

國立臺灣大學電機資訊學院生醫電子與資訊學研究所



碩士論文

Graduate Institute of Biomedical Electronics and Bioinformatics

College of Electrical Engineering and Computer Science

National Taiwan University

Master Thesis

機器學習應用於 MIMIC III 資料庫，使用呼吸訓練時呼吸電阻、心律變異與臨床資訊預測拔管成敗

Prediction of Post-Extubation Failure using Machine Learning
with Impedance Pneumography, Heart Rate Variability, and
Clinical Data during Spontaneous Breathing Trial: Analysis of the
MIMIC-III Database

張裕鑫

Yu-Hsin Chang

指導教授：賴飛熊博士

Advisor: Feipei Lai, Ph.D.

中華民國 112 年 5 月

May, 2023

誌謝



兩年半的時光，從原本一個完全不會寫程式語言的人，學會了 Python 和 Matlab 兩種程式語言，雖說不上精通，但至少可以自己獨立完成一整個計畫。美國著名勵志作家 Zig Ziglar 曾說過 "You don't have to be great to start, but you have to start to be great"，與其徬徨著思索這條路的可能性，不如想好就下來做看看吧，人生有幾個十年，我寧願做過了發現這不是我想要的，也不要為了當初沒有下去做而感嘆。

這段期間，工作、研究、課業、家庭與健康是我希望同時能夠兼顧的，老天爺很公平，每個人一天都只有 24 個小時可以使用，出了社會、結了婚、生了小孩，這個現象就更加的殘酷，你總是得要選擇並捨棄掉些什麼，才有機會能夠獲得更多的東西；非常感謝我的老婆，在我說出要去台北唸書的時候不僅沒有澆我一頭冷水，反倒是鼓勵我去進修，並且扛起家裡大部分的雜事與小孩的教育，讓我沒有後顧之憂地去做我想要做的事情，往往在趕進度的時候，我不是在用電腦，就是在前往用電腦的路上，上班用電腦，下班也用電腦，回家就是蹲在自己的小房間裡面不停地 coding、串資料、找 bug、準備報告...中間也經歷了好幾個月的低潮，每天早上起床就覺得自己好像做不出東西來，畢不了業，也多虧了老婆的支持，以及藉由規律且持續的跑步計畫，來重拾每天的節奏與自信，撐過那段時間。美國一位馬拉松運動員 Mitch Ammons 曾說 "The struggles I went through have made me the person that I am today." 每次的挑戰，都是讓自己更堅強的磨練。

感謝我親愛的爸媽，從小給我這麼好的環境，讓我有這樣樂天、願意嘗試的態度與堅持的精神；感謝我的同梯戰友劉任軒跟我一起撐過這個蠟燭多頭燒的碩士班；感謝我的工作上司，都很認同與體諒我的進修，讓我有充足的自我空間去做發展，尤其是我的精神導師施宏謀學長，在各個層面都給我很大的幫助；感謝研究所的指導老師賴飛熊教授，在我遇到瓶頸的時候給我指點迷津，當初會進行這個困難的主題時，曾經躊躇不前，也是老師的建議才讓我下定決心勇往前行。

碩士班就像是個橋樑、也更像是個轉運站，在這之後還連結著無數的地方等著去探索，認真的做出選擇並盡自己最大的努力走下去，不後悔自己走過的路，我想就夠了。下一站，要往哪去呢？


中文摘要



背景：侵入性機械通氣仍然是作為拯救重症患者生命的主要治療方法。然而，過久的侵入性機械通氣可能會增加呼吸器相關肺炎的風險，與增加醫院費用和死亡率。為評估是否適合脫離機械通氣，必須進行多因素評估，包括生理狀態、鎮靜劑使用、通氣機設置、插管的主要原因是否緩解，以及是否擁有通過自發性呼吸試驗的能力。不論是延長侵入性機械通氣的時間或因為拔管失敗都會產生相對應的風險，也因此，確定拔管的時機非常重要。到目前為止，還沒有關於拔管的統一評估準則。

目標：本研究的主要目的是調查：(1) 在自主呼吸訓練期間的電阻式呼吸訊號與心律變異度所提取的特徵，是否能夠用於訓練機器學習模型來預測拔管失敗；以及(2) 綜合結構化數據和上述信號特徵的模型是否能夠優於僅使用結構化數據的模型。

方法：本研究乃利用已公開的美國大型資料庫，分別被稱為多參數智能監測重症監護 (Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC-III)) 和 MIMIC-III 匹配的波形數據庫。其中包含各種資訊，包括結構化和波形數據 (電阻式呼吸訊號和心電圖)。納入已進行自主呼吸試驗並具有可用波形信號的患者。利用波形分析來提取電阻式呼吸訊號的特徵，和心律變異度分析來萃取心電圖的資訊。使用五種常見的機器學習模型進行拔管失敗的預測。此外，我們使用不同的數據類型組合來訓練模型，以評估它們對性能的影響。使用 DeLong test 的檢驗比較不同模型之間的 AUROC，並且用 Shapley value 來解釋特徵對於模型預測的影響程度。



結果：最終我們納入了 411 位滿足所有標準的拔管患者，有 46 位患者 (11.2%) 拔管失敗。我們發現，當使用所有的數據，包括電阻式呼吸訊號、心律變異度和結構化數據進行訓練時，XGBoost 分類器優於其他分類器，並且比使用單獨的數據類型（僅結構化數據、僅電阻式呼吸訊號、僅心律變異度或綜合電阻式呼吸訊號和心律變異度）進行訓練的 XGBoost 分類器具有顯著更高的成效。特徵重要性分析顯示，電阻式呼吸訊號和心律變異度的特徵佔了 XGBoost 分類器前二十個重要特徵的一半以上。

結論：這些發現表明，將從電阻式呼吸訊號和心律變異度獲取的特徵納入機器學習模型，可以提高預測機械通氣下重症患者拔管失敗的成效。然而，需要進一步研究以驗證這些結果在不同臨床背景下的有效性。

關鍵詞：人工智慧；機器學習；預測拔管失敗；加護病房；電阻式呼吸波形；心律變異度；MIMIC-III 資料庫

ABSTRACT



Background: A multifaceted evaluation is necessary for weaning patients off mechanical ventilation (MV), and identifying the optimal time for extubation is crucial. The objective of this study was to explore whether impedance pneumography (IP) and heart rate variability (HRV) features could be utilized to train or improve a machine learning (ML) model for predicting post-extubation failure (PF) in mechanically ventilated patients in the intensive care unit.

Methods: The MIMIC-III and MIMIC-III Waveform Database Matched Subset were the sources of data for this study, which consisted of clinical and waveform data, including IP signals and electrocardiogram (ECG) readings. Patients who underwent spontaneous breathing trials and had available waveform signals before extubation were eligible for inclusion. After a sequential assessment of waveform quality, the IP signal features were extracted, and HRV analysis was used to present the ECG features. Five common machine learning models were trained to predict PF. Moreover, the impact of different combinations of data types on the model's performance was evaluated. DeLong's test was utilized to compare the AUROC among individual models, and the SHapley Additive exPlanations method was applied to evaluate the models' expandability.

Results: The final cohort included 411 extubated patients who met all the criteria, of whom 46 patients (11.2%) failed extubation. Our findings indicated that XGBoost classifier performed better than other classifiers when trained with the entire dataset (including IP signal, HRV, and clinical data) and produced a significantly higher AUROC than when it was trained using only one type of data (clinical data alone, IP

signal alone, HRV alone, or both IP signal and HRV). Feature importance analysis revealed that IP waveform and HRV features constituted over half of the top 20 important features of the XGBoost classifier.

Conclusion: The results indicate that integrating IP signal and HRV features can enhance the performance of ML models in predicting PF among critically ill patients on MV. Nevertheless, additional research is required to verify these outcomes in diverse clinical scenarios.

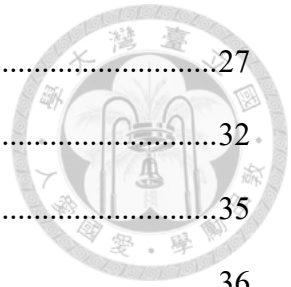
Keywords: Artificial intelligence, Machine learning, Extubation failure prediction, Intensive care unit, Impedance pneumography, Heart rate variability, MIMIC-III database.

CONTENTS



誌謝	i
中文摘要	ii
ABSTRACT	iv
CONTENTS	vi
LIST OF FIGURES	viii
LIST OF TABLES	ix
Chapter 1 Introduction.....	1
1.1 Background.....	1
1.2 Motivation.....	4
1.3 Objectives	4
Chapter 2 Method	5
2.1 Data Source.....	5
2.2 Selection of Participation and Definition	5
2.3 Preprocessing.....	6
2.3.1 Clinical data	6
2.3.2 IP signal.....	10
2.3.3 HRV signal	20
2.4 Feature Selection	22
2.5 Training Pipeline.....	23
2.6 Performance Evaluation.....	25
2.7 Experiment Environment.....	26
Chapter 3 Result.....	27

3.1	Participant Characteristics	27
3.2	Performance	32
3.3	Feature Importance	35
3.4	XGBoost Classifier with Different Input.....	36
Chapter 4	Discussion.....	40
4.1	Clinical Application.....	42
4.2	Limitation	43
Chapter 5	Conclusion	45
Reference	46	
Appendix	52	
	List of abbreviations	52
	Ethics approval and consent to participate.....	54
	Availability of data and materials	54
	Competing interests	54
	Funding	54



LIST OF FIGURES



Figure 1 Bedside patient monitor	3
Figure 2 3-Lead positions of electrodes.....	3
Figure 3 Flow chart for identification of relative peaks and troughs among impedance pneumography (IP) segment.....	12
Figure 4 Examples of different quality of impedance pneumography (IP) segments	15
Figure 5 Sequential assessment of IP waveform quality	17
Figure 6 Example of extracting waveform features	19
Figure 7 Training process of machine learning.	24
Figure 8 Flow chart of patient recruitment.	28
Figure 9 Prediction ability of different models for post-extubation failure.....	32
Figure 10 Feature importance (A) Top twenty most important features selected through mutual information. (B) the SHAP values of XGBoost classifier, and top 20 features are arranged based on the SHAP values across all samples.	36

LIST OF TABLES

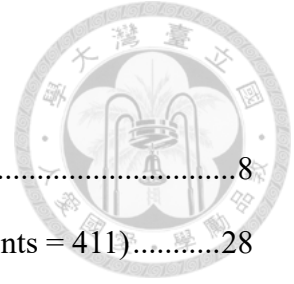


Table 1 Sampled criteria, missing rate, and exclusion of variables	8
Table 2 Demographic characteristics of enrolled patients (Total patients = 411).....	28
Table 3 Outcome analysis	31
Table 4 Performance of different machine learning models	33
Table 5 Hyperparameters for individual model	34
Table 6 Performance of XGBoost classifier with different input	38
Table 7 DeLong test to compare AUROC of different input in XGBoost.....	39

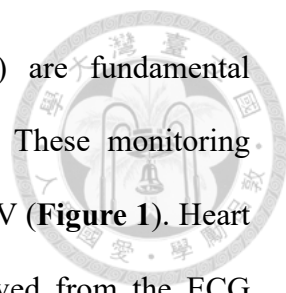
Chapter 1 Introduction



1.1 Background

In critically ill patients, invasive mechanical ventilation (MV) continues to be the primary life-saving intervention. Nonetheless, extended utilization of MV can lead to an elevated likelihood of ventilator-associated pneumonia, which in turn escalates hospital expenses and mortality rates. [1, 2]. To determine the readiness for discontinuing invasive MV, it is crucial to conduct a comprehensive assessment that considers various factors. These factors include the patient's physiological condition, sedative usage, ventilator settings, resolution of the underlying cause for MV, and the ability to successfully complete a spontaneous breathing trial (SBT). Typically, SBTs involve either intermittent T piece or low-level pressure support ventilation (PSV) for a minimum duration of 30 minutes [3]. Additionally, several parameters have been identified as indicators of a higher risk of post-extubation failure (PF). These parameters include lower inspiratory pressure, higher rapid shallow breathing index (RSBI) [4], weakened cough strength [5-7], and extended duration of MV [7-9].

The identification of the optimal timing for extubation is crucial in mitigating the risks associated with both prolonged MV and PF. Currently, there is no universal standard protocol for the weaning process. However, a meta-analysis has demonstrated that adherence to standardized weaning protocols can lead to a reduction in the duration of MV, the weaning period, and the length of stay in the intensive care unit (ICU) [10]. Despite the implementation of standardized weaning protocols, it is important to acknowledge that PF can still occur in approximately 10-23% of patients. PF is closely associated with prolonged MV, extended ICU stay, and elevated risks of morbidity and mortality [4, 6, 7, 11].



Electrocardiogram (ECG) and impedance pneumography (IP) are fundamental components of bedside monitoring for patients undergoing MV. These monitoring techniques are commonly employed for nearly every patient with MV (**Figure 1**). Heart rate variability (HRV) obtained from the inter-beat intervals derived from the ECG provides a non-invasive assessment of the autonomic function of the cardiovascular system. HRV analysis includes time-domain, frequency-domain, and non-linear domain analyses. This widely utilized approach is commonly employed in clinical research studies [12-14]. The IP signal is obtained by measuring the thoracic electrical impedance through ECG electrodes, which are placed in standardized positions (as depicted in **Figure 2**). The main mechanism underlying the elevated electrical resistance observed in IP is attributed to two factors. Firstly, an increase in gas volume within the chest reduces conductivity, leading to higher resistance. Secondly, during inspiration, the expansion of lung cavities increases the length of conductance, further contributing to the overall impedance changes [15]. IP offers several advantages, including its sustainability, non-invasiveness, and ease of accessibility. It has also demonstrated a high level of agreement with tidal volume measurements derived from a pneumotachograph [16, 17]. Several studies have examined the association between HRV, respiratory rate variability (RRV), tidal volume, and their relationship with extubation outcomes [18-20]. Seely et al. indicated that the predictive model utilizing RRV during the last SBT demonstrated optimal prediction accuracy for all patients, which further improved when combined with clinical impression or RSBI [21]. Additionally, some studies have found a significant correlation between respiratory signals and PF, primarily by extracting and analyzing parameters such as interbreath interval, tidal volume, and peak inspiratory flow [22-25]. However, to the best of our knowledge, there is currently no research specifically focusing on analyzing the

waveform of the IP signal itself in relation to extubation outcomes.



Figure 1 Bedside patient monitor

A. Electrocardiogram waveform and, **B.** impedance pneumography waveform

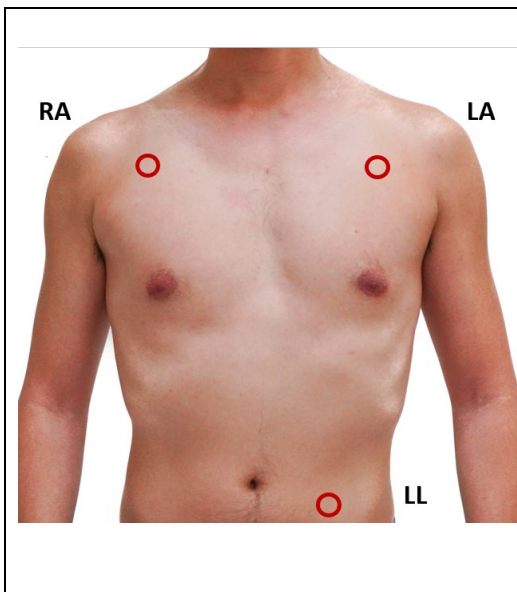


Figure 2 3-Lead positions of electrodes

RA placement: below the clavicle and near the right shoulder

LA placement: below the clavicle and near the left shoulder

LL placement: on the left lower abdomen

RA right arm; LA left arm; LL left lower abdomen



1.2 Motivation

Recent advancements in technology have led to the emergence of artificial intelligence (AI) models, including both machine learning (ML) and deep learning approaches. Several articles have indicated the superior predictive ability of AI models compared to conventional scoring systems alone in the context of PF. These AI models leverage the power of data analysis and pattern recognition to enhance prediction accuracy and provide valuable insights for clinical decision-making [26, 27]. In the development of AI models for predicting PF, clinical data such as vital signs, laboratory values, ventilator settings, and medication usage have commonly been utilized as parameters for training [28, 29]. However, the potential of incorporating features derived from the signals of heartbeat and breath to enhance the performance of AI models remains unknown. Currently, it is uncertain whether the inclusion of these signal-derived features could strengthen the predictive capabilities of AI models in the context of PF. Further research is needed to investigate the potential benefits and assess the impact of incorporating heartbeat and breath-related features in AI models.

1.3 Objectives

The goal of this study is to examine two key aspects. First, the investigation aims to determine whether features derived from the IP signal or HRV during SBT can be utilized for training the machine learning models to predict PF. Second, the study aims to evaluate whether the ML models trained using both clinical data and the signal features surpasses the performance of models trained solely on clinical data. By exploring these aspects, the study aims to assess the potential benefits of incorporating IP signal and HRV features in ML models and their impact on predicting PF.

Chapter 2 Method

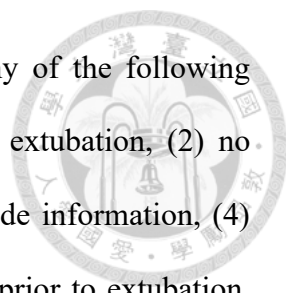


2.1 Data Source

The data used in this study is sourced from the Multi-parameter Intelligent Monitoring in Intensive Care (MIMIC-III) database, along with its subset called the MIMIC-III Waveform Database Matched Subset. MIMIC-III is an open database that provides access to de-identified electronic health records, including clinical measurements, laboratory results, and other relevant data from patients in ICUs. The MIMIC-III Waveform Database Matched Subset contains additional waveform data, such as ECG and IP signals, which can be used for more in-depth analysis and feature extraction in the study [30]. The Laboratory for Computational Physiology at Massachusetts Institute of Technology (MIT) developed the database, which is now accessible to the public through the PhysioNet platform. PhysioNet serves as a repository for sharing physiological signal and clinical data, providing researchers with open access to the database's resources [31]. The Beth Israel Deaconess Medical Center's ICU database contains extensive information regarding patients' admissions. This includes vital signs, medication details, laboratory measurements, care providers' notes, procedure records, length of stay, and image reports. Additionally, the MIMIC-III Matched Subset's Waveform database includes simultaneous recordings of various waveform signals such as IP signals, continuous arterial blood pressure, and ECG. These signals are digitized at a sampling rate of 125Hz, enabling detailed analysis and investigation.

2.2 Selection of Participation and Definition

Initially, we included all adult patients aged 20 or above who had experienced an




extubation event. Subsequently, we excluded patients who met any of the following criteria: (1) no waveform data available within 24 hours prior to extubation, (2) no definitive intubation record before extubation, (3) no ventilator mode information, (4) no SBT conducted using pressure support ventilation (PSV) mode prior to extubation, (5) no recorded peak inspiratory pressure (PIP) during SBT, (6) insufficient or conflicting information to determine the exact time of extubation (e.g., a time gap of over 1 hour between the timestamp of discontinuing MV and the extubation order), (7) use of noninvasive positive pressure ventilation (NIPPV) or high-flow nasal cannula (HFNC) within one hour after extubation due to hard to distinguish between curative use or prophylactic use and (8) waveform data not overlapping with the SBT interval. PF was defined by any of the following criteria: (1) reintubation, (2) use of NIPPV or HFNC, or (3) death within 48 hours following extubation [32].

In the records of MV, we considered the start time as the time of intubation. If the start time was not available, we used the time of the intubation order as a proxy for the time of intubation. When extubation time and the end time of MV duration were not consistent, to determine the time of extubation, we selected the earlier timestamp between the end of MV and the extubation order as the time of extubation. Because it was unclear whether the clinician prescribed the extubation order before or after the actual extubation, and this approach was adopted to avoid any potential errors in analyzing the post-extubation period.

2.3 Preprocessing

2.3.1 Clinical data

The parameters included in the analysis comprised:

- 
1. Demographic information: Age, gender, and body weight.
 2. Comorbidities: Diabetes mellitus, hypertension, chronic obstructive pulmonary disease, congestive heart failure, old myocardial infarction, moderate to severe renal disease, liver cirrhosis, old cerebrovascular disease, and organ recipients.
 3. Vital signs: Heart rate, respiratory rate (RR), body temperature, oxygen saturation, mean arterial pressure, and consciousness status.
 4. Laboratory data: FiO₂ (inspired fraction of oxygen), partial pressure of oxygen (PaO₂), PaO₂/FiO₂ ratio, pH value, partial pressure of carbon dioxide (PaCO₂), base excess (BE), anion gap, hemoglobin, white blood cell count, platelet count, blood urea nitrogen, and creatinine.
 5. Ventilator parameters: Pressure support level, tidal volume, RSBI, positive end-expiratory pressure (PEEP) level, mean airway pressure, duration of intubation, and SBT.
 6. Main diagnosis for ICU admission.
 7. Indication of intubation.

We only extracted the above features once, aiming to capture them at the time point closest to the timestamp of respiratory waveform analysis. However, if the nearest value occurs too early in relation to the respiratory waveform, that specific feature is regarded as a missing value. The time interval restrictions vary according to the type of parameter (**Table 1**). In addition, inconsistent values were also treated as missing values, such as MAP > 250 mmHg or < 20mmHg, pulse rate > 300/min or < 20/min, respiratory rate > 60/min, body temperature > 45 °C or < 30 °C, SpO₂ > 100 % or < 0%, and body weight > 300 kilograms or < 5 kilograms. Variables with a missing rate exceeding 20% were excluded from the analysis. These variables included body height, lactic acid, bilirubin,

sputum consistency/color/amount, cough effort/value/type, and central venous pressure. For the remaining parameters, missing values were imputed using the median values derived from other patients. This approach was chosen due to the abnormal distribution of the data.

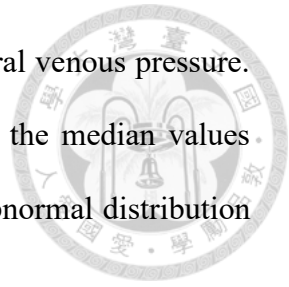


Table 1 Sampled criteria, missing rate, and exclusion of variables

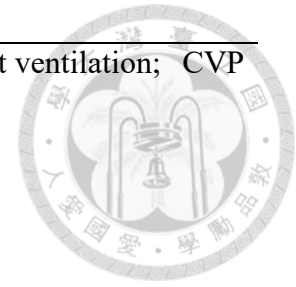
Variables	Sampled criteria	Missing value, n (%)	Inclusion
Body weight	During admission	32 (7.8)	O
Body height	During admission	156 (38.0)	X
Vital signs			
RR	1 hours within analysis	0	O
MAP	1 hours within analysis	11 (2.7)	O
HR	1 hours within analysis	0	O
BT	6 hours within analysis	79 (19.2)	O
SpO ₂	1 hours within analysis	0	O
GCS-motor			
GCS-motor	6 hours within analysis	11 (2.7)	O
GCS-eye			
GCS-eye	Consciousness, median (IQR)	28 (6.8)	O
Laboratory data			
FiO ₂	Closest to analysis	15 (3.6)	O
PaO ₂	12 hours within analysis	58 (14.1)	O
PH	12 hours within analysis	55 (13.4)	O
PaCO ₂	12 hours within analysis	69 (16.8)	O
BE	12 hours within analysis	69 (16.8)	O
Anion gap	12 hours within analysis	40 (9.7)	O
Hb	12 hours within analysis	37 (9.0)	O



WBC	12 hours within analysis	40 (9.7)	O
Platelet	12 hours within analysis	40 (9.7)	O
BUN	12 hours within analysis	32 (7.8)	O
Creatinine	12 hours within analysis	32 (7.8)	O
Lactic acid	12 hours within analysis	124 (30.2)	X
Bilirubin	12 hours within analysis	334 (81.3)	X
Ventilator data			
Mean P _{AW}	Closest to analysis	7 (1.7)	O
Tidal Volume	6 hours within analysis	62 (14.9)	O
PEEP level	Closest to analysis	0	O
PSV level	Closest to analysis	0	O
Clinical data			
Main diagnosis	According to the note	3 (<0.1)	O
Indication for intubation	recorded by doctors	5 (0.1)	O
Sputum Consistency	6 hours within analysis	268 (65.2)	X
Sputum color	6 hours within analysis	264 (64.2)	X
Sputum amount	6 hours within analysis	221 (53.8)	X
Cough effort	6 hours within analysis	136 (39.7)	X
Cough value	6 hours within analysis	402 (97.8)	X
Cough type	6 hours within analysis	240 (58.4)	X
CVP level	6 hours within analysis	162 (39.4)	X

RR respiratory rate; MAP mean arterial pressure; HR heart rate; BT body temperature; SpO₂ oxygen saturation from pulse oximeter; GCS Glasgow Coma Scale; FiO₂ inspired fraction of oxygen; PaO₂ partial pressure of oxygen in arterial blood; PaCO₂ partial pressure of carbon dioxide in arterial blood; BE base excess; Hb hemoglobin; WBC white blood cell; BUN blood urea nitrogen; P_{AW} airway pressure;

PEEP positive end-expiratory pressure; PSV Pressure support ventilation; CVP central venous pressure



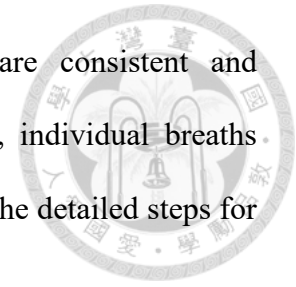
2.3.2 IP signal

The entire IP signal underwent a bandpass filter that eliminated frequency content outside the range of 0.05 Hz to 1.0 Hz. This filtering procedure aimed to exclude any irrelevant frequency components from the signal. Additionally, respiratory rates exceeding sixty breaths per minute or falling below three breaths per minute were considered artifacts and were therefore excluded from the analysis. These exclusion criteria were implemented to ensure that the filtered IP signal focused on the desired respiratory frequency range and to remove any abnormal respiratory rates that could potentially distort the analysis.

We analyzed only a 10-minute segment of the respiratory waveform at a time and extracted its corresponding feature values. The segment extraction started from the time closest to extubation, excluding the first five minutes before extubation to minimize the influence of interfering events such as sputum suction or extubation. After extracting the segment, an analytical approach was applied to ensure the quality of the IP signal that it could provide a high-quality representation of the respiratory activity for subsequent analysis.

The first stage of the analysis involved detecting individual breaths within the extracted segment of the IP signal. Before proceeding with the identification of individual breaths, the segment of the IP signal was detrended and normalized. Detrending involves removing any long-term trends or baseline variations from the signal, allowing for a focus on the respiratory variations. Besides, the signal was normalized to lead it with a mean of 0 and a standard deviation of 1. This normalization

process ensures that the signal's amplitude and distribution are consistent and comparable across different segments and patients. Subsequently, individual breaths were detected using a modified version of the Count-orig method. The detailed steps for detecting individual breaths are elaborated as follows [33, 34]:



1. Identifying peaks and troughs: A local extrema analysis approach was employed to identify peaks and troughs in the respiratory signal. Peaks were defined as points with amplitudes greater than the two neighboring points, and troughs were defined as points with amplitudes lower than the two neighboring points (**Figure 3B**).
2. Determining relevant peaks and troughs: Peaks with amplitudes above 0.2 times the 75th percentile of all peaks are considered relevant, while troughs with amplitudes below 0.2 times the 25th percentile of all troughs are classified as relevant (**Figure 3C**).
3. Selecting the highest amplitude peak and lowest amplitude trough: In cases where multiple relevant peaks were present between each pair of consecutive relevant troughs, only the peak with the highest amplitude was selected and retained for analysis. Similarly, among the adjacent relevant peaks, the relevant trough with the lowest amplitude is retained (**Figure 3D**).
4. Validating breath intervals: The interval between consecutive relevant peaks was defined as a valid breath, and the duration of each breath was measured by calculating the time between the start and end peaks.

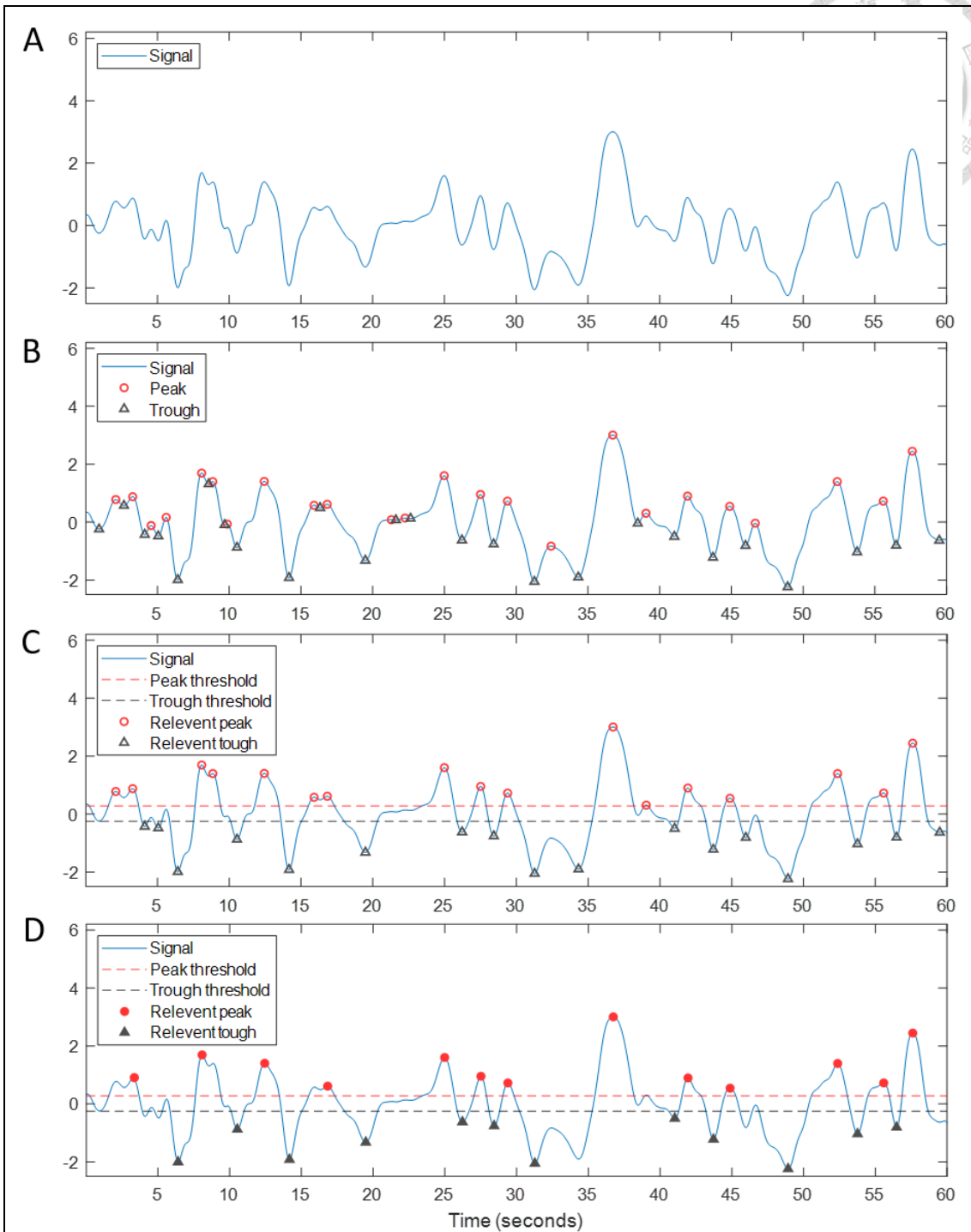
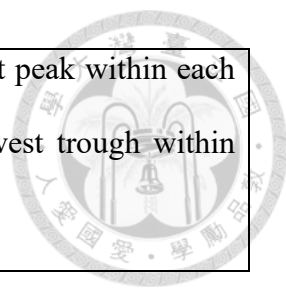


Figure 3 Flow chart for identification of relative peaks and troughs among impedance pneumography (IP) segment

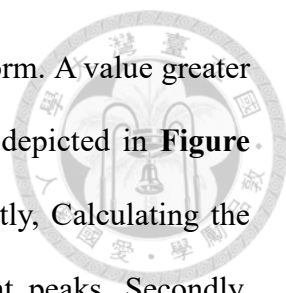
A. Original IP segment; **B.** Find all peaks and troughs; **C.** Yield relevant peaks and troughs by excluding peaks lower than 0.2 times the 75th of all peaks and troughs



higher than 0.2 times the 25th of all troughs; **D.** Retain one highest peak within each pair of consecutive relevant troughs, and similarly, retain the lowest trough within among the adjacent relevant peaks.

In the second stage, a set of thresholds and the method proposed by Charlton et al. were applied to assess the quality of the IP segment. The analysis was conducted in the following sequence [34]:

1. Mean RR and waveform occupancy: The mean RR within the IP segment should fall between 6 and 60 breaths per minute. Additionally, the analyzable waveform should occupy over 80% of the segment to ensure an adequate number of breaths for analysis (as depicted in **Figure 4B**).
2. Amplitude difference threshold: The difference between the maximum and minimum amplitudes of peaks or troughs should not exceed twenty times. This threshold is employed to exclude segments with interfering spikes (as illustrated in **Figure 4C**).
3. Normalized standard deviation of breath durations: The normalized standard deviation of breath durations should not exceed 0.25. This criterion is used to allow only a lower variation in breath duration (as shown in **Figure 4D**).
4. Outlier breath durations: Outlying breath durations are defined as those that are more than 1.5 times or less than 0.5 times the median breath duration. To avoid erroneous interpretation caused by outlying breaths, the proportion of outlying breaths should be less than 15% of all breaths, and the duration of outlier breaths should occupy less than 40% of the analyzed breathing duration.
5. Assessment of breath morphology similarity: The similarity of breath morphologies is evaluated by calculating the mean correlation coefficient between



the average breath waveform and each individual breath waveform. A value greater than 0.75 is considered indicative of high-quality breaths (as depicted in **Figure 4E**). To yield the average breath wave in the IP segment, firstly, Calculating the mean breath interval by averaging the time between relevant peaks. Secondly, extracting individual breaths by segmenting the signal based on the mean breath interval. Each segment is centered on a relevant peak, representing the start of a breath. Normalizing the individual breath segments by applying their Euclidean norm. Last, Computing the average waveform template by taking the mean of all individual breaths, and aligned each individual breath with its corresponding waveform template. The similarity of breath morphologies was assessed by quantifying the mean correlation coefficient between individual breaths and the average breath template.

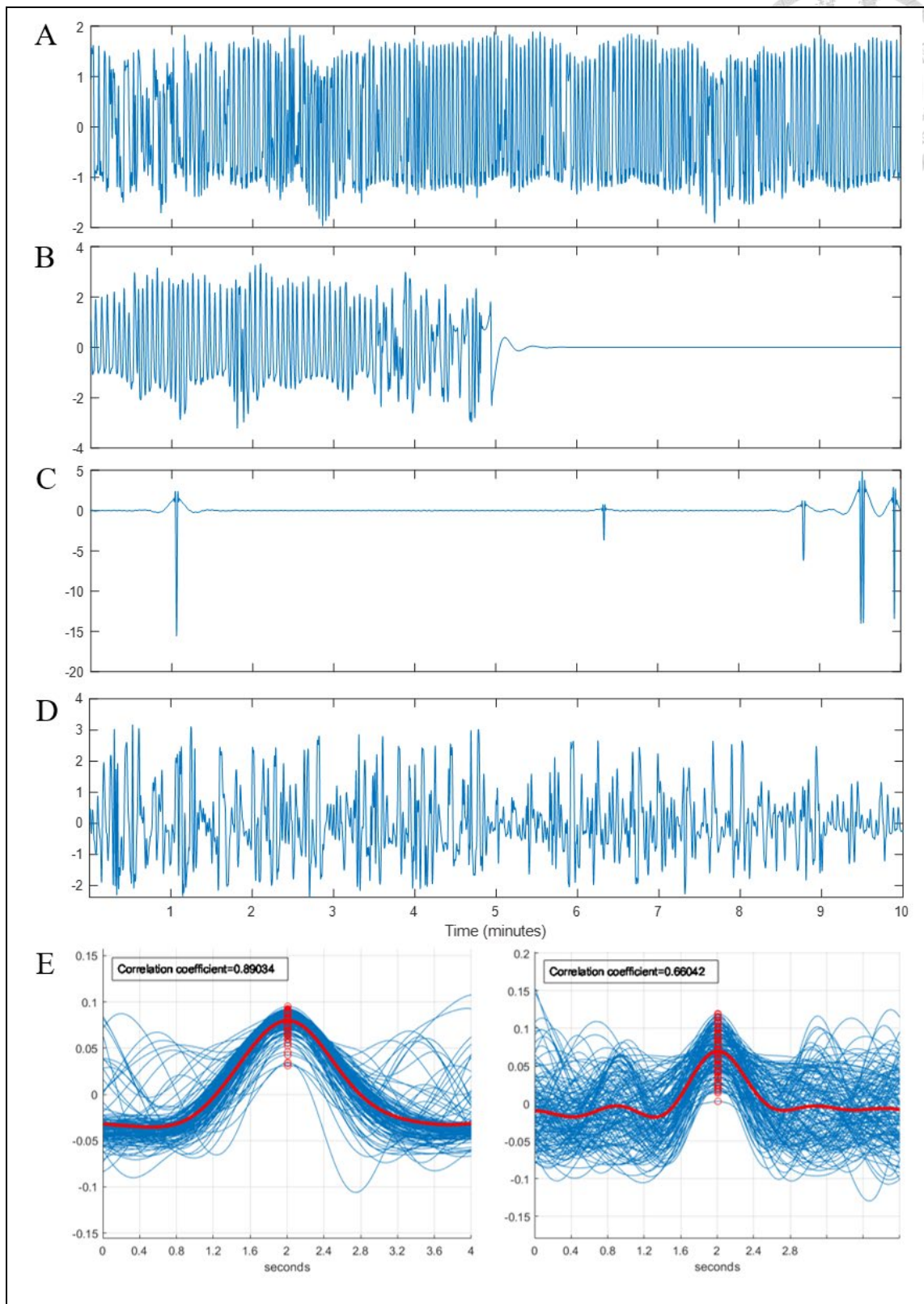


Figure 4 Examples of different quality of impedance pneumography (IP) segments
A. High-quality IP segment; **B.** Low-quality IP segment with less than 80% segment

occupied with breaths; **C.** Low-quality IP segment with interfering spikes; **D.** Low-quality IP segment with high standard deviation of breath duration (0.53); **E.** High-quality IP segment with mean correlation coefficient value of 0.89 on the left side, and low-quality IP segment with value of 0.66 on the right side

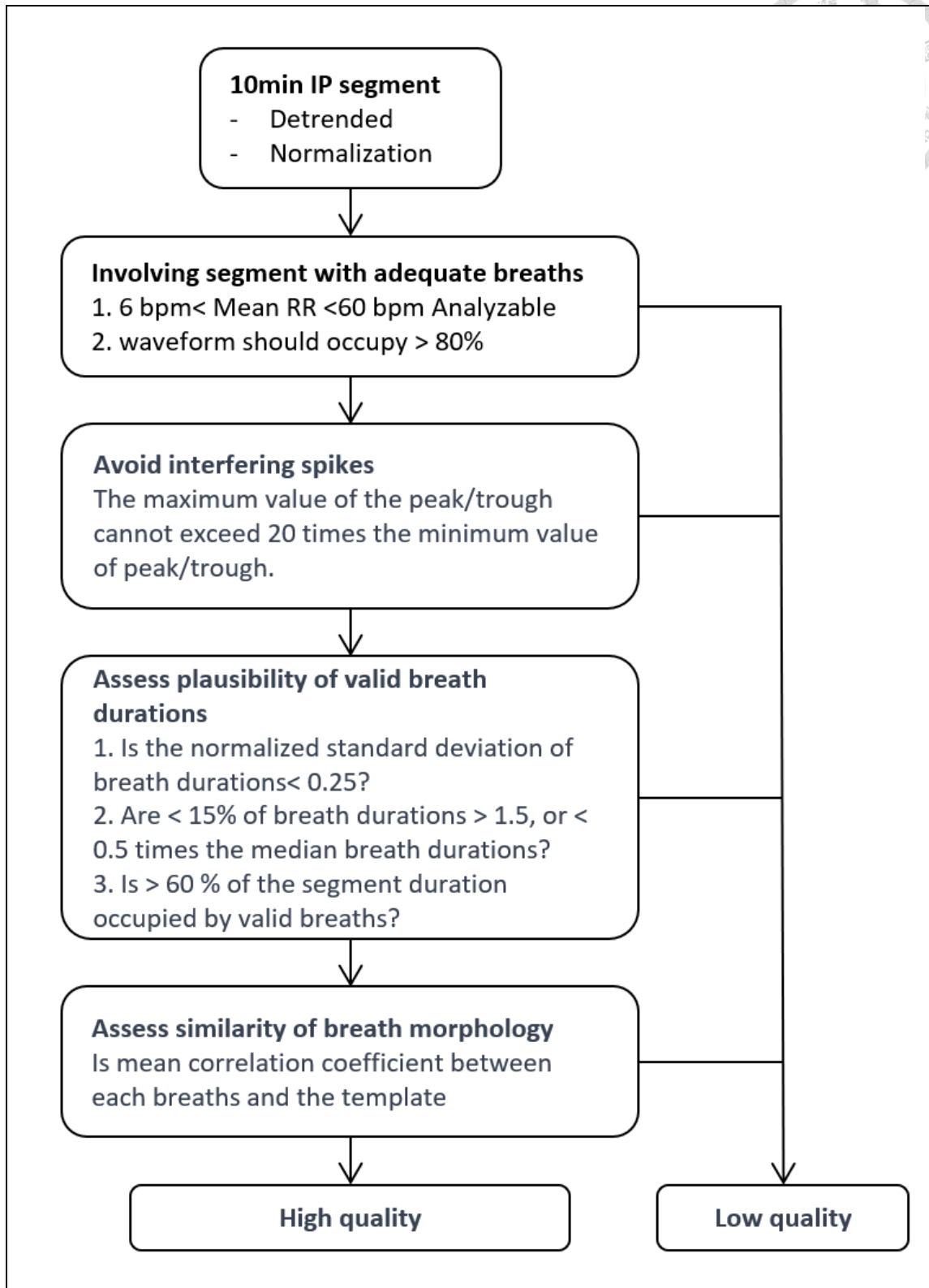
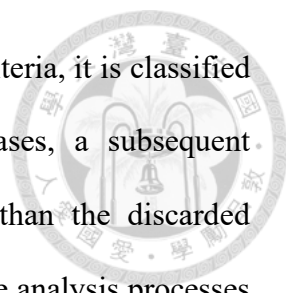


Figure 5 Sequential assessment of IP waveform quality

IP impedance pneumography; bpm breaths per minute; RR respiratory rate



As shown in **Figure 5**, if an IP segment fails to meet any of the criteria, it is classified as low quality and excluded from further analysis. In such cases, a subsequent 10-minute IP segment is generated, starting 30 seconds earlier than the discarded segment. This new segment then undergoes stages one and two of the analysis processes once again. This iterative process continues until an IP segment satisfies all the criteria in the second stage. Only IP segments meeting all the quality criteria proceed to the final stage of the analysis, ensuring that the selected segments are of sufficient quality and meet the established criteria for reliable and accurate assessment of the respiratory waveform.

In the final stage of the analysis, various features were extracted from the IP segment. These included the mean RR, standard deviation of RR, ratio of inspiration and total breath, waveform slopes of inspiration and expiration, inspiration amplitude variation, frequency-domain analysis, and sample entropy.

Additionally, we localized three quartiles (0.25, 0.50, and 0.75 percentile) in both inspiration and expiration in each breath along the time axis (**Figure 6**). Following the identification of quartiles within the inspiration and expiration periods, the waveform slope of each quartile was calculated using the two adjacent points. For instance, the slope of the first quartile in inspiration was determined by the coordinates of the nearest relevant trough and the second quartile within the same inhalation. To calculate the slope of the entire inspiration or expiration, the adjacent relevant trough and peak were utilized. The coefficient of variation (CoV) of slope among inspiration and expiration was obtained by dividing the mean slope by the standard deviation of slopes. Furthermore, the slope ratio of each quartile was computed by dividing the slope of the quartile by the slope of the entire corresponding period. For example, the slope ratio of the second quartile during expiration was obtained by dividing the slope of the second

quartile by the slope of the entire expiration.

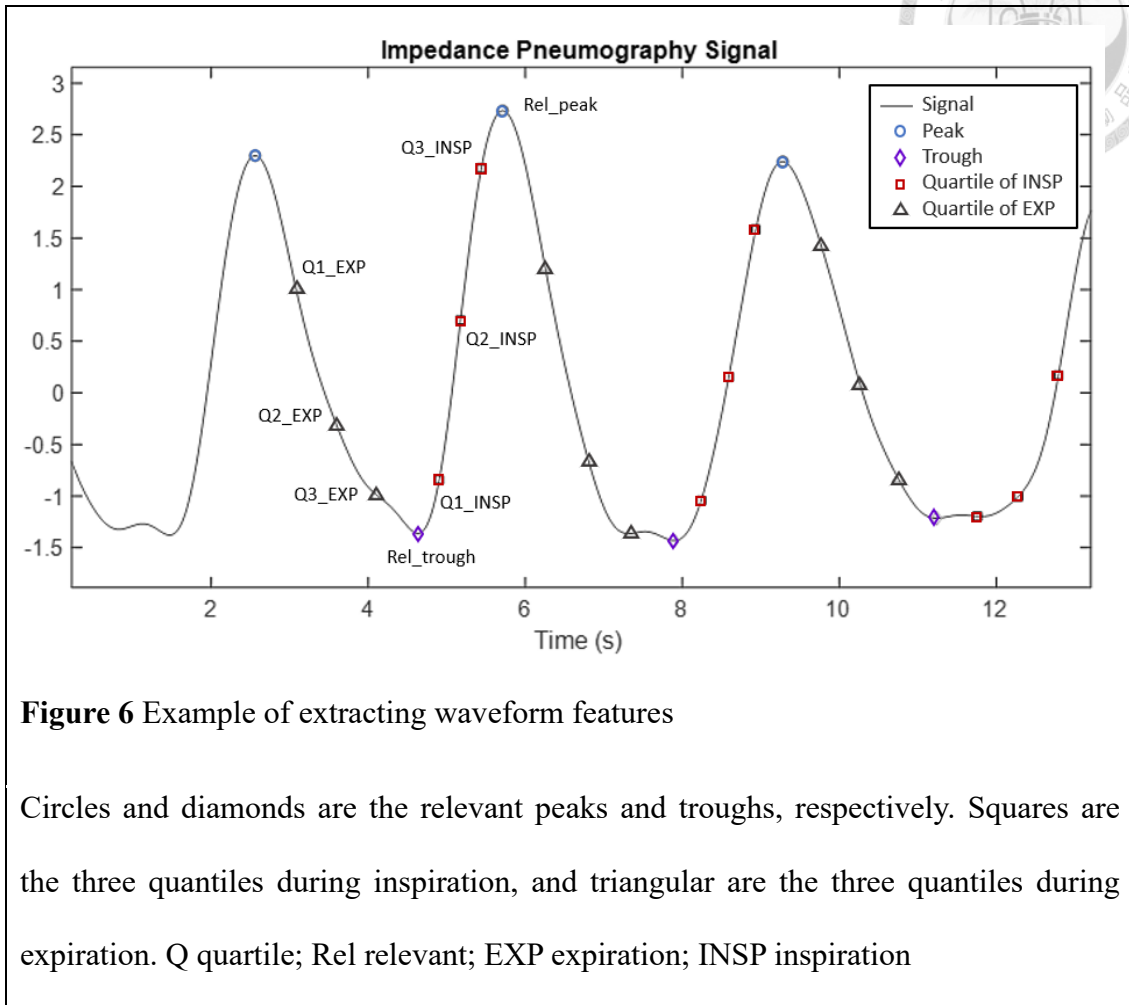
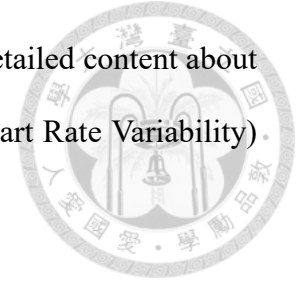


Figure 6 Example of extracting waveform features

Circles and diamonds are the relevant peaks and troughs, respectively. Squares are the three quartiles during inspiration, and triangular are the three quartiles during expiration. Q quartile; Rel relevant; EXP expiration; INSP inspiration

Frequency-domain analysis was conducted using the Fast Fourier Transform (FFT) algorithm to analyze the IP segment. The Fourier Transform is a mathematical technique used to transform a function of time, a temporal signal, into a function of frequency. The FFT is an algorithm that efficiently computes the Discrete Fourier Transform. It is termed 'fast' because it significantly reduces the computation time and complexity from $O(n^2)$ in DFT to $O(n \log n)$. This efficiency is gained by exploiting the symmetry in the computation of the DFT. This analysis provided insights into the distribution of power across different frequency components within the signal. Parameters such as total power (TP), very low frequency (VLF), low frequency (LF), high frequency (HF), peak of HF,

peak of LF, and LF/HF ratio were derived from this analysis. The detailed content about the frequency domain will be introduced in the following HRV (Heart Rate Variability) section.



2.3.3 HRV signal

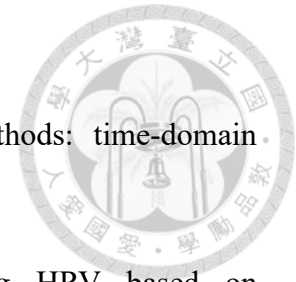
A ten-minute EKG segment was obtained at the timestamp of analysis for IP segment. R peak detection was performed using an optimized Pan–Thompkins algorithm, enabling the calculation of R-peak to R-peak intervals (RRi) [35, 36]. To ensure the inclusion of reliable signals, a threshold was applied to exclude segments with a heart rate below 40 beats per minute. To address the potential bias introduced by ectopic beats, which can distort HRV analysis, the ECG segment underwent additional processing. A modified threshold-based detection algorithm was employed to identify ectopic beats. Any heartbeat with an RRi that increased or decreased by more than 15% compared to the median value of the previous five RRi was considered an ectopic beat and subsequently discarded. This approach aimed to alleviate the influence of ectopic beats on the subsequent HRV analysis, as supported by previous research, and as showed in **Equation (1)** [37-40].

$$RRi = \{RRi \mid 0.85 \times RRi_{median} < RRi_n < 1.15 \times RRi_{median}\} \quad (1)$$

An acceptable range of ectopic beats was defined as two percent. If an ECG segment did not meet these criteria, it was excluded from further analysis. As a result, a subsequent 10-minute ECG segment was generated, starting 30 seconds earlier than the excluded segment. This new segment then underwent R peak detection again. Conversely, if a segment satisfied both the heart rate threshold and the ectopic ratio criteria, it underwent interpolation using the cubic spline method. This interpolation method was applied to replace the discarded beats and ensure a continuous and

uninterrupted ECG waveform [41].

The HRV analysis was conducted using three types of methods: time-domain analyses, frequency-domain analyses, and non-linear analyses.




1. Time-domain analyses: These methods involve analyzing HRV based on time-related measures. The following parameters were calculated:

- (a) Difference between the highest and lowest heart rate (HR Maximal - HR Minimal): Measures the range of heart rate fluctuations.
- (b) Standard deviation of individual normal-to-normal intervals (SDNN): Reflects the overall variability of the normal heartbeats.
- (c) Root mean square of successive differences between normal heartbeats (RMSSD): Provides a measure of short-term variability.
- (d) Proportion of NN50 divided by the total number of normal-to-normal intervals (pNN50): Evaluates the proportion of adjacent normal-to-normal intervals that differ by more than 50 milliseconds.
- (e) HRV triangular index: Represents the integral of the density distribution of normal-to-normal intervals.

2. Frequency-domain analyses: These methods involve analyzing HRV based on frequency components obtained through spectral analysis. The following parameters were derived:

- (a) TP: Represents the total power of the HRV signal across all frequency ranges.
- (b) LF: Reflects the power within the low-frequency range (typically 0.04 to 0.15 Hz).
- (c) HF: Represents the power within the high-frequency range (typically 0.15 to 0.4 Hz).

- 
- (d) VLF: Represents the power within the very low-frequency range (typically below 0.04 Hz).
 - (e) LF/HF ratio: Provides an indicator of the balance between sympathetic and parasympathetic influences on HRV.

3. Non-linear analyses: These methods evaluate HRV from a non-linear perspective, assessing the complexity and patterns in the RRi time series. The following parameters were calculated:

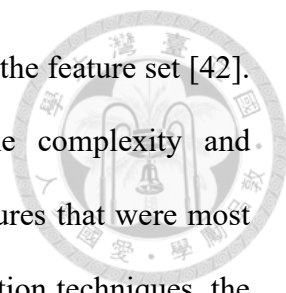
- (a) SD1 and SD2 derived from the Poincaré plot: Measures of short-term and long-term HRV, respectively.
- (b) Sample entropy: A measure of the complexity and irregularity of the RRi time series.

By employing these different analysis methods, a comprehensive evaluation of HRV was performed, capturing various aspects of HRV and providing insights into the autonomic modulation of the cardiovascular system.

To simulate real-life scenarios, clinical data were collected at the timestamp of analysis for waveform. Categorical variables were converted into one-hot encoding to ensure compatibility with the machine learning algorithms. Continuous features, including clinical data, IP signal, and HRV, were standardized through standardization scaling, resulting in variables with a mean of 0 and a standard deviation of 1.

2.4 Feature Selection

For feature selection, several steps were taken. Initially, correlation analysis was performed to identify highly correlated features, with a cutoff threshold of 0.85. Highly correlated features were removed to avoid redundancy and potential overfitting. Next, a



filter method using mutual information was applied to further refine the feature set [42]. This method helped filter out irrelevant features, reducing the complexity and redundancy of the dataset. The aim was to identify a subset of features that were most relevant to the classification task. After applying these feature selection techniques, the importance rank of each feature was assessed. From this ranking, the thirty most relevant features were selected to train the models. This approach ensured that the models were trained on the most informative and discriminative features, optimizing their performance for the classification task at hand.

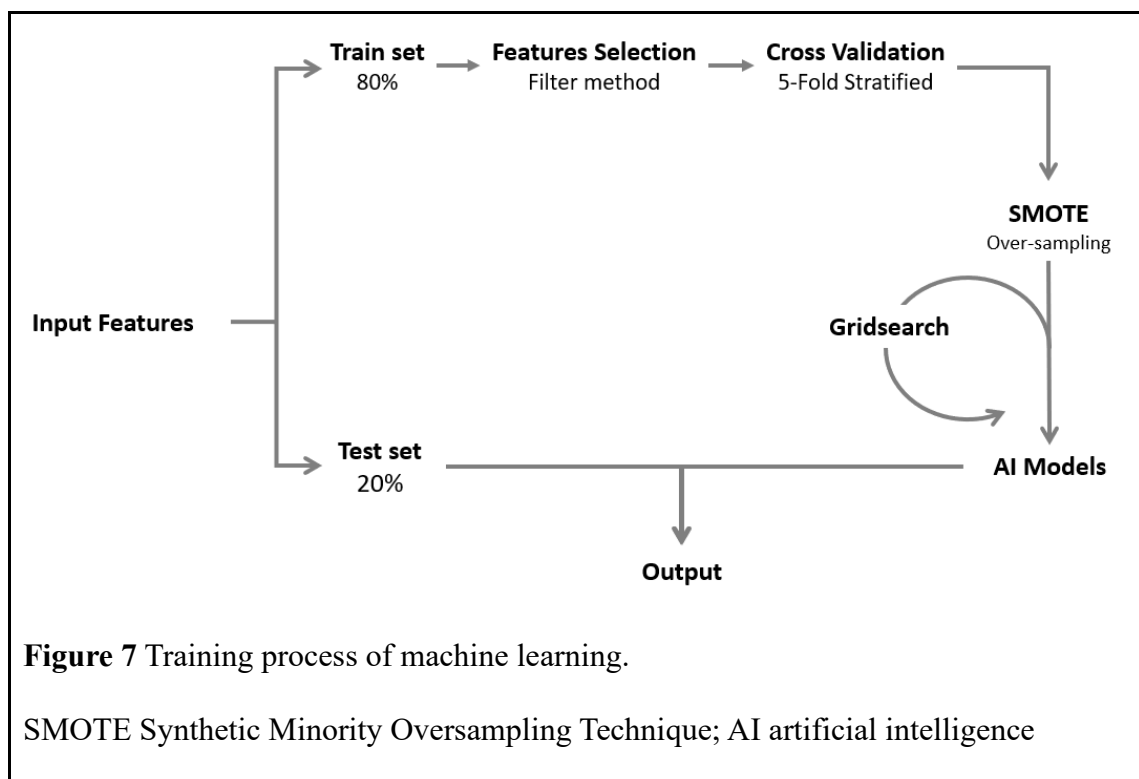
2.5 Training Pipeline

The training process was visually represented in **Figure 7**. The dataset was divided into an 80% training set and a 20% testing set. To ensure robustness and reliable performance assessment, 5-fold cross-validation was employed during the hyperparameter optimization, training, and evaluation stages.

Considering the imbalance in the distribution of classes (successful and failed extubation), the Synthetic Minority Oversampling Technique (SMOTE) was applied to resample the dataset. SMOTE is a technique that over-samples the minority class by generating synthetic samples, thereby balancing the number of patients between the two classes [43]. This adjustment was made to prevent AI algorithms from being biased towards the majority class in an imbalanced dataset. This facilitated more accurate training and classification by ensuring that the model was exposed to a balanced representation of both successful and failed extubation cases.

The combination of dataset splitting, cross-validation, and SMOTE resampling ensured a robust and fair evaluation of the AI models, accounting for the class

imbalance and minimizing the risk of bias in the analysis. We constructed binary classifiers using five well-known machine learning models: Extreme Gradient Boosting (XGBoost), LightGBM, CATegorical Boosting (CATBoost), Random Forest (RF), and Logistic Regression (LR) [44-48]. These models are widely used and have demonstrated strong performance in various classification tasks. By utilizing a diverse set of models, we aimed to capture different aspects of the data and leverage the unique advantages of each model to achieve accurate and robust predictions.



In this study, hyperparameter optimization was performed using the grid search method. This technique systematically explored all possible combinations of specified hyperparameters and selected the optimal values based on pre-defined performance metrics. The models were trained and assessed using 5-fold cross-validation, which ensured reliable performance estimation. These optimal hyperparameter values were then utilized to fine-tune the models and enhance their predictive capabilities.



2.6 Performance Evaluation

During the cross-validation, the mean performance of the models was assessed using various evaluation metrics. These metrics included the area under the receiver operating characteristic curve (AUROC), the area under the precision-recall curve (AUPRC), accuracy, precision (positive predictive value), recall (sensitivity), F1 score, negative predictive value, and specificity. These metrics provided a comprehensive evaluation of the models' performance in predicting extubation outcomes.

To further analyze the impact of different types of data on model performance, the models were trained using different combinations of data types. To compare the AUROC values between individual models, DeLong's test was applied to assist in determining if there is a significant difference in the performance of the models [49, 50].

To interpret and explain the output of the models, the SHapley Additive exPlanations (SHAP) method was applied. The SHAP method, implemented using the SHAP Python package (version 0.41.0), provides insights into the importance and contributions of different features in the models' predictions. The results of the SHAP analysis were visualized using bee swarm plots, where each dot represents an individual data point in the model, helping to understand the impact of various features on the model's predictions [51].

By employing these evaluation metrics, statistical tests, and interpretability techniques, a comprehensive assessment of the models' performance and their underlying explanations was conducted, enabling a better understanding of the predictive capabilities and feature importance in predicting extubation outcomes.



2.7 Experiment Environment

The preprocessing of clinical data and signal in this study was performed using MATLAB (version R2021a). MATLAB provided the necessary tools and functions for data manipulation, feature extraction, and signal processing. For constructing and training the AI models, Python (version 3.8.10) was utilized on the Google Colab platform. Python, with its rich ecosystem of ML libraries such as scikit-learn, offered a flexible and powerful environment for building and training the AI models. Google Colab, being a cloud-based platform, provided computational resources and easy access to libraries and packages required for machine learning tasks [52].

Chapter 3 Result



3.1 Participant Characteristics

In the final cohort derived from the MIMIC-III and MIMIC-III Waveform Database Matched Subset (**Figure 8**), a total of 411 extubated patients met all the inclusion criteria. For model training, 328 patients (80% of the dataset) were used, while the remaining 83 patients (20%) were reserved for model testing. Among the total population, the median age of the patients was 67.0 years, and 134 patients (32.7%) were female. The median duration of MV was 13.7 hours, and the median duration of SBT with PSV mode was 3.5 hours (**Table 2**). In **Table 3**, it is reported that out of the 411 patients, 46 patients (11.2%) experienced PF. Among these patients, one patient died, thirteen patients required re-intubation, thirteen patients required noninvasive positive pressure ventilation (NIPPV), and twenty patients required high flow nasal cannula (HFNC) within 48 hours after extubation. The median time to PF was 13.3 hours.

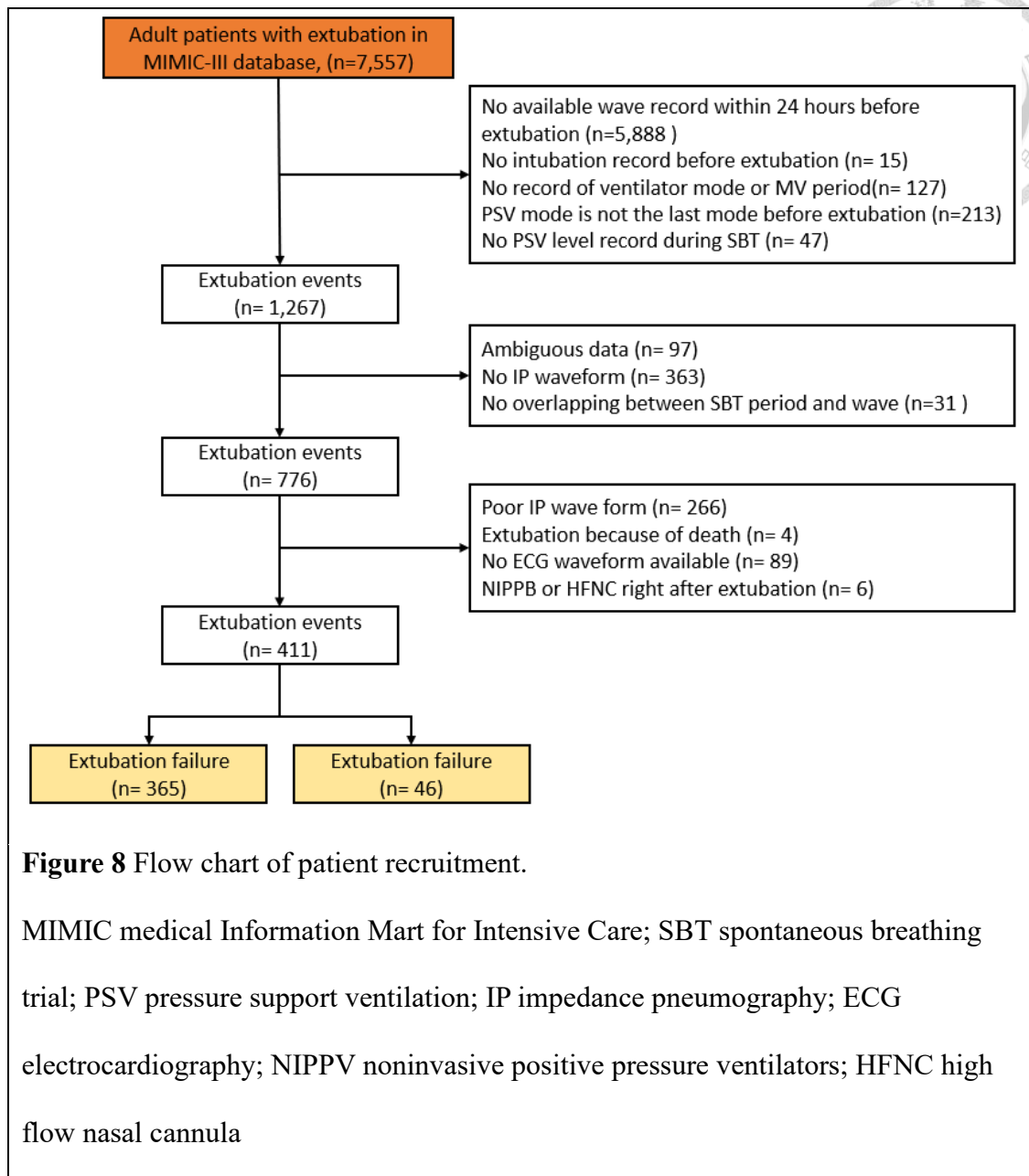



Figure 8 Flow chart of patient recruitment.

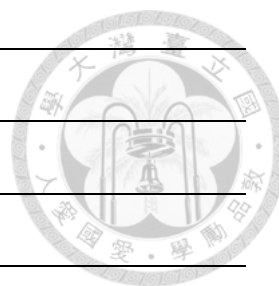
MIMIC medical Information Mart for Intensive Care; SBT spontaneous breathing trial; PSV pressure support ventilation; IP impedance pneumography; ECG electrocardiography; NIPPV noninvasive positive pressure ventilators; HFNC high flow nasal cannula

Table 2 Demographic characteristics of enrolled patients (Total patients = 411)

Variables	Value	MV (%)
Age, median (IQR)	67.0 (58.9-75.3)	0
Sex-female, No. (%)	134 (32.7)	0
BW, median (IQR)	83.0 (71.8-95.9)	32 (7.8)
Vital signs, median (IQR)		



RR (per min)	18.0 (14.6-21.1)	0
MAP (mmHg)	78 (70-88)	11 (2.7)
HR (bpm),	84 (73-95)	0
BT (Celsius)	37.1 (36.7-37.4)	79 (19.2)
SpO ₂ (%)	99 (97-100)	0
Consciousness, median (IQR)		
GCS-motor median	6 (5-6)	11 (2.7)
GCS-eye	3 (3-4)	28 (6.8)
Lab, median (IQR)		
PaO ₂ /FiO ₂ ratio	262 (222.5-329.6)	NA
FiO ₂	40 (40-50)	15 (3.6)
PaO ₂ (mmHg)	122 (97-152)	58 (14.1)
PH	7.39 (7.35-7.42)	55 (13.4)
PaCO ₂ (mmHg)	40 (36-44)	69 (16.8)
BE (mmol/L)	0 ((-2)-1)	69 (16.8)
Anion gap	11 (10-13)	40 (9.7)
Hb (g/dL)	10.7 (9.8-11.8)	37 (9.0)
WBC (10 ³ /μL)	11.5 (8.5-14.7)	40 (9.7)
Platelet (10 ³ /μL)	166 (124.3-224.8)	40 (9.7)
BUN (mg/dL)	17 (12-22)	32 (7.8)
Cre (mg/dL)	0.9 (0.7-1.1)	32 (7.8)
Underlying disease, No. (%)		
DM	129 (31.4)	NA
COPD	38 (9.2)	NA



Hypertension	255 (62.0)	NA
CHF	93 (22.6)	NA
Severe CKD	28 (6.8)	NA
Old MI	35 (8.5)	NA
Cirrhosis	39 (9.5)	NA
Stroke	21 (5.1)	NA
Organ transplant	9 (2.2)	NA
Ventilator data, median (IQR)		
Mean P _{AW} (mmHg)	7 (6-8)	7 (1.7)
RSBI	40.7 (30.9-52.1)	NA
Tidal Volume (ml)	466.0 (388.3-570.3)	62 (14.9)
PEEP level	5 (5-5)	0
PSV level	5 (5-5)	0
T _{VEN} (hours)	13.7 (5.6-20.5)	0
T _{PSV} (hours)	3.5 (1.3-8.8)	0
Main diagnosis		
Cardiovascular	225 (54.7)	
Central nerve system	73 (17.8)	
Gastrointestinal-related	56 (13.6)	
Oncology-related	16 (3.9)	3 (<0.1)
Pulmonary-related	14 (3.4)	
Vertebral	10 (2.4)	
Others	17 (4.1)	
Indication of intubation		
Surgery	300 (73.0)	5 (0.1)



Airway protection	66 (16.1)
Respiratory distress	32 (7.8)
Others	13 (3.2)

IQR interquartile range; NA not access; MV missing value; No. numbers; BW body weight; RR respiratory rate; MAP mean arterial pressure; HR heart rate; BT body temperature; SpO₂ oxygen saturation from pulse oximeter; GCS Glasgow Coma Scale; FiO₂ inspired fraction of oxygen; PaO₂ partial pressure of oxygen in arterial blood; PaCO₂ partial pressure of carbon dioxide in arterial blood; BE base excess; Hb hemoglobin; WBC white blood cell; BUN blood urea nitrogen; Cre Creatinine; COPD chronic obstructive pulmonary disease; CHF congestive heart failure; MI myocardial infarction; CKD chronic kidney disease; P_{AW} airway pressure; RSBI rapid shallow breathing index; T_{VEN} duration of invasive ventilation; T_{PSV} duration of PSV mode

Table 3 Outcome analysis

Item	Value
IP signal analysis time from extubation, minutes (median, IQR)	44.9 (8.0-40.5)
HRV analysis time from extubation, minutes (median, IQR)	48.7 (8.5-51.3)
Failure extubation (n, %)	47 (11.4)
Re-intubation	13
NIPPV	13
HFNC	20
Death	1
Mean failure time, hours (median, IQR)	13.3 (5.3-28.8)

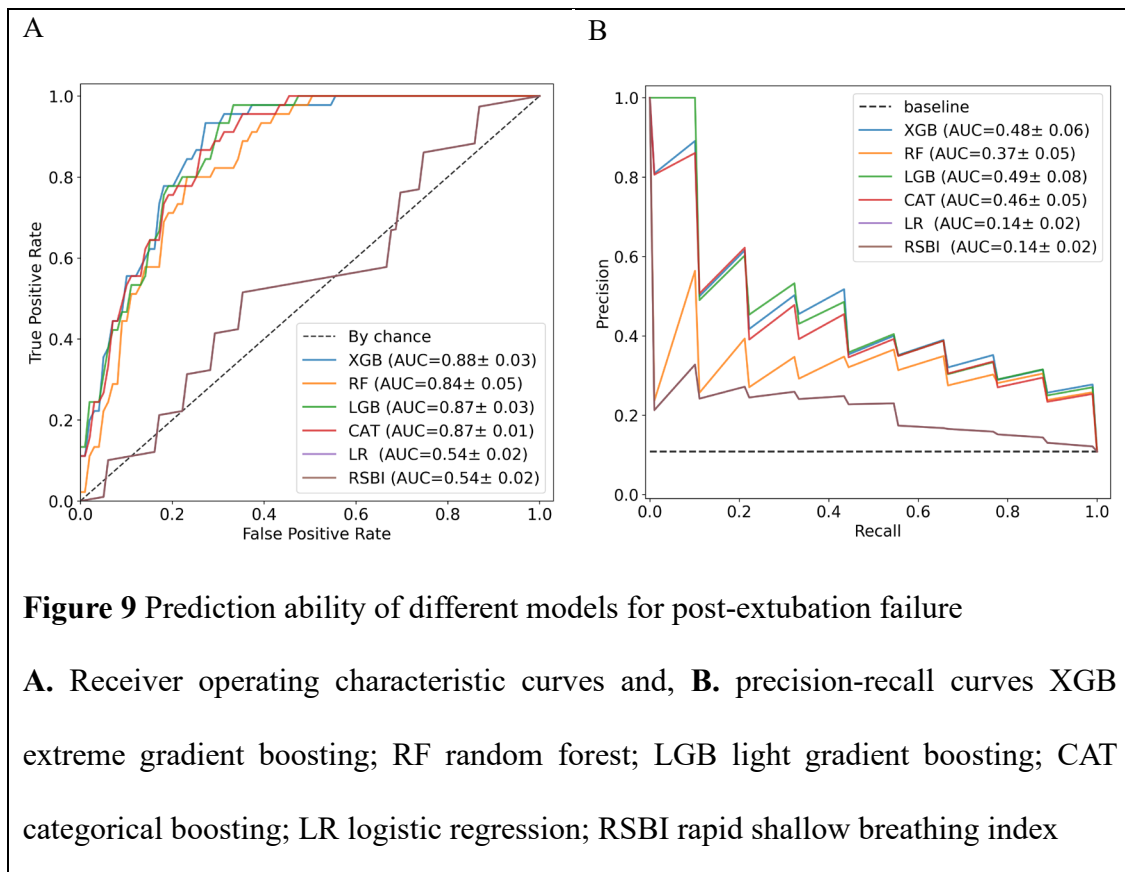
IP impedance pneumography; HRV heart rate variability; IQR interquartile range;

NIPPV non-invasive positive pressure ventilation; HFNC high flow nasal cannula



3.2 Performance

Figure 9 depicts the ROC curves and PRC curves on the test set, illustrating the prediction ability of different models for PF. Compared to the LR-RSBI model, other models had higher performance, The AUROC values for each model were as follows: XGBoost- 0.877, LightGBM - 0.871, CATBoost- 0.868, RF- 0.841, LR- 0.774, and LR-RSBI- 0.540. **Table 4** presents the performance metrics of the XGBoost classifier, which outperformed the other classifiers in terms of AUPRC (0.479), F1-score (0.398), sensitivity (0.333), specificity (0.962), positive predictive value (0.537), and negative predictive value (0.922). The hyperparameters of each machine learning model used in the study are presented in **Table 5**.



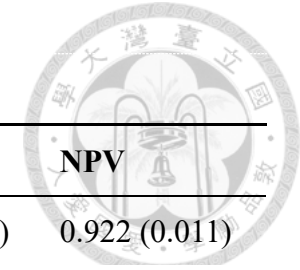


Table 4 Performance of different machine learning models

Models	Threshold	AUROC	AUPRC	F1-score	Sensitivity	Specificity	PPV	NPV
XGBoost	0.5	0.877 (0.032)	0.479 (0.055)	0.398 (0.087)	0.333 (0.111)	0.962 (0.020)	0.537 (0.111)	0.922 (0.011)
LGBM	0.3	0.871 (0.029)	0.486 (0.080)	0.297 (0.154)	0.267 (0.186)	0.951 (0.023)	0.393 (0.118)	0.915 (0.019)
CatBoost	0.5	0.868 (0.008)	0.458 (0.054)	0.366 (0.057)	0.333 (0.111)	0.946 (0.027)	0.446 (0.051)	0.921 (0.010)
RF	0.45	0.841 (0.049)	0.368 (0.052)	0.242 (0.0443)	0.222 (0.079)	0.930 (0.046)	0.300 (0.052)	0.908 (0.005)
LR	0.7	0.774 (0.025)	0.224 (0.024)	0.171 (0.070)	0.111 (0.111)	0.881 (0.032)	0.096 (0.103)	0.891 (0.013)
LR-RSBI	0.55	0.540 (0.022)	0.137 (0.016)	0.128 (0.009)	0.067 (0.061)	0.954 (0.043)	0.153 (0.024)	0.894 (0.002)

XGBoost extreme gradient boost; LGB light gradient boosting; RF random forest classifier; LR logistic regression; RSBI rapid shallow breathing index; PPV positive predictive value; NPV negative predictive value; AUROC area under receiver-operator curve; AUPRC area under precision-recall curve

Table 5 Hyperparameters for individual model

Model	Hyperparameter	Value	Model	Hyperparameter	Value
XGBoost	learning_rate	0.2	LightGBM	learning_rate	0.3
	max_depth	5		max_depth	10
	n_estimators	30		colsample_bytree	0.3
	colsample_bytree	0.5		n_estimators	60
CATBoost	Depth	7	Logistic Regression	C	0.1
	learning_rate	0.05		penalty	L1
	Iterations	300		solver	saga
	bagging_temperature	1			
Random Forest	bootstrap	False			
	min_sample_leaf	4			
	min_samples_split	4			
	n_estimators	100			

XGBoost extreme gradient boost; CATBoost CATegorical Boosting; LGB light gradient boosting

3.3 Feature Importance

Figure 10A illustrates the analysis of feature importance using the filter method of mutual information. The five most important features are the Mean_slope_insp_Q1_ratio (mean slope ratio of the first quartile during inspiration), HR_Max-HR_min, CoV of the first quartile during inspiration, MV duration, mean slope of the third quartile during inspiration. Notably, features derived from the IP waveform and HRV make up over half of the top ten important features. These results highlight the significant contribution of IP waveform and HRV features in predicting PF.

Figure 10B presents the SHAP values for the XGBoost classifier trained with the selected features mentioned above. The top twenty features are displayed, with each row representing a feature. The SHAP value is plotted on the horizontal axis, and each data sample is represented by a single point. Features with higher values are shown in red, indicating an elevated risk of PF, while features with lower values are displayed in blue, indicating a reduced risk. The length of the feature bar reflects the magnitude and direction of the feature's effect on the model's output. A longer bar suggests a greater impact on the model's predictions, while a shorter bar indicates a smaller impact. A positive SHAP value indicates an increased risk of PF, while a negative value suggests a decreased risk. By examining the top five SHAP values of the XGBoost classifier, we can observe that the duration of MV, mean slope during inspiration (mean_slope_insp), and mean slope ratio on the first quartile during inspiration (Mean_slope_insp_Q1_ratio) positively correlate with the risk of PF. Conversely, lower values of PaO₂ (PAO2_value) and motor part of Glasgow Coma scale (GCS_motor_full) are associated with a higher risk of PF.

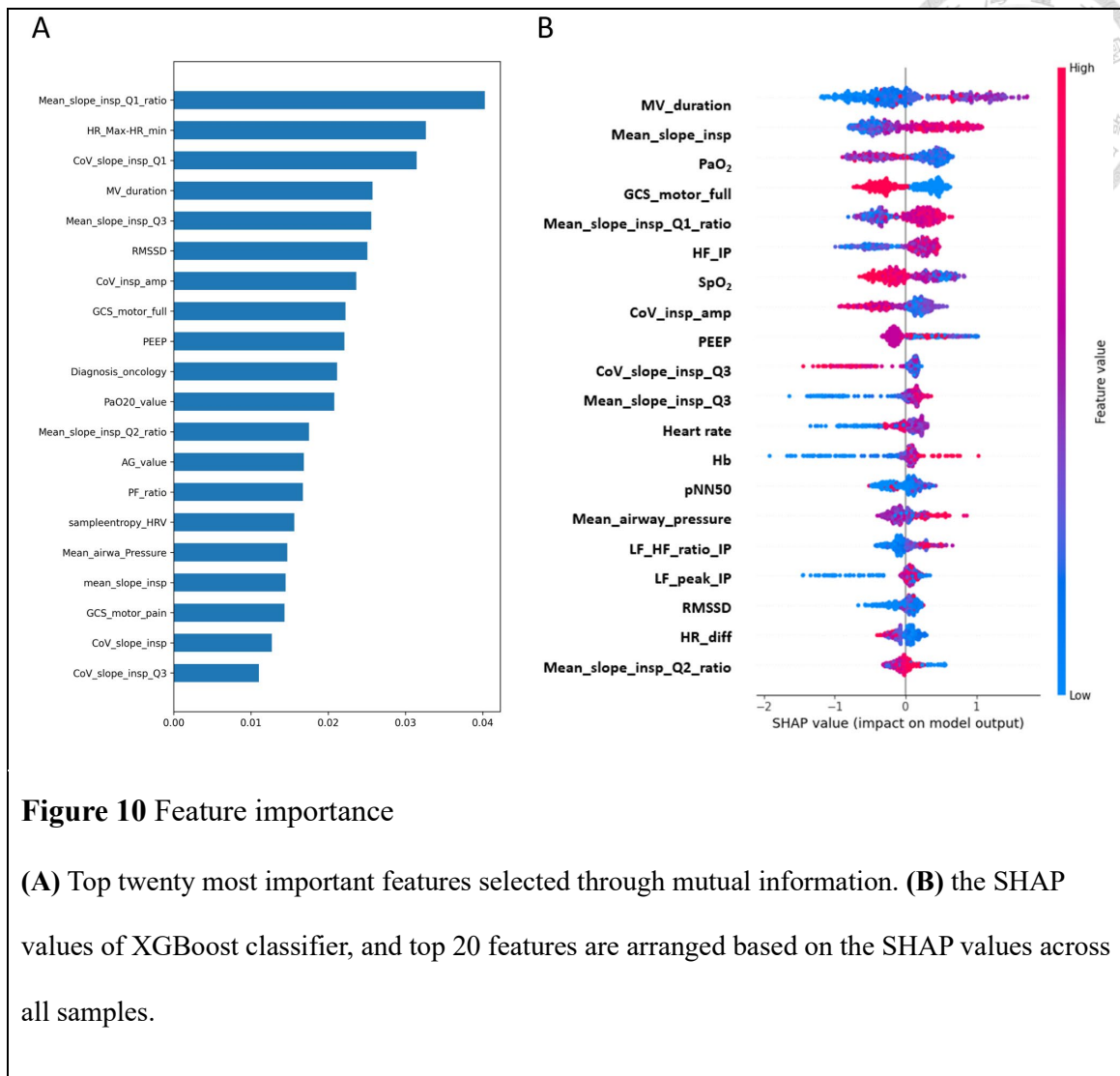


Figure 10 Feature importance

(A) Top twenty most important features selected through mutual information. (B) the SHAP values of XGBoost classifier, and top 20 features are arranged based on the SHAP values across all samples.

3.4 XGBoost Classifier with Different Input

In our study, we explored different combinations of input features for training the XGBoost classifier. These features included clinical data, HRV, and waveform features of the IP signal. The results of the different input configurations are presented in **Table 6**. The model trained with all three types of features (IP signal, HRV, and clinical data) achieved the highest AUROC, and it showed a significant difference compared to the models based solely on HRV or IP signals ($p < 0.05$). Furthermore, although it exhibited a higher AUROC trend compared to the combinations of clinical data with HRV and IP

signals, the difference was not statistically significant. The model trained with clinical data shows a tendency towards a higher AUROC of 0.813 (95% CI 0.773-0.853) compared to the models trained with either the HRV, IP signal alone or a combination of the IP signal and HRV features. However, this difference was not statistically significant, as indicated in **Table 7**.

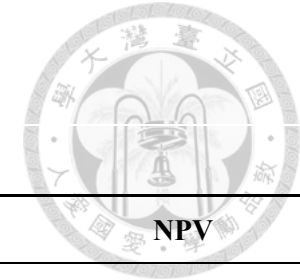


Table 6 Performance of XGBoost classifier with different input

Input type	AUROC	P value ^b	AUPRC	F1-score	Sensitivity	Specificity	PPV	NPV
Whole data ^a	0.877 (0.837-0.917)	reference	0.479 (0.410-0.547)	0.398 (0.290-0.506)	0.333 (0.195-0.471)	0.962 (0.937-0.987)	0.537 (0.399-0.674)	0.922 (0.909-0.936)
Clinical	0.813 (0.773-0.853)	0.185	0.495 (0.421-0.570)	0.298 (0.146-0.449)	0.222 (0.084-0.360)	0.973 (0.946-1.000)	0.582 (0.230-0.934)	0.912 (0.898-0.925)
IP & HRV	0.772 (0.683-0.861)	0.173	0.375 (0.222-0.528)	0.387 (0.265-0.509)	0.489 (0.332-0.646)	0.873 (0.829-0.917)	0.324 (0.214-0.434)	0.934 (0.914-0.953)
IP signal	0.712 (0.622-0.802)	0.011	0.243 (0.149-0.336)	0.224 (0.078-0.370)	0.289 (0.080-0.498)	0.849 (0.835-0.863)	0.183 (0.071-0.296)	0.908 (0.882-0.934)
HRV	0.678 (0.583-0.773)	0.016	0.240 (0.132-0.347)	0.279 (0.196-0.362)	0.467 (0.316-0.618)	0.773 (0.748-0.798)	0.200 (0.142-0.257)	0.923 (0.902-0.944)

IP impedance pneumography; HRV heart rate variability; PPV positive predictive value; NPV negative predictive value; AUC area under receiver-operator curve

^a Whole data were composed of clinical data, IP signal and HRV

^b The P value is obtained by comparing the AUROC of individual model with the reference model through the Delong test.



Table 7 DeLong test to compare AUROC of different input in XGBoost

Input type (AUROC)	Whole (0.877)	Clinical (0.813)	IP & HRV (0.772)	IP (0.712)	HRV (0.678)			
Whole	\							
Clinical						0.185		
IP & HRV						0.173	0.915	
IP						0.011	0.383	0.337
HRV						0.016	0.301	0.116

AUROC area under receiver-operator curve; IP impedance pneumography; HRV heart rate variability

Chapter 4 Discussion

The results of our study indicated the potential of ML models in predicting PF among critically ill patients receiving MV in the ICU. Importantly, this study is the first to extract features from the IP signal and incorporate them alongside clinical data and HRV for model construction. Our findings show that the XGBoost classifier performed the best in terms of AUROC, AUPRC, and F1-score. This indicates its superior ability to accurately predict PF compared to other models tested in this study. Furthermore, the inclusion of IP signal and HRV features in addition to clinical data significantly improved the predictive performance of the model. This highlights the importance of considering respiratory signals and autonomic status when developing predictive models for extubation outcomes.

While previous studies have explored the correlation between respiratory signals and extubation outcomes, they have primarily focused on parameters such as interbreath interval, tidal volume, and peak inspiratory flow [22-24]. However, none of these studies have specifically investigated the IP signal waveform itself. The IP signal has been shown to have a strong correlation with tidal volumes measured by pneumotachometry, which is considered the gold standard method [17, 53]. In our study, we hypothesized that the IP signal could provide valuable insights into a patient's breathing pattern, especially during the SBT when the patient relies partially on their own respiratory effort. To extract features from the IP signal, we identified three quartiles within the waveform for both exhalation and inhalation. We then examined several parameters related to the slope of each quartile, including the mean value, coefficient of variation (CoV), and slope ratio. By analyzing these IP signal features, we aimed to capture the characteristics of a patient's breathing and identify potential

indicators of PF. This approach appeared to add a novel dimension to prediction models of PF.

Previous studies have utilized AI models to predict PF within different time frames. Otaguro et al. and Hsieh et al. developed models to predict PF within 72 hours after extubation [26, 54]. In our study, we defined PF as either resuming MV or death within 48 hours, following the criteria set by the International Consensus Conference [32]. Comparing our results to a study conducted by Chen et al. in the MIMIC-III database, there are some notable differences in the AUROC values achieved by the XGBoost classifier. Chen et al. reported AUROC values ranging from 0.81 to 0.82 using clinical data alone for predicting PF within 48 hours [28]. There were several differences, including (1) our study had specific requirements regarding waveform quality, which led to the selection of a limited and different cohort compared to Chen et al.; (2) to align with the timing of waveform analysis, clinical data in our study were collected at an earlier time point before extubation, which may not accurately represent the actual extubation conditions; (3) our study employed a more stringent threshold for missing data, excluding features with a missing rate exceeding 20%. In contrast, Chen et al. removed features with a missing rate above 40%. Despite the differences, our study obtained a similar AUROC value of 0.813 (95% CI: 0.773-0.853) when using the XGBoost classifier with clinical data.

Combining clinical data, IP signal, and HRV significantly improved the model's performance (**Table 6**). Models trained with only IP signal or HRV alone had the lowest AUROC values of 0.712 and 0.678, respectively, but the difference was not significant (**Table 7**). The model trained with clinical data alone achieved a slightly higher AUROC value of 0.813, but the improvement was not statistically significant compared to the other models. By analyzing the feature importance using SHAP values (**Figure 10B**)

revealed that a longer duration of MV was positively correlated with PF, consistent with findings from previous studies [28, 29, 54]. Additionally, eight out of the top twenty important features of the XGBoost classifier were related to IP signal, with most of them being slope parameters during inspiration (**Figure 10A**). This suggests that the features of IP signal might provide crucial information for predicting PF.

4.1 Clinical Application

If external validation is further conducted to confirm the model's generalizability, the model could be integrated into ICU or other critical care settings for real-time prediction the risk of PF. However, we may encounter several challenges. First, extracting waveform data from the device itself can be difficult. Current clinical monitoring information, though collected at a central monitoring station, cannot be exported, and data older than 24 hours is abandoned. To store or export data, we need further cooperation with the device manufacturers or purchasing a larger scale information platform. Second, in clinical situations, there may be instances where patients are unable to cooperate with the measurement of vital signs. For example, continuous shivering, severe dyspnea causing excessive body movement, or a confused state of consciousness leading to constant posture changes. These situations can cause a lot of interference in the waveform, making analysis impossible. Lastly, integrating clinical and waveform data into an AI model requires a cross-team effort.

Despite the challenges, due to the widely use of IP signals and ECG as part of monitor for vital signs in critical care units (such as ICU and ED) and even pre-hospital settings, it can be considered to expand its application to other areas, including identifying specific respiratory patterns of diseases such as chronic obstructive

pulmonary disease, metabolic acidosis, and acute respiratory failure.



4.2 Limitation

Our study has several limitations that need to be acknowledged. Firstly, while previous research has shown that the inspiration and expiration phases of the IP signal demonstrate good consistency with the actual airway flow, it is important to note that the amplitude of the IP signal can be influenced by factors such as the location and distance of the electrodes [55]. As a retrospective study, we were unable to control or predefine the placement of the electrodes used to capture the IP signal. To address this limitation, we standardized the IP signal to minimize the potential impact of electrode positioning on the signal amplitude. However, through this process, the information about the amplitude of the waveform itself may be attenuated. Secondly, our study only included patients undergoing SBT with PSV mode and excluded those with CPAP mode or T-piece trials. This decision limits the generalizability of our findings to patients in specific ventilatory modes. Including these additional modes in future studies would be valuable to assess their impact on prediction models. Thirdly, the selection criteria used to obtain high-quality IP segment in our study may have resulted in the exclusion of some meaningful IP features, potentially limiting the comprehensiveness of our analysis. However, if we did not utilize these filtering criteria, it would lead to the inclusion of many waveforms with interference, thereby masking the true respiratory characteristics. This also emphasizes the need for future studies to compare the impact of different filtering conditions on the extraction of feature values. Finally, our study solely relied on data from the MIMIC-III database. External validation using independent datasets is crucial to evaluate the generalizability of our model across different settings and

populations. Such validation studies would provide a more robust assessment of the performance and clinical utility.



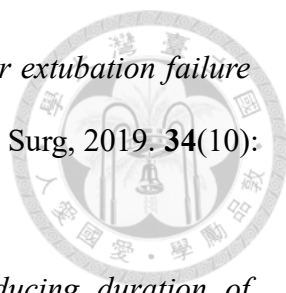
Chapter 5 Conclusion

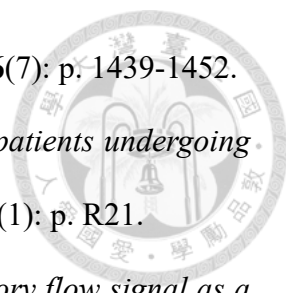
This study utilized the MIMIC-III and MIMIC-III Waveform Database Matched Subset to develop ML models for predicting PF within 48 hours in MV patients. By incorporating IP signal and HRV along with clinical data, the predictive performance of the model was significantly improved, indicating the potential value of these additional features. However, it is important to validate these findings in different clinical contexts through further research.



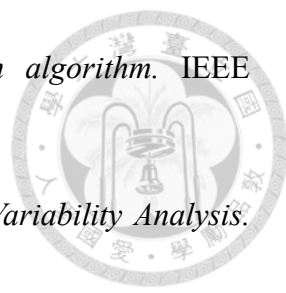
Reference


1. Lone, N.I. and T.S. Walsh, *Prolonged mechanical ventilation in critically ill patients: epidemiology, outcomes and modelling the potential cost consequences of establishing a regional weaning unit*. Critical Care, 2011. **15**: p. 1-10.
2. Zilberberg, M.D., et al., *Characteristics, Hospital Course, and Outcomes of Patients Requiring Prolonged Acute Versus Short-Term Mechanical Ventilation in the United States, 2014–2018**. Critical Care Medicine, 2020. **48**(11): p. 1587-1594.
3. Cabello, B., et al., *Physiological comparison of three spontaneous breathing trials in difficult-to-wean patients*. Intensive Care Med, 2010. **36**(7): p. 1171-9.
4. dos Santos Bien, U., et al., *Maximum inspiratory pressure and rapid shallow breathing index as predictors of successful ventilator weaning*. Journal of physical therapy science, 2015. **27**(12): p. 3723-3727.
5. Duan, J., X. Zhang, and J. Song, *Predictive power of extubation failure diagnosed by cough strength: a systematic review and meta-analysis*. Crit Care, 2021. **25**(1): p. 357.
6. Jaber, S., et al., *Risk factors and outcomes for airway failure versus non-airway failure in the intensive care unit: a multicenter observational study of 1514 extubation procedures*. Crit Care, 2018. **22**(1): p. 236.
7. Thille, A.W., et al., *Risk Factors for and Prediction by Caregivers of Extubation Failure in ICU Patients: A Prospective Study**. Critical Care Medicine, 2015. **43**(3): p. 613-620.
8. Chung, W.C., et al., *Novel mechanical ventilator weaning predictive model*. Kaohsiung J Med Sci, 2020. **36**(10): p. 841-849.

- 
9. Xie, J., et al., *To extubate or not to extubate: Risk factors for extubation failure and deterioration with further mechanical ventilation*. J Card Surg, 2019. **34**(10): p. 1004-1011.
 10. Blackwood, B., et al., *Use of weaning protocols for reducing duration of mechanical ventilation in critically ill adult patients: Cochrane systematic review and meta-analysis*. BMJ, 2011. **342**: p. c7237.
 11. Thille, A.W., et al., *Outcomes of extubation failure in medical intensive care unit patients*. Crit Care Med, 2011. **39**(12): p. 2612-8.
 12. Shaffer, F. and J.P. Ginsberg, *An Overview of Heart Rate Variability Metrics and Norms*. Front Public Health, 2017. **5**: p. 258.
 13. Tiainen, M., et al., *Arrhythmias and heart rate variability during and after therapeutic hypothermia for cardiac arrest*. Crit Care Med, 2009. **37**(2): p. 403-9.
 14. Ranard, B.L., et al., *Heart rate variability and adrenal size provide clues to sudden cardiac death in hospitalized COVID-19 patients*. J Crit Care, 2022. **71**: p. 154114.
 15. Gupta, A.K. *Respiration Rate Measurement Based on Impedance Pneumography*. 2011.
 16. Seppä, V.-P., et al., *Impedance pneumography for assessment of a tidal breathing parameter in patients with airway obstruction*. European Respiratory Journal, 2011. **38**(Suppl 55): p. p1199.
 17. Seppä, V.P., J. Viik, and J. Hyttinen, *Assessment of pulmonary flow using impedance pneumography*. IEEE Trans Biomed Eng, 2010. **57**(9): p. 2277-85.
 18. Arcentales, A., et al., *Classification of patients undergoing weaning from mechanical ventilation using the coherence between heart rate variability and*

- 
- respiratory flow signal*. *Physiological Measurement*, 2015. **36**(7): p. 1439-1452.
19. Huang, C.-T., et al., *Application of heart-rate variability in patients undergoing weaning from mechanical ventilation*. *Critical Care*, 2014. **18**(1): p. R21.
 20. Chaparro, J.A. and B.F. Giraldo. *Power index of the inspiratory flow signal as a predictor of weaning in intensive care units*. in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 2014. IEEE.
 21. Seely, A.J., et al., *Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients?* *Critical Care*, 2014. **18**(2): p. 1-12.
 22. Arcentales, A., et al., *Classification of patients undergoing weaning from mechanical ventilation using the coherence between heart rate variability and respiratory flow signal*. *Physiol Meas*, 2015. **36**(7): p. 1439-52.
 23. Seely, A.J., et al., *Do heart and respiratory rate variability improve prediction of extubation outcomes in critically ill patients?* *Crit Care*, 2014. **18**(2): p. 1-12.
 24. Bien, M.Y., et al., *Comparisons of predictive performance of breathing pattern variability measured during T-piece, automatic tube compensation, and pressure support ventilation for weaning intensive care unit patients from mechanical ventilation*. *Crit Care Med*, 2011. **39**(10): p. 2253-62.
 25. Engoren, M. and J.M. Blum, *A comparison of the rapid shallow breathing index and complexity measures during spontaneous breathing trials after cardiac surgery*. *J Crit Care*, 2013. **28**(1): p. 69-76.
 26. Hsieh, M.H., et al., *An Artificial Neural Network Model for Predicting Successful Extubation in Intensive Care Units*. *J Clin Med*, 2018. **7**(9).
 27. Kuo, H.J., et al., *Improvement in the Prediction of Ventilator Weaning Outcomes by an Artificial Neural Network in a Medical ICU*. *Respir Care*, 2015. **60**(11): p.

- 1560-9.
28. Chen, T., et al., *Prediction of Extubation Failure for Intensive Care Unit Patients Using Light Gradient Boosting Machine*. IEEE Access, 2019. 7: p. 150960-150968.
29. Zhao, Q.Y., et al., *Development and Validation of a Machine-Learning Model for Prediction of Extubation Failure in Intensive Care Units*. Front Med (Lausanne), 2021. 8: p. 676343.
30. Johnson, A.E., et al., *MIMIC-III, a freely accessible critical care database*. Sci Data, 2016. 3: p. 160035.
31. Goldberger, A.L., et al., *PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals*. circulation, 2000. 101(23): p. e215-e220.
32. Boles, J.M., et al., *Weaning from mechanical ventilation*. Eur Respir J, 2007. 29(5): p. 1033-56.
33. Schafer, A. and K.W. Kratky, *Estimation of breathing rate from respiratory sinus arrhythmia: comparison of various methods*. Ann Biomed Eng, 2008. 36(3): p. 476-85.
34. Charlton, P.H., et al., *An impedance pneumography signal quality index: Design, assessment and application to respiratory rate monitoring*. Biomed Signal Process Control, 2021. 65: p. 102339.
35. Sedghamiz, H., *Complete Pan Tompkins Implementation ECG QRS detector*. March 2014, MATLAB Central File Exchange. Available online: <https://www.mathworks.com/matlabcentral/fileexchange/45840-complete-pan-tompkins-implementation-ecg-qrs-detector> Retrieved (Retrieved February 20, 2023).

- 
36. Pan, J. and W.J. Tompkins, *A real-time QRS detection algorithm*. IEEE transactions on biomedical engineering, 1985(3): p. 230-236.
37. Zhao, L., et al., *Influence of Ectopic Beats on Heart Rate Variability Analysis*. Entropy (Basel), 2021. **23**(6).
38. Catai, A.M., et al., *Heart rate variability: are you using it properly? Standardisation checklist of procedures*. Braz J Phys Ther, 2020. **24**(2): p. 91-102.
39. Electrophysiology, T.F.o.t.E.S.o.C.t.N.A.S.o.P., *Heart Rate Variability*. Circulation, 1996. **93**(5): p. 1043-1065.
40. Clifford, G.D., P.E. McSharry, and L. Tarassenko. *Characterizing artefact in the normal human 24-hour RR time series to aid identification and artificial replication of circadian variations in human beat to beat heart rate using a simple threshold*. in *Computers in Cardiology*. 2002.
41. McKinley, S. and M. Levine, *Cubic spline interpolation*. College of the Redwoods, 1998. **45**(1): p. 1049-1060.
42. Guyon, I. and A. Elisseeff, *An introduction to variable and feature selection*. Journal of machine learning research, 2003. **3**(Mar): p. 1157-1182.
43. Chawla, N.V., et al., *SMOTE: synthetic minority over-sampling technique*. Journal of artificial intelligence research, 2002. **16**: p. 321-357.
44. Chen, T. and C. Guestrin. *Xgboost: A scalable tree boosting system*. in *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*. 2016.
45. Prokhorenkova, L., et al., *CatBoost: unbiased boosting with categorical features*. Advances in neural information processing systems, 2018. **31**.
46. Ke, G., et al., *Lightgbm: A highly efficient gradient boosting decision tree*.

- 
- Advances in neural information processing systems, 2017. **30**.
47. Breiman, L., *Random forests*. Machine learning, 2001. **45**: p. 5-32.
 48. Hosmer Jr, D.W., S. Lemeshow, and R.X. Sturdivant, *Applied logistic regression*. 3rd ed. Vol. 398. 2013, Hoboken, New Jersey: John Wiley & Sons.
 49. DeLong, E.R., D.M. DeLong, and D.L. Clarke-Pearson, *Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach*. Biometrics, 1988: p. 837-845.
 50. Sun, X. and W. Xu, *Fast Implementation of DeLong's Algorithm for Comparing the Areas Under Correlated Receiver Operating Characteristic Curves*. IEEE Signal Processing Letters, 2014. **21**(11): p. 1389-1393.
 51. Lundberg, S.M. and S.-I. Lee, *A unified approach to interpreting model predictions*. Advances in neural information processing systems, 2017. **30**.
 52. Bisong, E. and E. Bisong, *Google colabatory. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners*, 2019: p. 59-64.
 53. Mlynczak, M., et al., *Assessment of calibration methods on impedance pneumography accuracy*. Biomed Tech (Berl), 2016. **61**(6): p. 587-593.
 54. Otaguro, T., et al., *Machine Learning for Prediction of Successful Extubation of Mechanical Ventilated Patients in an Intensive Care Unit: A Retrospective Observational Study*. J Nippon Med Sch, 2021. **88**(5): p. 408-417.
 55. Wang, H.-b., et al., *A Robust Electrode Configuration for Bioimpedance Measurement of Respiration*. Journal of Healthcare Engineering, 2014. **5**: p. 696103.

Appendix



List of abbreviations

MV	Mechanical ventilation
IP	Impedance pneumography
HRV	Heart rate variability
AI	Artificial intelligence
ML	Machine learning
MIMIC-III	Medical Information Mart for Intensive Care III
ECG	Electrocardiogram
PF	Post-extubation failure
AUROC	Area under the receiver operating characteristic curve
ICU	Intensive care unit
SBT	Spontaneous breathing trial
RRV	Respiratory rate variability
NIPPV	Noninvasive positive pressure ventilation
HFNC	High flow nasal cannula
PIP	Peak inspiratory pressure
PSV	Pressure support ventilation
RSBI	Rapid shallow breathing index
PaO ₂	Partial pressure of oxygen
FiO ₂	Inspired fraction of oxygen
PaCO ₂	Partial pressure of carbon dioxide
BE	Base excess



PEEP	Positive end-expiratory pressure
RR	Respiratory rate
CoV	Coefficient of variation
TP	Total power
VLF	Very low frequency
LF	Low frequency
HF	High frequency
RRi	R-peak to R-peak interval
SDNN	Standard deviation of individual normal-to-normal intervals
	Root mean square of successive differences between normal
RMSSD	heartbeats
pNN50	Proportion of NN50 divided by total number of NNs
SD1	Standard deviation of the first Poincaré plot axis
SD2	Standard deviation of the second Poincaré plot axis
SMOTE	Synthetic Minority Oversampling Technique
AUPRC	Area under the precision recall curve
XGBoost	Extreme Gradient Boosting
LightGBM	Light Gradient Boosting Machine
CATBoost	CATegorical Boosting
RF	Random Forest
LR	Logistic Regression
SHAP	SHapley Additive exPlanations



Ethics approval and consent to participate

The Institutional Review Boards of Beth Israel Deaconess Medical Center (Boston, MA) and the Massachusetts Institute of Technology (Cambridge, MA) approved the establishment of the MIMIC-III database. In addition, this study was approved by the Institutional Review Board of the China Medical University (CMUH112-REC1-073). Since the project did not affect clinical care and all protected health information was deidentified, the need for individual patient consent was waived.

Availability of data and materials

The MIMIC-III data were available on the project website at <https://physionet.org/content/mimiciii/1.4/>, and data of MIMIC-III Waveform Database Matched Subset were available on the project website at <https://physionet.org/content/mimic3wdb-matched/1.0/>.

Competing interests

The authors declare that they have no competing interests.

Funding

No funding received in this study.