 ± 0.41 |
| INSGES5 | 2.00 ± 1.41 | 1.83 ± 1.17 | 0.67 ± 0.82 | EHFMAN | 0.83 ± 1.17 | 1.00 ± 1.10 | 0.67 ± 0.82 |
| AVOICE5 | 2.00 ± 0.71 | 1.50 ± 1.05 | 1.33 ± 0.82 | COTHMAN | 1.33 ± 1.03 | 0.00 ± 0.00 | 0.33 ± 0.82 |
| CIMIT | 2.40 ± 0.89 | | 1.60 ± 1.34 | EOTHMAN | 1.67 ± 1.03 | 0.00 ± 0.00 | 1.17 ± 1.47 |
| IMIT5 | 2.80 ± 0.45 | 2.17 ± 1.17 | 1.83 ± 1.33 | CMLHAND | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| CPLAY | 1.80 ± 1.30 | | 1.00 ± 1.00 | EMLHAND | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.17 ± 0.41 |
| PLAY5 | 2.60 ± 0.89 | 1.50 ± 1.38 | 1.33 ± 1.21 | CGAIT | 0.83 ± 1.33 | 0.17 ± 0.41 | 0.33 ± 0.52 |
| CPEERPL | 2.00 ± 0.71 | | 0.80 ± 0.84 | GAIT5 | 0.80 ± 0.84 | 0.67 ± 1.03 | 1.00 ± 1.41 |
| PEERPL5 | 2.40 ± 0.89 | 1.67 ± 1.37 | 1.17 ± 1.17 | CAGGFAM | 0.83 ± 0.75 | 1.00 ± 0.82 | 0.00 ± 0.00 |
| GAZE5 | 1.80 ± 1.10 | 1.50 ± 1.22 | 1.00 ± 0.63 | EAGGFAM | 2.00 ± 1.10 | 1.50 ± 0.58 | 0.00 ± 0.00 |
| CSSMILE | 1.00 ± 0.89 | 1.00 ± 1.26 | 0.50 ± 0.84 | CAGGOTH | 1.00 ± 1.10 | 0.00 ± 0.00 | 0.00 ± 0.00 |

| | | | | | | | |
|---------|-------------|-------------|-------------|---------|-------------|-------------|-------------|
| SSMILE5 | 1.60 ± 0.89 | 1.60 ± 1.14 | 0.67 ± 0.82 | EAGGOTH | 1.67 ± 0.82 | 0.50 ± 1.00 | 0.00 ± 0.00 |
| CSHOW | 1.67 ± 1.37 | 0.60 ± 0.89 | 0.67 ± 1.21 | CSLFINJ | 0.33 ± 0.52 | 0.00 ± 0.00 | 1.00 ± 0.00 |
| SHOW5 | 2.60 ± 0.55 | 1.33 ± 1.21 | 0.83 ± 1.17 | ESLFINJ | 1.00 ± 1.10 | 0.17 ± 0.41 | 2.00 ± 1.41 |
| COSHARE | 2.33 ± 0.52 | 1.17 ± 1.33 | 0.67 ± 1.21 | CHVENT | 0.33 ± 0.52 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| OSHARE5 | 2.80 ± 0.45 | 1.17 ± 1.33 | 1.17 ± 1.17 | EHVENT | 0.50 ± 0.55 | 0.00 ± 0.00 | 0.00 ± 0.00 |
| CSHARE | 1.17 ± 0.98 | 0.80 ± 0.45 | 0.33 ± 0.52 | CFAINT | 0.00 ± 0.00 | 0.00 ± 0.00 | 0.40 ± 0.89 |
| SHARE5 | 1.60 ± 0.89 | 1.33 ± 0.82 | 0.83 ± 0.98 | EFAINT | 0.00 ± 0.00 | 0.40 ± 0.89 | 0.40 ± 0.89 |
| COCOMF | 0.83 ± 0.75 | 0.67 ± 1.21 | 0.50 ± 0.55 | CVISSPZ | 1.00 ± 1.55 | 0.50 ± 1.22 | 3.00 ± 0.00 |
| OCOMF5 | 1.80 ± 1.30 | 1.33 ± 1.21 | 0.67 ± 0.52 | EVISSPZ | 1.00 ± 1.55 | 0.50 ± 1.22 | 3.00 ± 0.00 |
| CQUALOV | 1.17 ± 1.17 | 0.60 ± 1.34 | 0.50 ± 0.84 | CMEMZ | 0.17 ± 0.41 | 1.00 ± 1.55 | 3.00 ± 0.00 |
| QUALOV5 | 2.00 ± 1.00 | 1.40 ± 1.34 | 0.83 ± 0.98 | EMEMZ | 0.17 ± 0.41 | 1.00 ± 1.55 | 3.00 ± 0.00 |
| CRFACEX | 0.67 ± 0.52 | 0.83 ± 0.41 | 0.17 ± 0.41 | CMUSICZ | 0.33 ± 0.82 | 0.50 ± 1.22 | 3.00 ± 0.00 |
| RFACEX5 | 1.00 ± 1.22 | 1.33 ± 0.52 | 0.33 ± 0.52 | EMUSICZ | 0.33 ± 0.82 | 0.50 ± 1.22 | 3.00 ± 0.00 |
| CINAPFE | 1.00 ± 0.00 | 1.00 ± 0.63 | 1.17 ± 0.75 | CDRAWZ | 0.17 ± 0.41 | 0.00 ± 0.00 | 3.00 ± 0.00 |
| EINAPFE | 1.50 ± 0.84 | 1.33 ± 0.52 | 1.33 ± 0.82 | EDRAWZ | 0.17 ± 0.41 | 0.00 ± 0.00 | 3.00 ± 0.00 |
| CQRESP | 1.67 ± 1.03 | 1.20 ± 1.30 | 0.50 ± 0.84 | CREADZ | 0.83 ± 1.33 | 1.00 ± 1.55 | 3.00 ± 0.00 |
| QRESP5 | 2.40 ± 0.89 | 0.83 ± 1.33 | 0.83 ± 1.17 | EReadZ | 1.00 ± 1.55 | 1.00 ± 1.55 | 3.00 ± 0.00 |
| CINITIA | 1.33 ± 1.21 | 0.67 ± 1.03 | 0.83 ± 0.98 | CCOMPUZ | 0.00 ± 0.00 | 0.00 ± 0.00 | 3.00 ± 0.00 |
| INITIA5 | 1.80 ± 1.30 | 0.67 ± 0.82 | 0.83 ± 1.17 | ECOMPUZ | 0.00 ± 0.00 | 0.00 ± 0.00 | 3.00 ± 0.00 |
| CSOPLAY | 1.00 ± 0.71 | | 0.60 ± 0.55 | | | | |

ประวัติผู้เขียน

| | |
|-------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| ชื่อ-สกุล | นางสาว กัลยภัสตร์ หัสเสมอ |
| วัน เดือน ปี เกิด | 16 ตุลาคม 2538 |
| สถานที่เกิด | กรุงเทพมหานคร |
| วุฒิการศึกษา | วิทยาศาสตรบัณฑิต สาขาเทคนิคการแพทย์ จุฬาลงกรณ์มหาวิทยาลัย |
| ที่อยู่ปัจจุบัน | 28/12 ซ.นวมินทร์ 111 แยก 12-2 ถ.นวมินทร์ แขวงนวมินทร์ เขตบึงกุ่ม จ. กรุงเทพมหานคร 10230 |
| ผลงานตีพิมพ์ | <ol style="list-style-type: none"> Hussem, K., & Sarachana, T. (2019). Application of Artificial Intelligence (AI) Analysis to Identify Biomarker Candidates for Autism Spectrum Disorder (ASD). CASNIC, 7th. National and International Conference. Sukjamnong, S., Thongkorn, S., Kanlayaprasit, S., Saeliw, T., Hussem, K., Warayanon, W., Hu, V. W., Tencomnao, T., & Sarachana, T. (2020). Prenatal exposure to bisphenol A alters the transcriptome-interactome profiles of genes associated with Alzheimer's disease in the offspring hippocampus. Scientific reports, 10(1), 9487. https://doi.org/10.1038/s41598-020-65229-0 |
| รางวัลที่ได้รับ | Best Poster Presentation Award The 12th Thammasat-Chulalongkorn University Medical Technology Student Research Conference (2018). |