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—

Arteriovenous fistula (AVF) occlusion is a common complication in hemodialysis patients, often requiring repeated surgical interventions. To address the issue of limited patient data, we propose a machine learning approach combining photoplethysmography (PPG) signal analysis with transfer learning. A model was pre-trained on PPG signals from healthy individuals to extract second-order derivative features and then fine-tuned using patient-specific data. Classification was performed using support vector machines (SVM), k-nearest neighbors (KNN), and random forests (RF) to identify patients at high risk of AVF occlusion, defined as requiring two or more interventions within three months.

Experimental results from 32 participants (11 high-risk, 21 low-risk) show our method achieved an accuracy of 95.6%**, sensitivity of** 94.1%**, specificity of** 96.4%**, and F1-score of** 94.1%**. This demonstrates the potential of combining multi-order PPG features and transfer learning to enable non-invasive, efficient early detection of AVF dysfunction and support personalized clinical decision-making.**

Keywords—ArteriovenousFistula, Photoplethysmography, Transfer Learning, Second-Order Derivative, Machine Learning

I. Introduction (*Heading 1*)

The primary innovation of this study is the prediction of arteriovenous access dysfunction from the perspective of occlusion risk. By analyzing photoplethysmography (PPG) signals collected before and after surgery, we extract hemodynamic features and apply transfer learning combined with machine learning to develop a non-invasive predictive model. This approach enables early identification of occlusion risk, supporting timely clinical intervention and improving resource allocation in dialysis care.

II. Experimental Equipment and Data

A. Research Instruments

PPG signals were acquired using a fingertip sensor module connected to the PowerLab system (ADI Instruments) at a sampling rate of 1000 Hz. A custom Python-based pipeline was developed for signal preprocessing, feature extraction, and classification.

B. Participants

Participants were classified into high- and low-risk occlusion groups based on whether they had undergone two or more vascular access interventions within the past three months. All subjects were recruited from the Dialysis Center of the Department of Nephrology, National Cheng Kung University Hospital, Douliu Branch. Data were collected both pre- and post-surgery. Additionally, healthy adult data were included to facilitate transfer learning pre-training.

Eligible participants had received either autogenous arteriovenous fistula (AVF) or synthetic graft (AVG) placement in one arm and had no history of cardiovascular disease.

C. PPG Signals and Experimental Features

In this study, photoplethysmography (PPG) was utilized as the primary physiological signal due to its ability to capture dynamic blood flow characteristics. To ensure precise analysis, we extracted a set of clinically relevant and physiologically meaningful waveform features.[1,2]

Feature	Definition
Systolic Peak	Maximum amplitude of the systolic phase in the PPG waveform.
Diastolic Peak	Maximum amplitude during the diastolic phase.
Cardiac Cycle	Integral area under one PPG cycle.
Cycle Area	Time between successive PPG peaks.
SSI	The time from one PPG peak to the next PPG peak
Rise Time	Time from waveform valley to systolic peak.
Systolic Peak Height	Vertical distance from baseline to systolic peak.

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

Delta T	Time interval between systolic and diastolic peaks.
1st Derivative Cycle	Duration of one cycle in the first derivative of the PPG.
1st Derivative Peak	Maximum value in the first derivative waveform.
Ratio b/a, c/a, d/a	Relative amplitudes of characteristic points b, c, d to a in the second derivative of the PPG (SDPPG).
Ratio (b-d-c-e)/a, (b-e)/a	Composite ratios capturing curvature and morphological complexity in the SDPPG waveform.

III. RESEARCH METHODS

A. Signal Preprocessing

PPG signals were preprocessed using a fourth-order Butterworth filter (0.7–9 Hz) due to its smooth frequency response. Filtered signals were segmented into two-cycle intervals based on waveform periodicity, effectively augmenting the dataset for downstream classification.

B. PPG feature extraction

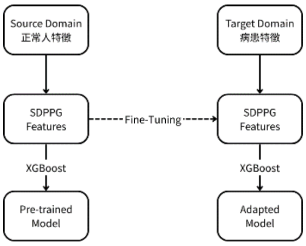
This study focuses on second-order derivative features of PPG signals to capture subtle hemodynamic variations. To enhance model performance, first-order and original waveform features were also incorporated, resulting in a more robust and discriminative feature set for classifying high- and low-risk occlusion groups.

C. Waveform Quality Screening

To ensure data quality and analytical reliability, three classification levels were applied to the PPG waveforms: (1) high-quality signals suitable for analysis, (2) waveforms with weak second-order derivative features that remain manually identifiable, and (3) signals heavily contaminated by motion artifacts and deemed unusable. Each waveform was evaluated using predefined criteria to determine its eligibility for further processing.

D. Transfer Learning

The proposed method initially extracts hemodynamic features from second-order derivative PPG (SDPPG) signals of healthy individuals. An XGBoost classifier is trained on these features to construct a robust baseline model. This model is then fine-tuned via transfer learning to align with the feature distribution of patient-specific PPG data, resulting in a customized occlusion risk prediction model.[3,4]

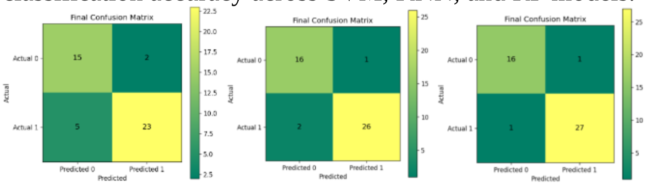


IV. RESULT

PPG feature data from 32 participants (11 high-risk, 21 low-risk) were analyzed using the Mann–Whitney U test. Subsequently, the optimal feature subset identified through exhaustive search was used to train and evaluate three supervised classifiers: 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 p-values less than 0.05 were selected for classification. The optimal feature subset included systolic peak, diastolic peak, cardiac cycle, SSI, Delta_T, first derivative metrics, and key second-order derivative ratios (e.g., Ratio_BA, Ratio_CA, Ratio_DA, Ratio_BDCE_A). These features yielded high classification accuracy across SVM, KNN, and RF models.

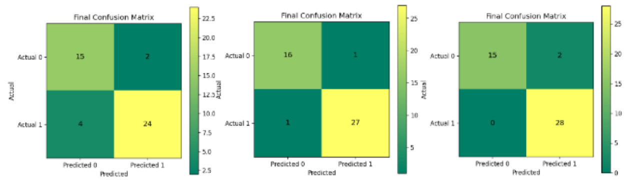


(a)SVM (b) KNN (c) RF

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

B. Exhaustive search

An exhaustive feature selection process was conducted to identify the subset yielding the highest classification accuracy. The optimal features included Diastolic Peak, Systolic Peak, Delta_T, Rise Time, and key second-order derivative ratios such as Ratio_BA, Ratio_CA, Ratio_DA, Ratio_CDB_A, and Ratio_BDCE_A. These features were subsequently used for model training with SVM, KNN, and RF classifiers



(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] N. H. Ab Hamid and M. Y. Nayan, "Analysis on Four Derivative Waveforms of Photoplethysmogram for Peak Detection," *Sensors*, vol. 22, no. 15, p. 5332, 2022

[2] S. Bansal and M. Singh, "Automatic Feature Extraction in Acceleration Plethysmography," *Procedia Computer Science*, vol. 46, pp. 567–574, 2014

[3] J. C. Vásquez-Correa et al., "Convolutional Neural Networks and a Transfer Learning Strategy to Classify Parkinson's Disease from Speech in Three Different Languages," *arXiv preprint arXiv:2002.04374*, 2020.

[4] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, CA, USA, 2016, pp. 785–794