การเรียนรู้ของเครื่องเพื่อทำนายระดับความรุนแรงของความผิดปกติของความยืดหยุ่นปอดของพนักงานโรงงาน

Machine learning for predicting the severity of restrictive lung defect among factory workers

ณัฐวุฒิ แถมเงิน[1], ปกรณ์ ล่องทอง[1], พงศศรัณย์ ทองหนูนุ้ย[1], กนกวรรณ ละอองศรี[2], อนามัย เทศกะทึก[2], พีรพล ศิริพงศ์วุฒิกร[3], ณฐนนท์ เทพตะขบ[1] และ วิริยะ มหิกุล[1*]
Nattawut Theamngoen[1], Pakorn Longthong[1], Phongsaran Thongnunuy[1], Kanokwan Laoongsri[2], Anamai Thetkathuek[2], Peerapon Siriipongwutikorn[3], Nathanon Theptakob[1] and Wiriya Mahikul[1*]

Received: 14 June 2023 ; **Revised**: 25 July 2023 ; **Accepted**: 4 August 2023

บทคัดย่อ

กลุ่มโรคที่มีความผิดปกติของความยืดหยุ่นของปอดโดยเฉพาะอย่างยิ่งโรคนิวโมโคนิโอซีส (Pneumoconiosis) เป็นโรคที่พบมากในผู้คนที่มีการสัมผัสสภาพแวดล้อมที่มีฝุ่นแร่ การตรวจสมรรถภาพปอดด้วยวิธีสไปโรเมตรีย์ (Spirometry) เป็นวิธีมาตรฐานในการทดสอบสมรรถภาพการทำงานของปอด อย่างไรก็ตาม วิธีดังกล่าวมีข้อจำกัดเนื่องจากค่าใช้จ่ายและอุปกรณ์ในการตรวจมีราคาแพง และประสบการณ์ของผู้อ่านผลการตรวจ ส่งผลให้ผู้พนักงานกลุ่มเสี่ยงไม่สามารถเข้าถึงการตรวจสมรรถภาพปอดได้ทันท่วงที การศึกษานี้มีวัตถุประสงค์เพื่อนำการเรียนรู้ของเครื่องมาใช้ทำนายระดับความรุนแรงของความผิดปกติของความยืดหยุ่นของปอดเบื้องต้น ก่อนที่จะนำไปสู่การตรวจสไปโรเมตรีย์ต่อไป โดยแบ่งระดับความรุนแรงของความยืดหยุ่นของปอดเป็น 3 กลุ่ม คือ กลุ่มปกติ รุนแรงน้อย และ รุนแรงปานกลางถึงมาก การศึกษาได้นำข้อมูลจากการตรวจสมรรถภาพปอดในกลุ่มพนักงานของโรงงานเฟอร์นิเจอร์ ทั้งหมด 685 คน จากศึกษาภาคตัดขวาง มาใช้สร้างแบบจำลองการเรียนรู้ของเครื่องด้วยเทคนิค 6 แบบ ได้แก่ Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost และ Support Vector Machine (SVM) พบว่าผลลัพธ์การฝึกแบบจำลองที่ดีที่สุด คือ แบบจำลอง Random Forest ร่วมกับเทคนิคการจัดการข้อมูลไม่สมดุล และการคัดเลือกตัวแปรที่สำคัญ 20 ตัวแปรด้วยวิธีการ Recursive Feature Elimination (RFE) พบว่า กลุ่มตัวแปรที่สำคัญในการทำนายระดับความรุนแรง ได้แก่ น้ำหนัก ส่วนสูง อายุ ประวัติการศึกษา ชั่วโมงการทำงาน การสูบบุหรี่ การใช้หน้ากากอนามัย และอาการบางอย่างเกี่ยวกับระบบทางเดินหายใจ เช่น หายใจติดขัด และอาการการมีเสมหะ โดยมีค่าเฉลี่ยของ F1-score, precision, recall และ accuracy เท่ากับ 0.74, 0.74, 0.76 และ 0.75 ตามลำดับ แบบจำลองทำนายประสิทธิภาพปอดถูกนำไปสร้างเป็นเว็บแอปพลิชันเพื่อให้สามารถใช้งานได้ง่าย และได้มีการนำไปให้พนักงานในโรงงานตรวจคัดกรองเบื้องต้น ซึ่งผู้ใช้มีความพึงพอใจต่อประสิทธิภาพการคัดกรอง ความสะดวกรวดเร็วในการใช้งาน และการประหยัดค่าใช้จ่ายจากการใช้แอปพลิเคชันตรวจคัดกรองนี้

คำสำคัญ: กลุ่มโรคที่มีการจำกัดการขยายตัวของปอด, พนักงานโรงงาน, สไปโรเมตรีย์, การเรียนรู้ของเครื่อง, เว็บแอปพลิเคชัน

Abstract

Restrictive lung disease such as pneumoconiosis is the most common disease among people working in dusty environment such as mines and in industry. The gold standard diagnosis for this disease is spirometry, which is used to evaluate the lung performance. However, this tool has certain limitations such as high service costs, limited access

[1] วิทยาลัยแพทยศาสตร์ศรีสวางควัฒน ราชวิทยาลัยจุฬาภรณ์ กรุงเทพฯ 10210
[2] คณะสาธารณสุขศาสตร์ มหาวิทยาลัยบูรพา จังหวัดชลบุรี 20131
[3] ภาควิชาวิศวกรรมคอมพิวเตอร์ คณะวิศวกรรมศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี กรุงเทพฯ 10140
[1] Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok 10210
[2] Faculty of Public Health, Burapha University, Chonburi Province 20131
[3] Department of Engineering Computer, Faculty of Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140
* Corresponding author: Wiriya Mahikul, Princess Srisavangavadhana College of Medicine, Chulabhorn Royal Academy, Bangkok, Thailand 10210
E-mail: wiriya.mah@cra.ac.th

to the device, and availability of specialists. These limitations impede early detection of this disease. The objective of this study is to utilize machine learning algorithms to predict the severity of restrictive lung defects among factory workers, aiding in early identification before proceeding to the spirometry test. Three severity classes considered. - Normal, Mild, and Moderate or Severe. By using spirometry's results and behavioral data among 685 workers from a cross-sectional study in a furniture factory in Thailand, six machine learning algorithms were developed. They were Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, XGBoost and Support Vector Machine (SVM). The best model was Random Forest with Synthetic Minority Oversampling (SMOTE) to deal with imbalance class and Recursive Feature Elimination (RFE) to select most important features. The important features for prediction were weight, height, age, education, hours of work, smoking and mask wearing at the f1-score = 0.746, precision = 0.743, recall = 0.756, and accuracy = 0.75. The model was deployed through a web application for ease of use and the application was used among the factory workers for early screening of the disease. The users were satisfied with the application for its effectiveness, ease of use, time, and cost savings.

Keywords: Restrictive lung disease, factory workers, spirometry, machine learning, web application

บทนำ

หนึ่งในกลุ่มโรคที่จำกัดการขยายตัวของปอดที่พบได้มากที่สุดในประเทศไทย ได้แก่ โรคฝุ่นจับปอด (pneumoconiosis) ซึ่งเป็นโรคที่มักเกิดในผู้ที่ประกอบอาชีพอยู่ในสภาพแวดล้อมที่มีฝุ่นมาก เช่น เหมือง หรือโรงงาน โรคฝุ่นจับปอดเป็นหนึ่งในโรคที่พบมากที่สุดในกลุ่มโรคหลอดลมอุดกั้นเรื้อรังทั้งในประเทศไทยและนานาชาติ การสัมผัสฝุ่นในโรงงานสามารถสร้างความเสียหายให้กับสมรรถภาพปอดอย่างถาวรได้ (Kunpeuk *et al*., 2021) การตรวจสมรรถภาพปอดมีวิธีการมาตรฐาน (gold standard) คือวิธีสไปโรเมตรีย์ (spirometry) ซึ่งเป็นการตรวจสมรรถภาพปอดด้วยเครื่อง Spirometer

ค่าความยืดหยุ่นของปอด (restrictive defect) ใช้เป็นมาตรฐานในการวินิจฉัยโรคฝุ่นจับปอดหรือโรคที่มีความผิดปกติของปอด เช่น โรคหลอดลมโป่งพอง (bronchiectasis), โรคหอบหืด (asthma) โรคถุงลมโป่งพอง (emphysema) และหลอมลมฝอยอักเสบเฉียบพลัน (acute bronchiolitis) เป็นต้น (Martinez-Pitre *et al*., 2022) อย่างไรก็ดี การตรวจด้วยวิธี Spirometry มีข้อจำกัดหลายประการ ได้แก่ งบประมาณในการจัดซื้อเครื่อง Spirometer ความเชี่ยวชาญของบุคลากรผู้ให้บริการ Spirometry มาตรฐานของอุปกรณ์ที่ต้องใช้ควบคู่กับเครื่องตรวจ เป็นต้น ส่งผลให้ทั่วประเทศ มีโรงพยาบาลเพียงร้อยละ 47 เท่านั้นที่มีเครื่อง Spirometer ไว้สำหรับให้บริการผู้ป่วย (Martinez-Pitre *et al*., 2022)

เพื่อแก้ไขข้อจำกัดเกี่ยวกับค่าใช้จ่ายและบุคลากรดังกล่าวข้างต้น ผู้วิจัยจึงนำเทคโนโลยีการเรียนรู้ของเครื่อง (machine learning) มาใช้ในการคัดแยกผู้มีความเสี่ยงโรคปอด โดยได้มีการนำเทคโนโลยีการเรียนรู้ของเครื่องมาใช้กับการวิเคราะห์ความเสี่ยงจากการศึกษาก่อนหน้า โดยเฉพาะโรคมะเร็งปอด (Gould *et al*., 2021, Guo *et al*., 2022, Chandran *et al*., 2023) โดยแนวคิดของการเรียนรู้ของเครื่อง (machine learning) คือการให้ระบบคอมพิวเตอร์เรียนรู้และสร้างแบบจำลองจากข้อมูลเพื่อแก้ปัญหา ซึ่งมีความแตกต่างหลักจากการแก้ปัญหาด้วยการเขียนโปรแกรมแบบดั้งเดิมที่ใส่ข้อมูลขาเข้า (input) และกำหนดกฎหรือชุด คำสั่ง (program) บางอย่างเพื่อให้ได้ผลลัพธ์ที่ต้องการ ในขณะที่การเรียนรู้ของเครื่องเป็นการใส่ข้อมูลให้คอมพิวเตอร์ประมวลผลเพื่อเรียนรู้ความสัมพันธ์ระหว่างข้อมูลขาเข้ากับข้อมูลขาออกเพื่อให้ได้ขั้นตอนการประมวลผลหรือแบบจำลองสำหรับการดำเนินการกับข้อมูลอื่นๆ ที่ตามมาในภายหลัง เช่น งานวิจัยการสร้างแบบจำลองเพื่อนำไปใช้ในการทำนายโอกาสเกิดโรคที่พบมาก (Kumar *et al*., 2021) งานของ Steven J Pascoe และคณะศึกษาการนำลักษณะทางคลินิกมาใช้สำหรับการทำนายผลตรวจสำหรับเครื่อง spirometry ของโรคปอดอุดกั้น (Pascoe *et al*., 2018) งานของ Alan Kaplan และคณะศึกษาการใช้งานการเรียนรู้ของเครื่องสำหรับการทำงานกับการแพทย์ที่เกี่ยวข้องกับระบบทางเดินหายใจและความเป็นไปได้ในการวินิจฉัยโรคหอบหืดและโรคปอดอุดกั้นเรื้อรัง (Kaplan *et al*., 2021) และงานของ Hongbo Liu ศึกษาการใช้ งานการเรียนรู้ของเครื่องสำหรับการวินิจฉัยโรคฝุ่นจับปอดจากข้อมูลแรงงานโรงงานถ่านหิน (Liu *et al*., 2009) จากการศึกษา พบว่าโดยส่วนใหญ่ของงานวิจัยดังที่กล่าวมาเน้นไปที่การวินิจฉัยกลุ่มโรคที่มีการอุดกั้นทางเดินทางเดินหายใจ (obstructive lung disease) โดยเฉพาะอย่างยิ่งโรคปอดอุดกั้นเรื้อรัง การเรียนรู้ของเครื่องโดยรวมให้ความแม่นยำที่น่าพอใจ โดยมีค่า Receiver Operating Characteristic curve (ROC Curve) มากกว่า 0.87แบบจำลองที่ใช้งานกันอย่างแพร่หลายในงานด้านวิทยาศาสตร์ข้อมูลในการจำแนกความผิดปกติแบบแบ่งกลุ่ม (classification) (Hazra *et al*., 2017) ได้แก่ Logistic Regression, Random Forest และ Support Vector Machine อย่างไรก็ตาม งานวิจัยส่วนมากยังไม่มีการศึกษาการคัดกรองโรคในกลุ่มโรคที่ปอดมีขนาดเล็กกว่าปกติ

(restrictive lung disease) เช่น โรคฝุ่นจับปอด (pneumoconiosis) ซึ่งพบมากในกลุ่มแรงงานในโรงงานที่มีฝุ่น โดยทั้ง 2 กลุ่มโรคสามารถวินิจฉัยได้จากค่า Obstructive (ภาวะมีการอุดกั้นหรือมีการตีบของหลอดลม) และ Restrictive (ภาวะมีการจำกัดการขยายตัวของปอด) จากเครื่อง spirometry ดังนั้นการศึกษานี้จึงนำการใช้การเรียนรู้ของเครื่องมาฝึกเพื่อใช้เป็นเครื่องมือคัดกรองเบื้องต้น สำหรับผู้เข้ารับการทดสอบโอกาสประสบภาวะและระดับความผิดปกติของความยืดหยุ่นของปอด (restrictive defect) อันเป็นการเพิ่มโอกาสในการเข้าถึงการประเมินประสิทธิภาพปอดในระยะแรกของความผิดปกติ และนำแบบจำลองมาพัฒนาเป็นเว็บแอปพลิเคชัน เพื่อใช้ในการวิเคราะห์สมรรถภาพปอดพนักงานกลุ่มเสี่ยงต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาการทำนายความผิดปกติในการขยายตัวของปอดโดยการใช้เทคนิกการเรียนรู้ของเครื่อง

2. เพื่อประยุกต์ใช้การเรียนรู้ของเครื่องร่วมกับการพัฒนาเว็บแอปพลิเคชันสำหรับการวิเคราะห์สมรรถภาพปอด

วิธีดำเนินการวิจัย

การจำแนกความรุนแรงของความผิดปกติ

โดยการให้ผู้รับการทดสอบหายใจผ่านเครื่อง Spirometer โดยเร็วและแรง เครื่องจะมีการประมวลค่าต่างๆ เกี่ยวกับสมรรถภาพปอด ค่าที่มีความสำคัญประกอบด้วยค่าแสดงปริมาตรอากาศที่ถูกขับออกมาในวินาทีแรกของการหายใจออกอย่างเร็วและแรงเต็มที่ (force expiratory volume in one second: FEV1) ค่าปริมาตรสูงสุดของอากาศที่หายใจออกจนกระทั่งหายใจเข้าเต็มที่ (force vital capacity: FVC) และค่าที่แสดงการอุดกั้นของหลอดลม (FEV1/FVC) ซึ่งจะบอกถึงระดับของการอุดตัน (obstructive defect) และความยืดหยุ่น (restrictive defect) ของปอด (HITAP, 2017) แต่ก็จำเป็นต้องใช้ข้อมูลอื่นๆ ในการคัดกรองด้วย เช่น ผลการซักประวัติหรือผลการตรวจร่างกายอื่นๆ โดย ในผู้ที่มีความยืดหยุ่นของปอดผิดปกติ (restrictive defect) จะมีค่า FEV1 และ FVC ที่น้อยกว่าปกติแต่ค่า FEV1/ FVC เป็นปกติ ดัง Table 1

การเรียนรู้ของเครื่อง

การเรียนรู้ของเครื่อง (machine learning) คือแนวคิดเพื่อให้ระบบคอมพิวเตอร์สามารถเรียนรู้จากข้อมูลเพื่อแก้ปัญหาด้วยตนเองโดยใช้ข้อมูล ซึ่งมีความแตกต่างจากการแก้ปัญหาด้วยการเขียนโปรแกรมแบบดั้งเดิม เนื่องจากการเขียนโปรแกรมแบบดั้งเดิมจะเป็นการใส่ข้อมูลขาเข้า (input) และกำหนดกฎหรือชุดคำสั่ง (program) บางอย่างเพื่อให้ได้ผลลัพธ์ที่ต้องการ ในขณะที่การเรียนรู้ของเครื่องเป็นการใส่ข้อมูลให้คอมพิวเตอร์ประมวลผลเพื่อเรียนรู้ความสัมพันธ์ระหว่างข้อมูลขาเข้ากับข้อมูลขาออกเพื่อให้ได้ขั้นตอนการประมวลผลหรือแบบจำลอง สำหรับการใช้กับข้อมูลอื่นๆ ที่จะเกิดขึ้นมาในภายหลัง โดยมีตัวอย่างคือ การสร้างแบบจำลองเพื่อนำไปใช้ในการทำนายโอกาสเกิดโรคหัวใจ เป็นต้นโดยการเรียนรู้ของเครื่องถูกแบ่งออกเป็น 3 ประเภท ได้แก่

1. การเรียนรู้แบบมีผู้สอน (supervised learning) คือ การเรียนรู้ผ่านโจทย์ที่มีเฉลยคำตอบชัดเจน กล่าวคือข้อมูลที่มีการระบุลักษณะของประเภทข้อมูลที่ต้องการทำนาย (labels) โดยมีรูปแบบ ทั้งหมด 2 รูปแบบได้แก่

1.1 Classification task เป็นโจทย์ที่มีการกำหนดเฉลยเป็นกลุ่ม ยกตัวอย่างเช่นใช้ข้อมูลส่วนบุคคลได้แก่ เพศ อายุ น้ำหนัก และส่วนสูงเพื่อทำนายว่าบุคคลเป็นหมายมีความเสี่ยงต่อการเป็นโรคเบาหวานหรือไม่ เป็นต้น

1.2 Regession task เป็นโจทย์ที่มีการกำหนดเฉลยเป็นตัวเลข และดำเนินการหาสมการทางคณิตศาสตร์เพื่อสร้างสมการที่สามารถทำนายค่าได้ใกล้เคียงกับค่าเฉลยมากที่สุด ยกตัวอย่างเช่นใช้ข้อมูลสภาพแวดล้อม และสิ่งอำนวยความสะดวกใกล้เคียง เพื่อทำนายราคาบ้านในพื้นที่ที่สนใจ เป็นต้น

โดยในงานวิจัยนี้เป็นการใช้การเรียนรู้ของเครื่องในรูปแบบการเรียนรู้แบบมีผู้สอน ในงานประเภท Classification task โดยใช้ข้อมูลจากแบบสำรวจเพื่อมาใช้ทำนายระดับความเสี่ยงของการประสบภาวะปอดยืดหยุ่นผิดปกตินั่นเอง

2. การเรียนรู้แบบไม่มีผู้สอน (unsupervised learning) คือการเรียนรู้ที่ไม่มีคำตอบตายตัวถูกกำหนดให้ถูกแบ่งเป็น 2 รูปแบบได้แก่

2.1 Clustering task คือการให้แบบจำลองศึกษาลักษณะข้อมูลและทำการแบ่งกลุ่มข้อมูลตามลักษณะที่แบบจำลองพบ

2.2 Dimentionality reduction คือการลดจำนวนมิติของข้อมูลโดยทำให้ลักษณะของข้อมูลเดิมน้อยที่สุด เช่น การลดข้อมูลจากลักษณะทรงกลม 3 มิติ เหลือเพียงวงกลมบนระนาบ 2 มิติ เป็นต้น

3. การเรียนรู้แบบเสริมกำลัง (reinforcement learning) คือ การเรียนรู้ที่ให้คอมพิวเตอร์ได้ เรียนรู้เพื่อหาแนวทางปฏิสัมพันธ์กับสิ่งแวดล้อมแล้วได้ผลลัพธ์ที่ดีที่สุด สำหรับการเรียนรู้แบบนี้จะต้องมีการกำหนดนโยบายการให้รางวัลว่าในสถานการณ์ไหนจะให้รางวัลเท่าไหร่ ตัวอย่างเช่น alphaGo ที่ถูกฝึกให้เล่นเกมโกะเอาชนะผู้เล่นแชมป์โลกได้ เป็นต้น

การทดสอบประสิทธิภาพแบบจำลองด้วย Confusion metrix

เป็นหนึ่งในวิธีที่นิยมใช้ในการประเมินความถูกต้องของแบบจำลองเพื่อวัดประสิทธิภาพสำหรับการนำมาใช้จริง โดยกกรณีที่ยกมาแสดงเป็นการวัดประสิทธิภาพสำหรับการแบ่งกลุ่มเป็น 2 กลุ่ม (binary classification) โดยเริ่มต้นจะมีค่าที่ต้องสนใจทั้งหมด 4 ค่าได้แก่

1. True Positive (TP) คือการที่แบบจำลองสามารถทำนายเหตุการณ์ที่สนใจเกิดขึ้นได้ถูกต้อง ซึ่งในกรณีของงานนี้คือ ทำนายว่าผู้รับการทดสอบว่ามีโอกาสเสี่ยงเป็นโรคฝุ่นจับปอด และผู้รับการทดสอบเสี่ยงเป็นโรคฝุ่นจับปอดจริง เป็นต้น

2. False Positive (FP) คือการที่แบบจำลองทำนายว่าเกิดเหตุการณ์ที่สนใจแต่ความจริงแล้วเหตุการณ์ที่สนใจไม่เกิดขึ้น

3. True Negative (TN) คือการที่แบบจำลองทำนายว่าเหตุการณ์ที่สนใจไม่เกิดขึ้นได้อย่างถูกต้อง เช่นทำนายว่าผู้รับการทดสอบไม่เป็นโรคโรคฝุ่นจับปอด และผู้รับการทดสอบนั้นไม่เป็นโรคฝุ่นจับปอดจริง เป็นต้น

4. False Negative (FN) คือการที่แบบจำลองทำนายว่าเหตุการณ์ที่สนใจไม่เกิดขึ้น แต่ในความจริงแล้วเกิดเหตุการณ์ที่สนใจขึ้น

โดยจากค่าทั้ง 4 ค่าที่ได้จากการทดสอบจะสามารถนำไปคำนวณหาค่าความแม่นยำได้ทั้งหมด 3 ค่า ได้แก่

1. Accuracy ($\frac{TP+}{TP+FP+TN+F}$) เป็นค่าที่ใช้สำหรับการวัดแบบจำลองในภาพรวมสามารถแบ่งข้อมูลได้ถูกกลุ่ม อย่างไรก็ตามค่านี้อาจไม่มีประสิทธิภาพมากนักหากมีความไม่สมดุลระหว่างกลุ่มเกิดขึ้น เช่น Positive 90% และ Negative 10% หากแบบจำลองทำนายทุกๆข้อมูลเป็น Positive จะพบว่ามีค่า Accuracy = 90% แต่ประสิทธิภาพของแบบจำลองไม่ดีนัก จึงควรใช้ค่าอื่นๆดังต่อไปนี้

2. **Precision** ($\frac{TP}{TP+FP}$) **หรือ True Positive Rate** เป็นค่าที่ใช้วัดความแม่นยำที่สนใจเฉพาะกับการทำนายกลุ่มที่สนใจเท่านั้นว่าจากที่ทำนายว่าเกิดเหตุการณ์ที่สนใจทั้งหมดมีความแม่นยำเพียงใด เหมาะสมกับการใช้ทำนายเหตุการณ์ที่ต้องใช้มูลค่าสูงในการดำเนินการ เช่น การตรวจร่างกายแบบ invasive เพราะเป็นกระบวนการที่ไม่ใช่การตรวจแบบทั่วไปและมีการรบกวนร่างกายของคนไข้ ดังนั้นจึงต้องมีความแม่นยำสูง

3. **Recall** ($\frac{TP}{TP+FN}$) เป็นค่าที่ใช้สำหรับการวัดประสิทธิภาพโดยคำนวณว่าจากกลุ่มที่เกิดเหตุการณ์ที่สนใจทั้งหมด แบบจำลองสามารถทำนายได้ถูกต้องเพียงใด เหมาะกับการใช้ในการวัดประสิทธิภาพแบบจำลองที่มีจุดมุ่งหมายในการคัดกรอง เนื่องจากไม่สนใจว่าแบบจำลองถูกต้องมากเพียงใด แต่สนใจว่าแบบจำลองครอบคลุมกลุ่มที่สนใจแค่ไหน

4. **F1-Score** ($2 * \frac{precision*recall}{precision+rec}$) เป็นค่าที่สามารถใช้เป็นตัวแทนในการวัดประสิทธิภาพของแบบจำลองทั้งหมดได้ในค่าเดียว เนื่องจากค่านี้ใช่การเฉลี่ยแบบ harmonic จากทั้งคะแนน precision และ recall มาแล้ว และมีประสิทธิภาพมากกว่าการใช้ค่า accuracy สำหรับการใช้วัดข้อมูลที่มีความไม่สมดุลระหว่างกลุ่ม

Table 1 Classifying the severity of the disorder.

Group	FVC (% Estimate value)	FEV1 (% Estimate value)	FEV1/FVC (%)	FEF25-75% (% Estimated value)
Normal	>80	>80	>70	>65
Mild	66-80	66-80	60-70	50-65
Moderate	50-65	50-65	45-59	35-49
Severe	<50	<50	<45	<35

ภาพรวมการทำงานของระบบ

ภาพรวมการทำงานของระบบประกอบด้วยการนำเข้าข้อมูลเพื่อปรับแต่งและทำความสะอาด ก่อนที่จะนำไปพัฒนาแบบจำลอง ทั้งหมด 6 รูปแบบ คือ Logistic Regression, Decision tree, Random Forest, Gradient boosting, XGBoost และ Support Vector Machine ซึ่งเมื่อทำการประเมินและได้แบบจำลองที่ดีที่สุดแล้วจะนำ Python script ไปใช้เป็นแกนเบื้องหลังสำหรับการทำนายข้อมูลทดสอบที่ได้ผ่านหน้าเว็บติดต่อผู้ใช้งานซึ่งไม่ซับซ้อนและสามารถปรับแต่งได้ง่าย รวมถึงมีข้อมูลเกี่ยวกับโรคเพื่อให้ความรู้กับผู้ใช้งาน โครงสร้างการทำงานของระบบเริ่มจากการให้ผู้ใช้ใส่ข้อมูลในรูปแบบของไฟล์ csv ที่มีค่าคอลัมน์ตามที่กำหนดไว้ลงผ่านทางเว็บแอปพลิเคชัน จากนั้นระบบจะทำนายโดยใช้โมเดล Machine Learning และแสดงผลการทำนายผ่านทางหน้าจอ โดยเว็บแอปพลิเคชันจะถูกสร้างด้วย Flask framework ที่ใช้ภาษา Python เป็นหลัก ระบบจะแบ่งเป็น 2 ส่วน คือ ส่วน

Frontend (user interface) ทำหน้าที่รับข้อมูลจากผู้ใช้และแสดงผลลัพธ์รวมถึงข้อมูลต่างๆ ของเว็บแอปพลิเคชัน เขียนด้วยภาษา HTML, CSS, และ JavaScript และ ส่วน Backend ซึ่งเป็นส่วนเรียกใช้โมเดล Machine Learning เพื่อทำนายผลลัพธ์ของข้อมูลที่รับเข้ามา และส่งผลลัพธ์ที่ได้กลับไปยังส่วน Frontend เขียนด้วยภาษา Python ใช้ไลบรารี Pandas ในการเก็บและจัดการกับข้อมูล และใช้ไลบรารี Scikit-learn ในการสร้างโมเดล Machine Learning

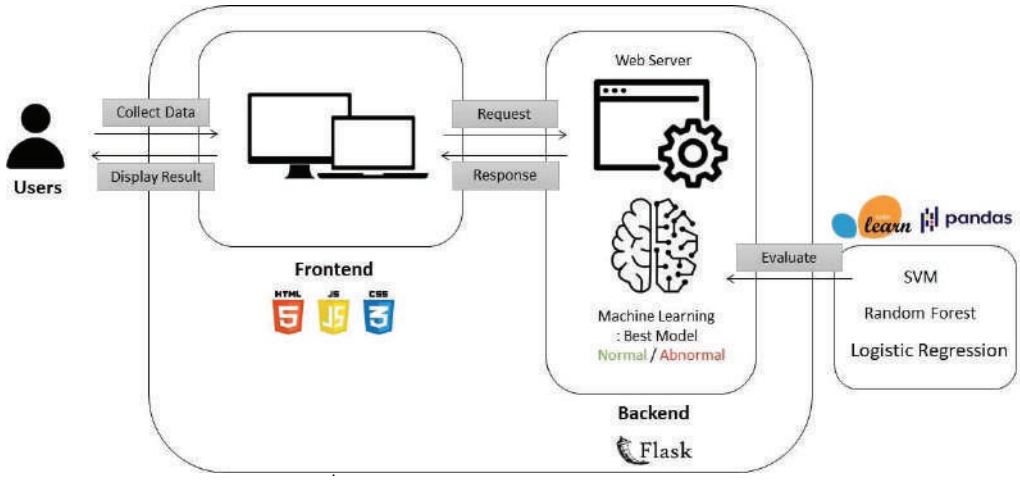

Figure 1 Overall working structure of the system.

การพัฒนาแบบจำลอง

ข้อมูลที่นำเข้า

ข้อมูลที่นำเข้าในการฝึกแบบจำลอง Machine learning เป็น ชุดข้อมูลที่ได้จากงานวิจัยเรื่อง Rubberwood dust and lung function among Thai furniture factory workers (Thetkathuek *et al.*, 2010) ซึ่งเป็นการศึกษาแบบภาคตัดขวาง โดยเก็บข้อมูลในช่วงเดือนเมษายน ถึงตุลาคม ปี พ.ศ. 2550 จากโรงงานเฟอร์นิเจอร์ไม้ยางพาราในภาคตะวันออก ที่ได้มาจากแบบสอบถามพนักงานที่ทำงานในโรงงานยางพารา จากผู้ตอบแบบสอบถามทั้งหมด 685 คน จากโรงงานยางพารา 8 แห่งในภาคตะวันออกของประเทศไทย อายุของผู้ตอบแบบสอบถามมีตั้งแต่อายุ 15-64 ปี โดยในชุดข้อมูลจะมีข้อมูลค่าฝุ่นของโรงงาน ข้อมูลทั่วไปของผู้ตอบรับแบบสอบถาม ประวัติการทำงาน ประวัติการเจ็บป่วย พฤติกรรมการป้องกันในขณะทำงาน และผลของการตรวจสมรรถภาพปอดด้วยวิธี Spirometry ชุดข้อมูลนี้มีคอลัมน์ทั้งหมด 110 คอลัมน์ ซึ่งเป็นคอลัมน์ที่สามารถนำมาใช้ในงานนี้ได้ทั้งหมด 86 คอลัมน์ ประกอบด้วยคอลัมน์ที่เป็นข้อมูลเชิงคุณภาพ (qualitative data) ทั้งหมด 64 คอลัมน์, ข้อมูลเชิงปริมาณ (quantitative data) จำนวน 13 คอลัมน์และคอลัมน์ที่เกี่ยวข้องกับการแปลผล (label data) ทั้งหมด 9 คอลัมน์

การเตรียมข้อมูล

1. การนำเข้าชุดข้อมูลที่เก็บมาจากแบบสอบถามและตรวจคุณภาพของข้อมูล

- ทำการอ่านไฟล์ Excel ด้วยไลบรารี่ pandas
- ศึกษาชนิดของข้อมูลในแต่ละคอลัมน์
- ศึกษาขอบเขตของคอลัมน์ที่ให้ค่าเป็นตัวเลข
- ศึกษาความสมดุลของคลาสจากการดูจำนวนหรือร้อยละของข้อมูลที่อยู่ในแต่ละคลาส

2. การจัดการข้อมูลที่ยังไม่สมบูรณ์

- ทำการตัดคอลัมน์ที่มีข้อมูลซ้ำกัน
- ทำการตัดแถวข้อมูลบางแถวที่มีคุณภาพต่ำ เช่น มีข้อมูลไม่สอดคล้องกับคอลัมน์
- ทำการแทนค่าลงในช่องว่าง เนื่องจากบางครั้งข้อมูลจากแบบสอบถามจะมีการเว้นหรือข้ามคำถามตามคำตอบในข้อก่อนหน้า
- แปลงชนิดของข้อมูลในบางคอลัมน์ที่เป็นตัวเลขให้เป็นตัวอักษร ให้ตรงกับความหมายของข้อมูลในเชิงลำดับไม่ใช่เชิงค่าตัวเลข

3. การสำรวจข้อมูลเบื้องต้น

- สำรวจความสัมพันธ์ระหว่างตัวแปรทำนาย (feature) กับผลลัพธ์ของการตรวจ Spirometry เช่น การใช้ crosstab เพื่อดูการกระจายตัวของผลตรวจเทียบกับตัวแปรทำนาย หรือการใช้เครื่องมือทางสถิติ เช่น การทดสอบ Chi-Square เพื่อทดสอบความสัมพันธ์ระหว่างตัวแปร
- คัดเลือกตัวแปรที่มีความเกี่ยวข้องกับผลลัพธ์ไปใช้ในการฝึกการเรียนรู้ของเครื่องต่อไป

• นำข้อมูลส่วนที่ใช้ในการฝึกไปใช้ฝึกด้วยแบบจำลอง Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, Gradient Boosting tree และ XGBoost

4. การสร้าง ปรับแต่ง และประเมินผลแบบจำลอง

• ทำการแบ่งข้อมูลที่คัดเลือกและผ่านการทดสอบมาแล้ว เป็นข้อมูลสำหรับฝึกและทดสอบแบบจำลองด้วยอัตราส่วน 70:30

• จับเวลาในช่วงที่ทำการฝึกแบบจำลอง เพื่อนำไปใช้วิเคราะห์ผลลัพธ์ในภายหลัง

• ทำการทดสอบประสิทธิภาพของแบบจำลองด้วยชุดข้อมูลสำหรับการทดสอบ และวัดผลประสิทธิภาพด้วย confusion matrix ร่วมกับการทำ k-fold cross validation เพื่อเลือกพารามิเตอร์ที่ดีที่สุด

• ทำการเปรียบเทียบประสิทธิภาพของแบบจำลองเพื่อตัดสินใจเลือกแบบจำลองที่ดีที่สุด

การพัฒนาเว็บแอปพลิเคชัน

ข้อกำหนดในการออกแบบ

เว็บแอปพลิเคชันจะถูกแบ่งออกเป็น 3 หน้า โดยจะมีรายละเอียดส่วนประกอบของแต่ละหน้าดังนี้

1. หน้าเว็บแอปพลิเคชันสำหรับอธิบายข้อมูลต่างๆ ของเว็บแอปพลิเคชัน

• ข้อมูลทั่วไปของเว็บแอปพลิเคชัน

• รายละเอียดและตัวอย่างของข้อมูลนำเข้าของเว็บแอปพลิเคชัน

• ผลลัพธ์และวิธีการแปลผลของเว็บแอปพลิเคชัน

2. หน้าเว็บแอปพลิเคชันสำหรับการวิเคราะห์ข้อมูล

• มีช่องสำหรับใส่ข้อมูลลงในเว็บแอปพลิเคชัน (มีการแสดงตัวอย่างของข้อมูลที่ใส่ไป)

• ผลของการวิเคราะห์ข้อมูลจากการเรียนรู้ของเครื่อง

3. หน้าเว็บแอปพลิเคชันสำหรับอธิบายข้อมูลของผู้จัดทำ

• ข้อมูลติดต่อของผู้จัดทำ และมหาวิทยาลัย

• อธิบายข้อมูลทั่วไปต่างๆ ในการจัดทำเว็บแอปพลิเคชัน

ผลการวิจัย

ผลการดำเนินงานของส่วนการเรียนรู้ของเครื่องแบบประเมินความพึงพอใจในการใช้เว็บแอปพลิเคชัน

การประเมินความพึงพอใจในการใช้เว็บแอปพลิเคชันได้จากการสัมภาษณ์ผู้ใช้เว็บแอปพลิเคชัน 3 ด้าน ได้แก่ ด้านการออกแบบและการจัดรูปแบบ ด้านเนื้อหา และ ด้านการใช้งาน หัวข้อการประเมินทั้งหมด ประกอบด้วย

1. ด้านการออกแบบและการจัดรูปแบบ จะประเมินความสวยงามและความทันสมัยของเว็บแอปพลิเคชัน ความน่าสนใจของหน้าเว็บแอปพลิเคชันพลิเคชัน และความง่ายในการอ่านข้อความในสีพื้นหลังและสีตัวอักษรที่เลือกใช้

2. ด้านเนื้อหา จะประเมินความถูกต้องและชัดเจนของเนื้อหา ปริมาณของเนื้อหาว่ามีความเพียงพอต่อการทำความเข้าใจหรือไม่ การจัดลำดับเนื้อหาเป็นขั้นตอน และความต่อเนื่องของเนื้อหา

3. ด้านการใช้งาน จะประเมินความง่ายของการใช้งานจริง ความสามารถในการนำไปใช้งานจริง ผลลัพธ์ที่ได้จากการทำนาย ประโยชน์ในการใช้งานเว็บแอปพลิเคชัน และภาพรวมของเว็บแอปพลิเคชันต่อความต้องการของผู้ใช้

1. ผู้ใช้สามารถทราบผลการทำนายโดยอาศัยข้อมูลนำเข้าคู่กับโครงสร้างในการวิเคราะห์จากผู้พัฒนา (prediction model)

2. ผู้ใช้และผู้พัฒนาสามารถดูรายละเอียดที่เกี่ยวกับเว็บแอปพลิเคชัน (about us)

ในภาพรวมของการดำเนินงานของส่วนการเรียนรู้ของเครื่อง แบ่งออกเป็น 5 ส่วน ได้แก่

การสำรวจข้อมูล

จากการสำรวจข้อมูลเบื้องต้นพบว่าข้อมูลสำหรับการดำเนินการมีลักษณะเป็นไฟล์ประเภท csv ขนาด 685 แถว 110 คอลัมน์ ซึ่งประกอบไปด้วยคอลัมน์ที่ไม่ได้ใช้ประกอบการฝึกแบบจำลอง เนื่องจากเป็นข้อมูลทับซ้อนหรือข้อมูลอ่อนไหว เช่น รายละเอียดชื่อผู้เข้ารับการทดสอบสไปโรเมตรีย์ ชื่อโรงงาน จังหวัดที่ตั้งของโรงงาน จำนวน 27 คอลัมน์ คอลัมน์ที่ใช้สำหรับการฝึกแบบจำลองทั้งหมด 77 คอลัมน์ แบ่งเป็นข้อมูลเชิงคุณภาพ (categorical data) ซึ่งเป็ยข้อมูลที่สามารถแบ่งเป็นกลุ่มได้ เช่นข้อมูลแผนกที่ทำงานอยู่ ข้อมูลสถานะการแต่งงาน เป็นต้นทั้งหมด 63 คอลัมน์ และข้อมูลเชิงปริมาณ (numerical data) ซึ่งเป็นข้อมูลที่มีลักษณะเป็นตัวเลข เช่น น้ำหนัก ส่วนสูง จำนวนชั่วโมงในการทำงานทั้งหมด 14 คอลัมน์ และคอลัมน์ผลการวินิจฉัยที่ได้จากการตรวจสไปโรเมตรีย์จำนวน 6 คอลัมน์ โดยภาพรวมการสำรวจข้อมูลเบื้องต้นเป็นดัง Figure 6

ข้อมูลของคอลัมน์ที่ใช้สำหรับการฝึกแบบจำลองสามารถแบ่งออกเป็นกลุ่มย่อยได้ทั้งหมด 4 กลุ่ม ได้แก่

• กลุ่มข้อมูลส่วนบุคคล เช่น เพศ อายุ ความสูง น้ำหนัก

• กลุ่มข้อมูลที่เกี่ยวข้องกับโรงงาน เช่น ฝ่ายที่ทำงาน ปริมาณฝุ่นในพื้นที่ทำงาน ชั่วโมงการทำงานต่อวัน

• กลุ่มข้อมูลประเภทพฤติกรรมด้านสุขภาพ เช่น การสูบบุหรี่ การดื่มสุรา การใช้หน้ากากอนามัย

• ข้อมูลประวัติการเจ็บป่วย เช่น ประวัติการเป็นโรคทางเดินหายใจต่างๆ ประวัติการแน่นหน้าอก ประวัติการไอ

ส่วนข้อมูลเป้าหมายเป็นข้อมูลกลุ่มผลตรวจสไปโรเมตรีย์ ได้แก่ ค่าร้อยละของค่า FVC ร้อยละของค่า FEV ร้อยละของค่า FEV ต่อ FVC ค่า FEF 25-75 ผลวินิจฉัย obstructive defect และผลวินิจฉัย restrictive defect ซึ่งเป็นข้อมูลที่จะถูกใช้สำหรับการฝึกแบบจำลองในงานนี้ ผลการสำรวจจำนวนผลวินิจฉัย restrictive defect แต่ละกลุ่มได้ผลลัพธ์ดัง Figure 2 และ 3

Figure 2 Overview of survey data and types of data

Figure 3 Survey of data groups in Restrictive defect test

จากการสำรวจพบว่าผลการวินิจฉัยแบ่งเป็น 4 ประเภทคือสภาพปอดปกติ (normal) มีจำนวนข้อมูลทั้งหมด 433 แถว, สภาพปอดเริ่มมีความผิดปกติ (mild) มีจำนวนข้อมูลทั้งหมด 210 แถว, สภาพปอดมีความผิดปกติ (moderate) มีจำนวนข้อมูล 36 แถว และสภาพปอดผิดปกติร้ายแรง (severe) มีจำนวนข้อมูล 6 แถว

การทำความสะอาดข้อมูล

แก้ไขค่าวินิจฉัย restrictive defect เนื่องจากหลังจากสำรวจข้อมูลแล้วพบว่ามีการบันทึกผลวินิจฉัยของคอลัมน์ restrictive defect ผิดพลาด ทำให้ต้องบันทึกเพื่อแก้ไขใหม่

แก้ไขการบันทึกค่าในตัวแปรอื่นๆ ที่มีปัญหาเนื่องจากพบว่าข้อมูลเชิงคุณภาพบางคอลัมน์มีข้อมูลเกินที่กำหนด หลังจากแปลงชนิดของข้อมูลให้เป็นตัวเลข

normalize data เนื่องจากข้อมูลเชิงปริมาณในชุดข้อมูลนี้มีระยะที่ห่างกันค่อนข้างมาก อาจส่งผลต่อประสิทธิภาพของแบบจำลองได้ จึงทำการ normalize ด้วยวิธี Min-Max scaling คือ การกำหนดให้ค่าทั้งหมดอยู่ระหว่างช่วง 0 ถึง 1

ทำ one-hot encoding เนื่องจากข้อมูลเชิงคุณภาพในงานนี้มีประเภทของข้อมูลมากกว่า 2 ประเภทจึงแยกแต่ละคอลัมน์ออกไปเป็นคอลัมน์ละ 1 ค่าเช่นจากคอลัมน์ Domicil ที่บรรจุค่าถิ่นกำเนิดไว้มีการกำหนดค่าเป็น 1 ถึง 6 จะถูกแบ่งออกเป็น 6 คอลัมน์ได้แก่ คอลัมน์ Domicil-1 ที่มีค่า เพียง 0, 1 เท่านั้น เป็นจำนวนทั้งหมด 6 ตาราง เป็นต้น เรียกวิธีการนี้ว่าการทำ One-hot encoding

การทดลองเพื่อหากลุ่มที่เหมาะสมที่สุดสำหรับการฝึกการเรียนรู้ของเครื่อง

การทดลองนี้ทำการฝึกฝนแบบจำลอง 6 แบบได้แก่ Logistic Regression, Decision tree, Random Forest, Gradient boosting, XGBoost และ SVM โดยใช้ไฮเปอร์พารามิเตอร์ตั้งต้น และใช้ค่า f1-score ในการประเมินผล โดยเป็นงานประเภท Multi-class classification task โดยมีข้อมูล label เป็นข้อมูลผลการตรวจประสิทธิภาพปอดเป็น 3 กลุ่มได้แก่ Normal, Mild และกลุ่มสุดท้ายคือข้อมูลที่รวมกลุ่ม Moderate และ Severe ไว้ด้วยกัน เรียกว่า Moderate+Severe โดยมีกลุ่ม Normal 432 แถว กลุ่ม Mild 208 แถว กลุ่ม Moderate+Severe 45 แถว ประสิทธิภาพของทั้ง 6 แบบจำลองแสดงดัง Table 2

Table 2 Results of the 3 groups training experiment using lung performance data.

Model	f1-score Normal group	f1-score Mild group	f1-score Moderate+Severe group	Training time
Logistic Regression	0.68	0.17	0.00	78 ms
Decision Tree	0.63	0.28	0.07	34 ms
Random Forest	0.75	0.20	0.00	299 ms
Gradient Boosting	0.72	0.21	0.22	920 ms
XGBoost	0.71	0.27	0.13	1270 ms
SVM	0.70	0.19	0.07	51 ms

Table 3 Amount of data in 3 groups comparism before and after oversampling of lung performance test data.

Lung Function Test	No. of Raws before oversampling	No. of Raws after oversampling
Normal	432	432
Mild	208	432
Moderate and Severe	45	432

Table 4 Results of the 3 groups training experiment using lung performance data.

Model	f1-score Normal group	f1-score Mild group	f1-score Moderate and Severe group	Training time
Logistic Regression	0.69	0.61	0.84	124 ms
Decision Tree	0.59	0.65	0.84	24 ms
Random Forest	0.76	0.78	0.95	310 ms
Gradient Boosting	0.72	0.69	0.92	1500 ms
XGBoost	0.72	0.73	0.94	729 ms
SVM	0.72	0.63	0.85	105 ms

จาก Table 2 พบว่าสามารถประเมินค่า f1-score ได้ทั้ง 3 กลุ่ม อย่างไรก็ตามประสิทธิภาพที่ได้ไม่น่าพอใจ แต่เนื่องด้วยข้อจำกัดของความต่างกันระหว่างกลุ่มไม่สามารถแก้ไขได้ด้วยเพียงการรวมกลุ่มของข้อมูล ดังนั้นจึงมีการใช้เครื่องมือที่ใช้จัดการข้อมูลที่มีขนาดไม่สมดุลดังการทดลองต่อไป

การจัดการข้อมูลที่มีขนาดไม่สมดุลด้วย SMOTE (synthetic minority oversampling technique)

เนื่องจากจำนวนข้อมูลในกลุ่ม Moderate and Severe มีปริมาณค่อนข้างน้อย งานวิจัยนี้จึงใช้เทคนิคการ oversampling เพื่อเพิ่มจำนวนข้อมูลส่วนน้อยให้มีจำนวนใกล้เคียงกับข้อมูลส่วนมากด้วยวิธี SMOTE (synthetic

minority oversampling technique) จำนวนข้อมูลก่อนและหลังประยุกต์ใช้วิธี SMOTE แสดงดัง Table 3

การทำวิธีสังเคราะห์ข้อมูลเพิ่ม จะเพิ่มจำนวนข้อมูลของกลุ่มที่มีข้อมูลน้อยทั้งหมดให้มีจำนวนเท่ากับจำนวนข้อมูลของกลุ่มที่ใหญ่ที่สุด ในส่วนข้อมูลที่ตรวจ ประสิทธิภาพปอดแบบ 3 กลุ่มดำเนินการฝึกแบบจำลองต่างๆ ด้วยค่า Hyperparameter ตั้งต้นของ Library scikit-learn และได้ผลลัพธ์ดัง Table 4

จาก Table 4 พบว่าหลังจากนำข้อมูลที่ผ่านการทำ Oversampling มาใช้ในการฝึกพบว่า เมื่อมีจำนวนกลุ่มที่น้อยลงทำให้ค่า f1-score ในกลุ่ม Normal และ Mild มีแนวโน้มที่จะมีค่าดีขึ้นเล็กน้อย ส่วนกลุ่ม Moderate and Severe ยังคงมีโอกาสเกิดปัญหา overfitting เนื่องจากข้อสมมติฐาน 2 ประการได้แก่การที่ข้อมูลมีค่า f1-score เพิ่มขึ้นอย่างมาก เมื่อเปรียบเทียบกับผลแบบจำลองก่อนการทำ oversampling เช่นในแบบจำลอง Logistic Regression (0.00 และ 0.84) เป็นต้น กอปรกับจำนวนข้อมูลตั้งต้นที่ในกลุ่มเยอะสุดมีจำนวน 432 ข้อมูล เมื่อเทียบกับกลุ่ม Moderate and Severe ที่มีจำนวนน้อยที่สุดมีเพียง 45 ข้อมูล (Table 3) ซึ่งด้วยหลักการของ Oversampling จะทำการเพิ่มข้อมูลจากลักษณะของจำนวนข้อมูลเดิมให้มีจำนวนเท่ากับข้อมูลกลุ่มใหญ่ ดังนั้นจึงมีสมมติฐานว่าจำนวนชุดข้อมูลที่น้อยอาจทำให้มีรูปแบบที่น้อยเกินไปในการทำ Oversampling จึงเกิดการ Overfitting ในแบบจำลองขึ้นได้

การปรับค่าไฮเปอร์พารามิเตอร์ของแบบจำลองร่วมกับ k-fold cross validation

การทำ 5-fold cross validation ร่วมกับการปรับไฮเปอร์พารามิเตอร์ถูกนำมาใช้เพื่อหาโครงสร้างแบบจำลองที่ให้ประสิทธิภาพของการเรียนรู้ของเครื่องได้ดีที่สุด โดยเริ่มจากการใช้วิธี Randomized search เพื่อการหาช่วงที่แบบจำลองจะสามารถให้ประสิทธิภาพได้ดี ตามด้วยการใช้ Grid search ในการหาค่าไฮเปอร์พารามิเตอร์ที่แบบจำลองให้ประสิทธิภาพดีในช่วงที่มีขนาดเล็กลง และประเมินผลด้วยการใช้ค่า f1-score โดยได้ผลลัพธ์ดัง Table 5

Table 5 Results after hyperparameter adjustment.

Model	Optimal hyperparameter value	Everage f1-score
Logistic Regression	Solvers = lbfgs	0.70023
Decision Tree	Max depth = 32, Max features = 20, Min sample leaf = 15	0.674431
Random Forest	Max features = 15, Min samples leaf = 2, Min samples split = 10, Max depth = 64	0.745796
Gradient Boosting	n_Estimators = 500, Max depth = 9, Learning rate = 1	0.731759
XGBoost	Max depth = 7, Gamma = 2, Min child weight = 4	0.714132
SVM	Kernel = rbf, Decision function shape = ovo, Gamma = scale	0.739056

จากผลลัพธ์ใน Table 5 พบว่าเมื่อทำการ Cross validate แล้วแบบจำลอง Random Forest ที่ผ่านการปรับปรุงไฮเปอร์พารามิเตอร์ให้ค่าเฉลี่ย f1-score มากที่สุดที่ 0.745796 โดยแบบจำลองใช้ตัวแปรทำนายที่มี ซึ่งเราสามารถปรับปรุงแบบจำลองโดยลดจำนวนตัวแปรลงและทำการปรับปรุงไฮเปอร์พารามิเตอร์อื่นๆเพื่อให้ได้แบบจำลองที่มีประสิทธิภาพที่สุดและง่ายต่อการใช้งานของผู้ใช้งานต่อไป

การเลือกตัวแปรที่สำคัญในการฝึกแบบจำลอง

การเลือกตัวแปรที่สำคัญสำหรับการฝึกแบบจำลองใช้วิธี RFE (recursive feature elimination) ซึ่งเป็นเครื่องมือสำหรับการคัดเลือกโดยกลวิธีการเลือกขึ้นอยู่กับประเภทแบบจำลองที่ใช้ (estimator) และตัวแปรที่ทำให้แบบจำลองนั้นผิดพลาดน้อยที่สุดในการทำนาย โดยทำให้ mean absolute error (MAE) ต่ำสุด จะถูกใช้เป็นกลุ่มตัวแปรที่สำคัญในการทำนายระดับความรุนแรงต่อไป โดยแต่ละแบบจำลองจะมีขั้นตอนการเลือกต่างกัน เช่น กลุ่มแบบจำลองเชิงเส้น (linear models) จะใช้ค่าสัมประสิทธิ์ของแต่ละตัวแปร (coefficient) ในขณะที่กลุ่มแบบจำลองที่ใช้ต้นไม้เป็นฐาน (tree-based models) จะใช้ค่าความสำคัญของตัวแปร (feature importance) ในการคัดเลือก หลังจากนั้นจะทำการวนทำซ้ำเพื่อหาจำนวนคู่ตัวแปรที่เหมาะสมที่สุด โดยในงานนี้กำหนดจำนวนตัวแปรไว้ที่ 20 ตัวแปร และในการทดลองนี้ทำการทดลองเฉพาะข้อมูล 2 กลุ่มที่ผ่านการทำ oversampling ด้วยวิธี SMOTE มาเท่านั้น ได้ผลดัง Table 6

Table 6 Results of 3 groups training with selecting variables in data.

Model	Everage f1-score
Logistic Regression	0.59
Decision Tree	0.62
Random Forest	0.74
Gradient Boosting	0.72
XGBoost	0.68
SVM	0.5

จาก Table 6 พบว่า แบบจำลอง Random Forest มีประสิทธิภาพที่ดีที่สุดซึ่งสามารถให้ค่าเฉลี่ย f1-score ระหว่างกลุ่ม Normal, Mild และ Moderate อยู่ที่ 0.74 และ 20 ตัวแปรที่ดีที่สุดประกอบด้วยตัวแปรเช่น แผนกที่ทำงาน ประวัติการสูบบุหรี่ และอายุระหว่าง 21 ถึง 25 ปี ซึ่งจากผลการทดลองในส่วนการเรียนรู้ของเครื่องทั้งหมดสรุปได้ว่า แบบจำลองที่เหมาะสมสำหรับการนำไปใช้ในเว็บแอปพลิเคชันคือ แบบจำลอง Random Forest แบบ 3 กลุ่มข้อมูล (normal, mlid และ moderate+severe) ที่ผ่านการทำ oversampling ด้วยวิธี SMOTE และคัดเลือกตัวแปรที่สำคัญด้วย RFE จำนวน 20 ตัวแปร

ผลการดำเนินงานของเว็บแอปพลิเคชัน

แบบจำลอง Random forest ในหัวข้อก่อนหน้าถูกนำไปใช้ร่วมกับเว็บแอปพลิเคชันที่ได้พัฒนาขึ้นตามที่ได้ออกแบบไว้ ซึ่งสามารถเข้าถึงได้จากลิงก์ https://whispering-wildwood-41614-c10f966ed49e.herokuapp.com/ โดยในส่วนของ Frontend หรือ UI ถูกเขียนด้วยภาษา HTML, CSS, และ JavaScript และทำงานร่วมกันโปรแกรมส่วน Backend ผ่าน Flask framework หน้าเว็บแอปพลิเคชันที่ได้ทำการออกแบบไว้ มีทั้งหมด 3 หน้า คือ 1. หน้าหลัก 2. การทำนาย 3. เกี่ยวกับเรา ตัวอย่างหน้าการทำนายของเว็บแอปพลิเคชันแสดงดัง Figure 4

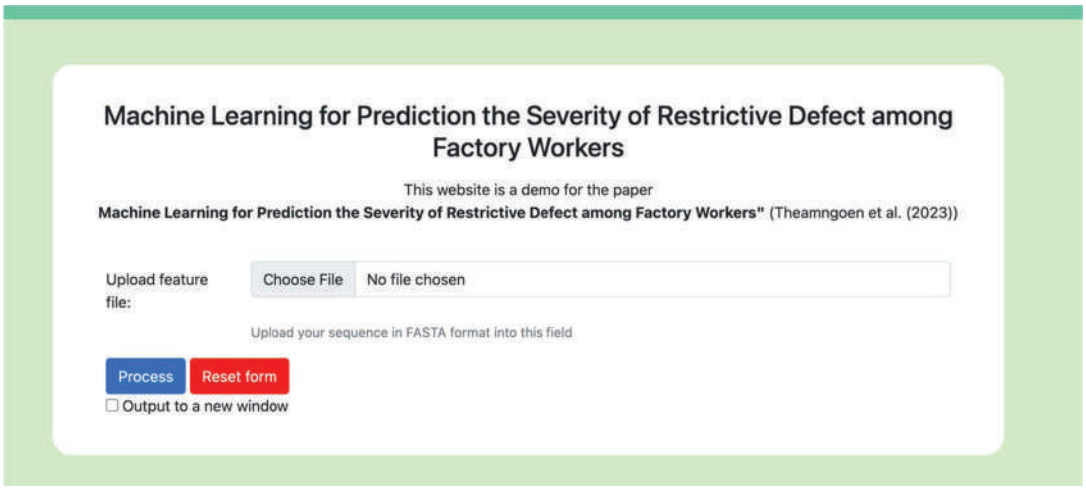

Figure 4 Prediction page of web appication

ผลการประเมินความพึงพอใจในการใช้เว็บแอปพลิเคชัน

จากการสัมภาษณ์เชิงลึกผู้ใช้งานเว็บแอปพลิเคชันจำนวน 10 ท่านซึ่งเป็นหัวหน้าแผนกโรงงานเกี่ยวกับการใช้งานแอปพลิเคชันใน 3 ด้าน สรุปผลการประเมินความพึงพอใจในแต่ละด้านได้ดังนี้

ในด้านความสวยงาม ผู้ใช้มีความพึงพอใจเนื่องจากหน้าเว็บแอปพลิเคชันมีรูปแบบที่ค่อนข้างเรียบง่าย ผู้ใช้ได้แนะนำให้เพิ่มรูปภาพหรือมีการตกแต่งหน้าเว็บเพื่อเพิ่มความสวยงามและความน่าใช้งาน

ในด้านเนื้อหา ผู้ใช้มีความพึงพอใจเนื่องจาก เนื้อหามีความถูกต้อง อ่านแล้วสามารถเข้าใจได้ง่าย มีประโยชน์ในการทำความเข้าใจและการใช้งาน แต่เนื้อหาบางหัวข้อมีปริมาณทำให้ผู้ใช้ไม่สามารถทำความเข้าใจได้ทั้งหมด จึงได้แนะนำให้มีการลดเนื้อหาบางหัวข้อให้น้อยลงหรือแบ่งเนื้อหาเป็นข้อ จะทำให้ผู้ใช้สามารถเข้าใจเนื้อหาได้ง่ายมากยิ่งขึ้น

ในด้านการใช้งาน ผู้ใช้พึงพอใจเว็บแอปพลิชันที่สามารถใช้งานได้สะดวก เข้าใจวิธีการใช้งานได้ง่าย วิธีการใช้งานไม่ยุ่งยาก โดยรวมเว็บแอปพลิเคชันมีประโยชน์ในการใช้งาน

สรุปและอภิปรายผล

ผลจากการสร้างแบบจำลองเพื่อวิเคราะห์ประสิทธิภาพปอดสำหรับการคัดกรองผู้มีโอกาสเกิดกลุ่มโรคที่มีการจำกัดการขยายตัวของปอดใช้เทคนิคการเรียนรู้ของเครื่องประเภทต่างๆ เพื่อหาแบบจำลองที่เหมาะสมที่สุดจากชุดข้อมูลจากงานวิจัยเรื่อง Rubberwood dust and lung function among Thai furniture factory workers พบว่าแบบจำลอง Random Forest ร่วมกับข้อมูลการบันทึกผลตรวจ Restrictive defect 3 กลุ่ม (normal, mild และ moderate + severe) จากเครื่อง Spirometry มีประสิทธิภาพดีที่สุดโดยมีค่า accuracy โดยรวม 0.75 และ แบบจำลองที่ได้ถูกนำไปใช้ในส่วน back-end ของเว็บแอปพลิเคชัน โดยใช้รูปแบบของข้อมูลนำเข้า เป็นการรับไฟล์ csv เพื่อทำนายระดับความรุนแรงของความผิดปกติของความยืดหยุ่นของปอดของพนักงานจำนวนมากพร้อมกันในครั้งเดียว โดยผลที่ได้จากแบบสอบถามความพึงพอใจในการใช้งานของผู้ใช้ มีความพึงพอใจอยู่ในระดับดี สอดคล้องกับงานวิจัยของ Kaplan *et al*. (2021) ที่ทำการสำรวจประสิทธิภาพของปัญญาประดิษฐ์และการเรียนรู้ของเครื่องในกลุ่มโรคปอดโดยในโรคกลุ่ม Chronic obstructive pulmonary disease (COPD) และหอบหืดให้ผลลัพธ์อยู่ระหว่าง 0.75-1.00 และมีค่า f1-score ของกลุ่ม Normal, Mild และ Moderate + Severe เท่ากับ 0.67 0.69 และ 0.88 ตามลำดับ โดยตัวแปรที่ช่วยทำนายที่ดีที่สุด ได้แก่ แผนกที่ทำงาน ประวัติการสูบบุหรี่ และอายุระหว่าง 21 ถึง 25 ปี โดยตัวแปรดังกล่าวบางตัวได้แก่ อายุปีที่ทำงาน และแผนกที่ทำงาน มีความสอดคล้องกับงานวิจัยของ Hongbo Liu และคณะ (Liu *et al*., 2009)

งานวิจัยนี้เลือกใช้เพียงแบบจำลองพื้นฐานและแบบจำลองในกลุ่ม Ensemble เท่านั้น ในอนาคตเมื่อมีจำนวนข้อมูลมากขึ้นอาจมีการทดลองใช้แบบจำลองในกลุ่มการเรียนรู้เชิงลึก (deep learning) และกลุ่มเครือข่ายประสาทเทียม (artificial neural network) เพื่อค้นหาแบบจำลองที่มีประสิทธิภาพดีที่สุดต่อไป

ผลการทดลองในงานวิจัยนี้ถูกใช้ในการทดสอบประสิทธิภาพกับข้อมูลตรวจจากตรวจสไปโรเมตรีย์ในช่วงเวลาเดียวกัน จึงมีข้อเสนอแนะให้มีการทดสอบประสิทธิภาพของแบบจำลองเทียบกับผลตรวจจริงอีกครั้งเมื่อถึงช่วงเวลาในการตรวจในโรงงานในอนาคต เพื่อทำการเปรียบเทียบประสิทธิภาพของการใช้งานจริงต่อไป

ข้อเสนอแนะ

แบบจำลองที่ได้สามารถใช้เพื่อเป็นเครื่องมือคัดกรองเบื้องต้นเพื่อลดเวลาและค่าใช้จ่ายก่อนตัดสินใจตรวจสไปโรมิเตอร์ ในด้านแนวทางการพัฒนา จำเป็นต้องเพิ่มปริมาณข้อมูลที่ใช้ฝึกแบบจำลอง รวมถึงเพิ่มความสมดุลของกลุ่มข้อมูลให้มีจำนวนใกล้เคียงกัน เพื่อให้อัลกอริทึมสามารถเรียนรู้รูปแบบในข้อมูลที่หลากหลายและได้แบบจำลองที่มีประสิทธิภาพสูงขึ้น

เอกสารอ้างอิง

Chandran, U., Reps, J., Yang, R., Vachani, A., Maldonado, F., & Kalsekar, I. (2023). Machine learning and real-world data to predict lung cancer risk in routine care. cancer epidemiology, biomarkers & prevention. *A Publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology, 32*(3), 337-343. https://doi.org/10.1158/1055-9965.EPI-22-0873

Gould, M. K., Huang, B. Z., Tammemagi, M. C., Kinar, Y., & Shiff, R. (2021). Machine learning for early lung cancer identification using routine clinical and laboratory data. *American journal of respiratory and critical care medicine, 204*(4), 445-453. https://doi.org/10.1164/rccm.202007-2791OC

Guo, Y., Yin, S., Chen, S., & Ge, Y. (2022). Predictors of underutilization of lung cancer screening: a machine learning approach. *European journal of cancer prevention, 31*(6), 523-529. https://doi.org/10.1097/CEJ.0000000000000742

Hazra, A., Bera, N., & Mandal, A. (2017). Predicting lung cancer survivability using SVM and logistic regression algorithms. *International Journal of Computer Applications, 174*, 19-24. doi:10.5120/ijca2017915325

HITAP. (2017). *การศึกษาการเข้าถึงบริการตรวจสมรรถภาพปอดด้วยวิธี spirometry ที่มีประสิทธิภาพ ในโรงพยาบาลชุมชน*. https://www.hitap.net/documents/173095

Kaplan, A., Cao, H., FitzGerald, J. M., Iannotti, N., Yang, E., Kocks, J. W. H., Mastoridis, P. (2021). Artificial intelligence/machine learning in respiratory medicine and potential role in asthma and COPD diagnosis. *The Journal of Allergy and Clinical Immunology: In Practice, 9*(6), 2255-2261. doi:10.1016/j.jaip.2021.02.014

Kumar, N., Narayan Das, N., Gupta, D., Gupta, K., & Bindra, J. (2021). Efficient automated disease diagnosis using machine learning models. *Journal of Healthcare Engineering, 2021*, 9983652. doi:10.1155/2021/9983652

Kunpeuk, W., Julchoo, S., Phaiyarom, M., Sosom, J., Sinam, P., Sukaew, T., Siriruttanapruk, S. (2021). A scoping review on occupational exposure of silica and asbestos among industrial workers in Thailand. *Outbreak, Surveillance, Investigation & Response (OSIR) Journal 14*(2), 41-51. http://osirjournal.net/index.php/ osir/article/view/231

Liu, H., Tang, Z., Yang, Y., Weng, D., Sun, G., Duan, Z., & Chen, J. (2009). Identification and classification of high risk groups for Coal Workers' Pneumoconiosis using an artificial neural network based on occupational histories: a retrospective cohort study. *BMC Public Health, 9*(1), 366. 10.1186/1471-2458-9-366

Luize, A. P., Menezes, A. M., Perez-Padilla, R., Muiño, A., López, M. V., Valdivia, G., Lisboa, C., Montes de Oca, M., Tálamo, C., Celli, B., Nascimento, O. A., Gazzotti, M. R., Jardim, J. R., & PLATINO Team (2014). Assessment of five different guideline indication criteria for spirometry, including modified GOLD criteria, in order to detect COPD: data from 5,315 subjects in the PLATINO study. NPJ primary care respiratory medicine, 24, 14075. https://doi.org/10.1038/npjpcrm.2014.75

Martinez-Pitre, P. J., Sabbula, B. R., & Cascella, M. (2022). Restrictive lung disease. In *StatPearls*. StatPearls Publishing.

Pascoe, S. J., Wu, W., Collison, K. A., Nelsen, L. M., Wurst, K. E., & Lee, L. A. (2018). Use of clinical characteristics to predict spirometric classification of obstructive lung disease. *Int J Chron Obstruct Pulmon Dis, 13*, 889-902. doi:10.2147/copd.S153426

Qi, X. M., Luo, Y., Song, M. Y., Liu, Y., Shu, T., Liu, Y., Pang, J. L., Wang, J., & Wang, C. (2021). Pneumoconiosis: current status and future prospects. Chinese medical journal, 134(8), 898-907. https://doi.org/10.1097/CM9.0000000000001461

Thetkathuek, A., Yingratanasuk, T., Demers, P. A., Thepaksorn, P., Saowakhontha, S., & Keifer, M. C. (2010). Rubberwood dust and lung function among Thai furniture factory workers. *Int J Occup Environ Health, 16*(1), 69-74. 10.1179/107735210800546281