บทที่ 2
ทบทวนวรรณกรรม และงานวิจัยที่เกี่ยวข้อง

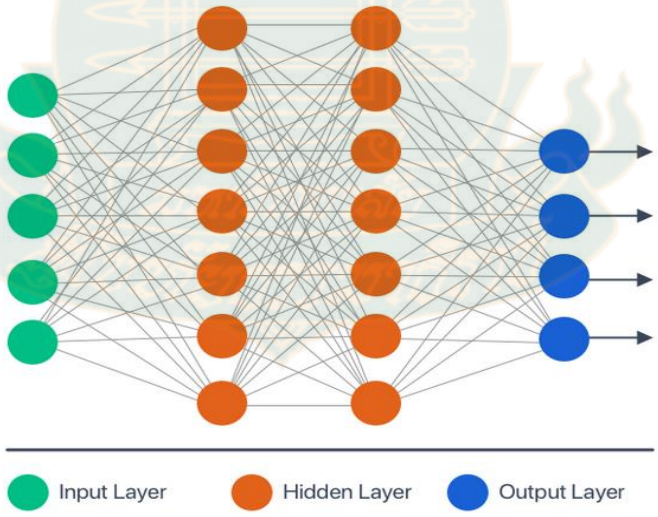
การวิจัยครั้งนี้เป็นการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ ทั้งนี้ผู้วิจัยขอนำเสนอแนวคิดและทฤษฎีที่เกี่ยวข้อง ซึ่งแบ่งออกได้เป็น 3 ตอน ดังนี้

1. ตอนที่ 1 แนวคิดเกี่ยวกับการวิเคราะห์ และจำแนกข้อมูลการแสดงออกทางสีหน้าสำหรับการทุจริตการสอบด้วยโครงข่ายประสาทเทียมคอนโวลูชัน
2. ตอนที่ 2 แนวคิดเกี่ยวกับการแสดงออกทางสีหน้า
3. ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

โดยในแต่ละตอนมีรายละเอียดดังต่อไปนี้

1. ตอนที่ 1 แนวคิดเกี่ยวกับการวิเคราะห์ และจำแนกข้อมูลการแสดงออกทางสีหน้าสำหรับการทุจริตการสอบด้วยโครงข่ายประสาทเทียมคอนโวลูชัน

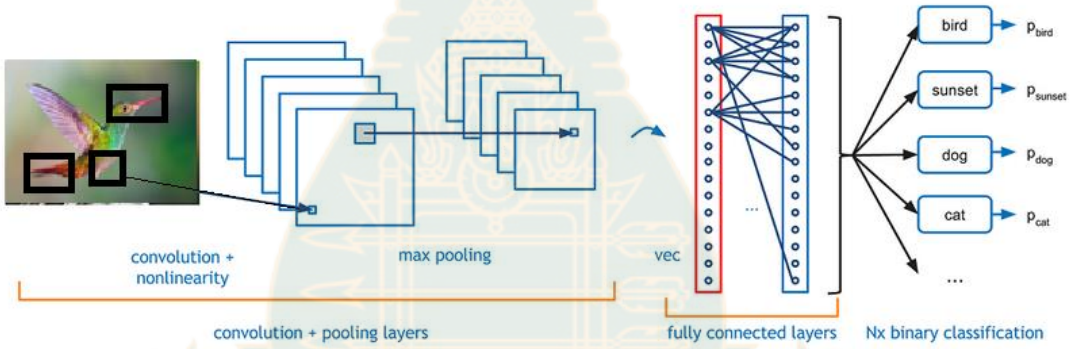
โครงข่ายประสาทเทียม (Neural Network: NN) เป็นโครงข่ายที่จำลองหลักการทำงานโครงข่ายประสาทที่อยู่ในสมองของมนุษย์ ซึ่งมีการแบ่งลักษณะของโครงสร้างการทำงานออกเป็นชั้นๆ ที่เรียกว่าเลเยอร์ (layer) การทำงานของโครงข่ายประสาทเทียมโดยทั่วไป ดังภาพที่ 2.1



ภาพที่ 2.1 โครงข่ายประสาทเทียม (Neural Network)
(ที่มา: https://www.xenonstack.com/blog/artificial-neural-network-applications)

จากภาพที่ 2.1 โครงข่ายประสาทเทียม ประกอบด้วย 3 ชั้นหลัก ได้แก่ ชั้นนำเข้า (input layer) ชั้นซ่อน (hidden layer) และชั้นแสดงผลลัพธ์ (output layer) ที่เรียกว่า โครงข่ายการรับรู้หลายชั้น หรือมัลติเลเยอร์เพอร์เซ็ปตรอน (Multi-Layer Perceptron: MLP) ซึ่งจำนวนหน่วย (unit) หรือโหนด (node) ของชั้นแสดงผลลัพธ์ขึ้นอยู่กับจำนวนหมวดหมู่ (category) ที่ทำการจำแนกประเภท (classification) งานวิจัยของ LeCun et al. (1998) ได้ปรับปรุงโครงข่ายโดยการเพิ่มการคำนวณที่มีลักษณะแบบคอนโวลูชัน (convolutional) เข้าไปในโครงข่าย ซึ่งเรียกว่า โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network: CNN)

โครงข่ายประสาทเทียมคอนโวลูชัน ถูกออกแบบมาโดยเฉพาะสำหรับงานที่ใช้ภาพเป็นข้อมูลนำเข้า (Behnke, 2003) ชั้นของโครงข่ายประสาทเทียมคอนโวลูชันไม่เหมือนกับโครงข่ายประสาทเทียม (ANN) ทั่วไปตรงที่มีชั้นเพอร์เซ็ปตรอนที่อยู่ในลักษณะ 3 มิติ ที่แสดงความกว้าง ความสูง และความลึก ซึ่งเหมาะสมสำหรับการใช้งานข้อมูลเชิงพื้นที่ การวิเคราะห์ภาพ และการรู้จำวัตถุ โดยใช้โครงสร้างเซลล์ประสาทแบบหลายมิติ (multidimensional neurons structures) ดังภาพที่ 2.2



ภาพที่ 2.2 โครงข่ายประสาทเทียมคอนโวลูชัน (Convolutional Neural Network: CNN) (ที่มา: ดัดแปลงจาก https://towardsdatascience.com/covolutional-neural-network-cb0883dd6529)

จากภาพที่ 2.2 โครงข่ายประสาทเทียมคอนโวลูชัน จะทำการจำลองภาพการมองเห็นของมนุษย์ที่มองพื้นที่เป็นส่วนย่อยๆ และนำกลุ่มของพื้นที่ส่วนย่อยๆ มาผสานกัน เพื่อแสดงผลสิ่งที่มองเห็นว่าคืออะไรตามหลักการสถิติความน่าจะเป็น โดยการทำงานของโครงข่ายประสาทเทียมคอนโวลูชัน ประกอบด้วย 3 ขั้นตอน ได้แก่ 1) การกำหนดตัวกรอง (filter) หรือเคอร์แนล (kernel) สำหรับกรองลักษณะสำคัญที่ใช้ในการรู้จำวัตถุ ในรูปแบบตารางสองมิติ 2) การกำหนดการเลื่อน (stride) สำหรับกำหนดการเลื่อนของตัวกรอง (filter) ไปกี่ขั้นตอน (step) และ 3) การกำหนดการรวมกันสูงสุด

(max pooling) สำหรับหาค่าสูงสุดในบริเวณที่ตัวกรองทาบอยู่บนวัตถุมาเป็นผลลัพธ์ ตามหลักการคำนวณการจำแนกข้อมูล (data classification)

ตัวอย่างงานวิจัยที่มีการนำ CNN มาใช้ในการวิเคราะห์ข้อมูล เช่น Akhand et al. (2021) นำชุดข้อมูล (dataset) ของการจดจำอารมณ์ใบหน้าของมนุษย์ (Facial Emotion Recognition: FER) และเทคนิคการถ่ายทอดการเรียนรู้ (Transfer Learning: TL) มาช่วยลดเวลาในการฝึกสอนแบบจำลองการเรียนรู้เชิงลึก ให้มีความรวดเร็วและถูกต้องสูง และในงานวิจัยของ Mehendale (2020) นำภาพคนจำนวน 154 คน ที่มีภาพประมาณ 10,000 ภาพ มาทำการวิเคราะห์ภาพอารมณ์บนใบหน้าด้วย CNN และเพิ่มการปรับขนาดสเกลของเวกเตอร์ (vector) เพื่อหาค่าความเหมาะสมที่สุดในการวิเคราะห์เชิงทำนายภาพ และ Gupta and Vishwamitra (2021) นำเสนอวิธีการจดจำอารมณ์ใบหน้าของมนุษย์ จากวิดีโอที่มีการระบุขนาดสตรีม (stream) และเวลาในการวิเคราะห์ภาพแบบเรียลไทม์ จากแนวคิดเกี่ยวกับโครงข่ายประสาทเทียมคอนโวลูชันที่กล่าวมาข้างต้น การวิจัยภาพอารมณ์บนใบหน้ายังคงมีการพัฒนาในด้านประสิทธิภาพ ความถูกต้อง โดยเฉพาะด้านการเรียนการสอน และการสอบออนไลน์ ดังนั้นในงานวิจัยนี้ จึงขยายขีดความสามารถของงานวิจัยที่เกี่ยวข้องมาประยุกต์ใช้ในแบบจำลองการตรวจจับพฤติกรรมทุจริตระหว่างการสอบออนไลน์ (detection of cheating at online examinations) ที่แสดงออกทางใบหน้าของผู้เข้าสอบ

2. ตอนที่ 2 แนวคิดเกี่ยวกับการแสดงออกทางสีหน้า

การแสดงออกทางสีหน้า (facial expressions) เป็นช่องทางหนึ่งในการทำความเข้าใจและตีความการแสดงออกทางสีหน้าระหว่างการปฏิสัมพันธ์ทางสังคม โดยงานวิจัยที่มีอยู่ในปัจจุบันส่วนใหญ่มุ่งเน้นไปที่การรับรู้การแสดงออกทางสีหน้าแบบอัตโนมัติ (Automatic Facial Expression Recognition: AFER) ตามทฤษฎีของ Ekman (Ekman and Keltner, 1997) ซึ่งแนะนำอารมณ์ความรู้สึกพื้นฐานที่ใช้อย่างเป็นสากลในทุกวัฒนธรรมไว้ 6 อารมณ์ ได้แก่ ความสุข (happiness) ความประหลาดใจ (surprise) ความโกรธ (anger) ความเศร้า (sadness) ความกลัว (fear) และการรังเกียจ (disgust) ในการตรวจสอบการแสดงออกทางอารมณ์ความรู้สึกที่แสดงออกทางใบหน้า จะอ้างอิงรูปแบบของใบหน้าที่มีการแสดงอารมณ์ความรู้สึกออกมาในลักษณะหรือรูปแบบต่างๆ ดังกล่าว โดยอ้างอิงได้จากตำแหน่งดวงตา คิ้ว จมูก และปาก แล้วมาทำการตรวจสอบเพื่อแสดงการตรวจจับใบหน้าว่าเป็นอารมณ์ความรู้สึกใดที่แสดงออกทางใบหน้า (Samal and Iyengar, 1992; Tarnowski et al., 2017)

ตัวอย่างงานวิจัยที่เกี่ยวข้องกับการแสดงออกทางสีหน้า เช่น Jia et al. (2021) นำเสนองานวิจัยที่เกี่ยวข้องกับการแสดงออกทางสีหน้าที่เกิดขึ้นเองโดยธรรมชาติเปรียบเทียบกับการแสดงออกด้วยท่าทางจริง (Spontaneous vs. Posed: SVP) เนื่องจากบางครั้งการแสดงออกทางสีหน้าของคน

อาจไม่ตรงกับความจริงได้ โดยงานวิจัยนี้ใช้ฐานข้อมูลอ้างอิงที่หลากหลายและเทคนิควิธีการตรวจจับการแสดงออกทางสีหน้าด้วยคอมพิวเตอร์วิทัศน์ (computer vision) ในรูปแบบต่างๆ ซึ่งกล่าวถึงงานวิจัยของ Racoviteanu et al. (2019) ที่นำเสนอการใช้สถาปัตยกรรม AlexNet CNN ในการตรวจจับอารมณ์จากการแสดงออกทางสีหน้าและแปลผลข้อมูลจากฐานข้อมูล DISFA และทดสอบบน SPOS โดยใช้วิธีการจำแนกข้อมูลแบบซัพพอร์ตเวกเตอร์แมชชีน (Support Vector Machine: SVM) ผลที่ได้มีค่าความถูกต้อง (accuracy) ร้อยละ 72.10 อีกทั้งยังมีงานวิจัยของ Mandal et al. (2016) ที่ได้สำรวจคุณลักษณะ (feature) หลายประการ รวมถึง CNN เชิงลึก (deep CNN) การหาโลคอลเฟสควอนไทเซชัน (Local Phase Quantization: LPQ) การไหลของแสงที่หนาแน่น (dense optical flow) และฮิสโตแกรมของการไล่ระดับสี (Histogram of Gradient: HOG) เพื่อจำแนกการแสดงออกทางสีหน้าจากรอยยิ้มด้วยวิธี SVP โดยใช้วิธีการจำแนกข้อมูลแบบ SVM เชิงเส้น (Linear SVM)

อย่างไรก็ตาม งานวิจัยในปัจจุบันมีช่องว่างงานวิจัย (research gap) สำหรับการคำนวณที่ซับซ้อนด้านการแสดงออกทางสีหน้าของคน เช่น วิธีการเรียนรู้เชิงลึกที่นำมาใช้ในการตรวจจับการแสดงออกทางสีหน้าและการประมวลผลทางอารมณ์ แบบ SVP ดังนั้น หากมีการนำวิธีการเรียนรู้เชิงลึกด้วย CNN มาประยุกต์ใช้ในการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์จากการแสดงออกทางใบหน้าเพื่อให้มีความถูกต้องและแม่นยำมากยิ่งขึ้น จะทำให้ผู้คุมสอบสามารถรับรู้ถึงการแสดงออกทางสีหน้าของผู้เข้าสอบในห้องสอบออนไลน์ได้แบบเรียลไทม์ (real time) และช่วยป้องกันการทุจริตในการสอบออนไลน์ได้อย่างเหมาะสมและทันท่วงที ซึ่งจะเป็นงานวิจัยที่มีประโยชน์ทางการศึกษา และสร้างนวัตกรรมที่สำคัญแก่องค์กร

3. ตอนที่ 3 งานวิจัยที่เกี่ยวข้อง

งานวิจัยนี้เป็นการพัฒนานวัตกรรมการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ ซึ่งอาศัยแนวคิดของโครงข่ายประสาทเทียมคอนโวลูชันมาพัฒนาร่วมกับแบบจำลองในการจดจำ วิเคราะห์ และจำแนกการเคลื่อนไหวของตำแหน่งต่างๆ บนใบหน้า เช่น การหันหน้า ดวงตา คิ้ว และปาก สำหรับจำแนกการทุจริตจากการแสดงออกทางสีหน้า เพื่อส่งเสริมให้เกิดความถูกต้องในการทำนาย เพิ่มความน่าเชื่อถือ โดยนำแนวคิดการตรวจจับใบหน้ามากกว่าการให้ความสนใจลักษณะทางอารมณ์ จาก 6 ลักษณะ (Yang et al., 2018) ซึ่งกล่าวได้ว่า การตรวจจับการทุจริตในการสอบออนไลน์เป็นเรื่องที่ท้าทายตั้งแต่ในอดีตจนถึงปัจจุบัน ดังงานวิจัยของ Roger (2006) ที่ใช้การควบคุมการทุจริตในการสอบออนไลน์โดยการติดตั้งโปรแกรมสำหรับการสอบออนไลน์ลงในเครื่องคอมพิวเตอร์เพื่อให้ผู้คุมสอบสามารถมองเห็นจอภาพของผู้สอบแต่ละคนได้ นอกจากนี้ยังมีงานวิจัยที่ใช้เทคนิคหลากหลายในการตรวจจับการทุจริตในการสอบออนไลน์ ดังนี้

Jadi (2021) ได้นำเสนอการใช้วิธีการวิเคราะห์ภาพโดยใช้การประมวลผลภาพและโครงข่ายประสาทเทียมคอนโวลูชัน (CNN) เพื่อใช้ในการสอบและตรวจสอบความถูกต้องของการสอบออนไลน์ (eExam) และการคุมสอบออนไลน์ (online proctoring) โดยมีลักษณะสำคัญของใบหน้าที่จับ ได้แก่ สัดส่วนของดวงตา คิ้ว ความกว้างของปาก และความสูงของปาก ร่วมกับการแสดงอารมณ์ปกติ มีความสุข และโกรธ มาใช้ในการพิจารณาร่วมด้วย ทั้งนี้ ภาพที่ใช้จับเป็นภาพสีที่มีความละเอียดของภาพสูง (high definition) ผลของการตรวจจับมีความไวที่ให้ค่าความถูกต้องสูง

Li et al. (2021) ได้นำเสนอการใช้วิธีการวิเคราะห์ด้วยภาพ (visual analytics approach) ในการคุมสอบออนไลน์ โดยเริ่มจากการกำหนดข้อมูลความต้องการในการออกแบบการคุมสอบออนไลน์จากคณาจารย์ในมหาวิทยาลัยและทีมผู้สอน จากนั้นทำการวิเคราะห์ข้อมูลด้วยภาพจากวิดีโอที่บันทึกไว้และการเคลื่อนที่ของเมาส์ของผู้เข้าสอบแต่ละคนในขณะทำการสอบออนไลน์ ทีมวิจัยได้ตรวจจับภาพของการกระทำที่ต้องสงสัยว่ามีการทุจริตในการสอบ โดยวิเคราะห์จากการเคลื่อนที่ของศีรษะ (head pose analysis) และภาพการเคลื่อนที่ของเมาส์ (mouse movement visualization) ว่าอยู่นอกโปรแกรมที่ใช้ทำข้อสอบหรือไม่ ผลการประเมินการวิจัยจากการทดลองจริงและการสัมภาษณ์ผู้เชี่ยวชาญพบว่าวิธีการวิเคราะห์ด้วยภาพนี้มีประสิทธิภาพและสามารถใช้ประโยชน์ได้ดี

Ousmane et al. (2019) ได้พัฒนาระบบการรับรู้อัตโนมัติ (automatic recognition system) ของอารมณ์ที่แสดงผ่านใบหน้า ซึ่งถูกนำมาประยุกต์ใช้กับการสอบปากคำของตำรวจหรือการจำลองการสัมภาษณ์ โดยระบบนี้มีพื้นฐานมาจากการแบ่งฟังก์ชันการทำงาน ได้แก่ การตรวจจับและติดตามใบหน้า การแสดงออกทางสีหน้า ซึ่งเริ่มจากลำดับภาพวิดีโอ การสกัดคุณลักษณะเฉพาะ (feature extraction) และการรับรู้การแสดงออกทางอารมณ์ต่างๆ โดยใช้ชุดข้อมูลทางด้านอารมณ์ FER2013 พัฒนาแบบจำลองเครือข่ายคอนโวลูชันเชิงลึก (deep convolution network model) ที่สามารถตรวจจับอารมณ์ทางใบหน้าด้วยอัลกอริทึม Viola and Jones

นอกจากนี้ วัชรชัย คงศิริวัฒนา และคณะ (2564) ได้พัฒนาระบบตรวจสอบการเข้าชั้นเรียนโดยการใช้กล้องเว็บแคมในการตรวจจับใบหน้าของนักศึกษาในชั้นเรียนแบบเรียลไทม์ และนำผลการแสดงอารมณ์ทางใบหน้าของนักศึกษาขณะเรียนในช่วงเวลาดังกล่าวมาทำการประเมินผลความสนใจในการเรียน โดยการตรวจจับใบหน้าของนักศึกษาจะมีการทำงานร่วมกับแบบจำลองที่สร้างขึ้น เพื่อจำแนกแยกแยะและรู้จำใบหน้าของนักศึกษาแต่ละคน จากนั้นนำมาทำการเปรียบเทียบกับชุดข้อมูลแบบจำลองที่สร้างขึ้นเพื่อตรวจจับอารมณ์ที่แสดงออกทางใบหน้า เพื่อดูว่านักศึกษาแต่ละคนมีการแสดงอารมณ์อยู่ในสถานะอะไร จากนั้นระบบจะทำการคำนวณร้อยละความสนใจเรียน และไม่สนใจเรียนของนักศึกษา รวมถึงสถานะการเข้าชั้นเรียนของนักศึกษาแต่ละคนออกเป็นผลสรุปในตอนสิ้นสุดการทำงาน

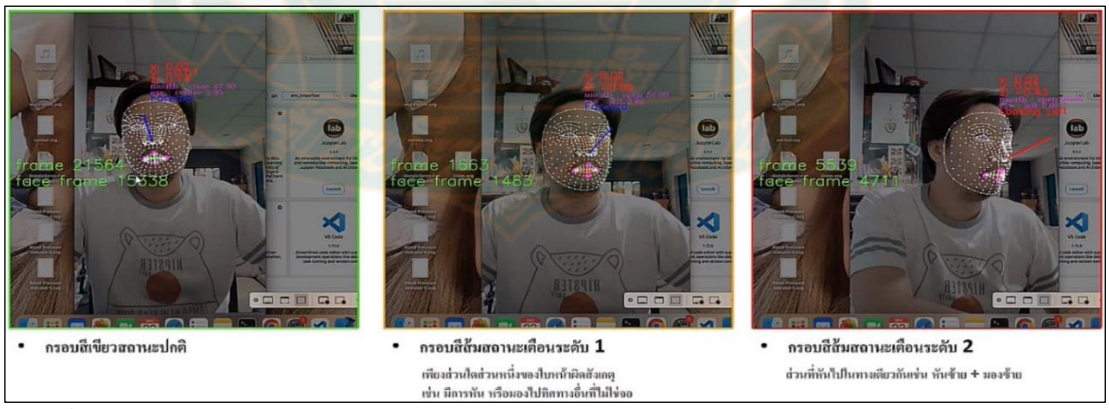
ดังนั้นทีมวิจัยจึงได้นำแนวคิดจากงานวิจัยดังกล่าวข้างต้นทั้งหมดมาต่อยอด เพื่อสร้างนวัตกรรมทางการศึกษา โดยการพัฒนานวัตกรรมสำหรับการประเมินพฤติกรรมทุจริตระหว่างการสอบออนไลน์ด้วยปัญญาประดิษฐ์บนระบบการรับรู้การแสดงออกทางสีหน้าของนักศึกษาแบบอัตโนมัติ เพื่อสนับสนุนการสอบออนไลน์ของมหาวิทยาลัยสุโขทัยธรรมาธิราช ให้มีประสิทธิภาพเพิ่มมากขึ้น และสามารถตรวจจับการทุจริตแบบอัตโนมัติในเวลาจริง โดยพฤติกรรมที่ส่อทุจริตที่กำหนดในเบื้องต้นสำหรับงานวิจัยนี้ อ้างอิงมาจากระเบียบมหาวิทยาลัยสุโขทัยธรรมาธิราช ว่าด้วย การสอบออนไลน์ของนักศึกษาระดับปริญญาตรีและระดับต่ำกว่าปริญญา พ.ศ. 2563 (https://www.stou.ac.th/Offices/ore/rere/goto/page349) ซึ่งระบบที่พัฒนาจะสามารถตรวจจับว่ามีพฤติกรรมที่ส่อทุจริตใน 3 ประเด็นหลักๆ ดังนี้

1) พยายามดูคำตอบจากผู้อื่นหรือแหล่งข้อมูลอื่นจากช่องทางต่างๆ สามารถตรวจจับพฤติกรรมที่ส่อทุจริตด้วยการตรวจจับการขยับของใบหน้าและลูกตา เช่น ตรวจจับการมองซ้ายหรือมองขวา มองต่ำหรือหลับตา ก้มหน้า หันหน้า หรือกระพริบตาถี่มาก

2) พูดคุยหรือติดต่อกับบุคคลอื่นระหว่างสอบ หรือมีการใช้โทรศัพท์ สามารถตรวจจับพฤติกรรมที่ส่อทุจริตด้วยการตรวจจับการขยับของปาก

3) มีบุคคลอื่นอยู่ในห้องสอบหรือนั่งใกล้ๆ ผู้เข้าสอบ สามารถตรวจจับพฤติกรรมที่ส่อทุจริตด้วยการตรวจจับว่ามีคนมากกว่า 1 คนอยู่ในเฟรมเดียวกัน

เนื่องจากการตรวจจับการหันของใบหน้าเพียงอย่างเดียว อาจไม่สามารถบอกได้ว่าผู้เข้าสอบมีพฤติกรรมที่ส่อทุจริตหรือไม่ จึงจำเป็นต้องมองควบคู่ไปกับตา เช่น ผู้เข้าสอบตำแหน่งหน้าตรงปกติ แต่ตามีการมองไปทางซ้าย ซึ่งการตรวจจับพฤติกรรมที่ส่อทุจริตมักจะกำหนดคุณลักษณะ (feature) ที่ส่อทุจริตด้วยตำแหน่งการมองของตา และการเปิดปิดของปาก ดังภาพที่ 2.3



ภาพที่ 2.3 ตัวอย่างการตรวจจับพฤติกรรมส่อทุจริต โดยดูจากตำแหน่งการมองของตาและการเปิดปิดของปาก