

เครื่องต้นแบบพัฒนาภาษามือโดยใช้เทคโนโลยี IoT

IoT-Based Sign Language Recognition System

ศุภกร กมลศรี¹ สุพัตรา เขียวพริ้ม² วรัญญา ชันบรรจง³ และสุรเทพ แป้นเกิด*

คณะบริหารธุรกิจ สาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล คณะบริหารธุรกิจ

มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ 2 ถนนนางลิ้นจี่ แขวงทุ่งมหาเมฆ เขตสาทร กรุงเทพฯ 10120

Email: 655051800143@mail.rmutk.ac.th¹, 655051800523@mail.rmutk.ac.th², 655051800515@mail.rmutk.ac.th³

*Corresponding Email: suratep.p@mail.rmutk.ac.th

Manuscript Received February 7, 2026

Revised March 13, 2026

Accepted April 8, 2026

บทคัดย่อ

งานวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาระบบแปลภาษามือตัวอักษรภาษาอังกฤษแบบเรียลไทม์โดยใช้แนวคิดการประมวลผลแบบไฮบริดระหว่างเว็บไคลเอนต์และอุปกรณ์ IoT ระบบใช้ MediaPipe สำหรับสกัดพิกัดจุดข้อต่อของมือจำนวน 42 ค่า แทนการใช้ภาพถ่ายดิบ เพื่อลดความซับซ้อนของข้อมูลและเพิ่มประสิทธิภาพในการประมวลผล จากนั้นนำข้อมูลพิกัดไปจำแนกด้วยโมเดลโครงข่ายประสาทเทียมแบบหลายชั้น (MLP) บนไมโครคอนโทรลเลอร์ ESP32 โดยประยุกต์ใช้แนวคิดรูปแบบมาตรฐาน (Canonical Form) เพื่อแก้ปัญหาความแตกต่างระหว่างมือซ้ายและขวา ผลการทดลองพบว่าระบบสามารถแปลภาษามือตัวอักษรภาษาอังกฤษ A – Z (ยกเว้น J และ Z) ได้อย่างมีประสิทธิภาพ โดยมีค่าความถูกต้องเฉลี่ยเท่ากับ 91.67% และสามารถทำงานได้อย่างเสถียรในสภาพแวดล้อมที่แตกต่างกัน งานวิจัยนี้แสดงให้เห็นว่าการใช้ข้อมูลเชิงโครงสร้างร่วมกับสถาปัตยกรรมแบบไฮบริดช่วยให้ระบบแปลภาษามือมีความแม่นยำสูง เหมาะสมกับการประยุกต์ใช้บนอุปกรณ์ IoT ที่มีข้อจำกัดด้านทรัพยากร

คำสำคัญ: ภาษามือ, MediaPipe, Internet of Things, ESP32, การเรียนรู้ของเครื่อง

ABSTRACT

This research aims to develop a real-time English alphabet sign language translation system based on a hybrid computing approach integrating a web client and an IoT device. The system employs MediaPipe to extract 42 hand landmark coordinates instead of raw images in order to reduce data complexity and improve processing efficiency. The extracted coordinates are classified using a lightweight Multilayer Perceptron (MLP) model deployed on an ESP32 microcontroller. To address the differences between left and right hands, the Canonical Form concept is applied by transforming right-hand inputs into a unified representation. Experimental results indicate that the proposed system can accurately recognize English alphabet signs from A to Z, excluding J and Z, achieving an average accuracy of 91.67% while maintaining stable performance under varying environmental conditions. The results demonstrate that combining structural hand features with a hybrid computing architecture effectively enhances recognition accuracy and is well suited for IoT-based sign language translation systems with limited computational resources.

Keywords: Sign Language, MediaPipe, Internet of Things, ESP32, Machine Learning

1. ข้อมูลทั่วไป

สถิติจำนวนประชากรทางการทะเบียนพบว่าประเทศไทยมีประชากรประมาณ 66.2 ล้านคน (Department of Provincial

Administration, 2022) โดยข้อมูลจากกระทรวงการพัฒนาสังคมและความมั่นคงของมนุษย์ ระบุว่าผู้พิการที่จดทะเบียนสูงถึง 2,108,536 คน ในจำนวนนี้เป็นผู้พิการทางการได้ยินจำนวน 393,998 คน หรือคิดเป็นร้อยละ 18.69 ของผู้พิการทั้งหมด (Department of Empowerment of Persons with Disabilities 2022) จากข้อมูลดังกล่าวสะท้อนให้เห็นว่า อุปสรรคด้านการสื่อสารเป็นปัญหาสำคัญที่ส่งผลกระทบต่อการดำเนินชีวิตของผู้พิการทางการได้ยิน โดยเฉพาะการสื่อสารกับบุคคลทั่วไป เนื่องจากภาษามือมีลักษณะเฉพาะตัวและมีความแตกต่างกันตามบริบทของสังคมและวัฒนธรรม ความซับซ้อนนี้อาจนำไปสู่ความเข้าใจที่คลาดเคลื่อนในการสื่อสารร่วมกับผู้อื่นในสังคม

ภาษามือเป็นระบบการสื่อสารที่ใช้การเคลื่อนไหวของมือและท่าทางแทนการใช้เสียง ซึ่งเป็นเครื่องมือหลักของกลุ่มผู้พิการทางการได้ยิน ในช่วงหลายทศวรรษที่ผ่านมา เทคโนโลยีด้านการประมวลผลภาพและการเรียนรู้ของเครื่อง Machine Learning ได้เข้ามามีบทบาทสำคัญในการพัฒนาระบบแปลภาษามือ เพื่อทำหน้าที่เป็นสื่อกลางในการเชื่อมโยงการสื่อสารให้มีประสิทธิภาพยิ่งขึ้น การพัฒนาเครื่องต้นแบบแปลภาษามือโดยประยุกต์ใช้เทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง จึงมีความสำคัญอย่างยิ่งในการเสริมสร้างความเสมอภาคและการเข้าถึงข้อมูลข่าวสารในสังคมอย่างยั่งยืน[1]

ในปัจจุบัน Google ได้พัฒนาซอฟต์แวร์ MediaPipe [2] ซึ่งเป็นเฟรมเวิร์กที่ใช้เทคโนโลยีปัญญาประดิษฐ์และการเรียนรู้ของเครื่องที่มีประสิทธิภาพสูง จุดเด่นของ MediaPipe คือเทคนิคการตรวจจับมือ (Hand Detection) และการกำหนดจุดแลนด์มาร์กบนมือจำนวน 21 จุดต่อหนึ่งข้าง ซึ่งสามารถตรวจจับได้พร้อมกันทั้งสองมืออย่างแม่นยำ จึงเหมาะสมอย่างยิ่งในการนำมาประยุกต์ใช้เพื่อการรู้จำท่าทางในงานวิจัยนี้

จากที่กล่าวมาข้างต้นผู้วิจัยจึงมีแนวคิดในการพัฒนาเครื่องต้นแบบพัฒนาภาษามือด้วยเทคโนโลยี IOT โดยมีวัตถุประสงค์เพื่อพัฒนาเครื่องมือที่สามารถตรวจจับและแปลภาษามือเป็นตัวอักษรภาษาอังกฤษผ่านเทคโนโลยี MediaPipe และแสดงผลลัพธ์ผ่าน IOT โดยระบบจะแสดงผลลัพธ์เป็นตัวอักษรพร้อมค่าความเชื่อมั่นเป็นเปอร์เซ็นต์ เพื่อยืนยันความถูกต้องแม่นยำของท่าทาง ซึ่งจะช่วยให้ความสะดวกและลดข้อผิดพลาดในการสื่อสาร [3], [4]

ดังนั้น เห็นได้ว่างานวิจัยแปลภาษามือมีความสำคัญต่องานในหลากหลาย ด้าน เช่น 1) ด้าน ภาษาศาสตร์ เพื่อช่วยสร้างความ

เข้าใจในการสื่อสารด้วยภาษามือภาษาอังกฤษ 2) ด้าน วิทยาการคอมพิวเตอร์และปัญญาประดิษฐ์ เพื่อช่วย ในการพัฒนาและประยุกต์ใช้ MediaPipe สำหรับการแปลภาษามือ [2] และช่วยให้ผู้วิจัยที่มีความสนใจงานทาง ด้านนี้มีความเข้าใจในวิธีการจำแนกข้อมูลภาพ 3) ด้านการศึกษาและการเรียนรู้เพื่อช่วยเสริมสร้างทักษะการสื่อสารของผู้ใช้ภาษามือทั้งผู้ใช้ภาษาอังกฤษ รวมถึงการสร้าง ความเข้าใจในการเรียนรู้ภาษามือเพื่อคนทั่วไป และ 4) ด้านการประยุกต์ใช้งานการแปลภาษามือสามารถนำไปใช้ประยุกต์ในหลากหลายสถานการณ์ เช่น การประยุกต์ใช้ได้ทั้ง ผู้พิการทางการได้ยินและบุคคลทั่วไปได้ใช้งานได้ อย่างสะดวกรวดเร็ว ลดปัญหาการทำความเข้าใจ คลาดเคลื่อน

2. การออกแบบและกระบวนการทดลอง

2.1 วิธีการดำเนินงานวิจัย

สำหรับการพัฒนาเครื่องต้นแบบ ผู้วิจัยใช้ไมโครคอนโทรลเลอร์ Esp 32 cam wrover พร้อมกับติดตั้งจอ led tft โดยเลือกใช้ภาษาไพธอน (Python) ในการพัฒนาโปรแกรม โดยเฟรมเวิร์คหลักที่ใช้ในการหาจุดสำคัญบนมือคือ MediaPipe และ tensorflow บนเครื่องคอมพิวเตอร์โน้ตบุ๊ก (Notebook Computer) ที่ติดตั้งระบบปฏิบัติการ MacOS

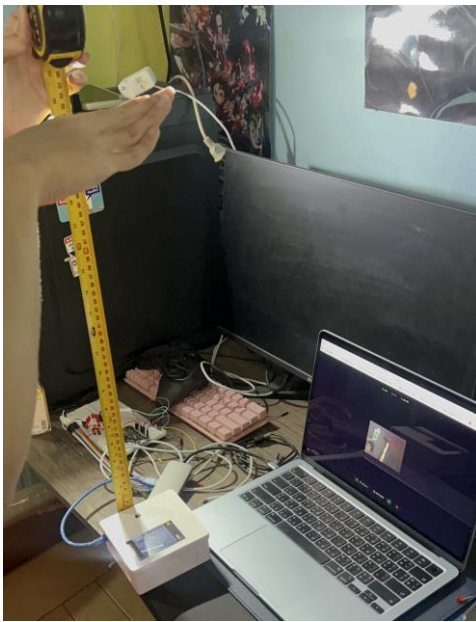
2.2 ขอบเขตของงานวิจัย

2.2.1 ด้านความสามารถของงานวิจัย

- 1) สามารถตรวจจับจุดบนฝ่ามือในพื้นที่ที่มีแสงเพียงพอ
- 2) สามารถตรวจจับจุดบนฝ่ามือได้ไกลถึงประมาณ 60 เซนติเมตร ในกรณีที่เจอจุดฝ่ามือแล้วเท่านั้นตัวอย่างการตรวจจับมือในระยะ 60 เซนติเมตรแสดงดังรูปที่ 1
- 3) สามารถตรวจจับจุดบนฝ่ามือได้แม้จะเปลี่ยนผู้ทดลองหรือฉากหลัง

2.2.2 ด้านข้อจำกัดของงานวิจัย

- 1) ไม่สามารถตรวจจับได้ในพื้นที่ที่มีแสงน้อยได้ ตัวอย่างภาพที่มีแสงน้อยแสดงดังรูปที่ 2



รูปที่ 1 การตรวจจับมือในระยะประมาณ 60 เซนติเมตร



รูปที่ 2 แสดงให้เห็นถึงภาพที่ถ่ายในพื้นที่แสงน้อย

2.2.3 สมการ

Accuracy (ความแม่นยำ) สัดส่วนของจำนวนข้อมูลที่ไม่เดลทำนายถูกต้องทั้งหมด ดังสมการที่ 1

$$\text{Accuracy} = \frac{(\text{TP}) + (\text{TN})}{\text{Total Instances}} \quad (1)$$

Precision (ความเที่ยงตรง) ความถูกต้องของข้อมูลที่ไม่เดลทำนายว่าเป็นค่าบวก ดังสมการที่ 2

$$\text{Precision} = \frac{(\text{TP})}{(\text{TP}) + (\text{FP})} \quad (2)$$

Recall (ความครอบคลุม) ความสามารถของโมเดลในการตรวจจับข้อมูลที่เป็ค่าบวกจริง ดังสมการที่ 3

$$\text{Recall} = \frac{(\text{TP})}{(\text{TP}) + (\text{FN})} \quad (3)$$

mAP (ค่าความแม่นยำเฉลี่ย) ค่าเฉลี่ยของความแม่นยำ (Average Precision) จากทุกหมวดหมู่ (Class) รวมกัน เพื่อใช้ประเมินประสิทธิภาพโดยรวมของโมเดล ดังสมการที่ 4

$$\text{mAP} = \frac{\sum_{i=1}^n \text{AP}_i}{n} \quad (4)$$

TP (True Positive) คือ จำนวนที่ไม่เดลทำนายว่า เป็นจริง/บวก และข้อมูลจริงก็ เป็นจริง/บวก (ทายถูก)

TN (True Negative) คือ จำนวนที่ไม่เดลทำนายว่า เป็นเท็จ/ลบ และข้อมูลจริงก็ เป็นเท็จ/ลบ (ทายถูกว่าไม่ใช่สิ่งที่ต้องการตรวจจับ)

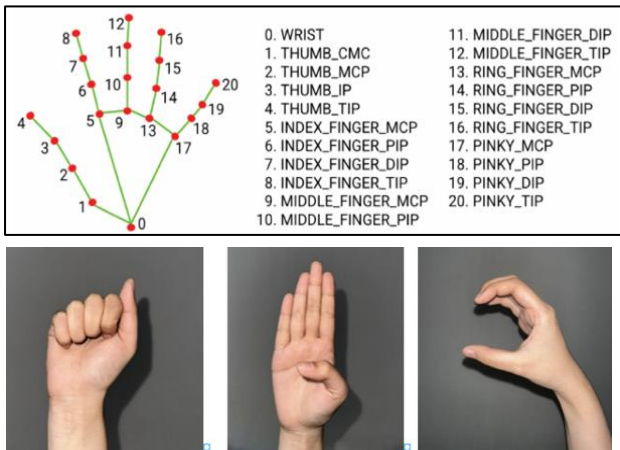
FP (False Positive) คือ จำนวนที่ไม่เดลทำนายว่า เป็นจริง/บวก แต่ข้อมูลจริงคือ เป็นเท็จ/ลบ (ทายผิดว่าใช่ ทั้งที่จริง ๆ ไม่ใช่)

FN (False Negative) คือ จำนวนที่ไม่เดลทำนายว่า เป็นเท็จ/ลบ แต่ข้อมูลจริงคือ เป็นจริง/บวก (ทายผิดว่าไม่ใช่ ทั้งที่จริง ๆ ใช่)

2.3 ขั้นตอนการดำเนินงาน

2.3.1 การเตรียมข้อมูล การวิจัยในครั้งนี้ ใช้ภาพมือทั้งหมด 7,221 ภาพโดยการแบ่งตามตัวอักษรภาษาอังกฤษทั้งหมด 24 ตัวอักษรโดยไม่รวมตัวอักษร J และ Z โดยจะเก็บภาพตัวอักษรภาษาอังกฤษอักษรละ 300 ภาพโดยมีป้ายกำกับทั้งหมด 25 ป้ายโดยจะมีป้าย

กำกับที่มีค่าว่างจำนวน 1 ป้ายดังรูปที่ 3 โดยจะมีการแบ่ง 5,776 ภาพ (80%) สำหรับการฝึกฝน 1,444 ภาพ (20%) สำหรับการทดสอบ



รูปที่ 3 ตัวอย่างตัวอักษรภาษามือภาษาอังกฤษตัว A B C และตำแหน่งของจุดทั้งสิ้น 20 จุด

2.3.2 การวิเคราะห์และออกแบบสำหรับแปลภาษามือตัวอักษรภาษาอังกฤษ

กระบวนการทำงานของระบบนี้เริ่มต้นจากการที่ผู้ใช้งานทำการเปิดโปรแกรมและ ระบุหมายเลข IP Address เพื่อทำการเชื่อมต่อสัญญาณภาพจากกล้องเข้าสู่ระบบประมวลผล เมื่อภาพปรากฏขึ้นระบบจะให้ผู้ใช้เลือกว่าต้องการจะวิเคราะห์ มือซ้ายหรือมือขวา ซึ่งจุดนี้มีความสำคัญในด้านการแสดงผล หากผู้ใช้เลือกมือขวา ระบบจะกำหนดให้ต้องมีกรกดปุ่ม Mirror บนหน้าเว็บเพื่อกลับด้านภาพให้เหมือนกับการส่องกระจก ช่วยให้ผู้ใช้งานขยับมือได้สอดคล้องกับภาพที่เห็นบนจอมากขึ้น แต่หากเลือกมือซ้ายก็สามารถเข้าสู่ขั้นตอนถัดไปได้ทันที

เมื่อระบบตรวจพบฝ่ามือในกล้องแล้ว เทคโนโลยี MediaPipe จะเข้ามาทำหน้าที่สำคัญในการสแกนและสร้างจุดเชื่อมโยงดิจิทัลลงบนมือทั้งหมด 21 จุด ซึ่งครอบคลุมตั้งแต่ข้อนิ้ว ปลายนิ้ว ไปจนถึงฐานของฝ่ามือ ข้อมูลพิกัดตำแหน่ง (Coordinates) เหล่านี้จะถูกส่งต่อไปยังบอร์ดไมโครคอนโทรลเลอร์ ESP32 เพื่อทำการประมวลผลและวิเคราะห์ผลตามอัลกอริทึมที่ตั้งไว้

ในขั้นตอนสุดท้ายก่อนการแสดงผล ระบบจะมีการตรวจสอบ

เงื่อนไขด้านความน่าเชื่อถือ โดยจะคำนวณว่าค่าทำนายที่ได้นั้นมี ความแม่นยำมากกว่า 60% หรือไม่ หากระบบประมวลผลแล้วพบว่าความแม่นยำไม่ถึงเกณฑ์ (ซึ่งอาจเกิดจากมือสั่นหรือแสงไม่พอ) ระบบจะทำการวนลูปกลับไปจับพิกัดที่มือใหม่ซ้ำๆ จนกว่าจะได้ค่าที่เสถียรและแม่นยำเพียงพอ เมื่อผ่านเกณฑ์แล้ว ผลการทำนายจะถูกส่งไปแสดงผลที่ จอ LED TFT เพื่อให้ผู้ใช้งานอ่านค่าได้ เป็นอันเสร็จสิ้นกระบวนการทำงานดังรูปที่ 4

2.3.3 การออกแบบงานวิจัยการออกแบบงานวิจัยฉบับนี้มุ่งเน้นการพัฒนาเครื่องต้นแบบพัฒนาภาษามือโดยใช้เทคโนโลยีเทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง (Internet of Things : IoT) สามารถทำงานได้จริง มีความถูกต้อง และสอดคล้องกับวัตถุประสงค์ของการวิจัย โดยแบ่งขั้นตอนการออกแบบออกเป็นลำดับดังต่อไปนี้

2.3.3.1 การทบทวนวรรณกรรมที่เกี่ยวข้อง

ผู้วิจัยได้ทำการศึกษางานวิจัยและเอกสารที่เกี่ยวข้องกับการแปลภาษามือ การประมวลผลภาพ การเรียนรู้ของเครื่อง และเทคโนโลยีไอโอที เพื่อนำแนวคิด หลักการ และข้อจำกัดจากงานวิจัยที่ผ่านมา มาใช้เป็นแนวทางในการออกแบบระบบและเลือกเทคโนโลยีที่เหมาะสมกับการพัฒนาเครื่องต้นแบบ

ในการศึกษาวรรณกรรมพบว่าการใช้ MediaPipe ร่วมกับโมเดลโครงข่ายประสาทเทียมช่วยเพิ่มความแม่นยำในการจำแนกภาษามือได้ดี อย่างไรก็ตาม โมเดลประเภท MLP ยังมีข้อจำกัดในการแปลตัวอักษรที่เป็นท่าทางเคลื่อนไหวอย่าง J และ Z เนื่องจากรองรับเฉพาะข้อมูลภาพนิ่ง

จึงมีความจำเป็นต้องศึกษาแนวทางการใช้โมเดล Long Short-Term Memory (LSTM) ซึ่งเป็นโครงข่ายประสาทเทียมแบบเรียกซ้ำ (RNN) ที่โดดเด่นในการประมวลผลข้อมูลอนุกรมเวลา (Time-series data) โดยโครงสร้างของ LSTM มีส่วนช่วยจดจำข้อมูลต่อเนื่องและคัดกรองข้อมูลสำคัญ ทำให้เหมาะกับการนำมาพัฒนาต่อยอดเพื่อรองรับภาษามือที่มีการเคลื่อนไหว (Dynamic Sign Language) ให้มีประสิทธิภาพมากยิ่งขึ้นในอนาคต

2.3.3.2 การศึกษาและประกอบอุปกรณ์ไอโอที

ขั้นตอนนี้เป็นการศึกษาโครงสร้างและคุณสมบัติของอุปกรณ์ฮาร์ดแวร์ที่ใช้ ได้แก่ โมดูล ESP32-CAM กล้องถ่ายภาพ และหน้าจอ

แสดงผล TFT โดยผู้วิจัยได้ทำการประกอบอุปกรณ์ ตรวจสอบการทำงาน และทดสอบการเชื่อมต่อเครือข่าย เพื่อให้มั่นใจว่าอุปกรณ์สามารถรับส่งข้อมูลภาพและแสดงผลลัพธ์ได้อย่างถูกต้องดังรูปที่ 5 และ รูปที่ 6

2.3.3.3 การออกแบบการตั้งค่าการฝึกโมเดล

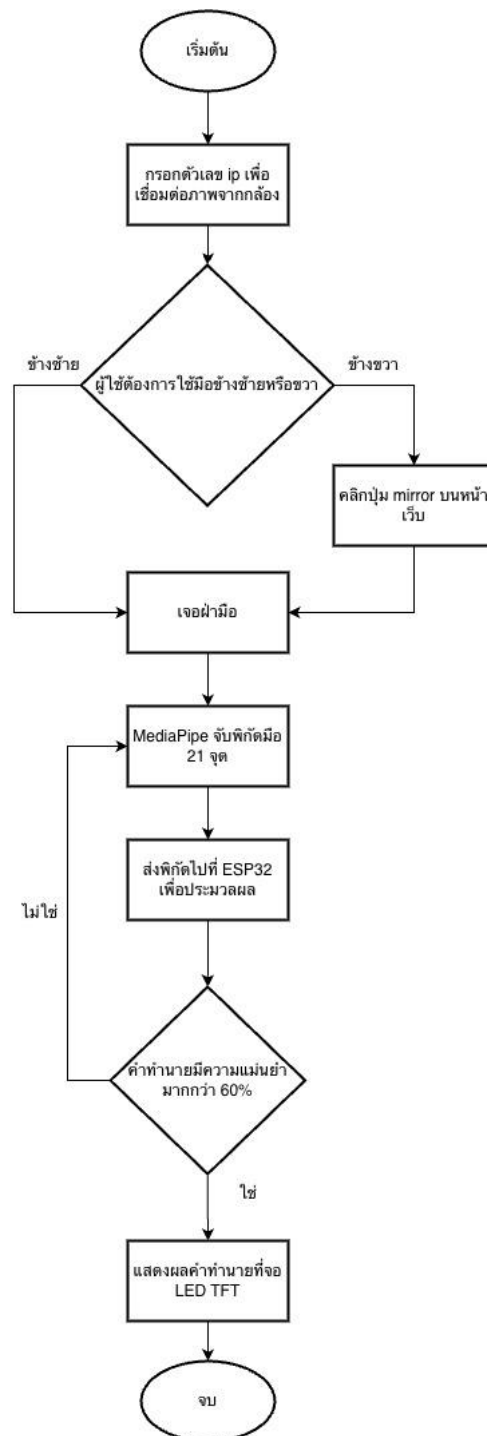
ผู้วิจัยได้กำหนดค่าพารามิเตอร์ที่ใช้ในการฝึกโมเดลการเรียนรู้ของเครื่อง เพื่อให้โมเดลมีความแม่นยำและลดความผิดพลาดในการทำนาย โดยค่าพารามิเตอร์ที่ใช้ประกอบด้วย จำนวนรอบการฝึก ค่าแบบตซ์ และค่าเริ่มต้นแบบสุ่ม ซึ่งแสดงรายละเอียดดังตารางที่ 1

ตารางที่ 1 ค่าพารามิเตอร์

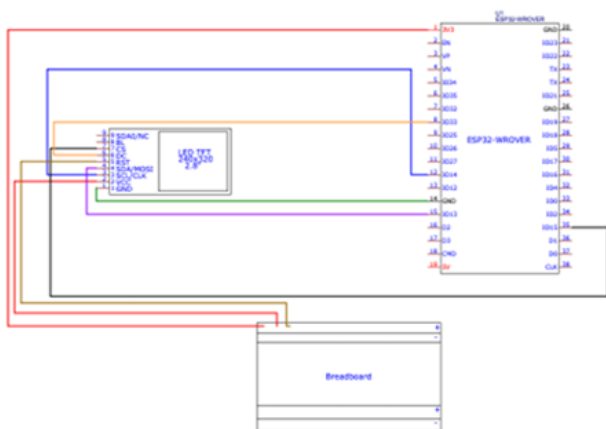
Name of parameter	Value
Epochs	100
BATCH_SIZE	32
SEED	42

2.3.3.4 การออกแบบและพัฒนาหน้าเว็บไซต์

ผู้วิจัยได้ออกแบบและพัฒนาหน้าเว็บไซต์เพื่อใช้เป็นส่วนติดต่อผู้ใช้ โดยใช้ภาษาไทป์สคริปต์ร่วมกับเฟรมเวิร์คเน็กซ์เจเอส เพื่อให้ผู้ใช้งานสามารถเชื่อมต่อกล้อง ดูภาพแบบเรียลไทม์ และภายในเว็บไซต์ติดตั้งสคริปต์การทำงานร่วมกับ Mediapipe โดยการทำงานเริ่มเมื่อตรวจพบมือของผู้ใช้งาน จะทำการสร้างจุดพิกัดของมือ และส่งต่อไปยัง ESP-32 เพื่อประมวลผลต่อไปดังรูปที่ 7



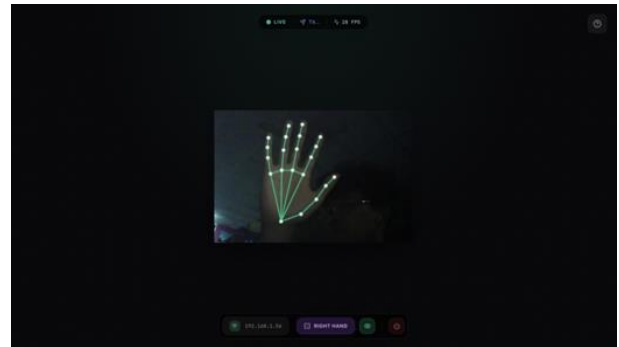
รูปที่ 4 การวิเคราะห์และออกแบบสำหรับแปลภาษามือตัวอักษรภาษาอังกฤษ



รูปที่ 5 การออกแบบ Hardware



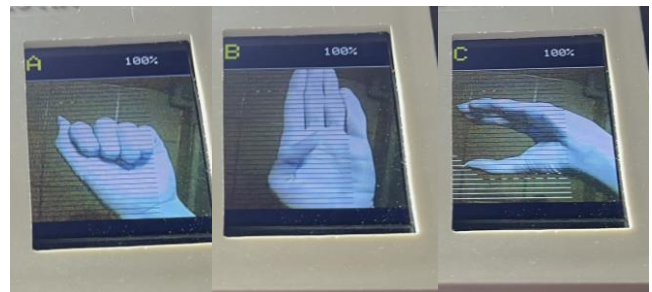
รูปที่ 6 การออกแบบเครื่องต้นแบบพัฒนาภาษามือโดยใช้เทคโนโลยี IOT ก่อนนำไปประกอบให้เรียบร้อย



รูปที่ 7 การสร้างจุดพิกัดของมือ

2.4 ผลการศึกษา

การทดลองการตรวจจับมือการค้นหาจุดบนมือและการแปลภาษามือตัวอักษรภาษาอังกฤษจากการทดลองพบว่า ระยะห่างระหว่างบุคคลและกล้อง ESP32 ที่เหมาะสม ที่สุดอยู่ที่ระยะไม่เกิน 60 เซนติเมตร ดังนั้น ในการวัด ประสิทธิภาพจึงกำหนดไว้ที่ระยะดังกล่าว ตัวอย่างของการทดลองการแปลภาษามือตัวอักษรภาษาอังกฤษซึ่งแสดงให้เห็นถึงการทดลองการแปล ตัวอย่างภาษามือภาษาอังกฤษของตัวอักษร a-c ดังรูปที่ 8



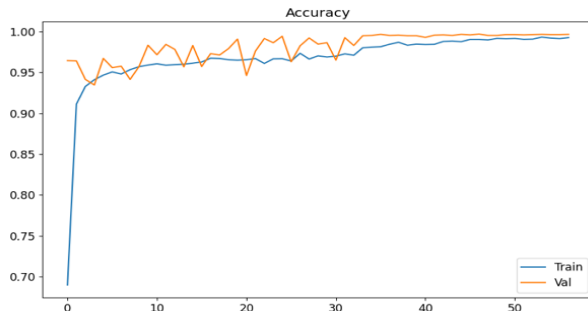
รูปที่ 8 หน้าจอแสดงผล IOT โดยแสดงตัวอย่างภาษามือตัวอักษรภาษาอังกฤษ a-c

ตารางที่ 3 สรุปผลการทดสอบของทั้ง 3 คน

คนที่ 1	คนที่ 2	คนที่ 3
91.67%	91.67%	91.67%
เฉลี่ยรวม	91.67%	

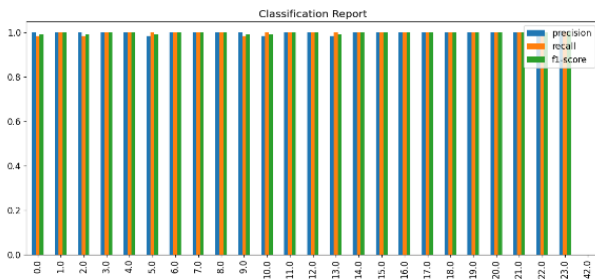
จากตารางที่ 3 แสดงให้เห็นว่าเมื่อทดสอบของทั้ง 3 คนพบว่า ความถูกต้องในการแปลภาษามือเฉลี่ยที่ 91.67% ฝ่ามือได้ โดย

ระยะที่เหมาะสมที่สุดที่จะตรวจจับและแปลภาษามือคือระยะ 60 เซนติเมตร จากผลการ ทดลองแต่ละตัวอักษรมีความแม่นยำถูกต้อง โดยเฉลี่ยที่ 91.67% แต่จะต้องวางมือในตำแหน่งที่ถูกต้องตาม เงื่อนไขที่ได้ตั้งไว้เท่านั้น



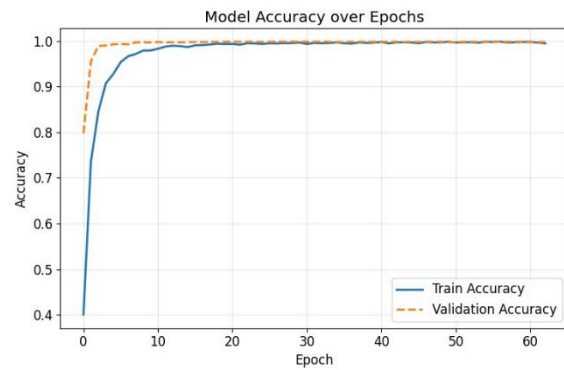
รูปที่ 9 ค่าความแม่นยำ

จากรูปที่ 9 แสดงความแม่นยำ (Accuracy) ของแบบจำลอง ในช่วงการฝึก (Trainings) และการทดสอบ (Validation) พบว่า แบบจำลองมีการเรียนรู้ที่รวดเร็วในช่วงต้น โดยค่า accuracy ของ ชุดฝึกเพิ่มขึ้นจากประมาณ 0.70 ไปมากกว่า 0.90 ภายในไม่กี่ epoch แรก และมีแนวโน้มเพิ่มขึ้นอย่างต่อเนื่องจนเข้าใกล้ 1.00 เมื่อสิ้นสุดกระบวนการฝึก

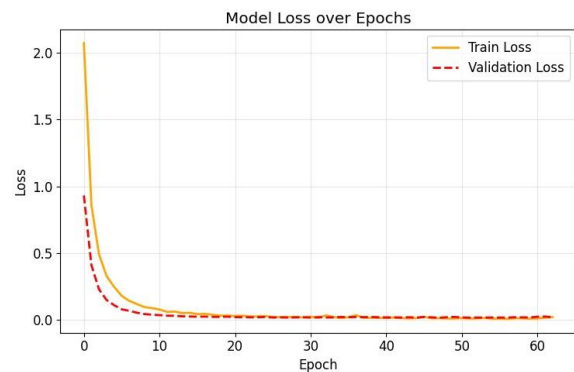


รูปที่ 10 ค่า Precision Recall และค่า F1-Score

จากรูปที่ 10 แสดงรายงานผลการจำแนกแสดงค่าความแม่นยำ (Precision) ความครอบคลุม (Recall) และค่า F1-score ของแต่ละ คลาส พบว่าค่าตัวชี้วัดทั้งสามมีค่าสูงใกล้เคียง 1.00 สำหรับทุกคลาส ซึ่งบ่งชี้ว่าแบบจำลองสามารถจำแนกข้อมูลในแต่ละคลาสได้อย่าง ถูกต้องและสม่ำเสมอ



รูปที่ 11 ค่า Model Accuracy over Epochs



รูปที่ 12 ค่า Model Loss over Epochs

จากรูปที่ 11 และรูปที่ 12 ผลการฝึกแบบจำลองแสดงให้เห็นว่า accuracy ของชุดฝึกและชุดทดสอบเพิ่มขึ้นอย่างรวดเร็วและมีค่า เข้าใกล้ 1.00 ขณะที่ค่า loss ลดลงอย่างต่อเนื่องจนเข้าใกล้ศูนย์ โดยเส้นกราฟของทั้งสองชุดมีแนวโน้มใกล้เคียงกัน สะท้อนถึง ประสิทธิภาพในการเรียนรู้ที่ดีและไม่เกิดปัญหา overfitting

3. อภิปรายผล

ผลการศึกษาพบว่าระบบสามารถแปลภาษามือตัวอักษร ภาษาอังกฤษ A – Z (ยกเว้น J และ Z) ได้อย่างมีประสิทธิภาพ โดยมีค่าความถูกต้องเฉลี่ยเท่ากับ 91.67% ซึ่งแสดงให้เห็นว่าการ ออกแบบระบบที่เน้นการลดความซับซ้อนของข้อมูลอินพุตและการ เลือกใช้สถาปัตยกรรมที่เหมาะสมกับอุปกรณ์ IoT สามารถช่วยเพิ่ม ประสิทธิภาพของการจำแนกภาษามือได้อย่างมีนัยสำคัญ การ เลือกใช้ข้อมูลพิกัดจุดของมือจำนวน 42 ค่าแทนการใช้ภาพถ่ายดิบ

ช่วยลดผลกระทบจากสัญญาณรบกวน เช่น แสง เงา และพื้นหลัง ทำให้โมเดลสามารถเรียนรู้โครงสร้างทางเรขาคณิตของมือได้โดยตรง ซึ่งสอดคล้องกับงานของวรินทร์ สุวรรณ และคณะ [2] และสหโชค คุ่มวงษา และวียดา ยะไวทย์ [1] ที่รายงานว่าการใช้ MediaPipe และคุณลักษณะเชิงโครงสร้างช่วยเพิ่มความแม่นยำของระบบ แปลภาษามือ อีกทั้งแนวทางดังกล่าวยังเหมาะสมกับข้อจำกัดด้านทรัพยากรของไมโครคอนโทรลเลอร์ ESP32 ตามแนวคิดการออกแบบระบบ IoT ของสุศักดิ์ สงวนพงษ์ [3] นอกจากนี้ การแก้ปัญหาความแตกต่างระหว่างมือซ้ายและมือขวาด้วยการประยุกต์ใช้แนวคิด Canonical Form โดยการกลับด้านข้อมูลจากมือขวาให้มีลักษณะทางเทคโนโลยีเดียวกับมือซ้าย ช่วยลดความซับซ้อนของชุดข้อมูลและลดความสับสนของโมเดล ทำให้โมเดลขนาดเล็กสามารถรองรับการใช้งานทั้งสองมือได้โดยไม่ต้องเพิ่มจำนวนพารามิเตอร์ ซึ่งสอดคล้องกับงานวิจัยด้านระบบ แปลภาษามือและ IoT ที่มุ่งเน้นการออกแบบโมเดลให้มีขนาดกะทัดรัดแต่ยังคงความแม่นยำสูง [5], [6], [7] ขณะเดียวกัน การออกแบบสถาปัตยกรรมแบบ Hybrid Computing โดยแบ่งภาระงานให้ฝั่ง Web Client ทำหน้าที่ประมวลผลภาพและสกัดพิกัดจุดผ่าน MediaPipe ก่อนส่งข้อมูลขนาดเล็กไปยัง ESP32 เพื่อทำการอนุมาน ช่วยลดภาระการประมวลผลบนอุปกรณ์ปลายทางและแก้ปัญหาคอขวดของระบบได้อย่างมีประสิทธิภาพ แนวทางนี้สอดคล้องกับงานวิจัยที่เสนอการใช้การประมวลผลแบบกระจายร่วมกับ IoT เพื่อเพิ่มความลื่นไหลและเสถียรภาพของระบบ แปลภาษามือแบบเรียลไทม์ [8], [9] อย่างไรก็ตาม ระบบยังมีข้อจำกัดในการแปลตัวอักษรที่เป็นท่าทางเคลื่อนไหว ได้แก่ J และ Z เนื่องจากโครงสร้างโมเดลปัจจุบันรองรับเฉพาะข้อมูลภาพนิ่ง ซึ่งเป็นข้อจำกัดที่พบในงานวิจัยก่อนหน้านี้ ปฏิภาณ สุขเกษม[10] และ รุ่งโรจน์ ศิริวัฒน์ [11] ดังนั้น การพัฒนาต่อยอดในอนาคตควรพิจารณาการนำโมเดลที่สามารถประมวลผลข้อมูลอนุกรมเวลา เช่น LSTM หรือ Transformer มาประยุกต์ใช้ เพื่อให้ระบบสามารถรองรับการแปลภาษามือที่มีการเคลื่อนไหวได้อย่างครบถ้วนและมีประสิทธิภาพมากยิ่งขึ้น [12], [13], [14]

4. สรุป

งานวิจัยนี้ประสบความสำเร็จในการพัฒนาเครื่องต้นแบบ แปลภาษามืออักษรภาษาอังกฤษ A-Z ยกเว้น J และ Z ที่มีความ

แม่นยำสูงถึง 91.67% โดยใช้เทคนิคสกัดพิกัดมือ 21 จุดผ่าน อัลกอริทึม MediaPipe เพื่อลดสัญญาณรบกวนจากสภาพแวดล้อม ส่งผลให้โมเดล MLP ขนาดเล็กสามารถทำงานบน ไมโครคอนโทรลเลอร์ ESP32-CAM ได้อย่างมีประสิทธิภาพ

ผู้วิจัยได้แก้ปัญหาความแตกต่างระหว่างมือซ้ายและขวาด้วยแนวคิด Canonical Form โดยการกลับด้านข้อมูลให้เป็นรูปแบบเดียวกัน เพื่อลดความซับซ้อนของชุดข้อมูล นอกจากนี้ยังใช้สถาปัตยกรรม Hybrid Computing เพื่อแบ่งภาระการประมวลผลภาพไปยัง Web Client และส่งเฉพาะพิกัดมายังชิปเพื่อแปลผล ช่วยให้ระบบทำงานได้แบบเรียลไทม์แม้ทรัพยากรจำกัด

อย่างไรก็ตาม เนื่องจากปัจจุบันรองรับเพียงภาพนิ่ง จึงยังไม่สามารถแปลตัวอักษร J และ Z ที่มีการเคลื่อนไหวได้ ซึ่งในอนาคตมีแผนจะนำโมเดล LSTM หรือ Transformer มาพัฒนาต่อยอด พร้อมปรับปรุงเพื่อสร้างเครื่องมือสื่อสารที่เข้าถึงง่ายและเสริมสร้างความเสมอภาคในสังคมดิจิทัลอย่างยั่งยืน

เอกสารอ้างอิง

- [1] สหโชค คุ่มวงษา และ วียดา ยะไวทย์, “โปรแกรมอัจฉริยะสำหรับ แปลภาษามือภาษาไทยและภาษาอังกฤษ”, Journal of Applied Informatics and Technology, 5(2), 178–194, 2566.
- [2] วรินทร์ สุวรรณ, ปิยะพงษ์ ทองดี และ สมชาย เลิศไพฑูรย์, “การประยุกต์ใช้เทคนิคการเรียนรู้เชิงลึกร่วมกับมีเดียไปป์เพื่อจำแนกภาษามือไทยสำหรับผู้พิการทางการได้ยิน”, วารสารวิชาการพระจอมเกล้าพระนครเหนือ, 31(3), 450–462, 2564.
- [3] สุศักดิ์ สงวนพงษ์, “การออกแบบและพัฒนาระบบ Internet of Things ด้วย ESP32”, กรุงเทพฯ: ซีเอ็ดดูเคชั่น, 2565.
- [4] อภิสิทธิ์ เหลืองอรุณ, “การออกแบบและพัฒนาเครื่องต้นแบบ IoT สำหรับ การเรียนรู้ภาษามือภาษาไทย”, วิทยานิพนธ์ครุศาสตรมหาบัณฑิต สาขา นวัตกรรมเทคโนโลยีการศึกษา, มหาวิทยาลัยราชภัฏสวนสุนันทา, 2565.
- [5] กิตติพงศ์ ศรีสุข, “การพัฒนาเครื่องต้นแบบแปลภาษามือภาษาไทยด้วย เทคโนโลยีอินเทอร์เน็ตของสรรพสิ่ง (IoT)”, วิทยานิพนธ์วิศวกรรมศาสตร มหาบัณฑิต สาขาวิศวกรรมไฟฟ้า, มหาวิทยาลัยเทคโนโลยีพระจอมเกล้า ธนบุรี, 2563.
- [6] ชัยวัฒน์ พงษ์ไพศาล, “ระบบสื่อสารภาษามือสำหรับผู้บกพร่องทางการได้ยินโดยใช้ไมโครคอนโทรลเลอร์”, วิทยานิพนธ์ครุศาสตรมหาบัณฑิต สาขา เทคโนโลยีการศึกษา, จุฬาลงกรณ์มหาวิทยาลัย, 2562.
- [7] ธนพล อินทรชัย, “การออกแบบและพัฒนาอุปกรณ์อัจฉริยะเพื่อช่วย แปลภาษามือไทยแบบเรียลไทม์”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต สาขาวิศวกรรมคอมพิวเตอร์, มหาวิทยาลัยเชียงใหม่, 2564.

- [8] ภาณุเดช วงศ์คำ, “การออกแบบระบบ IoT เพื่อสนับสนุนการสื่อสารภาษามือในชีวิตประจำวัน”, วิทยานิพนธ์ครุศาสตรมหาบัณฑิต สาขาเทคโนโลยีสารสนเทศ, มหาวิทยาลัยนเรศวร, 2564.
- [9] วรพล นาคทอง, “การประยุกต์ใช้ IoT เพื่อสร้างระบบช่วยแปลภาษามือภาษาไทย”, วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต, มหาวิทยาลัยบูรพา, 2563.
- [10] ปฏิภาณ สุขเกษม, “ระบบแปลภาษามือภาษาไทยด้วยการประมวลผลภาพและปัญญาประดิษฐ์”, วิทยานิพนธ์วิทยาศาสตรมหาบัณฑิต สาขาวิทยาการคอมพิวเตอร์, มหาวิทยาลัยศิลปากร, 2565.
- [11] รุ่งโรจน์ ศิริวัฒน์, “ระบบรู้จำภาษามือภาษาไทยโดยใช้กล้องและเซนเซอร์อัจฉริยะ”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต, สถาบันเทคโนโลยีพระจอมเกล้าเจ้าคุณทหารลาดกระบัง, 2565.
- [12] พรชัย แซ่ตั้ง, “การพัฒนาอุปกรณ์สวมใส่สำหรับตรวจจับท่าทางภาษามือของผู้พิการทางการได้ยิน”, วิทยานิพนธ์วิศวกรรมศาสตรมหาบัณฑิต, มหาวิทยาลัยเทคโนโลยีราชมงคลธัญบุรี, 2563.
- [13] มนัสวี รัตนพงษ์, “การพัฒนาแอปพลิเคชันแปลภาษามือภาษาไทยบนอุปกรณ์พกพา”, การศึกษาค้นคว้าอิสระ ระดับปริญญาโท, มหาวิทยาลัยมหิดล, 2562.
- [14] ลลิตา พัฒนานุกูล, “การพัฒนาเครื่องต้นแบบช่วยสื่อสารภาษามือสำหรับผู้พิการทางการได้ยินในสถานศึกษา”, วิทยานิพนธ์ศึกษาศาสตรมหาบัณฑิต, มหาวิทยาลัยศรีนครินทรวิโรฒ, 2561.



สุพัตรา เขียวพร้อม นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล คณะบริหารธุรกิจมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ



วรัญญา ชันบรรจง นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล คณะบริหารธุรกิจมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ



ผู้ช่วยศาสตราจารย์สุรเทพ แป้นเกิด อาจารย์ประจำสาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล คณะบริหารธุรกิจมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ มหานครงานวิจัยด้าน IOT ,AI และ Machine Learning (ML)



ศุภกร กมเลศร์ นักศึกษาสาขาวิชาเทคโนโลยีสารสนเทศและธุรกิจดิจิทัล คณะบริหารธุรกิจมหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ