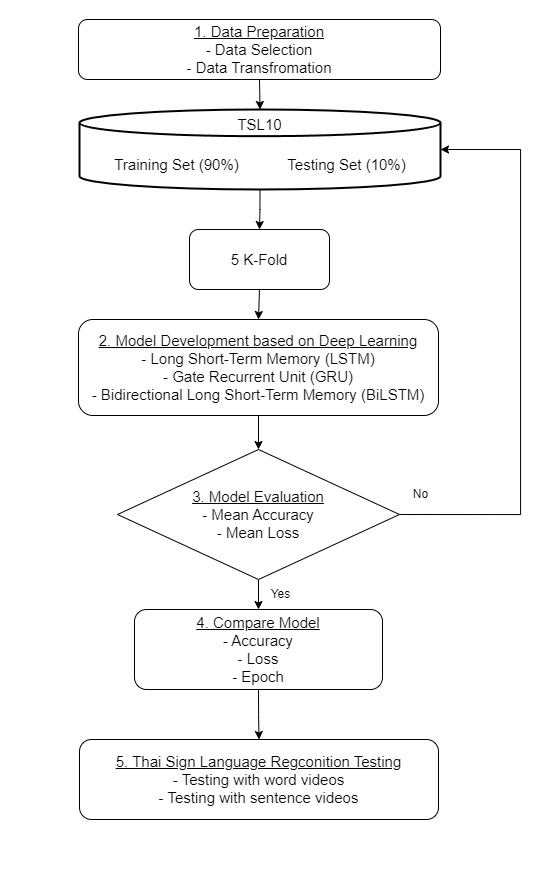
# **บทที่ 3**

# **วิธีดำเนินการวิจัย**

สำหรับวิธีการดำเนินการวิจัยการพัฒนาระบบการรู้จำท่าทางภาษามือไทยด้วยโครงข่ายประสาทเทียงแบบวนกลับ มีขั้นตอนดังภาพที่ 3.1



**ภาพที่ 3.1** กรอบการดำเนินการวิจัย

## **3.1 การเตรียมข้อมูล**

**3.1.1 การรวบรวมข้อมูล**

ในการเตรียมข้อมูลสำหรับการสร้างระบบรู้จำท่าทางภาษามือไทยด้วยโครงข่ายประสาทเทียมแบบวนกลับจะเก็บข้อมูลเป็นวิดีโอภาษามือไทยจำนวน 10 คำที่เป็นคำทั่วไปที่ใช้ในชีวิตประจำวันของผู้ที่ใช้ภาษามือในการสื่อสาร โดยจะเก็บวิดีโอต่อคำเป็น 85 วิดีโอต่อ 1 คำและใน 1 วิดีโออัตราเฟรมต่อวินาทีคือ 30 FPS ขนาดของวิดีโอคือ 640 x 480 ระยะของ 1 วิดีโอคือ 1 วินาทีต่อ 1 วิดีโอ โดยอัดวิดีโอจาก Laptop ของผู้วิจัย

รูปภาพประกอบด้วย เสื้อผ้า, หญิง, ใบหน้าของมนุษย์, คน

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 3.2** ปัจจัยควบคุมในการรวบรวมข้อมูล

1. ตัวของผู้ทำท่าทางภาษามือจะต้องอยู่ในเฟรม ดังในภาพที่ 3.1 ในข้อ A

2. ในการทำท่าทางจะต้องอยู่ในเฟรมไม่หลุดออกจากเฟรม ดังภาพที่ 3.2 ในข้อ B

3. ในการบันทึกวิดีโอแสงจะต้องไม่มืดเกินไป ดังภาพภาพที่ 3.2 ในข้อ C

4. คุณภาพของวิดีโอจะต้องมีความละเอียดตั้งแต่ 640 × 480 หรือสูงกว่าสำหรับกระบวนการบันทึกวิดีโอ ดังภาพที่ 3.2 ในข้อ D

ซึ่งคำที่จะใช้ในการวิจัยครั้งนี้ดังตารางที่ 3.1

**ตารางที่ 3.1** คำศัพท์ภาษามือที่ใช้ในโครงงาน

|  |  |  |
| --- | --- | --- |
| **คำภาษาไทย** | **คำภาษาอังกฤษ** | **ความหมาย** |
| ขอบคุณ | Thank You | กล่าวแสดงความรู้สึกถึงบุญคุณหรือกล่าวเมื่อได้รับความช่วยเหลือ |
| ขอโทษ | Sorry | ขออภัยเมื่อได้ทำผิดพลาดอย่างใดอย่างหนึ่ง |
| 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 | IAm | ใช้สำหรับการเรียกแทนตัวเอง |
| คุณ | You | ใช้สำหรับเรียกแทนผู้ที่เราพูดด้วย |

**3.1.2 การแปลงข้อมูล**

ในขั้นตอนนี้คือการแปลงข้อมูลเพื่อให้เหมาะสมกับโมเดลที่จะนำไปเทรนได้แก่โมเดล ซึ่งก็คือการนำวิดีโอที่ได้จากการรวบรวมข้อมูลมาแปลงใหม่ด้วยการสกัดลักษณะเด่นของวิดีโอเด่นภาษามือนั้นขึ้นอยู่กับการใช้มือและท่าทาง การนำวิดีโอที่เป็นภาษามือมาใช้ในการเทรนโมเดลนั้นจึงเป็นเรื่องยาก ผู้วิจัยจึงได้ใช้เครื่องมือ MediaPipe ที่เป็น Framework มาใช้ในการแก้ปัญหา ซึ่งวิธีการคือการใช้ MediaPipe ในการ Keypoints ขึ้นตามจุดต่าง ๆ ของร่างกายเป็นค่า มิติ X, Y, Z ของหน้า, มือและ

ในมือแต่ละข้างนั้น MediaPipe จะสกัดออกมาได้ 21 Keypoints ซึ่ง Keypoint จะถูกคำนวณแบบ 3 มิติ X, Y, Z ของมือทั้งสองข้าง โดยจะได้ Keypoints จากการสกัดจากมือดังนี้

Keypoins in hand x Three dimensions x No. of hands = (21 x 3 x 2) = 126 Keypoints ดังภาพที่ 3.3

รูปภาพประกอบด้วย ข้อความ, ตัวอักษร, แผนภาพ, ไลน์

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 3.3** ลำดับและป้ายกำกับ Keypoints ของมือใน MediaPipe

**ที่มา :** MediaPipe (2023: Online)

ในส่วนของท่าทางนั้น MediaPipe จะสกัดออกมาได้ 33 Keypoints คำนวณแบบ 3 มิติ X, Y, Z และเพิ่มค่า Visibility เข้าไปซึ่งเป็นค่าที่จะระบุว่าจุดนั้นมองเห็นหรือซ่อนอยู่ (ที่ถูกปิดโดยจุดอื่นของร่างกาย) บนเฟรมดังนั้นจะได้ค่า Keypoints ดังนี้

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

รูปภาพประกอบด้วย ข้อความ, แผนภาพ, โครงกระดูก

คำอธิบายที่สร้างโดยอัตโนมัติ

**ภาพที่ 3.4** ลำดับและป้ายกำกับ Keypoints ของท่าทางใน MediaPipe

**ที่มา :** MediaPipe (2023: Online)

สำหรับหน้านั้น Mediapipe สกัดออกมาได้ 468 Keypoints ได้แก่ รูปทรงรอบหน้าและหน้า, ตา, ปากและคิ้ว ซึ่งคำนวณค่า 3 มิติ X, Y, Z ได้ดังนี้

Keypoints in face x Three dimensions = (468 x 3) = 1404 Keypoints ดังภาพที่ 3.5

รูปภาพประกอบด้วย ร่าง, ขาว

คำอธิบายที่สร้างโดยอัตโนมัติ

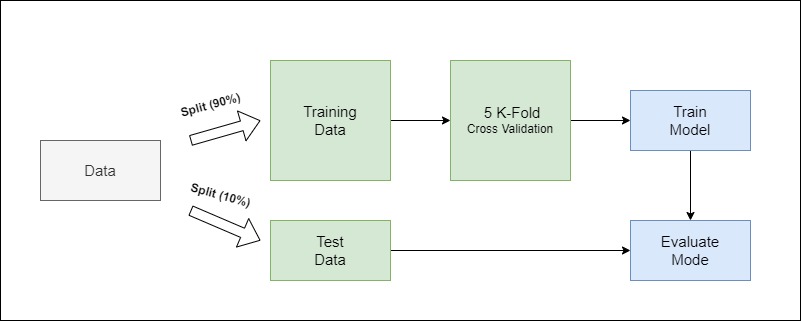
**ภาพที่ 3.5** Keypoints บนหน้า

ดังนั้นเมื่อรวม Keypoint ทั้งหมดเข้าด้วยกันไม่ว่าจะเป็นจาก หน้า ท่าทางและมือจะสามารถคำนวณได้ดังนี้

Keypoints in hands + in pose + inface = (126 + 132 + 1404) = 1662 Keypoints

**3.1.3 การแบ่งข้อมูล**

ในขั้นตอนนี้ผู้วิจัยจะแบ่งข้อมูลข้อมูลออกเป็น 2 ส่วนเพื่อสำหรับการในการไปเทรนและสำหรับการนำไปทดสอบ โดยข้อมูลทั้งหมดคือ 850 วิดีภาษามือ จะทำการเป็นข้อมูลเป็นอัตราส่วน 90:10 และนำข้อมูลข้อมูลในอัตราส่วน 90% นั้นมาทำการแบ่งสำหรับการทำ K-Fold 5 Fold เพื่อให้โมเดลฝึกฝน ดังภาพที่ 3.6



**ภาพที่ 3.6** การแบ่งข้อมูลสำหรับเทรนและทดสอบ

## **3.2 การฝึกฝนโมเดล**

ผู้วิจัยได้ใช้โมเดลในการเทรนทั้งหมด 3 โมเดลได้แก่ LSTM, GRU, BiSLTM ในงานวิจัยครั้งนี้   
Number of Nodes คือ จำนวนของ Input Node ซึ่งผู้วิจัยกำหนดขั้นต่ำไว้ 64 จนถึง 256

Activation คือตัวฟังก์ชันที่ใช้ในการรับผลรวมจากการประมวลผลทั้งหมดจากทุก Input Node เข้ามาพิจารณาตามกลไกลการคำนวณของ Activation Function นั้น ๆ แล้วส่งต่อไปเป็น Output ซึ่งในงานวิจัยนี้ได้เลือกใช้ 2 ตัว คือ Rectified Linear Unit (ReLU) และ Softmax

Optimizer คือ อัลกอริทึมการเพิ่มประสิทธิภาพ (Optimizer) ทำหน้าที่เป็นกลไกการปรับปรุงค่าน้ำหนักของตัวแปรต้นต่าง ๆ รวมถึงค่าความคลาดเคลื่อน (Bias) ในงานวิจัยนี้ได้เลือกใช้ Optimizer ได้แก่ Adagrad. Adamax, Adam or RMSprop ดังตารางที่ 3.2

**ตารางที่ 3.2** 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 ผู้วิจัยได้ใช้ตัวชี้วัดคือค่า Accuracy หรือก็คือค่าอัตราความถูกต้องของการทำนายของโมเดลโดยในการวิจัยครั้งนี้ ผู้วิจัยตั้งเป้าหมายของค่าความถูกต้องไว้ที่ > 90% และจะทำการทดสอบค่าความถูกต้องในการทำนายของโมเดลที่เทรนด้วยวิธี Cross Validation โดยทำการแบ่งข้อมูลออกเป็น 2 ส่วน ได้แก่ ส่วนที่เอาไว้ใช้สำหรับการเทรนและอีกส่วนคือส่วนสำหรับการทดสอบ จะทำการสุ่มข้อมูลตามอัตราส่วนร้อยละ 90:10 และแบ่งข้อมูลสำหรับทำ K-Fold 5 Fold

## **3.4 การเปรียบเทียบประสิทธิภาพโมเดล**

ในขั้นตอนการเปรียบประเทียบประสิทธิภาพ ผู้วิจัยจะนำโมเดลที่ผ่านการเทรนทั้งหมด 3 โมเดล ได้แก่ LSTM, GRU, BiLSTM ซึ่งจะเปรียบประสิทธิภาพเรื่องของ ค่า Accuracy, ค่า Loss และ จำนวนรอบที่ใช้ในการเทรนโมเดล (epochs) เพื่อหาว่าโมเดลใด มีความแม่นยำมากที่สุด แล้วจะนำโมเดลที่มีความแม่นยำมากที่สุดนั้นมาทดสอบทำนายท่าท่างภาษามือไทย

## **3.5 การทดสอบโมเดล**

หลังจากได้รับโมเดลที่มีประสิทธิภาพที่ดีที่สุดแล้ว ผู้วิจัยจะนำโมเดลนั้นมาทดสอบด้วยวิดีโอที่เตรียมไว้สำหรับทดสอบเป็นคำศัพท์ภาษามือจำนวน 10 คำ ซึ่งจะเป็นการทดสอบเพื่อหาประสิทธิภาพของโมเดลด้วยการทำ Confusion Matrix เพื่อหา Accuracy