Vascular Ageing Prediction with Visibility Graphs

Jiaqi Yan

{kaede.interpreter@outlook.jp}

Department of Life & Health Sciences, Faculty of Life & Environmental Sciences, Teikyo University of Science (Senju Campus), Tokyo, Japan

Abstract. Vascular ageing is a crucial metric of cardiovascular system health. Numerous studies have explored predicting vascular ageing using Photoplethysmography (PPG) signals and employing deep learning techniques. Nevertheless, these studies exhibit limitations such as reliance on human-involved processing, susceptibility to signal corruption, and dependence on signal amplitude. In this study, we propose a novel approach for detecting vascular ageing with PPG signals. The proposed algorithm combines visibility graphs with deep learning, offering a robust estimation with affine-invariant and amplitude-independent characteristics. We tested our method on multiclass classification, binary classification, and age regression. Our model has demonstrated superior performance compared to other well-established baselines.

Keywords: Vascular Aging, Photoplethysmography, Deep Learning, Visibility Graphs.

1 Introduction

Vascular ageing is an important biomarker reflecting the overall health of the human cardiovascular system, characterized by structural and functional changes in blood vessels. These changes are central to the development of a wide range of age-related cardiovascular diseases, such as hypertension, atherosclerosis, and heart failure.

Meanwhile, with the rapid development of portable sensors, Photoplethysmography (PPG) sensors have been widely integrated into devices such as smartwatches to monitor cardiovascular health by measuring blood volume changes in the microvascular bed of tissue. PPG is a photo-optical, non-invasive technique for measuring hemodynamics and has become a more frequently used diagnostic tool for cardiovascular diseases [1]. PPG utilizes light emitted through an LED in the green optical spectrum to measure tissue volume changes beneath the skin's microvasculature. The light intensity reflected from or transmitted through the skin is detected by the photodetector, typically a photodiode or image sensor, resulting in the PPG waveform, which is crucial for diagnosis [1]. Specifically, alterations in the morphology of the PPG waveform, along with certain features, have been demonstrated to reflect the arterial ageing process and the increase in arterial stiffness, enabling assessment of the patient's vascular health condition in the future [2]. Consequently, pulse waveform changes occur as arteries lose elasticity with age, such as a reduction in the amplitude of peripheral waves and increased stiffness in the upstrokes of the primary pulse [2].

Previous methods for the vascular ageing assessment using PPG data include manually defining mathematical characteristics of the waveform after applying signal processing on the raw

waveform before feeding it to the machine learning classification or regression models. Common features include waveform amplitudes, the proportion between the peaks and valleys, timing characteristics, and contour characteristics of the waveform and its derivative, which are intended to capture the ageing effects of the waveform [2][3][4]. For example, Pilt et al. used signal processing to manually select 12 features in the second derivative PPG waveform related to wave amplitudes and contour points.[3] Implementing multiple linear regression with these parameters yielded an R2 variance of 0.78-0.87 for estimating arterial stiffness indices relative to chronological age.

Thus, due to its convenience and importance, it is natural to bring up the idea of diagnosing vascular ageing using PPG signals. Among them, deep learning approaches have attracted much attention due to the rapid progression of deep learning algorithms. Convolutional neural networks (CNNs), for example, are of relatively recent development and use deep learning, hence enabling direct extraction of robust hierarchical representations from the raw sensor data without much need for feature extraction [5]. The learned filters and accumulated feature maps enable CNNs to preserve the key features connected to the target output at different scales and locations in the input signal.

However, these algorithms share different limits, such as heavy reliance on human-crafted features, amplitude dependence, and sensitivity to affine transformation. This work proposes a novel approach for detecting vascular ageing with PPG signals. The proposed algorithm combines visibility graphs with deep learning, providing a robust estimation with affine-invariant and amplitude-independent characteristics.

2 Related work

2.1 Early Studies Using Manual Feature Extraction

Yousef et al. were among the first authors to investigate the relationship between study variables in a study involving 60 young and elderly subjects [2]. The researchers defined and extracted ten time-domain and amplitude features from the PPG waveform, including the ratios of the highest peaks and deepest troughs. Logistic regression analysis of the manually extracted features yielded R2 values ranging from 0.71 to 0.79 for predicting arterial compliance, indicating the features that captured age-related changes. Pilt et al. conducted pulse waveform analysis on 65 subjects aged 18 to 75 years [3]. Twelve waveform features were identified from the second derivative of the PPG waveform (SDPPG), including contour points and wave amplitude. By using manually defined features for linear regression, they achieved R2 values in the range of 0.78 to 0.87 for estimating arterial stiffness indices [6]. Similar research was carried out by Ahn [4] with 111 participants aged 20-39 years and 60-79 years. Fifteen time and amplitude features were manually extracted from the PPG waveform and its first derivative. Correlation analysis of the calculated "ageing index" with chronological age ranged from 0.69 to 0.74 [7].

2.2 Algorithm Development for Feature Extraction

Another well-known study by Pilt et al. marked a significant advancement by introducing a new method to derive quantitative characteristics of plethysmographic waveforms from PPG for

estimating ASI values [3]. The researchers involved 65 young and elderly male participants aged between 18 and 75 years with no history of cardiovascular diseases. Blood samples were collected through pulsed oximetry from the left index finger of the subjects in a seated position using the pulse oximeter system commonly available in the market. In this study, a new subject-specific algorithm was developed and utilized to extract multiple parameters that characterize discriminative information from the SDPPG, which represents the second derivative of the raw PPG waveform. This process involved employing algorithms to derive the SDPPG traces from the raw PPG signals to enhance the clearly identifiable contour variations that are more prominent in the derived waveform compared to the primary derivative of the waveform alone. In total, 12 features, including amplitudes and the precise positions of significant points within the SDPPG waveform, were obtained using signal processing methods to identify these features.

Linear regression models were then developed using the 12 manually extracted SDPPG features as predictor variables and the paired arterial stiffness indices obtained from the brachial-ankle PWV as dependent variables. The extracted features exhibited high R2 values in the range of 0.78-0.87, indicating their strong potential to predict arterial stiffness accurately in relation to chronological age. The developed preprocessing algorithm was considered superior in its ability to facilitate the extraction of parameters sensitive to vascular ageing from PPG signals by providing numerical descriptions of waveform features. This established methodology set a reference point for further related research investigations.

2.3 Computation of an "Aging Index"

Tang et al. introduced a vascular ageing index based on the PPG signals of 100 subjects aged 20 to 80 years. Initially, the PPG signal and its derivative were analyzed, yielding twelve time and frequency domain features [8]. An ageing score was derived from these features using a machine learning model, showing a negative correlation with actual age and a reliability of 0.89. Srinivasan et al. examined PPG signals from 120 individuals categorized as young (aged 18-30 years) and older (aged 60-80 years) [9]. They developed a deep learning model directly on the raw PPG signal to classify subjects as young or old, achieving an accuracy of 85%. Shin et al. gathered PPG signals from 90 participants aged 25 to 75 years [10]. They extracted a total of 20 time and frequency domain features and averaged them to compute an ageing score using a support vector machine. This approach showed a strong association with chronological age, with a correlation coefficient of 0.88. Charlton et al. conducted a study on 86 healthy subjects aged 20 to 70 years, analyzing recorded PPG signals [11]. They nonlinearly combined the twelve time and frequency domain features into a vascular ageing biomarker, which exhibited a positive correlation with age (r = 0.92, significance level = 0.000).

2.4 Employment of Manual Feature Extraction

It is worth pointing out that in the case of the works reviewed in this paper by Yousef et al., Pilt et al., and Ahn regarding the PPG-based vascular ageing prediction, it originated from the manual identification of predetermined quantitative parameters of the raw signal [2][3][4]. This preprocessing was intended to capture the ageing effect changes in a minimum yet effective dimensionality of features. In all the investigations, the number of extracted features was within ten to fifteen based on the PPG waveform's time, frequency, and amplitude characteristics and its extension. Extracted features often used included ratios of pulse points, peak amplitudes, timing values, and the relative position of contours on the waveforms. Although this approach

worked well, it became apparent that several issues arose in machine learning related to scalability, complexity of features, and the kinds of discriminative information that can be extracted from raw data in a preprocessing stage in a feasible manner. Equally, the laborintensive approach to sample analysis also limited rigorous validation and clinical translation capacity.

2.5 Potential for Deep Learning Approaches

Advanced deep-learning methods like CNNs alleviate the need to perform feature extraction on PPG signals using manual preprocessing. Without the need for manually extracting features, CNNs perform feature extraction through raw signal waveforms and learn hierarchal representations that best suit the particular prediction task at hand. This systematic analysis utilizes several layers: convolution, activation, and pooling. Convolutional filters that are run across the input at multiple resolutions pull implicit local features, activation and pooling down samples' spatial and temporal dimensions for salient embedded characteristics. Fully connected layers are the subsequent layers, allowing for classification or regression modelling based on aggregated convolutional features.

CNNs learn their weights and architectures through backpropagation and gradient descent training algorithms, thus creating end-to-end learning hierarchies of enhancing abstract representations directly from raw input data [12]. This avoids the limitations of heuristically selecting features for use in a model by using significantly more discriminative information that is spread out across the whole input signal. In the case of vascular ageing analysis, CNNs can potentially obtain reliable latent PPG-derived indices associated with slow waveform variations superior to hand-crafted features. Previous studies showed that CNN-based models had better results compared to standard machine learning on the raw medical time series [13]. Future application of CNNs to raw multichannel PPG may improve vascular status assessment based on DFL without an exhaustive preprocessing framework.

3 Methods

3.1 Graphs

Graphs are robust mathematical concepts that represent intricate relationships between entries in diverse domains. Consisting of vertices and edges, graphs can depict relationships between vertices (such as dependencies) by connecting them with edges. Graphs find applications in various fields, including finance, social networks, and biological systems. Henceforth, we will denote the graph as $\mathbf{G} = (\mathbf{V}, \mathbf{E})$, where *V* stands for vertices and *E* stands for edges. Figure 1 illustrates an example graph with vertices and edges.



Fig. 1. A graph with vertices and edges.

Adjacency matrices encapsulate all the connection information of the corresponding graphs. For a graph of N vertices, the corresponding adjacency matrix A is of shape $N \times N$. If two vertices v_i and v_j are connected, then $A_{i,j} = w_{i,j}$ where $w_{i,j}$ is the corresponding edge weight between vertices v_i and v_j . For this work, all graphs will be unweighted and bidirectional, which means all edge weights will be one or zero, and the adjacency matrix will be symmetric over its diagonal line.

3.2 From PPG Signals to Graphs

PPG is famous for its convenience and non-invasiveness. However, it suffers badly from the vulnerability of light sensors, leading to corruptions like baseline wandering. Also, the magnitudes of the PPG signals are highly sensitive to factors that do not reflect cardiovascular health, such as skin thickness and skin tones, increasing the difficulty of extracting cardiovascular-related information. Although common approaches like min-max normalization could rearrange the signal magnitudes, the signal morphology is inevitably changed. Thus, it is necessary to develop a method to process the PPG signals, which shows benefits including:

- No need for manual feature extraction.
- Amplitude invariant.
- Show robustness to corruption.

Thus, we introduce a visibility graph to transform 1D time series signals into graphs. The visibility graphs encode structural information from the time series signals and discard the amplitude-related information. Given a time series signal y, each signal point y_i is now converted into a vertex. Two vertices will be connected if a line connecting two signal points will no intersect with a third signal point. In other words, one signal point will see another one without being blocked by a third signal point, corresponding to the visibility term. More formally, two vertices (signal samples), y_a and y_b , are connected by an edge if

$$y_c < y_b + (y_a - y_b) \frac{t_b - t_c}{t_b - t_a}$$
(1)

for any other signal sample y_c , where t_a , t_b and t_c are the time indices corresponding to the signal samples y_a , y_b and y_c . It has been proven that the visibility graph can encode structural information [14][15], including:

- Periodic time series are converted into regular graphs, where the degree distribution correlates with the signal's periodicity.
- Random time series will be converted into graphs that resemble exponential random graphs.

• Fractal series are converted into scale-free networks, showcasing the algorithm's effectiveness in maintaining intricate statistical features like self-similarity and scale invariance within the graph structure. The method also successfully identifies the hub repulsion phenomenon typical of fractal networks, allowing for a clear distinction between scale-free visibility graphs that demonstrate the small-world effect and those that preserve scale invariance.

Thus, time series PPG signals can now be transformed into visibility graphs, which are amplitude-independent but preserve structural information. In this way, we can also extract import geometry information such as time delay, stiffness index, dicrotic notch, etc.

3.3 Convolutional Neural Networks and Graph Convolutional Neural Networks

Convolutional Neural Network (CNN) is a type of neural network that specifically works with regular-structure data, such as images and 1D signals. CNNs use a series of convolutional layers to automatically learn hierarchies of spatial features. The extracted features are further processed for future tasks. CNNs are naturally suitable to work on 1D PPG signals for vascular ageing detection due to the regular structure of PPG time series.

However, CNNs do not work on irregular data domains such as graphs. For graphs, each vertex has a non-fixed number of neighbours (connected nodes), leading to a dynamic kernel size. Thus, graph convolution has been introduced to compensate for this inadequacy.

Graph convolution adapts the idea of convolution to the graph domain by aggregating information from a vertex's neighbours based on graph's connectivity. This resembles the convolution ideas in CNNs to aggregate local spatial features for further actions, including but not limited to classification, regression, generation, etc. Typical graph convolutional neural networks include GCN, GraphSAGE, GAT, etc. In this work, we will utilize ChebyConv networks, which will be further introduced in the chapter afterwards.

3.4 Chebyshev Convolution (ChebyConv)

The Chebyshev Convolution (ChebyConv) is a graph convolutional layer designed to improve the efficiency of spectral graph convolutional networks (GCNs) by leveraging the properties of Chebyshev polynomials. Spectral GCNs traditionally define convolution operations in the spectral domain by applying filters to the eigenvalues of the graph Laplacian matrix, $\mathbf{L} = \mathbf{D} - \mathbf{A}$, where **A** is the adjacency matrix of the graph and **D** is the degree matrix. However, computing the eigendecomposition of the Laplacian, which requires $\mathcal{O}(N^3)$ operations, is computationally expensive for large graphs, where N is the number of nodes.

To address this, ChebyConv approximates the spectral filter $g_{\theta}(\mathbf{L})$ using a truncated expansion in terms of Chebyshev polynomials $T_k(\tilde{\mathbf{L}})$ up to order K, where $\tilde{\mathbf{L}} = \frac{2}{\lambda_{\text{max}}} \mathbf{L} - \mathbf{I}$ is the scaled and normalized Laplacian, and λ_{max} is the largest eigenvalue of **L**. The spectral filter can be expressed as:

$$g_{\theta}(\mathbf{L}) \approx \sum_{k=0}^{K} \theta_k T_k(\tilde{\mathbf{L}}), \qquad (2)$$

where θ_k are the trainable parameters of the model.

The Chebyshev polynomials are defined recursively as:

$$T_0(\tilde{\mathbf{L}}) = \mathbf{I}, \quad T_1(\tilde{\mathbf{L}}) = \tilde{\mathbf{L}}, \quad T_k(\tilde{\mathbf{L}}) = 2\tilde{\mathbf{L}}T_{k-1}(\tilde{\mathbf{L}}) - T_{k-2}(\tilde{\mathbf{L}}) \quad \text{for}k \ge 2.$$
(3)

Thus, the convolution operation on the node feature matrix $\mathbf{X} \in \mathbb{R}^{N \times F}$, where *F* is the number of input features per node, is given by:

$$\mathbf{Z} = \sum_{k=0}^{K} \theta_k T_k(\tilde{\mathbf{L}}) \mathbf{X},\tag{4}$$

where $\mathbf{Z} \in \mathbb{R}^{N \times F'}$ is the output feature matrix with F' output features per node.

This approach reduces the computational complexity from $O(N^3)$ for the eigendecomposition to O(K|E|) for the Chebyshev polynomial expansion, where |E| is the number of edges in the graph and K is typically much smaller than N. The ChebyConv layer efficiently captures multihop information within the graph structure while remaining computationally scalable by restricting the filter to be localised. This makes ChebyConv particularly suitable for large-scale graph learning tasks, where balancing computational efficiency and expressive power is crucial.

3.5 VasVG

Now we introduce the proposed model VasVG. The VasVG contains four ChebBlock modules and final classification layers. We will now denote a graph as *G* and its node feature as $F \in \mathbb{R}^{n \times d}$, where *n* is the number of nodes and *d* is the dimension of the feature.

ChebBlock. Each ChebBlock module consists of three ChebyConv layers stacked together, namely f_1 , f_2 and f_3 . The input graph node feature F is processed as:

$$F' = \sigma\left(f_3\left(G, \sigma\left(f_2\left(G, \sigma\left(f_1(G, F)\right)\right)\right)\right)\right)$$
(5)

where $F' \in \mathbb{R}^{n \times d_h}$. d_h represents the feature dimension after the third ChebyConv layer. Then, three graph global pooling layers, namely max pooling, sum pooling and average pooling, are further applied to give three graph embeddings of dimension $1 \times d_h$, and they are concatenated to form the final embedding of dimension $1 \times 3d_h$. The final embedding of the four ChebyConv layers is concatenated together to form the final graph embedding $F_{final} \in \mathbb{R}^{1 \times 12d_h}$.

Classification Layer. The F_{final} graph embedding will be fed to an MLP with the output size of the final number of classes.

4 Experiment

4.1 Vascular Ageing Prediction

For the first experiment, we aim to predict vascular ageing through PPG signals and visibility graphs.

4.2 Data prerparation

In this study, we utilized the Real-World PPG dataset, comprising PPG signals from 35 subjects captured using an IoT sensor. The subjects' ages range from 10 to 75 years old. We structured our investigation into three tasks: binary classification, multiclass classification, and age

regression. For binary classification, we categorized the subjects into two age groups: under 30 years old and 30 years old and above, aiming to predict these distinctions. Additionally, we segmented the subjects into four age brackets: 0-20, 20-30, 30-40, and 40+, conducting a fourclass classification task. Finally, we evaluated our model's performance on a regression task to predict the ages of the subjects.

No preprocessing was conducted as visibility graphs are amplitude-invariant and exhibit robustness to noise. The PPG signals were segmented into individual pulses, and corresponding visibility graphs were constructed. White noise was introduced as node features. The resulting graph, combined with the node features, was input into the VasVG model to generate the final predictions.

4.3 Preprocessing

Preprocessing transforms the signal's raw form into forms the algorithm can directly work on [16]. The signal is first detrended using the moving average method. Afterwards, the Hilbert transform is performed to demodulate the input signal. Furthermore, the demodulated signal is divided by the extracted envelope and normalized.

In contrast to this cumbersome preprocessing pipeline, VasVG requires no preprocessing. The **FindPeaks** function in Matlab is applied to extract pulses, and no further action is taken before feeding the data into the VasVG model.

4.4 Results

Here, we introduce the performance of our proposed VasVG model on binary classification, multiclass classification and age regression. We compare the results with the baseline models proposed by Dall'Olio *et al.* [16] and Chiarelli *et al.* [17].

Table 1. The performance of our proposed VasVG mode and the baseline model proposed by Dall'Olio *et al.* on Binary classification

Method	Accuracy: Four Classes
VasVG	94.3%
Dall'Olio et al.[16]	93.5%

Table 2. The performance of our proposed VasVG mode and the baseline model proposed by Dall'Olio
 et al. on multiclass classification

Method	Accuracy: Four Classes
VasVG	89.35%
Dall'Olio et al.[16]	87.14%

Table 3. The performance of our proposed VasVG mode and the baseline model proposed by Chiarelli *et al.* on age regression.

Method	Accuracy: Four Classes
VasVG	2.85%
Chiarelli et al.[17]	3.5%

Our model achieves better performance in terms of vascular ageing prediction, outperforming the well-established previous baseline in both binary and multiclass classification, as shown in Table 3. We have also demonstrated superior performance in the age regression tasks. Moreover, our model exhibits the advantage of faster training speed. We employed an early-stopping training method, where the training process halts if the training loss fails to decrease for five consecutive epochs. Our model has shown significantly faster convergence speed.

5 Conclusion

In summary, our proposed VasVG model significantly advances the use of PPG signals in predicting vascular ageing, surpassing established baseline models while reducing training time. The combination of visibility graphs with deep learning enhances accuracy and eliminates complex pre-processing steps, rendering the method more efficient and robust. However, the study has limitations, including the potential for overfitting to diverse datasets and the necessity for extensive validation across different populations. Future work will address these limitations through the integration of larger, more diverse datasets and by exploring the model's applicability in real-world clinical settings to further validate its effectiveness in predicting vascular health.

References

- John Allen. Photoplethysmography and its application in clinical physiological measurement. Physiological measurement, 28(3):R1, 2007.
- [2] Q Yousef, MBI Reaz, and Mohd Alauddin Mohd Ali. The analysis of ppg morphology: investigating the effects of aging on arterial compliance. Measurement Science Review, 12(6):266–271, 2012.
- [3] Kristjan Pilt, Rain Ferenets, Kalju Meigas, Lars-Göran Lindberg, Kristina Temitski, and Margus Viigimaa. New photoplethysmographic signal analysis algorithm for arterial stiffness estimation. The scientific world journal, 2013(1):169035, 2013.
- [4] Jae Mok Ahn. New aging index using signal features of both photoplethysmograms and acceleration plethys- mograms. Healthcare informatics research, 23(1):53–59, 2017.
- [5] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. Journal of big data, 6(1):1–48, 2019.
- [6] Y Tardy, JJ Meister, F Perret, HR Brunner, and M Arditi. Non-invasive estimate of the mechanical properties of peripheral arteries from ultrasonic and photoplethysmographic measurements. Clinical Physics and Physiological Measurement, 12(1):39, 1991.
- [7] Wei Xia, Yu Zhou, Xingchen Yang, Keshi He, and Honghai Liu. Toward portable hybrid surface electromyography/a-mode ultrasound sensing for human–machine interface. IEEE Sensors Journal, 19(13):5219–5228, 2019.
- [8] Qingfeng Tang, Zhiqiang Pan, Changlong Tao, Jing Jiang, Benyue Su, Hui An, Guodong Liu, and Zhigeng Pan. Vascular age acquired from the pulse signal: A new index to screen early vascular aging. Computers in Biology and Medicine, 151:106355, 2022.
- [9] Venkatesh Bharadwaj Srinivasan and Foroohar Foroozan. Deep learning based non-invasive diabetes predictor using photoplethysmography signals. In 2021 29th European Signal Processing Conference (EUSIPCO), pages 1256–1260. IEEE, 2021.

- [10] Hangsik Shin, Gyujeong Noh, and Byung-Moon Choi. Photoplethysmogram based vascular aging assessment using the deep convolutional neural network. Scientific Reports, 12(1):11377, 2022.
- [11] Peter H Charlton, Birute' Paliakaite', Kristjan Pilt, Martin Bachler, Serena Zanelli, Daniel Kulin, John Allen, Magid Hallab, Elisabetta Bianchini, Christopher C Mayer, et al. Assessing hemodynamics from the photoplethys- mogram to gain insights into vascular age: a review from vascagenet. American Journal of Physiology-Heart and Circulatory Physiology, 322(4):H493–H522, 2022.
- [12] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436–444, 2015.
- [13] S Suganyadevi, V Seethalakshmi, and Krishnasamy Balasamy. A review on deep learning in medical image analysis. International Journal of Multimedia Information Retrieval, 11(1):19–38, 2022.
- [14] Bartolo Luque, Lucas Lacasa, Fernando Ballesteros, and Jordi Luque. Horizontal visibility graphs: Exact results for random time series. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 80(4):046103, 2009.
- [15] Mutua Stephen, Changgui Gu, and Huijie Yang. Visibility graph based time series analysis. PloS one, 10(11):e0143015, 2015.
- [16] Lorenzo Dall'Olio, Nico Curti, Daniel Remondini, Yosef Safi Harb, Folkert W Asselbergs, Gastone Castellani, and Hae-Won Uh. Prediction of vascular aging based on smartphone acquired ppg signals. Scientific reports, 10(1):19756, 2020.
- [17] Antonio M Chiarelli, Francesco Bianco, David Perpetuini, Valentina Bucciarelli, Chiara Filippini, Daniela Car done, Filippo Zappasodi, Sabina Gallina, and Arcangelo Merla. Data-driven assessment of cardiovascular ageing through multisite photoplethysmography and electrocardiography. Medical engineering & physics, 73:39–50, 2019