

Received: August 17, 2023; Revised: November 3, 2023; Accepted: November 13, 2023

การตรวจจับท่าทางมือเพื่อเป็นสัญญาณขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์
โดยใช้วิธีการระบุพิกัดของมือด้วยมีเดียไปป์
Utilizing hand landmarks and MediaPipe for deep neural network-based
recognition of signal for help

ณรงค์ฤทธิ์ ภิรมย์นก^{1*}
Narongrit Piromnok^{1*}

¹ คณะวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสวนดุสิต กรุงเทพฯ

¹ Faculty of Science and Technology, Suan Dusit University, Bangkok

*Corresponding Author E-mail Address: narongrit_phi@dusit.ac.th

บทคัดย่อ

ปัญหาความไม่ปลอดภัยเป็นปัญหาที่ส่งผลกระทบต่อคุณภาพชีวิตของคนในสังคมเป็นอย่างมาก โดยเฉพาะผู้พิการทางการได้ยินที่มีจำนวนมากเป็นอันดับ 2 ของผู้พิการทั้งหมด ซึ่งเกิดขึ้นได้ทั้งทางร่างกายและจิตใจ และปัญหาความรุนแรงภายในครอบครัวที่อาจเกิดขึ้นได้ทั้งภายในบ้าน โรงเรียน หรือที่สาธารณะ ซึ่งเป็นปัญหาสำหรับบุคคลทุกกลุ่ม ทุกช่วงวัย จึงเป็นที่มาของการพัฒนาระบบการตรวจจับท่าทางมือเพื่อเป็นสัญญาณขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์โดยใช้วิธีการระบุพิกัดของมือด้วยมีเดียไปป์ โดยช่วยให้พวกเขาสามารถขอความช่วยเหลือในกรณีเกิดเหตุฉุกเฉินได้โดยไม่ต้องพูดหรือส่งเสียงออกมา โดยมีวัตถุประสงค์เพื่อ 1) เพื่อตรวจจับท่าทางมือที่เป็นสัญญาณขอความช่วยเหลือและแจ้งเตือน 2) เพื่อศึกษาประสิทธิภาพของการตรวจจับท่าทางมือขอความช่วยเหลือ โดยจะใช้ MediaPipe framework ซึ่งได้รับการฝึกฝนจากข้อมูล 30,000 ภาพ และทดสอบโดยการรับข้อมูลจากกล้องวิดีโอมาทำการตรวจจับท่าทางมือขอความช่วยเหลือ จำนวน 6 ท่าทาง จำนวนท่าทางละ 100 ภาพ รวมเป็น 600 ภาพ โดยการกำหนดค่าพิกัดของฝ่ามือและนิ้วมือลงในโปรแกรมภาษา Python เมื่อตรวจพบท่าทางมือขอความช่วยเหลือจะส่งการแจ้งเตือนไปยังผู้ดูแล จากผลการทดสอบ พบว่า ค่าความถูกต้องของท่าทางมือ 1-6 ของชุดข้อมูลอยู่ในช่วง 95-99% ความแม่นยำประมาณ 97%

คำสำคัญ: การรู้จำท่าทางมือ โค้ดข่ายประสาทเทียม มีเดียไปป์

Abstract

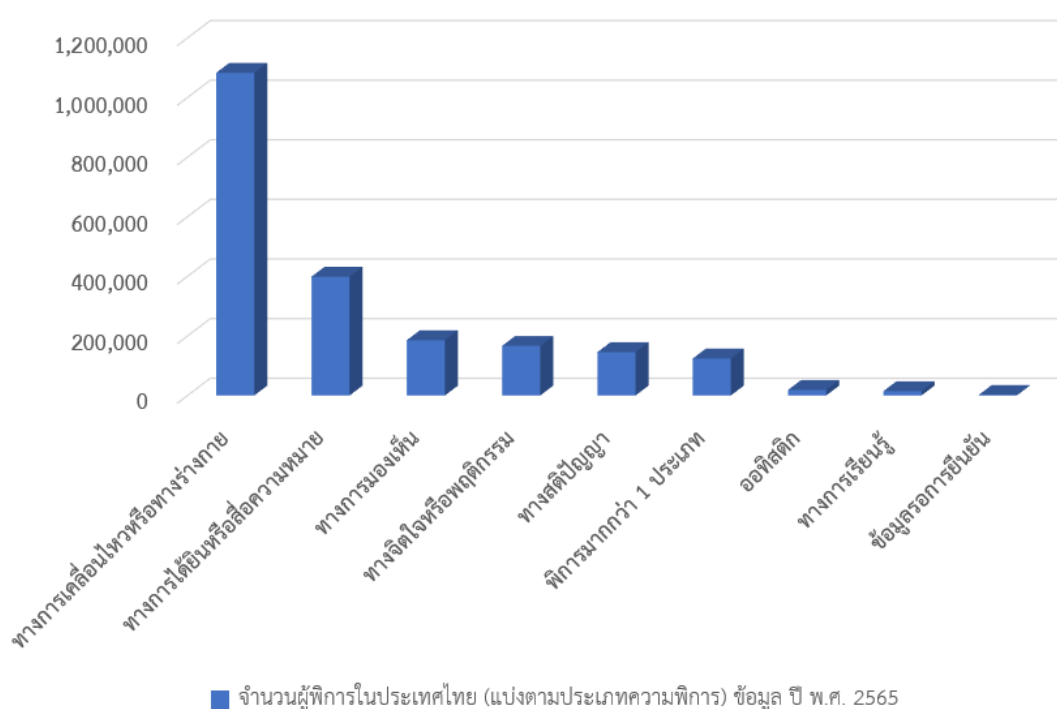
The unsafe condition is a problem that has greatly affected the quality of life of people in society, especially those people with hearing loss which accounts for the second largest number of all disabled people. The unsafe conditions may take place physically and mentally, as well as problems of family violence that may occur at home, school, or public places, thereby vastly affected people of all groups and all ages. Thus, the issue is attributed to the real-time “signal for help” hand gesture recognition development with the aid of the MediaPipe to track the location. The victim can ask for help in an

emergency case without words or shouting out. The objectives of the study include: 1) to detect the “signal for help” hand gestures and provide notifications, and 2) to examine the performance of the “signal for help” hand gesture recognition. The MediaPipe framework is used and practiced with 30,000 images dataset and tested by the obtained video data to detect six “signal for help” hand gestures, each gesture of 100 images, a total of 600 images. The palm and finger coordinates are configured into a Python programming language. When “signal for help” hand gestures are detected, a notification is provided to the administrator. The test results found that the accuracy of hand gestures 1-6 of the dataset ranged between 95-99%, and the accuracy was approximately 97%.

Keywords: Hand gesture recognition, Neural networks, MediaPipe

บทนำ

จากข้อมูลรายงานสถานการณ์ด้านคนพิการในประเทศไทย ประจำปี พ.ศ. 2565 โดยกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ ได้ให้ข้อมูลเกี่ยวกับคนพิการไว้ว่ามีคนพิการที่ได้รับการออกบัตรประจำตัวคนพิการในประเทศไทย จำนวนทั้งสิ้น 2,180,178 คน โดยคิดเป็นร้อยละ 3.29 ของประชากรทั้งประเทศ โดยข้อมูลประชากรประเทศไทยของสำนักทะเบียนกลาง กรมการปกครอง ณ วันที่ 2 กรกฎาคม 2566 จำนวนประชากรทั้งประเทศมีจำนวน 66,073,904 คน โดยแบ่งเป็นคนพิการ เพศชาย จำนวน 1,136,836 คน คิดเป็นร้อยละ 52.14 และเพศหญิง จำนวน 1,043,432 คน คิดเป็นร้อยละ 47.85 จากคนพิการทั้งหมด ซึ่งหากแบ่งตามประเภทของความพิการ จะสามารถจัดลำดับได้ดังรูปที่ 1 (กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ, 2565)



รูปที่ 1 ข้อมูลสถิติคนพิการในประเทศไทย ปี พ.ศ. 2565 แบ่งตามประเภทความพิการ

จากข้อมูลดังกล่าวจะเห็นได้ว่าจำนวนผู้พิการอันดับที่ 1 คือ ทาง การ เคลื่อน ไหว หรือ ทาง ร่าง กาย จำนวน 1,112,763 คน อันดับที่ 2 ผู้พิการทางการได้ยิน จำนวน 405,920 คน จากผู้พิการทั้งหมด 2,180,178 คน คิดเป็นร้อยละ 18.61 ซึ่งเป็นสัดส่วนที่สูงมาก เราจึงควรให้ความสำคัญในการพัฒนาสิ่งสนับสนุนหรือระบบที่ช่วยให้คุณภาพชีวิตของผู้พิการดีขึ้น ปัญหาของความบกพร่องทางการได้ยินทำให้เกิดปัญหาเกี่ยวกับพัฒนาการด้านภาษาและการสื่อสารที่ถือว่าเป็นส่วนสำคัญในพัฒนาการทางสังคม อารมณ์ และสติปัญญา นอกจากนี้ยังพบว่าผู้ที่มีความบกพร่องทางการได้ยินมีแนวโน้มในการต้องอยู่อย่างโดดเดี่ยว และมักจะตกเป็นผู้ถูกรังแกหรือใช้ความรุนแรง มีผลกับการเรียนและการทำงาน ปัญหาความไม่ปลอดภัยเป็นปัญหาที่ส่งผลต่อคุณภาพชีวิตของคนในสังคมเป็นอย่างมาก ซึ่งเกิดขึ้นได้ทั้งทางร่างกายและจิตใจ เช่น การถูกลั่นทรวง การทำร้ายร่างกาย การกักขัง หน่วงเหนี่ยว การข่มขืน การถูกล่วงละเมิดทางเพศ และปัญหาความรุนแรงภายในครอบครัวที่อาจเกิดขึ้นได้ทั้งภายในบ้าน โรงเรียน ห้างสรรพสินค้า หรือที่สาธารณะ ซึ่งเป็นปัญหาสำหรับบุคคลทุกกลุ่ม ทุกช่วงวัย ไม่ว่าจะเป็นผู้พิการ หรือบุคคลปกติ จึงเป็นที่มาของการคิดค้นสัญญาณมือขอความช่วยเหลือสากล (Signal for help) ที่ได้ถูกพัฒนาขึ้นโดย มูลนิธิสตรีแคนาดา เป็นสัญลักษณ์ที่ถูกเผยแพร่ในปี พ.ศ. 2563 โดยมูลนิธิสตรีแห่งแคนาดา จากประเทศแคนาดา (Canadian women's foundation) เพื่อเป็นสัญญาณมือในการขอความช่วยเหลือ จากการถูกทำร้ายร่างกาย จิตใจ หรือการถูกคุกคามต่าง ๆ การส่งสัญญาณโดยใช้มือเพียงข้างเดียว โดยเป็นการส่งสัญญาณแสดงท่าที่ว่าการขอความช่วยเหลือ และต้องการความช่วยเหลือโดยด่วน ซึ่งในการใช้ท่าทางมือนี้จะเป็นการส่งสัญญาณอย่างลับ ๆ เพื่อให้คนที่ผ่านไปผ่านมาเห็นท่าทางมือที่แสดงออกมาผ่านทางมุมมองเห็นแผนการส่งเสียงร้องขอความช่วยเหลือ ซึ่งบางครั้งคนที่ถูกทำร้ายหรือถูกคุกคาม อาจไม่สามารถที่จะขอความช่วยเหลือผ่านทาง การส่งเสียงร้องขอความช่วยเหลือได้ หรือไม่มีโอกาสที่จะขอความช่วยเหลือ เพราะมักจะถูกจับตามองอยู่ตลอดเวลา สัญญาณขอความช่วยเหลือไม่ได้เกี่ยวข้องกับภาษามืออเมริกัน (ASL) หรือภาษามืออื่น ๆ โดยท่าทางสัญญาณมือนี้เป็นเพียงท่าทางที่แสดงขึ้นเพื่อขอความช่วยเหลือ ที่มูลนิธิสตรีแห่งแคนาดาต้องการเผยแพร่ให้ทุกคนเข้าใจเท่านั้น จึงเป็นการแสดงท่าทางมือแบบง่าย ๆ ไม่ซับซ้อนแต่อย่างใด (The Canadian Women's Foundation, 2020)

ทั้งนี้ความบกพร่องทางการได้ยินเป็นความผิดปกติที่สามารถนำเอาเทคโนโลยีมาช่วยเหลือให้สามารถสื่อสารกับคนปกติได้สะดวกยิ่งขึ้น เนื่องจากในปัจจุบันเทคโนโลยีที่เกี่ยวข้องกับการมีปฏิสัมพันธ์ระหว่างมนุษย์กับคอมพิวเตอร์ ได้เพิ่มขึ้นอย่างแพร่หลายและรวดเร็ว และวิธีการที่มนุษย์กับคอมพิวเตอร์จะมีปฏิสัมพันธ์โต้ตอบกันได้สามารถทำได้หลายรูปแบบ เช่น การโต้ตอบด้วยเสียง และการโต้ตอบด้วยท่าทาง เนื่องจากท่าทางประกอบด้วยข้อมูลจำนวนมากที่สามารถสื่อความหมาย อารมณ์ และสอดคล้องกับพฤติกรรมในชีวิตประจำวันของมนุษย์ ท่าทางสามารถอยู่ในรูปของท่าทางมือที่ช่วยให้เราสามารถสื่อสารและโต้ตอบกันในชีวิตประจำวันได้ การรู้จำท่าทางมือ คือ การตีความทางคณิตศาสตร์ของการเคลื่อนไหวของมนุษย์โดยใช้อุปกรณ์คอมพิวเตอร์ในตรวจจับท่าทางหรือสัญญาณต่าง ๆ โดยใช้กล้องในการตรวจจับเป็นภาพนิ่งและภาพเคลื่อนไหว โดยการรู้จำท่าทางมือสามารถช่วยให้ผู้พิการทางร่างกาย โดยเฉพาะผู้พิการทางการได้ยิน หรือแม้แต่บุคคลทั่วไปที่ไม่สามารถสื่อสารด้วยเสียงได้ให้สามารถสื่อสารออกมาได้เมื่อตกอยู่ในอันตราย โดยผู้ทำการวิจัยได้พัฒนาระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ โดยการใช้วิธีการระบุพิกัดของมือด้วยมือเดียวไปป์โดยมุ่งเน้นไปที่การตรวจจับผ่านระบบกล้องวิดีโอ หรือกล้องวงจรปิด ที่มีการติดตั้งอยู่ทั่วไป ทั้งที่บ้าน ที่ทำงาน โรงเรียน ห้างสรรพสินค้า หรือที่สาธารณะ เพื่อตรวจจับท่าทางมือขอความช่วยเหลือในรูปแบบต่าง ๆ และสามารถแจ้งเตือนไปยังแอปพลิเคชันไลน์ (LINE) รวมทั้งสามารถทำการแจ้งเตือนเข้าไปที่เจ้าหน้าที่ ที่ดูแลความปลอดภัยผ่านทางเสียงแจ้งเตือน หรือสัญญาณไฟกระพริบ โดยติดต่อผ่านอุปกรณ์ไมโครคอนโทรลเลอร์เพื่อส่งสัญญาณไปยัง Output แบบอื่น ๆ ต่อไป

ภัทรณัฐ และคณะ (2565) ได้วิจัยเกี่ยวกับโปรแกรมแปลภาษามือเป็นข้อความและเสียงพูดโดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe ได้ศึกษาประสิทธิภาพของการรู้จำภาษามือไทยด้วย Long Short Term Memory (LSTM) อัลกอริทึมบนเทคนิคโครงกระดูกของ MediaPipe ในการทดลองสามารถแบ่งออกได้ 3 หัวข้อ 1) การสะกดด้วยนิ้วสำหรับตัวเลข (1-9), 2) ภาษามือธรรมชาติ จำนวน 20 คำ และ 3) โปรแกรมแปลภาษามือด้วยซอฟต์แวร์ โดยผลการทดลองแสดงความแม่นยำของการสะกดด้วยนิ้วและแบบภาษามือธรรมชาติด้วยโปรแกรมแปลภาษามืออยู่ระหว่าง 94, 99 และ 80.5% ตามลำดับ

ณัฏฐา และธนภัทร (2565) ได้วิจัยเกี่ยวกับระบบแปลคำศัพท์ภาษามือไทยโดยการเรียนรู้เชิงลึกบนข้อมูลบางส่วน โดยมีวัตถุประสงค์ 1) เพื่อพัฒนาระบบแปลคำศัพท์ภาษามือไทยในชีวิตประจำวัน โดยที่แปลคำศัพท์ภาษามือไทยในชีวิตประจำวันได้เบื้องต้น 2) เพื่อสร้างโมเดลโดยใช้ตัวอย่างข้อมูลจำนวนน้อย โดยการสร้างโมเดลที่เหมาะสมเพื่อแปลคำศัพท์ภาษามือไทยโดยแบ่งออกเป็น 3 หมวด ได้แก่ หมวดความรู้สึกจำนวน 21 คำ หมวดรสชาติจำนวน 9 คำ และหมวดคำกริยาจำนวน 17 คำ รวมทั้งหมด 47 คำ ผลการทดลองให้ความแม่นยำในการแปลคำศัพท์ภาษามือไทยทั้ง 47 คำ วัดค่า Accuracy ได้ถึง 74%

พิศณุ (2564) ได้วิจัยเกี่ยวกับระบบรู้จำท่าทางมือทางทหารด้วยการประมวลผลภาพดิจิทัล โดยนำเสนอระบบรู้จำท่าทางมือทางทหารด้วยการประมวลผลภาพดิจิทัล โดยเริ่มจากการเรียนรู้จดจำท่าทางมือโดยใช้คอมพิวเตอร์ด้วยการประมวลผลภาพดิจิทัล ทำการตรวจจับและรู้จำท่าทางของมือจากอัลกอริทึม จำนวน 3 วิธีคือ โครงข่ายประสาทเทียม วิธีการเพื่อนบ้านใกล้ที่สุดและการจับคู่แม่แบบ ผลลัพธ์ที่ได้พบว่า วิธีการโครงข่ายประสาทเทียมให้ผลความถูกต้องมากที่สุดคือ 82%

Indriani et al. (2021) ได้ทำการวิจัยเกี่ยวกับการพัฒนาแอปพลิเคชันคู่มือของผู้ใช้อย่างง่ายโดยใช้เฟรมเวิร์ก MediaPipe คู่มือผู้ใช้ (User manual) เป็นที่รู้จักกันทั่วไปว่าเป็นเอกสารเกี่ยวกับการสื่อสารทางเทคนิคหรือคู่มือในระบบบางอย่างเพื่อช่วยเหลือผู้ใช้ให้ใช้งานอย่างถูกต้อง ในการศึกษาจะทำการจับภาพแบบเรียลไทม์โดยใช้ Kinect จากนั้นทำการฝึกข้อมูล (Training) ท่าทางมือที่หลากหลาย ระบุท่าทางมือแต่ละท่าทาง และท่าทางมือที่รู้จักเพื่อถ่ายทอดข้อมูลตามท่าทางมือในแอปพลิเคชันคู่มือผู้ใช้ระบบ เปลี่ยนคู่มือผู้ใช้ที่ยังคงเป็นแบบแมนนวลให้กลายเป็นแอปพลิเคชันแบบที่สามารถโต้ตอบกับผู้ใช้ได้มากขึ้น โดยผลลัพธ์ที่ได้มีประสิทธิภาพที่แม่นยำถึง 95%

Sundar and Bagyammal (2022) ได้ทำการวิจัยเกี่ยวกับการรู้จำภาษามือแบบอเมริกันสำหรับตัวอักษรโดยใช้ MediaPipe และ LSTM โดยศึกษาถึงแนวทางการรู้จำตัวอักษรภาษามือแบบอเมริกัน (ASL) ที่ง่ายและมีประสิทธิภาพโดยใช้การมองเห็นเพื่อจดจำท่าทางสัมผัสทั้งแบบคงที่และแบบไดนามิก โดยใช้ Mediapipe ที่เปิดตัวโดย Google ถูกนำมาใช้เพื่อตรวจจับจุดสังเกตบนฝ่ามือ และชุดข้อมูลที่กำหนดเองได้ถูกสร้างขึ้นและใช้สำหรับการศึกษาเชิงทดลอง การรู้จำท่าทางมือทำได้โดยใช้หน่วยความจำระยะสั้นแบบยาว (LSTM) ระบบนี้ได้รับการตรวจสอบด้วยตัวอักษร 26 ตัว งานนี้สามารถใช้เพื่อแปลงท่าทางมือเป็นข้อความ และมีความแม่นยำถึง 99%

งานวิจัยที่อ้างอิงข้างต้นจะเห็นได้ว่าโครงข่ายประสาทเทียมที่ใช้ในการตรวจจับท่าทางมือในงานวิจัยต่าง ๆ ที่กล่าวมาให้ผลความถูกต้องแม่นยำค่อนข้างสูง โดยเฉพาะงานวิจัยของ พิศณุ (2565) ได้วิจัยเกี่ยวกับ ระบบรู้จำท่าทางมือทางทหารด้วยการประมวลผลภาพดิจิทัล จำนวน 3 วิธีคือ โครงข่ายประสาทเทียม วิธีการเพื่อนบ้านใกล้ที่สุดและ การจับคู่แม่แบบ พบว่า วิธีการโครงข่ายประสาทเทียมให้ผลความถูกต้องมากที่สุด สอดคล้องกับ Indriani et al. (2021) และ Sundar and Bagyammal (2022) ที่ได้ทำการวิจัยโดยใช้ MediaPipe เช่นเดียวกันซึ่งทั้งสองระบบต่างก็ให้ผลความแม่นยำที่สูงมาก เนื่องจากอัลกอริทึมบนเทคนิคโครงกระดูกของ MediaPipe เป็นโมเดลที่ได้รับการฝึกฝนโดยใช้ข้อมูลภาพประมาณ 30,000 ภาพ ให้ความแม่นยำที่สูงกว่าเทคนิคอื่น ๆ รวมทั้งมีความรวดเร็วในการทำนาย ผู้พัฒนาจึงเลือกใช้ MediaPipe ในการพัฒนาระบบการตรวจจับท่าทางมือเพื่อเป็นสัญญาณขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe

วัสดุ อุปกรณ์ และวิธีการ

ภาษามือ

ภาษามือ (Sign language) เป็นอวัจนภาษาที่มีการสื่อสาร ประกอบด้วย การสื่อสารด้วยมือ การสื่อสารด้วยร่างกาย และการใช้ริมฝีปากในการสื่อความหมายแทนการใช้เสียงพูด การสื่อสารจะใช้ลักษณะของมือที่ทำเป็นสัญลักษณ์ในการสื่อสารเป็นหลัก การเคลื่อนไหวมือ แขนและร่างกาย และการแสดงความรู้สึกทางใบหน้าเพื่อช่วยในการสื่อสารความคิดของผู้สื่อ ภาษาสัญลักษณ์ส่วนใหญ่มักใช้ในกลุ่มผู้บกพร่องทางการได้ยิน ภาษามือจึงมีความสำคัญในการสื่อสารระหว่างผู้บกพร่องทางการได้ยินและบุคคลธรรมดา ทำให้สามารถสื่อสารได้ในชีวิตประจำวัน เช่น การสนทนา การซื้อสินค้า การแสดงความคิดเห็น เป็นต้น รวมทั้งยังเป็นการเพิ่มโอกาสให้บุคคลที่เป็นผู้บกพร่องทางการได้ยินได้รับการศึกษาอีกด้วย

ราชบุรณ์ (2551) ได้อธิบายว่า ภาษามือในแต่ละประเทศมีความแตกต่างกัน สำหรับประเทศไทย ภาษามือไทยมีความแตกต่างจากภาษามือประเทศอื่น ๆ ภาษามือที่ใช้ในประเทศไทยมีหลายลักษณะ แบ่งออกเป็น 4 ประเภท คือ ภาษาท่าทาง ภาษามือครอบครัว ภาษามือตามไวยากรณ์ภาษาไทย และภาษามือไทย

1. ภาษาท่าทาง (Gesture) เป็นการสื่อสารโดยการเคลื่อนไหวของร่างกายที่ไม่จำเป็นต้องพูดก็สามารถเข้าใจได้ตรงกัน
2. ภาษามือครอบครัว (Home sign) เป็นภาษามือที่คิดและใช้กันภายในครอบครัว เพื่อการสื่อสารระหว่างพ่อแม่และลูกที่มีความบกพร่องทางการได้ยิน
3. ภาษามือตามไวยากรณ์ภาษาไทย (Signed thai) เป็นการใช้ภาษามือที่คัดลอกคำต่อคำจากประโยคภาษาไทย สำหรับการสื่อสารระหว่างครูและนักเรียนที่มีความบกพร่องทางการได้ยิน
4. ภาษามือไทย (Thai sign language) เป็นภาษาแรกของคนหูหนวกไทย ซึ่งได้รับอิทธิพลมาจากภาษามืออเมริกัน

เครื่องมือที่ใช้ในการพัฒนา

ในการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe นี้จะมีการใช้เครื่องมือต่าง ๆ ที่เป็นไลบรารีที่ทำงานหลัก ๆ ดังต่อไปนี้

1. MediaPipe framework คือ Machine learning solutions (ML) ที่สามารถใช้งานได้หลากหลายทั้งการตรวจจับท่าทาง ตรวจจับมือ ใบหน้าของมนุษย์ และตรวจจับวัตถุต่าง ๆ โดยในระบบนี้จะเป็นการตรวจจับท่าทางมือแบบ Real-time โดยในการพัฒนาระบบนี้จะใช้โมเดล 2 โมเดลด้วยกัน เพื่อให้สามารถตรวจจับท่าทางมือได้ คือ Palm detection model ที่ใช้ในการตรวจจับฝ่ามือ และ Hand landmarks detection model ที่ใช้ตรวจหาตำแหน่งของจุด Landmark แต่ละจุดบนฝ่ามือนั้น โดยในการพัฒนาระบบนี้จะใช้โมเดลที่ได้เรียนรู้จุดสังเกตของมือ (Hand landmarks) ซึ่งจะตรวจจับตำแหน่งคีย์พอยต์ของพิกัดของมือ และนิ้วจำนวน 21 ตำแหน่ง ภายในบริเวณมือที่ตรวจพบ ซึ่งโมเดลนี้ได้รับการฝึกฝนโดยใช้ภาพจริงประมาณ 30,000 ภาพ จึงมีความแม่นยำสูง “MediaPipe framework เป็นคอมโพเนนต์ระดับต่ำที่ใช้สร้างไปป์ไลน์แมชชีนเลิร์นนิงบนอุปกรณ์ที่มีประสิทธิภาพ โดยลดความซับซ้อนของการสร้าง ML บนอุปกรณ์ที่ปรับแต่งได้ พร้อมใช้งานจริงและเข้าถึงได้ในทุกแพลตฟอร์ม” (Google Inc., 2023a)

ความสามารถในการรับรู้ รูปร่างและการเคลื่อนไหวของมือ สามารถเป็นองค์ประกอบสำคัญในการปรับปรุงประสบการณ์ผู้ใช้ในแพลตฟอร์มทางเทคโนโลยีต่าง ๆ โดยที่ MediaPipe เป็นเฟรมเวิร์กข้ามแพลตฟอร์มแบบโอเพ่นซอร์สสำหรับการสร้างไปป์ไลน์เพื่อประมวลผลข้อมูลการรับรู้ของรูปแบบต่าง ๆ เช่น วิดีโอและเสียง โดยการติดตามมือและนิ้วที่มีความแม่นยำสูงโดยใช้การเรียนรู้ของเครื่อง (ML) เพื่ออนุมานจุดสำคัญ 3 มิติ 21 จุดของมือจากเฟรมเดียว (Valentin and Fan, 2019).

2. OpenCV (Open-source computer vision) คือ ไลบรารีซอฟต์แวร์โอเพ่นซอร์สการมองเห็นด้วยคอมพิวเตอร์ (Computer vision) และการเรียนรู้ของเครื่อง (Machine learning) OpenCV ถูกสร้างขึ้นเพื่อให้เป็นโครงสร้างพื้นฐานทั่วไปสำหรับแอปพลิเคชัน การมองเห็นด้วยคอมพิวเตอร์ โดยมีเป้าหมายในการแสดงผลด้วยคอมพิวเตอร์แบบเรียลไทม์ (Real-time) โดย OpenCV ถูกพัฒนาขึ้นโดย Intel และเป็นไลบรารีแบบข้ามแพลตฟอร์ม และสามารถใช้งานได้ฟรีภายใต้ลิขสิทธิ์ของ BSD แบบ Open source โดย OpenCV ยังรองรับ Framework การเรียนรู้เชิงลึก Deep learning frameworks เช่น TensorFlow, Torch, PyTorch เป็นต้น OpenCV นั้นถูกเขียนขึ้นมาโดยภาษา C++ และยังมีการรองรับภาษา Python, Java และ Matlab โดยในงานวิจัยนี้จะใช้ภาษา Python ในการพัฒนาระบบ (OpenCV, 2023)

ขั้นตอนการพัฒนา

ในส่วนของการพัฒนาระบบการตรวจจับท่าทางมือเพื่อเป็นสัญญาณขอความช่วยเหลือ ซึ่งเป็นเป้าหมายของงานวิจัยนี้ ผู้วิจัยเลือกที่จะเริ่มต้นจากสัญญาณมือขอความช่วยเหลือพื้นฐาน อีกทั้งผู้วิจัยยังตัดสินใจที่จะนำเอาแบบจำลองที่ถูกสร้างไว้อย่างดีแล้วของทีมนักวิจัยจากกูเกิลคือ MediaPipe มาต่อยอดเพื่อให้สามารถสร้างระบบต้นแบบที่นำไปสู่การตรวจจับท่าทางมือขอความช่วยเหลือและส่งการแจ้งเตือนไปยังผู้ที่เกี่ยวข้องได้อย่างรวดเร็ว โดยผู้วิจัยได้นำผลลัพธ์ของมือข้างหนึ่ง ๆ ออกมาเป็นพิกัดสามมิติของจุดจำนวน 21 จุด ดังรูปที่ 3 โดยแต่ละจุด คือ ตัวแทนของจุดสังเกตของมือและนิ้ว (Hand landmarks) ซึ่งออกแบบมาโดยทีมนักวิจัยจากกูเกิล อย่างไรก็ตามพิกัดมือนี้จำเป็นต้องผ่านการวิเคราะห์อีกขั้นเพื่อให้สามารถสรุปได้ว่าเป็นการทำท่าทางมือขอความช่วยเหลือท่าทางใด โดยผู้วิจัยได้ทำการกำหนดจุดพิกัดต่าง ๆ ในแต่ละท่าทางมือด้วยตนเอง ซึ่งระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe ได้แบ่งการทำงานออกเป็น 4 ขั้นตอนด้วยกัน ดังนี้ 1) การกำหนดท่าทางมือเพื่อขอความช่วยเหลือ 2) การติดตามท่าทางมือ (Hand gesture tracking) 3) การตรวจจับท่าทางมือ (Gesture recognition) 4) การส่งการแจ้งเตือน (Send notification)

1. การกำหนดท่าทางมือเพื่อขอความช่วยเหลือ

การกำหนดท่าทางมือเพื่อขอความช่วยเหลือ ที่ระบบจะรับรู้พร้อมกับแปลความหมายของท่าทางมือเหล่านั้นออกมา โดยในงานวิจัยนี้ได้ใช้ตัวอย่างท่าทางมือขอความช่วยเหลือสากล (Signal for help) ที่ถูกกำหนดขึ้นโดยมูลนิธิสตรีแห่งแคนาดา จากประเทศแคนาดา (Canadian women's foundation) เพื่อเป็นท่าทางมือในการขอความช่วยเหลือ จากการถูกทำร้ายร่างกายจิตใจ หรือการถูกคุกคามต่าง ๆ โดยเป็นการส่งท่าทางมือโดยใช้มือเพียงข้างเดียว เพื่อให้สะดวกในการส่งสัญญาณขอความช่วยเหลือโดยท่าทาง ซึ่งท่าทางมือเป็นการส่งสัญญาณแสดงท่าทีว่าต้องการขอความช่วยเหลือโดยด่วน แต่ไม่สามารถพูดหรือใช้เสียงขอความช่วยเหลือได้โดยตรง ซึ่งท่าทางมือทั้งหมดจะมีจำนวน 6 ท่าทาง ดังแสดงในตารางที่ 1 ตามลำดับดังนี้

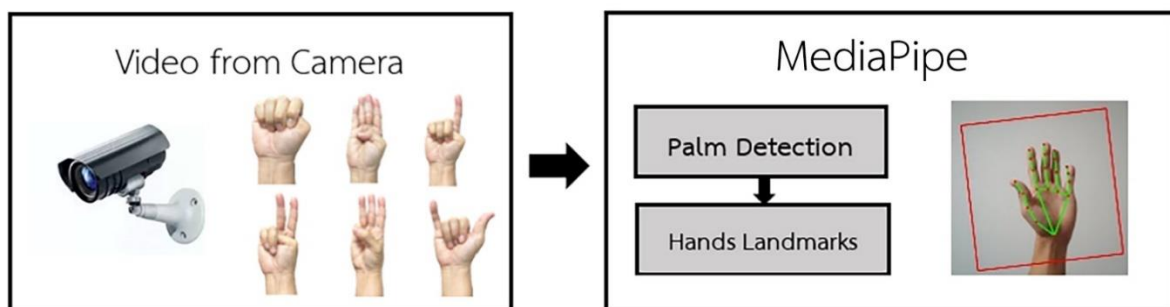
ขอความช่วยเหลือ (Signal for help) ต้องการความช่วยเหลือ (Need help) ต้องการตำรวจ (Need police) ต้องการรถพยาบาล (Need ambulance) มาหาฉัน (Visit me) โทรหาฉัน (Call me)

ตารางที่ 1 การกำหนดท่าทางมือและความหมาย

ท่าทางมือ	ความหมาย	ท่าทางมือ	ความหมาย
	ขอความช่วยเหลือ (Signal for help)		ต้องการความช่วยเหลือ (Need help)
	ต้องการตำรวจ (Need police)		ต้องการรถพยาบาล (Need ambulance)
	มาหาฉัน (Visit me)		โทรหาฉัน (Call me)

2. การติดตามท่าทางมือ (Hand gesture tracking)

การติดตามท่าทางมือในงานวิจัยนี้จะใช้ MediaPipe framework ซึ่งเป็น Machine learning ที่ใช้ในการจับท่าทางของมือ หรือ Hand tracking ข้อดี คือ ความรวดเร็วแบบ Real-time และความสะดวกในการเรียกใช้งาน MediaPipe ที่พัฒนาโดยกูเกิ้ลนั้น เพื่อที่จะฝึกระบบให้เรียนรู้จับภาพนิ้วมือได้อย่างแม่นยำ นักวิจัยต้องกำหนดพิกัดทั้ง 21 จุดเองก่อน ลงในภาพมือ จำนวน 30,000 ภาพ ในหลายอิริยาบถและในสภาพแสงที่ต่างไป เมื่อตำแหน่งท่าทางของมือถูกระบุได้แล้ว ระบบจะเปรียบเทียบกับรูปแบบมือกับฐานข้อมูลท่าทางตั้งแต่สัญลักษณ์ตัวอักษรและตัวเลขแบบภาษามือ รวมถึงคำอื่น ๆ ผลลัพธ์คืออัลกอริทึมทำงานได้เร็วและแม่นยำ



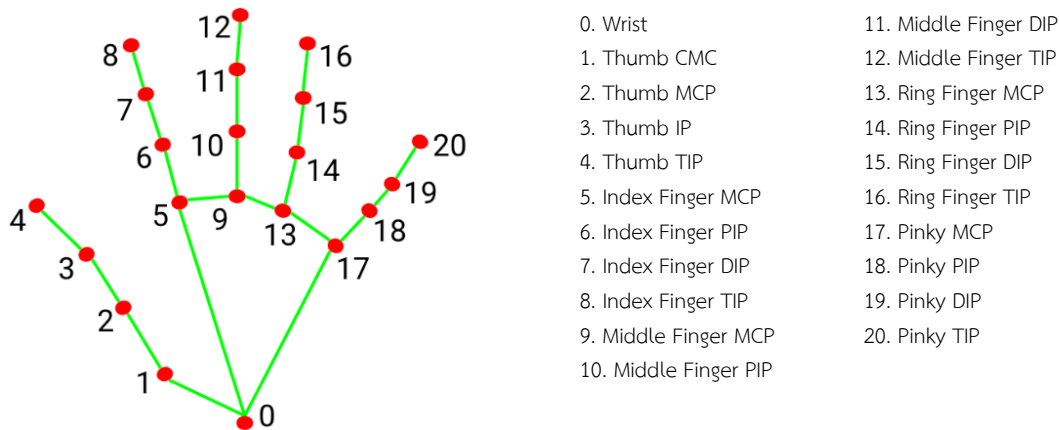
รูปที่ 2 ขั้นตอนการติดตามมือโดยใช้ MediaPipe

จากรูปที่ 2 แสดงขั้นตอนการทำงานโดยรับภาพวิดีโอจากกล้อง แล้วนำเฟรมภาพที่ได้มาตรวจหาฝ่ามือและตำแหน่งของนิ้วมือโดย MediaPipe framework ซึ่งจะประกอบด้วย 2 โมเดลหลัก ในการทำงาน คือ

1) โมเดลการตรวจจับฝ่ามือในภาพ (Palm detection model) โดยในงานวิจัยนี้จะใช้การตรวจจับจากวิดีโอ โดยการนำเฟรมภาพในวิดีโอมาประมวลผลเพื่อหาตำแหน่งของฝ่ามือ และทำการตีกรอบฝ่ามือ เพื่อให้ระบบสนใจแค่บริเวณฝ่ามือเท่านั้น

2) โมเดลตรวจจับตำแหน่งสำคัญ (Hand landmarks detection model) จะทำการระบุตำแหน่งสำคัญของฝ่ามือและนิ้วมือ จำนวน 21 จุด ดังรูปที่ 3 ที่เชื่อมโยงกับข้อนี้ ปลายนิ้ว รวมถึงระยะห่างของแต่ละนิ้วโดยอาศัยขนาดและองศาของฝ่ามือ

โดยจะทำการเรียกใช้ Palm detection model เฉพาะเฟรมแรกเท่านั้น และจะทำการเรียกอีกครั้งหากไม่พบมือในเฟรมภาพ เพื่อให้ระบบไม่ต้องตรวจหาฝ่ามือทุกครั้งและทุกเฟรม และขั้นตอนสุดท้ายคือการแสดงผลหรือแจ้งเตือน

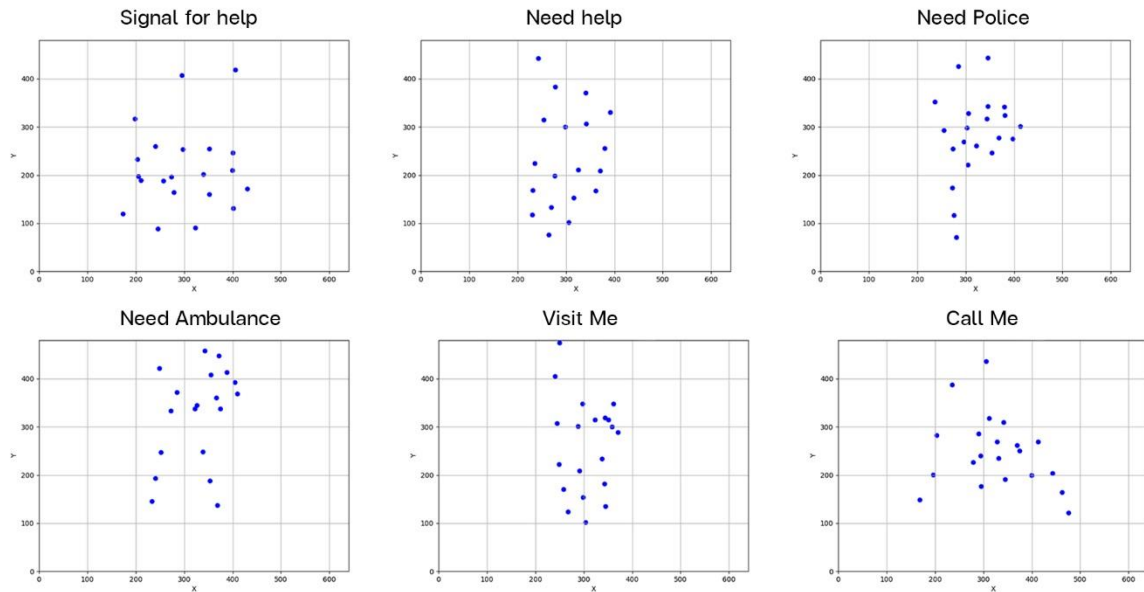


รูปที่ 3 Hand Landmarks (Google Inc, 2023)

3. การตรวจจับท่าทางมือ (Gesture recognition)

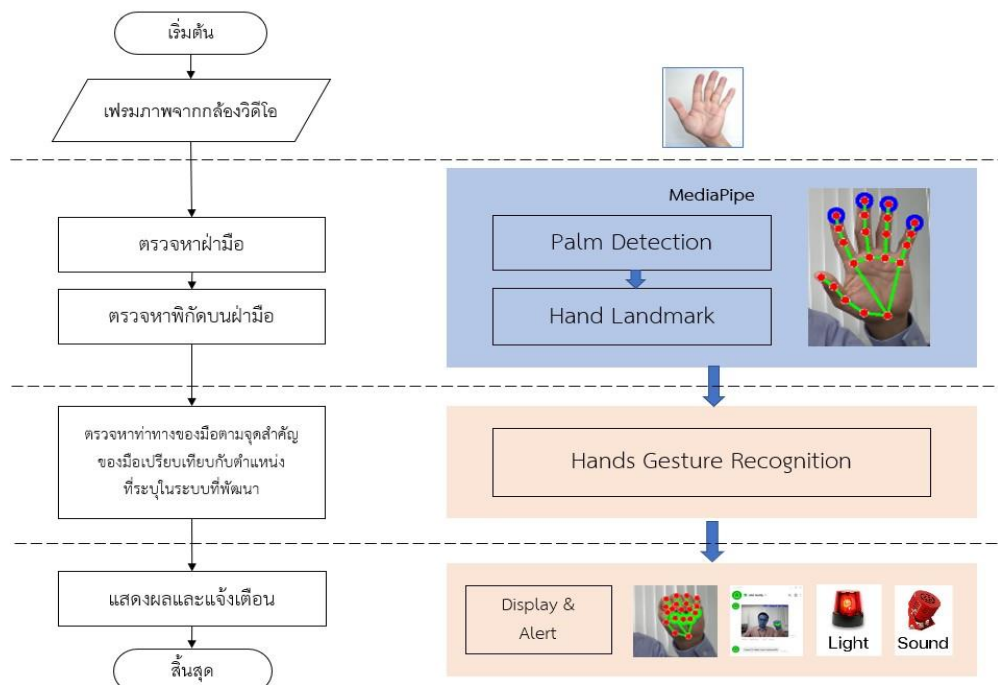
ขั้นตอนการตรวจจับท่าทางมือ โดยหลังจากกำหนดจุด Landmarks แล้วระบบจะสามารถตรวจจับท่าทางมือได้จากการระบุตำแหน่งแกน x และแกน y ตามท่าทางมือแบบต่าง ๆ ในรูปแบบ 3 มิติ ซึ่งทำงานร่วมกันระหว่าง Palm detection model และ Hand landmarks model โดยใช้ภาษา Python ในการพัฒนาระบบในการกำหนดจุดพิกัดบนแกน X, Y ของมือในท่าทางมือขอความช่วยเหลือทั้ง 6 ท่าทาง ตัวอย่างเช่น พิกัด [4, 8, 12, 16, 20] เป็นพิกัดของปลายนิ้วและการกำหนดค่าก็จะกำหนดให้เป็นปลายนิ้ว และจะทำการเปรียบเทียบพิกัดของปลายนิ้วตามตำแหน่ง x (แนวนอน) และ y (แนวตั้ง) กับจุดกึ่งกลางพิกัด [2, 6, 10, 14, 18] หากพิกัดเปรียบเทียบปลายนิ้วมีค่าสูงกว่าจุดกึ่งกลาง แสดงว่านิ้วอยู่ในสภาพเปิด (แบ่มือ) และในทางกลับกันก็คือการกำมือ เป็นต้น

ในขั้นตอนต่อไปหลังจากกำหนดเงื่อนไขของนิ้วแต่ละนิ้วว่าเปิดหรือปิดตามที่กล่าวมาแล้วจะทำการกำหนดค่าของนิ้วทั้งหมดและเปรียบเทียบทุกตำแหน่งของนิ้วมือ (1 สำหรับนิ้วเปิดและ 0 สำหรับนิ้วปิด) ถ้าตรงตามเงื่อนไขและผลลัพธ์เป็นจริง โปรแกรมจะทำงานตามคำสั่งที่ได้กำหนดไว้ ซึ่งขั้นตอนการทำงานของระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ จะมีขั้นตอนดังแสดงในรูปที่ 4



รูปที่ 4 Hand landmarks จำนวน 21 จุด บนแกน X, Y ของมือและนิ้วเมื่อทำท่าทางมือขอความช่วยเหลือ

จากรูปที่ 4 ตัวอย่างของพิกัดจุดสามมิติจำนวน 21 จุด ในงานวิจัยนี้ซึ่งเป็นผลลัพธ์จากการใช้เฟรมเวิร์ค MediaPipe เพื่อวิเคราะห์มือที่กำลังทำท่าทางมือขอความช่วยเหลือ ทั้ง 6 ท่าทาง ทั้งนี้เนื่องจากแกน Y ของ MediaPipe กับโปรแกรมของผู้วิจัยแตกต่างกัน ผลลัพธ์ในภาพจึงเป็นภาพที่กลับหัวบนล่างดังที่ปรากฏ ทั้งนี้จะสังเกตได้ว่าภาพจุดต่าง ๆ บนตำแหน่งแกน X และ Y ในรูปที่ 4 นั้นจะตรงกับท่าทางมือขอความช่วยเหลือทั้ง 6 ท่า เพียงแต่เป็นลักษณะกลับหัวในแกน Y เท่านั้น



รูปที่ 5 ขั้นตอนการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์

ในรูปที่ 5 เป็นขั้นตอนการทำงานของระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe โดยเริ่มจากกล้องอ่านเฟรมภาพจากกล้องวิดีโอ จากนั้นระบบจะตรวจจับว่า

มีฝ่ามืออยู่ในภาพหรือไม่ เมื่อตรวจพบฝ่ามือระบบจะคำนวณกรอบของฝ่ามือที่ตรวจพบและครอบตัดรูปภาพตามกรอบ เมื่อตรวจพบฝ่ามือระบบจะพิจารณาจุดสังเกตของมือ (Hands landmarks) และดึงจุดสังเกตของมือออกมา หลังจากแยกจุดสังเกตของมือออกมาแล้ว ระบบจะเปรียบเทียบท่าทางมือที่ตรวจจับได้ว่าตรงกับท่าทางมือขอความช่วยเหลือหรือไม่ ถ้าตรวจพบระบบจะทำการส่งการแจ้งเตือนเพื่อขอความช่วยเหลือ และแจ้งเตือนตามแต่ละท่าทางที่ตรวจพบต่อไป

4. การส่งแจ้งเตือน (Send notification)

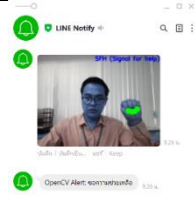
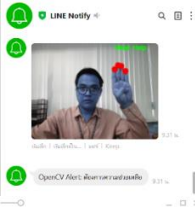
ในระบบการตรวจจับท่าทางมือขอความช่วยเหลือฯ จะเป็นขั้นตอนสุดท้ายหลังจากรับภาพวิดีโอจากกล้อง โดยนำเฟรมภาพมาตรวจจับหาฝ่ามือ ตีกรอบฝ่ามือ ตรวจจับหาพิภัก 21 จุด บนฝ่ามือ และทำการตรวจจับท่าทางมือขอความช่วยเหลือแล้ว ระบบจะทำการส่งการแจ้งเตือน โดยจะสามารถส่งการแจ้งเตือนได้ 3 วิธี คือ

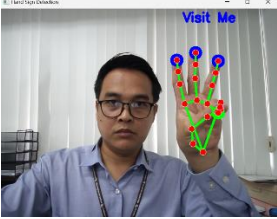
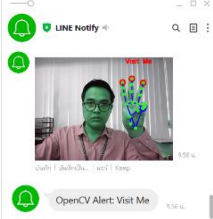
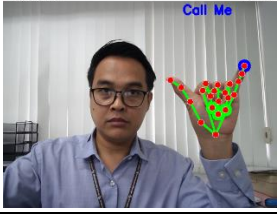
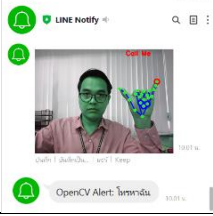
1) ส่งการแจ้งเตือนไปยังแอปพลิเคชันไลน์ (LINE) ที่เป็นที่ยอมรับกันอย่างแพร่หลายในประเทศไทย โดยการส่งข้อความการแจ้งเตือนตามท่าทางมือที่ตรวจสอบได้ รวมทั้งส่งภาพของมือและภาพของบุคคลที่ทำท่าทางมือเพื่อระบุตัวบุคคล เพื่อให้ทราบถึงตำแหน่งที่บุคคลนั้น ๆ อยู่และส่งสัญญาณขอความช่วยเหลือมา

2) การส่งการแจ้งเตือนในรูปแบบเสียงแจ้งเตือนผ่านลำโพง ซึ่งอาจติดตั้งไว้ในห้องควบคุมที่ดูแลความปลอดภัยผ่านไมโครคอลโทรลเลอร์ เพื่อไปควบคุมการเปิดปิดสัญญาณที่ส่งไปยังกริ่งหรือลำโพงแจ้งเตือน

3) การแจ้งเตือนผ่านไฟกระพริบ ผ่านไมโครคอลโทรลเลอร์ ที่จะส่งสัญญาณไปควบคุมการเปิด-ปิดสัญญาณไฟกระพริบหรือไฟหมุนฉุกเฉิน ซึ่งวิธีการนี้อาจเหมาะสำหรับผู้ดูแลความปลอดภัยที่มีความบกพร่องทางการได้ยิน หรือหน่วยงานที่มีผู้พิการทางการได้ยิน ซึ่งรับรู้ผ่านทางรูปแบบของภาพและแสงได้ดี

ตารางที่ 2 ผลของการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ไปยังแอปพลิเคชันไลน์ (LINE)

ลำดับ	ท่าทางมือ	การแสดงผลทางหน้าจอ	การแจ้งเตือนด้วยข้อความและภาพ
1	Signal for help		
2	Need help		
3	Need police		
4	Need ambulance		

5	Visit me		
6	Call me		

จากตารางที่ 2 แสดงท่าทางมือที่ทำการตรวจจับได้และแสดงผลออกทางจอภาพที่แสดงให้เห็นฝ่ามือที่ตรวจจับได้จากโมเดล Palm detection และจุดคีย์พอยต์ของมือที่ตรวจจับโดยโมเดล Hand landmarks พร้อมกับการแสดงผลผ่านแอปพลิเคชันไลน์ (LINE) โดยใช้ภาษา Python โดยกำหนดค่า Token ของ Line notify ซึ่งการแจ้งเตือนประกอบไปด้วยท่าทางมือขอความช่วยเหลือ จำนวน 6 ท่าทาง คือ 1) ขอความช่วยเหลือ (Signal for help) 2) ต้องการความช่วยเหลือ (Need help) 3) ต้องการตำรวจ (Need police) 4) ต้องการรถพยาบาล (Need ambulance) 5) มาหาฉัน (Visit me) และ 6) โทรหาฉัน (Call me) ตามลำดับ

การประเมินประสิทธิภาพ

การประเมินประสิทธิภาพของระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe จะใช้วิธี Confusion matrix ซึ่งเป็นเครื่องมือที่ใช้ในการวัดประสิทธิภาพของการทำนาย (Prediction) โดยใช้ True label และ Predict label เป็นตัวชี้วัด โดยมีรูปแบบเป็นตาราง 2x2 ดังตารางที่ 3

ตารางที่ 3 ข้อมูล Label

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

- True Positive (TP) คือ จำนวนของเคสที่ True label และ Predict label เป็น Positive
- True Negative (TN) คือ จำนวนของเคสที่ True label และ Predict label เป็น Negative
- False Positive (FP) คือ จำนวนของเคสที่ True label เป็น Negative แต่ Predict label เป็น Positive
- False Negative (FN) คือ จำนวนของเคสที่ True label เป็น Positive แต่ Predict label เป็น Negative

ตาราง Confusion matrix ใช้วัดประสิทธิภาพของโมเดลที่ถูกสร้างขึ้นมาว่ามีการทำนายถูกต้องหรือไม่ โดยจะสามารถคำนวณหาค่า Precision, Recall, F1-Score และ Accuracy ได้จากจำนวน TP, FP, FN และ TN ดังนั้นข้อมูลใน Confusion matrix นี้ ถูกใช้เพื่อปรับแต่งโมเดลหรือเลือกโมเดลที่ดีที่สุดสำหรับงานที่กำลังทำได้ (ScienceDirect, 2019)

Accuracy หมายถึง ค่าที่บ่งบอกถึงความสามารถของเครื่องมือวัด ในการอ่านค่าหรือแสดงค่าที่วัดได้เข้าใกล้ค่าจริง
Precision หมายถึง ค่าความแม่นยำที่เกิดจากการนำค่า TP มาเทียบกับ FN
Recall หมายถึง ค่าความถูกต้องที่เกิดจากการนำค่า TP มาเทียบกับ FN
F1 Score หมายถึง ค่าเฉลี่ยของค่าความแม่นยำและค่าความถูกต้อง

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

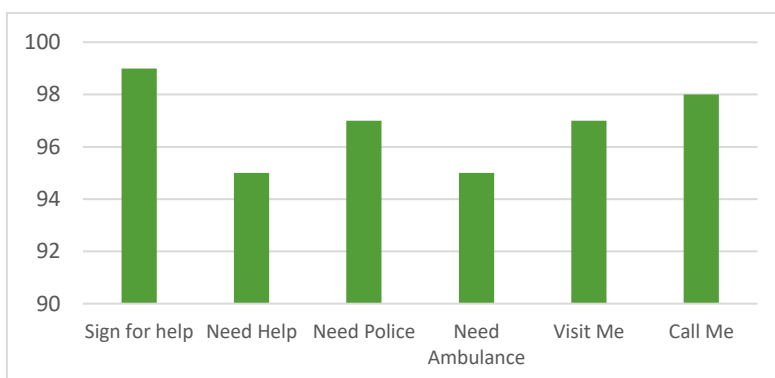
$$\text{Precision} = \frac{TP}{TP+FP}$$

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

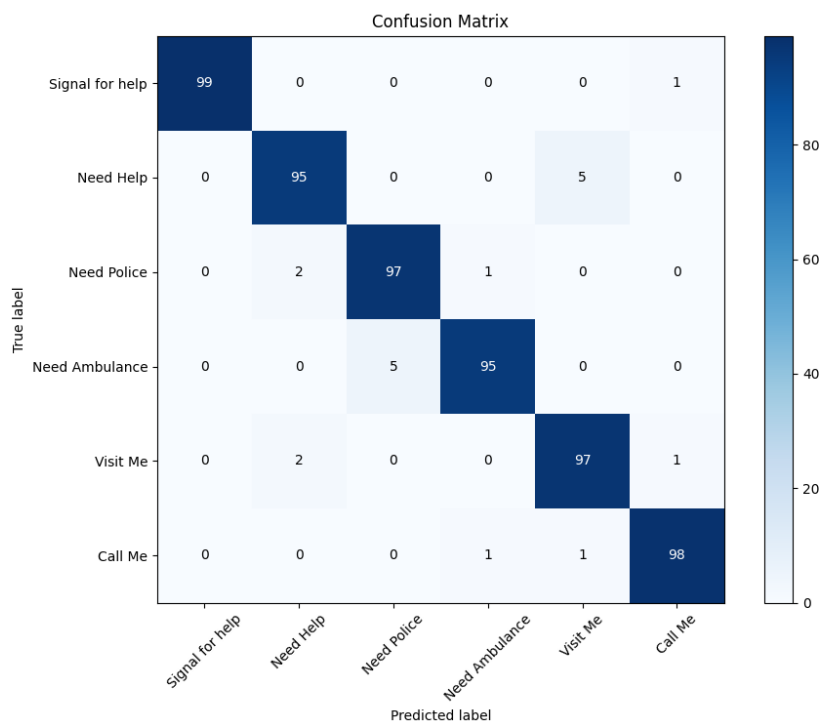
$$\text{F1 Score} = 2 \times \frac{\text{Pr} \times \text{Re}}{\text{Pr} + \text{Re}}$$

ผลการวิจัย

ผลการวิจัยจากการวัดประสิทธิภาพโดยใช้ Confusion matrix ในภาษา Python โดยจะใช้ไลบรารี Scikit-learn ในการสร้างตาราง Confusion matrix จะเห็นได้จากการตรวจจับและทำนายท่าทางมือขอความช่วยเหลือ โดยใช้ท่าทางมือขอความช่วยเหลือทั้ง 6 ท่าทาง จากการทดสอบโดยการถ่ายภาพท่าทางมือทดสอบ ท่าทางมือละ 100 ภาพ รวมทั้งหมด 6 ท่าทาง รวมเป็น 600 ภาพ ความถูกต้องของท่าทางมือขอความช่วยเหลือทั้ง 6 ท่า อยู่ในช่วง 95–99% และเมื่อนำมาหาค่าเฉลี่ยในแต่ละท่าจะได้ดังกราฟในรูปที่ 6 ค่าเฉลี่ยความถูกต้องของทั้ง 6 ท่าจะเป็น 97% เมื่อเปรียบเทียบประสิทธิภาพกับงานวิจัยอื่น ๆ ที่ใช้โครงข่ายประสาทเทียมกับท่าทางมือหรือภาษามือเหมือนกันถือว่ามีความถูกต้องที่ดีมาก เหมาะที่จะนำไปทดลองท่าทางอื่น ๆ ต่อไป



รูปที่ 6 ค่าเฉลี่ยผลการทดลองท่าทางขอความช่วยเหลือ



รูปที่ 7 Confusion matrix แสดงประสิทธิภาพการจำแนกของการรู้จำท่าทางของมือ

จากรูปที่ 7 Confusion matrix ประสิทธิภาพการจำแนกประเภทของการรู้จำท่าทางของมือ ซึ่งแสดงผลลัพธ์ของการทำนายและป้ายกำกับที่ถูกต้อง โดยใช้สีเข้มแสดงค่าที่สูงขึ้นและสีอ่อนแสดงค่าที่ต่ำลงเพื่อให้แสดงผลให้อ่านง่ายขึ้น

จากตาราง Confusion matrix แสดงความถูกต้องของการทำนายผลของการตรวจจับท่าทางมือ จำนวน 6 ท่าทาง พบว่าบางท่าทางที่มีตำแหน่งการวางมือคล้ายกัน หรือใกล้เคียงกัน เช่น Need help ที่ได้ผลการทำนาย 95% กับ Visit me ที่ได้ผลการทำนาย 97% มีลักษณะใกล้เคียงกัน หรือ Need ambulance ที่ได้ผลการทำนาย 95% กับ Visit me ที่ได้ผลการทำนาย 97% มีลักษณะการวางมือและนิ้วใกล้เคียงกัน จะเกิดความผิดพลาดเล็กน้อยหากว่าระยะห่างของมือภายในเฟรมภาพอยู่ห่างและมีขนาดภาพฝ่ามือที่เล็ก ความชัดเจนของภาพ สภาพแสง ล้วนมีผลกับการทำนายท่าทางมือทั้งสิ้น

Classification Report:				
	precision	recall	f1-score	support
Signal for help	1.00	0.99	0.99	100
Need Help	0.96	0.95	0.95	100
Need Police	0.95	0.97	0.96	100
Need Ambulance	0.98	0.95	0.96	100
Visit Me	0.94	0.97	0.96	100
Call Me	0.98	0.98	0.98	100
accuracy			0.97	600
macro avg	0.97	0.97	0.97	600
weighted avg	0.97	0.97	0.97	600

รูปที่ 8 ผลการคำนวณท่าทางมือ

จากรูปที่ 8 ผลการคำนวณท่าทางมือ จะได้ค่าความถูกต้องของท่าที่ 1-6 ของชุดข้อมูลอยู่ในช่วง 95-99% โดยข้อมูลที่โปรแกรมจัดสร้างแบ่งขึ้นมา 600 ชุด ค่า F1 score ซึ่งจากค่าที่บ่งบอกมานั้นถือว่าความผิดพลาดน้อยมาก ค่าความแม่นยำ 97%

การอภิปรายผล

งานวิจัยนี้ได้ทำการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe โดยได้ตรวจจับท่าทางมือขอความช่วยเหลือ จำนวน 6 ท่าทาง จากการทดสอบโดยใช้ภาพท่าทางมือทดสอบ ท่าทางมือละ 100 ภาพ รวมทั้งหมด 6 ท่าทาง รวมเป็น 600 ภาพ ความถูกต้องของท่าทางมือขอความช่วยเหลือทั้ง 6 ท่า อยู่ในช่วง 95-99% ค่าเฉลี่ยความถูกต้องของทั้ง 6 ท่าจะเป็น 97% เมื่อเปรียบเทียบกับประสิทธิภาพกับงานวิจัยของ พิศณุ (2564) ที่ได้วิจัยเกี่ยวกับระบบรู้จำท่าทางมือทางทหารด้วยการประมวลผลภาพดิจิทัล ทำการตรวจจับและรู้จำท่าทางของมือทางทหารจากอัลกอริทึม จำนวน 3 วิธีคือ โครงข่ายประสาทเทียม วิธีการเพื่อนบ้านใกล้ที่สุด และการจับคู่แม่แบบผลลัพธ์ที่ได้พบว่า วิธีการโครงข่ายประสาทเทียมให้ผลความถูกต้องมากที่สุดคือ 82% สอดคล้องกับงานวิจัยของ Indriani et al. (2021) ที่ได้ทำการวิจัยเกี่ยวกับการพัฒนาแอปพลิเคชันคู่มือของผู้ใช้อย่างง่ายโดยใช้เฟรมเวิร์ก MediaPipe โดยทำการจับภาพแบบเรียลไทม์โดยใช้ Kinect จากนั้นทำการฝึกข้อมูล (Training) และนำมาพัฒนาแอปพลิเคชันเพื่อถ่ายทอดข้อมูลตามท่าทางมือในแอปพลิเคชันคู่มือผู้ใช้ระบบ โดยผลลัพธ์ที่ได้มีประสิทธิภาพที่แม่นยำถึง 95% จะเห็นได้ว่าการใช้โครงข่ายประสาทเทียม โดยเฉพาะ MediaPipe ที่ได้ผ่านการเรียนรู้จากภาพจำนวนมากทำให้สามารถจำแนกท่าทางมือได้อย่างถูกต้องและรวดเร็วเป็นอย่างมาก โดยในงานวิจัยนี้ปัจจัยที่มีผลต่อคุณภาพการจำแนกท่าทางมือที่มีความแตกต่างกับงานวิจัยอื่น ๆ คือ ระยะห่างของนิ้วมือแต่ละนิ้วจะมีผลต่อการจำแนกท่าทางมือ ซึ่งขึ้นอยู่กับข้อกำหนดตำแหน่งพิกัดมือที่ได้กำหนดตำแหน่งไว้ภายในโปรแกรมภาษา Python ในแต่ละท่าทางมือจะมีการกำหนดค่าที่แตกต่างกัน รวมทั้งความใกล้เคียงและความใกล้เคียงของภาพมือภายในภาพที่ตรวจจับได้ ความคมชัดของภาพ รวมทั้งสภาพแสง ก็จะมีผลต่อผลการจำแนกท่าทางมือเนื่องจากระยะของภาพที่ไกลออกไปจะทำให้ได้ภาพถ่ายมือขนาดเล็ก เมื่อโปรแกรมนำมา Crop แต่ภาพของฝ่ามือจะทำให้ความละเอียดของภาพฝ่ามือมีความละเอียดที่ต่ำ จึงทำให้เกิดความคลาดเคลื่อนในการจำแนกท่าทางมือ จึงทำให้เกิดความแตกต่างกันของผลการจำแนกท่าทาง ซึ่งถ้าบางท่าทางที่มีลักษณะการวางนิ้วมือใกล้เคียงกันระบบอาจจำแนกผิดพลาดได้

บทสรุป

ระบบการตรวจจับท่าทางมือขอความช่วยเหลือและแจ้งเตือนแบบเรียลไทม์ โดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe โดยได้ตรวจจับท่าทางมือขอความช่วยเหลือ จำนวน 6 ท่าทาง จากการทดสอบโดยใช้ภาพท่าทางมือทดสอบ ท่าทางมือละ 100 ภาพ รวมทั้งหมด 6 ท่าทาง รวมเป็น 600 ภาพ ความถูกต้องของท่าทางมือขอความช่วยเหลือทั้ง 6 ท่าทาง อยู่ในช่วง 95-99% ซึ่งถือว่าค่อนข้างสูงเมื่อเทียบกับวิธีการอื่น ๆ โดยท่าทางมือบางท่าทางที่มีความคล้ายหรือใกล้เคียงกัน อาจเกิดความผิดพลาดจากการทำนายได้ หากระยะห่างของบุคคลกับกล้องห่างไกลกันมากทำให้ภาพฝ่ามือที่ตรวจจับได้มีขนาดเล็กและเมื่อนำมาจำแนกจะไม่มีคมชัดทำให้ความไม่แม่นยำในการจำแนกของโมเดลลดลงได้ โดยในระบบนี้ให้ค่าเฉลี่ยความถูกต้องเท่ากับ 97% ซึ่งในการวิจัยระยะต่อไปในอนาคตควรปรับปรุงให้สามารถจำแนกภาพที่อยู่ห่างไกลจากกล้องได้มากยิ่งขึ้นโดยอาจใช้เทคนิคการประมวลผลภาพเข้ามาช่วยให้ภาพคมชัดและสามารถจำแนกได้ง่ายขึ้นกว่าในปัจจุบัน

เอกสารอ้างอิง

- กรมส่งเสริมและพัฒนาคุณภาพชีวิตคนพิการ. (2565). รายงานข้อมูลสถานการณ์ด้านคนพิการในประเทศไทย. ค้นเมื่อ 16 กันยายน 2566. https://dep.go.th/images/uploads/files/Data_situa-mar66.xlsx.
- ภัทรณัฐ ศรีบุญเรือง, พันธกรณ์ ทะนันชัย, คเชนทร์เทพ แขกระโทก, วุฒิชัย วิจิตรกุลสวัสดิ์ และปราโมทย์ อนันต์วรพงษ์. (2565). โปรแกรมแปลภาษามือเป็นข้อความและเสียงพูดโดยการใช้วิธีการระบุพิกัดตำแหน่งด้วย MediaPipe. วารสารวิชาการเทคโนโลยีอุตสาหกรรม มหาวิทยาลัยราชภัฏสวนสุนันทา. 10(2): 66-76.
- ณัฏยา เปลี่ยนวงศ์ และธนภัทร ชั่งคะจิตร. (2565). ระบบแปลคำศัพท์ภาษามือไทยโดยการเรียนรู้เชิงลึกบนข้อมูลบางส่วน. วารสารบัณฑิตวิทยาลัย. 2: 119-132.
- พิศณุ คูมีชัย. (2564). ระบบรู้จำท่าทางมือทางทหารด้วยการประมวลผลภาพดิจิทัล. วารสารวิทยาศาสตร์ เทคโนโลยี และนวัตกรรม. 2(5): 28-36.
- ราษฎร์ บุญญา. (2551). ภาษามือ: ภาษาของคนหูหนวก. วารสารวิทยาลัยราชสุดาเพื่อการศึกษาและพัฒนาคนพิการ. 1: 77-94.
- Google Inc. (2023a). MediaPipe Framework. Accessed 12 Aug. 2023. <https://developers.google.com/mediapipe/framework>.
- Google Inc. (2023b). Hand landmarks detection guide. Accessed 10 Sep. 2023. https://developers.google.com/mediapipe/solutions/vision/hand_landmarker.
- Indriani M.H. and Ali S.A. (2021). Applying hand gesture recognition for user guide application using MediaPipe. Proceedings of the 2nd International Seminar of Science and Applied Technology (ISSAT 2021). 207: 101–108.
- OpenCV. (2023). About OpenCV. Accessed 12 Aug. 2023. <https://opencv.org>.
- ScienceDirect. (2019). Confusion matrix. Accessed 16 Sep. 2023. <https://www.sciencedirect.com/topics/engineering/confusion-matrix>
- Sundar B. and Bagyammal T. (2022). American sign language recognition for alphabets using MediaPipe and LSTM. 4th International Conference on Innovative Data Communication Technology and Application, Procedia Computer Science. 215: 642–651.
- The Canadian Women's Foundation. (2020). Signal for help responder. Accessed 16 Sep. 2023. <https://canadianwomen.org/signal-for-help>.
- Valentin B. and Fan Z. (2019). On-device, Real-time hand tracking with MediaPipe. Accessed 16 Sep. 2023. <https://ai.googleblog.com/2019/08/on-device-realtime-hand-tracking-with.html>.