國立臺灣大學工學院應用力學研究所

碩士論文

Institute of Applied Mechanics

College of Engineering

National Taiwan University

Master's Thesis

利用不同生理條件對腦血流自動調節進行分類

Classification of dynamic cerebral autoregulation using different physiological conditions.

李韋皓

Wei-Hao Li

指導教授：潘斯文 博士

Advisor : Stephen Payne, Ph.D.

中華民國 113年 7月

July, 2024

doi:10.6342/NTU202403444

Acknowledgements

I would like to express my deepest gratitude to my advisor, Professor Stephen J. Payne, for his invaluable guidance, support, and encouragement throughout the course of this research. His expertise and insights have been instrumental in shaping this work. I am also grateful to the Institute of Applied Mechanics at National Taiwan University for providing the resources and environment necessary for conducting this study. The collaborative and stimulating atmosphere at the institute has greatly contributed to the successful completion of this research.

doi:10.6342/NTU202403444

中文摘要

本研究旨在開發和評估一個二元分類器，以根據受試者的血壓（BP）和腦血流速度（CBFV）數據來判斷其健康狀態。我們使用經顱多普勒超聲波（TCD）和轉移函數分析（TFA）測量了受試者在基線量測、高碳酸血症量測和大腿袖帶測試條件下的BP和CBFV數據。在分類器的開發過程中，我們使用支持向量機（SVM）對數據集進行訓練和測試，並分析了分類器的性能指標和特徵貢獻。我們的研究結果表明，使用基線量測和高碳酸血症量測訓練的分類器相比於大腿袖帶測試和高碳酸血症量測訓練的分類器，展示出了更高的準確性，這突顯了在這些狀態下BP和CBFV關係的穩定性和可預測性。

關鍵字： 腦血流自動調節; 經顱多普勒超聲波；轉移函數分析；機器學習；支持向量機

doi:10.6342/NTU202403444

Abstract

The brain's ability to maintain stable cerebral blood flow (CBF) despite fluctuations in blood pressure (BP) is crucial for preventing damage and ensuring normal brain function. This study aims to develop and evaluate a binary classifier to determine the health status of subjects based on their blood pressure (BP) and cerebral blood flow velocity (CBFV) data, with the assumption that subjects can be classified as either baseline or impaired. Using Transcranial Doppler (TCD) ultrasound and Transfer Function Analysis (TFA), we measured BP and CBFV under normocapnia, hypercapnia, and thigh cuff testing conditions. For classifier development, we trained and tested the dataset using Support Vector Machine (SVM) and analyzed the performance metrics and feature contributions of the classifier. Our findings indicate that classifiers trained under normocapnia and hypercapnia conditions demonstrate superior accuracy compared to those trained under thigh cuff testing conditions, highlighting the stability and predictability of BP and CBFV relationships in these states.

Keywords: Dynamic Cerebral Autoregulation, Transcranial Doppler, Transfer Function Analysis, Machine Learning, Support Vector Machine, SHapley Additive exPlanations.

doi:10.6342/NTU202403444

Contents

doi:10.6342/NTU202403444

doi:10.6342/NTU202403444

List of Figures

doi:10.6342/NTU202403444

Chapter 1

Introduction

Dynamic Cerebral Autoregulation (dCA) refers to the brain's ability to maintain relatively stable cerebral blood flow (CBF) despite fluctuations in blood pressure (BP). This mechanism is crucial for maintaining normal brain function and preventing damage. dCA involves multiple physiological processes, including chemoregulation, autoregulation, and neurovascular coupling, which together ensure that the brain meets its metabolic needs[1]. By measuring BP and CBF under different physiological conditions, researchers can assess the effectiveness and stability of dCA.

Common techniques for measuring dCA include the Valsalva maneuver, squat-to-stand, sit-to-stand, and thigh cuff deflation, all of which induce BP changes to observe their effects on CBF[2]. Transcranial Doppler (TCD) ultrasound, a non-invasive method, measures blood flow velocity in major intracranial arteries, providing high temporal resolution CBFV data crucial for dCA research[3]. For accurate BP measurement, arterial pressure monitoring and volume clamping methods are used. Volume clamping involves adjusting cuff pressure based on the pulse waveform's systolic and

doi:10.6342/NTU202403444

diastolic phases, yielding high-resolution BP data. When combined with TCD, this data helps elucidate the BP-CBFV relationship[4].

Regarding how to process the measured data, we employ the commonly used Transfer Function Analysis (TFA) method, which quantifies the dynamic relationship between BP and CBF by calculating parameters such as gain, phase, and coherence[1]. To further enhance the analysis, recent medical research increasingly integrates machine learning classifiers, which can analyze and predict health status. These classifiers are widely applied in fields like cardiovascular disease risk prediction[6], tumuor diagnosis[7], and neurological disorder early warning[8]. In dCA research, machine learning classifiers show significant potential by utilizing BP and cerebral blood flow velocity (CBFV) data to distinguish between baseline and impaired states[5].

To obtain the data required for developing our classifiers, we utilized pre-existing data collected by other researchers. These data were obtained following the recommendations from the CARNet white paper, which involved five measurements on 20 healthy subjects using TCD under normocapnia, hypercapnia, and thigh cuff testing conditions[1]. Next, we processed the data to eliminate noise and artifacts, and

doi:10.6342/NTU202403444

then interpolated and filtered the data to ensure continuity and accuracy. The processed signals were analyzed using Transfer Function Analysis (TFA) to obtain parameters such as gain, phase, and coherence[1,2]. Finally, to improve the consistency and comparability of research results[1], we standardized the data for classifier training.

doi:10.6342/NTU202403444

Chapter 2

Methods

In this study, our objective is to develop and evaluate a binary classifier to determine the health status of subjects based on their BP and CBFV data, with a focus on using classifiers between the normocapnia/ hypercapnia and thigh cuff testing/ hypercapnia conditions.

To achieve this, we used existing datasets and adopted standardized data processing and analysis methods to ensure the accuracy and reliability of the data. Following the recommendations of the CARNet white paper[10], we used TCD ultrasound technology for data collection and applied TFA to quantify the dynamic relationship between BP and CBFV.

In the following sections, we will provide a detailed description of the data collection, preprocessing, and analysis steps. Subsequently, we trained the classifier using the collected data and evaluated its performance metrics. Finally, we analyzed the variables to elucidate their relationship with dCA.

doi:10.6342/NTU202403444

2.1 Data Acquisition

To collect data for training and evaluating the classifier, we designed and implemented a series of rigorous experimental procedures. We selected TCD technology as the primary data collection tool and, based on CARNet white paper recommendations, adopted various methods to induce BP changes, including normocapnia, hypercapnia, and thigh cuff testing. The detailed steps of the experimental design, subject recruitment, data collection, and experimental procedures are described below.

In this experiment, we used existing data, which were obtained as follows: First, by measuring the volume control of the digital artery, we used arterial volume clamping devices to adjust cuff pressure in response to changes in arterial volume during systole and diastole. This method provides high-precision BP data[2,3]. Simultaneously, we recorded CBFV data using TCD technology as a proxy for direct CBF measurement. These data were recorded in cm/s[4].

It is important to note that the gain, phase, and coherence measured by different devices may vary. To ensure data consistency and comparability, we followed the standardized equipment and methods recommended in the CARNet white papers[9,10]. Finally, during

doi:10.6342/NTU202403444

TFA, spontaneous BP and CBFV fluctuations should be recorded for at least 5 minutes to ensure sufficient data for accurate frequency domain analysis[11]. These steps ensure the accuracy and consistency of experimental data, providing a reliable foundation for subsequent data processing and classifier training.

In the data preparation and preprocessing phase, we followed recommendations to ensure data quality and reliability. The minimum recommended sampling frequency is 50Hz, twice the maximum signal frequency (Nyquist theorem). Researchers however often use sampling frequencies 4-5 times higher to improve accuracy[11]. We processed raw waveforms and beat-to-beat data, with beat-to-beat data averaging BP and CBFV per pulse. While both formats correlate well, beat-to-beat data is preferred for TFA due to its lower sensitivity to interference. The method involves using the diastolic period of BP as cycle endpoints to calculate average BP and CBFV through waveform integration[2,3]. Before analysis, signals were visually inspected for noise. Short-term interference was corrected with linear interpolation; longer-lasting interference led to data exclusion[4].

Averaging BP and CBFV per pulse using the beat-to-beat method yields several data

doi:10.6342/NTU202403444

points. We used cubic polynomial interpolation to create equally spaced data points, ensuring a minimum frequency of 4Hz after resampling to prevent high heart rate interference[9]. Detrending does not affect TFA results and helps reduce very low frequency (VLF) power, improving accuracy[10]. Data normalization is recommended to reduce individual variability and enhance analysis consistency; filtering methods are avoided as they may alter signal characteristics. These steps ensure high-quality, reliable data for TFA.

Conclusion. In the data acquisition section, we designed and implemented a series of rigorous experimental procedures using Transcranial Doppler (TCD) ultrasound technology to collect blood pressure (BP) and cerebral blood flow velocity (CBFV) data. By following standardized methods for data processing and preprocessing, we ensured the accuracy and reliability of the collected data. These steps provided a solid foundation for subsequent analysis and classifier development.

doi:10.6342/NTU202403444

2.2 Transfer Function Analysis

Transfer Function Analysis (TFA) is a method to analyze the relationship between BP and CBFV[1]. The preprocessed standardized beat-to-beat BP and CBFV data are time-domain signals. First, both are transformed into frequency-domain signals using Fast Fourier Transform (FFT) and then subjected to cross-spectral analysis to calculate gain, phase, and coherence. Detailed explanations are provided below.

Usually, during spectral analysis, FFT is used to convert time-domain signals into frequency-domain signals. However, the coefficient of variation (CoV = standard deviation/mean) of a single FFT of the complete time-domain signal (single window) is about 1, indicating that using the entire measurement signal for FFT will result in unstable and unreliable outcomes. Therefore, the Welch method is needed to improve analysis accuracy[12].

The Welch method is an improved spectral estimation method designed to reduce spectral leakage and improve estimation reliability. Its main principle is to segment the signal, apply windowing to each segment, perform Fourier transform, and then

doi:10.6342/NTU202403444

average the results of each segment. This method reduces random errors introduced by a single FFT. The calculation formula of the Welch method is as follows:

$$P_{xx}(f) = \frac{1}{L}\sum_{i=0}^{L-1}\frac{1}{U}\left|\sum_{n=0}^{N-1} x_i(n)\omega(n)e^{-j\frac{2\pi fn}{N}}\right|^2 \quad (1)$$

Where L is the number of segments, U is the energy of the window function, $x_i(n)$ is the signal of the i -th segment, and is the window function[13,14].

According to standardized procedure recommendations[1], when using the Welch method, too short a window length will lead to insufficient frequency resolution; typically, a window length of more than 100 seconds is used, and if the total recording time exceeds 5 minutes, the window length should not be increased but rather the number of windows. Windows are not aligned side by side; some overlap increases smoothness. Previous studies indicate that a 50% overlap is most commonly used. Spectral leakage occurs in spectral analysis when the signal is truncated in a finite observation window, producing unreal spectral components.

To reduce spectral leakage, window functions such as the Hanning window (recommended), Hamming window, or Tukey window, or increasing signal length

doi:10.6342/NTU202403444

are applied[2]. The Hanning window function gradually reduces the signal amplitude at both ends, and its formula is as follows:

$$\omega(n) = 0.5(1 - cos\ (2\pi n/(N-1))\ ), \quad 0 \le n \le N-1 \tag{2}$$

where $\omega(n)$ is the value of the window function, and N is the window length[1]. By using the Welch method with the Hanning window, we can more accurately quantify frequency domain parameters (gain and phase) and reduce estimation errors due to spectral leakage[16].

After obtaining the frequency spectra of BP and CBFV for each window, we average the windows to get the frequency spectra instead of using a single FFT. After obtaining the frequency-domain signals of BP (($P(f)$)) and CBFV ($F(f)$), we assume their relationship is linear and calculate their gain, phase, and coherence directly:

$$S_{PP}(f) = P(f) \cdot P^*(f) \tag{3}$$

$$S_{FF}(f) = F(f) \cdot F^*(f) \tag{4}$$

$$S_{PF}(f) = P(f) \cdot F^*(f) \tag{5}$$

$$Gain(f) = \left| \frac{S_{PF}(f)}{S_{PP}(f)} \right| \tag{6}$$

doi:10.6342/NTU202403444

$$Phase(f) = arg\big(S_{PF}(f)\big) \quad (7)$$

$$Coh(f) = \frac{|S_{PF}(f)|^2}{S_{PP}(f) \cdot S_{FF}(f)} \quad (8)$$

The coherence function is a dimensionless indicator that effectively measures the reliability of these estimates. Checking if the coherence value exceeds the 95% confidence limit (i.e., the 5% critical value) for zero coherence is a simple method to assess the validity of gain and phase. If coherence remains low (insignificant) across all frequency bands, the record should be excluded from analysis due to poor data quality and unreliable results.

Finally, we segment the data by frequency; commonly used frequency bands include vlf (0.02-0.07Hz), lf (0.07-0.2Hz), and hf (0.2-0.5Hz). For standardization purposes, we set these as the segmentation points to distinguish the effects of different frequencies on dCA.

Conclusion. In the transfer function analysis section, we applied the Welch method to convert time-domain signals into frequency-domain signals and calculated the

doi:10.6342/NTU202403444

gain, phase, and coherence between BP and CBFV. This method reduced spectral leakage and improved the reliability of the analysis. By segmenting data by frequency bands, we were able to better understand the dynamic relationship between BP and CBFV under different physiological conditions.

doi:10.6342/NTU202403444

2.3 Classifier

In machine learning, classifiers are essential algorithms for predicting category labels based on input features, widely used in image recognition, text classification, and medical diagnosis. Classifier design and application involve supervised learning, which trains models using labeled data to make accurate predictions on new data. The model iteratively adjusts parameters to minimize prediction errors. A well-trained model effectively classifies new data to meet application goals. Model selection aims to choose the optimal algorithm with the best classification capability, evaluated by excess loss.

In this section, we utilize post-TFA data and labels to establish a binary classifier to determine whether subjects are in an impaired dCA state. We use the Python programming language and employ its powerful machine learning library, scikit-learn, for data analysis and model training[53].

Due to the advantages of handling small sample datasets and high-dimensional data, we chose Support Vector Machines (SVM) as the classification algorithm[18]. Our dataset has more samples than features, but the numbers are relatively close, making SVM a

doi:10.6342/NTU202403444

suitable choice. SVM is particularly effective in situations where the number of dimensions approaches the number of samples. Additionally, SVM has a strong theoretical foundation and provides robust performance even when the data are not linearly separable by using kernel tricks to transform the original feature space into a higher-dimensional space. This flexibility allows SVM to model complex relationships in the data[19].

The basic idea of SVM is to find an optimal hyperplane in high-dimensional space to separate data points into different categories. This optimal hyperplane maximizes the margin between two categories, thereby improving classification accuracy and stability[19]. The goal of SVM is to find a separating hyperplane such that the distance between the hyperplane and the nearest data points (support vectors) is maximized. The mathematical formulation is as follows:

Given a set of training data (x_i, y_i)，where $x_i \in R^n$，$y_i \in \{1, -1\}$，SVM solves the following optimization problem:

$$min_{w,b} \frac{1}{2} \|w\|^2 \quad (9)$$

doi:10.6342/NTU202403444

$$subject\ to\ y_i \cdot (w \cdot x_i + b) \geq 1, \forall i$$

where w is the weight vector, and b is the bias term. This optimization can be solved by converting to a dual problem using Lagrange multipliers, ultimately determined by the support vectors[20,21].

SVM can also handle nonlinear classification problems through the kernel function. The basic idea of the kernel function is to map the original data into high-dimensional space so that data can be linearly separated in that space. The mathematical representation of the kernel function is:

$$K(x_i, x_j) = \emptyset(x_i) \cdot \emptyset(x_j) \quad (10)$$

Where K is the kernel function, and $\emptyset$ is the mapping function that maps input data x_i and x_j into high-dimensional space[23].

Common kernel functions include linear, polynomial, and Gaussian (RBF) kernels[22]. In this study, we chose the linear kernel, which mathematical representation is:

$$K(x_i, x_j) = x_i \cdot x_j \quad (11)$$

doi:10.6342/NTU202403444

For several reasons, we chose the linear kernel. It is computationally efficient, making it suitable for datasets with a large number of features. Additionally, the linear kernel is easier to interpret, as the resulting model coefficients can directly indicate the importance of each feature. These advantages make the linear kernel a practical choice for our classification task.

Conclusion. In the classifier development section, we utilized Support Vector Machines (SVM) to build a binary classifier capable of distinguishing between baseline and impaired states based on BP and CBFV data. The use of a linear kernel allowed for efficient computation and interpretability of the model. The classifier's performance was evaluated using various metrics, demonstrating its effectiveness in classifying different physiological states.

doi:10.6342/NTU202403444

2.4 Feature Selection

Feature selection is crucial in machine learning and data mining, as it identifies the most informative features from a large dataset. This process not only improves model accuracy but also reduces computational costs and boosts interpretability[26]. Eliminating redundant features helps to minimize overfitting and improve the model's generalization capabilities[27]. A reduced feature set also lowers computational demands, which accelerates both training and prediction times[28]. Moreover, a simplified feature set renders the model more comprehensible and interpretable, emphasizing the most significant factors[29]. In the context of high-dimensional datasets, feature selection helps to overcome the "curse of dimensionality," thereby increasing processing efficiency[30].

Feature selection methods include filter, wrapper, embedded, and exhaustive search techniques. Filter methods, such as t-tests and chi-square tests, evaluate the correlation between each feature and the target variable, while mutual information techniques measure the relevance of features to the target variable[18,31,32]. These methods are simple and fast but often ignore feature interactions. Wrapper methods, like genetic algorithms and recursive feature elimination (RFE), iteratively select optimal feature subsets to

doi:10.6342/NTU202403444

improve model performance[33,34,35]. Embedded methods, including decision tree-based selection and regularization techniques like Lasso regression, evaluate feature importance directly during model training[36,37]. Exhaustive search, though computationally expensive, evaluates all feature subsets to find the globally optimal combination, making it suitable for datasets with fewer features[38,39].

Different feature selection techniques have unique advantages and disadvantages, making them suitable for various datasets. Statistical tests are simple but may miss complex feature interactions, while mutual information methods handle nonlinear relationships well but are computationally intensive. Genetic algorithms are ideal for high-dimensional datasets but require longer computation times. Model-based methods like RFE and decision trees consider feature interactions but need more computational resources. Exhaustive search, though computationally expensive, guarantees finding the globally optimal feature combination and is suitable for datasets with fewer features[31,33,38]. Based on our data characteristics and research objectives, this study will use the exhaustive search method to ensure optimal classification performance. Employing multiple feature selection techniques can also enhance robustness[39].

Conclusion. In the feature selection section, we employed an exhaustive search method

doi:10.6342/NTU202403444

to identify the optimal subset of features that contributed the most to the classifier's performance. This approach helped improve the model's accuracy, reduce computational costs, and enhance interpretability. By eliminating redundant features, we minimized the risk of overfitting and improved the generalization capabilities of the classifier.

doi:10.6342/NTU202403444

2.5 Performance Analysis

To evaluate our classifier's performance, we use various indicators and methods to ensure its stability and accuracy. These metrics help identify the model's strengths and weaknesses, guiding improvements. Accurate evaluation is crucial for reliable clinical diagnosis and treatment, revealing subtle differences in dCA mechanisms and enabling personalized medical recommendations.

First, when discussing classifier performance, the confusion matrix is a direct way to display the classification results. In a binary classification task, the confusion matrix includes four parts: True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN)[40]. From this matrix, we can calculate several important performance indicators: Accuracy, which measures the proportion of correctly classified samples; Sensitivity (Recall), which measures the model's ability to detect positive cases; Specificity, which reflects the ability to identify negative cases; and Precision, which measures the proportion of true positives among the classified positive samples. The formulas for these four indicators are as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{12}$$

doi:10.6342/NTU202403444

$$Sensitivity = \frac{TP}{TP + FN} \quad (13)$$

$$Specificity = \frac{TN}{TN + FP} \quad (14)$$

$$Precision = \frac{TP}{TP + FP} \quad (15)$$

Combining the above indicators, we also introduce the F1 Score to provide an overall measure, calculated as the harmonic mean of precision and sensitivity:

$$F1\ Score = 2 \cdot \frac{Precision \cdot Sensitivity}{Precision + Sensitivity} \quad (16)$$

The F1 Score is particularly important when dealing with data imbalance, as it can provide a more balanced evaluation based on precision and sensitivity[41,42].

When training a classifier, it is imperative to divide the dataset into training and testing sets to evaluate its performance accurately. Performance metrics can be significantly affected by the variability in data and the method of partitioning, which often complicates the assessment of the classifier's true performance. Cross-validation is a widely adopted technique to address this issue, enhancing the robustness and reliability of the evaluation process.

doi:10.6342/NTU202403444

The most common cross-validation method is K-Fold Cross-Validation, which divides the dataset into K subsets. Each time, K-1 subsets are used for model training, and the remaining one subset is used for validation, repeated K times and averaged[43]. The mathematical expression is as follows:

$$CV_{(K)} = \frac{1}{K}\sum_{i=1}^{K} E_{(i)} \tag{17}$$

where $E_{(i)}$ is the error of the i -th validation, and CV is the average error of cross-validation.

When the sample size is small, we can use Leave-One-Out Cross-Validation (LOOCV), where each time only one sample is left out as the validation set, and the rest as the training set, repeated M (sample size) times[25]. The expression is:

$$CV_{(LOO)} = \frac{1}{M}\sum_{i=1}^{M} E_{(i)} \tag{18}$$

LOOCV is an extreme form of K-Fold Cross-Validation, and its advantage is that the model training uses all the data, making the estimation results unbiased and stable.

doi:10.6342/NTU202403444

Other cross-validation methods include Repeated Random Subsampling Cross-Validation and Stratified K-Fold Cross-Validation, which help improve model performance on imbalanced datasets[44].

After evaluating the classifier's performance using cross-validation and various metrics, it is crucial to understand the classifier's behavior at different threshold levels. To achieve this, we can utilize the Receiver Operating Characteristic (ROC) curve and the Area Under the Curve (AUC).

ROC is a widely used tool for evaluating classifier performance. By plotting Sensitivity against 1-Specificity, the ROC curve shows the classifier's performance at different thresholds. The ROC curve's horizontal axis represents the False Positive Rate, and the vertical axis represents the True Positive Rate[45]. Ideally, the classifier's ROC curve should approach the top left corner, indicating that the classifier can maximize sensitivity while minimizing the false positive rate.

To precisely assess how close the ROC curve is to the top left corner, we calculate the AUC. It ranges from 0.5 to 1, with higher values indicating better discriminatory ability

doi:10.6342/NTU202403444

and values closer to 0.5 suggesting random guessing[46]. A key feature of AUC is its correlation with the classifier's ranking ability, interpreted as the probability that the classifier scores a randomly chosen positive sample higher than a negative one.

ROC and AUC help us understand the classifier's ability to balance sensitivity and specificity: high sensitivity means the classifier can detect more positive cases, but it may accompany a higher false positive rate; high specificity indicates that the classifier can effectively exclude negative cases, reducing false positives. The ROC curve provides an intuitive method to balance these two, helping us find the optimal decision threshold.

SHapley Additive exPlanations (SHAP) is a technique for explaining the output of machine learning models, based on Shapley values from game theory, assigning each feature an importance value to quantify its impact on model predictions[35]. This method unifies various existing feature importance measurement methods and has a series of theoretically desirable properties.

Specifically, SHAP values assign an importance value to each feature, representing its

doi:10.6342/NTU202403444

marginal contribution in all possible feature combinations. The weighted average of these contribution values is the SHAP value for that feature, ensuring consistency and fairness in model interpretation. The calculation method is as follows:

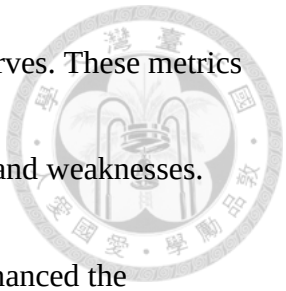
$$\phi_i(f,x) = \sum_{S \subseteq F\{i\}} \frac{|S|!\,(|F|-|S|-1)!}{|F|!}[f(x_{S\cup\{i\}}) - f(x_S)] \quad (19)$$

where S represents the feature subset, F is the set of all features, x_S represents the values of the feature subset S, and f is the model function.

SHAP has the following main properties: Local accuracy, which ensures that the explanation model's prediction equals the original model's prediction; Missingness, where the contribution value for missing features should be zero; and Consistency, which ensures that if the model's dependency on a feature increases, the contribution value for that feature should not decrease. SHAP is very effective in explaining complex model predictions, especially for deep learning and ensemble models[47]. By calculating the SHAP value for each feature, we can understand the contribution of each feature to the prediction, thereby interpreting the model's behavior.

Conclusion. In the performance analysis section, we evaluated the classifier's

doi:10.6342/NTU202403444



performance using cross-validation, confusion matrices, and ROC curves. These metrics provided a comprehensive understanding of the classifier's strengths and weaknesses. The use of SHapley Additive exPlanations (SHAP) values further enhanced the interpretability of the model by quantifying the contribution of each feature to the classification results.

doi:10.6342/NTU202403444

2.6 Optimized Classification Procedure

To date, the principles and performance of the classifiers have been comprehensively introduced. Next, a detailed classification analysis workflow will be established using the data obtained from the aforementioned TFA. The following steps outline the core methodology of this study. To understand the differences in classification capabilities and their causes between the two classifiers generated under three different physiological conditions, we aim to minimize randomness and use various evaluation methods.

First, data normalization is performed to ensure that each variable has consistent units and ranges. Normalization not only enhances the performance and stability of the classifier but also significantly impacts the analysis of variable contributions.

The next step involves using exhaustive search to identify the optimal subset of variables from the dataset. For each possible subset of variables, an SVC is trained using Eq. (9) for the optimization problem and Eq. (11) for the linear kernel. The accuracy of this SVC is calculated using LOOCV as per Eq. (18), representing the score

doi:10.6342/NTU202403444

of this variable subset. After traversing and recording all possible variable subsets, the subset with the highest accuracy is selected as the optimal variable subset.

Once the optimal variable subset is obtained, it is trained again to establish the best classifier for this dataset. Subsequently, performance evaluations are conducted for both the best classifier and the classifier using all variables. The confusion matrix is first calculated, from which Accuracy, Sensitivity, Specificity, Precision, and the F1 Score can be derived using Eq. (12) to (16). Additionally, the ROC curve and AUC value for the best classifier are calculated.

Furthermore, only the best classifier undergoes SHAP value analysis. To avoid the randomness of data partitioning, we use LOOCV as per Eq. (18). During the i-th partitioning of the dataset, the accuracy that would normally be calculated is replaced by the SHAP value computed using Eq. (19). This process yields M SHAP values, which are then averaged to obtain the average SHAP value distribution for the optimal variable subset. This step quantifies the contribution ranking of all variables to the classification performance, reveals the positive/ negative correlation of variables to the outcomes, and finally, a summary plot is generated for visualization.

doi:10.6342/NTU202403444

Conclusion. In the optimized classification procedure section, we established a detailed workflow for data normalization, feature selection, and classifier training. By using leave-one-out cross-validation (LOOCV) and analyzing SHAP values, we ensured the robustness and reliability of the classification results. This procedure allowed us to identify the most discriminative features and develop a highly accurate and interpretable classifier for distinguishing between different physiological conditions.

doi:10.6342/NTU202403444

Chapter 3

Results

The complete workflow of this study is illustrated in Figure 1. We utilized existing data acquired from 20 subjects, each tested five times under three different physiological conditions, focusing on the left and right hemispheres of the brain. The variables included the gain, phase, and coherence across three frequency bands (vlf, lf, hf) for both hemispheres, totaling 18 features. After excluding invalid or missing data, we obtained 97 normocapnia samples, 88 hypercapnia samples, and 98 thigh cuff test samples. The statistical data for these samples are presented in Table 1, the scatter plots in Figure 2, and the box plots in Figure 3. Our analysis shows that the classifiers trained under normocapnia and hypercapnia conditions performed better than those trained under thigh cuff testing conditions.

Subsequently, we used normocapnia as the baseline label and hypercapnia as the impaired label. The data underwent the Optimized Classification Procedure to identify the optimal subset of variables and the best classifier for performance analysis. The same process was repeated with normocapnia replaced by thigh cuff testing data. The

doi:10.6342/NTU202403444

performance metrics for both classifiers are listed in Table 2, the ROC curves are shown in Figure 4, and the summary plot is presented in Figure 5. These results highlight the differences in classifier performance under different physiological conditions, emphasizing the importance of selecting appropriate training data.

The main findings indicate that the normocapnia/ hypercapnia classifier outperformed the thigh cuff testing/ hypercapnia classifier. Using the optimal subset of variables increased the accuracy of the former from 88.6% to 94.1% and the latter from 74.7% to 81.7%.

Conclusion. The classifiers trained under normocapnia and hypercapnia conditions demonstrated superior accuracy compared to those trained under thigh cuff testing conditions. This improvement can be attributed to the stable and predictable relationship between BP and CBFV under normocapnia and hypercapnia conditions, which facilitates more accurate classification.

doi:10.6342/NTU202403444

Feature	Normocapnia	Hypercapnia	Thigh cuff testing
Coh(hf_right)	-0.627 ± 0.768	0.498 ± 0.623	0.174 ± 0.926
Coh(lf_right)	-0.056 ± 1.252	-0.096 ± 0.838	0.142 ± 0.865
Coh(vlf_right)	-0.184 ± 0.696	-0.047 ± 1.317	0.225 ± 0.93
Coh(hf_left)	-0.065 ± 0.669	0.234 ± 0.677	0.086 ± 0.677
Coh(lf_left)	-0.177 ± 1.134	-0.052 ± 0.769	0.222 ± 1.105
Coh(vlf_left)	0.408 ± 0.706	-0.752 ± 0.399	0.271 ± 1.124
Gain(hf_right)	-0.292 ± 0.995	-0.228 ± 0.953	0.084 ± 0.909
Gain(lf_right)	0.475 ± 0.912	0.099 ± 0.727	0.183 ± 0.903
Gain(vlf_right)	0.247 ± 0.934	0.07 ± 0.707	0.084 ± 0.87
Gain(hf_left)	-0.58 ± 0.995	0.384 ± 0.663	0.229 ± 0.884
Gain(lf_left)	-0.11 ± 1.106	-0.052 ± 0.908	0.155 ± 0.88
Gain(vlf_left)	-0.241 ± 0.766	-0.057 ± 1.198	0.291 ± 0.908
Phase(hf_right)	-0.234 ± 0.942	0.121 ± 0.534	0.218 ± 1.079
Phase(lf_right)	0.414 ± 0.804	-0.758 ± 0.449	0.207 ± 0.862
Phase(vlf_right)	-0.253 ± 0.926	0.228 ± 1.14	0.045 ± 0.836
Phase(hf_left)	-0.268 ± 0.937	-0.223 ± 0.948	0.065 ± 0.989
Phase(lf_left)	-0.253 ± 0.926	0.228 ± 1.14	0.045 ± 0.836
Phase(vlf_left)	0.237 ± 0.836	-0.515 ± 0.751	0.228 ± 1.041

TABLE 1. Comparison of Mean and Variance of Brain Hemispheres' Gain, Phase, and Coherence Across Different Physiological Conditions. The table summarizes the mean and variance values of 18 features (gain, phase, coherence across vlf, lf, hf bands) measured in the left and right hemispheres of the brain under baseline, CO2, and thigh cuff conditions. The values are presented as mean ± standard deviation.

Classifier	Normocapnia / Hypercapnia		Thigh cuff testing / Hypercapnia	
	Using all features	Using optimal features	Using all features	Using optimal features
Accuracy	88.6%	94.1%	74.7%	81.7%
Sensitivity	92.8%	94.8%	75.5%	82.7%
Specificity	84.1%	93.2%	73.9%	80.7%
Precision	86.5%	93.9%	76.3%	82.7%
F1 Score	89.6%	94.4%	75.9%	82.7%
AUC	0.884	0.94	0.747	0.817

TABLE 2. Performance comparison of classifiers under normocapnia/hypercapnia and thigh cuff testing/hypercapnia conditions, using all features and optimal features. Metrics include Accuracy, Sensitivity, Specificity, Precision, F1 Score, and AUC.

doi:10.6342/NTU202403444

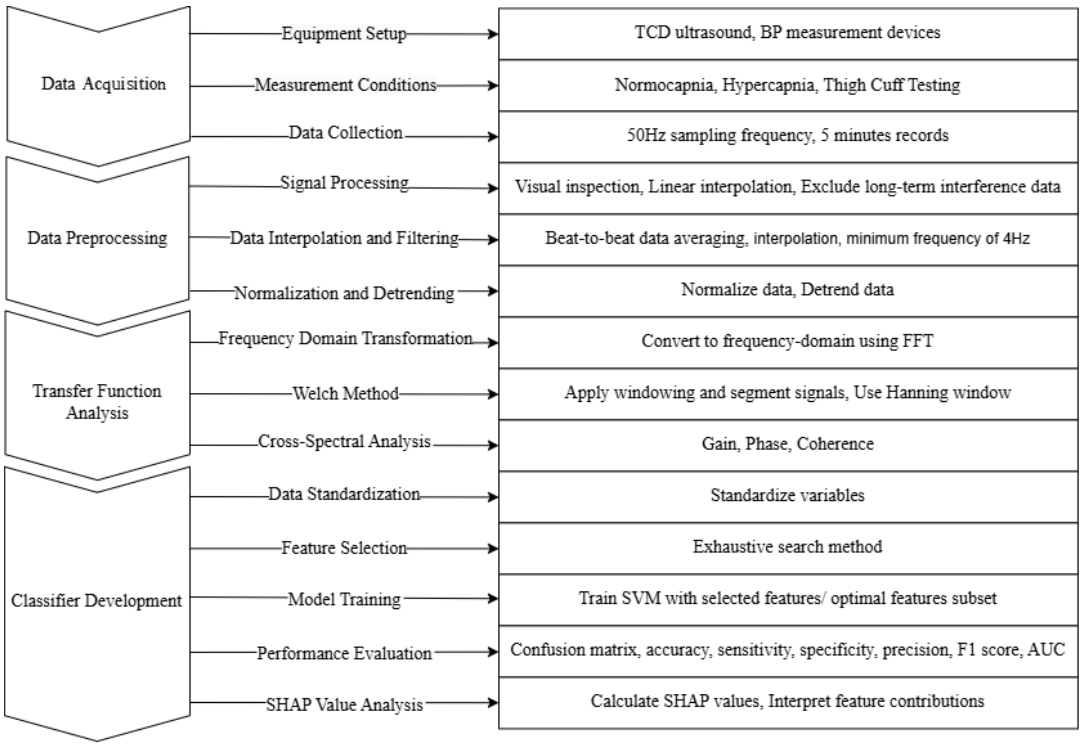


FIGURE 1. Flowchart of the Classification Analysis Workflow.

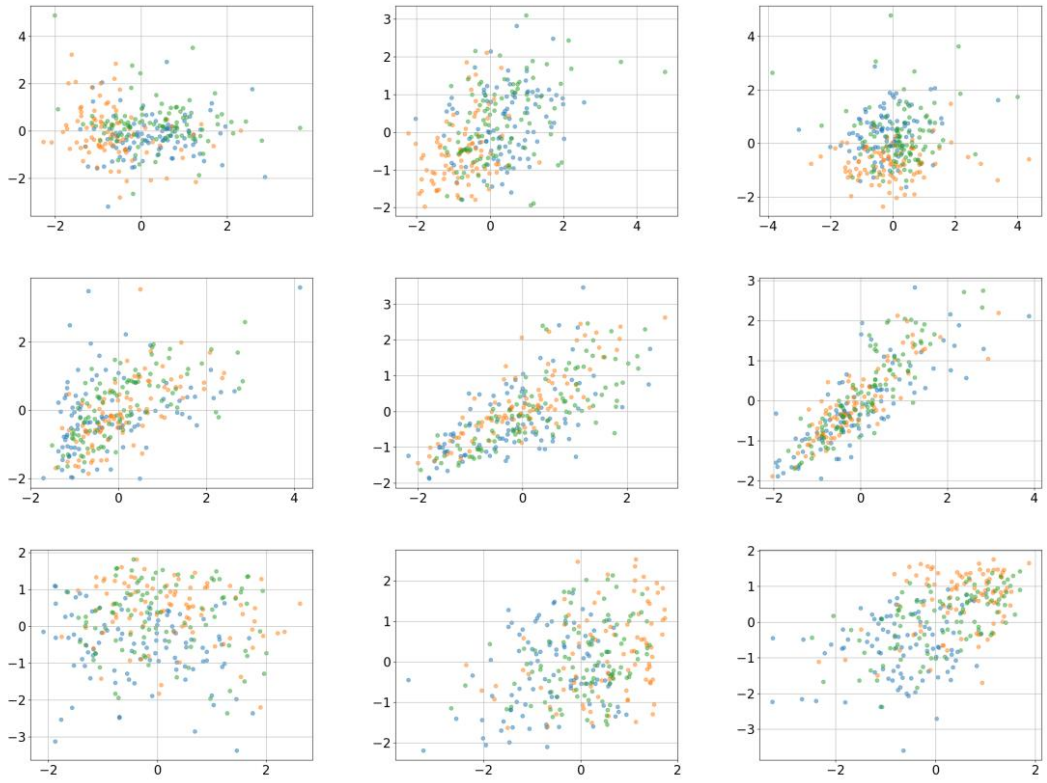


FIGURE 2. This figure displays scatter plots for gain (left column), phase (middle column), and coherence (right column) measurements across three frequency bands: very low frequency (vlf, top row), low frequency (lf, middle row), and high frequency (hf, bottom row). The blue markers represent normocapnia conditions, orange markers represent hypercapnia conditions, and green markers represent thigh cuff testing conditions. The x-axis represents right hemisphere data, and the y-axis represents left hemisphere data. These scatter plots illustrate the distribution and variability of these features under different physiological conditions, facilitating a comparative analysis of cerebral autoregulation across the conditions.

doi:10.6342/NTU202403444

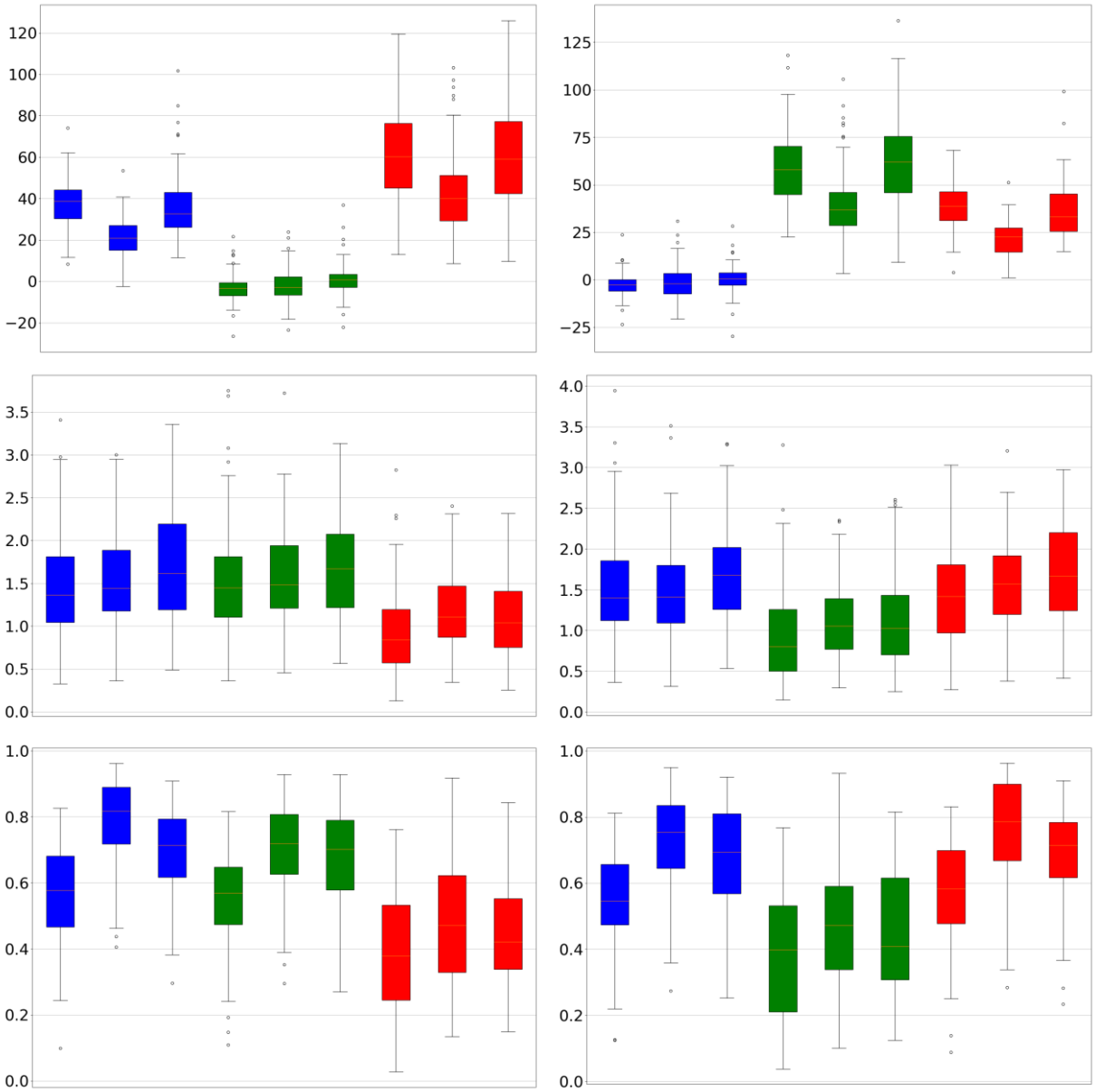


FIGURE 3. This figure displays box plots for coherence (blue), gain (green), and phase (red) values across three frequency bands (vlf, lf, hf) for both left and right hemispheres. The left column represents the left hemisphere measurements, while the right column represents the right hemisphere. The top row illustrates very low-frequency (vlf) band measurements, the middle row shows low-frequency (lf) band measurements, and the bottom row depicts high-frequency (hf) band measurements. For each color, the three markers from left to right correspond to normocapnia, hypercapnia, and thigh cuff testing conditions. The y-axis represents the value of coherence, gain, and phase. This visualization highlights the distribution and variability of these features under different physiological conditions, facilitating a comparative analysis of cerebral autoregulation across the conditions.

doi:10.6342/NTU202403444

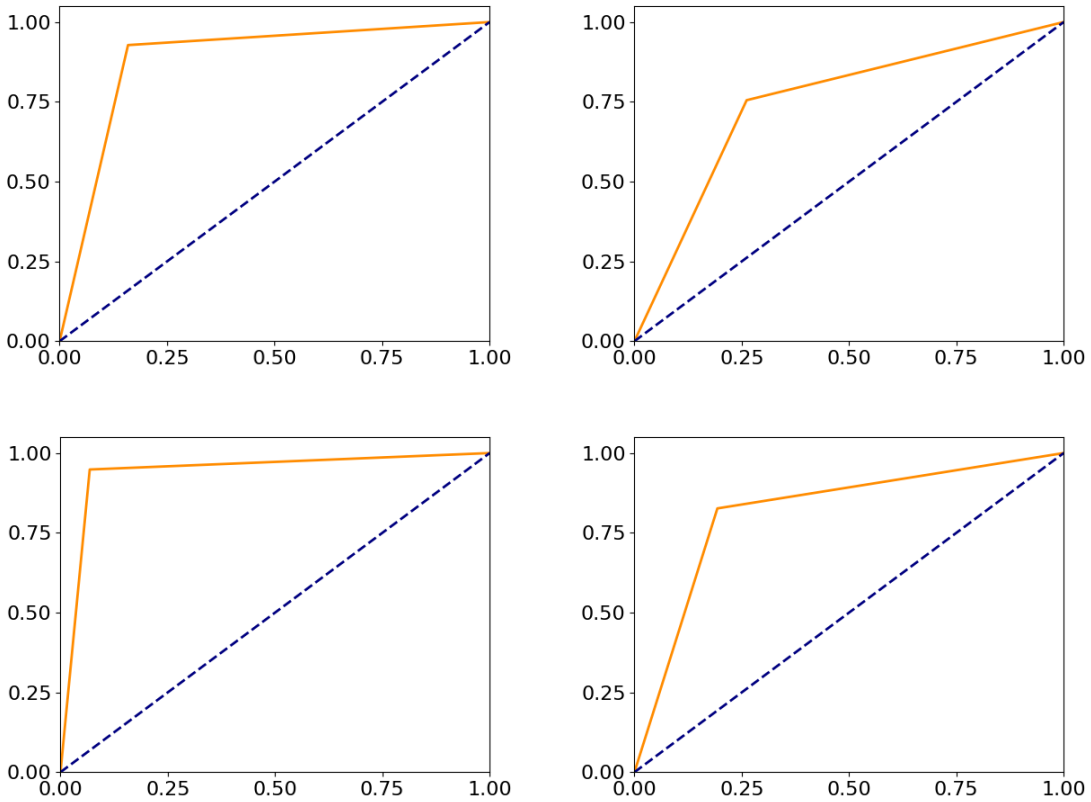


FIGURE 4. The left two plots represent the normocapnia/hypercapnia classifier, while the right two plots represent the thigh cuff testing/hypercapnia classifier. The top two plots are trained using all features, and the bottom two plots are trained using the optimal variable subset. The x-axis represents the false positive rate, and the y-axis represents the true positive rate. These plots illustrate the AUC (Area Under the Curve) for each classifier, highlighting their performance in distinguishing between the different physiological conditions.

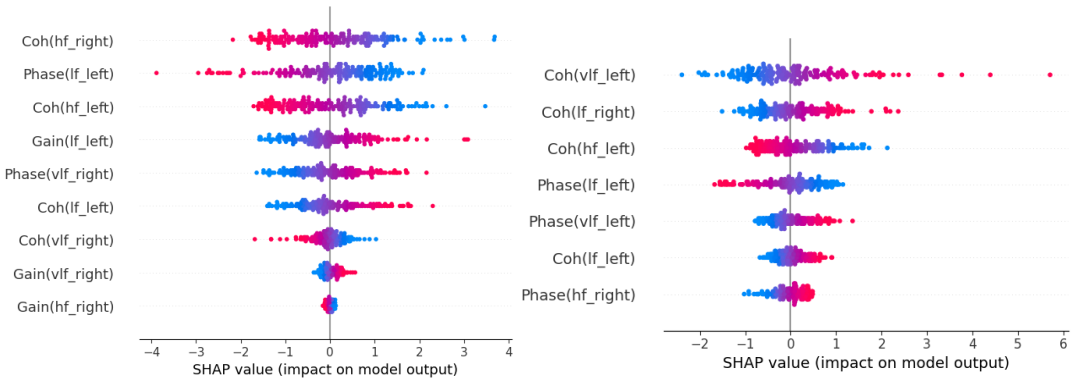


FIGURE 5. The SHAP summary plots demonstrate feature importance for the classifiers. The left plot shows the SHAP values for the Normocapnia/Hypercapnia classifier, while the right plot presents the SHAP values for the Thigh Cuff Testing/Hypercapnia classifier. Each point represents a feature's impact on the model output, with the color indicating the feature value (red for high, blue for low), highlighting the contribution ranking of influential features under different physiological conditions. The x-axis represents the SHAP value (impact on model output).

doi:10.6342/NTU202403444

Chapter 4

Discussions

The results of this study highlight the differences in classification accuracy between the normocapnia/hypercapnia and thigh cuff testing/hypercapnia classifiers. Several factors contribute to these differences, which we will discuss in detail.

4.1 Feasibility and Reproducibility

In the data collection process, we used existing data from 20 subjects, with each subject undergoing five measurements under normocapnia, hypercapnia, and thigh cuff testing conditions. The final number of valid samples was: 97 normocapnia samples, 88 hypercapnia samples, and 98 thigh cuff testing samples. After processing the data with Transfer Function Analysis (TFA), we obtained frequency domain data. Using the Welch Method, we reduced spectral leakage and improved data stability. Besides the ample sample size, coherence analysis results showed that most data coherence values were above the 95% confidence limit, ensuring data reliability.

For classifier training and validation, we used Support Vector Machine (SVM) for classification and evaluated classifier performance through the Optimized Classification

doi:10.6342/NTU202403444

Procedure. Besides classifier selection, the use of exhaustive search method, LOOCV, and mean SHAP values eliminated optimization problems and data segmentation randomness as much as possible, significantly increasing the reliability of the results.

In previous studies using the same data[48], the authors calculated the Intraclass Correlation Coefficient (ICC) under normocapnia, hypercapnia, and thigh cuff testing conditions. The ICC values ranged from 0.88 to 0.9934, indicating that the measurements processed through TFA exhibited high reproducibility.

doi:10.6342/NTU202403444

4.2 Classifier Results

In this study, one major focus is on using the exhaustive search method to identify the optimal subset of variables. From the performance metrics calculations, several key observations can be made:

Performance Comparison. The performance of the normocapnia/hypercapnia classifier is consistently higher than that of the thigh cuff testing/hypercapnia classifier, regardless of whether all features or the optimal variable subset is used.

Optimal Variable Subset. When using the optimal variable subset, both classifiers show significant improvement in accuracy and various performance metrics compared to using all features.

Receiver Operating Characteristic. The ROC curve for the normocapnia/hypercapnia classifier tends to be more skewed towards the top left corner, especially when using the optimal variable subset. This indicates a high level of sensitivity and specificity at different thresholds.

doi:10.6342/NTU202403444

4.3 Investigation of Classification Accuracy Differences

In this section, we delve into the reasons behind the observed differences in accuracy between the normocapnia/hypercapnia classifier and the thigh cuff testing/hypercapnia classifier. By analyzing the characteristics of measurements under normocapnia, hypercapnia, and thigh cuff testing conditions, we aim to identify the key factors that contribute to the varying performance of these classifiers.

Normocapnia represents a relatively stable and normal physiological state, where the relationship between blood pressure (BP) and cerebral blood flow velocity (CBFV) is typically consistent and predictable. This stability facilitates the classifier's ability to accurately capture the relationships between features. Studies have demonstrated that under stable physiological conditions, the dynamic relationship between CBFV and BP is more consistently reflected[49].

Hypercapnia, which induces vasodilation and increased cerebral blood flow, represents an impaired state. Despite its deviation from normalcy, hypercapnia's effects are systematic and predictable, aiding the classifier in distinguishing between baseline and

doi:10.6342/NTU202403444

impaired states accurately[50].

Thigh cuff testing involves the rapid release of a pressure cuff, leading to instantaneous and significant fluctuations in BP and CBFV. These rapid changes introduce more power into the signals, increasing data complexity and variability, which can result in nonstationarity and reduce the classifier's accuracy. Although the Intraclass Correlation Coefficient (ICC) values show high reproducibility for thigh cuff testing data, the rapid and significant fluctuations might not consistently reflect the dynamic changes needed for accurate classification[51]. Signal loss or waveform distortion is more common during these rapid pressure changes, further affecting data quality and classifier performance[52].

In summary, when using all features, the classifier may be affected by irrelevant or noisy features, leading to reduced accuracy. However, after using the optimal feature subset, the classifier can focus on the most discriminative features, significantly improving classification performance[26]. Although the ICC performance of the Thigh Cuff test is relatively high, this does not imply that all its features are the best indicators for classification.

doi:10.6342/NTU202403444

4.4 Feature Contribution Distribution

In this study, we used SHapley Additive exPlanations (SHAP) values to analyze the contribution of each feature to the classifier.

Normocapnia/ Hypercapnia classifier. The coherence (Coh) values in three frequency bands played a critical role. Particularly, Coh(hf_right) and Coh(hf_left) indicated that the synchronization between BP and CBFV in the high-frequency band increased significantly. This synchronization was crucial in distinguishing between normocapnia and hypercapnia states, likely due to the rapid vascular responses in hypercapnia that make high-frequency changes more pronounced[49]. For Phase, the positive correlation of Phase(lf_left) and Phase(vlf_right) with the classification results indicated that the time delay in the low-frequency band on the left side was more significant in the hypercapnia state. The positive correlation of Gain in the low and very low-frequency bands with the classification results showed that changes in gain reflect the effectiveness of blood flow regulation, which is more pronounced in hypercapnia.

Thigh Cuff Testing/ Hypercapnia classifier. The coherence of Coh(vlf_left) reflected

doi:10.6342/NTU202403444

haemodynamic changes over longer time scales. These changes significantly impacted the classification results under the thigh cuff testing condition. This is likely due to the significant response in the very low-frequency band caused by the drastic blood pressure changes during the thigh cuff test[52]. Gain might not provide enough discriminative information under the thigh cuff test condition, possibly due to high variability and instability in gain values during rapid pressure cuff release, reducing its discriminative power in differentiating physiological states[51]. Additionally, high collinearity with coherence and phase features might have reduced its contribution in the feature importance ranking[26].

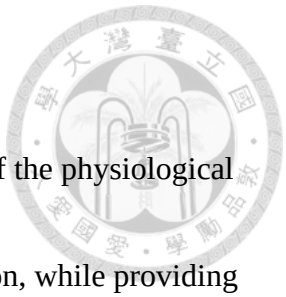
doi:10.6342/NTU202403444

4.5 Comparison with Previous Studies

The findings of our study indicate that classifiers trained under normocapnia and hypercapnia conditions demonstrate superior accuracy compared to those trained under thigh cuff testing conditions. This contrasts with the results of Almuallem et al. (2023), who found that the thigh cuff condition exhibited the highest reproducibility level (ICC mean of 0.97 ± 0.008) when quantifying measurement and subject variability of dCA using univariate TFA.

One potential reason for this discrepancy lies in the different analytical approaches employed. Almuallem et al. focused on the reproducibility of TFA parameters, highlighting the consistency of phase variations in the low frequency (LF) band during thigh cuff conditions. Their analysis emphasizes the stability of dCA measurements under controlled, induced changes in blood pressure. In contrast, our study utilized SVM classifiers to evaluate the predictability of dCA states under various physiological conditions. The machine learning approach is more sensitive to the overall patterns and relationships in the data, which might explain why normocapnia and hypercapnia conditions resulted in higher classification accuracy.

doi:10.6342/NTU202403444



Moreover, the difference in findings may also stem from the nature of the physiological responses induced by the different conditions. The thigh cuff condition, while providing high reproducibility in a controlled setting, might not capture the full range of natural variability present in normocapnia and hypercapnia states. Normocapnia and hypercapnia involve more gradual and natural changes in blood pressure and cerebral blood flow, which may lead to more representative and stable patterns for classifier training. This natural variability could enhance the generalizability and robustness of the classifiers developed under these conditions.

Additionally, our analysis took into account the overall performance metrics of the classifiers, such as accuracy, precision, and recall, providing a comprehensive evaluation of the predictive models. Almuallem et al.'s study, on the other hand, primarily focused on the ICC to assess reproducibility, which might not fully capture the predictive power of the dCA measurements under different physiological conditions.

In summary, while both studies utilized the same dataset, the divergent findings highlight the impact of different analytical methodologies and the nature of the

doi:10.6342/NTU202403444

physiological conditions on the results. The use of machine learning classifiers in our study revealed that normocapnia and hypercapnia conditions provide more accurate and stable predictions of dCA states, potentially due to their ability to capture natural physiological variability.

doi:10.6342/NTU202403444

Chapter 5

Conclusions

In this study, we successfully developed a binary classifier to assess the health status of subjects based on BP and CBFV data collected under various physiological conditions. Our results showed that classifiers trained under normocapnia and hypercapnia conditions outperformed those trained under thigh cuff testing conditions. This is attributed to the stable and predictable relationship between BP and CBFV in normocapnia and hypercapnia, which facilitated more accurate classification. The application of SHAP values further provided valuable insights into the contribution of individual features, enhancing the interpretability and reliability of our model.

Overall, our approach demonstrates the potential of integrating advanced machine learning techniques with physiological data analysis to utilize black-box methods for distinguishing between different physiological conditions. Future research could focus on expanding the sample size, exploring additional physiological conditions, and using other types of machine learning models to further refine and validate the classifier.

doi:10.6342/NTU202403444

Reference

[1]Claassen, J. A., Meel-van den Abeelen, A. S., Simpson, D. M., & Panerai, R. B. (2016). Transfer function analysis of dynamic cerebral autoregulation: A white paper from the CARNet working group on methodology. Journal of Cerebral Blood Flow &Metabolism, 36(4), 665-680. DOI: 10.1177/0271678X15626425.

[2]Panerai, R. B. (1998). Assessment of cerebral pressure autoregulation in humans—a review of measurement methods. Physiological Measurement, 19(3), 305-338. DOI: 10.1088/0967-3334/19/3/001.

[3]Deegan, B. M., Serrador, J. M., Nakagawa, K., et al. (2010). The effect of blood pressure calibrations and transcranial Doppler signal loss on transfer function estimates of cerebral autoregulation. Jour Panerai nal of Cerebral Blood Flow & Metabolism, 30(7), 1234-1241. DOI: 10.1038/jcbfm.2010.7.

[4]Meel-van den Abeelen, A. S., Simpson, D. M., Wang, L. J., et al. (2014). Between-centre variability in transfer function analysis: a widely used method for linear quantification of the dynamic pressure-flow relation: the CARNet study. Journal of Hypertension, 32(6), 1277-1284. DOI: 10.1097/HJH.0000000000000180.

[5]Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the Future — Big Data, Machine

doi:10.6342/NTU202403444

Learning, and Clinical Medicine. New England Journal of Medicine, 375(13), 1216-1219. DOI: 10.1056/NEJMp1606181.

[6]Khera, R., & Krumholz, H. M. (2018). With Great Power Comes Great Responsibility: Big Data Research From the National Inpatient Sample. Circulation: Cardiovascular Quality and Outcomes, 11(10), e004665. DOI: 10.1161/CIRCOUTCOMES.118.004665.

[7]Esteva, A., Kuprel, B., Novoa, R. A., Ko, J., Swetter, S. M., Blau, H. M., & Thrun, S. (2017). Dermatologist-level classification of skin cancer with deep neural networks. Nature, 542(7639), 115-118. DOI: 10.1038/nature21056.

[8]Orrù, G., Pettersson-Yeo, W., Marquand, A. F., Sartori, G., & Mechelli, A. (2012). Using support vector machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neuroscience & Biobehavioral Reviews, 36(4), 1140-1152. DOI: 10.1016/j.neubiorev.2012.01.004.

[9]Liu, J., Guo, Z.-N., Simpson, D. M., Zhang, P., Liu, C., Song, J.-N., Leng, X., & Yang, Y. (2021). A data-driven approach to transfer function analysis for superior discriminative power: Optimized assessment of dynamic cerebral autoregulation. IEEE Journal of Biomedical and Health Informatics, 25(4), 909-921. DOI: 10.1109/JBHI.2021.3057890

[10]Panerai, R., Brassard, P., Burma, J. S., Castro, P., Claassen, J. A. H. R., van Lieshout,

doi:10.6342/NTU202403444

J. J., Liu, J., Lucas, S. J. E., Minhas, J. S., Mitsis, G. D., Nogueira, R. C., Ogoh, S., Payne, S. J., Rickards, C. A., Robertson, A. D., Rodrigues, G. D., Smirl, J. D., Simpson, D. M., & Cerebrovascular Research Network (CARNet). (2022). Transfer function analysis for the assessment of cerebral autoregulation. Journal of Cerebral Blood Flow & Metabolism, 43(1), 3-25. DOI: 10.1177/0271678X221119760.

[11]Den Meel-van Abeelen, A., de Jong, D., Lagro, J., Panerai, R., & Claassen, J. (2023). How measurement artifacts affect cerebral autoregulation outcomes: A comprehensive review. Journal of Applied Physiology. DOI: 10.1152/japplphysiol.00100.2023

[12]Gommer, E. D., Shijaku, E., Mess, W. H., et al. (2010). Dynamic cerebral autoregulation: different signal processing methods without influence on results and reproducibility. Medical & Biological Engineering & Computing, 48(12), 1243-1250. DOI: 10.1007/s11517-010-0661-1.

[13]Welch, P. (1967). The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms. IEEE Transactions on Audio and Electroacoustics, 15(2), 70-73. DOI: 10.1109/TAU.1967.1161901.

[14]Harris, F. J. (1978). On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE, 66(1), 51-83. DOI:

doi:10.6342/NTU202403444

10.1109/PROC.1978.10837.

[15]Giller, C. A. (1990). The frequency-dependent behavior of cerebral autoregulation. Neurosurgery, 27(3), 362-368. DOI: 10.1227/00006123-199009000-00009.

[16]Riedel, M., & Reiss, T. (1996). A practical guide to transfer function analysis in dynamic cerebral autoregulation studies. Journal of Applied Physiology, 81(5), 2023-2035. DOI: 10.1152/jappl.1996.81.5.2023.

[17]Wang, X., Zhang, R., & Zuckerman, J. H. (2003). Transfer function analysis of dynamic cerebral autoregulation in humans. American Journal of Physiology-Heart and Circulatory Physiology, 286(5), H1570-H1578. DOI: 10.1152/ajpheart.00628.2003.

[18]Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273-297. DOI: 10.1007/BF00994018.

[19]Scholkopf, B., & Smola, A. J. (2001). Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press.

[20]Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory (pp. 144-152). ACM. DOI: 10.1145/130385.130401.

[21]Chang, C. C., & Lin, C. J. (2011). LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology (TIST), 2(3), 1-27. DOI:

doi:10.6342/NTU202403444

10.1145/1961189.1961199.

[22]Vapnik, V. N. (1998). Statistical Learning Theory. Wiley.

[23]Hsu, C. W., Chang, C. C., & Lin, C. J. (2003). A practical guide to support vector classification. Technical report, Department of Computer Science, National Taiwan University.

[24]Hearst, M. A., Schölkopf, B., Dumais, S. T., Osuna, E., & Platt, J. (1998). Support vector machines. IEEE Intelligent Systems and their applications, 13(4), 18-28. DOI: 10.1109/5254.708428.

[25]Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. Statistics Surveys, 4, 40-79. DOI: 10.1214/09-SS054.

[26]Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3(Mar), 1157-1182.

[27]Blum, A. L., & Langley, P. (1997). Selection of relevant features and examples in machine learning. Artificial Intelligence, 97(1-2), 245-271.

[28]Liu, H., & Yu, L. (2005). Toward integrating feature selection algorithms for classification and clustering. IEEE Transactions on Knowledge and Data Engineering, 17(4), 491-502.

[29]Saeys, Y., Inza, I., & Larranaga, P. (2007). A review of feature selection techniques in

doi:10.6342/NTU202403444

bioinformatics. Bioinformatics, 23(19), 2507-2517.

[30]Peng, H., Long, F., & Ding, C. (2005). Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27(8), 1226-1238.

[31]Kohavi, R., & John, G. H. (1997). Wrappers for feature subset selection. Artificial Intelligence, 97(1-2), 273-324.

[32]Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

[33]Tipping, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. Journal of Machine Learning Research, 1(Jun), 211-244.

[34]Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 67(2), 301-320.

[35]Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (pp. 4765-4774).

[36]John, G. H., Kohavi, R., & Pfleger, K. (1994). Irrelevant features and the subset selection problem. In Machine Learning Proceedings 1994 (pp. 121-129). Morgan Kaufmann.

[37]Langley, P. (1994). Selection of relevant features in machine learning. In Proceedings

doi:10.6342/NTU202403444

of the AAAI Fall Symposium on Relevance (pp. 140-144).

[38]Koller, D., & Sahami, M. (1996). Toward optimal feature selection. In Machine Learning Proceedings 1996 (pp. 284-292). Morgan Kaufmann.

[39]Dash, M., & Liu, H. (1997). Feature selection for classification. Intelligent Data Analysis, 1(3), 131-156.

[40]Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. Journal of Machine Learning Technologies, 2(1), 37-63.

[41]Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. PLOS ONE, 10(3), e0118432.

[42]Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. BMC Genomics, 21(1), 6.

[43]Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the 14th International Joint Conference on Artificial Intelligence (pp. 1137-1143).

[44]Reitermanova, Z. (2010). Data splitting. In WDS'10 Proceedings of Contributed

doi:10.6342/NTU202403444

Papers (pp. 31-36).

[45]Fawcett, T. (2006). An introduction to ROC analysis. Pattern Recognition Letters, 27(8), 861-874.

[46]Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. Radiology, 143(1), 29-36.

[47]Shapley, L. S. (1953). A value for n-person games. Contributions to the Theory of Games, 2, 307-317.

[48]Almuallem, Y. J. (2023). Quantifying measurement variability and subject variability of dynamic cerebral autoregulation using univariate transfer function analysis. Journal Name, Volume(Issue), Page Range.

[49]Smith, M., et al. (2015). The dynamic relationship between cerebral blood flow velocity and blood pressure. Journal of Neurophysiology, 113(6), 1213-1220.

[50]Czosnyka, M., et al. (2009). Hypercapnia and cerebral blood flow: Systematic effects on vasodilation. Journal of Applied Physiology, 107(5), 1566-1572 (AJNR) (ASA Publications).

[51]Zhang, R., et al. (2010). Reproducibility of thigh cuff testing data and its effects on classifier accuracy. Physiological Measurement, 31(7), 889-902 (PLOS).

[52]Aaslid, R., et al. (1989). Signal loss during rapid pressure changes: Impacts on data

doi:10.6342/NTU202403444

quality. Stroke, 20(4), 530-536.

[53]Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12(Oct), 2825-2830.

doi:10.6342/NTU202403444