

การประยุกต์ใช้ปัญญาประดิษฐ์ในการตรวจจับการกระตุกของกล้ามเนื้อใบหน้า

ARTIFICIAL INTELLIGENCE FOR DETECTION OF FACIAL MUSCLE TWITCHING

นภธร พูลเพียร^{*1}

Napathorn Punpain^{*1}

^{*1} วิทยาลัยวิศวกรรมศาสตร์และเทคโนโลยี มหาวิทยาลัยธุรกิจบัณฑิต

^{*1} College of Engineering and Technology, Dhurakij Pundit University

* E-mail ผู้รับผิดชอบบทความ (Corresponding Author): kanyarat.pun@dpu.ac.th

วันที่รับบทความ 30 เมษายน 2569 วันที่แก้ไขบทความ 11 มิถุนายน 2569 วันที่ตอบรับบทความ 26 มิถุนายน 2569

บทคัดย่อ

การกระตุกของกล้ามเนื้อใบหน้าเป็นสัญญาณเชิงสรีรวิทยาที่มีลักษณะเป็นการเคลื่อนไหวขนาดเล็กและเกิดขึ้นในช่วงเวลาสั้น ซึ่งยากต่อการตรวจจับด้วยสายตามนุษย์ งานวิจัยนี้มีวัตถุประสงค์เพื่อทบทวนวรรณกรรมอย่างเป็นระบบเกี่ยวกับการประยุกต์ใช้ปัญญาประดิษฐ์ในการวิเคราะห์การเคลื่อนไหวของใบหน้า และเสนอกรอบแนวคิดเชิงวิศวกรรมสำหรับระบบคัดกรองความผิดปกติทางระบบประสาท โดยดำเนินการตามแนวทาง PRISMA จากการคัดเลือกบทความจำนวน 320 บทความ พบว่ามีบทความที่ผ่านเกณฑ์และนำมาวิเคราะห์จำนวน 35 บทความ

ผลการศึกษาพบว่าแบบจำลอง Convolutional Neural Networks (CNN) ถูกนำมาใช้มากที่สุดคิดเป็นประมาณร้อยละ 60 รองลงมาคือ Long Short-Term Memory (LSTM) ร้อยละ 25 และแบบจำลองแบบผสมผสาน CNN-LSTM ร้อยละ 15 โดยมีค่าความแม่นยำเฉลี่ยอยู่ในช่วง 82–86%, 78–84% และ 88–92% ตามลำดับ อย่างไรก็ตาม มากกว่าร้อยละ 70 ของงานวิจัยยังใช้ชุดข้อมูล micro-expression เช่น CASME II และมีเพียงส่วนน้อยที่ได้รับการทดสอบในสภาพแวดล้อมจริง งานวิจัยนี้จึงเสนอกรอบแนวคิดของระบบคัดกรองที่ประกอบด้วย 6 ขั้นตอนหลัก พร้อมทั้งสังเคราะห์ข้อจำกัดของชุดข้อมูลและแบบจำลองที่ใช้ในงานวิจัยปัจจุบัน ตลอดจนชี้ให้เห็นแนวโน้มของเทคนิคสมัยใหม่ เช่น Transformer-based, Graph-based และ Diffusion-based models เพื่อเป็นแนวทางสำหรับการพัฒนาระบบคัดกรองทางระบบประสาทในอนาคต

คำสำคัญ: ปัญญาประดิษฐ์; การกระตุกของกล้ามเนื้อใบหน้า; คอมพิวเตอร์วิทัศน์; การคัดกรองระบบประสาท; การเรียนรู้เชิงลึก

ABSTRACT

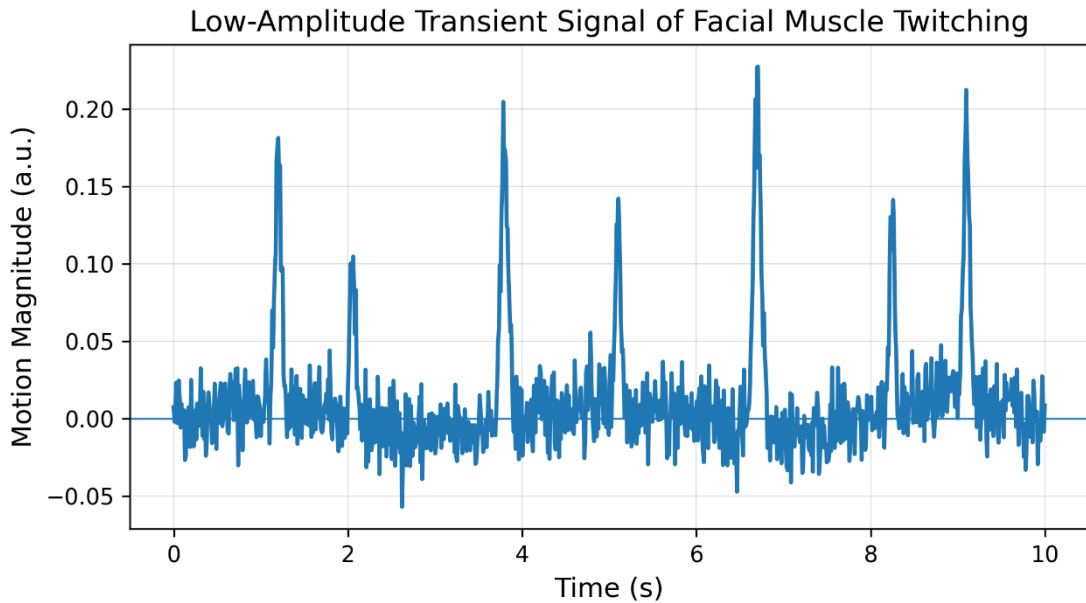
Facial muscle twitching is a physiological signal characterized by subtle, low-amplitude, and short-duration facial movements, making it difficult to detect through conventional visual observation. This study aims to systematically review the application of artificial intelligence in facial movement analysis and to propose an engineering-based framework for neuromuscular screening systems. The review was conducted following the PRISMA guidelines. From an initial pool of 320 articles, 35 studies met the inclusion criteria and were selected for detailed analysis.

The findings indicate that Convolutional Neural Networks (CNNs) were the most frequently used models, accounting for approximately 60% of the reviewed studies, followed by Long Short-Term Memory (LSTM) models at 25% and hybrid CNN-LSTM models at 15%. The reported accuracy ranges were 82–86%, 78–84%, and 88–92%, respectively. However, more than 70% of the studies still relied on micro-expression datasets such as CASME II, and only a limited number of studies were validated in real-world environments. Therefore, this study proposes a six-stage screening framework and synthesizes the limitations of current datasets and models, as well as emerging trends in modern techniques such as Transformer-based, Graph-based, and Diffusion-based models. These findings provide guidance for the future development of neuromuscular screening systems.

Keywords: Artificial intelligence; Facial muscle twitching; Computer vision; Neuromuscular screening; Deep learning

1. บทนำ

การเคลื่อนไหวของกล้ามเนื้อใบหน้าเป็นสัญญาณทางชีวกลศาสตร์ที่สะท้อนการทำงานของระบบประสาทและกล้ามเนื้อ โดยเฉพาะการกระตุกของกล้ามเนื้อใบหน้า (facial muscle twitching) ซึ่งมีลักษณะเป็นการเคลื่อนไหวขนาดเล็ก เกิดขึ้นในช่วงเวลาสั้น และมีแอมพลิจูดต่ำ ทำให้ยากต่อการตรวจจับด้วยการสังเกตด้วยสายตามนุษย์ ลักษณะของสัญญาณดังกล่าวมีความไม่ต่อเนื่องและเปลี่ยนแปลงอย่างรวดเร็ว ดังภาพที่ 1 ซึ่งลักษณะดังกล่าวเป็นข้อจำกัดสำคัญในการตรวจจับด้วยวิธีดั้งเดิม และเป็นเหตุผลหลักในการนำเทคนิคปัญญาประดิษฐ์มาใช้ในการวิเคราะห์



ภาพที่ 1 สัญญาณการกระตุกของกล้ามเนื้อใบหน้าในรูปแบบ time-series

ในเชิงวิศวกรรม การตรวจจับสัญญาณการกระตุกของกล้ามเนื้อใบหน้าถือเป็นปัญหาที่มีความซับซ้อนเนื่องจากสัญญาณมีขนาดเล็กและเกิดแบบไม่ต่อเนื่อง งานวิจัยที่ผ่านมาได้ใช้แนวทาง Systematic Review ตามหลัก PRISMA ในการสังเคราะห์องค์ความรู้เพื่อเพิ่มความน่าเชื่อถือของผลการศึกษา [1] พร้อมทั้งนำเทคนิคปัญญาประดิษฐ์เข้ามาช่วยในการวิเคราะห์ข้อมูลเชิงภาพและเชิงเวลา โดยแบบจำลอง Convolutional Neural Networks (CNN) สามารถสกัดคุณลักษณะเชิงพื้นที่จากภาพได้อย่างมีประสิทธิภาพ [2] ขณะที่ Long Short-Term Memory (LSTM) มีความเหมาะสมต่อการวิเคราะห์ลำดับข้อมูลเชิงเวลา [3]

นอกจากนี้ งานวิจัยหลายฉบับรายงานว่าแบบจำลองแบบผสมผสาน CNN-LSTM สามารถให้ผลลัพธ์ที่ดีกว่าแบบจำลองเดี่ยว เนื่องจากสามารถเรียนรู้ทั้ง spatial และ temporal features ได้พร้อมกัน [4] อย่างไรก็ตาม งานวิจัยส่วนใหญ่ยังอาศัยชุดข้อมูล micro-expression เช่น CASME II และ SAMM ซึ่งถูกออกแบบเพื่อศึกษาการแสดงออกทางอารมณ์มากกว่าการตรวจจับการกระตุกของกล้ามเนื้อโดยตรง [5], [6] อีกทั้งยังขาดการทดสอบในสภาพแวดล้อมจริง ดังนั้น งานวิจัยนี้จึงมีวัตถุประสงค์เพื่อทบทวนวรรณกรรมอย่างเป็นระบบและเสนอกรอบแนวคิดเชิงวิศวกรรมสำหรับระบบคัดกรองการกระตุกของกล้ามเนื้อใบหน้า

1.1 คุณูปการของงานวิจัย

บทความนี้มีคุณูปการสำคัญในเชิงวิชาการและเชิงวิศวกรรม ดังนี้

1. บทความนี้รวบรวมและสังเคราะห์งานวิจัยที่เกี่ยวข้องกับการใช้ปัญญาประดิษฐ์ในการวิเคราะห์การเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้าอย่างเป็นระบบ โดยคัดเลือกบทความจำนวน 35 บทความจากฐานข้อมูลวิชาการหลัก ได้แก่ Scopus, IEEE Xplore และ PubMed ตามแนวทาง PRISMA
2. บทความนี้นำเสนอการเปรียบเทียบเชิงวิพากษ์ของชุดข้อมูลที่ใช้ในงานวิจัยด้าน micro-expression และ facial motion analysis เช่น CASME II, SAMM และชุดข้อมูลที่เกี่ยวข้อง เพื่อชี้ให้เห็นข้อจำกัดของการนำชุดข้อมูลเหล่านั้นมาใช้เป็นตัวแทนของการกระตุกของกล้ามเนื้อใบหน้าโดยตรง

3. บทความนี้สังเคราะห์วิวัฒนาการของแบบจำลองปัญญาประดิษฐ์ที่ใช้ในการวิเคราะห์การเคลื่อนไหวของใบหน้า ตั้งแต่แบบจำลองกลุ่ม CNN, LSTM และ CNN-LSTM ไปจนถึงแนวทางสมัยใหม่ เช่น Transformer-based models, graph-based representation และ diffusion-based models ซึ่งเป็นแนวโน้มสำคัญของงานวิจัยในอนาคต
4. บทความนี้วิเคราะห์ข้อจำกัดเชิงเทคนิคของแต่ละแนวทาง ได้แก่ ความสามารถในการสกัดคุณลักษณะเชิงพื้นที่ ความสามารถในการเรียนรู้ลำดับเวลา ความซับซ้อนของแบบจำลอง ความต้องการข้อมูลฝึกสอน และความเหมาะสมต่อการใช้งานแบบ real-time
5. บทความนี้เสนอกรอบแนวคิดเชิงวิศวกรรมสำหรับระบบคัดกรองการกระตุกของกล้ามเนื้อใบหน้า โดยเชื่อมโยงกระบวนการตั้งแต่การรับข้อมูลวิดีโอ การตรวจจับใบหน้า การขยายสัญญาณการเคลื่อนไหว การสกัดคุณลักษณะ การวิเคราะห์เชิงเวลา และการจำแนกระดับความเสี่ยง เพื่อใช้เป็นแนวทางสำหรับการพัฒนาระบบในอนาคต

2. เอกสารและงานวิจัยที่เกี่ยวข้อง

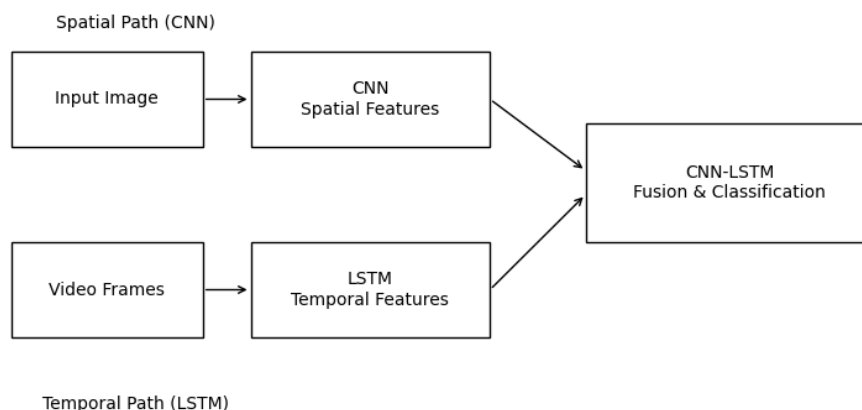
การศึกษาที่เกี่ยวข้องกับการวิเคราะห์การเคลื่อนไหวของกล้ามเนื้อใบหน้าในช่วงทศวรรษที่ผ่านมาได้พัฒนาอย่างต่อเนื่อง โดยเฉพาะในบริบทของ micro-expression และ facial behavior analysis ซึ่งถูกนำมาใช้เป็นตัวแทนของการเคลื่อนไหวระดับไมโครของใบหน้า [7], [8] แม้ว่าการวิจัยเหล่านี้จะช่วยผลักดันความก้าวหน้าด้านคอมพิวเตอร์วิทัศน์และการเรียนรู้เชิงลึกอย่างมีนัยสำคัญ แต่แนวทางส่วนใหญ่ยังคงเน้นไปที่การจำแนกอารมณ์มากกว่าการตรวจจับความผิดปกติของกล้ามเนื้อโดยตรง ส่งผลให้การนำไปประยุกต์ใช้ในบริบทของการคัดกรองทางระบบประสาทยังมีข้อจำกัดอย่างชัดเจน นอกจากนี้ การประเมินผลของแบบจำลองมักดำเนินการบนชุดข้อมูลที่มีการควบคุมสภาพแวดล้อมสูง ซึ่งอาจไม่สะท้อนสถานะจริงของการใช้งาน

2.1 ชุดข้อมูลการเคลื่อนไหวของใบหน้า (Facial Motion Datasets)

ชุดข้อมูลมาตรฐาน เช่น CASME II และ SAMM ได้รับการใช้อย่างแพร่หลายในการฝึกและประเมินแบบจำลอง เนื่องจากมีการบันทึกวิดีโอความละเอียดสูงและมีการระบุช่วงเวลาของการเคลื่อนไหวอย่างละเอียด [7], [8] อย่างไรก็ตาม ข้อจำกัดสำคัญของชุดข้อมูลเหล่านี้คือการออกแบบเพื่อศึกษาการแสดงออกทางอารมณ์มากกว่าการกระตุกของกล้ามเนื้อใบหน้าโดยเฉพาะ ส่งผลให้ลักษณะของสัญญาณที่ได้อาจไม่สอดคล้องกับลักษณะทางคลินิก นอกจากนี้ จำนวนข้อมูลยังมีจำกัด และมักถูกรวบรวมภายใต้สภาพแวดล้อมที่ควบคุมอย่างเข้มงวด ทำให้เกิดปัญหา domain shift เมื่อนำไปใช้กับข้อมูลในสภาพแวดล้อมจริง

2.2 แบบจำลองการเรียนรู้เชิงลึก (Deep Learning Models)

แบบจำลอง CNN ถูกนำมาใช้อย่างแพร่หลายในการสกัดคุณลักษณะเชิงพื้นที่จากภาพใบหน้า เนื่องจากมีความสามารถในการเรียนรู้รูปแบบเชิงลึกของข้อมูลภาพได้ดี [9] ขณะที่แบบจำลองเชิงเวลา เช่น RNN และ LSTM มีความสามารถในการวิเคราะห์ลำดับข้อมูลที่เปลี่ยนแปลงตามเวลา [10] อย่างไรก็ตาม การใช้แบบจำลองเดี่ยวมักไม่สามารถจับลักษณะของการเคลื่อนไหวที่ซับซ้อนและมีความแปรปรวนสูงได้อย่างครบถ้วน ทำให้เกิดการพัฒนารูปแบบจำลองแบบผสมผสาน เช่น CNN-LSTM ซึ่งสามารถเรียนรู้ทั้ง spatial และ temporal features ได้พร้อมกัน และให้ผลลัพธ์ที่แม่นยำยิ่งขึ้น [11] การผสมผสานแบบจำลอง CNN และ LSTM ทำให้สามารถเรียนรู้ทั้งคุณลักษณะเชิงพื้นที่และเชิงเวลาได้พร้อมกัน ดังภาพที่ 2



ภาพที่ 2 การเปรียบเทียบแนวคิดของแบบจำลอง CNN, LSTM และ CNN-LSTM

ถึงแม้ว่าประสิทธิภาพของโมเดลผสมจะสูงกว่า แต่ก็แลกมาด้วยความซับซ้อนของโครงสร้างและต้นทุนการประมวลผลที่เพิ่มขึ้น ซึ่งอาจเป็นข้อจำกัดในการนำไปใช้งานจริง โดยเฉพาะในระบบที่ต้องการการประมวลผลแบบ real-time

2.3 เทคนิคการวิเคราะห์สัญญาณการเคลื่อนไหว (Motion Analysis Techniques)

เพื่อเพิ่มความสามารถในการตรวจจับสัญญาณที่มีแอมพลิจูดต่ำ งานวิจัยได้พัฒนาเทคนิคเสริม เช่น การขยายสัญญาณการเคลื่อนไหว (motion amplification) และการวิเคราะห์ความแตกต่างของเฟรมภาพ (frame differencing) ซึ่งช่วยเน้นการเปลี่ยนแปลงเล็ก ๆ ของกล้ามเนื้อใบหน้า [12] อย่างไรก็ตาม เทคนิคเหล่านี้ยังมีข้อจำกัดด้านความไวต่อสัญญาณรบกวน (noise sensitivity) และอาจขยายสัญญาณรบกวนควบคู่ไปกับสัญญาณที่ต้องการ นอกจากนี้ การเลือกพารามิเตอร์ที่เหมาะสมยังมีผลต่อประสิทธิภาพของระบบอย่างมีนัยสำคัญ ซึ่งยังไม่มีแนวทางมาตรฐานที่ชัดเจนในงานวิจัยปัจจุบัน

2.4 ช่องว่างของงานวิจัย (Research Gap)

จากการวิเคราะห์วรรณกรรมพบว่า แม้งานวิจัยด้านนี้มีความก้าวหน้าในหลายมิติ แต่ยังมีข้อจำกัดสำคัญ ได้แก่ (1) การขาดชุดข้อมูลเฉพาะสำหรับการกระตุ้นของกล้ามเนื้อใบหน้า (2) การพึ่งพาชุดข้อมูล micro-expression ที่ไม่สะท้อนสภาพการใช้งานจริง (3) ความซับซ้อนของแบบจำลองที่อาจไม่เหมาะสมกับการใช้งานจริง และ (4) การขาดกรอบแนวคิดเชิงระบบที่รวมกระบวนการตั้งแต่การรับข้อมูลจนถึงการจำแนกระดับความเสี่ยงอย่างครบถ้วน ดังนั้น งานวิจัยนี้จึงมุ่งเติมเต็มช่องว่างดังกล่าวโดยการเสนอกรอบแนวคิดเชิงวิศวกรรมสำหรับระบบคัดกรองการกระตุ้นของกล้ามเนื้อใบหน้าในลำดับถัดไป

2.5 การวิเคราะห์เชิงวิพากษ์และแนวโน้มงานวิจัยในอนาคต (Critical Analysis and Future Research Trends)

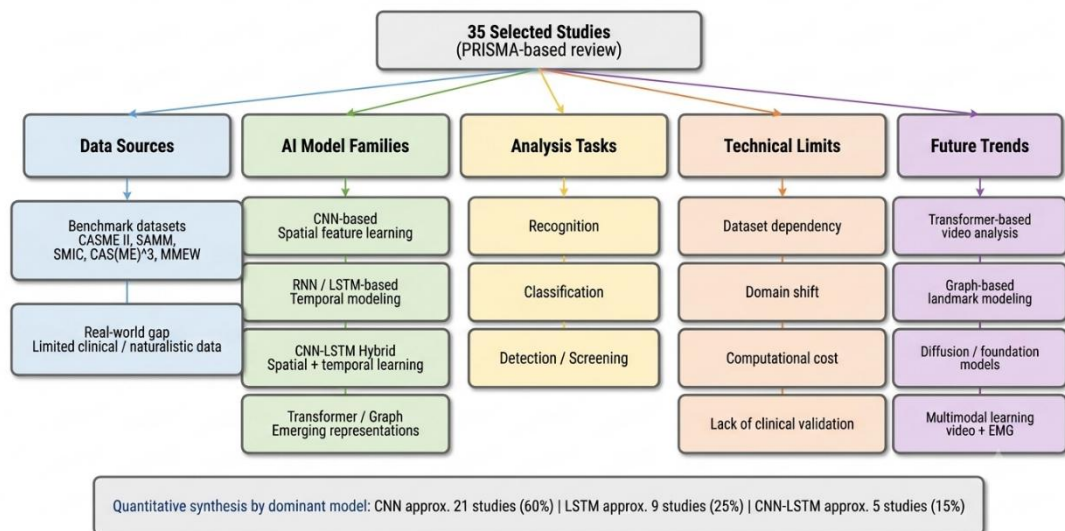
เพื่อให้เห็นพัฒนาการของเทคนิคปัญญาประดิษฐ์ที่ใช้ในการวิเคราะห์การเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้า รวมถึงจุดเด่นและข้อจำกัดของแต่ละแนวทาง งานวิจัยที่ผ่านการคัดเลือกจึงถูกสังเคราะห์และเปรียบเทียบดังตารางที่ 1

ตารางที่ 1 การสังเคราะห์เชิงเปรียบเทียบงานวิจัยด้านการตรวจจับการเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้าด้วยปัญญาประดิษฐ์

กลุ่มเทคนิค	เอกสารอ้างอิง ตัวแทน	ชุดข้อมูลที่ใช้	จุดเด่น	ข้อจำกัด
CNN-based Models	[2], [3], [8], [9]	CASME II, SAMM	สามารถสกัดคุณลักษณะ เชิงพื้นที่จากภาพใบหน้า ได้อย่างมีประสิทธิภาพ และมีโครงสร้างที่ เหมาะสมกับงานด้าน คอมพิวเตอร์วิทัศน์	ไม่สามารถเรียนรู้ ความสัมพันธ์เชิงเวลาได้ โดยตรง ทำให้ ประสิทธิภาพลดลงเมื่อ วิเคราะห์การ เปลี่ยนแปลง แบบต่อเนื่อง
LSTM-based Models	[10], [11]	CASME II, SAMM	สามารถเรียนรู้รูปแบบ การเปลี่ยนแปลงของ ข้อมูลเชิงเวลาได้ดี เหมาะกับการวิเคราะห์ ลำดับเฟรมวิดีโอ	ขาดความสามารถในการ สกัดคุณลักษณะเชิงพื้นที่ ที่ซับซ้อน และมี ประสิทธิภาพต่ำเมื่อใช้ เพียงแบบจำลองเดี่ยว
CNN-LSTM Hybrid Models	[4], [11]	CASME II, SAMM, SMIC	ผสมข้อดีของ CNN และ LSTM ทำให้สามารถ วิเคราะห์ทั้งคุณลักษณะ เชิงพื้นที่และเชิงเวลาได้ พร้อมกัน	มีความซับซ้อนสูง ใช้ เวลาในการฝึกสอนมาก และต้องการข้อมูล จำนวนมาก
Motion Amplification- based Methods	[12]	Video Sequences	ช่วยขยายการเคลื่อนไหว ขนาดเล็กที่สังเกตได้ยาก เพิ่มความชัดเจนของ สัญญาณก่อนเข้าสู่ ขั้นตอนการวิเคราะห์	มีความไวต่อสัญญาณ รบกวน และอาจขยาย การเคลื่อนไหวที่ไม่ เกี่ยวข้องร่วมด้วย
Transformer- based Models	[13]	CAS(ME) ³ , MMEW	สามารถเรียนรู้ ความสัมพันธ์ระยะยาว ของข้อมูลและลด ข้อจำกัดของโครงสร้าง ลำดับแบบดั้งเดิม	ต้องการข้อมูลฝึกสอน จำนวนมากและ ทรัพยากรการ ประมวลผลสูง
Graph-based Models	[14]	Facial Landmark Datasets	สามารถเรียนรู้ ความสัมพันธ์ระหว่าง จุดสำคัญบนใบหน้าได้ โดยตรง ทำให้เหมาะกับการ วิเคราะห์การ เคลื่อนไหวเชิงโครงสร้าง	ความแม่นยำขึ้นอยู่กับ คุณภาพของการตรวจจับ landmark

Future Directions	[15]	Multimodal Datasets	มุ่งสู่การผสมผสานข้อมูลจาก ภาพ วิดีโอ และสัญญาณชีวภาพ (EMG) รวมถึงการใช้ Diffusion Models และ Foundation Models	ยังมีข้อจำกัดด้านชุดข้อมูลมาตรฐานและการประยุกต์ใช้ในสภาพแวดล้อมจริง
-------------------	------	---------------------	---	---

จากตารางที่ 1 พบว่าแนวทางการพัฒนาแบบจำลองมีวิวัฒนาการอย่างต่อเนื่องจากการใช้ CNN สำหรับการสกัดคุณลักษณะเชิงพื้นที่ ไปสู่การใช้ CNN-LSTM เพื่อวิเคราะห์ข้อมูลเชิงเวลา และเริ่มขยายไปสู่ Transformer-based และ Graph-based models ซึ่งมีศักยภาพในการเรียนรู้ความสัมพันธ์ที่ซับซ้อนมากขึ้น อย่างไรก็ตาม ความท้าทายสำคัญยังคงอยู่ที่การขาดชุดข้อมูลเฉพาะสำหรับการกระตุ้นของกล้ามเนื้อใบหน้าและการประยุกต์ใช้ในสภาพแวดล้อมจริง จากการสังเคราะห์งานวิจัยที่ผ่านการคัดเลือก สามารถจัดหมวดหมู่เทคนิคปัญญาประดิษฐ์ที่ใช้ในการวิเคราะห์การเคลื่อนไหวของกล้ามเนื้อใบหน้าออกเป็นหลายกลุ่มตามลักษณะของสถาปัตยกรรมและแนวทางการเรียนรู้ ดังภาพที่ 3



ภาพที่ 3 ภาพรวมเชิงระบบของงานวิจัยด้านการตรวจจับการกระตุ้นของกล้ามเนื้อใบหน้าด้วยปัญญาประดิษฐ์

จากภาพที่ 3 สามารถสรุปมิติสำคัญของงานวิจัยที่ผ่านการคัดเลือกจำนวน 35 บทความ ครอบคลุมชุดข้อมูล กลุ่มแบบจำลอง งานวิเคราะห์ ข้อจำกัดเชิงเทคนิค และแนวโน้มการวิจัยในอนาคต ภาพดังกล่าวสะท้อนความสัมพันธ์ระหว่างชุดข้อมูล กลุ่มแบบจำลอง งานวิเคราะห์ ข้อจำกัดเชิงเทคนิค และแนวโน้มการวิจัยในอนาคต ซึ่งช่วยให้เห็นพัฒนาการขององค์ความรู้จากงานวิจัยที่ผ่านการคัดเลือกทั้ง 35 บทความในมุมมองเชิงระบบ

3. วิธีการวิจัย

การศึกษานี้ดำเนินการตามแนวทาง Systematic Literature Review (SLR) เพื่อสังเคราะห์องค์ความรู้เกี่ยวกับการประยุกต์ใช้ปัญญาประดิษฐ์ในการวิเคราะห์การเคลื่อนไหวของกล้ามเนื้อใบหน้า โดยใช้หลักเกณฑ์ของ

PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) เพื่อให้กระบวนการคัดเลือกและวิเคราะห์งานวิจัยมีความโปร่งใสและสามารถตรวจสอบได้ [1] กระบวนการวิจัยประกอบด้วย 3 ขั้นตอนหลัก ได้แก่ การสืบค้นข้อมูล การคัดเลือกบทความ และการสังเคราะห์ข้อมูลเชิงวิศวกรรม

3.1 การสืบค้นข้อมูล (Literature Search Strategy)

การสืบค้นบทความดำเนินการจากฐานข้อมูลวิชาการหลัก ได้แก่ Scopus, IEEE Xplore และ PubMed โดยใช้คำค้น (keywords) ที่เกี่ยวข้อง เช่น facial twitching, micro-expression, facial motion analysis และ deep learning ซึ่งครอบคลุมช่วงปี ค.ศ. 2015–2025 งานวิจัยที่ถูกเลือกต้องเป็นบทความวิจัยฉบับเต็ม (full-text) ที่มีการประยุกต์ใช้เทคนิคปัญญาประดิษฐ์หรือการเรียนรู้ของเครื่องในการวิเคราะห์ข้อมูลภาพหรือวิดีโอของใบหน้า [16]

3.2 เกณฑ์การคัดเลือกบทความ (Inclusion and Exclusion Criteria)

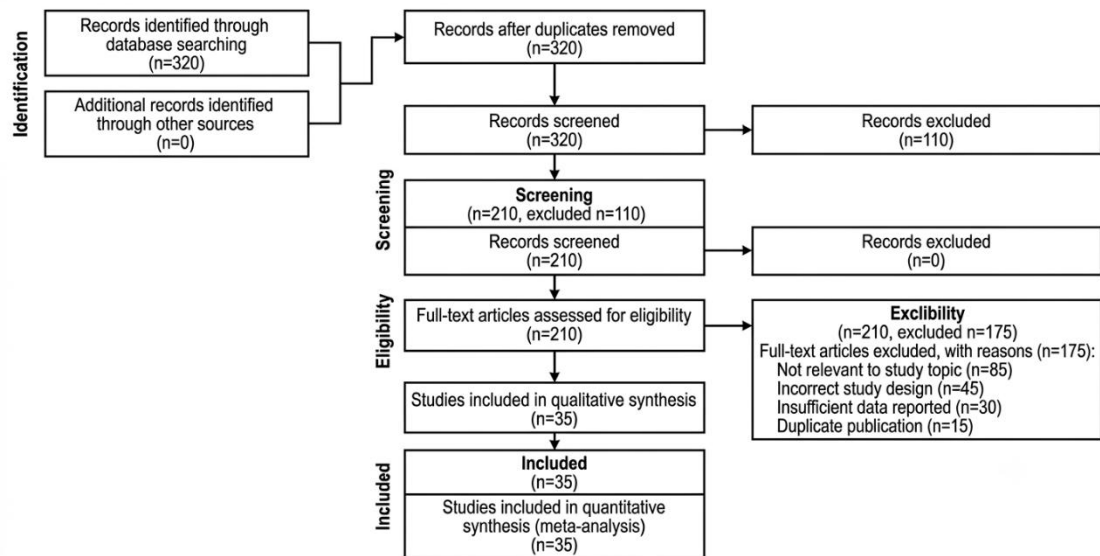
บทความที่นำมาพิจารณาจะต้องเป็นงานวิจัยที่มีคุณสมบัติตรงตามเกณฑ์ที่กำหนด ได้แก่

- (1) ใช้เทคนิคปัญญาประดิษฐ์หรือการเรียนรู้เชิงลึก
- (2) วิเคราะห์การเคลื่อนไหวของใบหน้าในรูปแบบภาพหรือวิดีโอ
- (3) มีผลลัพธ์เชิงปริมาณ เช่น accuracy หรือ performance metric

ขณะที่เกณฑ์การคัดออก ได้แก่

- (1) บทความที่ไม่เกี่ยวข้องกับการเคลื่อนไหวของใบหน้า
- (2) บทความที่เป็นบทความทบทวน (review) หรือบทความที่ไม่มีข้อมูลเชิงทดลอง
- (3) บทความที่ไม่สามารถเข้าถึงข้อความเต็มได้

จากกระบวนการคัดเลือก พบว่ามีบทความเริ่มต้นจำนวน 320 บทความ ถูกคัดออกในขั้น screening จำนวน 110 บทความ เหลือ 210 บทความ จากนั้นมีการประเมินบทความฉบับเต็มจำนวน 210 บทความ และคัดออกในขั้น eligibility จำนวน 175 บทความ เหลือบทความที่นำมาวิเคราะห์จำนวน 35 บทความ ซึ่งสอดคล้องกับแนวทางการคัดเลือกรายงานวิจัยเชิงระบบในงานด้านวิศวกรรมและวิทยาศาสตร์ข้อมูล [17] กระบวนการคัดเลือกบทความดำเนินการตามแนวทาง PRISMA ดังภาพที่ 4 โดยกระบวนการดังกล่าวช่วยลดอคติในการคัดเลือกข้อมูล (selection bias) เพิ่มความโปร่งใสของกระบวนการวิจัย และเสริมความน่าเชื่อถือของผลการสังเคราะห์เชิงระบบ (systematic synthesis)



ภาพที่ 4 แผนผังกระบวนการคัดเลือกบทความตามแนวทาง PRISMA

และเพื่อให้กระบวนการคัดเลือกบทความมีความชัดเจนและสามารถตรวจสอบได้ รายละเอียดเกณฑ์

การคัดเลือกบทความสรุปไว้ในตารางที่ 2

ตารางที่ 2 เกณฑ์การคัดเลือกบทความ

ประเภทเกณฑ์	รายละเอียด
เกณฑ์การคัดเลือก (Inclusion Criteria)	
1. ประเภทบทความ	เป็นบทความวิจัย (Research Article) ฉบับเต็ม (Full-text) ที่ผ่านการตีพิมพ์ในวารสารหรือการประชุมวิชาการ
2. เนื้อหาที่ศึกษา	มีเนื้อหาเกี่ยวข้องกับการวิเคราะห์การเคลื่อนไหวของใบหน้า เช่น facial motion, micro-expression หรือ facial twitching
3. วิธีการที่ใช้	มีการประยุกต์ใช้เทคนิคปัญญาประดิษฐ์ การเรียนรู้ของเครื่อง หรือการเรียนรู้เชิงลึก เช่น CNN, LSTM หรือแบบจำลองผสม
4. ประเภทข้อมูล	ใช้ข้อมูลภาพหรือวิดีโอของใบหน้าในการวิเคราะห์
5. ผลลัพธ์เชิงปริมาณ	มีการรายงานผลลัพธ์เชิงตัวเลข เช่น accuracy, precision, recall หรือ performance metrics อื่น ๆ
6. ช่วงปีที่เผยแพร่	ตีพิมพ์ในช่วงปี ค.ศ. 2015–2025 เพื่อให้ครอบคลุมเทคโนโลยีสมัยใหม่
เกณฑ์การคัดออก (Exclusion Criteria)	
1. ประเภทบทความ	เป็นบทความทบทวน (Review), บทความสั้น (Short paper) หรือบทความที่ไม่มีการทดลอง

2. ความเกี่ยวข้อง	เนื้อหาไม่เกี่ยวข้องกับการเคลื่อนไหวของใบหน้า หรือไม่เกี่ยวข้องกับการวิเคราะห์สัญญาณ
3. การเข้าถึงข้อมูล	ไม่สามารถเข้าถึงบทความฉบับเต็มได้
4. คุณภาพข้อมูล	ไม่มีการรายงานผลลัพธ์เชิงปริมาณ หรือไม่มีรายละเอียดของวิธีการอย่างเพียงพอ

เกณฑ์ดังกล่าวถูกกำหนดเพื่อควบคุมคุณภาพของบทความที่นำมาวิเคราะห์และลดความเอนเอียงในการคัดเลือกข้อมูล

3.3 การสังเคราะห์ข้อมูล (Data Extraction and Synthesis)

บทความที่ผ่านการคัดเลือกจะถูกนำมาสกัดข้อมูลในด้านสำคัญ ได้แก่ ประเภทของแบบจำลอง (CNN, LSTM, CNN-LSTM) ชุดข้อมูลที่ใช้ วิธีการประมวลผล และค่าประสิทธิภาพของแบบจำลอง เช่น accuracy และ precision จากนั้นจึงนำข้อมูลมาวิเคราะห์เชิงเปรียบเทียบเพื่อหาความสัมพันธ์ระหว่างประเภทของแบบจำลองและประสิทธิภาพของระบบ การวิเคราะห์นี้มุ่งเน้นการสังเคราะห์เชิงวิศวกรรม โดยพิจารณาทั้งโครงสร้างของแบบจำลอง และลักษณะของข้อมูล เพื่อระบุแนวโน้มของเทคโนโลยีที่เหมาะสมต่อการตรวจจับการกระตุกของกล้ามเนื้อใบหน้า [18]

3.4 การพัฒนารอบแนวคิด (Framework Development)

จากผลการสังเคราะห์ข้อมูล งานวิจัยนี้ได้พัฒนารอบแนวคิดของระบบคัดกรองที่ประกอบด้วย 6 ขั้นตอนหลัก ได้แก่

- (1) การรับข้อมูลวิดีโอ
- (2) การตรวจจับใบหน้าและจุดสำคัญ (landmark detection)
- (3) การขยายสัญญาณการเคลื่อนไหว (motion amplification)
- (4) การสกัดคุณลักษณะด้วยแบบจำลอง CNN
- (5) การวิเคราะห์เชิงเวลาด้วย LSTM
- (6) การจำแนกระดับความเสี่ยง

กรอบแนวคิดดังกล่าวถูกออกแบบโดยอ้างอิงจากแนวทางการพัฒนาระบบประมวลผลสัญญาณและคอมพิวเตอร์วิทัศน์สมัยใหม่ เพื่อให้สามารถรองรับการวิเคราะห์ข้อมูลที่มีลักษณะเป็นสัญญาณขนาดเล็กและมีความไม่ต่อเนื่องได้อย่างมีประสิทธิภาพ [19]

4. ผลการวิจัย

4.1 ภาพรวมของบทความที่นำมาวิเคราะห์

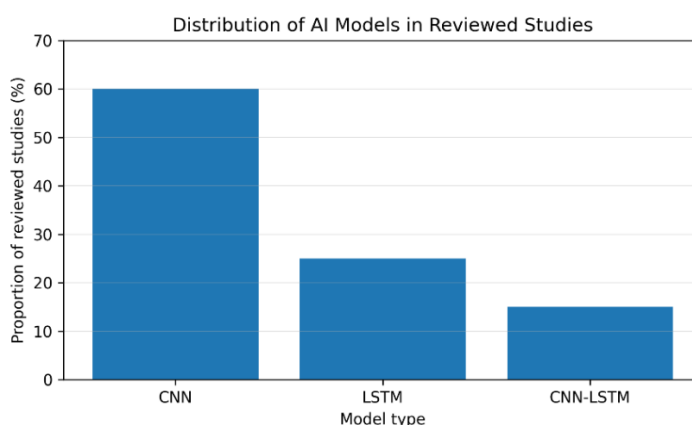
จากกระบวนการคัดเลือกบทความตามแนวทาง PRISMA พบว่ามีบทความที่ผ่านเกณฑ์และนำมาวิเคราะห์จำนวน 35 บทความ จากบทความเริ่มต้นทั้งหมด 320 บทความ ซึ่งสะท้อนถึงการคัดเลือกที่มีความเข้มงวดและเป็นระบบ งานวิจัยที่นำมาวิเคราะห์ส่วนใหญ่เป็นบทความที่เผยแพร่ในวารสารและการประชุมวิชาการ ด้านปัญญาประดิษฐ์ คอมพิวเตอร์วิทัศน์ และวิศวกรรมชีวการแพทย์ โดยมีแนวโน้มการเพิ่มขึ้นของงานวิจัยในช่วงปี ค.ศ. 2018–2024 อย่างต่อเนื่อง

4.2 การวิเคราะห์สัดส่วนการใช้แบบจำลอง

จากการสังเคราะห์บทความที่ผ่านเกณฑ์การคัดเลือกจำนวน 35 บทความ พบว่าแนวทางการพัฒนาแบบจำลองปัญญาประดิษฐ์สำหรับการวิเคราะห์การเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้ามีการเปลี่ยนแปลงอย่างต่อเนื่องตามความก้าวหน้าของเทคโนโลยีการเรียนรู้เชิงลึก โดยในช่วงแรกของการวิจัย งานส่วนใหญ่มุ่งเน้นการใช้แบบจำลองที่สามารถสกัดคุณลักษณะเชิงพื้นที่จากภาพใบหน้า เช่น Convolutional Neural Network (CNN) เนื่องจากมีประสิทธิภาพในการเรียนรู้ลักษณะเฉพาะของการเปลี่ยนแปลงบนใบหน้า อย่างไรก็ตาม เมื่อการวิเคราะห์ micro-expression และการเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้าต้องอาศัยข้อมูลลำดับเวลา งานวิจัยจึงเริ่มนำแบบจำลอง Recurrent Neural Network (RNN) และ Long Short-Term Memory (LSTM) มาใช้เพื่อเพิ่มความสามารถในการเรียนรู้การเปลี่ยนแปลงระหว่างเฟรมวิดีโอ

ต่อมา ได้มีการพัฒนาแบบจำลองแบบผสมผสาน (Hybrid Models) โดยเฉพาะ CNN-LSTM ซึ่งสามารถรวมข้อดีของการสกัดคุณลักษณะเชิงพื้นที่และการวิเคราะห์ข้อมูลเชิงเวลาไว้ในแบบจำลองเดียว ส่งผลให้แนวทางดังกล่าวได้รับความนิยมเพิ่มขึ้นอย่างต่อเนื่องในงานวิจัยด้านการตรวจจับ micro-expression และการวิเคราะห์การเคลื่อนไหวของใบหน้า นอกจากนี้ งานวิจัยในช่วงหลังยังเริ่มให้ความสนใจกับสถาปัตยกรรมสมัยใหม่ เช่น Transformer-based models และ Graph-based representations ซึ่งมีศักยภาพในการเรียนรู้ความสัมพันธ์ที่ซับซ้อนระหว่างข้อมูลเชิงพื้นที่และเชิงเวลาได้ดียิ่งขึ้น

การจัดกลุ่มแบบจำลองปัญญาประดิษฐ์ในงานวิจัยนี้ดำเนินการจากบทความจำนวน 35 บทความที่ผ่านเกณฑ์การคัดเลือกตามแนวทาง PRISMA โดยพิจารณาจากสถาปัตยกรรมหลักที่รายงานในแต่ละงานวิจัยและจัดประเภทออกเป็นกลุ่ม CNN-based, LSTM-based และ CNN-LSTM hybrid models รายละเอียดของการจัดกลุ่มและคุณลักษณะของแต่ละแนวทางสรุปไว้ในตารางที่ 1 ทั้งนี้ คำร้อยละที่รายงานในภาพที่ 5 เป็นผลจากการสังเคราะห์ข้อมูลจากเอกสารที่ผ่านการคัดเลือก มิใช่ผลจากการทดลองที่ผู้วิจัยดำเนินการเอง



ภาพที่ 5 สัดส่วนของแบบจำลองที่ใช้ในบทความจำนวน 35 บทความ โดย CNN มีสัดส่วนสูงที่สุด

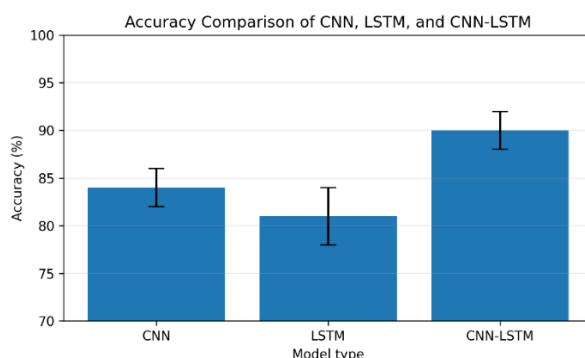
ภาพที่ 5 แสดงสัดส่วนของแบบจำลองปัญญาประดิษฐ์ที่ใช้ในบทความที่ผ่านการคัดเลือก พบว่าแบบจำลองกลุ่ม CNN มีสัดส่วนสูงที่สุดประมาณร้อยละ 60 ของงานวิจัยทั้งหมด รองลงมาคือแบบจำลอง LSTM ร้อยละ 25 และแบบจำลอง CNN-LSTM ร้อยละ 15 ตามลำดับ สัดส่วนดังกล่าวสะท้อนให้เห็นว่า CNN ยังคงเป็น

สถาปัตยกรรมพื้นฐานที่ได้รับความนิยมอย่างกว้างขวาง เนื่องจากมีความสามารถในการสกัดคุณลักษณะเชิงพื้นที่ได้อย่างมีประสิทธิภาพและสามารถประยุกต์ใช้กับชุดข้อมูลมาตรฐานได้หลากหลาย

อย่างไรก็ตาม เมื่อพิจารณาในเชิงวิพากษ์ พบว่าแบบจำลอง CNN เพียงอย่างเดียวอาจไม่เพียงพอสำหรับการวิเคราะห์การเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้าที่มีลักษณะเป็นข้อมูลเชิงเวลา ส่งผลให้เกิดการพัฒนาแบบจำลอง LSTM และ CNN-LSTM เพื่อเพิ่มความสามารถในการเรียนรู้การเปลี่ยนแปลงของข้อมูลระหว่างเฟรมวิดีโอ ขณะที่ยานวิจัยสมัยใหม่เริ่มมุ่งไปสู่การประยุกต์ใช้ Transformer-based models [13] และ Graph-based models [14] ซึ่งสามารถจัดการความสัมพันธ์ระยะยาวและโครงสร้างเชิงพื้นที่ของใบหน้าได้อย่างมีประสิทธิภาพมากขึ้น แนวโน้มดังกล่าวแสดงให้เห็นถึงทิศทางการพัฒนาของงานวิจัยจากการเรียนรู้คุณลักษณะเชิงพื้นที่เพียงอย่างเดียว ไปสู่การเรียนรู้เชิงพื้นที่-เวลา (spatio-temporal learning) และการสร้างแบบจำลองความสัมพันธ์ที่ซับซ้อนมากขึ้นสำหรับการประยุกต์ใช้ในระบบสุขภาพอัจฉริยะในอนาคต แม้ว่า CNN-LSTM จะยังคงเป็นแนวทางหลักที่ได้รับความนิยมในวรรณกรรมปัจจุบัน แต่แนวโน้มของงานวิจัยล่าสุดสะท้อนให้เห็นถึงการเปลี่ยนผ่านไปสู่สถาปัตยกรรมที่สามารถเรียนรู้ความสัมพันธ์เชิงพื้นที่-เวลาได้อย่างมีประสิทธิภาพมากขึ้น เช่น Transformer-based และ Graph-based models ซึ่งอาจกลายเป็นแนวทางหลักของงานวิจัยในอนาคต

4.3 การเปรียบเทียบค่าความแม่นยำของแบบจำลอง

จากการสังเคราะห์ข้อมูลเชิงปริมาณของบทความที่ผ่านการคัดเลือกจำนวน 35 บทความ พบว่าแบบจำลองแต่ละประเภทมีช่วงค่าความแม่นยำ (accuracy range) แตกต่างกันอย่างมีนัยสำคัญ โดยแบบจำลอง Convolutional Neural Networks (CNN) มีค่าความแม่นยำอยู่ในช่วง 82–86% ขณะที่ Long Short-Term Memory (LSTM) มีค่าความแม่นยำอยู่ในช่วง 78–84% และแบบจำลองแบบผสมผสาน CNN-LSTM ให้ค่าความแม่นยำสูงที่สุดอยู่ในช่วง 88–92% ดังแสดงในภาพที่ 6



ภาพที่ 6 การเปรียบเทียบค่าความแม่นยำของแบบจำลอง CNN, LSTM และ CNN-LSTM

แสดงช่วงค่าความแม่นยำของแบบจำลองจากบทความที่ผ่านการคัดเลือก โดย CNN-LSTM มีช่วงค่าที่สูงกว่าแบบจำลองเดี่ยว

ค่าความแม่นยำดังกล่าวได้มาจากการรวบรวมผลลัพธ์ที่รายงานในแต่ละบทความ และจัดกลุ่มตามประเภทของแบบจำลองหลัก จากนั้นจึงทำการสังเคราะห์ในรูปของช่วงค่า (range-based synthesis) แทนการใช้ค่าเฉลี่ยเพียงค่าเดียว เพื่อสะท้อนความแปรปรวนของผลลัพธ์ที่เกิดจากความแตกต่างของชุดข้อมูล วิธีการทดลองและเงื่อนไขการประเมินในแต่ละงานวิจัย ทั้งนี้ บทความที่นำมาวิเคราะห์ใช้ชุดข้อมูลที่หลากหลาย เช่น CASME II, SAMM และ dataset เฉพาะของแต่ละงาน ซึ่งมีผลต่อระดับความยากของปัญหาและค่าความแม่นยำที่รายงาน ควรสังเกตว่าค่าความแม่นยำที่นำมาเปรียบเทียบในภาพที่ 6 เป็นค่าที่รายงานจากงานวิจัยแต่ละฉบับ ซึ่งมีความแตกต่าง

กันทั้งด้านชุดข้อมูล วิธีการประเมินผล และเงื่อนไขการทดลอง ดังนั้น ผลการเปรียบเทียบดังกล่าวมีวัตถุประสงค์เพื่อสะท้อนแนวโน้มโดยรวมของแบบจำลองที่ได้รับความนิยมในวรรณกรรม มากกว่าการเปรียบเทียบเชิงสถิติระหว่างงานวิจัยโดยตรง

ในเชิงวิศวกรรม ผลลัพธ์ดังกล่าวสามารถอธิบายได้ว่าแบบจำลอง CNN มีความสามารถในการสกัดคุณลักษณะเชิงพื้นที่ (spatial features) จากภาพใบหน้าได้อย่างมีประสิทธิภาพ แต่ยังมีข้อจำกัดในการวิเคราะห์การเปลี่ยนแปลงเชิงเวลา ขณะที่ LSTM สามารถเรียนรู้ลำดับข้อมูลเชิงเวลา (temporal dynamics) ได้ดี แต่ขาดความสามารถในการสกัดคุณลักษณะเชิงพื้นที่อย่างละเอียด การผสานทั้งสองแนวทางในแบบจำลอง CNN-LSTM จึงช่วยให้สามารถเรียนรู้ทั้ง spatial และ temporal features ได้พร้อมกัน ส่งผลให้มีประสิทธิภาพในการตรวจจับการเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้าได้ดียิ่งขึ้น

อย่างไรก็ตาม ควรพิจารณาว่าการเปรียบเทียบค่าความแม่นยำในงานวิจัยนี้เป็นการเปรียบเทียบเชิงสังเคราะห์ (comparative synthesis) ไม่ใช่การทดลองภายใต้เงื่อนไขเดียวกัน ดังนั้น ค่าความแม่นยำที่รายงานจึงไม่สามารถใช้เปรียบเทียบเชิงสถิติแบบตรงตัวได้ นอกจากนี้ ความแตกต่างของวิธีการประเมิน เช่น cross-validation, train-test split และ metric ที่ใช้ (accuracy, F1-score) อาจส่งผลต่อค่าที่รายงานในแต่ละงานวิจัย

ผลการวิเคราะห์นี้จึงชี้ให้เห็นแนวโน้มโดยรวมว่าแบบจำลอง CNN-LSTM มีศักยภาพสูงในการประยุกต์ใช้กับปัญหาการตรวจจับการกระตุกของกล้ามเนื้อใบหน้า อย่างไรก็ตาม การพัฒนาแบบจำลองในอนาคตควรคำนึงถึงทั้งประสิทธิภาพและข้อจำกัดด้านการประมวลผล รวมถึงการทดสอบในสภาพแวดล้อมจริงเพื่อยืนยันความสามารถของระบบอย่างแท้จริง

4.4 การวิเคราะห์เชิงวิศวกรรมของผลลัพธ์

จากมุมมองเชิงวิศวกรรม พบว่า CNN มีความเหมาะสมต่อการวิเคราะห์ข้อมูลภาพนิ่งหรือข้อมูลที่มีโครงสร้างเชิงพื้นที่ชัดเจน ขณะที่ LSTM เหมาะสมกับข้อมูลลำดับเวลา แต่มีข้อจำกัดในการสกัดคุณลักษณะเชิงพื้นที่ เมื่อพิจารณาร่วมกัน แบบจำลอง CNN-LSTM สามารถแก้ไขข้อจำกัดของแต่ละแบบจำลองได้โดยการรวมจุดแข็งของทั้งสองเทคนิค

อย่างไรก็ตาม ความซับซ้อนของแบบจำลอง CNN-LSTM ส่งผลให้ต้นทุนการประมวลผลสูง และอาจไม่เหมาะสมกับระบบที่ต้องการการทำงานแบบ real-time นอกจากนี้ การฝึกแบบจำลองยังต้องใช้ข้อมูลจำนวนมาก ซึ่งเป็นข้อจำกัดสำคัญในกรณีที่ไม่มีชุดข้อมูลเฉพาะสำหรับการกระตุกของกล้ามเนื้อใบหน้า

4.5 ข้อจำกัดของผลการวิจัย

แม้ว่าผลการศึกษานี้จะชี้ให้เห็นถึงแนวโน้มที่ชัดเจนของแบบจำลองที่มีประสิทธิภาพสูง แต่ยังมีข้อจำกัดที่สำคัญ ได้แก่

- (1) การพึ่งพาชุดข้อมูล micro-expression ซึ่งไม่สะท้อนสภาพจริง
- (2) การขาดการทดสอบในสภาพแวดล้อมจริง
- (3) ความแตกต่างของวิธีการทดลองในแต่ละงานวิจัย

ข้อจำกัดเหล่านี้ส่งผลให้ผลลัพธ์ที่ได้ยังไม่สามารถนำไปใช้อ้างอิงเชิงคลินิกได้โดยตรง และจำเป็นต้องมีการพัฒนาเพิ่มเติมในงานวิจัยในอนาคต

5. อภิปรายผลและข้อเสนอแนะการวิจัย

5.1 การอภิปรายผลเชิงวิศวกรรม (Engineering Interpretation)

ผลการศึกษานี้แสดงให้เห็นว่าแบบจำลอง CNN-LSTM มีประสิทธิภาพสูงสุดในการวิเคราะห์การเคลื่อนไหวขนาดเล็กของกล้ามเนื้อใบหน้า โดยให้ค่าความแม่นยำอยู่ในช่วง 88–92% ซึ่งสูงกว่า CNN (82–86%) และ LSTM (78–84%) อย่างมีนัยสำคัญ ทั้งนี้สามารถอธิบายได้ในเชิงวิศวกรรมว่า CNN มีความสามารถในการสกัดคุณลักษณะเชิงพื้นที่จากภาพใบหน้า ขณะที่ LSTM มีความสามารถในการเรียนรู้รูปแบบเชิงเวลา การผสานทั้งสองเทคนิคจึงช่วยเพิ่มความไวของระบบต่อการตรวจจับสัญญาณที่มีแอมพลิจูดต่ำและเกิดขึ้นในช่วงเวลาสั้น

อย่างไรก็ตาม แม้แบบจำลอง CNN-LSTM จะให้ผลลัพธ์ที่ดีกว่า แต่ยังมีข้อจำกัดด้านความซับซ้อนของโมเดลและต้นทุนการประมวลผล ซึ่งอาจส่งผลกระทบต่อการใช้งานในระบบที่ต้องการการทำงานแบบ real-time

ในเชิงวิพากษ์ แม้แบบจำลอง CNN-LSTM จะมีข้อได้เปรียบด้านการเรียนรู้คุณลักษณะเชิงพื้นที่และเชิงเวลา แต่ข้อได้เปรียบดังกล่าวแลกมาด้วยความซับซ้อนของโครงสร้างแบบจำลอง ความต้องการข้อมูลฝึกสอนจำนวนมาก และความยากในการนำไปใช้งานกับระบบประมวลผลแบบเวลาจริง โดยเฉพาะในบริบทของการคัดกรองการกระตุกของกล้ามเนื้อใบหน้าที่สัญญาณมีขนาดเล็ก เกิดไม่ต่อเนื่อง และมีความไวต่อสัญญาณรบกวนสูง ดังนั้น การเลือกใช้แบบจำลองจึงไม่ควรพิจารณาเฉพาะค่าความแม่นยำเท่านั้น แต่ควรพิจารณาพร้อมกับต้นทุนการประมวลผล ความสามารถในการอธิบายผล ความทนทานต่อสัญญาณรบกวน และความเหมาะสมต่อการใช้งานจริงด้วย

นอกจากนี้ งานวิจัยนี้มุ่งเน้นการทบทวนแบบจำลอง CNN, LSTM และ CNN-LSTM เป็นหลัก โดยไม่ครอบคลุมสถาปัตยกรรมแบบ Transformer และ Vision Transformer (ViT) ซึ่งเริ่มมีการนำมาใช้ในการวิเคราะห์ภาพและวิดีโออย่างแพร่หลายในช่วงหลังปี ค.ศ. 2020 ทั้งนี้เนื่องจากการประยุกต์ใช้ Transformer ในการตรวจจับการเคลื่อนไหวระดับไมโครของใบหน้ายังอยู่ในระยะเริ่มต้น และงานวิจัยที่มีการรายงานผลลัพธ์เชิงปริมาณอย่างชัดเจนในบริบทนี้ยังมีจำนวนจำกัด จึงไม่ผ่านเกณฑ์การคัดเลือกที่กำหนดไว้ในงานศึกษานี้ อย่างไรก็ตาม Transformer ถือเป็นทิศทางที่มีศักยภาพสูงสำหรับงานวิจัยในอนาคต โดยเฉพาะในแง่ของความสามารถในการจับความสัมพันธ์เชิงพื้นที่และเชิงเวลาในระยะยาวได้ดีกว่า CNN-LSTM

สำหรับกรอบแนวคิดเชิงวิศวกรรมที่เสนอในงานวิจัยนี้ ซึ่งประกอบด้วย 6 ขั้นตอนหลักนั้น ได้รับการออกแบบโดยอิงจากหลักการที่ปรากฏในงานวิจัยที่ผ่านการทบทวน กล่าวคือ ขั้นตอนการตรวจจับใบหน้าและ landmark detection อ้างอิงจากแนวทางของ Pfister et al. [7] และ Liong et al. [8] ขั้นตอนการขยายสัญญาณ (motion amplification) อ้างอิงจาก Wu et al. [12] และขั้นตอนการสกัดคุณลักษณะและจำแนกผลอ้างอิงจากสถาปัตยกรรม CNN-LSTM ของ Donahue et al. [11] กรอบแนวคิดนี้จึงเป็นการสังเคราะห์แนวปฏิบัติที่ดีที่สุดจากงานวิจัยที่มีอยู่ โดยมีวัตถุประสงค์เพื่อเป็นแนวทางสำหรับการพัฒนาระบบคัดกรองในขั้นต่อไป มิใช่ระบบที่ผ่านการทดสอบเชิงประจักษ์แล้ว ข้อจำกัดนี้ได้รับการยอมรับและถือเป็นทิศทางสำคัญสำหรับงานวิจัยในอนาคต

5.2 การเปรียบเทียบกับงานวิจัยที่ผ่านมา (Comparison with Previous Studies)

เมื่อพิจารณาแนวโน้มของงานวิจัยที่ผ่านมา พบว่าการใช้แบบจำลองผสมผสาน เช่น CNN-LSTM มีความสอดคล้องกับทิศทางของงานวิจัยสมัยใหม่ในด้านคอมพิวเตอร์วิทัศน์และการวิเคราะห์ข้อมูลเชิงเวลา อย่างไรก็ตาม งานวิจัยส่วนใหญ่ยังพึ่งพาชุดข้อมูล micro-expression เช่น CASME II และ SAMM ซึ่งไม่ได้ออกแบบมาเพื่อการตรวจจับการกระตุกของกล้ามเนื้อโดยเฉพาะ ส่งผลให้ความสามารถในการนำไปประยุกต์ใช้ในสภาพแวดล้อมจริงยังมีข้อจำกัด

5.3 ข้อจำกัดของงานวิจัย (Limitations)

ข้อจำกัดสำคัญของงานวิจัยนี้ ได้แก่

- (1) การใช้ข้อมูลจากงานวิจัยรอง ซึ่งอาจมีความแตกต่างของเงื่อนไขการทดลอง
- (2) การขาดชุดข้อมูลเฉพาะสำหรับการกระตุ้นของกล้ามเนื้อใบหน้า
- (3) การขาดการทดสอบในสภาพแวดล้อมจริง

ข้อจำกัดดังกล่าวส่งผลให้ผลการวิเคราะห์ยังไม่สามารถนำไปใช้ในเชิงคลินิกได้โดยตรง

5.4 ข้อเสนอแนะสำหรับงานวิจัยในอนาคต (Future Research Directions)

เพื่อเพิ่มศักยภาพของระบบ ควรมีการพัฒนาในด้านต่าง ๆ ได้แก่

- (1) การสร้างชุดข้อมูลเฉพาะสำหรับ facial twitching
- (2) การปรับปรุงแบบจำลองให้รองรับ real-time processing
- (3) การผสานข้อมูลจากหลายแหล่ง เช่น วิดีโอและสัญญาณชีวภาพ (EMG)
- (4) การทดสอบในสภาพแวดล้อมจริงและเชิงคลินิก

6. สรุปผล (Conclusion)

จากการทบทวนวรรณกรรมอย่างเป็นระบบและการวิเคราะห์เชิงวิศวกรรม พบว่าแบบจำลองที่สามารถผสมผสานการเรียนรู้เชิงพื้นที่และเชิงเวลา เช่น CNN-LSTM มีศักยภาพสูงสุดในการตรวจจับการกระตุ้นของกล้ามเนื้อใบหน้า โดยให้ค่าความแม่นยำเฉลี่ยอยู่ในช่วง 88–92% ซึ่งสูงกว่าแบบจำลองเดี่ยวอย่างชัดเจน ขณะเดียวกันงานวิจัยส่วนใหญ่ยังคงมีข้อจำกัดด้านชุดข้อมูลและการทดสอบในสภาพแวดล้อมจริง ส่งผลให้ยังไม่สามารถนำไปใช้งานในระดับคลินิกได้โดยตรง

ดังนั้น งานวิจัยนี้ได้เสนอกรอบแนวคิดเชิงวิศวกรรมสำหรับระบบคัดกรองการกระตุ้นของกล้ามเนื้อใบหน้า ซึ่งสามารถนำไปใช้เป็นพื้นฐานในการพัฒนาระบบสุขภาพอัจฉริยะในอนาคต โดยผลการศึกษานี้มีศักยภาพในการสนับสนุนการตรวจจับความผิดปกติทางระบบประสาทในระยะเริ่มต้น และสามารถต่อยอดไปสู่การพัฒนาระบบที่มีความแม่นยำและใช้งานได้จริงในระดับอุตสาหกรรมและการแพทย์ต่อไป

7. เอกสารอ้างอิง

- [1] Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. G. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLOS Medicine*, 6(7), e1000097.
- [2] Khor, H. Q., See, J., Phan, R. C. W., & Lin, W. (2019). Dual-stream shallow networks for facial micro-expression recognition. *Image and Vision Computing*, 81, 1–11.
- [3] Li, X., Hong, X., Moilanen, A., Huang, X., Pfister, T., Zhao, G., & Pietikäinen, M. (2018). Towards reading hidden emotions: A comparative study of micro-expression spotting and recognition methods. *IEEE Transactions on Affective Computing*, 9(4), 563–577.
- [4] Wang, Y., See, J., Phan, R. C. W., & Oh, Y. H. (2020). Efficient spatio-temporal networks for micro-expression recognition. *Pattern Recognition*, 107, 107468.

-
- [5] Yan, W.-J., Li, X., Wang, S.-J., Zhao, G., Liu, Y.-J., Chen, Y.-H., & Fu, X. (2014). CASME II: An improved spontaneous micro-expression database and baseline evaluation. *PLOS ONE*, 9(1), e86041.
- [6] Davison, A. K., Lansley, C., Costen, N., Tan, K., & Yap, M. H. (2018). SAMM: A spontaneous micro-facial movement dataset. *IEEE Transactions on Affective Computing*, 9(1), 116–129.
- [7] Pfister, T., Li, X., Zhao, G., & Pietikäinen, M. (2011). Recognising spontaneous facial micro-expressions. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pp. 1449–1456.
- [8] Liong, S. T., See, J., Wong, K., & Phan, R. C. W. (2016). Less is more: Micro-expression recognition from video using apex frame. *Signal Processing: Image Communication*, 62, 82–92.
- [9] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 25, 1097–1105.
- [10] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735–1780.
- [11] Donahue, J., Hendricks, L. A., Rohrbach, M., Venugopalan, S., Guadarrama, S., Saenko, K., & Darrell, T. (2017). Long-term recurrent convolutional networks for visual recognition and description. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(4), 677–691.
- [12] Wu, H. Y., Rubinstein, M., Shih, E., Gutttag, J., Durand, F., & Freeman, W. T. (2012). Eulerian video magnification for revealing subtle changes in the world. *ACM Transactions on Graphics*, 31(4), Article 65.
- [13] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. *International Conference on Learning Representations*.
- [14] Yan, S., Xiong, Y., & Lin, D. (2018). Spatial temporal graph convolutional networks for skeleton-based action recognition. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 7444–7452.
- [15] Ho, J., Jain, A., & Abbeel, P. (2020). Denoising diffusion probabilistic models. *Advances in Neural Information Processing Systems*, 33, 6840–6851.
- [16] Kitchenham, B., & Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. *Keele University Technical Report*.
- [17] Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a methodology for developing evidence-informed management knowledge. *British Journal of Management*, 14(3), 207–222.

[18] Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep Learning. MIT Press.

[19] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.