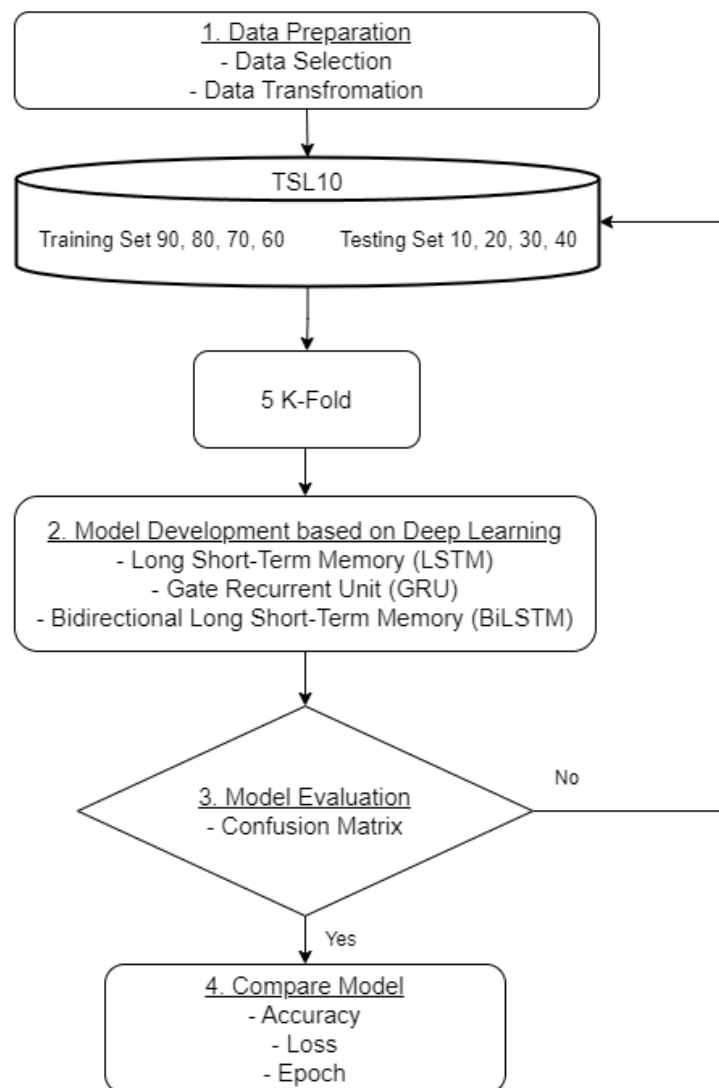


บทที่ 3

วิธีการดำเนินงาน

สำหรับวิธีการดำเนินการวิจัยการพัฒนากระบวนการรู้จำท่าทางภาษามือไทยด้วยโครงข่ายประสาทเทียมแบบวนกลับ มีขั้นตอนดังภาพที่ 3.1

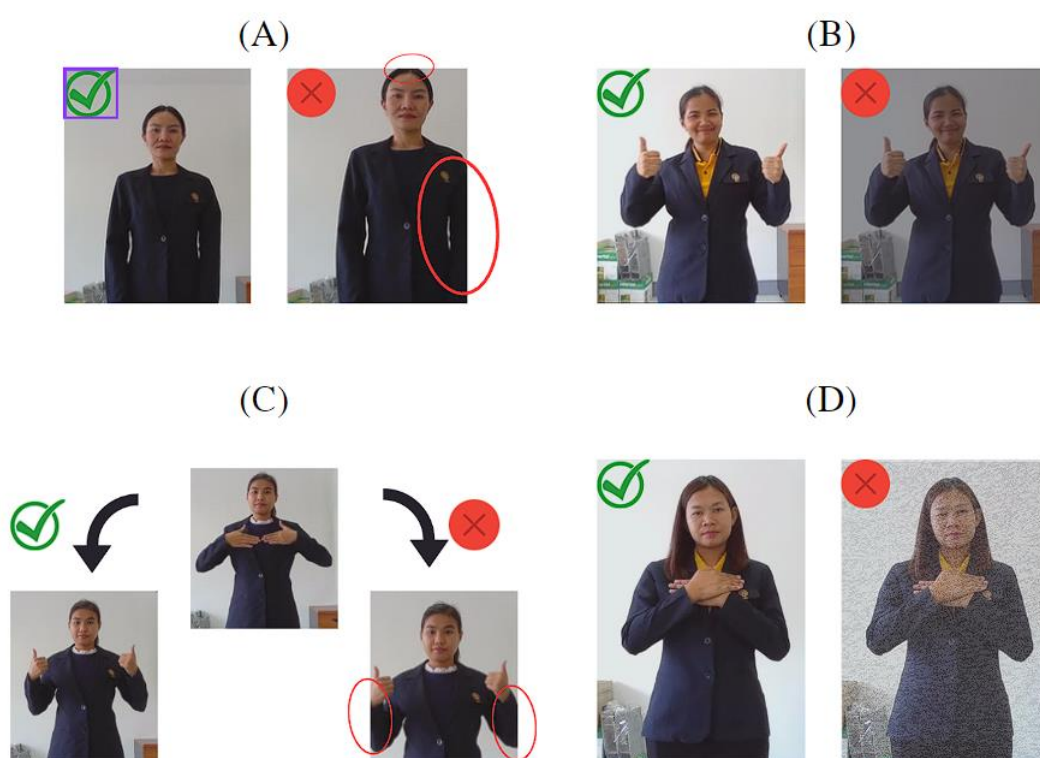


ภาพที่ 3.1 กรอบการดำเนินการวิจัย

3.1 การเตรียมข้อมูล

3.1.1 การรวบรวมข้อมูล

ในการเตรียมข้อมูลสำหรับการสร้างระบบรู้จำท่าทางภาษามือไทยด้วยโครงข่ายประสาทเทียมแบบวนกลับจะเก็บข้อมูลเป็นวิดีโอภาษามือไทยจำนวน 10 คำที่เป็นคำทั่วไปที่ใช้ในชีวิตประจำวันของผู้ที่ใช้ภาษามือในการสื่อสาร โดยจะเก็บวิดีโอต่อคำเป็น 100 วิดีโอต่อ 1 คำและใน 1 วิดีโออัตราเฟรมต่อวินาทีคือ 30 FPS ขนาดของวิดีโอคือ 640 x 480 ระยะของ 1 วิดีโอคือ 1 วินาทีต่อ 1 วิดีโอ โดยอัดวิดีโอจาก Laptop ของผู้วิจัย



ภาพที่ 3.2 ปัจจัยควบคุมในการรวบรวมข้อมูล

- 1) ตัวของผู้ทำท่าทางภาษามือจะต้องอยู่ในเฟรม ดังในภาพที่ 3.1 ในข้อ A
- 2) ในการทำท่าทางจะต้องอยู่ในเฟรมไม่หลุดออกจากเฟรม ดังภาพที่ 3.2 ในข้อ B
- 3) ในการบันทึกวิดีโอแสงจะต้องไม่มีดเกินไป ดังภาพภาพที่ 3.2 ในข้อ C
- 4) คุณภาพของวิดีโอจะต้องมีความละเอียดตั้งแต่ 640 x 480 หรือสูงกว่าสำหรับกระบวนการบันทึกวิดีโอ ดังภาพที่ 3.2 ในข้อ D

ซึ่งคำที่จะใช้ในการวิจัยครั้งนี้ดังตารางที่ 3.1

ตารางที่ 3.1 คำศัพท์ภาษามือที่ใช้ในโครงการ

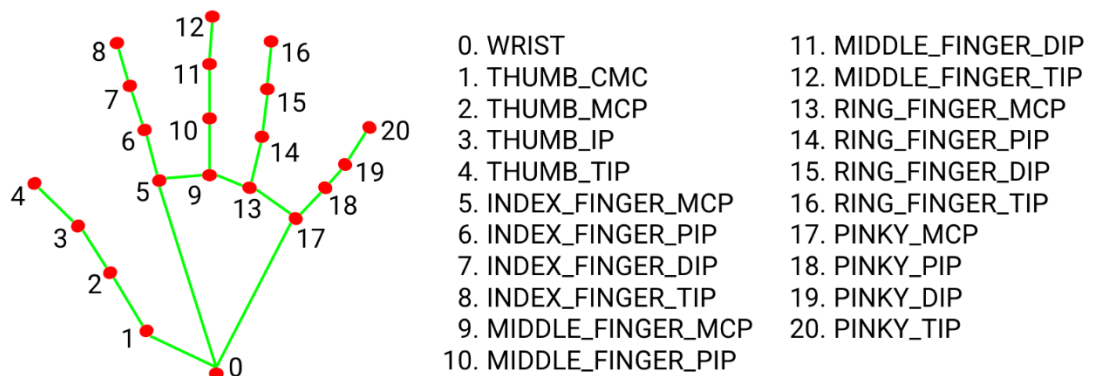
คำภาษาไทย	คำภาษาอังกฤษ	ความหมาย
ขอบคุณ	Thank You	กล่าวแสดงความรู้สึกถึงบุญคุณหรือกล่าวเมื่อได้รับความช่วยเหลือ
ขอโทษ	Sorry	ขอภัยเมื่อได้ทำผิดพลาดอย่างใดอย่างหนึ่ง
ไม่เป็นไร	That is OK	คำแสดงความรู้สึกที่ไม่ได้ถือโทษหรือโกรธเคืองใด ๆ เพื่อให้ผู้ฟังรู้สึกดีขึ้นหรือไม่ต้องรู้สึกผิด
สบายดี	Fine	สภาวะปกติของทั้งร่างกายและจิตใจ ร่างกายไม่เจ็บป่วย รวมทั้งอารมณ์ดี มีความสุข ไม่มีอะไรให้กังวล
ชอบ	Like	พอใจ แสดงอาการพึงพอใจ
รัก	Love	มีใจผูกพันอย่างมาก
ไม่สบาย	Sick	สภาวะที่ร่างกายและจิตใจไม่ปกติ หรือเกิดอาการป่วย
สวัสดี	Hello	ใช้สำหรับการทักทายผู้คน
ฉัน	I (Am)	ใช้สำหรับการเรียกแทนตัวเอง
คุณ	You	ใช้สำหรับเรียกแทนผู้ที่เราพูดด้วย

3.1.2 การแปลงข้อมูล

ในขั้นตอนนี้คือการแปลงข้อมูลเพื่อให้เหมาะสมกับโมเดลที่จะนำไปเทรน ซึ่งก็คือการนำวิดีโอที่ได้จากการรวบรวมข้อมูลมาแปลงใหม่ด้วยการสกัดลักษณะเด่นของวิดีโอภาษามือนั้นขึ้นอยู่กับการใช้มือและท่าทาง ผู้วิจัยจึงได้ใช้เครื่องมือ MediaPipe ที่เป็น Framework ซึ่งวิธีการคือการใช้ MediaPipe ในการสร้าง Key Points ขึ้นตามจุดต่าง ๆ ของร่างกายเป็นค่า มิติ X, Y, Z ของหน้า, มือ และตัว

ในการสกัดลักษณะเด่นส่วนของมือแต่ละข้างนั้นจะได้ 21 Key Points ซึ่ง Key Points จะถูกคำนวณแบบ 3 มิติ X, Y, Z ของมือทั้งสองข้าง โดยจะได้ Key Points จากการสกัดจากมือดังนี้

Keypoints in hand x Three dimensions x No. of hands = $(21 \times 3 \times 2) = 126$ Key Points ดังภาพที่ 3.3

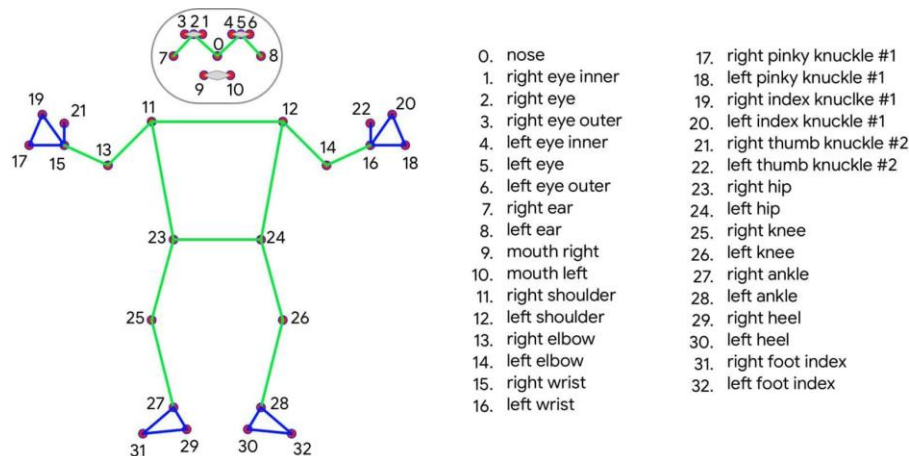


ภาพที่ 3.3 ลำดับและป้ายกำกับ Key Points ของมือใน MediaPipe

ที่มา : MediaPipe (2023: Online)

ในการสกัดลักษณะเด่นส่วนของตัวนั้นจะได้ 33 Key Points คำนวณแบบ 3 มิติ X, Y, Z และเพิ่มค่า Visibility เข้าไปซึ่งเป็นค่าที่จะระบุว่าจุดนั้นมองเห็นหรือซ่อนอยู่ (ที่ถูกปิดโดยจุดอื่นของร่างกาย) บนเฟรมดังนั้นจะได้ค่า Key Points ดังนี้

Key Points in pose x (Three dimenstions + Visibility) = (33 + (33 + 1)) = 132 Key Points ดังภาพที่ 3.4



ภาพที่ 3.4 ลำดับและป้ายกำกับ Key Points ของตัวใน MediaPipe

ที่มา : MediaPipe (2023: Online)

ในการสกัดลักษณะเด่นของใบหน้านั้นจะได้ 468 Key Points ได้แก่ รูปทรงรอบหน้าและหน้า, ตา, ปากและคิ้ว ซึ่งคำนวณค่า 3 มิติ X, Y, Z ได้ดังนี้

Key Points in face x Three dimensions = $(468 \times 3) = 1404$ Key Points ดังภาพที่ 3.5



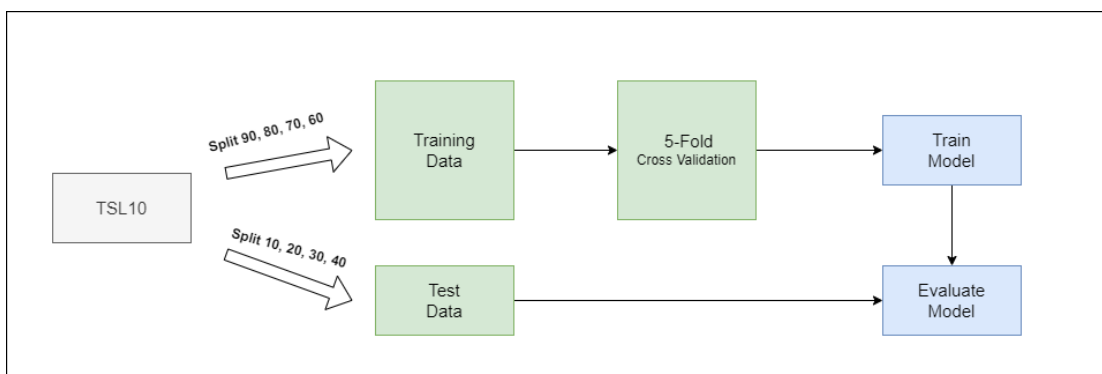
ภาพที่ 3.5 Key points บนหน้า

ดังนั้นเมื่อรวม Key Points ทั้งหมดเข้าด้วยกันไม่ว่าจะเป็นจาก หน้า ตัวและมือจะสามารถคำนวณได้ดังนี้

Key Points in hands + in pose + inface = $(126 + 132 + 1404) = 1662$ Key Points

3.1.3 การแบ่งข้อมูล

ในขั้นตอนนี้ผู้วิจัยจะแบ่งข้อมูลออกเป็น 2 ส่วนในการเทรนและสำหรับการนำไปทดสอบ โดยข้อมูลทั้งหมดคือ 1000 จะทำการแบ่งข้อมูลเป็นอัตราส่วน 60:40, 70:30, 80:20 และ 90:10 แล้วนำข้อมูลข้อมูลเทรนมาทำการแบ่งสำหรับการทำ 5-Fold Cross Validation เพื่อให้โมเดลฝึกฝน ดังภาพที่ 3.6



ภาพที่ 3.6 การแบ่งข้อมูลสำหรับเทรนและทดสอบ

ตารางที่ 3.2 การแบ่งข้อมูลสำหรับเทรนและทดสอบ

อัตราส่วน	ข้อมูลสำหรับเทรน	ข้อมูลสำหรับทดสอบ
60:40	600	400
70:30	700	300
80:20	800	200
90:10	900	100

ตารางที่ 3.3 ข้อมูลสำหรับเทรนใน 5-Fold Cross Validation

อัตราส่วน	5-Fold Cross Validation	ข้อมูลในแต่ละคลาส
60:40	120	12
70:30	140	14
80:20	160	16
90:10	180	18

ในคอลัมน์ 5-Fold Cross Validation คือจำนวนข้อมูลในแต่ละ Fold ของข้อมูลสำหรับเทรน เช่น อัตราส่วน 60:40 มีข้อมูลสำหรับเทรน 600 ข้อมูลเมื่อแบ่งเป็น 5 Fold คือการนำ 600 มาแบ่งออกเป็น 5 ส่วนจึงได้ Fold ละ 120 ข้อมูล ส่วนในคอลัมน์ ข้อมูลในแต่ละคลาส คือจำนวนข้อมูลในแต่ละ Fold มาแบ่งออกให้เท่ากับจำนวนคลาสอย่างในงานวิจัยนี้มีทั้งหมด 10 คลาสจึงแบ่งได้เป็น 12 ข้อมูลต่อ 1 คลาส

3.2 การฝึกฝนโมเดล

ผู้วิจัยได้ใช้โมเดลในการเทรนทั้งหมด 3 โมเดลได้แก่ LSTM, GRU, BiLSTM ในงานวิจัยครั้งนี้ Number of Nodes คือ จำนวนของ Input Node ซึ่งผู้วิจัยกำหนดขั้นต่ำไว้ 64 จนถึง 256 Activation คือตัวฟังก์ชันที่ใช้ในการรับผลรวมจากการประมวลผลทั้งหมดจากทุก Input Node เข้ามาพิจารณาตามกลไกการคำนวณของ Activation Function นั้น ๆ แล้วส่งต่อไปเป็น Output ซึ่งในงานวิจัยนี้ได้เลือกใช้ 2 ตัว คือ Rectified Linear Unit (ReLU) และ Softmax Optimizer คือ อัลกอริทึมการเพิ่มประสิทธิภาพ (Optimizer) ทำหน้าที่เป็นกลไกการปรับปรุงค่าน้ำหนักของตัวแปรต้นต่าง ๆ รวมถึงค่าความคลาดเคลื่อน (Bias) ในงานวิจัยนี้ได้เลือกใช้ได้แก่ Adagrad, Adamax, Adam or RMSprop ดังตารางที่ 3.4

ตารางที่ 3.4 พารามิเตอร์ของเลเยอร์โมเดล

Parameters	Value
RNN Model	GRU, LSTM, BiLSTM
Number of Nodes	Between (64, 256)
Activation	'Relu' or 'Softmax'
Optimizer	'Adagrad', 'Adamax', 'Adam' or 'RMSprop'

3.3 การวัดประสิทธิภาพโมเดล

การวัดประสิทธิภาพของโมเดล ผู้วิจัยได้ใช้ตัวชี้วัดได้แก่ค่า Accuracy คือค่าอัตราความถูกต้องของการทำนายของโมเดลโดยในการวิจัยครั้งนี้ โดยจะนำโมเดลที่มีค่าเทรน Accuracy ที่มากที่สุดของแต่ละโมเดลที่มีการแบ่งข้อมูลเท่ากันและจะทำการทดสอบค่าความถูกต้องในการทำนายของโมเดลที่เทรนด้วยวิธี Confusion Matrix โดยใช้ข้อมูลสำหรับทดสอบที่ได้เตรียมไว้ตั้งแต่ขั้นตอนการแบ่งข้อมูล

3.4 การเปรียบเทียบประสิทธิภาพโมเดล

ในขั้นตอนการเปรียบเทียบประสิทธิภาพ ผู้วิจัยจะนำโมเดลที่ผ่านการเทรนทั้งหมด 3 โมเดล ได้แก่ LSTM, GRU, BiLSTM ซึ่งจะเปรียบเทียบประสิทธิภาพเรื่องของ ค่า Accuracy, ค่า Loss และ จำนวนรอบที่ใช้ในการเทรนโมเดล (epochs)