Application of machine learning and artificial intelligence in diagnosis of clogged ducts

line 1: 1st Chuo Hong Xu
line 2: *Department of Computer Science and Information Engineering*
line 3: *National Yunlin University of Science and Technology*
line 4: Douliu, Taiwan
line 5: m11217045@yuntech.edu.tw

line 1: 2nd Wen-Feng Wang*[1]
line 2: *Department of Computer Science and Information Engineering*
line 3: *National Yunlin University of Science and Technology*
line 4: Douliu, Taiwan
line 5: wwf@yuntech.edu.tw

***Abstract*—**

To address the issue of limited data, transfer learning was applied by pre-training the model with PPG signals from healthy individuals to extract a large number of second-order derivative features. These features were then fine-tuned for patient-specific data. Finally, support vector machines (SVM), k-nearest neighbors (KNN), and random forests (RF) were employed for classification, determining whether patients were prone to occlusion (requiring two surgeries within three months) or not.

The results demonstrate that this method achieves high accuracy in predicting and classifying AVF occlusion risks, highlighting the potential of combining multi-order derivative features of PPG signals with transfer learning in the medical diagnostic field. This approach provides a non-invasive and efficient solution for early diagnosis and personalized medical interventions for AVF occlusion.

Keywords—Arteriovenous Fistula, Photoplethysmography, Second-Order Derivative Features, KNN, SVM, RF

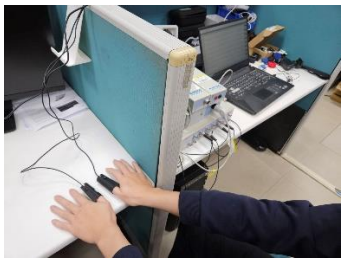
I. INTRODUCTION (*HEADING 1*)

The innovation of this study lies in predicting the potential need for surgery from the perspective of "occlusion risk." By analyzing photoplethysmogram (PPG) signals, we extract hemodynamic characteristics before and after surgery and, in combination with transfer learning and machine learning techniques, develop a non-invasive predictive model. With this model, doctors can identify potential surgical risks before the patient exhibits obvious occlusion symptoms, enabling more effective preventive treatment, reducing health threats to patients, and optimizing healthcare resource allocation.

II. EXPERIMENTAL EQUIPMENT AND DATA

A. Research Instruments

In this study, the PowerLab system produced by ADI Instruments. For this research, a fingertip-based PPG sensor module was used, with a sampling rate set at 1000 Hz. For subsequent signal processing and analysis, we developed a Python-based program, including feature extraction, data processing, and classification, among other stages.



B. Participants

In this study, participants were categorized into two groups: those who had undergone two or more fistula intervention surgeries within the past three months were classified as "high-risk for occlusion" patients, while those with fewer interventions were classified as "low-risk for occlusion" patients. Participants were recruited from the dialysis center at the Department of Nephrology, National Cheng Kung University Hospital, Douliu Branch. The inclusion criteria were as follows: participants had to be at least 20 years of age, with no gender restrictions. Additionally, selected participants had to have undergone either autogenous arteriovenous fistula (AVF) or synthetic arteriovenous graft (AVG) placement on the left or right arm and must have no history of cardiovascular diseases.

C. PPG Signals and Experimental Features

In this study, PPG serves as the core signal source, and its waveform characteristics directly reflect the dynamic changes in blood flow. To accurately analyze the features of the PPG signal, we extracted the following physiologically meaningful characteristics (parameters):

Feature	Definition
Systolic Peak	The maximum value of systolic pressure in the PPG waveform
Diastolic Peak	The maximum value of diastolic pressure in the PPG waveform
Cardiac Cycle	The period of the PPG waveform
Cycle Area	The area under one cycle of the PPG waveform
SSI	The time from one PPG peak to the next PPG peak
Peak to Valley	The time from the valley to the peak in the PPG waveform
Systolic Peak Height	The height of the systolic peak in the PPG waveform
Delta T	The distance between the systolic and diastolic peaks in the PPG waveform
1st Derivative Cycle	The period time of the first derivative of the PPG waveform.
1st Derivative Peak	The peak value of the first derivative of the PPG waveform.
Ratio b/a	The ratio of points b and a in the second derivative of the PPG waveform.
Ratio c/a	The ratio of points c and a in the second derivative of the PPG waveform.
Ratio d/a	The ratio of points d and a in the second derivative of the PPG waveform.
Ratio (b-d-c-e)/a	The ratio of the second derivative (b-d-c-e) and point a in the PPG waveform.
Ratio (b-e)/a	The ratio of the second derivative (b-e) and point a in the PPG waveform.

[1] Corresponding Author. Email: wwf@yuntech.edu.tw

III. RESEARCH METHODS

A. Signal Preprocessing

In this study, the PPG signals underwent preprocessing using a fourth-order Butterworth filter for signal filtering. The Butterworth filter was chosen for its smooth frequency response characteristics, with a cutoff frequency set between 0.7 Hz and 9 Hz. Following filtering, the signals were segmented based on the periodic characteristics of the PPG waveform, dividing the signal into segments corresponding to two complete cardiac cycles. This segmentation method effectively increases the sample size of the signal, providing sufficient data support for subsequent classification models.

B. PPG feature extraction

This study primarily employs second-order derivative features for analysis in order to capture subtle changes related to hemodynamics. After extracting the second-order derivative features, we further derive the first-order derivative and the original waveform features to create a more comprehensive feature set. This approach enhances the classification model's ability to distinguish between patients at high risk of occlusion and those at low risk.

C. Waveform Quality Screening

In this study, three classification methods were employed to categorize the PPG waveforms in order to assess their suitability for subsequent analysis. The classification objectives include: (1) identifying high-quality waveforms; (2) classifying waveforms with less distinct second-order derivative features that can still be identified manually; and (3) categorizing waveforms with excessive motion artifacts where features cannot be manually identified. During the waveform classification process, the data were screened and assessed based on the following criteria to ensure data quality and improve the reliability and accuracy of the analysis.

D. Transfer Learning

The method proposed in this study first extracts features from the second-order derivative PPG (SDPPG) signals of a healthy population, focusing on capturing second-order derivative feature points related to hemodynamics. Next, an XGBoost classification model is used to train and predict these feature points from the healthy population, thereby generating a robust preliminary classification model. Subsequently, transfer learning techniques are applied to fine-tune the model, adapting it to the feature distribution of patient PPG signals. This process results in a patient-specific classification model

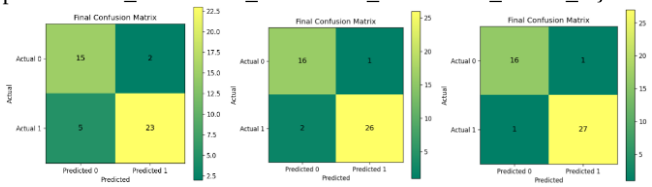
IV. RESULT

the PPG feature data from 32 experimental participants (11 with high risk of occlusion and 21 with low risk of occlusion) were analyzed using the Mann-Whitney U test. In the second part, the optimal feature combination obtained through exhaustive search will be applied to three supervised machine learning algorithms: K-Nearest Neighbors (KNN), Support Vector Machine (SVM), and Random Forest (RF).

A. Mann-Whitney U test

In the Mann-Whitney U test, features with a p-value < 0.05 were used for classification. The optimal subsets and accuracy results for SVM, KNN, and RF are as follows:

The optimal subsets :{systolic peak、Diastolic peak、Cardiac cycle、SSI、Delta_T、1st Derivative cycle、1st Derivative peak、Ratio_BA、Ratio_CA、Ratio_DA、Ratio_BDCE_A}



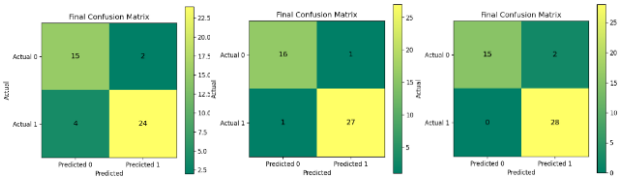
(a)SVM (b) KNN (c) RF

The Best Accuracy: 0.956,
'Sensitivity (Recall)': 0.941, 'Specificity': 0.964,
'Precision': 0.941, 'F1-Score': 0.941

B. Exhaustive search

In the exhaustive search, the feature subset with the highest accuracy was selected for classification. The results for SVM, KNN, and RF are as follows:

The optimal subsets :{ Diastolic peak、Ratio_BA、Ratio_CDB_A 、Ratio_CA、Peak to Valley、Ratio_DA、Ratio_BDCE_A、Delta_T、systolic peak }



(a)SVM (b) KNN (c) RF

The Best Accuracy: 0.956,
'Sensitivity (Recall)': 0.882, 'Specificity': 1.0,
'Precision': 1.0, 'F1-Score': 0.938

REFERENCES

[1] Elgendi, M. (8 C.E., February 20). On the Analysis of Fingertip Photoplethysmogram Signals. National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC3394104/

[2] Baruch, M. C. (2011, January 12). Pulse Decomposition Analysis of the Digital Arterial Pulse during Hemorrhage Simulation. National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC3025935/.

[3] Wu , J. (1015, May 1). Bilateral Photoplethysmography Analysis for Arteriovenous Fistula Dysfunction Screening with Fractional-Order Feature and Cooperative Game-Based Embedded Detector. National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC4614111/.

[4] Song, J. (2019, January). PQR Signal Quality Indexes: A Method for Real-Time Photoplethysmogram Signal Quality Estimation Based on Noise Interferences. ScienceDirect. https://pmc.ncbi.nlm.nih.gov/articles/PMC3025935/.

[5] Tsai, P. (2021, June 24). Coherence between Decomposed Components of Wrist and Finger PPG Signals by Imputing Missing Features and Resolving Ambiguous Features. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/34202597/

[6] Suboh, M. Z. (2022, June 30). Analysis on Four Derivative Waveforms of Photoplethysmogram (PPG) for Fiducial Point Detection. National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC9280335/