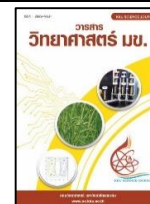




KKU SCIENCE JOURNAL

Journal Home Page : <https://ph01.tci-thaijo.org/index.php/KKUSciJ>

Published by the Faculty of Science, Khon Kaen University, Thailand



การพัฒนาระบบตรวจจับและแปลงภาษามือไทยเป็นภาษาไทย

โดยใช้การเรียนรู้เชิงลึก

Development of Thai Sign Language Detection and Conversion System into Thai with Deep Learning

ชัชชน ดำรงเอกอรุณ¹ ลูกเกต พิสิฐพิพัฒน์¹ สัจจาภรณ์ ไวจรรยา^{1*} และ ณัฐโชติ พรหมฤทธิ์^{1*}

Chatchon Damrongekarun¹, Lukket Pisitpipattana¹, Sajjaporn Waijanya^{1*} and

Nuttachot Promrit^{1*}

¹ภาควิชาคอมพิวเตอร์ คณะวิทยาศาสตร์ มหาวิทยาลัยศิลปากร วิทยาเขต พระราชวังสนามจันทร์ จังหวัดนครปฐม 73000

¹Department of Computing, Faculty of Science, Silpakorn University, Nakhon Pathom, 73000, Thailand

บทคัดย่อ

สำหรับผู้บกพร่องทางการได้ยินและการพูดจะสามารถสื่อสารได้โดยใช้ภาษามือแทนคำพูด ทำให้เป็นอุปสรรคต่อการสื่อสารกับผู้คนทั่วไป เพื่อช่วยให้ผู้บกพร่องทางการได้ยินและการพูดสามารถสื่อสารกับผู้คนทั่วไปได้ บทความนี้เสนอการพัฒนาระบบตรวจจับและแปลงภาษามือไทยเป็นภาษาไทยโดยใช้การเรียนรู้เชิงลึก โดยรับข้อมูลเข้าเป็นจุดสำคัญ (Keypoints) ของมือ ใบหน้า และท่าทาง ใช้ MediaPipe Framework ในการตรวจจับจุดสำคัญ รวบรวมข้อมูลจากวิดีโอที่เป็นท่าทางภาษามือไทยนำมาสกัดหาพิกัดของจุดสำคัญ แล้วแบ่งชุดข้อมูลสำหรับการฝึก (Train set) 80% ชุดข้อมูลทดสอบ (Test set) 20% นำข้อมูลมาฝึกฝนโมเดลโดยใช้อัลกอริทึมโครงข่ายประสาทเทียมหน่วยความจำระยะสั้นยาวแบบสองทาง (BiLSTM) ในการวิเคราะห์ท่าทาง ปรับเปลี่ยนชั้น (layer) ของโมเดลและทดลองการปรับเปลี่ยนค่าของพารามิเตอร์ต่างๆ พบว่าโมเดลที่มีการเพิ่มโหนด (node) และปรับเปลี่ยนชั้น (layer) ทำให้โมเดลสามารถป้องกันการเกิดปัญหา Underfitting และมีพารามิเตอร์ที่ใช้ในการเรียนรู้เพิ่มขึ้น ได้ค่าความถูกต้อง คือ 0.83 หลังจากได้ผลลัพธ์เป็นคำไทย จะนำมาแสดงผลในแอปพลิเคชันมือถือที่พัฒนาด้วย Flutter Framework ซึ่งเชื่อมต่อการทำงานระหว่างโมเดลและแอปพลิเคชันด้วย API ที่พัฒนาด้วย Flask API

ABSTRACT

The hearing and speech impaired can communicate using sign language instead of word, which causes obstacles in communicating with unimpaired person. To help the hearing and speech impaired to communicate with the general public, this paper presents. It takes in key points of hands, faces and gestures, using the MediaPipe Framework to detect key points. The information was gathered from video gestures in Thai Sign Language to extract the coordinates of important points. The data were then divided into 80% of the training dataset and 20% of the test dataset. The data were used to train the model using the Long

*Corresponding Author, E-mail: waijanya_s@silpakorn.edu, promrit_n@silpakorn.edu

Received: date: 2 June 2023 | Revised date: 20 June 2023 | Accepted date: 26 July 2023

Short-Term Memory (LSTM) neural network algorithm for posture analysis. The model layers and experiment were modified by modifying the parameters. It was found that the model required node addition and class modification. This allowed the model to prevent underfitting problems and increased learning parameters. The accuracy was 0.83 after the results were in Thai words. It will be rendered in a mobile application developed with the Flutter Framework, which connects the model and the application with an API developed with the Flask API.

คำสำคัญ: ภาษามือ โครงข่ายประสาทเทียมหน่วยความจำระยะสั้นยาว แอปพลิเคชันบนโทรศัพท์เคลื่อนที่

Keywords: Sign Language, Long Short-term Memory, Mobile Application

บทนำ

ในปัจจุบันสังคมได้ให้ความสำคัญกับผู้บกพร่องทางการได้ยินมากขึ้น เปิดโอกาสให้ผู้บกพร่องทางการได้ยินได้แสดงความสามารถ และเรียนรู้วิธีการใหม่ๆ เพื่อให้ผู้บกพร่องทางการได้ยินมีสิทธิ โอกาส และได้รับความเท่าเทียม เช่นเดียวกับบุคคลปกติสำหรับผู้บกพร่องทางการพูดและการได้ยินจะสามารถสื่อสารได้โดยการใช้ภาษามือแทนคำพูด คนทั่วไปต้องการสื่อสารกับผู้บกพร่องทางการได้ยินจำเป็นต้องเข้าใจภาษามือ แต่คนทั่วไปนั้นยังไม่มีความรู้ความเข้าใจและยังไม่เข้าใจวิธีการสื่อสารด้วยภาษามือรวมถึงความหมายของภาษามืออย่างชัดเจน ทำให้ยากต่อการช่วยเหลือและการติดต่อสื่อสารกันให้เข้าใจ ผู้บกพร่องทางการได้ยินจึงเป็นกลุ่มคนที่ถูกละเมิดสิทธิ และถูกเลือกปฏิบัติมากที่สุดกลุ่มหนึ่ง นอกจากนี้ยังเป็นกลุ่มคนที่มีปัญหาในการสื่อสารและการรับรู้ข้อมูลข่าวสาร จึงทำให้สังคมเปิดโอกาสให้กับผู้บกพร่องทางการได้ยินน้อย มักมองว่าผู้บกพร่องทางการได้ยินไม่มีความสามารถ ผู้บกพร่องทางการได้ยินจึงจำเป็นต้องใช้ล่ามภาษามือในการรับรู้ข้อมูลข่าวสาร ใช้เป็นสื่อกลางในการสื่อสารให้สังคมเข้าใจ วัฒนธรรมของผู้บกพร่องทางการได้ยินที่แสดงออกทั้งสีหน้าและท่าทาง รวมไปถึงการได้รับโอกาสจากสังคมเพื่อจะได้แสดงว่าผู้บกพร่องทางการได้ยินมีความสามารถ มีศักยภาพ ไม่เป็นภาระของสังคม และสามารถพึ่งพาตนเองได้ ในปัจจุบันมีการนำสื่อมัลติมีเดียเข้ามามีบทบาทในชีวิตประจำวันมากขึ้น ซึ่งหากมีโปรแกรมที่สามารถช่วยแปลความหมายของภาษามือไทยผ่านเทคโนโลยีการประมวลผลด้วยภาพ อาจจะช่วยให้ผู้บกพร่องทางการได้ยินสามารถสื่อสารกับคนปกติได้สะดวกและเกิดความเข้าใจที่ตรงกันมากขึ้น

เมื่อกับท่าทางของมนุษย์ มีบทบาทในการสื่อสาร ตอบโต้และกิจกรรมที่ทำในชีวิตประจำวัน การใช้เทคโนโลยีการเรียนรู้เชิงลึก (Deep Learning) ได้ถูกนำมาประยุกต์ใช้ในการตรวจจับมือและท่าทาง โดยการเริ่มต้นของกระบวนการนี้คือเครื่องมือที่ใช้ในการตรวจจับเพื่อเก็บข้อมูลที่เป็นภาพหรือตัวเลขให้อยู่ในรูปแบบที่สามารถนำไปใช้ในโมเดลเพื่อจำแนก และวินิจฉัยท่าทางที่ผู้ต้องการสื่อสาร งานวิจัยที่มีการพูดถึงเครื่องมือที่ใช้ในการตรวจจับและท่าทาง เช่น การรู้จำภาษามือไทยโดยใช้โครงข่ายประสาทเทียม (Chaikaew *et al.*, 2021) ระบบการรู้จำท่าทางมือที่ปรับปรุงใหม่โดยใช้จุดสำคัญและกรอบมือ และการรู้จำภาษามือท้องถิ่นตามเวลาจริงโดยใช้ MediaPipe (Halder and Tayade, 2021) เป็นต้น เมื่อได้ข้อมูลจากการใช้เครื่องมือในการตรวจจับ ต้องมีการเตรียมข้อมูล (Preprocess) และเรียงลำดับให้พร้อมก่อนนำเข้าสู่ตัวโมเดลโดยใช้โครงข่ายหน่วยความจำระยะสั้นยาว (Long Short-Term Memory: LSTM) ซึ่งจากงานวิจัยพบว่าโมเดลที่ใช้ LSTM BLSTM และ GRU มีความแม่นยำมากกว่า 90% กับข้อมูลประเภทจุดสำคัญและมีการฝึกโมเดลที่ไวขึ้นการจดจำภาษามือโดยอัตโนมัติ งานวิจัยที่กล่าวถึงโครงข่ายหน่วยความจำระยะสั้นยาว เช่น แบบฝึกทักษะพื้นฐานแบบมีต้นตอโดยใช้การเคลื่อนไหวการตรวจจับท่าทางจากการฝังวิดีโอท่าทางและเทคนิคการเรียนรู้แบบ One-Shot (Promrit and Waijanya, 2019) การจดจำภาษามือโดยอัตโนมัติ (Venugopalan and Reghunadhan, 2021) การรู้จำภาษามือแบบอเมริกันสำหรับตัวอักษรโดยใช้ MediaPipe และ LSTM (Sundar and Bagyammal, 2022) การรู้จำภาษามือแบบเรียลไทม์โดยใช้ MediaPipe และการ

เรียนรู้เชิงลึก (Bora *et al.*, 2023) และการจำแนกภาษามืออราบิกแบบเรียลไทม์อัจฉริยะ โดยใช้การเริ่มตามความสนใจและ BiLSTM (Abdul *et al.*, 2021)

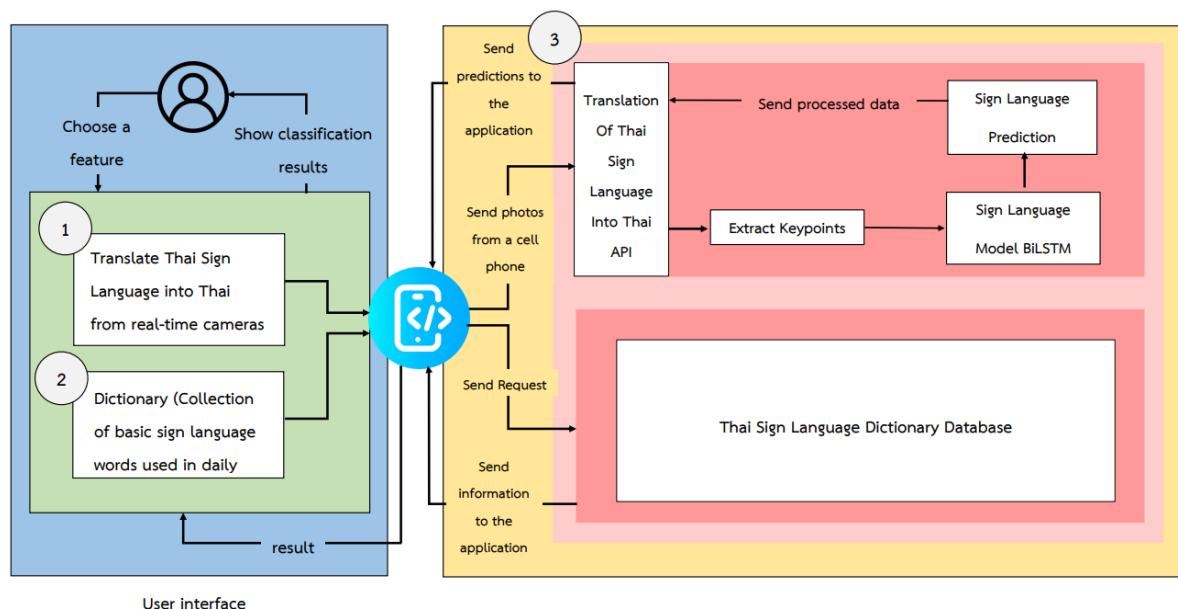
หลังจากการนิยามโมเดล จะเข้าสู่กระบวนการปรับปรุงโมเดลให้เหมาะสมกับข้อมูลของงานวิจัยนี้ด้วยวิธีการเพิ่มชั้น (layer) เพื่อให้ทำงานออกมาได้แม่นยำยิ่งขึ้น มีงานวิจัยที่เกี่ยวข้อง คือ ระบบรู้จำภาษามือสำหรับสื่อสารกับผู้บกพร่องทางการได้ยิน (Obi *et al.*, 2023) และนำมาพัฒนาให้เป็นแอปพลิเคชันบนมือถือเพื่อให้ผู้ใช้ประกอบชีวิตประจำวันได้สะดวกยิ่งขึ้น เช่น สอนภาษามือจีนด้วยสมาร์ตโฟน (Zhang *et al.*, 2021) และ การพัฒนาแอปพลิเคชันระบบการแปลภาษามือให้กับผู้พิการทางการได้ยิน (Chumchim and Maneerat, 2020)

ในบทความนี้ จึงเลือกนำเทคโนโลยีต่างๆ มาแก้ปัญหาที่เกิดขึ้น และเป็นที่มาของโครงงาน “การพัฒนาระบบตรวจจับและแปลงภาษามือไทยเป็นภาษาไทยโดยใช้การเรียนรู้เชิงลึก” โดยใช้อัลกอริทึมโครงข่ายประสาทเทียมหน่วยความจำระยะสั้นยาว (LSTM) ในการวิเคราะห์ท่าทาง และสร้างแอปพลิเคชันที่สามารถตรวจจับภาษามือไทยได้แบบ Real-Time และมีพจนานุกรมภาษามือไทย โดยมีแหล่งความรู้ คือ ฐานข้อมูลจากเว็บไซต์ของโรงเรียนเศรษฐเสถียร เพื่อให้คนทั่วไปสามารถสื่อสารกับผู้พิการทางการได้ยินและการพูดได้ทันที นอกจากนี้ยังสามารถเรียนรู้การทำท่าทางภาษามือที่ใช้ในชีวิตประจำวันได้อีกด้วย

วิธีการดำเนินการวิจัย

1. ภาพรวมของงานวิจัย

เพื่อแปลงภาษามือไทยเป็นภาษาไทยโดยสามารถรับภาพภาษามือจากผู้ใช้ได้ ผู้วิจัยจึงได้ออกแบบภาพรวมของงานวิจัยนี้โดยมีองค์ประกอบหลัก 2 ส่วน คือ 1) ส่วนติดต่อผู้ใช้ และ 2) ส่วนประมวลผลภาพ และแปลภาษามือ ซึ่งองค์ประกอบย่อยๆ ของทั้ง 2 ส่วน และการเชื่อมโยงทั้ง 2 ส่วน แสดงเป็นภาพรวม ดังรูปที่ 1



รูปที่ 1 ภาพรวมของงานวิจัย

จากรูปที่ 1 ส่วนติดต่อผู้ใช้ ประกอบด้วย feature คือ พจนานุกรมภาษามือไทยพื้นฐาน การค้นหาคำศัพท์ และการแปลงภาษามือไทยเป็นภาษาไทย โดยพจนานุกรมภาษามือไทย มีการทำงาน คือ ดึงข้อมูลคำศัพท์ภาษามือจากฐานข้อมูล ส่วนของการแปลงภาษามือไทยเป็นภาษาไทย เป็นส่วนที่นำภาพที่ได้จากกล้องโทรศัพท์มือถือแล้วส่งผ่าน API ไปยังส่วนที่ 2

ส่วนประมวลผลภาพ และแปลภาษามือ ประกอบด้วยส่วนย่อย 3 ส่วน คือ 1) การรับภาพจากกล้องโทรศัพท์มือถือ มีการแสดง Key point ของมือ ท่าทาง และใบหน้า จากนั้นระบบจะบันทึกพิกัด Key point ลงในไฟล์เตอร์ 2) การทำนายจากพิกัดที่ถูกบันทึกลงในไฟล์เตอร์ โดยจะทำนายพิกัดตลอดเวลาที่มีการเปิดกล้องโทรศัพท์มือถือ ซึ่งการทำนายนี้จะได้ผลออกมาเป็นคำ 3) การแสดงผลเป็นภาษาไทย หลังจากมีการทำนายพิกัดและได้ผลลัพธ์ออกมา จะนำมาแสดงในส่วนติดต่อผู้ใช้งาน

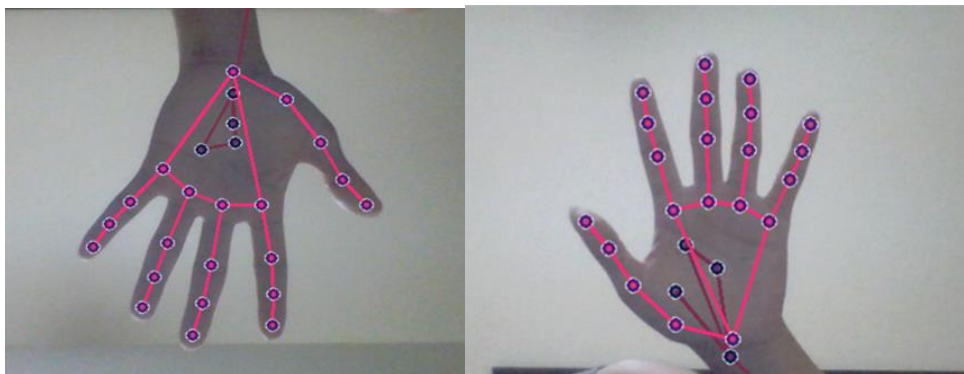
2. รายละเอียดการดำเนินงาน

การใช้พจนานุกรมภาษามือไทย

งานวิจัยนี้จะดึงข้อมูลเกี่ยวกับคำศัพท์ภาษามือไทยพื้นฐานจากฐานข้อมูล โดยข้อมูลที่เก็บในฐานข้อมูลมีชื่อท่าทางภาษามือ วิดีโอแสดงท่าทางภาษามือรวบรวมมาจากฐานข้อมูลของเว็บไซต์โรงเรียนเศรษฐเสถียร และคำอธิบายการทำท่าทางภาษามือรวบรวมมาจากหนังสือภาษามือเชิงปฏิบัติการ 8 เพื่อนำมาแสดงในส่วนติดต่อผู้ใช้งานแอปพลิเคชัน

เตรียมข้อมูลฝึกโมเดล

บทความนี้ใช้ข้อมูลเข้าเป็นพิกัดจุดสำคัญ (Keypoints) ของใบหน้า ลำตัว และมือทั้งสองข้างเพื่อใช้ในการฝึกฝนโมเดล เริ่มต้นจากการรับข้อมูลเข้าโดยใช้กล้องโน้ตบุ๊กของผู้ใช้ของผู้ใช้ผ่านไลบรารี MediaPipe Holistic สามารถตรวจจับพิกัดจุดสำคัญของใบหน้า ลำตัว และมือทั้งสองข้างของมนุษย์เป็นแบบเรียลไทม์ในเวลาเดียวกัน เมื่อโปรแกรมตรวจจับภาพของผู้ใช้ได้แล้วจะแสดงจุดสำคัญ 543 จุด ประกอบด้วย ส่วนของใบหน้า 468 จุด ลำตัว 33 จุด และมือข้างละ 21 จุด จากนั้นนำมาสกัดหาพิกัดจุดสำคัญ จะได้ผลลัพธ์ออกมาเป็นพิกัด x y และ z (ค่าความลึก) ของจุดนั้นๆ โดยใช้พิกัดจุดสำคัญที่ผู้วิจัยรวบรวมจากวิดีโอท่าทางภาษามือไทยจำนวน 100 ท่า ประกอบด้วยท่านิ่ง 16 ท่า และท่าเคลื่อนไหว 84 ท่า โดยมีแหล่งความรู้เป็นฐานข้อมูลจากเว็บไซต์ของโรงเรียนเศรษฐเสถียร



รูปที่ 2 ตัวอย่างการแสดง key point มือขวาและมือซ้ายของ MediaPipe Holistic

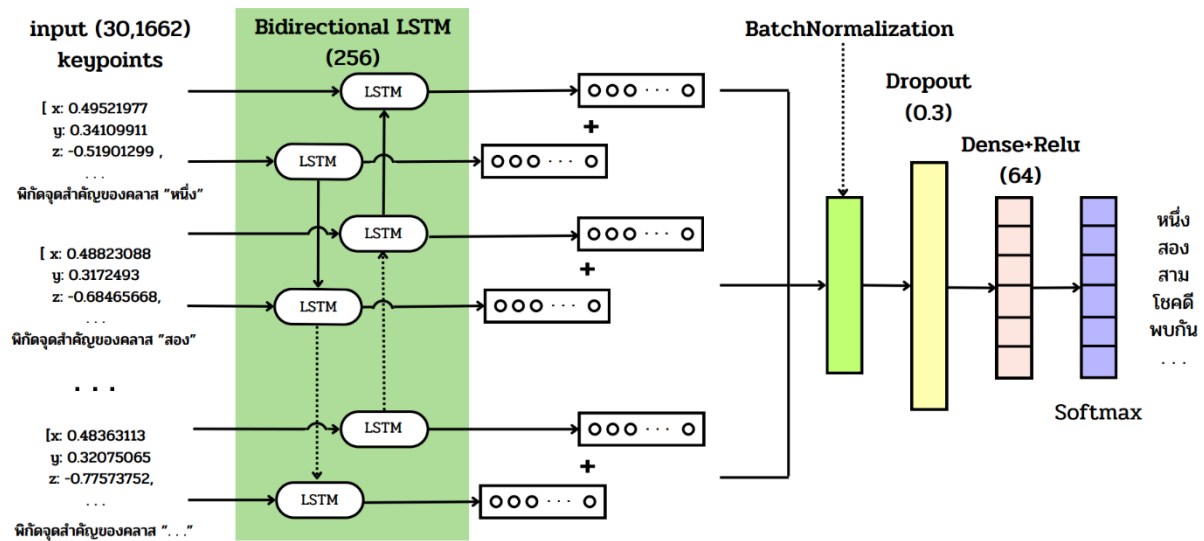
การรวบรวมข้อมูล

ผู้วิจัยได้บันทึกวิดีโอสำหรับการฝึกสอนโมเดล (มีการเรียกใช้ MediaPipe Holistic) ในหนึ่งท่าจะมี 60 คลิป โดย 1 คลิปมีทั้งหมด 30 frame บันทึกเป็นไฟล์ numpy ในไฟล์จะเก็บพิกัดของ Keypoint ท่าภาษามือนั้นๆ เอาไว้ ในหนึ่งไฟล์ หรือ 1 frame จะมีพิกัด Keypoint บันทึกไว้รวม 1,662 ค่า

การเตรียมข้อมูลและการนิยามโมเดล

ในขั้นตอนนี้เริ่มจากการอ่านไฟล์ numpy จากที่บันทึกไว้ในขั้นตอนที่ 2 เตรียมโดยการนำพิกัดของท่าที่บันทึกไว้มารวมเป็น array ก้อนเดียวกัน โดยก่อนเริ่มฝึกฝนโมเดล ในการนิยามโมเดลได้นำข้อมูลที่ผ่านขั้นตอนการเตรียมข้อมูลมาแบ่งเป็นชุดข้อมูลสำหรับฝึกฝน (Train set) 80% และชุดข้อมูลทดสอบ (Test set) 20% จากนั้นเข้าสู่กระบวนการ Data augmentation โดยใช้ SMOTETomek ให้สุ่มสร้างข้อมูลในแต่ละคลาสตามที่เรากำหนดเพื่อเพิ่มความหลากหลายของข้อมูล

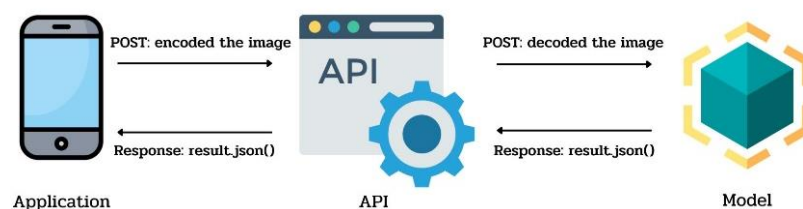
ที่จะนำไปฝึกฝนช่วยลดการเกิดปัญหา Overfitting ของโมเดล จากนั้นดำเนินการสร้างโมเดลเป็นโครงข่ายหน่วยความจำระยะสั้นยาว (LSTMs) แบบ Bidirectional ดังรูปที่ 3



รูปที่ 3 bidirectional long short-term memory (BiLSTM) model

จากรูปที่ 3 กำหนด input ที่ (30,1662) มีขนาด 128 โหนด (node) เพื่อป้องกันการเกิดปัญหา Underfitting ใช้ฟังก์ชัน Activate ReLU หลังจากนั้นทำ BatchNormalization เพื่อปรับข้อมูลให้อยู่ในขอบเขต ถัดมาเป็น Dense Layer มีขนาด 64 โหนด ใช้ฟังก์ชัน Activate ReLU ตามด้วย Dropout ที่ 3 เพื่อสุ่มปิดการทำงานบางโหนด และ Output Layer ใช้ softmax เป็น activation function เพื่อหาความน่าจะเป็นของแต่ละ class ใช้ Optimizer Adam กับค่า Learning Rate เท่ากับ 0.00001

การทำงานส่วน API



รูปที่ 4 การทำงานของ API

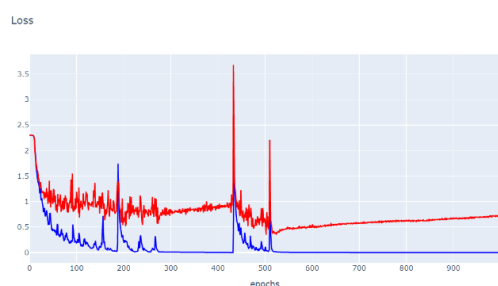
จากรูปที่ 4 แอปพลิเคชันทำหน้าที่จับภาพจากกล้องของโทรศัพท์มือถือทุกๆ 100 มิลลิวินาที รูปภาพที่ได้นั้นนำมาเข้ารหัสรูปภาพโดยใช้วิธีการแปลงรูปภาพเป็นฟอร์แมตที่เหมาะสมสำหรับการส่งผ่าน HTTP โดยใช้ Base64 encoding ซึ่งเป็นวิธีที่คล้ายกับการแปลงข้อความพื้นฐานของ 64 อักขระ การใช้ Base64 จะแปลงข้อมูลให้เป็นชุดของตัวอักษรและตัวเลขที่ไม่สามารถอ่านออกได้โดยตรง โดยรูปภาพจะถูกแปลงเป็นชุดของตัวอักษร และตัวเลข ซึ่งจะข้อมูลในรูปแบบ JSON ส่งผ่าน HTTP Request ไปยัง API เมื่อ API รับภาพจากแอปพลิเคชันแล้ว การถอดรหัสภาพจะถูกทำโดยการแปลงสตริง Base64 เป็นไบนารีภาพต้นฉบับ จะส่งต่อรูปภาพไปยังโมเดลเพื่อทำนายผลลัพธ์ เมื่อโมเดลทำนายผลลัพธ์ของภาพเสร็จสิ้น ผลลัพธ์นั้นจะถูกส่งกลับไปยัง API ในรูปแบบของ JSON โดยผลลัพธ์จะเป็นข้อมูลสตริงที่แสดงผลการทำนายของภาพ หลังจากนั้น API จะนำผลลัพธ์นี้ไปส่งกลับให้แอปพลิเคชันผ่านทาง HTTP response ในรูปแบบ JSON ที่ประกอบด้วยข้อมูลผลลัพธ์ของโมเดล โดยแอปพลิเคชันจะนำผลลัพธ์ที่ได้รับมาแสดงผลและแสดงข้อความสำหรับผู้ใช้งาน

ผลการวิจัยและวิจารณ์ผล

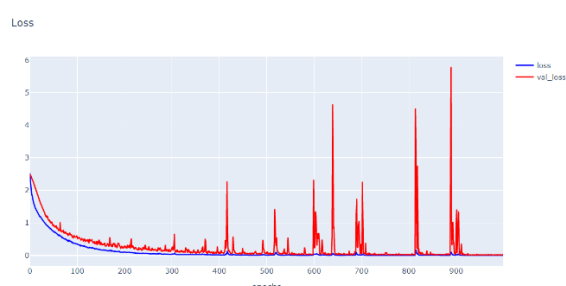
งานวิจัยนี้ทดสอบการทำงานของโมเดลโดยใช้ 10 ท่าทาง ได้แก่ ท่าเคลื่อนไหว 7 ท่า คือ ท่าไม ท่าไหน โชคดี ไข่ไม้ไข่ ป่วย พบ และภาพนิ่ง 3 ท่า ที่มีความหมายถึงการนับเลข คือ หนึ่ง สอง สาม นำข้อมูลที่ได้จากขั้นตอนการเตรียมข้อมูล แบ่งข้อมูลสำหรับฝึกฝน (Train Set) 80% และชุดข้อมูลสำหรับทดสอบ (Test Set) 20% มี 600 วิดีโอแบ่งออกเป็น 480 คำ สำหรับชุดข้อมูลการฝึกอบรม และ 120 คำ สำหรับชุดข้อมูลการตรวจสอบ จากนั้นทดลองฝึกโมเดลแบบ LSTM เปรียบเทียบกับการปรับเปลี่ยนเป็น Model แบบ BiLSTM โดยในการประเมินประสิทธิภาพของโมเดล ผู้วิจัยประเมินโดยแปรผลจาก Loss Value Accuracy Confusion Matrix และสร้างโมบายแอปพลิเคชันเพื่อทดลองการนำโมเดลไปใช้ โดยมีรายละเอียดดังนี้

การประเมินประสิทธิภาพของโมเดลโดยใช้ Loss Value

ผลลัพธ์จากการทดสอบการปรับเปลี่ยนชั้นและพารามิเตอร์ของโมเดล LSTM กับโมเดล BiLSTM พบว่า Loss Value ที่ได้จากการฝึกโมเดล 1000 รอบ คือ 0.77 และ 0.0125 ตามลำดับ สามารถสรุปได้ว่าผลลัพธ์ของโมเดล BiLSTM มี Loss Value ที่ดีกว่า LSTM ดังรูปที่ 5 (ก) และ (ข)



โมเดลแบบ LSTM (ก)

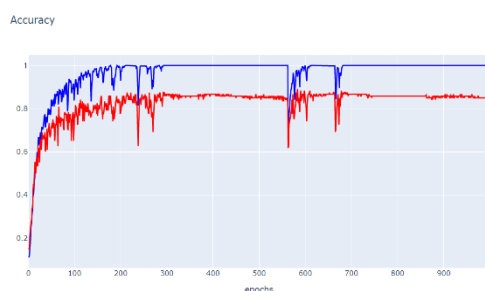


โมเดลแบบ BiLSTM (ข)

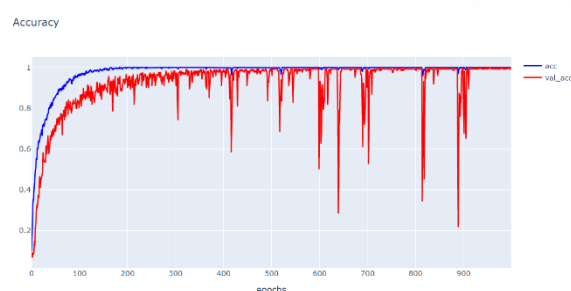
รูปที่ 5 Loss value ของโมเดลแบบ LSTM เทียบกับ BiLSTM

การประเมินประสิทธิภาพของโมเดลโดยใช้ Accuracy Value

ผลลัพธ์จากการทดสอบการปรับเปลี่ยนชั้นและพารามิเตอร์ของโมเดล LSTM กับโมเดล BiLSTM พบว่า Accuracy Value ที่ได้จากการฝึกโมเดล 1000 รอบ คือ 0.83 และ 0.9117 ตามลำดับ สามารถสรุปได้ว่าผลลัพธ์ของโมเดล BiLSTM มี Accuracy Value ที่ดีกว่า LSTM ดังรูปที่ 6 (ก) และ (ข)



โมเดล แบบ LSTM (ก)



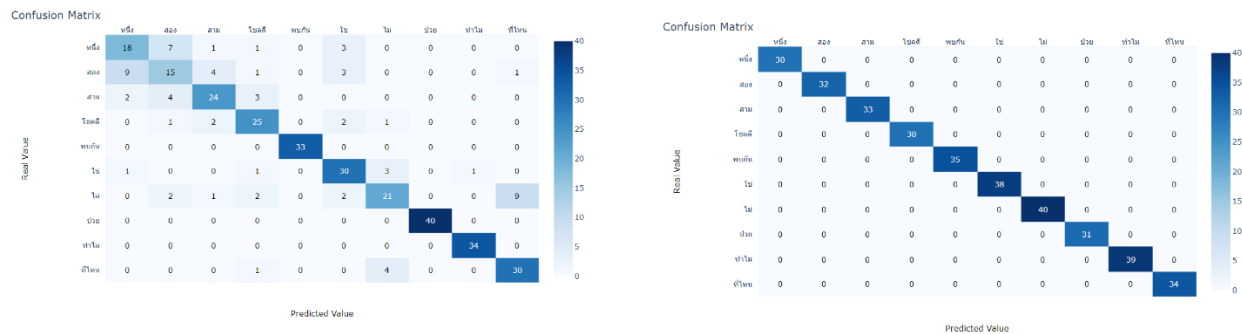
โมเดล แบบ BiLSTM (ข)

รูปที่ 6 Accuracy Value ของโมเดลแบบ LSTM เทียบกับ BiLSTM

การประเมินประสิทธิภาพของโมเดลโดยใช้ Confusion Matrix

ผู้วิจัยใช้ Confusion Matrix ที่แสดงถึงความแม่นยำของแต่ละ Class ดังรูปที่ 7 มาใช้คำนวณหาค่าที่ใช้วัดประสิทธิภาพโมเดลทั้งหมด 4 ค่า ได้แก่ ค่าความถูกต้อง (Accuracy) ค่าความแม่นยำ (Precision) ค่าความระลึก (Recall) และ ค่า F1-Score ซึ่งพบว่าโมเดลที่มีค่า Precision Recall และ F1-Score มากที่สุด คือ BiLSTM ด้วยค่า 0.92 0.91

0.91 และ 0.91 ตามลำดับ ทำให้สามารถสรุปได้ว่าโมเดลที่ดีที่สุด ได้แก่ โมเดล BiLSTM ที่มีการปรับใช้พารามิเตอร์และเทคนิค Augmentation Batch Normalization และ Dropout โดยผลการคำนวณ ค่า Precision Recall และ F1-Score แสดงดังตารางที่ 1 ด้วยเหตุนี้ผู้วิจัยจึงเลือกโมเดลนี้เพื่อใช้ทำนายภาษามือไทย ให้ได้ผลลัพธ์จากการทำนายที่มีประสิทธิภาพมากที่สุด



โมเดล แบบ LSTM (ก)

โมเดล แบบ BiLSTM (ข)

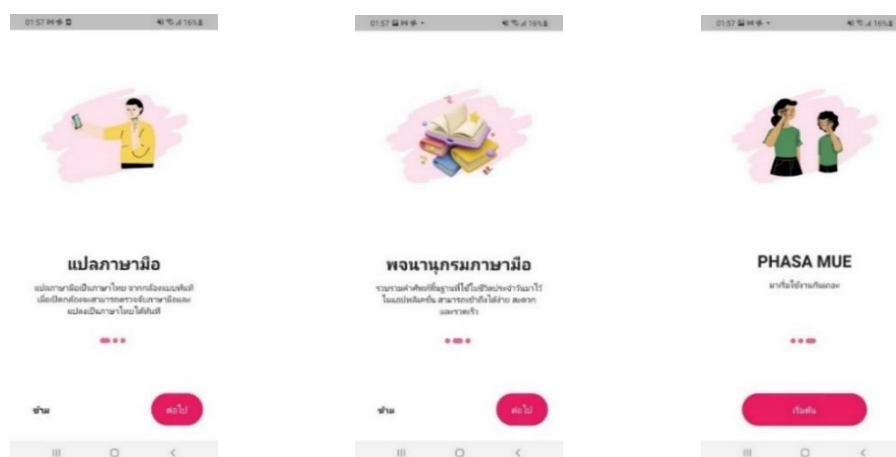
รูปที่ 7 Confusion Matrix ของโมเดลแบบ LSTM เทียบกับ BiLSTM

ตารางที่ 1 ผลลัพธ์จากการคำนวณค่า Precision Recall F1-Score Accuracy และ Loss ของแต่ละโมเดล

Model	Precision	Recall	F1-Score	Accuracy	Loss
LSTM	0.89	0.85	0.85	0.83	0.77
BiLSTM	0.92	0.91	0.91	0.91	0.01

การทำงานส่วนโมบายแอปพลิเคชัน

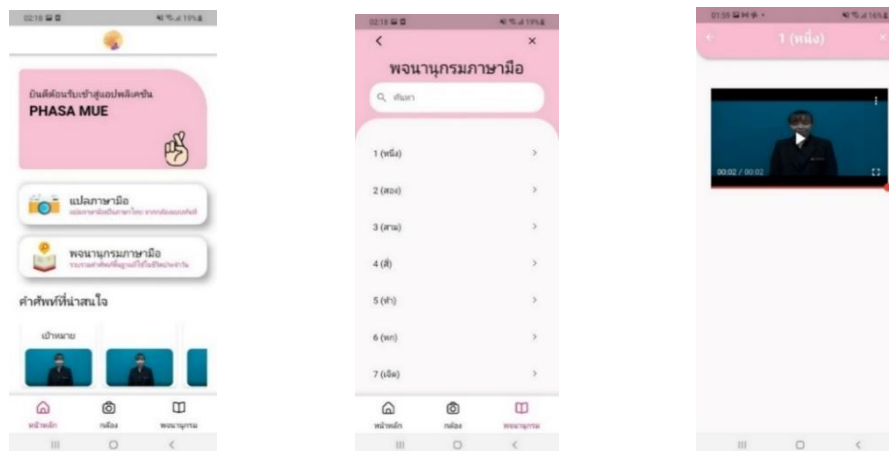
เพื่อทดลองนำโมเดลไปใช้งานจริง ผู้วิจัยจึงได้สร้างโมบายแอปพลิเคชันขึ้น โดยมีผลการทำงานของแอปพลิเคชัน คือ เมื่อเปิดใช้งานแอปพลิเคชันจะเข้าสู่หน้าจออธิบายวิธีการใช้งานแอปพลิเคชัน ดังรูปที่ 8



รูปที่ 8 หน้าจออธิบายวิธีการใช้แอปพลิเคชัน

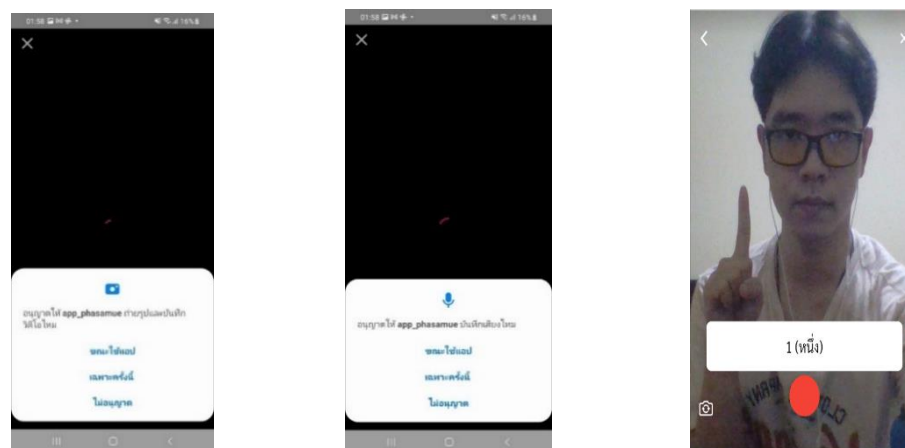
จากนั้นจะเข้าสู่หน้าหลัก ในหน้าหลักแสดงปุ่มที่สามารถกดไปยังหน้ากล้องเพื่อให้ในการตรวจจับท่าทางภาษามือ และมีปุ่มที่กดไปยังหน้าพจนานุกรมภาษามือไทย นอกจากนี้ยังแสดงท่าทางภาษามือที่สุ่มมาจากฐานข้อมูล ดังตัวอย่างในรูปที่ 9 (ก) โดยถ้าผู้ใช้กดไปยังหน้าพจนานุกรมภาษามือไทย จะแสดงท่าทางภาษามือไทยมากกว่า 300 คำ สามารถค้นหาท่าทาง

ภาษามือได้ด้วยชื่อท่าทางภาษามือ ดังรูปที่ 9 (ข) และถ้าผู้ใช้กดไปที่คำศัพท์จะแสดงชื่อของท่าทางภาษามือ และแสดงวิดีโอการทำท่าทางภาษามือ ดังรูปที่ 9 (ค)



หน้าหลัก (ก) หน้าพจนานุกรมภาษามือไทย (ข) หน้าแสดงรายละเอียดคำศัพท์ (ค)
รูปที่ 9 การทำงานส่วนของแอปพลิเคชัน และพจนานุกรม

จากรูปที่ 9 (ก) ซึ่งเป็นหน้าหลัก หากผู้ใช้กดปุ่มไปยังหน้ากล้องเพื่อให้ในการตรวจจับท่าทางภาษามือ จะมีข้อความขออนุญาตในการใช้งานกล้องและไมโครโฟน เมื่อผู้ใช้งานอนุญาตให้มีการใช้งาน แอปพลิเคชันจะสตรีมกล้องไปที่ API แล้วส่งภาพไปที่โมเดล ซึ่งโมเดลจะตรวจจับภาษามือ จากนั้นจะส่งชื่อท่าของภาษามือมาที่ API สุดท้ายจะส่งมาที่แอปพลิเคชันเพื่อแสดงผล ดังรูปที่ 10



รูปที่ 10 หน้าจอการแปลภาษามือเป็นภาษาไทยผ่านแอปพลิเคชันที่เชื่อมต่อกับโมเดล

สรุปผลการวิจัย

บทความนี้นำเสนอเรื่อง การพัฒนาระบบตรวจจับและแปลงภาษามือไทยเป็นภาษาไทยโดยใช้การเรียนรู้เชิงลึก โดยรวบรวมข้อมูลเป็นพิกัดจุดสำคัญ (Keypoints) ของใบหน้า ลำตัว และมือทั้งสองข้าง ใช้กล้องคอมพิวเตอร์และใช้ MediaPipe Holistic ที่เป็นไลบรารีที่สามารถตรวจจับพิกัดจุดสำคัญของใบหน้า ลำตัว และมือทั้งสองข้างของมนุษย์เป็นแบบเรียลไทม์ในเวลาเดียวกัน เพื่อตรวจจับจุดสำคัญ เมื่อโปรแกรมตรวจจับภาพของผู้ใช้ได้แล้วจะแสดงจุดสำคัญ 543 จุด ประกอบด้วยส่วนของใบหน้า 468 จุด ลำตัว 33 จุด และมือข้างละ 21 จุด ถัดมาสาธิตหาพิกัดจุดสำคัญ จะได้ผลลัพธ์ออกมาเป็นพิกัด x y และ z (ค่าความลึก) ของจุดนั้นๆ โดยเก็บรวบรวมพิกัดจุดสำคัญจากวิดีโอท่าทางภาษามือไทยที่ผู้วิจัยเป็นคนบันทึก

ทั้งหมด 100 ท่า ประกอบด้วยท่าหนึ่ง 16 ท่า และท่าเคลื่อนไหว 84 ท่า มีแหล่งความรู้เป็นฐานข้อมูลจากเว็บไซต์ของโรงเรียน เศรษฐศาสตร์และผู้เชี่ยวชาญให้คำแนะนำอย่างใกล้ชิด เพื่อตรวจสอบความถูกต้องของการทำท่าทางภาษามือ แบ่งชุดข้อมูล สำหรับฝึกฝน (Train set) 80% และชุดข้อมูลสำหรับทดสอบ (Test set) 20% จากนั้นนำข้อมูลมาฝึกฝนโมเดล BiLSTM ที่มีการปรับใช้พารามิเตอร์และเทคนิค Augmentation Batch Normalization และ Dropout จากการทดลองพบว่าโมเดลที่มีการเพิ่มโหนด (node) และปรับเปลี่ยนชั้น (layer) ทำให้โมเดลสามารถป้องกันการเกิดปัญหา Underfitting และพารามิเตอร์ที่ใช้ในการเรียนรู้เพิ่มขึ้น ได้ค่าความถูกต้องคือ 0.9117 ซึ่งผลลัพธ์ที่ได้จากการประมวลผลของโมเดลจะถูกแสดงผลบนแอปพลิเคชัน PHASA MUE โดยเชื่อมการทำงานระหว่างโมเดลและแอปพลิเคชันด้วย API ในอนาคตสามารถนำงานวิจัยนี้ไปพัฒนาให้มีความแม่นยำในการทำนายมากขึ้น เพื่อให้สื่อสารสะดวกยิ่งขึ้นระหว่างคนทั่วไปกับผู้บกพร่องทางการได้ยินและการพูด และยังสามารถนำไปประยุกต์ใช้ในสถานศึกษาที่นักศึกษาเป็นผู้พิการทางการได้ยินได้ หรือนำไปติดตั้งในองค์กรหรือสถานที่ที่ต้องการบริการผู้บกพร่องทางการได้ยินและการพูด เพื่ออำนวยความสะดวกในการใช้บริการต่างๆ ของผู้บกพร่องทางการได้ยินและการพูดได้มากยิ่งขึ้น

เอกสารอ้างอิง

- Abdul, W., Alsulaiman, M., Amin, S.U., Faisal, M., Muhammad, G., Albogamy, F.R., Bencherif, M.A. and Ghaleb, H. (2021). Intelligent real-time Arabic sign language classification using attention-based inception and BiLSTM. *Computers and Electrical Engineering* 95: 107 - 395. doi: 10.1016/j.compeleceng.2021.107395.
- Bora, J., Dehingia, S., Boruah, A., Chetia, A.A. and Gogoi, D. (2023). Real-time Assamese Sign Language Recognition using MediaPipe and Deep Learning. *Procedia Computer Science* 218: 1384 - 1393. doi: 10.1016/j.procs.2023.01.117.
- Chaikaew, A., Somkuan, K. and Yuyen, T. (2021). Thai Sign Language Recognition: an Application of Deep Neural Network. In: 2021 Joint International Conference on Digital Arts, Media and Technology with ECTI Northern Section Conference on Electrical, Electronics, Computer and Telecommunication Engineering. IEEE, Cha-am. 128 - 131. doi: 10.1109/ECTIDAMTNCN51128.2021.9425711.
- Chumchim, P. and Maneerat, P. (2020). Application development of sign language translation systems for the hearing-impaired people. *PKRU SciTech Journal* 4(1): 22 - 32.
- Halder, A. and Tayade, A. (2021). Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. *International Journal of Research Publication and Reviews* 2(5): 9 - 17.
- Obi, Y., Claudio, K.S., Budiman, V.M., Achmad, S. and Kurniawan, A. (2023). Sign language recognition system for communicating to people with disabilities. *Procedia Computer Science* 216: 13 - 20. doi: 10.1016/j.procs.2022.12.106.
- Promrit, N. and Waijanya, S. (2019). Model for Practice Badminton Basic Skills by using Motion Posture Detection from Video Posture Embedding and One-Shot Learning Technique. In: *Proceedings of the 2019 2nd Artificial Intelligence and CloudComputing Conference*. Kobe, Japan. 117 - 124. doi: 10.1145/3375959.3375981.

- Venugopalan, A. and Reghunadhan, R. (2021). Applying deep neural networks for the automatic recognition of sign language words: A communication aid to deaf agriculturists. *Expert Systems with Applications* 185: 115601. doi: 10.1016/j.eswa.2021.115601.
- Sundar, B. and Bagyammal, T. (2022). American Sign Language Recognition for Alphabets Using MediaPipe and LSTM. *Procedia Computer Science* 215: 642 - 651. doi: 10.1016/j.procs.2022.12.066.
- Zhang, Y., Min, Y. and Chen, X. (2021). Teaching Chinese Sign Language with a Smartphone. *Virtual Reality & Intelligent Hardware* 3(3): 248 - 260. doi: 10.1016/j.vrih.2021.05.004.

