

ข้อเสนอโครงการวิศวกรรมไฟฟ้า วิชา 2102490 ปีการศึกษา 2562
การเปรียบเทียบวิธีการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมอง
A comparison of EEG artifact removal methods

นายธนภุต แสนเจริญกิจ ID 5930218721
อาจารย์ที่ปรึกษา ผศ.ดร. จิตโกมุท ส่งศิริ
ภาควิชาวิศวกรรมไฟฟ้า คณะวิศวกรรมศาสตร์
จุฬาลงกรณ์มหาวิทยาลัย

สารบัญ

1	บทนำ	2
2	วัตถุประสงค์ของโครงการ	4
3	หลักการและทฤษฎีที่เกี่ยวข้อง	4
3.1	การตรวจจับบริเวณที่มีสัญญาณรบกวน	5
3.1.1	การแยกหาคุณลักษณะเฉพาะ	6
3.1.2	การจัดการข้อมูลก่อนการประมวลผล	6
3.1.3	แบบจำลองสำหรับตรวจจับบริเวณที่มีสัญญาณรบกวน	7
3.2	การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้	9
3.3	การประเมินประสิทธิภาพของวิธีการกำจัดสัญญาณรบกวน	10
4	ผลลัพธ์จากการดำเนินการ	11
5	รายละเอียดของหัวข้อโครงการ	13
5.1	ขอบเขตของโครงการ	13
5.2	แผนการดำเนินการ	14
5.3	อุปสรรคและปัญหาที่พบ และการแก้ไข	15
6	เอกสารอ้างอิง	15

1 บทนำ

Electroencephalography (EEG) เป็นเทคนิคการบันทึกสัญญาณคลื่นไฟฟ้าสมองที่มีการนำไปประยุกต์ใช้อย่างแพร่หลายโดยเฉพาะการศึกษาทางคลินิก การทดลองในห้องทดลอง การดูแลผู้ป่วย การวินิจฉัยโรค และ brain-computer interfaces (BCI) ข้อได้เปรียบของเทคนิคนี้คือมีค่าใช้จ่ายที่ไม่สูงเกินไป อย่างไรก็ตาม ปัญหาที่พบในการบันทึกสัญญาณคือการรบกวนอย่างมีนัยสำคัญจากแหล่งสัญญาณอื่น ๆ ที่ไม่ได้มาจากสมอง เช่น สัญญาณรบกวนที่เกิดจากการเคลื่อนไหวของตา สัญญาณรบกวนที่เกิดจากการเคลื่อนไหวของกล้ามเนื้อ สัญญาณรบกวนจากคลื่นไฟฟ้าหัวใจ สัญญาณรบกวนจากสิ่งแวดล้อมภายนอก ทำให้การวินิจฉัยโรคมีความผิดพลาดได้ ใน BCI สัญญาณรบกวนอาจส่งผลให้เกิดการตีความที่ผิดพลาดทำให้เกิดการควบคุมอุปกรณ์ที่ไม่เป็นไปตามความต้องการ [2] ดังนั้นการตรวจจับและกำจัดสัญญาณรบกวนจึงเป็นขั้นตอนที่สำคัญในการเตรียมข้อมูลเพื่อนำไปประยุกต์ใช้ต่อ

ในงานวิจัยก่อนหน้า วิธีการหลายวิธีถูกพัฒนาขึ้นเพื่อกำจัดสัญญาณรบกวน วิธีหนึ่งที่เป็นที่นิยมในการกำจัดสัญญาณรบกวนที่เกิดจากตา คือวิธีการถดถอย (Regression methods) วิธีการนี้เป็นการประมาณสัญญาณรบกวนจากสัญญาณ Electrooculography (EOG) จากนั้นสัญญาณที่ถูกประมาณจะถูกนำไปลบออกจากสัญญาณคลื่นไฟฟ้าสมองเพื่อกำจัดสัญญาณรบกวน ปัญหาของวิธีการนี้คือ วิธีการนี้อาจจะกำจัดสัญญาณสมองบางส่วนออกไปด้วย เนื่องจากในสัญญาณ EOG จะมีสัญญาณสมองรวมอยู่ปริมาณหนึ่ง [15] วิธีการที่สองคือ วิธีการกำจัดสัญญาณรบกวนด้วยตัวกรอง ด้วยวิธีการนี้ ตัวกรองจะกำจัดสัญญาณรบกวนได้อย่างมีประสิทธิภาพก็ต่อเมื่อย่านความถี่ของสัญญาณรบกวนและย่านความถี่ของสัญญาณสมองไม่ซ้อนทับกัน อย่างไรก็ตาม ความถี่ของสัญญาณรบกวนมีโอกาสซ้อนทับกับย่านความถี่ของสัญญาณสมอง ตัวกรองแบบปรับตัวได้ถูกพัฒนาสำหรับการกำจัดสัญญาณรบกวนที่มีความถี่ซ้อนทับกับย่านความถี่ของสัญญาณสมอง ซึ่งจะมีตัวถ่วงน้ำหนักที่ถูกปรับค่าด้วยค่าความผิดพลาดระหว่างสัญญาณรบกวนที่ถูกประมาณมาจากสัญญาณอ้างอิง กับสัญญาณคลื่นไฟฟ้าสมอง ตัวถ่วงน้ำหนักนี้จะมีหน้าที่เป็นตัวกรองสัญญาณอ้างอิง สัญญาณอ้างอิงเมื่อผ่านตัวกรองนี้ก็จะได้สัญญาณที่ประมาณเป็นสัญญาณรบกวน วิธีการนี้จะทำวนซ้ำการปรับค่าตัวถ่วงน้ำหนักเพื่อลดค่าความผิดพลาดระหว่างสัญญาณรบกวนที่ถูกประมาณมาจากสัญญาณอ้างอิงกับสัญญาณคลื่นไฟฟ้าสมอง [4] อย่างไรก็ตาม ตัวกรองนี้ยังต้องการสัญญาณ EOG เพื่อใช้ในการประมาณสัญญาณรบกวน วิธีการต่อมาคือ วิธีการวิเคราะห์องค์ประกอบอิสระ (Independent component analysis, ICA) [10] ที่ใช้กันอย่างแพร่หลายสำหรับการแยกองค์ประกอบของสัญญาณออกเป็นองค์ประกอบที่เป็นอิสระต่อกัน วิธีหนึ่งที่ใช้หาเมทริกซ์ของการแปลงสำหรับใช้ในการแยกองค์ประกอบ คือ InfoMax [6] InfoMax จะมีกระบวนการหาเมทริกซ์ของการแปลง เมื่อนำเมทริกซ์ของการแปลงนี้ไปทำการแปลงเชิงเส้นกับเมทริกซ์ของสัญญาณคลื่นไฟฟ้าสมองหลายช่องสัญญาณ ผลลัพธ์ที่ได้จะเป็นการประมาณสัญญาณจากแหล่งสัญญาณต่าง ๆ ในสมอง โดย InfoMax จะพยายามลดสารสนเทศร่วม (Mutual Information) ของสัญญาณที่ได้จากผลลัพธ์การแปลงเชิงเส้นนั้น [6] ผู้ใช้งานสามารถนำสัญญาณที่ถูกแยกเป็นองค์ประกอบต่าง ๆ ไปใช้สำหรับการพิจารณาว่าองค์ประกอบนั้นเป็นสัญญาณรบกวนหรือไม่ แต่วิธีการนี้จำเป็นจะต้องมีผู้เชี่ยวชาญเป็นผู้ระบุว่าจะองค์ประกอบไหนเป็นองค์ประกอบของสัญญาณรบกวน เพื่อจะกำจัดองค์ประกอบนั้นทิ้งออกไป สมมติฐานของวิธีการนี้คือ จำนวนขั้วอิเล็กโทรดจะต้องมีจำนวนมากกว่าหรือเท่ากับจำนวนแหล่งกำเนิดที่เป็นอิสระกันของสัญญาณคลื่นไฟฟ้าสมอง [2] เป็นเหตุให้วิธีการนี้ไม่เหมาะสมกับการนำไปประยุกต์ใช้ใน BCI และอุปกรณ์ดูแลสุขภาพที่ติดอยู่กับร่างกาย เนื่องจากแหล่งกำเนิดสัญญาณคลื่นไฟฟ้าสมองมีแหล่งกำเนิดหลายแหล่งจากในสมอง แต่อุปกรณ์เหล่านี้ส่วนใหญ่แล้วจะมีขั้วอิเล็กโทรดเพียงขั้วเดียวในอุปกรณ์หนึ่งชิ้น [8]

วิธีการที่ได้กล่าวไป เราจะใช้วิธีการเหล่านั้นก็ต่อเมื่อสัญญาณคลื่นไฟฟ้าสมองมีหลายช่องสัญญาณ วิธีการต่อไปนี้ ผู้ใช้สามารถนำไปใช้ได้กับสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ วิธีการหนึ่งที่ใช้คือ Empirical mode decomposition (EMD) เป้าหมายของวิธีการนี้คือ วิธีการนี้ต้องการแยกสัญญาณออกเป็นผลรวมของฟังก์ชันพื้นฐาน ที่เรียกว่า intrinsic mode functions (IMFs) โดยสัญญาณรบกวนและสัญญาณสมองสามารถถูกอธิบายได้ด้วยฟังก์ชันพื้นฐานแต่ละฟังก์ชัน [15] แต่วิธีการนี้จำเป็นจะต้องมีผู้เชี่ยวชาญเป็นผู้ระบุว่าจะองค์ประกอบไหนเป็นองค์ประกอบของสัญญาณรบกวน เพื่อจะกำจัดองค์ประกอบนั้นทิ้งออกไป นักวิจัยพัฒนาวิธีการนี้มาสำหรับ non-stationary processes ด้วยเหตุผลที่วิธีการนี้จึงเหมาะกับการนำไปใช้วิเคราะห์สัญญาณคลื่นไฟฟ้าสมอง [2] อย่างไรก็ตาม วิธีนี้มีข้อเสียในขั้นตอนของการแยกสัญญาณออกเป็นผลรวมของฟังก์ชันพื้นฐานซึ่งจะทำผ่าน Sifting process กระบวนการนี้เป็นกระบวนการที่ทำให้ EMD มีความซับซ้อนในการประมวลผลข้อมูลที่สูง [2] เราพบว่ายังมีวิธีการกำจัดสัญญาณรบกวนด้วยตัวกรองที่ไม่ต้องการสัญญาณอ้างอิงอาทิ วิธีการกำจัดสัญญาณรบกวนด้วยตัวกรอง Savitzky-Golay (Savitzky-Golay filter) [13] วิธีการนี้จะต้องระบุบริเวณที่มีสัญญาณรบกวนก่อน จากนั้นจึงทำการกำจัดสัญญาณรบกวนที่เกิดจากตาในสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณในบริเวณที่พบว่ามีสัญญาณรบกวนจะถูกประมาณด้วยวิธีการถดถอยพหุนาม สัญญาณที่ได้จากการประมาณจะถูกนำไปลบออกจากบริเวณที่พบว่ามีสัญญาณรบกวน วิธีการที่ใช้สำหรับแยกองค์ประกอบของสัญญาณเพื่อนำองค์ประกอบนั้นไประบุว่าเป็นสัญญาณรบกวนหรือไม่อีกวิธีการหนึ่งคือ วิธีการแปลงเวฟเลต (Wavelet transform, WT) วิธีการนี้เป็นวิธีการที่มีประโยชน์มากสำหรับการนำไปประยุกต์ใช้ในการวิเคราะห์สัญญาณคลื่นไฟฟ้าสมอง วิธีการนี้จะแยกสัญญาณออกเป็นหลายอำนาจจำแนก (Multiresolution) โดยแต่ละอำนาจจำแนกจะนำเสนอคุณลักษณะทางกายภาพที่แตกต่างกัน คุณลักษณะทางกายภาพที่ไม่พบในอำนาจจำแนกหนึ่ง อาจพบในอีกอำนาจจำแนกหนึ่ง ซึ่งมีประโยชน์สำหรับการใช้แยกหาคุณลักษณะเฉพาะของสัญญาณอย่างมาก [15] วิธีการนี้แตกต่างกับวิธีวิเคราะห์องค์ประกอบอิสระตรงที่ วิธีการนี้สามารถใช้กับสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณได้ ทำให้มีประโยชน์ต่อการนำไปประยุกต์ใช้ใน BCI และอุปกรณ์ดูแลสุขภาพที่ติดอยู่กับร่างกาย [8] และวิธีการนี้สามารถนำไปประยุกต์ใช้กับ real-time processing ได้ [14] ด้วยเหตุผลเหล่านี้วิธีการนี้จึงถูกนำไปประยุกต์ใช้เพื่อหาคุณลักษณะเฉพาะของสัญญาณสมอง ซึ่งจะเห็นได้ใน [1] ในตัวอย่างนี้สัญญาณคลื่นไฟฟ้าสมองในแต่ละช่วงเวลาจะถูกนำมาแปลงเวฟเลต เพื่อหาคุณลักษณะทางกายภาพในแต่ละอำนาจจำแนก จากนั้นคุณลักษณะทางกายภาพจะถูกนำไปหากรอบของสัญญาณด้วยวิธีการกล้ำสัญญาณเชิงขนาด (Amplitude modulation) บนพื้นฐานของ Hilbert transform ต่อมานักวิจัยจะนำกรอบ

ของสัญญาณไปใช้เป็นคุณลักษณะเฉพาะเพื่อใช้ในการเรียนรู้ของเครื่องต่อไป

วิธีการส่วนใหญ่ที่กล่าวมาข้างต้นจำเป็นต้องมีขั้นตอนหลัก คือ จะต้องมีการเตรียมข้อมูลซึ่งทำหน้าที่ระบุบริเวณหรือองค์ประกอบที่มีสัญญาณรบกวน งานวิจัยอีกกลุ่มหนึ่งจึงสนใจการตรวจจับบริเวณที่มีสัญญาณรบกวนด้วยวิธีการอัตโนมัติ อย่างไรก็ตาม การระบุบริเวณที่มีสัญญาณรบกวนอัตโนมัติก็ยังคงต้องการค่า Threshold สำหรับการจำแนกบริเวณที่เกิดสัญญาณรบกวน เพื่อที่จะกำจัดสัญญาณรบกวนในบริเวณนั้นต่อไป ทำให้วิธีการเรียนรู้ของเครื่องเป็นวิธีการที่ให้ความสนใจในโครงการนี้ วิธีการเรียนรู้ของเครื่องที่ถูกใช้ในการระบุบริเวณที่มีสัญญาณรบกวนวิธีการหนึ่งคือ Support vector machine (SVM) วิธีการนี้ถูกนำไปใช้ตรวจจับสัญญาณรบกวน ดังตัวอย่างที่ให้เห็นใน [1] การแยกคุณลักษณะเฉพาะจะถูกแยกด้วยวิธีการแปลงเวฟเลตซึ่งได้อธิบายไปแล้วข้างต้น คุณลักษณะเฉพาะจะถูกนำไปใช้ในการเรียนรู้ของเครื่อง เพื่อการตรวจจับการขยับนิ้วในสัญญาณคลื่นไฟฟ้าสมอง วิธีการที่ถูกเลือกใช้กันน้อยคือ Random forest อย่างไรก็ตาม วิธีการนี้มีความสามารถในการเลือกคุณลักษณะเฉพาะซึ่งเป็นประโยชน์มากเพราะการเลือกคุณลักษณะเฉพาะเป็นเรื่องที่ยากสำหรับคุณลักษณะเฉพาะของสัญญาณรบกวนและสมองที่มีความซับซ้อน งานวิจัยที่เลือกใช้วิธีการนี้คือ [11] ในงานวิจัยนี้วิธีการ Random forest จะถูกใช้ในการตรวจจับสัญญาณรบกวนโดยสัญญาณจะถูกตัดเป็นช่วง ๆ จากนั้นคุณลักษณะเฉพาะจากสัญญาณช่วงนั้นจะถูกคำนวณหาค่าต่าง ๆ อันได้แก่ ค่าขนาดสัญญาณมากที่สุด ค่าพลังงานในช่วงความถี่ 0.5-50 เฮิรตซ์ วิธีการต่อมาคือ Convolutional neural network (CNN) เป็นอีกวิธีการหนึ่งที่ได้ได้รับความสนใจเนื่องจากวิธีนี้จะมีการจัดชั้นที่ใช้สำหรับแยกคุณลักษณะเฉพาะ ซึ่งเป็นข้อดีที่จะนำไปใช้กับสัญญาณคลื่นไฟฟ้าสมองที่มีคุณลักษณะเฉพาะที่ซับซ้อน การแยกคุณลักษณะเฉพาะนี้จะอยู่ในระดับชั้นของ Convolutional layer วิธีการนี้ถูกใช้ใน [9] ในการทดลองนี้ CNN จะถูกใช้ในการตรวจจับสัญญาณรบกวนที่เกิดใน intracerebral EEG โดยวิธีการนี้จะสนใจสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ สัญญาณจะถูกแบ่งเป็นช่วง ๆ จากนั้นสัญญาณแต่ละช่วงจะถูกนำไปผ่านตัวกรองแถบความถี่ผ่านในช่วงความถี่ต่าง ๆ ต่อมาสัญญาณที่ผ่านตัวกรองแต่ละช่วงความถี่จะถูกนำไปหากรอบของสัญญาณ เพื่อนำกรอบของสัญญาณนี้ไปเป็นข้อมูลขาเข้าของ CNN เพื่อการตรวจจับช่วงที่มีสัญญาณรบกวน

จากวิธีที่ศึกษามาทั้งหมด โครงการนี้จะให้ความสนใจการกำจัดสัญญาณรบกวนที่เกิดจากตา เนื่องจากอุปกรณ์บันทึกสัญญาณคลื่นไฟฟ้าสมองที่มีอยู่ ตำแหน่งของอิเล็กโทรดจะอยู่ใกล้กับดวงตา ส่งผลให้สัญญาณรบกวนที่เกิดจากตามีอิทธิพลมากกว่าสัญญาณรบกวนชนิดอื่น ข้อมูลสัญญาณคลื่นไฟฟ้าสมองจะเป็นข้อมูลที่ได้จากการบันทึกสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ ในโครงการนี้ข้อมูลจะถูกส่งเคราะห์ขึ้น วิธีที่จะศึกษาจะต้องสามารถตรวจจับสัญญาณรบกวนได้อัตโนมัติ ขั้นตอนจะถูกแบ่งออกเป็นสองส่วน 1) ขั้นตอนของการตรวจจับบริเวณที่มีสัญญาณรบกวน 2) ขั้นตอนของการกำจัดสัญญาณรบกวนในบริเวณที่ถูกตรวจจับ วิธีการดังต่อไปนี้จะถูกใช้สำหรับการตรวจจับสัญญาณรบกวน

1. Thresholding เลือกวิธีการนี้สำหรับเป็นวิธีพื้นฐานเพื่อนำมาใช้สำหรับเปรียบเทียบประสิทธิภาพของวิธีการ Random forest และ CNN โดยวิธีการ Thresholding จะใช้วิธีการตรวจจับสัญญาณรบกวนด้วยการคำนวณพลังงานในย่านความถี่ 0.5-5 เฮิรตซ์ของสัญญาณในแต่ละช่วงเวลา ถ้าบริเวณไหนมีพลังงานที่สูงกว่าค่าหนึ่งบริเวณนั้นจะเป็นบริเวณสัญญาณรบกวน
2. Random Forest เลือกวิธีการนี้เพราะวิธีการนี้มีความสามารถในการเลือกคุณลักษณะเฉพาะ วิธีการนี้ถูกใช้ใน [11] และถูกนำไปเปรียบเทียบประสิทธิภาพกับวิธี Fast [16] ผลลัพธ์ที่ได้วิธีการ Random forest ให้ประสิทธิภาพที่สูงกว่า และงานวิจัยเกี่ยวกับการตรวจจับสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมองยังไม่มีหลักฐานยืนยันว่าวิธีการใดดีที่สุด ในโครงการนี้จึงอยากทดลองวิธีการที่ยังมีผู้ทดลองน้อย
3. CNN เลือกวิธีการนี้เนื่องจากเป็นวิธีการที่มีความสามารถในการแยกคุณลักษณะเฉพาะของข้อมูลได้ และยังมีงานวิจัยจำนวนน้อยที่ใช้วิธีการนี้ในการตรวจจับสัญญาณรบกวนหนึ่งช่องสัญญาณ และยังมีมีการนำวิธีการนี้ไปใช้กับการตรวจจับสัญญาณรบกวนใน Intracerebral EEG [9] แล้วให้ผลลัพธ์ที่ดี ผู้ศึกษาจึงอยากนำวิธีการนี้มาประยุกต์ใช้กับการตรวจจับสัญญาณรบกวนใน EEG

ทั้งสามวิธีเป็นวิธีการตรวจจับบริเวณที่มีสัญญาณรบกวน ต่อมาเมื่อวิธีการต่าง ๆ สามารถตรวจจับบริเวณที่มีสัญญาณรบกวนได้แล้ว เราจะต้องทำการกำจัดสัญญาณรบกวนในบริเวณนั้น ตัวกรองสวิตชิ่ง-โกลเลย์จะถูกใช้สำหรับกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้ว่ามีสัญญาณรบกวน ดังที่มีใช้ใน [13] และข้อดีของตัวกรองนี้คือสามารถใช้ได้กับสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ

โครงการฉบับนี้ จะศึกษาวิธีการตรวจจับและกำจัดสัญญาณรบกวนที่เกิดจากตา ที่เกิดขึ้นในการบันทึกสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ โดยโครงการนี้จะทดลองบนข้อมูลสังเคราะห์ และโครงการนี้มีเนื้อหาที่ศึกษาครอบคลุมดังนี้ 1) การศึกษาคุณลักษณะเฉพาะของสัญญาณไฟฟ้าในสมองและสัญญาณรบกวน 2) วิธีการกำจัดสัญญาณรบกวน 3) วิธีการประเมินประสิทธิภาพของวิธีการกำจัดสัญญาณรบกวน จากนั้นจะทำการทดลองเปรียบเทียบวิธีการกำจัดสัญญาณรบกวนที่ได้ศึกษานั้นคือวิธีการ Thresholding, Random forest และ CNN ต่อมาจะวิเคราะห์ผลที่ได้จากการเปรียบเทียบผลลัพธ์เพื่อนำเสนอวิธีการที่เหมาะสมสำหรับใช้กำจัดสัญญาณรบกวนซึ่งเกิดขึ้นในการบันทึกสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ

2 วัตถุประสงค์ของโครงการ

1. เพื่อศึกษาวิธีการกำจัดสัญญาณรบกวนที่ประยุกต์ใช้ได้กับการบันทึกสัญญาณคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ
2. เพื่อเปรียบเทียบผลลัพธ์และความซับซ้อนในการคำนวณของวิธีการกำจัดสัญญาณรบกวนที่ต่างกัน ซึ่งมีวิธีการกำจัดสัญญาณรบกวนดังต่อไปนี้ Thresholding, Random forest และ Convolutional neural network

3 หลักการและทฤษฎีที่เกี่ยวข้อง

ข้อมูลที่จะนำมาใช้กับการทดลองกำจัดสัญญาณรบกวนทั้ง 3 วิธี จะเป็นข้อมูลที่ถูกละทิ้งขึ้น โดยเป็นข้อมูลที่ได้จาก [5] ซึ่งการสังเคราะห์ข้อมูลของงานวิจัยนี้จะสังเคราะห์มาจากการวัดคลื่นไฟฟ้าสมองจริง ซึ่งมีอัตราการสุ่มเท่ากับ 200 เฮิรตซ์ โดยจะวัดจากผู้ร่วมทดลองจำนวน 54 คน เวลาที่วัดของแต่ละคนทั้งหมด 30 วินาที และจะเป็นการวัดสัญญาณคลื่นไฟฟ้าสมองหลายช่องสัญญาณ จะแบ่งขั้นตอนการวัดเป็น 2 ขั้นตอน ดังนี้

1. ขั้นตอนการสังเคราะห์สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน ขั้นตอนนี้สัญญาณจะถูกบันทึกขณะปิดตา และควบคุมการบันทึกไม่ให้เกิดสัญญาณรบกวน
2. ขั้นตอนการบันทึกสัญญาณรบกวน ขั้นตอนนี้สัญญาณจะถูกบันทึกขณะเปิดตาและสัญญาณคลื่นไฟฟ้าสมองจะถูกวัดพร้อมกับ Vertical Electrooculogram (VEOG) และ Horizontal Electrooculogram (HEOG) เพื่อใช้เป็นสัญญาณอ้างอิงสำหรับคำนวณสัมประสิทธิ์ a และ b ด้วยวิธี linear regression โดยสัมประสิทธิ์นี้คำนวณเพื่อต้องการทราบว่า VEOG และ HEOG ส่งผลต่อสัญญาณคลื่นไฟฟ้าสมองเท่าไร

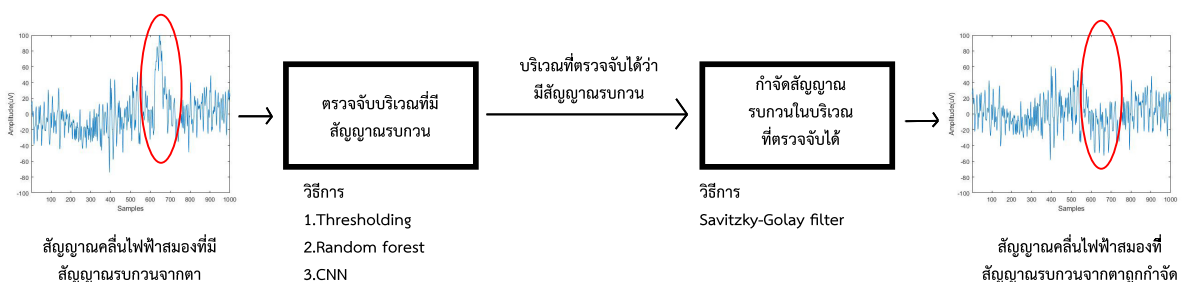
กำหนดให้ y คือ สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน, x คือ สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนซึ่งถูกวัดในขั้นตอนที่ 1, v คือ Vertical Electrooculogram, h คือ Horizontal Electrooculogram สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจะถูกสังเคราะห์ ดังสมการ

$$y = x + av + bh$$

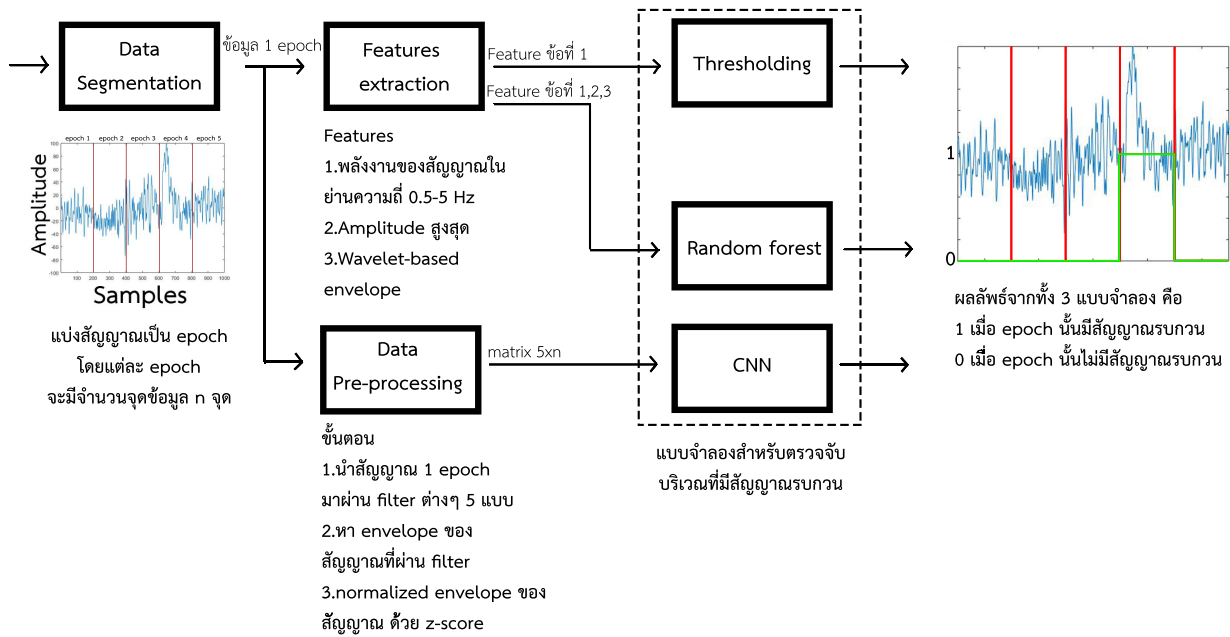
โครงการของเราสนใจการวัดคลื่นไฟฟ้าสมองเพียงหนึ่งช่องสัญญาณ สัญญาณคลื่นไฟฟ้าสมองช่องสัญญาณ FP1 จึงถูกนำมาใช้ในโครงการนี้เพียงช่องสัญญาณเดียว

ข้อมูลที่มียังมีจำนวนน้อย ในเทอมหน้าจะศึกษาวิธีสังเคราะห์สัญญาณรบกวนที่เกิดจากตาด้วยสมการคณิตศาสตร์เพื่อนำมาใช้เป็นข้อมูลเพิ่มเติมสำหรับแบบจำลองต่าง ๆ

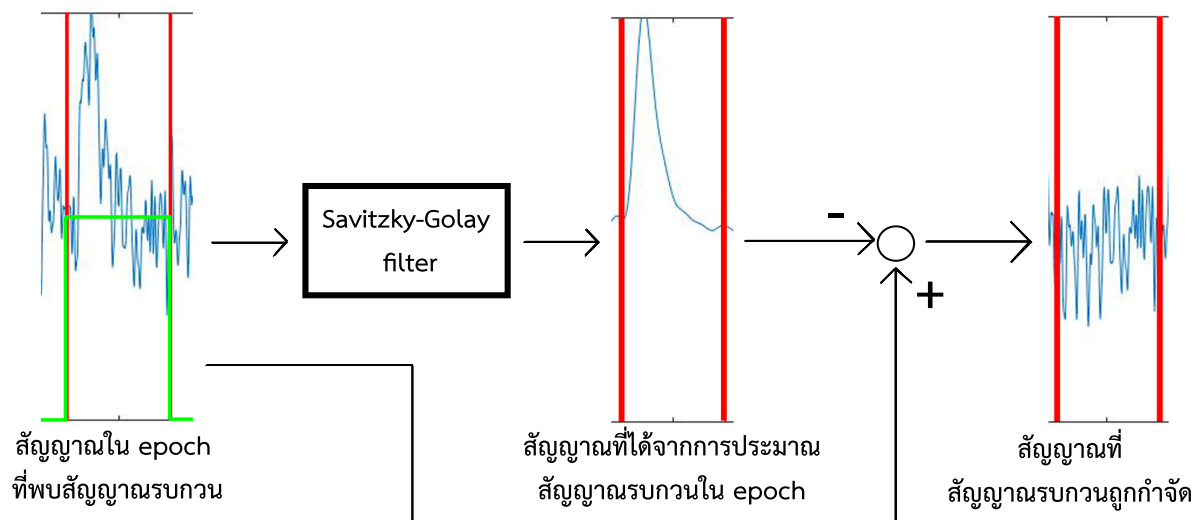
ข้อมูลจะมี 2 ประเภท 1) สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน 2) สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน ข้อมูลจะถูกแบ่งเป็น 80 เปอร์เซ็นต์สำหรับการฝึกของวิธีการต่างๆ และอีก 20 เปอร์เซ็นต์สำหรับการทดสอบวิธีการกำจัดสัญญาณรบกวนวิธีการต่างๆ ในโครงการฉบับนี้การทดลองจะถูกแบ่งเป็น 2 ส่วน ดังนี้ 1) การตรวจจับบริเวณที่มีสัญญาณรบกวน 2) การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้ ซึ่งแสดงดัง รูปที่ 1 ในหัวข้อสุดท้าย จะอธิบายวิธีการประเมินประสิทธิภาพของวิธีการกำจัดสัญญาณรบกวนทั้ง 3 วิธี



รูป 1: ขั้นตอนการกำจัดสัญญาณรบกวนในสัญญาณคลื่นไฟฟ้าสมอง ที่นำเสนอในโครงการนี้



รูป 2: ขั้นตอนของการตรวจจับบริเวณที่มีสัญญาณรบกวน



รูป 3: ขั้นตอนของการกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

3.1 การตรวจจับบริเวณที่มีสัญญาณรบกวน

การตรวจจับบริเวณที่มีสัญญาณรบกวน จะมีขั้นตอนต่าง ๆ เป็นไปดัง รูปที่ 2 สำหรับโครงงานนี้จะมีแบบจำลองสำหรับตรวจจับบริเวณที่มีสัญญาณรบกวน 3 แบบจำลอง คือ 1) Thresholding 2) Random forest และ 3) CNN แต่ในทุกแบบจำลองจะต้องมีขั้นตอนการแบ่งส่วนของข้อมูล (Data segmentation) ก่อนเสมอ นั่นคือ ข้อมูลจะถูกนำมาแบ่งออกเป็นท่อน (epoch) โดยแต่ละท่อนจะมีจำนวนจุดข้อมูล n จุด จากนั้นขั้นตอนในการเตรียมข้อมูลของแต่ละแบบจำลองจะต่างกัน สำหรับแบบจำลอง Thresholding และ Random forest จะมีขั้นตอนเหมือนกัน นั่นคือ ขั้นตอนการแยกหาคุณลักษณะเฉพาะ (Features extraction) ส่วนแบบจำลอง CNN จะเป็นขั้นตอนการจัดการข้อมูลก่อนการประมวลผล (Data pre-processing)

3.1.1 การแยกหาคุณลักษณะเฉพาะ

การแยกหาคุณลักษณะเฉพาะของข้อมูลจะเป็นการคำนวณบนข้อมูลแต่ละท่อนของสัญญาณ จะมีทั้งหมด 3 คุณลักษณะเฉพาะ

1. พลังงานในย่านความถี่ 0.5-5 เฮิร์ตซ์ สัญญาณในแต่ละท่อนจะถูกนำมาแปลงฟูรีเยไม่ต่อเนื่อง ดังสมการ

$$X(f) = \sum_{n=0}^{N-1} x[n]e^{-j2\pi n f}$$

โดยที่ $x[n]$ คือ สัญญาณที่จะนำมาแปลงฟูรีเยไม่ต่อเนื่อง, N คือ จำนวนจุดข้อมูลที่ทำกรแปลงฟูรีเยไม่ต่อเนื่อง จากนั้นจะนำผลลัพธ์ที่ได้ไปหาความหนาแน่นสเปกตรัมกำลัง ดังสมการ

$$S(f) = |X(f)|^2$$

ต่อมาจะนำความหนาแน่นสเปกตรัมกำลังไปหาพลังงานในย่านความถี่ 0.5-5 เฮิร์ตซ์ ดังสมการ

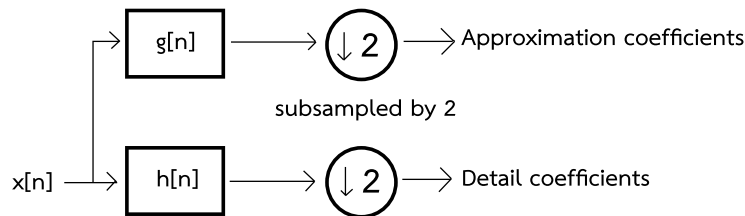
$$E = \sum_{f=0.5}^5 S(f)$$

ผลลัพธ์ที่ได้คือพลังงานในย่านความถี่ 0.5-5 เฮิร์ตซ์ซึ่งจะใช้เป็นคุณลักษณะเฉพาะในการเรียนรู้ของเครื่อง ที่เลือกใช้ย่านความถี่นี้เพราะว่าสัญญาณรบกวนจากตาจะเกิดในย่านความถี่นี้ [7]

2. ขนาดสัญญาณ (Amplitude) สูงที่สุด
3. Wavelet-based envelope [1] สัญญาณในแต่ละท่อนจะถูกนำมาแปลงเวฟเล็ตไม่ต่อเนื่องหนึ่งระดับชั้น ดัง รูปที่ 4 จากนั้น Approximation coefficients และ Detail coefficients ที่ได้จากการแปลงเวฟเล็ต จะถูกนำไปหากรอบของสัญญาณ (Envelope) ด้วยวิธีการกล้าสัญญาณเชิงขนาด ดังสมการ

$$AA(k) = \sqrt{cA(k)^2 + H(cA(k))^2} \text{ และ } AD(k) = \sqrt{cD(k)^2 + H(cD(k))^2}$$

โดยที่ $cA(k)$ คือ Approximation coefficient ที่ตำแหน่ง k , $cD(k)$ คือ Detail coefficient ที่ตำแหน่ง k , $AA(k)$ คือ กรอบของสัญญาณของ Approximation coefficient ที่ตำแหน่ง k , $AD(k)$ คือ กรอบของสัญญาณของ Detail coefficient ที่ตำแหน่ง k , $H(y)$ คือ การแปลงฮิลเบิร์ตของสัญญาณ y กรอบของสัญญาณที่ได้ทั้งสองจะถูกนำไปใช้เป็นตัวคุณลักษณะเฉพาะของแบบจำลอง Random forest



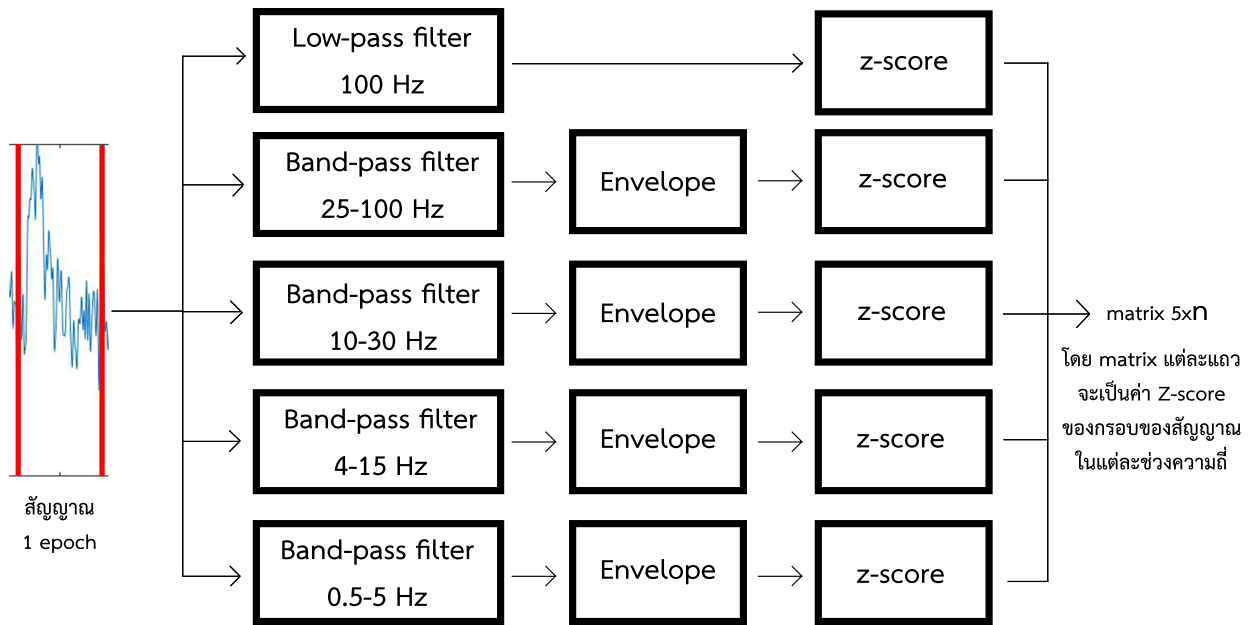
รูป 4: ขั้นตอนการแปลงเวฟเล็ตหนึ่งระดับชั้น โดยที่ $x[n]$ คือ สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน, $g[n]$ คือ ตัวกรองความถี่ต่ำผ่าน, $h[n]$ คือ ตัวกรองความถี่สูงผ่าน

3.1.2 การจัดการข้อมูลก่อนการประมวลผล

การจัดการข้อมูลก่อนการประมวลผลในหัวข้อนี้จะถูกประยุกต์มาจากงานวิจัยใน [9] ซึ่งจะมีขั้นตอนดัง รูปที่ 5 ในขั้นตอนนี้สัญญาณแต่ละท่อนจะถูกนำมาผ่านตัวกรองที่ความถี่ย่านต่างๆ จากนั้นสัญญาณในย่านความถี่ต่างๆจะถูกนำไปหากรอบของสัญญาณ ดังสมการ

$$\text{Envelope} = |x + jH(x)|^2$$

โดยที่ $H(x)$ คือ การแปลงฮิลเบิร์ตของสัญญาณ x จากนั้นกรอบของสัญญาณจะถูกนำไปทำนอร์มัลไลเซชัน (Normalization) ด้วย Z-Score ผลลัพธ์ที่ได้จะเป็นเมทริกซ์ โดยเมทริกซ์แต่ละแถวจะเป็นค่า Z-score ของกรอบของสัญญาณในแต่ละช่วงความถี่ สำหรับเป็นข้อมูลขาเข้าของแบบจำลอง CNN โดยที่ n คือ จำนวนจุดข้อมูลในแต่ละท่อน เหตุผลที่ต้องมีขั้นตอนนี้เพื่อแยกสัญญาณออกเป็นช่วงความถี่ต่างๆของสัญญาณคลื่นไฟฟ้าสมองที่สำคัญ และสัญญาณรบกวนจากตาจะเกิดในย่านความถี่ 0.5 - 5 เฮิร์ตซ์ [7] สัญญาณรบกวนที่เกิดจากตาจะมีขนาดสัญญาณที่สูงกว่าขนาดสัญญาณของสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน ทำให้การหากรอบของสัญญาณมีประโยชน์ต่อการตรวจจับสัญญาณรบกวนที่เกิดจากตาเพราะกรอบของสัญญาณจะแสดงถึงขนาดของสัญญาณ



รูป 5: ขั้นตอนของการจัดการข้อมูลก่อนการประมวลผล

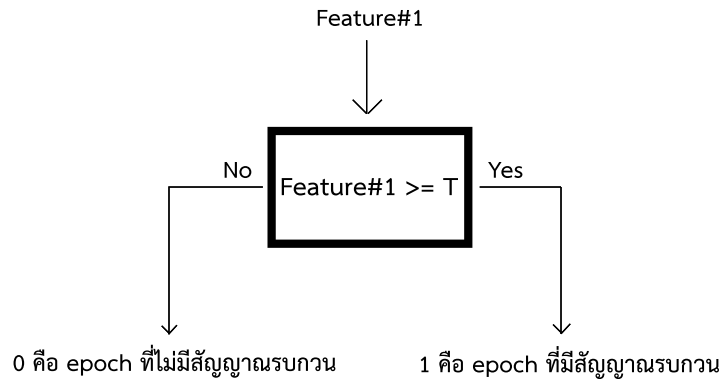
3.1.3 แบบจำลองสำหรับตรวจจับบริเวณที่มีสัญญาณรบกวน

แบบจำลองสำหรับตรวจจับบริเวณที่มีสัญญาณรบกวนในโครงงานนี้จะมีทั้งหมด 3 แบบจำลอง โดยในแบบจำลองแต่ละหัวข้อย่อยจะอธิบายหลักการพื้นฐานของแบบจำลองนั้นก่อน จากนั้นจะอธิบายวิธีการนำแบบจำลองนั้นไปใช้กับการตรวจจับบริเวณที่มีสัญญาณรบกวนของโครงงานนี้ แบบจำลองทั้ง 3 แบบจำลอง มีดังนี้

1. Thresholding

แบบจำลองนี้จะมีสมมติฐานสำหรับการตรวจจับบริเวณที่มีสัญญาณรบกวนคือ ถ้าพลังงานของสัญญาณในช่วงความถี่ 0.5 – 5 เฮิรตซ์ ในท่อนใดท่อนหนึ่ง มีค่าสูงกว่าค่า Threshold ค่าหนึ่ง ท่อนนั้นจะเป็นท่อนที่พบสัญญาณรบกวน ดัง รูปที่ 6 การหาค่า Threshold จะมีหลักการดังนี้ สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนในส่วนของคุณลักษณะเฉพาะในแต่ละท่อนตาม หัวข้อที่ 3.1.1 นั่นคือ พลังงานในย่านความถี่ 0.5 - 5 เฮิรตซ์ จากนั้นจะนำค่าพลังงานของแต่ละท่อนทุกค่า มาเรียงเพื่อเลือกค่า Threshold ซึ่งจะเลือกค่าที่เปอร์เซ็นต์ไทล์ที่ 85, 90, หรือ 95 เพื่อเป็นตัวแทนของค่าพลังงานสูงสุดในย่านความถี่ 0.5 - 5 เฮิรตซ์ ของสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน และใช้ค่าที่เลือกเป็นค่า Threshold ที่เลือกใช้ค่าพลังงานสูงสุดของสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนเพราะว่าค่าพลังงานของสัญญาณรบกวนควรมีค่าสูงกว่าค่าพลังงานสูงสุดของสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน ถ้าเลือกค่าที่ต่ำกว่านี้อาจจะทำให้มีการกำจัดสัญญาณคลื่นไฟฟ้าสมองไปด้วย เมื่อแบบจำลองถูกเลือกค่า Threshold ที่เหมาะสมแล้ว ต่อไปจะเป็นวิธีการนำแบบจำลองไปใช้จริง โดยสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจะถูกแบ่งเป็นท่อน จากนั้นสัญญาณในท่อนนั้นจะถูกนำไปแยกหาคุณลักษณะเฉพาะตาม หัวข้อที่ 3.1.1 ซึ่งใช้เฉพาะพลังงานในย่านความถี่ 0.5 - 5 เฮิรตซ์ คุณลักษณะเฉพาะที่ได้จะถูกนำไปใช้เปรียบเทียบกับค่า Threshold ในแบบจำลองดัง รูปที่ 6 ผลลัพธ์ที่ได้จะเป็นค่า 1 สำหรับท่อนที่มีสัญญาณรบกวน และค่า 0 สำหรับท่อนที่ไม่มีสัญญาณรบกวน

Feature#1 คือ ค่าพลังงานในย่านความถี่ 0.5 - 5 Hz ของแต่ละ epoch
T คือ ค่า Threshold



รูป 6: แบบจำลอง Thresholding

2. Random forest

แบบจำลองนี้เป็นแบบจำลองที่พัฒนามาจากแบบจำลองต้นไม้ตัดสินใจ (Decision tree) ในเบื้องต้นหลักการของ Classification trees [3] จะมีขั้นตอนดังนี้

- (a) จะแบ่งพื้นที่ของคุณลักษณะเฉพาะต่างๆออกเป็น J พื้นที่ที่ไม่มีารซ้อนทับกัน โดยเรียกแต่ละพื้นที่ย่อยว่า R_1, R_2, \dots, R_J และเรียกคุณลักษณะเฉพาะต่างๆว่า X_1, X_2, \dots, X_p , p คือ จำนวนคุณลักษณะเฉพาะทั้งหมด
- (b) สำหรับทุกๆข้อมูลที่ตกลงไปในพื้นที่ R_j กลุ่ม (Class) ของข้อมูลนั้นจะถูกทำนายจากกลุ่มของข้อมูลฝึกทั้งหมดที่มีมากที่สุดในพื้นที่ R_j นั้น

สำหรับในขั้นตอนที่ 1 จะแบ่งพื้นที่ของคุณลักษณะเฉพาะต่างๆออกเป็นพื้นที่ย่อย R_1, R_2, \dots, R_J โดยพื้นที่ย่อยที่แบ่งนั้น จะทำให้ค่า Gini index น้อยที่สุด โดย Gini index นิยามดังสมการ

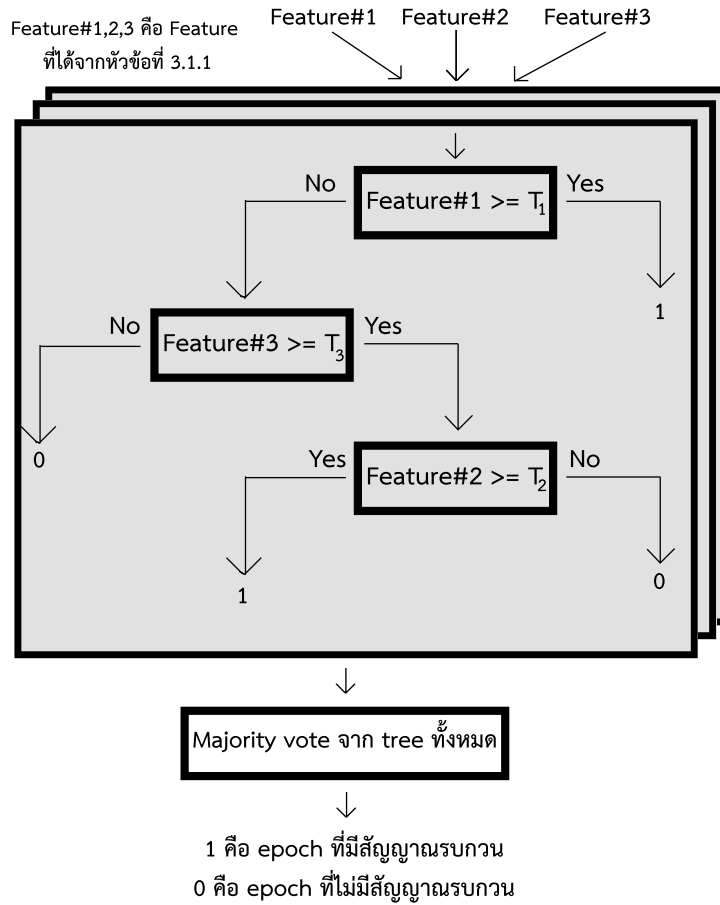
$$G = \sum_{k=1}^K \hat{p}_{mk}(1 - \hat{p}_{mk})$$

โดยที่ K คือ จำนวนกลุ่มทั้งหมด, \hat{p}_{mk} คือ สัดส่วนของกลุ่มที่ k ที่อยู่ในพื้นที่ที่ m จะเห็นได้ว่าค่า Gini index จะน้อยที่สุดก็ต่อเมื่อในพื้นที่ที่ถูกแบ่งมีจำนวนกลุ่มอยู่เพียงกลุ่มเดียวในพื้นที่นั้น ซึ่งพื้นที่จะถูกแบ่งไปเรื่อยๆ แต่ในแต่ละครั้งของการแบ่งจะมีเงื่อนไขที่พยายามทำให้ Gini index มีค่าน้อยที่สุด

จากที่กล่าวมาข้างต้น แบบจำลอง Random forest เป็นการรวมแบบจำลองต้นไม้ตัดสินใจ หลาย ๆ แบบจำลอง โดยแต่ละแบบจำลองของแบบจำลองต้นไม้ตัดสินใจ จะสุ่มเลือกคุณลักษณะเฉพาะมาเพียงบางส่วนเพื่อสร้างแบบจำลองต้นไม้ตัดสินใจ [3] ในแต่ละต้นไม้ตัดสินใจก็จะทำนายผลลัพธ์ออกมาเป็นกลุ่มต่าง ๆ จากนั้นจะทำการเลือกกลุ่มที่ถูกทำนายจำนวนมากที่สุด เพื่อเป็นผลลัพธ์การทำนายของแบบจำลอง Random forest กระบวนการดังกล่าวเป็นการลดค่าสหสัมพันธ์ระหว่างต้นไม้ตัดสินใจต่างๆ

ในขั้นตอนของการฝึกแบบจำลอง สัญญาณท่อนต่าง ๆ ของข้อมูลส่วนที่แบ่งไว้สำหรับการฝึก จะถูกติดป้าย (label) ไว้ว่าท่อนใดเป็นท่อนที่มีสัญญาณรบกวนและท่อนใดไม่มีสัญญาณรบกวน จากนั้นจะนำคุณลักษณะเฉพาะของท่อนนั้น ซึ่งได้จากหัวข้อที่ 3.1.1 และป้ายของท่อนนั้นเป็นข้อมูลขาเข้าสำหรับฝึกแบบจำลอง ผลลัพธ์ที่ได้จากการฝึกแบบจำลองก็คือแบบจำลองที่สามารถตรวจจับหรือจำแนกท่อนที่มีสัญญาณรบกวนหรือไม่มีสัญญาณรบกวนได้

เมื่อได้แบบจำลองแล้ว ต่อมาจะเป็นวิธีการนำแบบจำลองไปใช้จริง วิธีการนำไปใช้จะเป็นดัง รูปที่ 7 โดยสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจะถูกแบ่งเป็นท่อน จากนั้นสัญญาณในท่อนต่าง ๆ จะถูกนำไปแยกหาคุณลักษณะเฉพาะตาม หัวข้อที่ 3.1.1 คุณลักษณะเฉพาะที่ได้จะถูกนำไปใช้สำหรับเป็นข้อมูลขาเข้าของแบบจำลอง Random forest ผลลัพธ์ที่ได้จะเป็นค่า 1 สำหรับท่อนที่มีสัญญาณรบกวน และค่า 0 สำหรับท่อนที่ไม่มีสัญญาณรบกวน



รูป 7: แบบจำลอง Random forest

3. Convolutional neural network

ขั้นตอนของการฝึกแบบจำลอง สัญญาณท่อนต่าง ๆ ของข้อมูลส่วนที่แบ่งไว้สำหรับการฝึก จะถูกติดป้าย ไว้ว่าท่อนใดเป็นท่อนที่มีสัญญาณรบกวนและท่อนใดไม่มีสัญญาณรบกวน จากนั้นจะนำสัญญาณท่อนต่าง ๆ ไปทำการจัดการข้อมูลก่อนการประมวลผล ซึ่งเป็นไปตามหัวข้อที่ 3.1.2 ผลลัพธ์ที่ได้จะเป็นเมทริกซ์ โดยเมทริกซ์แต่ละแถวจะเป็นค่า Z-score ของกรอบของสัญญาณในแต่ละช่วงความถี่ จากนั้นจะนำเมทริกซ์ที่ได้และป้ายที่คู่กับเมทริกซ์นั้นเป็นข้อมูลขาเข้าสำหรับฝึกแบบจำลอง ผลลัพธ์ที่ได้จากการฝึกแบบจำลองก็คือแบบจำลองที่สามารถตรวจจับหรือจำแนกท่อนที่มีสัญญาณรบกวนหรือไม่มีสัญญาณรบกวนได้

เมื่อได้แบบจำลองแล้ว ต่อมาจะเป็นวิธีการนำแบบจำลองไปใช้จริง โดยสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนจะถูกแบ่งเป็นท่อน จากนั้นสัญญาณในท่อนต่าง ๆ จะถูกนำไปทำการจัดการข้อมูลก่อนการประมวลผล ซึ่งเป็นไปตาม หัวข้อที่ 3.1.2 เมทริกซ์ที่ได้จะถูกนำไปใช้สำหรับเป็นข้อมูลขาเข้าของแบบจำลอง CNN โดยในเลเยอร์แรกของแบบจำลองจะเป็นคอนโวลูชันเลเยอร์ เลเยอร์นี้มีความสำคัญ เพราะเป็นเลเยอร์ที่ไว้ใช้หาคุณลักษณะเฉพาะของเมทริกซ์ขาเข้า ผลลัพธ์ของแบบจำลองที่ได้จะเป็นค่า 1 สำหรับท่อนที่มีสัญญาณรบกวน และค่า 0 สำหรับท่อนที่ไม่มีสัญญาณรบกวน

3.2 การกำจัดสัญญาณรบกวนในบริเวณที่ตรวจจับได้

การกำจัดสัญญาณรบกวนในบริเวณที่พบว่ามีสัญญาณรบกวน จะใช้ตัวกรองสอวิทซกี-โกเลย์ส์สำหรับกำจัดสัญญาณรบกวน ซึ่งจะเลือกพหุนามอันดับที่ 3 สำหรับประมาณจุดที่สนใจ และจำนวนจุดข้างเคียงที่จะนำมาใช้ประมาณจำนวน 41 จุด [13] โดยจะมีขั้นตอนการกำจัดสัญญาณรบกวน ดัง รูปที่ 3 ข้อมูลแต่ละจุดของสัญญาณในบริเวณที่พบว่ามีสัญญาณรบกวนจะถูกประมาณค่าในแต่ละจุด ดัง สมการ (1) จากนั้นสัญญาณที่ได้จากการประมาณด้วยตัวกรองสอวิทซกี-โกเลย์ส์จะถูกนำไปลบออกจากบริเวณที่พบว่ามีสัญญาณรบกวน ผลลัพธ์ที่ได้จะเป็นสัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนถูกกำจัด

$$\hat{y}_j = \sum_{i=\frac{1-m}{2}}^{\frac{m-1}{2}} C_i y_{j+i}, \quad \frac{m-1}{2} \leq j \leq n - \frac{m-1}{2} \quad (1)$$

Polynomial Degree	Quadratic or Cubic (2 or 3)			Quartic or Quintic (4 or 5)	
Window size	5	7	9	7	9
y_{j-4}			-21		15
y_{j-3}		-2	14	5	-55
y_{j-2}	-3	3	39	-30	30
y_{j-1}	12	6	54	75	135
y_j	17	7	59	131	179
y_{j+1}	12	6	54	75	135
y_{j+2}	-3	3	39	-30	30
y_{j+3}		-2	14	5	-55
y_{j+4}			-21		15
Normalization	35	21	231	231	429

ตาราง 1: ตารางสัมประสิทธิ์คอนโวลูชัน

โดยที่ \hat{y}_j คือ ค่าที่ถูกประมาณจากจุดข้างเคียง, n คือ จำนวนจุดของข้อมูลในบริเวณที่พบสัญญาณรบกวน $j = 1, 2, \dots, n$, m คือ จำนวนจุดข้างเคียงที่จะนำมาใช้ประมาณ \hat{y}_j , y_{j+i} คือ ข้อมูลข้างเคียงที่นำมาใช้ประมาณ \hat{y}_j , C_i คือ สัมประสิทธิ์คอนโวลูชัน (Convolution coefficients) [12] โดยตัวอย่างค่าสัมประสิทธิ์คอนโวลูชันสามารถดูได้จาก ตารางที่ 1 ค่าสัมประสิทธิ์คอนโวลูชันจะมีหลักการหาตั้งนี้ จากคู่อันดับ (x_j, y_j) สมมติให้ พหุนามอันดับที่ k คือ $y_{j+i} = a_0 + a_1 z_i + a_2 z_i^2 + \dots + a_k z_i^k$ เป็นแบบจำลองของ y_{j+i} โดยที่ $z_i = \frac{x_{i+j} - x_j}{h}$ เมื่อ $h = x_{j+1} - x_j$ ถ้าจุดที่จะนำมาประมาณ y_j มีทั้งหมด m จุด จะได้ $i = \frac{1-m}{2}, \frac{3-m}{2}, \dots, 0, \dots, \frac{m-3}{2}, \frac{m-1}{2}$ นั่นคือ $z_{\frac{1-m}{2}} = \frac{1-m}{2}, \dots, z_0 = 0, \dots, z_{\frac{m-1}{2}} = \frac{m-1}{2}$ และจะได้ $y = J\theta$ โดยที่แต่ละแถวของ J มีสมาชิก $1, z_i, z_i^2, \dots, z_i^k, y = [y_{j+\frac{1-m}{2}} \quad y_{j+\frac{3-m}{2}} \quad \dots \quad y_{j+\frac{m-1}{2}}]^T$ สัมประสิทธิ์ $\theta = [a_0 \quad a_1 \quad \dots \quad a_k]^T$ จะหาได้จากการพยายามทำให้ค่า $\| \hat{y} - y \|^2$ มีค่าน้อยที่สุด ซึ่งมีผลเฉลยรูปแบบปิด (Closed form solution) คือ $\hat{\theta} = (J^T J)^{-1} J^T y$ สมมติให้แถวที่ 1 ของ $(J^T J)^{-1} J^T$ เท่ากับ $[C_{\frac{1-m}{2}} \quad \dots \quad C_0 \quad \dots \quad C_{\frac{m-1}{2}}]$ ซึ่งจุด y_j จะถูกประมาณจาก $\hat{y}_j = a_0 + a_1 z_0 + a_2 z_0^2 + \dots + a_k z_0^k$ ซึ่ง $z_0 = 0$ ดังนั้น $\hat{y}_j = a_0$ นั่นคือ $\hat{y}_j = a_0 = \hat{\theta}_1 = C_{\frac{1-m}{2}} y_{j+\frac{1-m}{2}} + C_{\frac{3-m}{2}} y_{j+\frac{3-m}{2}} + \dots + C_0 y_j + \dots + C_{\frac{m-3}{2}} y_{j+\frac{m-3}{2}} + C_{\frac{m-1}{2}} y_{j+\frac{m-1}{2}}$ จึงสรุปได้ว่า $C_{\frac{1-m}{2}}, \dots, C_0, \dots, C_{\frac{m-1}{2}}$ คือสัมประสิทธิ์คอนโวลูชัน

3.3 การประเมินประสิทธิภาพของวิธีการกำจัดสัญญาณรบกวน

กำหนดให้ x คือ สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน, \tilde{x} คือ สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน, \hat{x} คือ สัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนถูกกำจัด การประเมินประสิทธิภาพของวิธีการกำจัดสัญญาณรบกวนหนึ่ง ๆ จะใช้ดัชนีดังต่อไปนี้

1. กำหนดให้ ρ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง \tilde{x} กับ x และ $\hat{\rho}$ คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง \tilde{x} กับ \hat{x} ดังนั้นเราจะใช้ดัชนีความแตกต่างของสัมประสิทธิ์สหสัมพันธ์

$$\Delta\rho = \hat{\rho} - \rho$$

เพื่อแสดงว่าหลังจากที่สัญญาณรบกวนถูกกำจัด ความมีลักษณะร่วมกันของ \tilde{x} กับ \hat{x} มีมากขึ้นหรือน้อยลง เมื่อเทียบกับก่อนสัญญาณรบกวนถูกกำจัด

2. ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน (Normalized mean square error, NMSE)

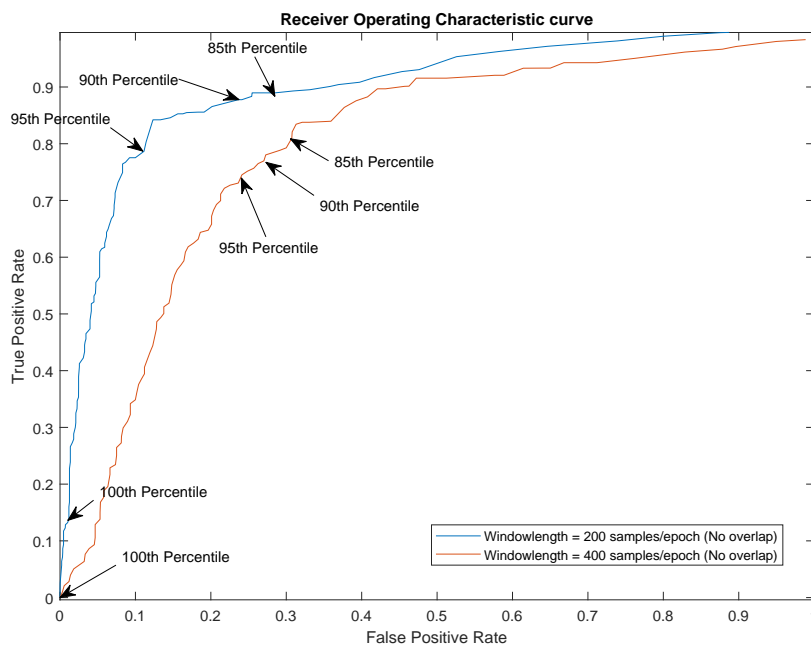
$$NMSE = 10 \log_{10} \left(\frac{1}{M} \sum_{i=1}^M \left(\frac{\sum_{j=1}^N |\hat{x}_j - \tilde{x}_j|^2}{\sum_{j=1}^N |\tilde{x}_j|^2} \right) \right)$$

ค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐานมีไว้ใช้หาความคลาดเคลื่อนระหว่างสัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนถูกกำจัดกับสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน โดย N คือ จำนวนจุดของข้อมูลทั้งหมด, M คือ จำนวนครั้งของการกำจัดสัญญาณรบกวนของแบบจำลองหนึ่ง

4 ผลลัพธ์จากการดำเนินการ

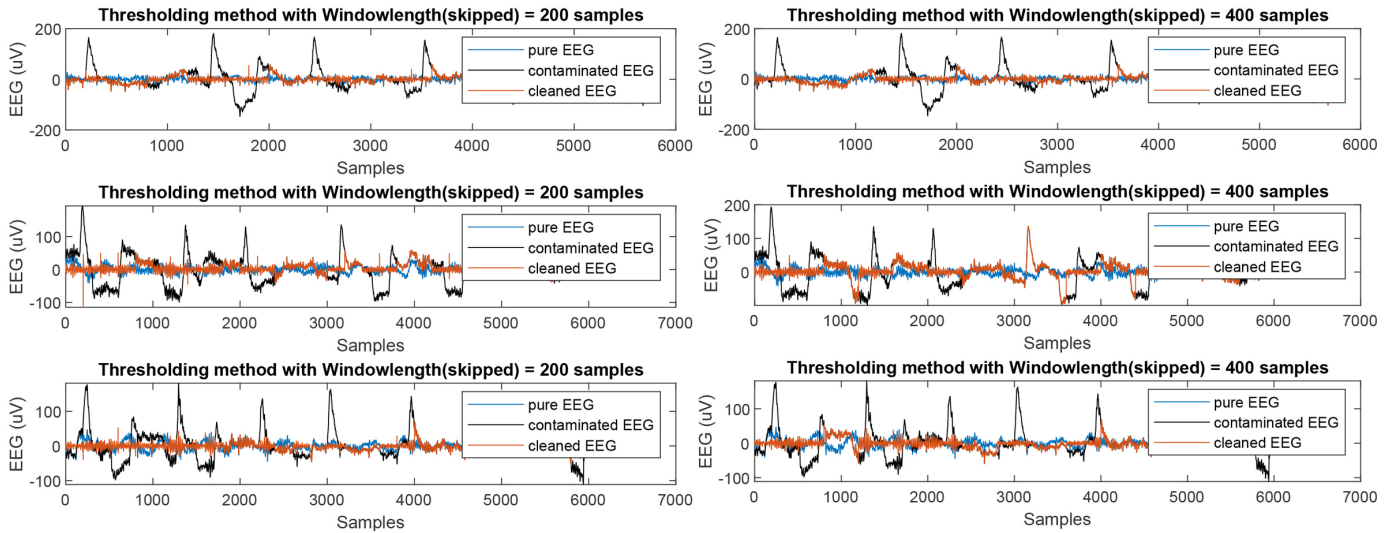
ผลลัพธ์จากการดำเนินงานนี้เป็นผลลัพธ์ที่ได้จากการทดลองกำจัดสัญญาณรบกวนด้วยวิธี Thresholding โดยจะมีการแบ่งการทดลองเป็น 2 การทดลอง ดังนี้ 1) แบ่งสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนให้จำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล การแบ่งท่อนไม่มีการซ้อนทับกัน 2) แบ่งสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนให้จำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล การแบ่งท่อนไม่มีการซ้อนทับกัน เหตุผลที่เลือกจำนวนจุดในแต่ละท่อนตามจำนวนดังกล่าวเพราะว่าสัญญาณรบกวนที่เกิดจากตาจะเกิดในย่านความถี่ 0.5 - 5 เฮิร์ตซ์ และการบันทึกข้อมูลมีอัตราการสุ่มเท่ากับ 200 เฮิร์ตซ์ ดังนั้นจำนวนจุดในหนึ่งท่อนที่ครอบคลุมการเกิดสัญญาณรบกวนหนึ่งครั้งคือ 400 จุดข้อมูล จึงทำการทดลองที่ 400 จุดข้อมูล และ 200 จุดข้อมูลซึ่งเป็นความยาวท่อนที่แคบลง ในขั้นตอนแรกสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนในส่วนของข้อมูลที่แบ่งไว้สำหรับการฝึกจะถูกนำมาหาคุณลักษณะเฉพาะในแต่ละท่อนนั้นคือ พลังงานในย่านความถี่ 0.5 - 5 เฮิร์ตซ์ จากนั้นจะนำค่าพลังงานของแต่ละท่อนทุกค่ามาเรียงเพื่อเลือกค่า Threshold ซึ่งจะเลือกค่าที่เปอร์เซ็นต์ไทล์ที่ 85, 90, หรือ 95 เพื่อเป็นตัวแทนของค่าพลังงานสูงสุดที่ย่านความถี่ 0.5 - 5 เฮิร์ตซ์ ของสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน จากนั้นจะนำค่าที่เปอร์เซ็นต์ไทล์ต่างๆไปพิจารณาบน Receiver operating characteristic curve (ROC curve) เพื่อดูประสิทธิภาพของการตรวจจับบริเวณที่มีสัญญาณรบกวน

1. ผลลัพธ์จากการพล็อต ROC curve ดัง รูปที่ 8 จะเห็นได้ว่าการแบ่งจำนวนจุดข้อมูลในแต่ละท่อนส่งผลให้ได้ ROC curve ที่แตกต่างกัน การแบ่งสัญญาณให้จำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล ให้ผลลัพธ์ที่น่าพอใจมากที่สุด โดยจะมีค่า True Positive Rate (TPR) ที่สูงกว่าการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล เมื่อเทียบกับจุด False Positive Rate (FPR) เดียวกัน และค่า Threshold ที่เปอร์เซ็นต์ไทล์ต่างๆ ก็ให้ประสิทธิภาพที่แตกต่างกัน โดยการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล จะให้ค่า TPR ที่สูงกว่า และ FPR ที่ต่ำกว่าการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล เมื่อเทียบกับเปอร์เซ็นต์ไทล์เดียวกัน

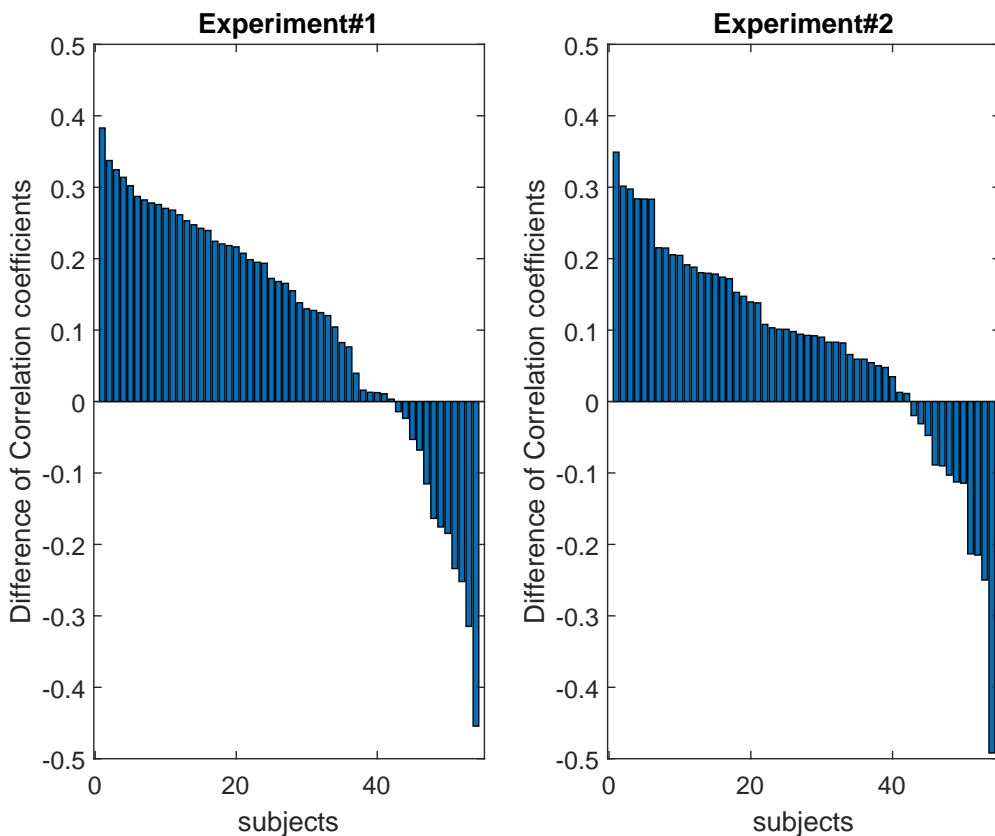


รูป 8: Receiver operating characteristic curve เมื่อ เส้นสีน้ำเงิน จะได้จากการแบ่งสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนให้จำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล การแบ่งท่อนไม่มีการซ้อนทับกัน, เส้นสีแดง จะได้จากการแบ่งสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนให้จำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล การแบ่งท่อนไม่มีการซ้อนทับกัน

2. ผลลัพธ์จากการกำจัดสัญญาณรบกวนของแต่ละการทดลองโดยใช้ค่า Threshold ค่าที่เปอร์เซ็นต์ไทล์ที่ 90 ดัง รูปที่ 9 จะเห็นได้ว่า ให้ผลลัพธ์ที่ใกล้เคียงกัน แต่การทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล จะมีการกำจัดสัญญาณรบกวนในบริเวณที่จริง ๆ แล้วบริเวณนั้นไม่มีสัญญาณรบกวน มากกว่าการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล ทำให้มีการคำนวณที่เพิ่มมากขึ้นเกินความจำเป็น อย่างไรก็ตามจากผลลัพธ์ดัง รูปที่ 10 แสดงให้เห็นว่าการทดลองทั้ง 2 วิธี ส่วนใหญ่แล้วให้ผลลัพธ์ของค่าความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ ($\Delta\rho = \hat{\rho} - \rho$) ที่มีค่าเป็นบวก นั่นคือสัญญาณคลื่นไฟฟ้าสมองที่ถูกกำจัดสัญญาณรบกวนมีคุณลักษณะทางกายภาพที่คล้ายกับสัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนมากขึ้น แต่จะเห็นได้ว่าการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูลส่วนใหญ่แล้วมีค่าที่เป็นบวกมากกว่าการทดลองที่แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล



รูป 9: ตัวอย่างผลจากการกำจัดสัญญาณรบกวนด้วยวิธีการ Thresholding (ซ้าย) แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล และขวา) แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล เมื่อ pure EEG คือ สัญญาณคลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวน, contaminated EEG คือ สัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวน และ cleaned EEG คือ สัญญาณคลื่นไฟฟ้าสมองที่สัญญาณรบกวนถูกกำจัด



รูป 10: ความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ ($\Delta \rho = \hat{\rho} - \rho$) เมื่อ (ซ้าย) แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล และขวา) แบ่งจำนวนจุดข้อมูลในท่อนเท่ากับ 400 จุดข้อมูล

จากผลการทดลองของการพล็อต ROC curve การกำจัดสัญญาณรบกวน และความแตกต่างของค่าสัมประสิทธิ์สหสัมพันธ์ในการทดลองต่างๆ จึงสรุปได้ว่า การตรวจจับและกำจัดสัญญาณรบกวนด้วยวิธี Thresholding ให้ผลลัพธ์ที่น่าพึงพอใจ และการแบ่งสัญญาณคลื่นไฟฟ้าสมองที่มีสัญญาณรบกวนให้จำนวนจุดข้อมูลในท่อนเท่ากับ 200 จุดข้อมูล การแบ่งท่อนไม่มีการซ้อนทับกัน ให้ผลลัพธ์ที่เหมาะสมสำหรับเป็นวิธีหนึ่งที่น่าไปใช้กับการแบ่งสัญญาณคลื่นไฟฟ้าสมองสำหรับเป็นข้อมูลขาเข้าของวิธีการกำจัดสัญญาณรบกวนวิธีการอื่น

5 รายละเอียดของหัวข้อโครงการ

5.1 ขอบเขตของโครงการ

1. การกำจัดสัญญาณรบกวนจะกำจัดสัญญาณรบกวนที่เกิดจากตาเท่านั้น อาทิ การเคลื่อนที่ของตา การกะพริบตา ไม่สนใจสัญญาณรบกวนที่เกิดจากกล้ามเนื้อ และสัญญาณรบกวนที่เกิดจากสัญญาณคลื่นไฟฟ้าหัวใจ
2. วิธีการที่จะศึกษาสำหรับกำจัดสัญญาณรบกวนคือ Thresholding, Random forest และ Convolutional neural network
3. ข้อมูลที่นำมาใช้จะเป็นข้อมูลจำลองการบันทึกสัญญาณจากเครื่องบันทึกคลื่นไฟฟ้าสมองหนึ่งช่องสัญญาณ และไม่มีการบันทึกสัญญาณอ้างอิงร่วมด้วย
4. การเปรียบเทียบของวิธีการกำจัดสัญญาณรบกวนที่เกิดจากตา จะทำบนข้อมูลสังเคราะห์
5. การเปรียบเทียบประสิทธิภาพของการกำจัดสัญญาณรบกวนที่เกิดจากตา ผู้ศึกษาจะเปรียบเทียบด้วยความแตกต่างของสัมประสิทธิ์สหสัมพันธ์ และค่าความคลาดเคลื่อนกำลังสองเฉลี่ยแบบบรรทัดฐาน ของวิธีการต่างๆ

5.2 แผนการดำเนินการ

แผนการดำเนินการจะเป็นไปตาม ตารางที่ 2 สีเทา คือ ความก้าวหน้าที่วางแผนไว้, สีดำ คือ ความก้าวหน้าปัจจุบัน

ขั้นตอนการดำเนินงาน	วิชา 2102490				วิชา 2102499				
	ส.ค.	ก.ย.	ต.ค.	พ.ย.	ธ.ค.	ม.ค.	ก.พ.	มี.ค.	เม.ย.
1. ศึกษาวิธีการกำจัดสัญญาณรบกวนในคลื่นไฟฟ้าสมอง และลักษณะเฉพาะของสัญญาณคลื่นไฟฟ้าสมองและสัญญาณรบกวน									
2. เรียนรู้วิธีการ Principal component analysis และ Independent component analysis									
3. ทดลองตรวจจับสัญญาณรบกวนด้วยวิธีการ Thresholding และกำจัดสัญญาณรบกวนด้วย Svitzy-Golay filter									
4. ศึกษาวิธีการกำจัดสัญญาณรบกวนด้วยวิธีการเรียนรู้ของเครื่อง									
5. สรุปรวมโครงงาน และเลือกวิธีการกำจัดสัญญาณรบกวนที่จะนำมาเปรียบเทียบ									
6. จัดทำรูปเล่ม proposal และเตรียมตัวนำเสนอ									
7. ศึกษาวิธีการสังเคราะห์สัญญาณคลื่นไฟฟ้าสมอง									
8. ทดลองตรวจจับสัญญาณรบกวนด้วยแบบจำลอง Random forest และปรับจูน Hyper parameters ของแบบจำลอง									
9. ทดลองตรวจจับสัญญาณรบกวนด้วยแบบจำลอง CNN และปรับเปลี่ยน layer ของแบบจำลอง									
10. ศึกษาและเปรียบเทียบ Computational complexity ของแต่ละวิธีการ									
11. สรุปลงการทดลอง และจัดทำรูปเล่มรายงาน เตรียมตัวนำเสนอ									

ตาราง 2: แผนการดำเนินการ การทำโครงงานวิศวกรรม ปีการศึกษา 2562

5.3 อุปสรรคและปัญหาที่พบ และการแก้ไข

เนื่องจากข้อมูลที่น้อยเกินไปทำให้ข้อมูลไม่เพียงพอสำหรับการฝึกของแบบจำลอง Random forest และ CNN ในตอนนี้มีฟังก์ชันสำหรับสังเคราะห์คลื่นไฟฟ้าสมองที่ไม่มีสัญญาณรบกวนแล้ว ในเทอมหน้าจะศึกษาวิธีสังเคราะห์สัญญาณรบกวนที่เกิดจากตา

6 เอกสารอ้างอิง

- [1] WY. Hsu, CH. Lin, HJ. Hsu, PH. Chen, and IR. Chen. Wavelet-based envelope features with automatic EOG artifact removal: Application to single-trial EEG data. *Expert Systems with Applications*, 39(3):2743–2749, 2012.
- [2] M. K. Islam, A. Rastegarnia, and Z. Yang. Methods for artifact detection and removal from scalp EEG: A review. *Neurophysiologie Clinique/Clinical Neurophysiology*, 46(4-5):287–305, 2016.
- [3] G. James, D. Witten, T. Hastie, and R. Tibshirani. *An introduction to statistical learning*. Springer, 2013.
- [4] S. Kanoga and Y. Mitsukura. Review of artifact rejection methods for electroencephalographic systems. *Electroencephalography*, page 69, 2017.
- [5] M. A. Klados and P. D. Bamidis. A semi-simulated EEG/EOG dataset for the comparison of EOG artifact rejection techniques. *Data in brief*, 8:1004–1006, 2016.
- [6] T. Lee, M. Girolami, and T. Sejnowski. Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural computation*, 11(2):417–441, 1999.
- [7] O. G. Lins, T. W. Picton, P. Berg, and M. Scherg. Ocular artifacts in EEG and event-related potentials I: Scalp topography. *Brain topography*, 6(1):51–63, 1993.
- [8] J. Minguillon, M. A. Lopez-Gordo, and F. Pelayo. Trends in EEG-BCI for daily-life: Requirements for artifact removal. *Biomedical Signal Processing and Control*, 31:407–418, 2017.
- [9] P. Nejedly, J. Cimbalnik, P. Klimes, F. Plesinger, J. Halamek, V. Kremen, I. Viscor, B. Brinkmann, M. Pail, M. Brazdil, et al. Intracerebral EEG artifact identification using convolutional neural networks. *Neuroinformatics*, 17(2):225–234, 2019.
- [10] T. Radüntz, J. Scouten, O. Hochmuth, and B. Meffert. EEG artifact elimination by extraction of ICA-component features using image processing algorithms. *Journal of neuroscience methods*, 243:84–93, 2015.
- [11] E. Saifutdinova, D. Dudysová, L. Lhotská, V. Gerla, and M. Macaš. Artifact detection in multichannel sleep EEG using random forest classifier. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2803–2805. IEEE, 2018.
- [12] A. Savitzky and M. J. Golay. Smoothing and differentiation of data by simplified least squares procedures. *Analytical chemistry*, 36(8):1627–1639, 1964.
- [13] D. Szibbo, A. Luo, and T. Sullivan. Removal of blink artifacts in single channel EEG. In *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 3511–3514. IEEE, 2012.
- [14] M. Unser and A. Aldroubi. A review of wavelets in biomedical applications. *Proceedings of the IEEE*, 84(4):626–638, 1996.
- [15] J. A. Urigüen and B. Garcia-Zapirain. EEG artifact removal—state-of-the-art and guidelines. *Journal of neural engineering*, 12(3):031001, 2015.
- [16] D. Coppieters’t Wallant, V. Muto, G. Gaggioni, M. Jaspar, S. L. Chellappa, C. Meyer, G. Vandewalle, P. Maquet, and C. Phillips. Automatic artifacts and arousals detection in whole-night sleep EEG recordings. *Journal of neuroscience methods*, 258:124–133, 2016.