การประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ

Assessment of fraudulent conduct during online exams using an artificial intelligence-based automatic student facial expression recognition system

วฤษาย์ ร่มสายหยุด[1*], สุภาวดี ธีรธรรมากร[1], พิมผกา ประเสริฐศิลป์[2] และ ภิรมย์ คงเลิศ[2]
Walisa Romsaiyud[1*], Supawadee Theerathammakorn[1], Pimpaka Prasertsilp[2] and Pirom Konglerd[2]
Received: 11 October 2024 ; **Revised**: 21 September 2024 ; **Accepted**: 16 December 2024

บทคัดย่อ

ปัญหาหลักจากการเปลี่ยนแปลงการสอบแบบเผชิญหน้าที่ห้องสอบมาเป็นการสอบแบบออนไลน์ คือ การแสดงออกทางสีหน้าและพฤติกรรมเฉพาะตัวของนักศึกษาแต่ละคนซึ่งยากต่อการเข้าใจด้วยมนุษย์ งานวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) พัฒนานวัตกรรมทางการศึกษาที่รองรับระบบการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ 2) ประเมินประสิทธิผลของแบบจำลอง จากค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน และค่าประสิทธิผลโดยรวม และ 3) ประเมินผลประสิทธิผลของระบบการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์ในการทดสอบการใช้งานจริง ผลการวิจัยพบว่า 1) แบบจำลองที่พัฒนาขึ้นเป็นการแก้ปัญหาด้วยวิธีการเรียนรู้เชิงลึกของปัญญาประดิษฐ์ ประกอบด้วย 6 ขั้นตอน ได้แก่ การเก็บรวบรวมข้อมูลจากไฟล์วิดีโอการสอบออนไลน์ การเตรียมข้อมูลโดยตัดไฟล์ภาพวิดีโอออกเป็นเฟรมภาพ การสร้างแบบจำลองการรับรู้อัตโนมัติจากการแสดงออกทางใบหน้า หรือ ชื่อว่า STOU-ASFER โดยใช้อัลกอริธึม 2 ตัว ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน และโครงข่ายประสาทเทียมแบบหลายชั้น การประเมินแบบจำลองด้วย 4 เมตริกหลัก ได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน และค่าประสิทธิผลโดยรวม การปรับค่าพารามิเตอร์ และการนำไปใช้เพื่อแจ้งเตือนแบบเรียลไทม์และสร้างรายงานสรุป 2) การประเมินแบบจำลองพบว่า มีค่าความถูกต้อง 86.2% ค่าความแม่นยำ 77.34% ค่าความครบถ้วน 95.7% และค่าประสิทธิผลโดยรวม 85.6% และ 3) แบบจำลองมีประสิทธิผลดีในการทำนายพฤติกรรมทุจริตในสภาพแวดล้อมการสอบจำลอง อย่างไรก็ตามแบบจำลองยังต้องการการปรับปรุงในส่วนของการตรวจจับใบหน้าในกรณีที่ใบหน้าอยู่ในตำแหน่งแบบสุ่มหรือมีขนาดภาพเล็กเพื่อเพิ่มความแม่นยำและประสิทธิผลในอนาคต

คำสำคัญ: การแสดงออกทางสีหน้า, การสอบออนไลน์, ปัญญาประดิษฐ์, โครงข่ายประสาทเทียมคอนโวลูชัน, โครงข่ายประสาทเทียมแบบหลายชั้น

Abstract

The primary issues arising from the disruption of traditional face-to-face examinations in the exam room and the shift to online exams are the unique facial expressions and behaviors of students, which are difficult for humans to understand. The purposes of this research were to 1) develop a model of educational innovation for the assessment of fraudulent conduct during online exams, using Artificial Intelligence (AI) based on an Automatic Student Facial Expression Recognition (ASFER) system; 2) evaluate the effectiveness of the model through metrics such as accuracy, precision, recall, and F-measure; and 3) evaluate the effectiveness of the model in assessing fraudulent

[1] รองศาสตราจารย์ สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช นนทบุรี 11120
[2] ผู้ช่วยศาสตราจารย์ สาขาวิชาวิทยาศาสตร์และเทคโนโลยี มหาวิทยาลัยสุโขทัยธรรมาธิราช นนทบุรี 11120
[1] Associate Professor, School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, 11120
[2] Assistant Professor, School of Science and Technology, Sukhothai Thammathirat Open University, Nonthaburi, 11120
* Corresponding author E-mail: walisa.rom@stou.ac.th

conduct during online exams, using AI based on an ASFER system during its actual use. The results revealed that 1) the model that was developed addressed the problem based on a deep learning method in artificial intelligence, consisting of six steps: data collection from online exam videos, data preparation by extracting frames, development of the Automatic Student Facial Expression Recognition Model, also known as STOU-ASFER using two algorithms: a convolutional neural network (CNN) and a multilayer perceptron (MLP) for classifying the results into those exhibiting a regular face and those exhibiting a face showing signs of fraudulent conduct, evaluation of the model using the four main metrics of accuracy, precision, recall, and F-measure, parameter optimization, and deployment for real-time alerts and summary reporting. 2) The evaluation of the model showed an accuracy value of 86.2%, a precision value of 77.34%, a recall value of 95.7%, and an F-measure of 85.6%, and 3) the evaluation of the model performed well in predicting fraudulent behavior in the simulated examination environment. However, the model needs to be improved for face detection when faces are randomly positioned and when small image sizes are encountered.

Keywords: Facial expressions, online examination, artificial intelligence, convolutional neural network, multi-layer perceptron

บทนำ

จากสถานการณ์การแพร่ระบาดของเชื้อโคโรนา 2019 หรือโควิด-19 (COVID-19) ส่งผลกระทบอย่างหนักต่อเศรษฐกิจและสังคมของโลก โดยเฉพาะผลกระทบต่อระบบการศึกษาในทุกระดับชั้น ตั้งแต่ชั้นอนุบาล ชั้นประถมศึกษา ชั้นมัธยมศึกษา ถึงชั้นอุดมศึกษา ทำให้หน่วยงานที่เกี่ยวข้อง จำเป็นต้องมีการปรับเปลี่ยนรูปแบบการเรียนและการสอบจากเดิมที่เป็นแบบเผชิญหน้าในห้องเรียน มาเป็นรูปแบบออนไลน์ (online) ทั้งการจัดการเรียนการสอนแบบออนไลน์ (online courses) สัมมนาออนไลน์ (online seminar) และโดยเฉพาะอย่างยิ่งการสอบออนไลน์ (online examinations) ที่นำเทคโนโลยีคอมพิวเตอร์ เครือข่ายระบบอินเทอร์เน็ต และแพลตฟอร์ม (platforms) ต่างๆ เข้ามาช่วยสนับสนุนการดำเนินการที่เป็นมาตรฐาน สร้างความน่าเชื่อถือและเป็นที่ยอมรับ โดยในหลายมหาวิทยาลัยได้มีการเตรียมความพร้อมสำหรับการสอบออนไลน์ และการป้องกันการทุจริตในการสอบหลายรูปแบบ อาทิ มหาวิทยาลัยเกษตรศาสตร์กำหนดให้นักศึกษาติดตั้งกล้อง 2 ตัว หรือกล้อง 360 องศา เพื่อให้เห็นภาพที่ครอบคลุมทั้งใบหน้านักศึกษา และด้านหลังของนักศึกษารวมถึงโต๊ะและหน้าจอคอมพิวเตอร์ที่นักศึกษาใช้สอบ เพื่อป้องกันไม่ให้มีเอกสารและอุปกรณ์ที่ไม่เกี่ยวข้องกับการสอบอยู่บนโต๊ะนักศึกษาในขณะทำการสอบ และมหาวิทยาลัยเชียงใหม่ มีการใช้กล้องจากคอมพิวเตอร์ที่นักศึกษาใช้สอบ แต่มีการกำหนดให้นักศึกษาติดตั้งโปรแกรมป้องกันการใช้งานเบราว์เซอร์ (browser) อื่นระหว่างการสอบ (Safe Exam Browser: SEB) บนอุปกรณ์ของตนเองที่ใช้สอบด้วย แต่ระหว่างการสอบเมื่อใช้กล้องจากคอมพิวเตอร์ของนักศึกษา ผู้คุมสอบจะไม่สามารถเห็นหน้าจอที่นักศึกษาทำสอบได้อย่างครอบคลุมว่ามีการเปิดอ่านข้อมูลจากแหล่งอื่นบนหน้าจอหรือไม่ ซึ่งส่งผลต่อการนำไปสู่การทุจริตในการสอบได้ ทั้งนี้ในการสอบของทั้งสองมหาวิทยาลัยจะมีการจำกัดจำนวนของนักศึกษาที่ผู้คุมสอบออนไลน์สามารถควบคุมได้ไม่เกิน 30 คน อย่างไรก็ตามในการคุมสอบเจ้าหน้าที่ประจำห้องสอบจะต้องเฝ้ามองหน้าจอเพื่อสังเกตการแสดงออกทางสีหน้า (facial expression) และพฤติกรรมของนักศึกษาที่สงสัยว่าจะส่อทุจริตหรือไม่ อาจทำให้เกิดข้อผิดพลาดได้จากจำนวนนักศึกษาบนหน้าจอที่จำนวนมาก และมีความหลากหลายของมุมมองของภาพนักศึกษาในแต่ละราย

งานวิจัยต่างๆ มีการประยุกต์ใช้หลักการของคอมพิวเตอร์วิทัศน์ (computer vision) เพื่อจดจำใบหน้า (face recognition) และตรวจจับใบหน้า (face detection) ดังงานวิจัยของ Ekundayo and Viriri (2021) ทำการสกัดคุณลักษณะและการจำแนกคุณลักษณะจากชุดข้อมูลการจดจำการแสดงออกบนใบหน้า (Facial Expression Recognition: FER) และงานวิจัยของ Guodong and Na (2019) ศึกษาหลักการเรียนรู้เชิงลึก (deep learning) และอัลกอริธึมต่างๆ เช่น โครงข่ายประสาทเทียม (Artificial Neural Networks: ANN) โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network: CNN) โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron: MLP) โครงข่ายประสาทแบบเวียนซ้ำ (Recurrent Neural Network: RNN) และ หน่วยความจำระยะสั้นแบบยาว (Long Short-Term Memory: LSTM) สำหรับการจดจำใบหน้าทั้งจากภาพและวิดีโอตามอนุกรมเวลา(time series) โดยในงานวิจัยของ Ramzan *et al*. (2024) มีการประยุกต์ใช้โครงข่ายประสาทเทียมคอนโวลูชันที่ได้รับการฝึกอบรมล่วงหน้า (pre-trained) ในการตรวจจับกิจกรรมที่ผิดปกติในการสอบออนไลน์จาก

วิดีโอ พบว่าการสร้างแบบจำลองด้วยโครงข่ายประสาทเทียมคอนโวลูชันมีค่าความถูกต้อง (accuracy) 92% มีค่าความแม่นยำ (precision) 92% มีค่าความครบถ้วน (recall) 92% และมีค่าประสิทธิผลโดยรวม 91%

นอกจากนี้ในงานวิจัยต่างๆ นำเสนอการตรวจจับจุดสังเกตบนใบหน้า (facial landmark detection) มาช่วยในการเพิ่มประสิทธิผลของแบบจำลอง เช่น งานวิจัยของ Iqbal *et al*. (2023) นำเสนอวิธีการตรวจติดตามสายตา (eye tracking) และการตรวจจับวัตถุ (object detection) บนไฟล์วิดีโอในระหว่างการสอบออนไลน์บนโครงข่ายประสาทเทียมคอนโวลูชันซึ่งมีค่าความถูกต้อง (accuracy) สูงถึง 92% และงานวิจัยของ Soltane and Laouar (2021) เพิ่มเติมวิธีการตรวจจับเสียง (voice detection) ตรวจจับใบหน้า (face detection) การจดจำใบหน้า (face recognition) และการตรวจม่านตา (iris detection) บนไฟล์วิดีโอในระหว่างการสอบออนไลน์บนโครงข่ายประสาทเทียมคอนโวลูชัน อย่างไรก็ตามการจดจำการแสดงออกบนใบหน้าแบบเรียลไทม์ (real-time) เป็นเรื่องที่มีความยุ่งยากและซับซ้อนมากในการจำแนกข้อมูล (data classification) และสกัดคุณลักษณะ (feature extraction) ของแต่ละใบหน้าเพื่อทำให้แบบจำลองมีความถูกต้องและน่าเชื่อถือ จึงมีการผสมผสานอัลกอริธึมต่างๆ เช่น งานวิจัยของ Abdullah *et al*. (2020) โดยผสมผสานวิธีการของ 2 อัลกอริธึม ได้แก่ โครงข่ายประสาทเทียมคอนโวลูชัน และหน่วยความจำระยะสั้นแบบยาวสำหรับสร้างแบบจำลองการแสดงออกทางสีหน้าในวิดีโอที่มีความถูกต้องสูงสำหรับข้อมูลทดสอบ (test data) และงานวิจัยของ Haghpanah *et al*. (2022) ผสมผสานวิธีการของ 2 อัลกอริธึม ได้แก่ โครงข่ายประสาทเทียมคอนโวลูชันและโครงข่ายประสาทเทียมแบบหลายชั้นมาประยุกต์ใช้ในการสร้างแบบจำลองเพื่อทำให้แบบจำลองมีความถูกต้องสูงมากถึง 96% ดังนั้นในงานวิจัยนี้จึงนำผลการศึกษาการดำเนินการจากงานวิจัยต่างๆ ดังกล่าว มาดำเนินการพัฒนาแบบจำลองสำหรับการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ

ดังนั้นงานวิจัยนี้ขอนำเสนอ 2 ประเด็นหลัก ได้แก่ 1) การพัฒนาระบบการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ จากปัญญาประดิษฐ์ (AI) ด้วยอัลกอริธึมการเรียนรู้เชิงลึก 2 อัลกอริธึม ได้แก่ โครงข่ายประสาทเทียมแบบคอนโวลูชัน และโครงข่ายประสาทเทียมแบบหลายชั้น เพื่อจำแนกการแสดงออกทางสีหน้าว่าเป็นใบหน้าปกติหรือใบหน้าที่ส่อทุจริต และ 2) การประเมินผลแบบจำลองด้วยคอนฟิวชัน เมทริก (confusion matrix) โดยใช้ตัวชี้วัดมาตรฐาน 4 ค่าได้แก่ค่าความถูกต้อง (accuracy) ค่าความแม่นยำ (precision) ค่าความครบถ้วน (recall) และค่าประสิทธิผลโดยรวม (F-measure) ที่สามารถนำไปใช้จริงในสภาพแวดล้อมการสอบจำลองเพื่อแสดงให้เห็นถึงความเป็นไปได้ในทางปฏิบัติและสิ่งที่ต้องปรับปรุงแก้ไข เพื่อให้ระบบมีความถูกต้องและน่าเชื่อถือ

ทฤษฎีและงานวิจัยที่เกี่ยวข้อง

เนื้อหาในส่วนนี้ขออธิบายทฤษฎี และงานวิจัยที่เกี่ยวข้อง จำนวน 3 ประเด็น ได้แก่ การแสดงออกทางสีหน้า โครงข่ายประสาทเทียมคอนโวลูชัน และโครงข่ายประสาทเทียมแบบหลายชั้น รายละเอียดดังนี้

1. การแสดงออกทางสีหน้า (Facial expressions)

การแสดงออกทางสีหน้า คือ การแสดงอารมณ์หรือความรู้สึกของบุคคลที่สามารถมองเห็นได้จากการเปลี่ยนแปลงของกล้ามเนื้อบนใบหน้า เช่น การยิ้ม การขมวดคิ้ว หรือการย่นหน้าผาก ซึ่งเป็นการสื่อสารที่สำคัญในคน (Frank, 2001) โดยงานวิจัยส่วนใหญ่มุ่งเน้นไปที่การรับรู้การแสดงออกทางสีหน้าแบบอัตโนมัติ (Automatic Facial Expression Recognition: AFER) ตามทฤษฎีของ Ekman (1992) ซึ่งแนะนำอารมณ์ความรู้สึกพื้นฐานที่ใช้อย่างเป็นสากลในทุกวัฒนธรรมไว้ 6 อารมณ์ ได้แก่ ความสุข (happiness) ความประหลาดใจ (surprise) ความโกรธ (anger) ความเศร้า (sadness) ความกลัว (fear) และการรังเกียจ (disgust) ในการตรวจสอบการแสดงออกทางอารมณ์ความรู้สึกที่แสดงออกทางใบหน้า จะอ้างอิงรูปแบบของใบหน้าที่มีการแสดงอารมณ์ความรู้สึกออกมาในลักษณะหรือรูปแบบต่างๆ ดัง Figure 1

(ก.) ความสุข (ข.) ความประหลาดใจ (ค.) ความโกรธ

(ง.) ความเศร้า (จ.) ความกลัว (ฉ.) การรังเกียจ

Figure 1 Examples from the FER2013 dataset for each emotion with labeled data

จาก Figure 1 (ก.)-(ฉ.) แสดงตัวอย่างอารมณ์ทั้ง 6 ประเภทจากชุดข้อมูล FER2013 ที่มีการกำหนดป้ายกำกับข้อมูลแต่ละประเภท (labeled data)

จากงานวิจัยของ Mohamad Nezami *et al*. (2020) นำเสนอการพัฒนาแบบจำลองการจดจำการมีส่วนร่วมของนักเรียนแบบอัตโนมัติด้วยอัลกอริธึมการเรียนรู้เชิงลึก โดยใช้ชุดข้อมูลของ FER2013 จำนวน 35,887 ตัวอย่าง ประกอบด้วยข้อมูลสำหรับฝึกสอนแบบจำลองจำนวน 28,709 ตัวอย่าง และข้อมูลสำหรับทดสอบแบบจำลองจำนวน 3,589 ตัวอย่าง ด้วยโครงข่ายประสาทเทียมคอนโวลูชัน พบว่าแบบจำลองสามารถทำนายผลการมีส่วนร่วมของนักเรียนได้ถูกต้องถึง 87.06% และในงานวิจัยของ Gavade *et al*. (2023) ดำเนินการพัฒนาระบบการระบุและจำแนกการแสดงออกทางสีหน้าจากวิดีโอแบบอัตโนมัติ ซึ่งแบ่งชุดข้อมูลเป็น 2 ประเภท คือ ไฟล์วิดีโอที่ถ่ายในป่า (มีข้อจำกัดเรื่องแสงเงา) และไฟล์วิดีโอที่ถ่ายในสภาพแวดล้อมที่ควบคุมได้ โดยใช้อัลกอริธึมโครงข่ายประสาทเทียมคอนโวลูชัน และหน่วยความจำระยะสั้นแบบยาว และทำการทดสอบโดยการเปรียบเทียบกับชุดข้อมูลต่างๆ และชุดข้อมูลจาก FER พบว่าแบบจำลองมีค่าความถูกต้อง 80%

โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network Algorithm หรือ CNN)

โครงข่ายประสาทเทียมคอนโวลูชันเป็นโครงข่ายประสาทเทียมที่ถูกออกแบบมาสำหรับการประมวลผลภาพ เช่น การจำแนกรูปภาพ การตรวจจับวัตถุ หรือการประมวลผลข้อมูลภาพต่างๆ (Indolia *et al*., 2018) โดยใช้วิธีการคำนวณแบบคอนโวลูชัน (convolutional) เข้าไปในโครงข่าย ทำให้สามารถวิเคราะห์ข้อมูลของภาพหรือวิดีโอได้อย่างถูกต้องและรวดเร็ว การทำงานของโครงข่ายประสาทเทียมคอนโวลูชัน ประกอบด้วย 4 ขั้นตอน ได้แก่ 1) ชั้นอินพุต (input layer) 2) ชั้นคอนโวลูชัน (convolutional layer) 3) ชั้นพูลลิ่ง (pooling layer) และ 4) ชั้นเชื่อมโยงแบบสมบูรณ์ (fully-connected layer) (Lapan , 2020) ดัง Figure 2

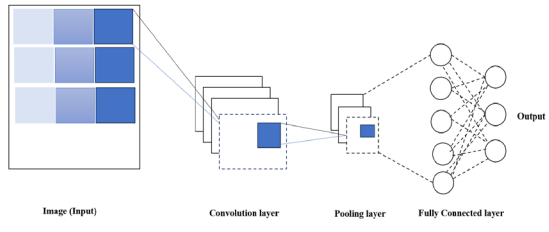

Figure 2 Convolutional Neural Network

จาก Figure 2 ขั้นตอนการทำงานของอัลกอริธึมโครงข่ายประสาทเทียมคอนโวลูชัน ได้แก่ 1) ชั้นอินพุต (input layer) เป็นชั้นแรกที่นำข้อมูลเข้า จากนั้นทำการแปลงข้อมูลให้เป็นเวกเตอร์ 2) ชั้นคอนโวลูชัน (convolutional layer) เป็นชั้นที่สองของแบบจำลอง ซึ่งทำหน้าที่สกัดคุณลักษณะสำคัญจากเวกเตอร์ ประกอบด้วย 2 ขั้นตอน ได้แก่ การกำหนดตัวกรอง (filters) หรือเคอร์เนล สำหรับกรองลักษณะสำคัญที่ใช้ในการรู้จำวัตถุ ในรูปแบบตารางสองมิติ และ แผนผังคุณลักษณะ (feature maps) เป็นผลลัพธ์จากขั้นตอนการกำหนดตัวกรอง เพื่อส่งข้อมูลที่ได้ไปในชั้นต่อไป 3) ชั้นพูลลิ่ง (pooling layer) เป็นชั้นที่สามที่ทำหน้าที่คำนวณผลรวมของข้อมูล เพื่อลดขนาดของเวกเตอร์ ทำให้ข้อมูลมีขนาดเล็กลงเมื่อเทียบกับข้อมูลต้นฉบับ ส่งผลให้สามารถเพิ่มความเร็วในการประมวลผลได้เป็นอย่างมาก และ 4) ชั้นเชื่อมโยงแบบสมบูรณ์ (fully-connected layer) เป็นชั้นสุดท้ายสำหรับแสดงผลลัพธ์ของโครงข่ายประสาทเทียมคอนโวลูชัน

โครงข่ายประสาทเทียมแบบหลายชั้น (Multi-Layer Perceptron หรือ MLP)

โครงข่ายประสาทเทียมแบบหลายชั้นเป็นโครงข่ายประสาทเทียมที่ถูกออกแบบมาสำหรับการจำแนกภาพ (image classification) การจดจำรูปแบบเสียง (speech recognition) การพยากรณ์ข้อมูลเวลา (time series prediction) และการวิเคราะห์ข้อความ (text analysis) ที่มีจุดเด่นในการเรียนรู้จากข้อมูลที่ไม่เป็นเส้นตรง (non-linear data) ได้ เนื่องจากใช้ฟังก์ชัน activation ที่ไม่เป็นเชิงเส้น จึงเหมาะกับปัญหาที่ต้องการการจำแนก (classification) และการทำนาย (prediction) (Zhang *et al*., 2018) การทำงานของโครงข่ายประสาทเทียมแบบหลายชั้น ประกอบด้วย 3 ขั้นตอน ได้แก่ ชั้นนำเข้า (input layer) จำนวน 1 ชั้น มีจำนวนชั้นผลลัพธ์ (output layer) 1 ชั้น และมีจำนวนชั้นซ่อน (hidden layer) อย่างน้อย 1 ชั้น โดยในแต่ละชั้นเชื่อมต่อกัน (Lapan , 2020) ดัง Figure 3

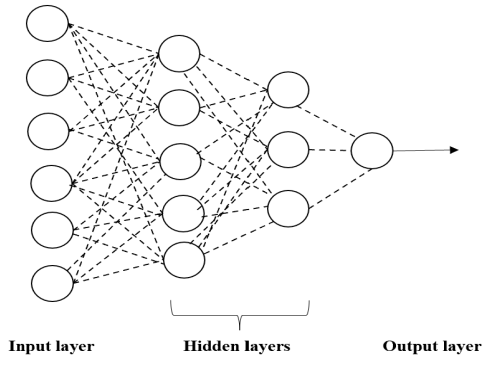

Figure 3 Multi-Layer Perceptron

จาก Figure 3 ขั้นตอนการทำงานของอัลกอริธึมการรับรู้หลายระดับชั้น โดยข้อมูลเข้าสู่ชั้นอินพุต (input layer) เพื่อทำการฝึกสอน จากนั้นส่งข้อมูลเข้าสู่ชั้นซ่อน (hidden layer) เพื่อจดจำรูปแบบข้อมูล แยกแยะภาพตามค่าน้ำหนัก (weight)

ของแต่ละโหนด (ซึ่งมีค่าน้ำหนักเฉพาะสำหรับคำนวณค่าถ่วงน้ำหนัก) จากนั้นนำค่าน้ำหนักมารวมกัน และแปลงให้ออกมาเป็นผลลัพธ์ และแสดงผลลัพธ์ที่ชั้นเอาต์พุต (output layer) จากค่าน้ำหนัก เพื่อทำนายผลลัพธ์ตามความน่าจะเป็น (Aggarwal, 2018)

วิธีดำเนินการวิจัย

เนื้อหาในส่วนนี้ขออธิบายขั้นตอนการทำงานของระบบการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ ดัง Figure 4

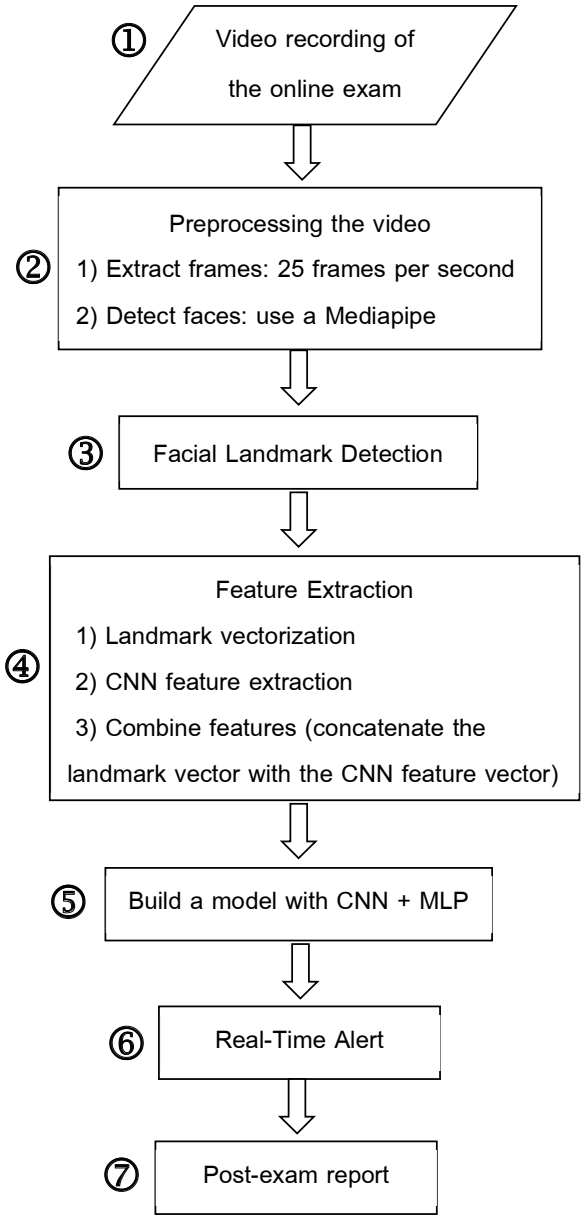

Figure 4 The workflow of STOU-ASFER

จาก Figure 4 แสดงขั้นตอนการทำงานของ STOU-ASFER ซึ่งประกอบด้วย 7 ขั้นตอน ได้แก่ 1. การบันทึกวิดีโอการสอบออนไลน์ 2. การประมวลผลวิดีโอ ซึ่งประกอบด้วย 2 ขั้นตอนย่อย ได้แก่ การแยกเฟรม และการตรวจจับใบหน้าโดยใช้ Mediapipe 3. การตรวจจับจุดสังเกตบนใบหน้า 4. การสกัดคุณลักษณะ ซึ่งประกอบด้วย 3 ขั้นตอนย่อย ได้แก่ 1) การสร้างเวกเตอร์แลนด์มาร์ก 2) การสกัดคุณลักษณะด้วย CNN และ 3) รวมคุณลักษณะเข้าด้วยกัน โดยการเชื่อมเวกเตอร์แลนด์มาร์กกับเวกเตอร์คุณลักษณะของ CNN 5. การสร้างแบบจำลองแบบผสานด้วย CNN+MLP 6. การแจ้งเตือนแบบเรียลไทม์ และ 7. การสร้างรายงานหลังสอบ รายละเอียดดังนี้

1. การบันทึกวิดีโอการสอบออนไลน์ (Video recording of the online exam)

เป็นการนำเข้าข้อมูลจากการรวบรวมข้อมูลของกลุ่มตัวอย่าง ซึ่งประกอบด้วยข้อมูลไฟล์วิดีโอใบหน้าของอาสาสมัครผู้เข้าร่วมในโครงการวิจัยนี้จำนวน 80 คน โดยแบ่งเป็น 2 กลุ่ม กลุ่มแรกสำหรับฝึกสอนแบบจำลอง และทดสอบแบบจำลอง จำนวน 50 คน และกลุ่มที่สองสำหรับทดสอบการใช้งานจริง จำนวน 30 คน (ตามหนังสือแสดงความยินยอมเข้าร่วมโครงการวิจัยสำหรับผู้เข้าร่วมโครงการอายุ 18 ปีขึ้นไป)

ตัวอย่างไฟล์วิดีโอของกลุ่มตัวอย่าง ที่ทำการสอบเสมือนจริง โดยทำการสอบครั้งละ 5, 10 และ 15 นาที ตามจำนวนเฟรม โดย 5 นาที = 5*60*25 เฟรม, 10 นาที = 10*60*25 และ 15 นาที = 15*60*25 เฟรม รวมเป็นชุดข้อมูลทั้งหมด 7,500+15,000+22,500 = 45,000 เฟรม แบ่งเป็นการสอบแบบปกติ หรือสุจริต (innocent) = 35,000 เฟรม และการสอบแบบทุจริต (cheat) = 10,000 เฟรม และกำหนดป้ายกำกับ (labeled data) โดยใช้วิธีการระบุช่วงเวลาของพฤติกรรมที่เกิดขึ้นในวิดีโอ พร้อมทั้งมีผู้เชี่ยวชาญที่มีความรู้และประสบการณ์ จากสำนักทะเบียนและวัดผล มหาวิทยาลัยสุโขทัยธรรมาธิราช (2566) (บนระบบการสอบแบบออนไลน์) ตัดสินการกำหนดข้อมูลที่มีป้ายกำกับที่ถูกต้อง สำหรับการกำหนดใบหน้าที่เป็นแบบสุจริต (innocent) คือใบหน้าที่มีองศาใบหน้า ตำแหน่งตา รูปแบบการมอง การหันศีรษะ อยู่ในบริเวณหน้าจอ ตามระยะเวลาที่กำหนด และใบหน้าที่เป็นแบบทุจริต (cheat) คือการแสดงใบหน้าที่มีตำแหน่งองศาใบหน้าผิดรูปแบบการมองหน้าจอคอมพิวเตอร์ ดำเนินการกำหนดใบหน้าที่เป็นแบบทุจริตตามระเบียบมหาวิทยาลัยสุโขทัยธรรมาธิราช ว่าด้วยการสอบออนไลน์ของนักศึกษาระดับปริญญาตรีและระดับต่ำกว่าปริญญา พ.ศ. 2563 หมวด 4 แนวปฏิบัติตนในการเข้าสอบของนักศึกษา และจากงานวิจัยของ Noorbehbahani *et al*. (2022) กำหนดพฤติกรรมการทุจริต

ในการสอบออนไลน์ ดังนี้ 1) การใช้หนังสือ หรือแหล่งข้อมูลออนไลน์ขณะทำการสอบ 2) การกำหนดให้บุคคลอื่นสอบแทน 3) การขอความช่วยเหลือจากบุคคลที่สามในการสอบ 4) การรับหรือเผยแพร่คำถามจากบุคคลอื่น 5) การคัดลอกและจำหน่ายข้อสอบให้นักศึกษาท่านอื่นหรือนำไปใช้ในปีต่อ ๆ ไป และ 6) การใช้อุปกรณ์เคลื่อนที่เพื่อการสื่อสารแบบคัดลอก ดัง Figure 5

Figure 5 Examples of Cheating Behaviors During Online Exams

จาก Figure 5 ตัวอย่างพฤติกรรมการทุจริตระหว่างการสอบออนไลน์ โดย (ก.) แสดงการใช้โทรศัพท์เพื่อติดต่อสอบถาม (ข.) แสดงการใช้อุปกรณ์เคลื่อนที่เพื่อการสื่อสารหรือค้นหาข้อมูลเพิ่มเติม และ (ค.) แสดงการขอความช่วยเหลือจากบุคคลอื่นหรือมีบุคคลอื่นร่วมอยู่ด้วยในระหว่างการสอบ

2. การประมวลผลวิดีโอ (Preprocessing the video)

เป็นการนำไฟล์วิดีโอจากขั้นตอนที่ 1 มาดำเนินการ1) การแยกเฟรม (extract frames) และ 2) การตรวจจับใบหน้า (detect faces) โดยใช้ Mediapipe ซึ่งเป็นเฟรมเวิร์กของ Google ซึ่งผู้วิจัยได้ใช้ส่วนของฟังก์ชัน Mediapipe Face Mesh (Google AI for Developers, 2022) ซึ่งสามารถพล็อต (plot) จุดบนใบหน้าได้สูงสุดถึง 468 ตำแหน่ง มาทำการตรวจจับใบหน้า และนำค่าการเคลื่อนไหวของแต่ละภาพ ดัง Figure 6

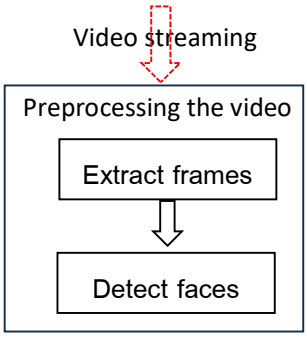

Figure 6 The video preprocessing step

จาก Figure 6 การแยกเฟรมดำเนินการโดยการแปลงไฟล์วิดีโอให้เป็นเฟรมด้วยอัตราเฟรมคงที่ (fixed frame rate) ซึ่งกำหนดให้มีค่าเป็น 25 เฟรมต่อวินาที สำหรับเพิ่มประสิทธิผลการจดจำการแสดงออกบนใบหน้าได้อย่างเหมาะสม และทำการลบหรือข้ามเฟรมที่ซ้ำซ้อนเพื่อลดภาระในการคำนวณ จากนั้นการตรวจจับใบหน้าด้วย Mediapipe เพื่อระบุตำแหน่งใบหน้าในแต่ละเฟรม และแยกเฉพาะส่วนของใบหน้าเพื่อใช้ในการวิเคราะห์ ดัง Figure 7

Figure 7 Cropped face images extracted from individual frames

จาก Figure 7 (ก.) จากต้นฉบับ จากนั้นทำการครอบตัดเฉพาะส่วนใบหน้า ซึ่งแสดงในภาพ (ข.) และจัดเก็บข้อมูลเป็นไฟล์ภาพ

3. การตรวจจับจุดสังเกตบนใบหน้า (Facial Landmark Detection)

เป็นการนำภาพใบหน้าจากขั้นตอนที่ 2 มาดำเนินการกำหนดการตรวจจับจุดสังเกตบนใบหน้า ซึ่งกำหนดที่ 68 จุด สำหรับการตรวจ 1) กราม (jawline) โดยกำหนดจุดสังเกตตั้งแต่เลขที่ 0–16 เพื่อกำหนดรูปร่างของกรามจากหูข้างหนึ่งไปยังอีกข้างหนึ่ง 2) คิ้ว (eyebrows) โดยกำหนดจุดสังเกตที่คิ้วขวาด้วยเลข 17–21 และคิ้วซ้ายด้วยเลข 22–26 3) จมูก (nose) โดยกำหนดจุดสังเกตของสันจมูกด้วยเลข 27–30 และรูปทรงรูจมูกด้วยเลข 31–35 4) ตา (eyes) โดยกำหนดจุดสังเกตของตาขวาด้วยเลข 36–41 และตาซ้ายด้วยเลข 42–47 เพื่อแสดงรูปร่างของดวงตา และใช้เพื่อตรวจจับการเคลื่อนไหวของดวงตาหรือการกระพริบตา 5) ปาก (mouth) โดยกำหนดจุดสังเกตที่รูปทรงริมฝีปากด้านนอกด้วยเลข 48–59 และรูปทรงริมฝีปากด้านในด้วยเลข 60–67 เพื่อตรวจจับการเคลื่อนไหวของริมฝีปาก เช่น การพูดคุยกับบุคคลอื่น ดัง Figure 8

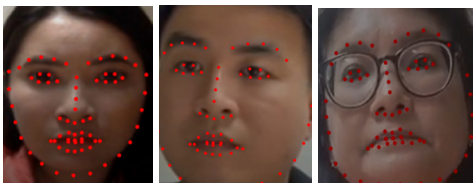

Figure 8 Setting up facial landmark detection with 68 points on the face

4. การสกัดคุณลักษณะ (Feature extraction)

เป็นการนำจุดสังเกตบนใบหน้าทั้ง 68 จุด มาดำเนินการดังนี้ 1) การสร้างเวกเตอร์จุดสังเกต (landmark vectorization) โดยแบ่งจุดสังเกตให้แบนราบเป็นเวกเตอร์เดียว [, , , , ..., , ,] 2) การดึงคุณลักษณะของ CNN (CNN feature extraction) นำภาพใบหน้าส่งเข้าสู่แบบจำลอง CNN และ 3) รวมคุณลักษณะ (combine features) ทำการเชื่อมเวกเตอร์จุดสังเกตกับเวกเตอร์คุณลักษณะของ CNN ให้เป็นเวกเตอร์คุณลักษณะอินพุตเดียว ดัง Figure 9

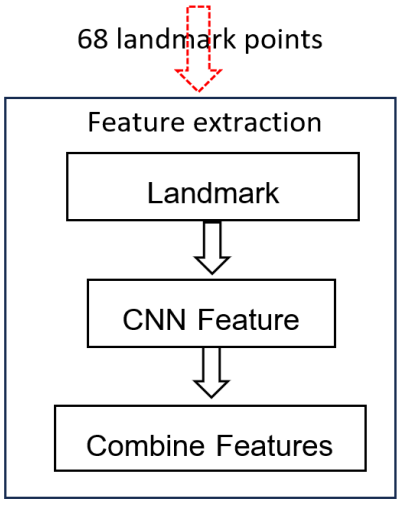

Figure 9 The Feature Extraction Step

5. การสร้างแบบจำลองแบบผสานด้วย CNN+MLP (Build a model with CNN + MLP)

เป็นการนำข้อมูลจากการรวมคุณลักษณะ (combine features) ของขั้นตอนที่ 4 ที่แสดงเป็นอาร์เรย์ (array) แบบ 2 มิติประกอบด้วย โดย คือความสูงของภาพ คือความกว้างของภาพ และ คือจำนวนช่องสัญญาณสี มาดำเนินการฝึกสอน (training) ซึ่งใน CNN เริ่มต้นจากชั้นคอนโวลูชันที่ *1, 2,…* ตามลำดับ จากนั้นดำเนินการโดยใช้ตัวกรองของภาพและสร้างแผนผังคุณลักษณะ ดังสมการที่ (1)

$$F^{(n)} = W^{(n)} * X + b^{(n)} \quad (1)$$

จากสมการที่ (1) คือชั้นคอนโวลูชันที่ คือตัวกรองของภาพ และ คือการกำหนดค่า bias จากนั้น Activation (ReLU) จะทำการปรับผลลัพธ์ให้อยู่ในช่วงที่ต้องการ ดังสมการที่ (2)

$$A^{(n)}=ReLU(F^{(n)}) = max\ (0, F^{(n)}) \quad (2)$$

ชั้นพูลลิ่ง ทำการคำนวณผลรวมค่าสูงสุดของข้อมูลเพื่อลดขนาดเชิงพื้นที่ของภาพ ดังสมการที่ (3)

$$P^{(n)}=MaxPool(A^{(n)}) \quad (3)$$

และทำการปรับให้เป็นแบนราบ (flattening) ทำให้ผลลัพธ์จากแต่ละชั้นถูกรวมในเวกเตอร์ 1 มิติ ดังสมการที่ 4 และจะถูกนำไปใช้เป็นข้อมูลเข้าของ MLP

$$Z = Flatten(A^{(n)}) \quad (4)$$

ที่ MLP เริ่มต้นจากชั้นอินพุตไปยังชั้นซ่อน $H^{(n)}$ จะทำการคำนวณโดยใช้ค่าถ่วงน้ำหนัก $W^{(n)}$ และค่าเอนเอียงไปที่เวกเตอร์ Z ที่ถูกปรับให้แบนราบ ดังสมการที่ (5)

$$H^{(n)} = \sigma(W^{(n)}Z + b^{(n)}) \quad (5)$$

โดย σ คือ activation function (ReLU) จากนั้นที่ชั้นเอาต์พุต จะคำนวณค่าความน่าจะเป็นของคลาสซึ่งในงานวิจัยนี้เป็นแบบ 2 คลาสหรือไบนารีคลาส (binary class) ได้แก่ คลาสสุจริต (Innocent) และคลาสทุจริต (Cheat) โดยใช้ฟังก์ชัน Softmax ดังสมการที่ (6)

$$H^{(k)} = \sigma(W^{(k)}H^{(k-1)} + b^{(k)}) \quad (6)$$

จากสมการที่ (6) กำหนดให้ k คือจำนวนชั้นซ่อน และฟังก์ชัน = $Softmax\ (x_i) = \frac{e^{x_i}}{\sum_{j=1}^{C_{classes}} e^{x_j}}$

6. การแจ้งเตือนแบบเรียลไทม์ (Real-Time Alert)

เป็นการวิเคราะห์ภาพแบบเฟรมต่อเฟรม สำหรับตรวจจับใบหน้า สกัดคุณลักษณะ (landmark detection + CNN) และการจำแนกประเภทคลาสด้วย MLP เพื่อทำการแจ้งเตือนแบบเรียลไทม์หากตรวจพบพฤติกรรมการทุจริต โดยจะปรากฏกรอบสีแดงล้อมรอบภาพของนักศึกษาแต่ละคนสำหรับแจ้งเตือนแก่ผู้คุมสอบ ให้เฝ้าระวังเป็นพิเศษ พร้อมทั้งแจ้งเตือนนักศึกษาหากมีพฤติกรรมที่กำลังจะนำไปสู่การทุจริต

7. การสร้างรายงานหลังสอบ (Post-exam report)

เป็นขั้นตอนสุดท้าย ซึ่งดำเนินการโดยทำการสร้างรายงานสรุปผลการสอบและพฤติกรรมต่างๆ ของนักศึกษาแต่ละคน ตลอด 3 ชั่วโมง ซึ่งประกอบด้วยข้อมูลนักศึกษา จำนวนครั้งของการทุจริต และเวลา

ผลการดำเนินงาน

1. ชุดข้อมูล และสภาพแวดล้อม

ชุดข้อมูลวิชาจำนวน 4 ชุดวิชา ได้แก่ ชุดวิชา 96408 การจัดการระบบฐานข้อมูล 99419 ความมั่นคงปลอดภัยไซเบอร์ 99420 การโปรแกรมเว็บ และ 96412 การบริหารโครงการด้านเทคโนโลยีสารสนเทศ ที่มีการสอบเป็นเฉพาะแบบปรนัยซึ่งนักศึกษาจะต้องมองข้อสอบและเลือกคำตอบที่ถูกต้องผ่านหน้าจอ โดยชุดวิชาเหล่านี้มีจำนวนข้อสอบเท่ากันทั้งหมด 120 ข้อ และใช้เวลาสอบ 3 ชั่วโมง สำหรับเครื่องเซิฟเวอร์ (server) GPU NVIDIA A100 Tensor Core ที่มีหน่วยความจำ GPU 40 GB. และติดตั้งโปรแกรมไพธอนรุ่น 3.10.15 โดยกำหนดให้ใน 1 หน้าจอการสอบ แสดงรูปภาพแบบตารางขนาด 5x5 = 25 ช่อง (สำหรับ 25 คน) และแต่ละหน้าจอย่อยมีขนาดประมาณ 200 x 200 พิกเซล (แสดงผลแบบสี่เหลี่ยมจัตุรัส)

2. การกำหนดค่าพารามิเตอร์ของแต่ละอัลกอริธึม

Table 1 Example of Parameter Settings for the CNN and MLP Algorithms

Algorithms	Types	Settings
CNN	Conv1	32 filters, kernel size (3, 3), ReLU
	MaxPool1	Pool size (2, 2)
	Conv2	64 filters, kernel size (3, 3), ReLU
	MaxPool2	Pool size (2, 2)
	Conv3	128 filters, kernel size (3, 3), ReLU
	MaxPool3	Pool size (2, 2)
	Dropout	0.25 after pooling layers
MLP	Dense1	512 units, ReLU
	Dropout	0.5
	Dense2	256 units, ReLU
	Dropout	0.3
	Dense3	128 units, ReLU
	Dropout	0.3
	Final Dense layer	1 unit, sigmoid (for binary classification)

จาก Table 1 ที่ CNN layer กำหนดจำนวนของการกรองเริ่มต้นที่ 32, 64 และ 128 ตัวกรอง และกำหนดขนาดของพูลลิ่งเป็น 2*2 ของแต่ละชั้น Convolution และที่ MLP layer เริ่มต้นกำหนดแผนที่คุณลักษณะที่ 512, 256 และ 128 หน่วย และในชั้น final layer ใช้ฟังก์ชัน sigmoid สำหรับไบนารีคลาส

3. การแบ่งข้อมูล

การแบ่งข้อมูลออกเป็น 3 ส่วนได้แก่ข้อมูลฝึกสอน (training data) ข้อมูลตรวจสอบ (validation data) และข้อมูลทดสอบ (testing data) และ โดยกำหนดเป็น 64:16:20 ตามลำดับ ซึ่งเป็นสัดส่วนมาตรฐานที่เหมาะสมกับข้อมูลชุดนี้ที่มีจำนวนมากพอในการสร้างและทดสอบแบบจำลอง และลดความเอนเอียงหรืออคติในการแบ่งข้อมูลโดยการใช้ cross validation เป็น 5-fold cross-validation เพื่อประเมินผลแบบจำลองในระหว่างการฝึกสอนแบบจำลอง โดยการใช้ 5-fold cross validation กับชุดข้อมูลนี้ที่มีขนาดใหญ่ทำให้แต่ละ fold มีขนาดที่เพียงพอสำหรับการฝึกสอนแบบจำลอง (หากกำหนดเป็น 10-fold cross validation ทำให้ข้อมูลในแต่ละ fold เล็กเกินไป ซึ่งผลที่ได้ไม่ถูกต้องเท่ากับ 5-fold cross validation)

4. การประเมินประสิทธิผลแบบจำลอง

การประเมินผลแบบจำลองด้วย 4 อัลกอริธึม ได้แก่ CNN, CNN+RNN โดย RNN คือโครงข่ายประสาทแบบเวียนซ้ำ (Recurrent Neural Network: RNN), CNN+LSTM และ CNN+MLP ซึ่งในงานวิจัยนี้วิดีโอไฟล์บันทึกการสอบจะเกี่ยวข้องกับอนุกรมเวลา จึงนำอัลกอริธึม RNN และ LSTM มาทดสอบการสร้างแบบจำลอง และประเมินผลแบบจำลองด้วย 4 เมตริกได้แก่ ค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน ดัง Figure 10

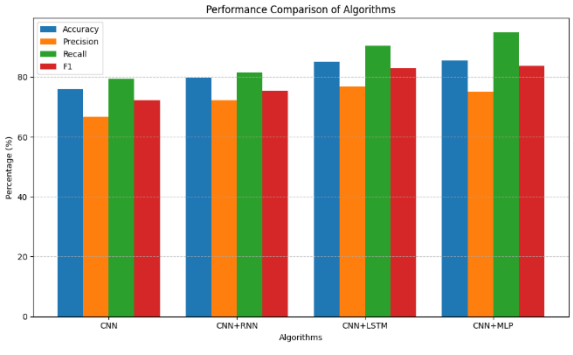

Figure 10 Compares four algorithms using a confusion matrix

จาก Figure 10 จะเห็นได้ว่าอัลกอริธึม CNN+MLP มีค่าความถูกต้อง = 85.5% ค่าความแม่นยำ = 75.0% ค่าความครบถ้วน = 94.9% และค่าประสิทธิผลโดยรวม = 83.7% สูงกว่าอัลกอริธึมอื่น ดังนั้นงานวิจัยนี้จึงเลือกใช้อัลกอริธึม CNN+MLP สำหรับสร้างแบบจำลอง

5. การปรับแต่พารามิเตอร์

นำแบบจำลองมาดำเนินการปรับแต่งพารามิเตอร์ต่างๆ ด้วย GridSearchCV พบว่า Optimizer choices กำหนดเป็น adam, conv_filter = 32, dense_units = 512, dropout_rate = 0.3, epochs = 200 ซึ่งทำให้แบบจำลองมีประสิทธิผลดีขึ้นดัง Table 2

Table 2 Hyperparameter tuning for CNN+MLP models

	CNN+MLP	
	Before Hyperparameter tuning	**After Hyperparameter tuning**
Accuracy	85.5	86.2
Precision	75.0	77.3
Recall	94.9	95.7
F1	83.7	85.6

จาก Table 2 การเปรียบเทียบประสิทธิผลแบบจำลองก่อนและหลังจากการปรับแต่พารามิเตอร์ แสดงให้เห็นว่าแบบจำลองมีประสิทธิผลสูงขึ้น โดยค่าความถูกต้อง = 86.2% ค่าความแม่นยำ = 77.34% ค่าความครบถ้วน = 95.7% และค่าประสิทธิผลโดยรวม = 85.6%

6. การประเมินค่าการสูญเสีย และค่าความถูกต้อง

นำแบบจำลองที่ปรับแต่พารามิเตอร์เรียบร้อยมาดำเนินการประเมินค่าการสูญเสีย (loss) และค่าความถูกต้องของแต่ละ epoch โดยกำหนดค่าเป็น 50, 100, 200 และ 500 ดัง Figure 11

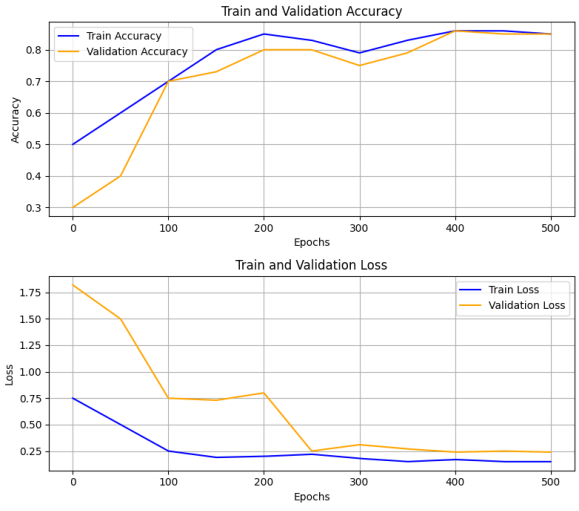

Figure 11 Test Loss and Accuracy for Each Epoch (50, 100, 200, 500)

7. การปรับใช้แบบจำลอง

เป็นกระบวนการนำแบบจำลองการเรียนรู้เชิงลึกที่ได้พัฒนา และทดสอบเรียบร้อยแล้ว ไปประยุกต์ปรับใช้งานในสภาพแวดล้อมจริง โดยในการทดลองนี้ ทางทีมผู้วิจัยใช้ทดสอบการใช้งานจริงในสภาพแวดล้อมการสอบจำลอง กับไฟล์วิดีโอ 4 ชุดวิชาได้แก่ ชุดวิชา 96408 การจัดการระบบฐานข้อมูล 99419 ความมั่นคงปลอดภัยไซเบอร์ 99420 การโปรแกรมเว็บชุดวิชา และ 96412 การบริหารโครงการด้านเทคโนโลยีสารสนเทศ ดัง Table 3

Table 3 compares the model performance results with four courses

Course ID	**Accuracy**	**Precision**	**Recall**	**F1**
96408	85.4	76.1	95.2	84.6
99419	87.4	**79.4**	96.5	87.1
99420	84.5	78.2	92.7	84.9
96412	**87.5**	76.0	**98.7**	**85.8**

จาก Table 3 จะเห็นได้ว่าแบบจำลองมีประสิทธิผลสูงเมื่อนำไปประยุกต์ใช้งานจริง โดยมีค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน และค่าประสิทธิผลโดยรวมของ 4 ชุดวิชาได้แก่ ชุดวิชา 96408 การจัดการระบบฐานข้อมูล 99419 ความมั่นคงปลอดภัยไซเบอร์ 99420 การโปรแกรมเว็บชุดวิชา และ 96412 การบริหารโครงการด้านเทคโนโลยีสารสนเทศ มีค่าเฉลี่ยของค่าความถูกต้อง = 86.2 ค่าเฉลี่ยของ

ค่าความแม่นยำ = 77.4 ค่าเฉลี่ยของค่าครบถ้วน = 95.7 และค่าเฉลี่ยของค่าประสิทธิผลโดยรวมของระบบในการใช้งานจริง = 85.6

อธิปรายผลและข้อเสนอแนะ

1. อภิปรายผล

จากผลการดำเนินการพัฒนาระบบการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ ที่มีชื่อเรียก STOU-ASFER ย่อมาจาก STOU-Automatic Student Facial Expression Recognition Model เป็นการนำข้อมูลเข้าจากไฟล์วิดีโอที่ได้จากการสอบออนไลน์ส่งเข้าไปในกระบวนการเตรียมข้อมูล (data preparation) เพื่อตัด VDO ออกเป็นเฟรมภาพใบหน้า แล้วส่งเฟรมภาพนี้เข้าสู่การสร้างแบบจำลองการรับรู้อัตโนมัติจากท่าทางของหน้า (STOU-ASFER) ด้วยอัลกอริธึม Convolution Neural Network เพื่อกรองคุณลักษณะ (feature) ที่สำคัญที่อยู่บนใบหน้าตามท่าทางที่แตกต่างกัน ซึ่งคล้ายกับงานวิจัยของ Bargal *et al*. (2016) ที่นำเสนอการใช้วิธีการวิเคราะห์ภาพโดยใช้การประมวลผลภาพและโครงข่ายประสาทเทียมคอนโวลูชัน (CNN) เพื่อใช้ในการสอบและตรวจสอบความถูกต้องของการสอบออนไลน์ (eExam) และการคุมสอบออนไลน์ (online proctoring) โดยมีลักษณะสำคัญของใบหน้าที่จับ และงานวิจัยของ Panlima and Sukvichai (2023) ที่พัฒนาระบบการรับรู้อัตโนมัติ (automatic recognition system) ของอารมณ์ที่แสดงผ่านใบหน้าด้วยอัลกอริธึม CNN+MLP แต่ในงานวิจัยนี้ดำเนินการขยายขีดความสามารถโดยการเพิ่มประสิทธิผลการทำงานของแบบจำลอง โดยใช้อัลกอริธึม MLP เพื่อหาความสัมพันธ์ของแต่ละเฟรมภาพตามอนุกรมเวลา โดยความสัมพันธ์นี้จะนำไปใช้ในการให้น้ำหนักและผลทำนายทุจริตการสอบ

การประเมินประสิทธิผลของแบบจำลอง จากค่าความถูกต้อง ค่าความแม่นยำ ค่าความครบถ้วน และค่าประสิทธิผลโดยรวมของระบบในการใช้งานจริง (จากไฟล์วิดีโอการสอบออนไลน์) จำนวน 4 ชุดวิชาได้แก่ ชุดวิชา 96408 การจัดการระบบฐานข้อมูล 99419 ความมั่นคงปลอดภัยไซเบอร์ 99420 การโปรแกรมเว็บชุดวิชา และ 96412 การบริหารโครงการด้านเทคโนโลยีสารสนเทศ มีค่าเฉลี่ยของค่าความถูกต้อง = 86.2 ค่าเฉลี่ยของค่าความแม่นยำ = 77.4 ค่าเฉลี่ยของค่าครบถ้วน = 95.7 และค่าเฉลี่ยของค่าประสิทธิผลโดยรวมของระบบในการใช้งานจริง = 85.6 และมีค่าความถูกต้องของแบบจำลองสำหรับการใช้งานจริง

2. ข้อเสนอแนะ

เครื่องคอมพิวเตอร์สำหรับรันโปรแกรมมีพื้นที่จัดเก็บไฟล์มากกว่า 25 GB. (ต่อการสอบ 1 ครั้งสำหรับนักศึกษา 25 - 50 คน และเป็นข้อสอบปรนัย)

กรณีการสอบออนไลน์ ที่มีนักศึกษาเกิน 25 คน โปรแกรม WebeX จะทำการสุ่ม (random) สลับใบหน้ากลับไปกลับมาของแต่ละตำแหน่งใหม่ ทำให้แบบจำลองเกิดข้อจำกัดของงานวิจัย ที่ต้องมีการระบุตำแหน่งของนักศึกษาแบบคงที่ เพื่อให้แบบจำลองแสดงผลได้ถูกต้อง

รายงานสรุปผลการทำงาน สามารถใช้เป็นเอกสารเพื่อสนับสนุนหรือประกอบการพิจารณาตัดสินพฤติกรรมส่อทุจริตเท่านั้น

การสอบออนไลน์ของนักศึกษา นอกจากจะต้องแสดงภาพใบหน้าทั้งหมด 25 ใบหน้าต่อครั้งแล้ว จะต้องมีพื้นที่สำหรับการสนทนา หรือ chat ของกรรมการคุมสอบและนักศึกษาอีกด้วย ทำให้พื้นที่แสดงใบหน้าของนักศึกษามีขนาดเล็กลง ดังนั้นในการพัฒนาต่อยอดงานวิจัยนี้ ควรพัฒนาให้โปรแกรมมีความสามารถเพิ่มมากขึ้นในจดจำภาพใบหน้าของนักศึกษาที่มีขนาดเล็กได้อย่างถูกต้อง และแม่นยำ

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบพระคุณกลุ่มตัวอย่างที่เข้าร่วมโครงการวิจัย ซึ่งโครงการวิจัยนี้ดำเนินการภายใต้การขอรับรองจริยธรรมการวิจัยในมนุษย์ รหัส STOUIRB 2565/019.1703

เอกสารอ้างอิง

มหาวิทยาลัยสุโขทัยธรรมาธิราช. (2566, 5 พฤษภาคม). *ระเบียบและแนวปฏิบัติในการสอบออนไลน์*. https://www.stou.ac.th/offices/ore/rere/goto/newsfeed353

Abdullah, M., Ahmad, M., & Han, D. (2020). Facial expression recognition in videos: An CNN-LSTM based model for video classification. *2020 International Conference on Electronics, Information, and Communication (ICEIC)*, 1–3.

Aggarwal, C. (2018). *Neural networks and deep learning*. Springer.

Bargal, A., Barsoum, E., Ferrer, C., & Zhang, C. (2016, October). Emotion recognition in the wild from videos using images. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (pp. 433–436).

Ekman, P. (1992). Facial expressions of emotion: An old controversy and new findings. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences*, *335*(1273), 63–69.

Ekundayo, O. S., & Viriri, S. (2021). Facial expression recognition: A review of trends and techniques. *IEEE Access*, *9*, 136944–136973.

Frank, M. G. (2001). Facial expressions. In *International encyclopedia of the social & behavioral sciences* (pp. 5230–5234).

Gavade, P. A., Bhat, V. S., & Gavade, A. B. (2023). Learning face expression features from video using spatio-temporal feature extractor and CNN-LSTM. *2023 Seventh International Conference on Image Information Processing (ICIIP)*, 46–50.

Google AI for Developers. (2022, November 11). *Face landmark detection guide*. https://ai.google.dev/edge/mediapipe/solutions/vision/face_landmarker

Guodong, G., & Na, Z. (2019). A survey on deep learning-based face recognition. *Computer Vision and Image Understanding*, *189*, 1–37.

Haghpanah, M., Saeedizade, E., Masouleh, M. T., & Kalhor, A. (2022). Real-time facial expression recognition using facial landmarks and neural networks. *2022 International Conference on Machine Vision and Image Processing (MVIP)*, 1–7.

Indolia, S., Kumar, A., Goswami, M., & Pooja, A. (2018). Conceptual understanding of convolutional neural network- A deep learning approach. *Procedia Computer Science*, *132*, 679–688.

Iqbal, T., Ali, T., Shaf, A., & Ali, M. S. (2023). Enhancing online exam security: Deep learning algorithms for cheating detection. *2023 International Conference on Frontiers of Information Technology (FIT)*, 126–131.

Lapan, M. (2020). *Deep reinforcement learning hands-on* (2nd ed.). Packt Publishing.

Mohamad Nezami, O., Dras, M., Hamey, L., Richards, D., Wan, S., & Paris, C. (2020). Automatic recognition of student engagement using deep learning and facial expression. In *Machine learning and knowledge discovery in databases. ECML PKDD 2019. Lecture notes in computer science* (Vol. 11908, pp. 273–289). Springer.

Noorbehbahani, F., Mohammadi, A., & Aminazadeh, M. (2022). A systematic review of research on cheating in online exams from 2010 to 2021. *Education and Information Technologies*, *27*(6), 8413–8460.

Panlima, A., & Sukvichai, K. (2023). Investigation on MLP, CNNs and Vision Transformer models performance for extracting a human emotions via facial expressions. *Proceedings of the IEEE International Conference on Instrumentation, Control, Artificial Intelligence, and Robotics (ICA-SYMP)*, 127–130.

Ramzan, M., Abid, A., Bilal, M., Aamir, K. M., Memon, S. A., & Chung, T.-S. (2024). Effectiveness of pre-trained CNN networks for detecting abnormal activities in online exams. *IEEE Access*, *12*, 21503–21519.

Soltane, M., & Laouar, M. R. (2021). A smart system to detect cheating in the online exam. *2021 International Conference on Information Systems and Advanced Technologies (ICISAT)*, 1–5.

Zhang, C., Pan, X., Huapeng, L., Gardiner, A., Sargent, I., Hare, J., & Atkinson, P. (2018). A hybrid MLP-CNN classifier for very fine resolution remotely sensed image classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, *140*, 133–144.