Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Explainable Machine-Learning Analysis of Carotid Doppler Waveform Biomarkers for Chronological-Age Estimation

Azran Azhim

azran@kct.ac.jp

National Institute of Technology Kitakyushu College https://orcid.org/0009-0007-8625-7681

Research Article

Keywords: Carotid Arteries, Ultrasonography, Doppler, Machine Learning, Biomarkers, Aging

Posted Date: November 1st, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11505581/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: The authors declare no competing interests.

Abstract

Background

Physiological sensing and machine learning are increasingly combined for health assessment, but many systems emphasize predictive accuracy over physiologically grounded explanation.

Objective

To describe a small explainable biomarker-discovery framework linking signal processing, explainable machine learning, and physiological interpretation, illustrated and stress-tested with carotid Doppler waveforms.

Methods

In 227 healthy adults (20–89 years) with portable carotid Doppler recordings, eight waveform features were related to chronological age using Pearson correlation and Random Forest (RF) regression with repeated 10-fold cross-validation. Feature importance was assessed by Gini and permutation methods (in-sample and held-out), and ablation quantified the unique contribution of the top-ranked feature, the Velocity Reflection Index (VRI).

Results

VRI correlated most strongly with age ($r = 0.739$, $p < 0.001$). The RF model estimated age with MAE 8.27 ± 1.40 years and $R^2 = 0.609 \pm 0.140$. VRI dominated both Gini importance (50.4%) and held-out permutation importance (an even larger share), yet removing VRI and retraining cost only 0.043 R^2 (0.609 to 0.566), since VRI correlates strongly with other retained landmarks ($r = -0.64$). A missingness-bias check showed complete-body-composition participants were significantly younger than others (31.4 vs. 45.7 years, $p < 0.001$).

Conclusion

Carotid Doppler waveforms carry substantial, explainably identifiable, age-related information, but a single importance metric can overstate a correlated feature's unique contribution; combining importance methods with ablation gives a more defensible account.

1. Introduction

Advances in physiological sensing now allow continuous acquisition of signals such as electrocardiograms, photoplethysmograms, blood-pressure waveforms, and Doppler blood-flow-velocity waveforms for health monitoring, disease screening, and personalized healthcare [1–6], and artificial intelligence (AI) has accelerated the extraction of clinically relevant information from these datasets [7–8].

A persistent challenge is translating physiological measurements into information a clinician can interpret: many AI-based monitoring systems emphasize predictive accuracy while giving little insight into why an output was produced, limiting trust and adoption [9–10]. This paper does not attempt that broad challenge, which would require evaluation with clinicians or patients; it instead adopts a narrower aim — a small conceptual framework connecting signal processing, explainable machine learning, and biomarker interpretation — illustrated with one worked example: carotid Doppler biomarkers of age-related hemodynamic change.

Carotid Doppler ultrasonography was chosen for comparatively simple, portable instrumentation [11–12], and because its waveforms carry hemodynamic information related to arterial stiffness, vascular resistance, wave reflection, and cardiovascular aging — why these changes matter clinically is established by a separate tonometry/pulse-wave-velocity literature [13–17], cited for clinical motivation only. The Doppler-specific claim that these waveforms carry age-related information rests instead on the earlier report of this methodology and dataset [18].

General advances in explainable machine learning [19–21] provide tools, in principle, for identifying which waveform features carry the most information about an outcome, but Random Forest importance scores are known to be inflated for predictors correlated with other predictors in the model [22]; a further aim here is to test, empirically, whether that caveat holds for this feature set.

Most existing work on physiological sensing treats signal acquisition, biomarker development, predictive modeling, and device engineering as separate problems. Relatively little work connects explainable analysis of a physiological signal to biomarker-level interpretation within one traceable pipeline, and fewer still test whether the resulting explanation survives a second, independent explainability check. This paper describes a small framework — the Explainable Biomarker-Discovery Framework — connecting signal measurement, feature extraction, explainable machine learning, and biomarker interpretation, and demonstrates and stress-tests it using carotid Doppler waveforms. The contributions are:

1) a small conceptual framework connecting signal processing, explainable machine learning, and physiologically interpretable biomarker discovery;

2) a worked illustration of that framework using carotid Doppler waveforms and three established hemodynamic indices (RI, VRI, VEI), evaluated in 227 adults;

3) a multi-method explainability stress test — combining impurity-based importance, permutation importance, and ablation — that shows these methods can disagree sharply on how “important” a correlated feature really is, with direct consequences for what an explainable-AI pipeline should be allowed to claim; and

4) an explicit accounting of what this illustration does and does not establish — in particular, that the resulting age estimate is a Doppler-based chronological-age estimate, not an independently validated vascular-age biomarker.

2. Materials and Methods

This was a cross-sectional secondary analysis of a previously collected dataset of carotid Doppler waveforms [18]; it did not involve new data collection, and the analysis plan was organized around the five-stage Explainable Biomarker-Discovery Framework summarized below and in Fig. 1. Reporting follows the STROBE statement for cross-sectional observational studies [23].

2.1 Explainable Biomarker-Discovery Framework

The framework keeps physiological meaning attached to each analytical step rather than treating the signal purely as model input [19–21]. It has five linked stages: (i) signal acquisition; (ii) feature extraction; (iii) explainable analysis of feature contributions; (iv) biomarker-level interpretation; and (v) reporting of an interpretable, qualified estimate (Fig. 1). Unlike a pipeline optimizing predictive performance alone, biomarker-level interpretation is an explicit intermediate step, and — as implemented here — an explainable-analysis result must be checked against at least one additional, independent method before being reported as a finding (Section 2.5).

2.2 Study Population

This cross-sectional dataset included 233 healthy adults aged 20–89 years who had voluntarily participated in community-based health screening. Complete carotid Doppler biomarker data were available for 227 participants (136 males [59.9%], 91 females [40.1%]; mean age 40.2 ± 18.3 years), used in the analyses below; six participants were excluded for incomplete recordings or missing demographic data, and participants with a known history of cardiovascular disease, arrhythmia, diabetes mellitus, or treated hyperlipidemia were also excluded. Baseline anthropometric and hemodynamic characteristics, where available, are reported in Section 3.1 and Table 1; these measurements were not complete for the full cohort, and the implications are examined directly in Section 3.6.

Participants gave written informed consent at the time of original data collection, under approval from the Ethics Committee of Tokushima University Hospital, Tokushima, Japan, as previously reported in Azhim et al. [18], consistent with the Declaration of Helsinki.

2.3 Physiological Signal Measurement

Blood-flow-velocity waveforms were obtained from the left common carotid artery using a portable Doppler ultrasound system operating at a center frequency of 2 MHz. The Doppler probe was positioned over the common carotid artery with an insonation angle of approximately 50°. The Doppler audio signal was digitized at a sampling frequency of 10 kHz and stored for subsequent analysis. The sample-volume setting and the procedure used to confirm the insonation angle at acquisition time follow the protocol described in the original data-collection report and the same group's earlier carotid Doppler measurement studies [24–25]; readers requiring that level of instrumentation detail should consult those sources.

Blood velocity was estimated from the standard Doppler frequency-shift equation [6]:

$V = cf_d / (2f_0 \cos \theta)$ (1)

where V is blood velocity, c is the speed of sound in tissue, fd is the Doppler frequency shift, f0 is the transmitted ultrasound frequency, and θ is the insonation angle.

2.4 Signal Processing and Biomarker Extraction

Fast Fourier Transform (FFT) analysis was applied to generate a time-frequency spectrogram of the Doppler signal, and a threshold-based envelope-detection method extracted the maximum-velocity contour, giving a beat-to-beat blood-flow-velocity waveform. From this waveform, eight landmark and derived features were used in the modeling reported below: the primary systolic peak (S1), secondary systolic peak (S2), incisura velocity (I), peak diastolic velocity (D), end-diastolic velocity (d-min), and the three composite indices RI, VRI, and VEI derived from them. RI, VRI, and VEI are the three indices singled out for biomarker-level interpretation in Section 3.7; the Random Forest model itself was trained on the full eight-feature set.

Resistance Index, following the classical formulation of the systolic/end-diastolic velocity ratio [26]:

RI = (S1 − d-min) / S1 (2)

The Velocity Reflection Index (VRI) and the Vascular Elasticity Index (VEI) were first proposed by the author in the 2011 report on this cohort's carotid blood-flow-velocity envelope [18]. Velocity Reflection Index is interpreted descriptively below as a *wave-reflection-related index* rather than a direct measurement of cerebral wave reflection (Section 4.4):

VRI = S2 / S1 − 1 (3)

Because S2 is normally smaller than S1, VRI as defined above is normally negative; its positive correlation with age (r = 0.739, Section 3.2) therefore means VRI becomes less negative — moves toward zero, and in some participants crosses to positive — with advancing age, not that its absolute magnitude increases from zero.

Vascular Elasticity Index, interpreted descriptively below as a *waveform-derived elasticity-related index* rather than a direct measurement of arterial elasticity:

VEI = 1 − I / D (4)

RI, VRI, and VEI are each calculated from pairs of the same underlying landmarks (S1, S2, I, D, d-min), so they are not statistically independent of one another or of the raw landmarks. The full pairwise correlation matrix among all eight features is reported in Section 3.5 together with the ablation analysis it motivates.

2.5 Statistical and Explainable Analysis

Pearson correlation related each of the eight waveform features to chronological age. Random Forest (RF) regression (500 trees) related all eight features jointly to chronological age, evaluated with repeated 10-fold cross-validation (10 repetitions, 100 runs), reporting MAE, RMSE, and R^2 as mean ± SD. RF was chosen for its ability to capture nonlinear relationships; its oft-cited robustness to multicollinearity is not absolute, a point this study returns to below. All models used scikit-learn 1.9.1 (Python 3), a fixed random-state seed (42) for every fold split and every RF fit, and library-default hyperparameters otherwise (max_features = 1.0, i.e. all features considered at each split; min_samples_leaf = 1; no bootstrap-fraction or depth tuning was performed). The ablation analysis below uses this same configuration, so all cross-validated R^2 figures are directly comparable; analysis code is available on request (Data Availability).

Feature importance was computed three ways, to test whether a single metric could be trusted alone: (i) *Gini importance*, the RF model's mean decrease in impurity, averaged across runs; (ii) *in-sample permutation importance*, from one model fit on the full dataset, shuffling each feature (30 repeats); and (iii) *held-out permutation importance*, within a separate 10-fold cross-validation (20 shuffles per fold, evaluated only on held-out data). Agreement between Gini and held-out permutation rankings was summarized with Spearman rank correlation.

Because both importance methods are known to be unreliable for predictors correlated with others in the model [22], an *ablation analysis* tested the unique contribution of the top-ranked feature (VRI): cross-validated R^2 (10-fold × 10 repeats, 500-tree RF — the same configuration as the main model above) was compared across five feature subsets — all eight; seven with VRI removed; S1 + S2 only; VRI only; and S1 + S2+VRI — so VRI's contribution could be assessed both alone and against cheaper alternatives.

A missingness-bias check compared participants with complete body-composition measurements (n = 88) against those without (n = 139) on age (Welch's t-test) and sex distribution (chi-square test). Sex was recorded for all 227 participants but was deliberately excluded from the RF feature set, which was restricted by design to waveform-derived landmarks and indices (Section 2.4); as a sensitivity check, adding sex as a ninth feature to the all-eight-feature model modestly improved cross-validated performance (R^2 0.609 to 0.632; MAE 8.27 to 7.96 years), indicating sex carries some age-related

information not fully captured by the waveform features alone, consistent with the sex-specific associations noted in Section 3.7. A two-sided $p < 0.05$ was considered significant throughout.

3. Results

3.1 Baseline Characteristics

Table 1 summarizes baseline characteristics, overall and by sex, for the subsets of the cohort with available measurements. Age was available for all 227 participants (mean 40.2 ± 18.3 years; range 20–89; by decade: 92 in their twenties, 36 in their thirties, 20 in their forties, 42 in their fifties, 21 in their sixties, 10 in their seventies, 6 in their eighties). Height, weight, and BMI were available for 145 participants (89 male, 56 female); blood pressure and heart rate for 169 and 227, respectively. Men had higher BMI and systolic blood pressure than women (Table 1), reported descriptively and not as outcomes of the modeling below.

Table 1
Baseline characteristics, overall and by sex (mean ± SD).

Characteristic	Overall	Male	Female
Age, years (n)	40.2 ± 18.3 (227)	40.2 ± 17.9 (136)	40.1 ± 18.9 (91)
BMI, kg/m² (n)	21.4 ± 2.8 (145)	22.2 ± 3.0 (89)	20.2 ± 2.1 (56)
Systolic BP, mmHg (n)	119.6 ± 11.8 (169)	121.8 ± 11.6 (107)	115.8 ± 11.1 (62)
Diastolic BP, mmHg (n)	73.8 ± 9.7 (169)	74.6 ± 10.4 (107)	72.4 ± 8.2 (62)
Mean arterial BP, mmHg (n)	89.1 ± 9.5 (169)	90.3 ± 9.9 (107)	86.9 ± 8.5 (62)
Heart rate, bpm (n)	72.9 ± 10.7 (227)	72.3 ± 10.8 (136)	73.9 ± 10.4 (91)

3.2 Association Between Doppler Biomarkers and Age

Pearson correlation showed significant associations between chronological age and several carotid Doppler features: strongest for the Velocity Reflection Index (VRI) ($r = 0.739$, $p < 0.001$); strong negative associations for the primary systolic peak (S1) ($r = -0.661$), the Vascular Elasticity Index (VEI) ($r = -0.494$), and peak diastolic velocity (D) ($r = -0.489$, all $p < 0.001$); moderate negative associations for end-diastolic velocity (d-min) ($r = -0.297$) and the Resistance Index (RI) ($r = -0.267$, both $p < 0.001$); and no significant association for the secondary systolic peak (S2) ($r = 0.081$, $p = 0.22$). These associations are descriptively consistent with the arterial-stiffness literature's general picture of age-related change in wave reflection and elastic recoil [13–17], cited for physiological context only, since that literature does not itself measure Doppler waveforms (Section 1).

3.3 Doppler-Based Chronological-Age Estimation

A Random Forest regression model using the eight carotid Doppler features estimated chronological age with repeated 10-fold cross-validation (10 repetitions, 100 runs). The model achieved a mean absolute error (MAE) of 8.27 ± 1.40 years, a root-mean-square error (RMSE) of 10.88 ± 1.86 years, and a coefficient of determination (R^2) of 0.609 ± 0.140 (Fig. 2).

These numbers show how well the model recovers chronological age from Doppler features — reported as a *Doppler-based chronological-age estimate*, not a validated vascular-age biomarker. Establishing the latter would require evidence that the estimate (a) reflects vascular status independently of chronological age, (b) correlates with an established vascular measurement such as pulse-wave velocity or carotid intima-media thickness, (c) distinguishes healthy from abnormal vascular aging, (d) relates to a clinical or cardiovascular outcome, and (e) adds information beyond age and conventional risk factors — none of which this cross-sectional, healthy-volunteer dataset can address.

Error was not uniform across age: decade-stratified MAE (same held-out predictions) was 6.1 years for participants in their twenties (n = 92) and 9.3, 7.3, 8.6, and 10.5 years for the thirties through sixties (n = 36, 20, 42, 21), rising to 15.3 years (n = 10) and 16.0 years (n = 6) for the seventies and eighties — the age range where the cohort is smallest (Table 1). These figures are examined further in Section 4.1.

3.4 Feature Importance: Gini Versus Permutation

VRI dominated impurity-based (Gini) feature importance, accounting for 50.4% ± 4.4% of total importance, roughly three times the next-ranked feature, S1 (17.5% ± 4.8%) (Table 2, Fig. 3). Held-out permutation importance agreed VRI led, but by a larger relative margin: its mean R^2 drop when shuffled in held-out folds (0.606 ± 0.118) was roughly fifteen times S1's (0.041 ± 0.034), and RI and d-min were indistinguishable from zero — not evidence those features carry no information, but evidence this method could not detect a contribution from them once VRI was available. In-sample permutation importance showed the same pattern even more strongly (VRI 78.8% ± 7.0%). The rank correlation between Gini and held-out permutation importance was moderate (Spearman $\rho = 0.667$) but did not reach significance ($p = 0.071$).

Table 2
Random Forest feature importance by three methods.

Feature	Gini importance	Permutation (in-sample)	Permutation (held-out CV)
d-min	0.083 ± 0.011	0.077 ± 0.007	0.004 ± 0.063
S1	0.175 ± 0.048	0.140 ± 0.012	0.041 ± 0.034
S2	0.035 ± 0.003	0.049 ± 0.005	0.008 ± 0.016
I	0.053 ± 0.009	0.062 ± 0.006	0.016 ± 0.035
D	0.070 ± 0.012	0.083 ± 0.007	0.015 ± 0.038
RI	0.030 ± 0.003	0.027 ± 0.002	−0.002 ± 0.012
VRI	0.504 ± 0.044	0.788 ± 0.070	0.606 ± 0.118
VEI	0.051 ± 0.007	0.070 ± 0.005	0.018 ± 0.027

3.5 Ablation: VRI's Unique Contribution

Using the same configuration as the main model (Section 2.5), all eight features gave cross-validated R^2 of 0.609 ± 0.140 — identical to Section 3.3, as expected. Removing VRI (seven features) gave R^2 of 0.566 ± 0.156 — a loss of only 0.043, far smaller than the ≈ 50% (Gini) or ≈ 79% (in-sample permutation) importance share VRI was assigned (Table 3, Fig. 4). S1 + S2 alone achieved R^2 of 0.520 ± 0.166, close to the no-VRI result; VRI alone achieved only 0.332 ± 0.253; and S1 + S2+VRI achieved 0.495 ± 0.170, not reliably higher than S1 + S2 alone. VRI correlates strongly with S1 (r = − 0.64) and moderately with several other landmarks (|r| 0.18–0.60); the implications for interpreting feature importance are discussed in Section 4.2.

Table 3
Ablation analysis: cross-validated R^2 by feature subset (same RF configuration as the main model, Section 2.5).

Feature subset	Cross-validated R^2 (mean ± SD)
All 8 features	0.609 ± 0.140
No VRI (7 features)	0.566 ± 0.156
S1 + S2 only	0.520 ± 0.166
VRI only	0.332 ± 0.253
S1 + S2 + VRI	0.495 ± 0.170

3.6 Missingness-Bias Check

Participants with complete body-composition measurements (n = 88) did not differ materially in sex distribution from those without (n = 139) (chi-square test, p = 0.54), but differed sharply in age: 31.4 ± 11.9 years (mean ± SD) in the complete-data subgroup versus 45.7 ± 19.4 years in the incomplete-data

subgroup (Welch's t-test, $p = 7.6 \times 10^{-11}$) (Table 4). Consequently, any analysis of this cohort restricted to participants with complete body-composition data — a common, almost reflexive way of handling missing covariates — would describe a systematically younger population than the full 227-participant cohort used for the age-estimation analyses above, without that restriction being visible anywhere in the model's reported performance or feature-importance values.

Table 4
Complete versus incomplete body-composition subgroups.

Mean age, years (mean ± SD)	**Complete body-composition data (n = 88)**	**Incomplete body-composition data (n = 139)**	**p-value**
	31.4 ± 11.9	**45.7 ± 19.4**	**7.6×10^{-11}**
Sex distribution	Not materially different (χ^2 test)		0.54

3.7 Biomarker Interpretation of RI, VRI, and VEI

Consistent with Section 2.4, the three composite indices are interpreted here in qualified, descriptive terms rather than as direct physiological measurements. VRI is a wave-reflection-related index — higher values track a larger secondary systolic component relative to the primary peak — not a direct measurement of cerebral wave reflection, which would need a validating measurement not available here. RI is a resistance-related index from the systolic/end-diastolic ratio, and VEI a waveform-derived elasticity-related index from the incisura/diastolic ratio, both without an independent reference to confirm the interpretation.

Within this same cohort, sex and age jointly modulate these indices: females showed higher VRI and lower VEI than males (both $p < 0.001$, two-sided t-test), and an age-by-sex interaction for VRI was detected ($p = 0.002$, linear regression). These two tests are reported directly as part of the present analysis, computed on the n = 227 dataset described in Section 2.2. A fuller multivariable determinant analysis of these indices — incorporating body-composition and blood-pressure covariates jointly rather than as separate significance tests — is beyond the scope of this paper and is left to dedicated future work.

4. Discussion

4.1 Principal Findings

This case study applies the proposed framework to carotid Doppler waveforms from 227 adults, stress-testing the explainable-analysis stage rather than accepting one importance ranking as final. A Random Forest model using eight waveform features produced a Doppler-based chronological-age estimate with MAE 8.27 years, and VRI consistently ranked top by correlation and both importance methods, yet its apparent dominance (≈ 50% of Gini importance, 61–79% of permutation importance) shrank to just a

0.043 R^2 loss (~ 7% relative) when removed and the model retrained — the central methodological result of this study. The underlying age-related slowing and waveform change in the common carotid artery is itself consistent with this group's earlier, smaller-sample observations of age- and exercise-training-related differences in common carotid blood velocities [27].

Against other waveform-based age-estimation approaches, this 8.27-year MAE is higher (worse) than two published figures. A Random Forest model using PPG waveform-morphology features in 173 healthy individuals aged 3–61 years achieved MAE = 6.97 years [28]. PWV/applanation-tonometry arterial-age models combining PWV, augmentation index, and carotid intima-media thickness reported MAE = 5.87–6.91 years (adjusted R^2 = 0.55–0.69) in a comparably sized cohort [29]. The higher error here may reflect the narrower feature set (eight Doppler landmarks only, versus PPG morphology or multimodal PWV-plus-structural measurements), the wider age range sampled (20–89 years), or a real difference in age-related information content; this analysis cannot distinguish between these explanations, and a head-to-head comparison on one cohort would be needed to.

The direction of the RI association also merits comment: RI fell with age in this cohort ($r = -0.267$, Section 3.2) — opposite the rise typically reported for resistance/pulsatility indices in more distal or cerebral vessels. This may reflect relative preservation of end-diastolic flow in this proximal, elastic conduit artery rather than a true rise in resistance; the interpretation is provisional and not confirmable from this cross-sectional dataset.

Prediction error was also not uniform across age (Section 3.3), and the pattern is worth discussing on its own terms rather than as a brief limitation. The model is least reliable at exactly the ages where arterial aging is of greatest clinical interest: 92 of 227 participants (41%) were in their twenties, the cohort's densest age bin, where MAE was 6.1 years, versus only 16 participants aged ≥ 70, the sparsest bins, where MAE rose to 15.3–16.0 years. The overall MAE and R^2 reported above are accordingly dominated by the easier, more populous younger decades, and performance in the clinically more relevant 70s-80s range is both underpowered and likely overstated relative to what a balanced-age sample would show. The non-monotonic wobble among the thirties-through-sixties decades (9.3–10.5 years, n = 20–42 each) most plausibly reflects sampling noise from these modest subgroup sizes rather than a real trend break; the broader pattern — lowest error where training data are densest, higher error as the cohort thins — matches the regression-to-the-mean bias well documented in biological-age prediction models generally [30], not a problem specific to this dataset.

This case study is part of a broader program of explainable-analysis work on this cohort's carotid Doppler waveforms; related companion conference papers — covering interpretable classification, explainable ML classification, and continuous age-regression on a related but non-identical subset of this cohort — are reported separately [31–33] and are not replications of this paper's framework or ablation stress-test.

4.2 What the Explainable-Analysis Stage Does and Does Not Show

Explainability is increasingly treated as a requirement for responsible AI use in healthcare [10, 19–21, 34], since predictive accuracy alone has limited clinical adoption when a model's output is opaque. This study's central finding is methodological: a single global importance metric, taken alone, would have supported the overstated claim that VRI is close to indispensable, whereas ablation shows the model loses comparatively little without it, because the information VRI carries is largely recoverable from S1 (r = − 0.64) and, to a lesser extent, several other landmarks (|r| 0.18–0.60**)** — matching documented bias in Random Forest importance toward correlated predictors [22], exactly why the framework requires multiple, independent explainability checks (Section 2.1) rather than one ranking. A counterfactual-explanation approach [35] would give a still more actionable account and is a natural next step.

4.3 Portability and Generalization

Portability here is not a future aspiration but an already-demonstrated property of the measurement platform: the carotid Doppler system used to collect this cohort's data (Section 2.3) is itself a portable device developed by the author and validated against commercially available diagnostic ultrasonography in an earlier report [18]. What remains untested is comparability with other wearable-biomarker modalities: this analysis has not yet been benchmarked against PPG-derived biomarkers specifically, and flexible/wireless ultrasound patches [11–12, 36] — with their own accuracy caveats [37] — represent a further step beyond the validated device used here. The framework (Section 2.1) is not specific to Doppler and could, in principle, apply to ECG- or PPG-derived biomarkers [4–5, 38–39], but this cross-modality comparison is untested, and the multi-method stress test (Sections 3.4–3.5) would need repeating for any new signal, since correlated-feature masking of importance is a general property of tree-based methods [22], not specific to Doppler indices.

4.4 Limitations

(1) Construct validity: the RF output is a Doppler-based chronological-age estimate, not a validated vascular-age biomarker, since no PWV, intima-media thickness, or outcome data were available (criteria a-e, Section 3.3). (2) Explainability depth: permutation importance and ablation strengthen the explainable-analysis stage beyond impurity-based importance, but instance-level and counterfactual explanations (Section 4.2) were not computed. (3) Cohort characterization: blood pressure, BMI, and body-composition data covered only subsets of the cohort (Table 1); Section 3.6 shows this incompleteness is age-related, not random, and smoking/medication status were not recorded, so some age association may reflect confounding. (4) Repeat-visit measurement reproducibility was not assessed. (5) This is one cross-sectional, single-site healthy-volunteer sample; generalizability to other modalities or to vascular disease is untested. (6) Sex was recorded but deliberately excluded from the RF feature set to keep the model waveform-only; a sensitivity check (Section 2.5) showed adding sex modestly improves performance, so the model likely does not capture all age-related information in this dataset. Limitations (1) and (3) especially would need new data or an independent reference measurement to resolve, and are left open rather than overstated.

4.5 Future Directions

Future work could pursue multimodal extension, wearable ultrasound platforms as they mature [11–12], counterfactual explainability [35], and — only once an independent vascular reference measurement is incorporated — a genuine test of whether the estimate behaves as a vascular-age biomarker. Clinical or commercial deployment claims should wait on that validation.

5. Conclusion

This paper described a small conceptual framework linking signal processing, explainable machine learning, and biomarker-level interpretation, illustrated and stress-tested with a secondary analysis of carotid Doppler waveforms from 227 healthy adults. VRI dominated by correlation and by two importance metrics, and a Random Forest model using eight waveform features produced a Doppler-based chronological-age estimate with MAE 8.27 years. Ablation, however, showed VRI's unique contribution beyond the landmarks it is calculated from is modest, since importance scores for correlated features can substantially overstate how indispensable any one of them is.

Declarations

Conflicts of interest:

The author declares no conflicts of interest.

Declarations

Ethical Considerations

Original data collection was approved by the Ethics Committee of Tokushima University Hospital, Tokushima, Japan, as previously reported in Azhim et al. [18], and conducted in accordance with the Declaration of Helsinki; written informed consent was obtained from all participants prior to data collection.

Consent for Publication:

Not applicable; no individually identifiable patient information is presented.

Conflicts of Interest:

The author declares no conflicts of interest.

Funding

This work was supported by an internal grant from the National Institute of Technology, Kitakyushu College.

Use of Artificial Intelligence

During preparation of this manuscript, the author used an AI-based writing and data-analysis assistant (Claude, Anthropic; Claude Sonnet model family, 2026) which helped identify relevant literature, draft statistical-analysis scripts (Python, scikit-learn/SciPy/statsmodels), build the figures, and polish the language, all under the author's direction and with iterative review. The author reviewed and verified all content, analyses, and citations, and takes full responsibility for the final manuscript.

Funding:

This work was supported by an internal grant from the National Institute of Technology, Kitakyushu College.

Authors' Contribution:

A. Azhim (A.A.): conceptualization, methodology, software, formal analysis, writing – original draft, writing – review and editing.

Acknowledgments:

The authors would like to acknowledge the support provided by National Institute of Technology, Kitakyushu College for facilitating this research work. The authors also would like to express their appreciation to Tokushima University and Tokushima University Hospital for their support in the original Doppler hemodynamic data collection and physiological investigations.

Data Availability:

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

1. Pantelopoulos A, Bourbakis NG (2010) A survey on wearable sensor-based systems for health monitoring and prognosis. IEEE Trans Syst Man Cybern C 40(1):1–12
2. Majumder S, Mondal T, Deen MJ (2017) Wearable sensors for remote health monitoring. Sensors 17(1):130
3. Piwek L, Ellis DA, Andrews S, Joinson A (2016) The rise of consumer health wearables: promises and barriers. PLoS Med 13(2):e1001953

4. Clifford GD, Azuaje F, McSharry P (2006) Advanced Methods and Tools for ECG Data Analysis. Artech House, Boston
5. Allen J (2007) Photoplethysmography and its application in clinical physiological measurement. Physiol Meas 28(3):R1–39
6. Evans DH, McDicken WN (2000) Doppler Ultrasound: Physics, Instrumentation and Signal Processing, 2nd edn. Wiley, Chichester
7. Topol EJ (2019) High-performance medicine: the convergence of human and artificial intelligence. Nat Med 25(1):44–56
8. Esteva A, Robicquet A, Ramsundar B, Kuleshov V, DePristo M, Chou K et al (2019) A guide to deep learning in healthcare. Nat Med 25(1):24–29
9. Shneiderman B (2020) Human-centered artificial intelligence: reliable, safe and trustworthy. Int J Hum Comput Interact 36(6):495–504
10. Reddy S, Allan S, Coghlan S, Cooper P (2020) A governance model for the application of AI in health care. J Am Med Inf Assoc 27(3):491–497
11. Amado-Rey AB, Goncalves Seabra AC, Stieglitz T (2025) Towards ultrasound wearable technology for cardiovascular monitoring: from device development to clinical validation. IEEE Rev Biomed Eng 18:93–112
12. Hofer LM, Kenny JS, Munding CE, Kerrebijn I, Atwi S, Younan A et al (2025) Defining the physiological bounds of left ventricular ejection time with a wireless, wearable ultrasound: an analysis of over 137,000 cardiac cycles. Digit Health. ;11
13. Nichols WW, O'Rourke MF, Vlachopoulos C (2011) McDonald's Blood Flow in Arteries: Theoretical, Experimental and Clinical Principles, 6th edn. Hodder Arnold, London
14. O'Rourke MF, Hashimoto J (2007) Mechanical factors in arterial aging: a clinical perspective. J Am Coll Cardiol 50(1):1–13
15. Laurent S, Cockcroft J, Van Bortel L, Boutouyrie P, Giannattasio C, Hayoz D et al (2006) Expert consensus document on arterial stiffness: methodological issues and clinical applications. Eur Heart J 27(21):2588–2605
16. Mitchell GF, Hwang SJ, Vasan RS, Larson MG, Pencina MJ, Hamburg NM et al (2010) Arterial stiffness and cardiovascular events: the Framingham Heart Study. Circulation 121(4):505–511
17. Vlachopoulos C, Aznaouridis K, Stefanadis C (2010) Prediction of cardiovascular events and all-cause mortality with arterial stiffness: a systematic review and meta-analysis. J Am Coll Cardiol 55(13):1318–1327
18. Azhim A, Ueno A, Tanaka M, Akutagawa M, Kinouchi Y (2011) Evaluation of blood flow velocity envelope in common carotid artery for reference data. Biomed Signal Process Control 6(2):209–215
19. Holzinger A, Langs G, Denk H, Zatloukal K, Muller H (2019) Causability and explainability of artificial intelligence in medicine. WIREs Data Min Knowl Discov 9(4):e1312

20. Tjoa E, Guan C (2021) A survey on explainable artificial intelligence (XAI): toward medical XAI. IEEE Trans Neural Netw Learn Syst 32(11):4793–4813
21. Arrieta AB, Diaz-Rodriguez N, Del Ser J, Bennetot A, Tabik S, Barbado A et al (2020) Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inf Fusion 58:82–115
22. Strobl C, Boulesteix AL, Kneib T, Augustin T, Zeileis A (2008) Conditional variable importance for random forests. BMC Bioinformatics 9:307
23. von Elm E, Altman DG, Egger M, Pocock SJ, Gotzsche PC, Vandenbroucke JP et al (2007) The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 370(9596):1453–1457
24. Azhim A, Yamaguchi H, Hirao Y, Kinouchi Y, Yoshizaki K, Ito S et al (2005) Monitoring carotid blood flow and ECG for cardiovascular disease in elder subjects. In: Proceedings of the 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. ;5:5495-8
25. Azhim A, Katai M, Akutagawa M, Hirao Y, Yoshizaki K, Obara S et al (2008) Measurement of blood flow velocity waveforms in the carotid, brachial and femoral arteries during head-up tilt. J Biomed Pharm Eng 2(1):1–6
26. Pourcelot L (1974) Applications cliniques de l'examen Doppler transcutane. In: Peronneau P (ed) Velocimetrie Ultrasonore Doppler. INSERM, Paris, pp 213–240
27. Azhim A, Akioka K, Akutagawa M, Hirao Y, Yoshizaki K, Obara S et al (2007) Effects of aging and exercise training on the common carotid blood velocities in healthy men. In: Proceedings of the 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. :989 – 93
28. Saran Khalid M, Shahid Quraishi I, Wasim Nawaz M, Sajjad H, Yaseen H, Mehmood A et al (2025) A low-cost PPG sensor-based empirical study on healthy aging based on changes in PPG morphology. Physiol Meas. ;46(2)
29. Fedintsev A, Kashtanova D, Tkacheva O, Strazhesko I, Kudryavtseva A, Baranova A et al (2017) Markers of arterial health could serve as accurate non-invasive predictors of human biological and chronological age. Aging 9(4):1280–1292
30. de Lange AM, Cole JH (2020) Commentary: Correction procedures in brain-age prediction. Neuroimage Clin 26:102229
31. Azhim A, Matsukubo J (2026) Interpretable AI framework for vascular health assessment using physiological waveform biomarkers. In: Proceedings of the International Conference on Artificial Intelligence for Everyone (IICAIET). In press
32. Azhim A, Matsukubo J (2026) Explainable machine learning classification of vascular aging using carotid velocity waveform features. In: Proceedings of the International Conference on Biomedical Imaging and Processing (ICBIP). In press
33. Azhim A (2026) AI-based vascular age regression using carotid Doppler hemodynamic features. In: Proceedings of the Biomedical Engineering International Conference (BMEiCON). In press

34. Markus AF, Kors JA, Rijnbeek PR (2021) The role of explainability in creating trustworthy artificial intelligence for health care: a comprehensive survey of the terminology, design choices, and evaluation strategies. J Biomed Inf 113:103655
35. Wachter S, Mittelstadt B, Russell C (2018) Counterfactual explanations without opening the black box: automated decisions and the GDPR. Harv J Law Technol 31(2):841–887
36. Kim J, Campbell AS, de Avila BEF, Wang J (2019) Wearable biosensors for healthcare monitoring. Nat Biotechnol 37(4):389–406
37. Bent B, Goldstein BA, Kibbe WA, Dunn JP (2020) Investigating sources of inaccuracy in wearable optical heart rate sensors. NPJ Digit Med 3:18
38. Seals DR, Justice JN, LaRocca TJ (2016) Physiological geroscience: targeting function to increase healthspan and achieve optimal longevity. J Physiol 594(8):2001–2024
39. North BJ, Sinclair DA (2012) The intersection between aging and cardiovascular disease. Circ Res 110(8):1097–1108

Figures

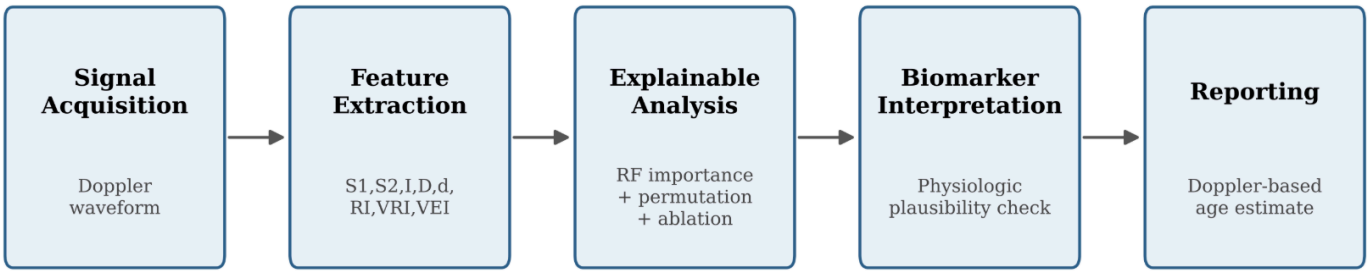


Figure 1

The Explainable Biomarker-Discovery Framework. A physiological signal is progressively transformed into an interpretable, feature-traceable estimate through feature extraction, multi-method explainable analysis, and biomarker-level interpretation.

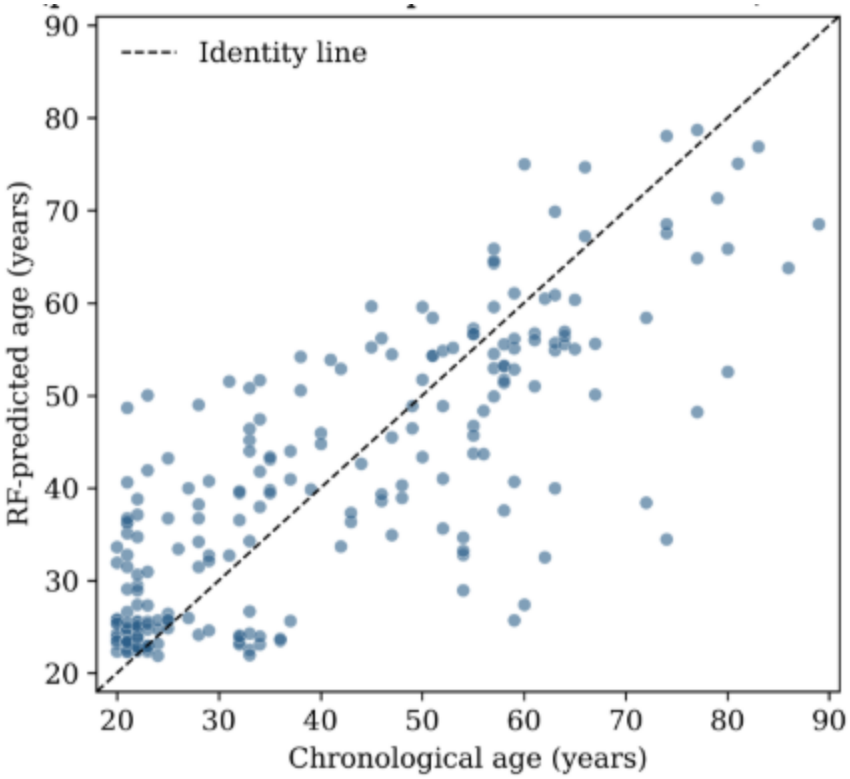


Figure 2

Predicted versus chronological age (n = 227). Each point is a participant's predicted age, averaged across the repeats of the repeated 10-fold cross-validation in which that participant was held out. The dashed line is the identity line.

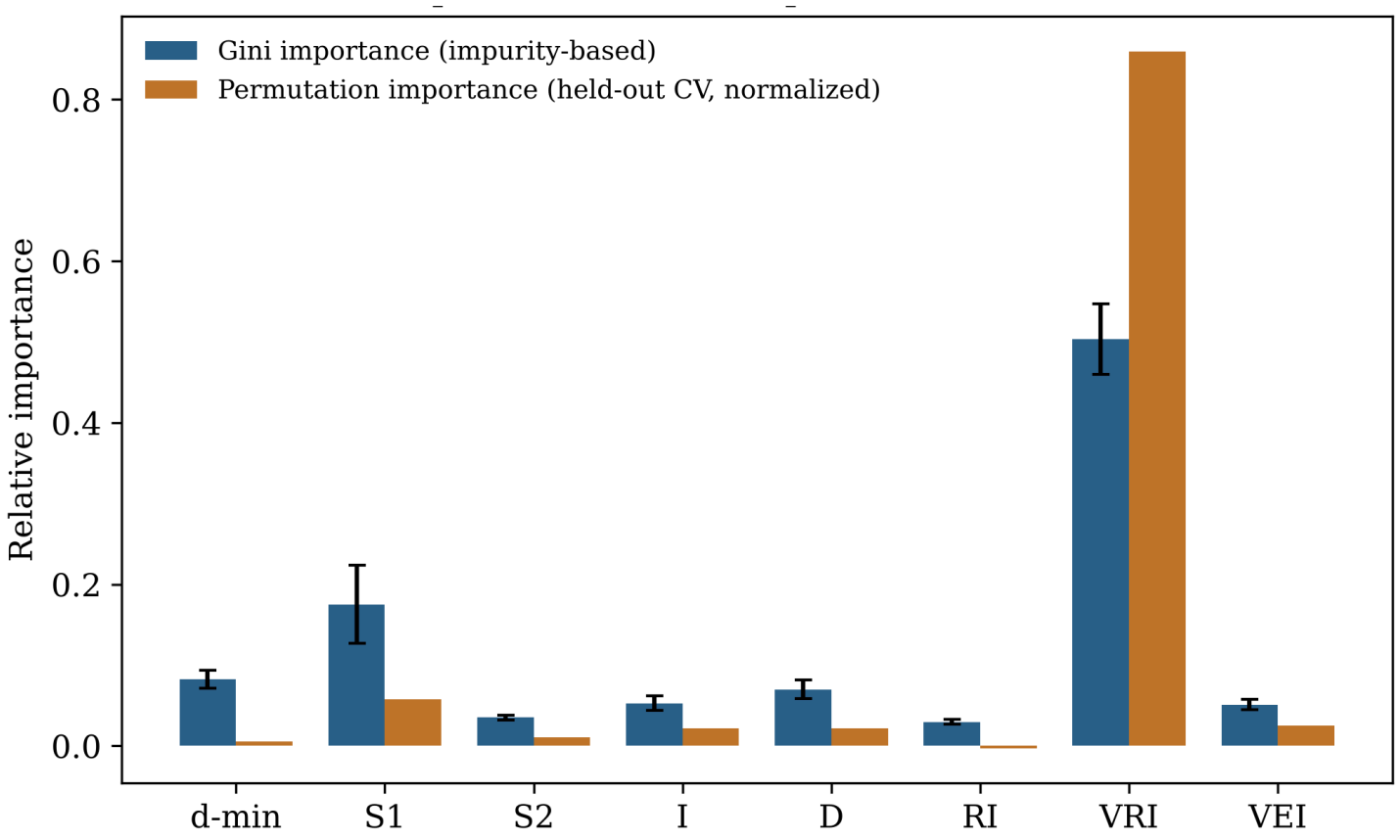


Figure 3

Feature importance by Gini (impurity-based) and permutation (held-out cross-validation, normalized to sum to 1 for visual comparability) methods. Error bars show SD across cross-validation runs.

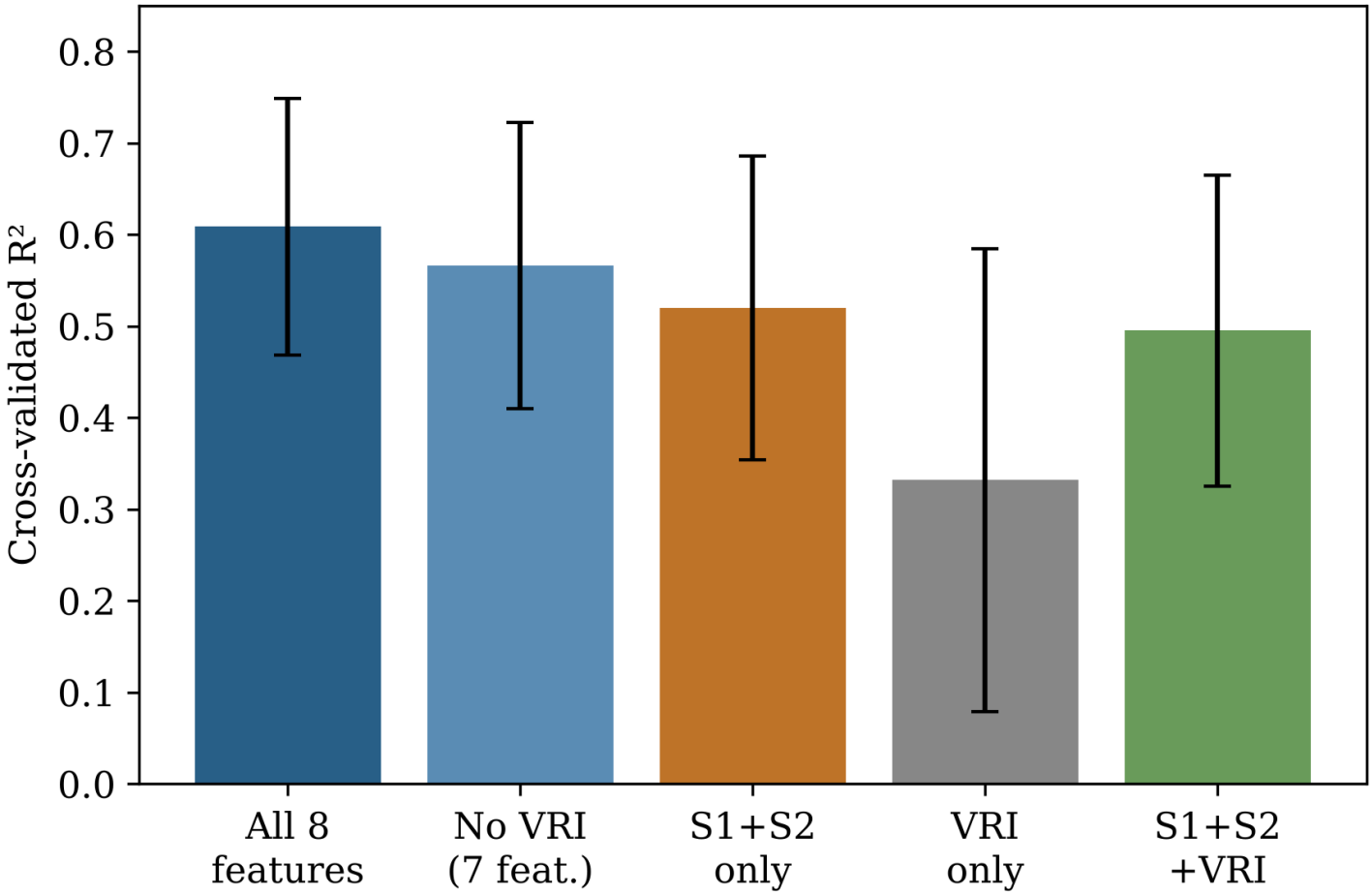


Figure 4

Cross-validated R² for five feature subsets (500-tree RF, repeated 10-fold cross-validation, 10 repeats — identical configuration to the main model). Error bars show SD across cross-validation runs.