

An Explainable Machine Learning Framework for Identifying Left Bundle Branch Block Mechanisms via Vectorcardiographic Features

Beatriz del Cisne Macas Ordóñez, Javier Garrigós, José Javier Martínez, José Manu

Abstract

Left Bundle Branch Block (LBBB) diagnosis is crucial for patient stratification and the selection of individuals who are likely to respond to Cardiac Resynchronization Therapy (CRT). The pathophysiological distinction between LBBB and strict LBBB (sLBBB) is investigated in this research with a view to optimizing diagnostic criteria and therapy. Data were retrieved from the E-OTH-12-0602-024 database of the Telemetric and Holter ECG Warehouse (THEW) and the Large-Scale 12-Lead Electrocardiogram Dataset. The ECG signals were transformed into the vectorcardiographic (VCG) domain, where QRS loops were divided into two halves at the time of the velocity peak computed over the discrete derivatives of the x, y, and z leads. From each half, QRS angles and maximal norms were extracted in all VCG planes. Furthermore, features representing relations between VCG peak velocity and VCG fiducial points were extracted. These were used to train a series of machine learning models for classification into Healthy, LBBB, and sLBBB categories. The analysis identified four most significant features that best discriminated between the classes: (1) peak velocity time relative to QRS onset and offset, (2) maximum amplitude of the initial portion of the QRS loop in the frontal plane, and (3) QRS angle in the horizontal plane. The reduced set of pathophysiological features preserved essential differences in conduction dynamics and electrical disturbances among the three groups. In particular, the time from velocity peak to QRS offset was the most discriminative feature, with progressive prolongation from Healthy to LBBB to sLBBB classes. This reduced 4-feature set achieved an accuracy of 0.85 and an F1-score of 0.83, which was on par with 15-feature-based models. Finally, the integration of explainable artificial intelligence (xAI) into these simplified models enabled the derivation of transparent diagnostic rules for LBBB, improving clinical interpretability and supporting more reliable diagnostic decisions.

1 1. Introduction

The definition of strict left bundle branch block (LBBB) was first introduced by Strauss in 2011 based on the analysis of the ECG phenotypes of patients who significantly benefited from cardiac resynchronization therapy (CRT) in the MADIT-CRT trial (1). This subgroup exhibited specific characteristics including a QRS duration ≥ 140 ms in men or ≥ 130 ms in women, QS- or rS-configurations of the QRS complex in leads V1 and V2, and mid-QRS notching or slurring in ≥ 2 of leads V1, V2, V5, V6, I, and aVL(2). Since then, the diagnosis of strict LBBB has emerged as a potent predictor of CRT efficacy.

The objective of this study is to investigate the pathophysiological aspects associated with left bundle branch block (LBBB) and strict left bundle branch block (sLBBB), aiming to refine treatment strategies and improve clinical outcomes in patients affected by these conditions.

Incorporating findings related to maximum conduction velocity might provide important insights. Understanding whether LBBB shows delayed peak conduction velocity compared to sLBBB could imply a more proximal occurrence of these blocks, with profound clinical implications. Confirmation of such distinctions would not only deepen our understanding of the underlying mechanisms but also their associations with underlying heart diseases.

While refining treatment strategies may be a longer-term outcome, the immediate value lies in advancing our physiological understanding of LBBB and sLBBB. This could pave the way for more targeted therapies, thereby enhancing clinical practice and patient outcomes.

1.1 Related Work and Contribution

The differentiation of Left Bundle Branch Block (LBBB) from its strict subtype (sLBBB) is a critical challenge in cardiology, as it directly impacts the selection of patients for Cardiac Resynchronization Therapy (CRT) (1). While strict electrocardiographic criteria were introduced to identify patients more likely to respond to CRT, their application can be subjective and may not fully capture the underlying electrophysiological dynamics (3). Moreover, a more accurate LBBB diagnosis might help identify candidates for other cardiac stimulation schemes, such as His bundle pacing or other biventricular configurations (4; 5).

To address these challenges, many studies have turned to machine learning (ML) and deep learning (DL) for the automated analysis of ECG signals (6; 7; 8; 9; 10). The field has seen significant advances in computational methods for cardiac monitoring, ranging from mobile cardiac monitoring systems (8) to sophisticated prediction models for sudden cardiac death (9; 10). Initial efforts in automated sLBBB detection, such as those presented in the International Society for Computerized Electrocardiology (ISCE) initiative, reported accuracies around 82% (2), demonstrating the potential of computational approaches.

However, a significant limitation of many advanced models is their "black box" nature. This lack of transparency makes it difficult for clinicians to understand the reasoning behind a diagnosis, which hinders clinical trust and adoption (11; 12). In response, the field of Explainable Artificial Intelligence (xAI) has become crucial across various medical domains, including EEG analysis for impulsivity classification (13), cognitive impairment assessment through clock drawing tests (14), neonatal seizure detection (15), and Alzheimer's disease evaluation (16).

Recent works have begun to incorporate explainability into LBBB classification. For instance, Macas et al. (2024) developed a system using a reduced set of bio-inspired features derived from the vectorcardiogram (VCG), such as QRS-T angles and areas. They employed XAI techniques like SHAP to provide feature-level explanations, successfully identifying the most influential parameters for a ternary classification task (*Healthy*, *LBBB*, *sLBBB*) and achieving an accuracy of up to 82.63%. While providing valuable insights into feature importance, this approach's explanations are based on feature attribution scores, which can still be complex to interpret directly in a clinical workflow (17).

Concurrently, novel signal representation methods and advanced learning algorithms have been explored. Graph theory has emerged as a powerful paradigm to model the complex inter-lead relationships within a 12-lead ECG. In this line, Macas Ordóñez et al. (2025) proposed and compared two distinct graph-based methodologies. The first approach, using Graph Signal Processing (GSP) and a Support Vector Machine, achieved superior diagnostic accuracy (mean balanced accuracy of 0.8317) by leveraging features from the graph spectral domain, although these features can be difficult to interpret clinically. The second approach converted connectivity matrices into images for a Convolutional Neural Network (CNN), incorporating xAI via Grad-CAM to visualize the inter-lead interactions influencing the model's decision. While this second method offered enhanced visual interpretability, its accuracy was lower, and like other xAI methods, it did not generate simple, direct clinical rules (18).

The field has also seen advances in neural dynamic classification algorithms (19),

finite element machines for fast learning (20), dynamic ensemble learning (21), and self-supervised learning approaches for electrophysiological data (22). However, while these previous works have made significant strides, a gap remains in developing a method that is both highly accurate and provides explanations in the form of simple, verifiable rules specifically for LBBB differentiation.

Our current study addresses this gap by introducing a framework that leverages a novel set of VCG features based on conduction velocity dynamics, specifically the timing of the peak velocity ($t_{v_{max}}$). Our main contribution is the subsequent application of advanced xAI techniques, such as the Anchors method, to distill the model’s complex decision-making process into a concise set of high-precision, human-readable rules. This approach moves beyond feature importance scores or complex spectral analysis to provide clinicians with transparent, physiologically grounded, and directly applicable criteria for distinguishing between *Healthy*, *LBBB*, and *sLBBB* patients, bridging the gap between predictive accuracy and true clinical interpretability.

Relative to our 2023 ICAE study (Macas et al., 2023), the present work introduces: (i) a velocity-based VCG loop segmentation (d1/d2) that yields physiologically grounded features; (ii) a 4-feature model that attains comparable performance to the 15-feature set while improving interpretability; (iii) consolidated, high-precision Anchor rules that provide global rule sets from local explanations; and (iv) by-class xAI analyses and spectral clustering that align with Strauss-defined phenotypes. Together these advances reduce model complexity, improve transparency, and sharpen physiologic insight.

2. Materials and Methods

The overall workflow of the methodology, from data preprocessing to model interpretation, is summarized in Figure 1. Two databases were utilized in this study. The first database, E-OTH-12-0602-024 (23), was obtained from the Telemetric and Holter ECG Warehouse (THEW) as part of the initiative of the International Society for Computerized Electrocardiology (ISCE) in 2018 (2). This database comprises 602 ECG records from the MADIT-CRT clinical trial, conducted at the University of Rochester (Rochester, NY). It includes 331 cases of strict left bundle branch block (sLBBB) and 192 cases of incomplete LBBB and 79 cases of other cardiopathies (*noLBBB*), each recording with a duration of 10 seconds, sampled at a frequency of 1 kHz with an amplitude resolution of 3.75 μ V. The *noLBBB* class was ruled out from analysis because it contained extremely heterogeneous pathologies. Also,

a preliminary analysis of the dataset revealed the presence of 41 duplicate cases in this dataset, 36 of which corresponded to the *LBBB* and *sLBBB* classes used in this study. We identified the 36 exact duplicate pairs and retained only one instance per pair, resulting in 184 *LBBB* and 302 *sLBBB* records. Due to the anonymization requirements of the database, it was not possible to determine if any records belonged to the same patient at different stages of the pathology (e.g., progressing from *LBBB* to *sLBBB*).

The second database, a Large Scale 12-lead Electrocardiogram Database for Arrhythmia Study from PhysioNet (24), includes ECGs from 45,152 patients in resting conditions, encompassing 64 different types of arrhythmias as well as healthy ECG records. The data were collected from Shaoxing People’s Hospital and Ningbo First Hospital. The ECGs are sampled at 500 Hz and have a duration of 10 seconds. From this database, 299 randomly selected ECG records of healthy subjects were extracted.

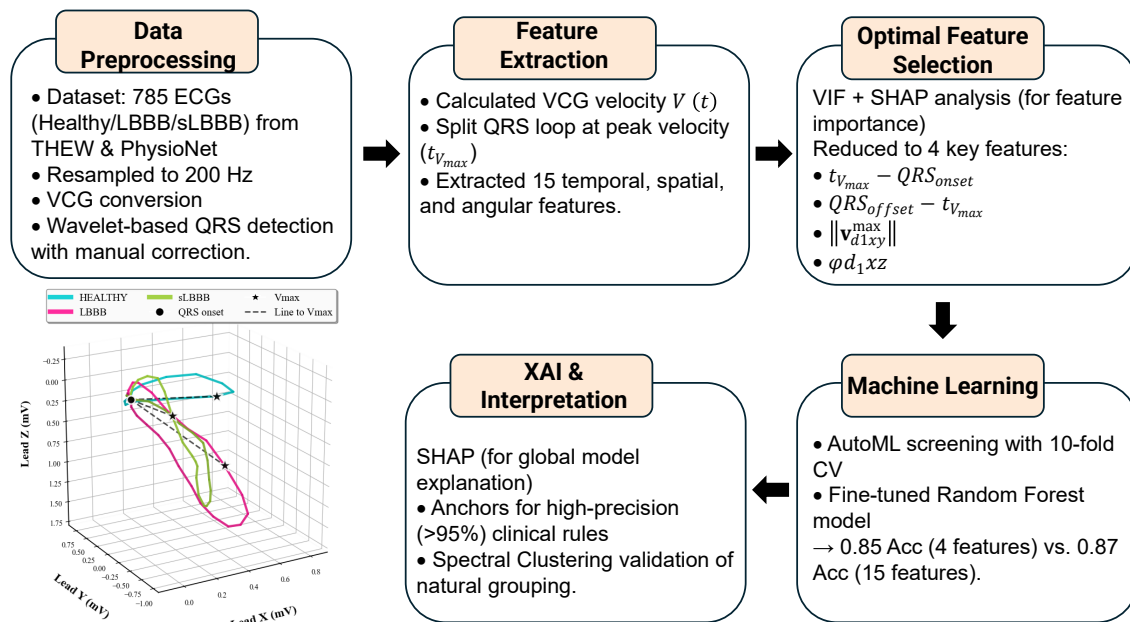


Figure 1: Methodology workflow: preprocessing, feature extraction/selection, ML, and XAI.

2.1 ECG Preprocessing

To ensure consistency in the analysis, all signals were resampled to 200 Hz. The ECG data were transformed into the vectorcardiographic (VCG) space using the inverse Dower matrix (25) and the derived VCG signals were delineated using a wavelet-based algorithm implemented with the WT-delineator library in Python (26). This algorithm identified the onset and offset of the QRS complex, enabling the construction of QRS loops. For ECGs with abnormal morphologies, manual corrections were frequently necessary to ensure accurate delineation. For each patient, a unique mean loop was obtained by averaging loops across all beats. From these mean loops, a set of 15 features were derived, as explained in the following sections.

2.1.1 Cardiac Velocity and loop segmentation

Cardiac velocity $V(t)$ was computed as the norm of the discrete derivatives of the VCG leads $x(t)$, $y(t)$, and $z(t)$. Afterwards, the time to peak ($\arg \max_t[V(t)]$) was estimated. This time reflects the instant when the instantaneous velocity reaches its maximum amplitude and was used to partition the VCG loops into two halves, as explained below.

Once the time of maximum velocity $t_{V_{\max}}$ during the QRS complex was found, segmentation occurred as follows. The first segment, d_1 , spanned from the onset of the QRS complex to the time of maximum velocity $t_{V_{\max}}$. This period is characterized by rapid depolarization and high velocity, indicating early ventricular activation. The second section, d_2 , extended from the time of maximum velocity $t_{V_{\max}}$ to the offset of the QRS complex. This interval represents the slower repolarization processes. The separation of these segments is detailed in Figure 3a. Here, the segmentation for lead z is depicted according to the three fiducial points, t_{QRS_onset} , t_{QRS_offset} and $t_{V_{\max}}$. Consistently, the segmentation on the loops is shown in Figure 3b, for the xz plane in the case of a *Healthy* subject, while the *LBBB* and *sLBBB* counterparts are shown in Figure 3c and Figure 3d respectively. The features *peak velocity intervals*, *maximum norms* and *QRS angles* were then computed for each half of the VCG waveforms and loops, totaling 15 features.

2.1.2 Peak velocity intervals

Based on the time for peak velocity $t_{V_{\max}}$ obtained, three temporal parameters were constructed. The first temporal parameter was $t_{V_{\max}}$ itself, while the remaining two $t_{V_{\max}-QRS_onset}$ and $t_{QRS_offset-V_{\max}}$ are references of the time for peak velocity with

respect to the fiducial points t_{QRS_onset} and t_{QRS_offset} respectively.

$$t_{V_{max}-QRS_{onset}} = t_{V_{max}} - t_{QRS_onset} \quad (1)$$

$$QRS_{offset} - t_{V_{max}} = t_{QRS_offset} - t_{V_{max}} \quad (2)$$

Figure 2 depicts the intervals constructed from the VCG fiducial points (QRS_{onset} and QRS_{offset}) and the time of peak velocity $T_{V_{max}}$. In this particular example, lead z (green) is shown simultaneously to cardiac velocity (blue). Notice the stretching of the $QRS_{offset}-T_{V_{max}}$ or the shortening of the $T_{V_{max}}-QRS_{onset}$ intervals with conduction impairment.

2.1.3 Maximum Norms

For each time t_j within the intervals d_1 and d_2 , the magnitude of the VCG vector in the xy plane, $\|\mathbf{v}_{xy}(t_j)\|$, was calculated as:

$$\|\mathbf{v}_{xy}(t_j)\| = \sqrt{x(t_j)^2 + y(t_j)^2} \quad (3)$$

The calculated norms were divided into two groups corresponding to the d_1 and d_2 intervals and the respective maximum norm for each interval was obtained as follows:

$$\begin{aligned} \|\mathbf{v}_{d1xy}^{max}\| &= \max_{t_j} \|\mathbf{v}_{xy}(t_j)\| \quad \text{for } t_j \in [t_{QRS_onset}, t_{V_{max}}] \\ \|\mathbf{v}_{d2xy}^{max}\| &= \max_{t_j} \|\mathbf{v}_{xy}(t_j)\| \quad \text{for } t_j \in [t_{V_{max}}, t_{QRS_offset}] \end{aligned} \quad (4)$$

These maximum norms represent the dominant vectors in d_1 and d_2 . Similar computations were applied to the xz and yz planes.

2.1.4 QRS Angles

The angle φ_{d_1} between the dominant vector \mathbf{v}_{d1xy}^{max} , extending from the origin of the VCG loop to the point of maximum norm $\|\mathbf{v}_{d1xy}^{max}\|$, and the horizontal axis, was determined using the dot product:

$$\varphi_{d1xy} = \left(\frac{180}{\pi} \right) \arccos \left(\frac{\langle \mathbf{v}_{d1xy}^{max}, \mathbf{i} \rangle}{\|\mathbf{v}_{d1xy}^{max}\| \cdot \|\mathbf{i}\|} \right)$$

where \mathbf{i} represents the unit vector along the horizontal axis x . The angle φ_{d_2} was calculated similarly for the vector \mathbf{v}_{d_2} and planes xz and yz .

Figure 3 shows a representative example of loop segmentation based on peak velocity for the three classes. Notice how the d_2 segment (blue) grows with pathology, from *Healthy* to *sLBBB*. Also, note the characteristic rotation of the loop in the xz plane for the *sLBBB* class.

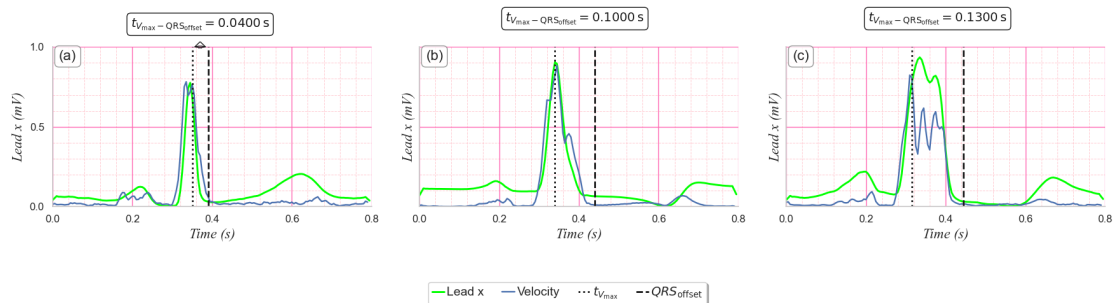


Figure 2: Comparison of VCG lead z morphology (top traces) and electrical conduction velocity constructed over the norm of the x, y, z discrete derivatives (bottom traces) for representative patients: a) *Healthy*, b) *LBBB*, and c) *sLBBB*. Panel a) illustrates the segmentation. The time from peak velocity to QRS offset ($t_{QRS_offset} - t_{V_{max}}$) visibly increases from *Healthy* to *sLBBB*.

2.2 Data conditioning for Machine Learning models

We initiated the process with a tabular dataset comprising 785 rows and 16 columns. Each of the 785 rows represented a unique patient, characterized by 15 features and a label column indicating the class, which was assigned one of three values: 0, 1, and 2 representing the categories *Healthy*, *LBBB*, and *sLBBB*, respectively. The distribution of subjects was balanced by class, with 299 *Healthy*, 184 *LBBB*, and 302 *sLBBB*. Therefore, no synthetic resampling or class-weighting technique was applied in order to preserve the natural composition of the clinically validated dataset.

Six out of fifteen features accounted for maximum norms at d_1 and d_2 in planes xy , xz and yz . Another six features consisted of QRS angles at d_1 and d_2 in every plane, and finally, the last three features were time-related ones, consisting of the time for peak velocity $t_{V_{max}}$, the interval from the beginning of the QRS to $t_{V_{max}}$: $t_{V_{max}-QRS_onset}$ and the interval from the time for $t_{V_{max}}$ to QRS offset: $QRS_offset - t_{V_{max}}$.

Following the feature engineering stage, we proceeded to a multicollinearity detection phase. Multicollinearity is detrimental to machine learning models as it can lead to unstable and unreliable estimates, inflated standard errors, and difficulties in interpreting the importance of individual predictors. To address this, we conducted a correlation analysis of the features and calculated the Variance Inflation Factor (VIF). VIF quantifies how much the variance of a regression coefficient increases due to multicollinearity among predictors, thereby measuring the extent of redundancy in the model (27). High VIF values indicate strong correlations between predictors and suggest the need for feature removal or transformation to stabilize model interpretation.

2.3 Statistical Analysis

To assess statistical differences in the QRS angles φ_{d_1} and φ_{d_2} in every plane, maximum norms $\|\mathbf{v}_{d1xy}^{\max}\|$ and $\|\mathbf{v}_{d2xy}^{\max}\|$ in every plane and $t_{V_{\max}}$ intervals between *Healthy* subjects and those patients with *LBBB* or *sLBBB*, we performed a non-parametric Kruskal-Wallis test. The Dunn’s test with Holm-Bonferroni correction was then employed for post-hoc pairwise comparisons when Kruskal-Wallis showed significant differences. Statistical significance was considered for $p < 0.05$.

3 3. Results

The segmentation of the VCG waveforms and loops produced clearly separated patterns across *Healthy*, *LBBB* and *sLBBB* patients. In healthy subjects, the time for peak velocity appeared in the delayed portions of the QRS complex, with a mean of 50 ms before QRS offset on average. This produced d_1 intervals much longer than their d_2 counterparts. Conversely, the time for peak velocity of *LBBB* patients appeared earlier and those of *sLBBB* even much earlier, at a mean of 91 ms and 134 ms with respect to the QRS offset, respectively. Figure 2a, Figure 2b, Figure 2c compare lead x and electrical conduction velocity for those patients chosen in Figure 2. Once again, the increasingly longer $t_{V_{\max}-QRS_{offset}}$ is evident from *Healthy* to *sLBBB* patients. A significantly extended d_2 interval might indicate a more proximal disruption in electrical conduction. These findings highlight differences in electrical conduction dynamics among the groups, consistent with the variations observed in the previously described VCG loops.

In addition, to compare $t_{V_{\max}}$ loop segmentation among the three classes analyzed, Figure 3 presents a paradigmatic segmented VCG loop for each cardiac condition: a *Healthy* subject (Figure 3b), a *LBBB* (Figure 3c), and a *sLBBB* patient (Figure

3d). In panel 3b, d_1 is the longest segment, reflecting preserved velocity activation during ventricular depolarization. In contrast, panel 3c shows a shorter d_1 interval, indicating a premature delay in ventricular conduction for this representative *LBBB* subject. Consistently, Panel 3d demonstrates an even shorter d_1 segment, signifying even more premature conduction impairment for a paradigmatic *sLBBB* patient. The opposite occurred with d_2 intervals. Panel 3b presents the shortest duration, indicating a rapid transition from $t_{V_{\max}}$ to the QRS offset. Conversely, in the *LBBB* patient (panel 3c), the d_2 segment is longer, and even longer in the *sLBBB* patient (panel 3d), suggesting a progressively slower transition. These findings highlight critical differences in QRS complex dynamics among the groups.

Consistently, Figure 4 represents the generalization of Figure 2 to the entire population. Here, an increasing $QRS_{offset} - t_{V_{\max}}$ can be noted from *Healthy* to *sLBBB* passing through the *LBBB* classes. This measure was significantly different in pairwise comparisons among all three classes by the Dunn’s test with Bonferroni correction, following the Kruskal-Wallis test.

Figure 5a shows the boxplots of the maximum norms from V_{\max} to the end of the QRS complex. A statistical analysis of these norms was conducted among the groups of subjects with *LBBB*, *sLBBB*, and *Healthy* controls, using the Kruskal-Wallis test and Dunn’s post-hoc test with Bonferroni correction. The Kruskal-Wallis test for d_1 yielded a statistic $H = 364.84$ with a p -value $< 5.98 \times 10^{-80}$, indicating significant differences between the groups. The Dunn’s test with Holm-Bonferroni correction for d_1 revealed significant differences between *Healthy* and *LBBB* subjects ($p = 7.24e-30$), *Healthy* and *sLBBB* ($p = 1.65e-78$), and *LBBB* and *sLBBB* ($p = 3.06e-06$). For d_2 , the Kruskal-Wallis test showed $H = 93.98$ with a p -value $< 3.91 \times 10^{-21}$. The Dunn’s test with Holm-Bonferroni correction indicated significant differences between *Healthy* and *LBBB* ($p = 1.36e-12$) and between *Healthy* and *sLBBB* ($p = 5.26e-19$), while no significant differences were observed between *LBBB* and *sLBBB* ($p = 1.00$).

Analogously, Figure 5b illustrates the statistical analysis of QRS angles in the xz plane, showing significant differences among the study groups: *Healthy*, with left bundle branch block (*LBBB*), and with strict left bundle branch block (*sLBBB*), evaluated at two distinct time points (d_1 and d_2). The Kruskal-Wallis test for d_1 revealed a statistic $H = 434.57$ with a p -value $< 4.32 \times 10^{-95}$, indicating substantial variability between groups. Dunn’s post-hoc tests highlighted significant differences between *LBBB* and *sLBBB* ($p = 1.06e-79$), as well as between *Healthy* and *LBBB* ($p = 2.48e-59$). However, for d_2 , the analysis found no significant differences ($p = 0.39$).

3.1 Identification of relevant Bio-inspired Features

In the context of classification models, situations where two or more features are highly correlated with each other can lead to several issues that affect the performance and interpretation of the model. Our multicollinearity analysis revealed high correlations (exceeding 75% in some cases) between several maximum norms and angles, respectively. Consequently, we decided to eliminate some of these features. To determine which ones to remove, we defined and fitted several basic models (based on XGBoost and Random Forest) and performed a feature importance analysis using various methods (SHAP Feature Importance (28), Random Forest Split Entropy Importance (29) and Univariate Selection Tests (30), among others. This comprehensive approach ensured that we retained the most informative features while mitigating the effects of multicollinearity on our models.

Feature elimination was conducted in multiple complementary steps to ensure statistical rigor. First, pairwise correlations were inspected, and one feature of each pair with $|r| > 0.75$ was removed. The Variance Inflation Factor (VIF) was then computed for the remaining features, and only those with extremely high VIF values ($VIF > 50$) were excluded to preserve potentially informative but correlated features for further analysis. Subsequently, several independent methods were applied to evaluate feature relevance:

1. **Random Forest feature importance**, which ranks predictors according to their mean contribution to impurity reduction across all decision trees, reflecting how frequently and effectively each variable is used to split the data.
2. **SHAP values**, which quantify each feature’s mean marginal contribution to the model prediction, allowing class-specific interpretability.
3. **Univariate selection tests (ANOVA)**, which assess the statistical association between each feature and the output class.
4. **Recursive Feature Elimination (RFE) with cross-validation**, a backward selection procedure that iteratively removes the least important features based on model performance. The cross-validation accuracy curve exhibited a clear inflection (“elbow”) at four features, indicating that this number achieved nearly the highest accuracy while minimizing model complexity.

Additionally, kernel density estimates (KDE) by class were inspected to confirm discriminative patterns, particularly between *LBBB* and *sLBBB*. The final subset of four features was selected by consensus among these methods, prioritizing variables

recurrently ranked as most important and physiologically interpretable. This multi-method approach aligns with the recommendations of Guyon and Elisseeff (31), who emphasize that there is no single universally valid framework for feature selection. Instead, combining complementary statistical, wrapper, and embedded methods provides a more robust and generalizable basis for selecting relevant predictors.

From the former analysis, we extracted a reduced set of just 4 features that achieved comparable classification metrics as the full 15-feature dataset, as will be explained in the next section. The reduced subset of features was composed of:

- the time of peak velocity to QRS onset ($t_{v_{\max}-QRS_onset}$)
- the time of peak velocity to QRS offset ($QRS_{offset} - t_{v_{\max}}$)
- the maximum norm of the d_1 segment in the xy plane ($\|\mathbf{v}_{d_1xy}^{\max}\|$)
- the angle of the d_1 dominant vector in the xz plane (φ_{d_1xz})

3.2 Models

To evaluate the performance of various machine learning models on our 15-feature dataset, we employed an automated machine learning (AutoML) tool, specifically PyCaret (32), to systematically assess a wide range of algorithms. Tree-based ensemble models such as Random Forest, Gradient Boosting, and XGBoost were prioritized because extensive empirical evidence shows that they outperform deep neural networks on structured, low-dimensional tabular datasets typical of biomedical studies (33; 34; 35). Deep learning architectures were not adopted since our dataset comprised fewer than one thousand subjects, a regime where such models tend to overfit and provide limited interpretability. The experimental setup utilized a 10-fold cross-validation (CV) strategy with an 80/20 train-test split. We noticed a notable superiority of ensemble-based decision tree models and gradient boosting algorithms in this particular classification task. Specifically, the Gradient Boosting Classifier (GBC) demonstrated the highest overall performance, achieving an accuracy of 0.8324, matched by its sensitivity, along with a precision of 0.8382 and an F1-score of 0.8297. The same procedure was applied to the reduced 4-feature dataset (described in the section “Identification of relevant Bio-inspired Features”), with the results shown in Table 1. The outcomes were remarkably similar to those obtained with the extended dataset, suggesting that the reduced 4-feature model retains the discriminatory power of the extended dataset.

The reduction in dimensionality from 15 features to a mere 4, achieved with just a small decrease in classification accuracy, represents a significant advancement in

our research. Firstly, it facilitates the development of an explainable model, wherein a concise set of rules can be derived to elucidate individual classification outcomes. This interpretability is particularly important for medical practitioners and specialists who need to comprehend the reasoning behind each prediction. Secondly, the reduced feature space enables a more tractable analysis of feature importance and their global impact on the classification process. To this end, we employed state-of-the-art interpretability techniques such as SHAP (SHapley Additive exPlanations) and analogous methodologies, as will be seen in the next sections.

Table 1: Comparison of 4-feature Machine Learning Models.

Abrev.	Model	Accuracy	Recall	Prec.	F1
et	Extra Trees Classifier	0.8343	0.8343	0.8331	0.8281
gbc	Gradient Boosting Classifier	0.8270	0.8270	0.8286	0.8244
rf	Random Forest Classifier	0.8269	0.8269	0.8258	0.8211
lda	Linear Discriminant Analysis	0.8216	0.8216	0.8151	0.8135
catboost	CatBoost Classifier	0.8178	0.8178	0.8183	0.8142
xgboost	Extreme Gradient Boosting	0.8124	0.8124	0.8134	0.8095
lightgbm	Light Gradient Boosting Machine	0.8033	0.8033	0.8074	0.8030
qda	Quadratic Discriminant Analysis	0.7997	0.7997	0.7989	0.7896
nb	Naive Bayes	0.7906	0.7906	0.7893	0.7832
ada	Ada Boost Classifier	0.7687	0.7687	0.7820	0.7707
dt	Decision Tree Classifier	0.7505	0.7505	0.7606	0.7528
ridge	Ridge Classifier	0.7233	0.7233	0.6297	0.6456
lr	Logistic Regression	0.7160	0.7160	0.6753	0.6823
knn	K Neighbors Classifier	0.6811	0.6811	0.6496	0.6576
svm	SVM - Linear Kernel	0.6141	0.6141	0.5567	0.5328

To obtain the definitive models, an ensemble of decision trees was employed, optimized through a hyperparameter search across 864 candidate configurations using 5-fold CV. The results of this fine-tuning process for both the original and reduced datasets are presented in Table 2, compared with the 15-feature model, the 4-feature model showed small absolute decreases across metrics (e.g., accuracy 0.87 vs 0.85; macro-F1 0.85 vs 0.83). Once again, it can be observed that the metrics obtained by the optimized models are remarkably similar in both cases, with the 15-feature model being slightly superior to the 4-feature model.

Although combinations of weaker classifiers in a meta-algorithmic or ensemble framework could theoretically enhance predictive accuracy, such approaches were not pursued in this work since our main objective was to preserve interpretabil-

ity while maximizing accuracy. By restricting the model to a compact set of 15 features, reduced to only 4 with small performance loss, we ensured compatibility with explainability techniques such as SHAP values and Anchors, which provide a clearer physiological interpretation of model decisions. Increasing the complexity of the model would likely reduce this clarity and hinder the extraction of clinically meaningful rules.

Table 2: Comparison of final finetuned models.

Model	Class	Precision	Sensitivity	F1-score	Accuracy
rf-15feat	0	0.94	1.00	0.97	0.87
	1	0.77	0.65	0.71	
	2	0.84	0.87	0.85	
rf-4feat	0	0.95	0.97	0.96	0.85
	1	0.72	0.70	0.71	
	2	0.83	0.83	0.83	

3.3 Explainability

Explainable Artificial Intelligence (xAI) was utilized to gain insights into the individual contributions of the four features contributing to the final model. Explainable AI was incorporated as a central component of the study design. Beyond enhancing interpretability, the goal was to derive rule-based insights grounded in cardiac electrophysiology, enabling clinicians to relate the model’s decisions to measurable conduction phenomena. By means of these tools, we were able to define the specific ranges for each feature and each class, as well as to define feature importance globally and by class. Moreover, we obtained sets of rules that explained the models’ decisions for individual and global instances. These rules proved to be very useful, since they could be applied by the physician in a straightforward and simple way, beyond any computational algorithm.

We can conclude via xAI that the most important feature for $sLBBB$ classification was $QRS_{offset} - t_{V_{max}}$, while $\varphi_{d_{1xz}}$ was the most important feature for $LBBB$ classification.

To interpret the decision process of the classification models, we used SHAP (SHapley Additive exPlanations) values, which quantify the contribution of each feature to the model’s output for each individual prediction. SHAP assigns an additive importance value to each feature based on cooperative game theory, comparing the model output with and without each feature.

For each patient, the model prediction $f(x)$ was decomposed as:

$$f(x) = \mathbb{E}[f(x)] + \sum_i \text{SHAP}_i,$$

where $\mathbb{E}[f(x)]$ is the mean model output (baseline prediction) and SHAP_i represents the contribution of the i -th feature to the deviation from the mean prediction.

In this study, SHAP values were computed using the *TreeExplainer* algorithm for the Random Forest model trained on the four selected features: $t_{V_{\max}} - t_{QRS_{\text{onset}}}$, $QRS_{\text{offset}} - t_{V_{\max}}$, $\|\mathbf{v}_{d1xy}^{\max}\|$, φ_{d1xz} .

Figure 6 shows such feature importance by class, with the contribution for each class visually grouped and labeled in the legend, for two different models. The classification power for the *Healthy* class is plotted, with an area proportional to its discrimination for each feature. Accordingly, other areas show the classification power for the *sLBBB* class and for *LBBB* discrimination. It is important to remark that this feature importance was present regardless of the model. In fact, the same order of feature importance was determined on each model separately. Thus, both models (XGboost and Random Forest) assessed $QRS_{\text{offset}} - t_{V_{\max}}$ as the best feature for *sLBBB* classification, and φ_{d1xz} as the most important feature for *LBBB* on both models. Regarding the third class, *Healthy*, discordance on the most important feature was found when comparing the models.

Thanks to the use of explainable techniques, it becomes possible to infer the most likely class of a given observation based on the specific range of values taken by each of the four features. This relationship is clearly illustrated in the dependence plot, where distinct classes occupy different regions of the feature space.

To further explore these inter-class differences, Figure 7 presents the distribution of the feature $QRS_{\text{offset}} - t_{V_{\max}}$ across the three classes. As shown, the distribution progressively shifts toward higher values as the class transitions from healthy to LBBB and then to SLBBB. In particular, the healthy class can be characterized by values between 0 and 0.08 ms, the LBBB class by values between 0.08 and 0.10 ms, and the SLBBB class by values ranging from 0.10 to 0.16 ms. Notably, these ranges are consistent with the simplified decision rules previously identified for the SLBBB class (see Table 3).

3.4 Clustering

To evaluate whether these four physiological features tended to naturally group similarly to Strauss criteria, we implemented a Spectral Clustering algorithm on these

features and clustered them into three classes. The result is displayed in Figure 8. Here, samples for each class are represented in different colors for each pair of features $\{t_{V_{\max}} - QRS_{onset}$ vs $QRS_{offset} - t_{V_{\max}}\}$ for the upper panels and $\{\varphi_{d_1xz}$ vs $\|\mathbf{v}_{d_1xy}^{\max}\|\}$ for the lower panels. The spectral clustering algorithm presented quite similar classes as the Strauss criteria visually. To quantify this natural grouping of the features, we computed the centroids for each cluster and compared them to the actual grouping by Strauss criteria. Centroids for the *Healthy* true classes and their corresponding clusters coincided for both pairs of features (upper and lower panels). True and clustered classes for *LBBB* (class 1) however, presented the most distant centroids, as can be seen in Figure 8, black square markers.

3.5 Anchors

The anchor method (36) is an *xAI* technique that explains individual predictions of any black box classification model. Anchors provide rule-like explanations that are easy to understand by highlighting parts of the input data that are sufficient to "anchor" the model's prediction. This means that changes in other feature values not present in the rule predicates do not affect the prediction.

The method was explicitly designed to be highly precise, ensuring that the generated rules, when satisfied, lead to the same prediction with a (customizable) precision level.

As an example, this is the Anchor obtained for sample1 in our dataset:

```

IF ( $QRS_{offset} - t_{V_{\max}} \leq 0.04$  AND
 $\varphi_{d_1xz} \leq 37.37^\circ$ ) THEN
PREDICT: Healthy
WITH PRECISION: 0.9845
AND COVERAGE: 0.2428

```

Precision represents the relative number of correct predictions. In this case, approximately 98% of the samples for which the anchor holds are classified as class *Healthy*. The coverage refers to the proportion of instances that satisfy the anchor conditions. By incorporating coverage, Anchors aim to produce explanations that are not only precise but also sufficiently general to be meaningful and actionable for human interpretation.

Anchor rules were generated in a stratified cross-validation setting to maximize precision subject to coverage, then consolidated by removing duplicates and retaining

rules with 95% precision and highest coverage. Thresholds (e.g., $QRS_{offset} - t_{V_{max}} > 0.12$ s; $\varphi_{d_{1xz}} \leq 37.37^\circ$) arise from this optimization, not from manual tuning. Table 3 shows rule precision, coverage, and the number of supporting samples.

However, Anchor Explanations provide local interpretability, meaning they explain individual predictions rather than the overall behavior of the model. In an attempt to derive global explanations from local anchor-based interpretations, we employed a novel consolidation approach. Initially, we generated anchors for all samples in the dataset using the Anchor algorithm. The Anchor algorithm obtained a total of 210 different anchors for the whole dataset. These local explanations were then aggregated to obtain a reduced set of consolidated anchors representative of the entire dataset.

The rule consolidation process involved several steps, such as removing duplicates and incorrect predictions as well as retaining only those with over 95% precision and the highest coverage. This final step resulted in just 27 rules (10, 14, and 3 for each class, respectively). The reduced set of anchors obtained for the *sLBBB* class is shown in Table 3, together with their respective precision, coverage and the number of *sLBBB* samples that satisfied it. For example, approximately 50% of the samples can be classified as *sLBBB* based solely on the condition that their $QRS_{offset} - t_{V_{max}}$ is larger than 0.12. This fact is consistent with Figure. 6, identifying $QRS_{offset} - t_{V_{max}}$ as the most important feature for the *sLBBB* class. Note also that for a more relaxed threshold on $QRS_{offset} - t_{V_{max}}$, the remaining 3 features need to be incorporated as well, producing rules with less coverage. These feature combinations basically state that the angle from the horizontal in the *xz* plane should be within the first quadrant, or anatomically, in the posterior left quadrant ($-7.94 < \varphi_{d_{1xz}} \leq 63.25$), the time of velocity peak should be located closer to the beginning than to the end of the QRS complex ($t_{V_{max}} > 0.08$ and $t_{V_{max}} - QRS_{onset} \leq 0.04$) and present a large maximum instantaneous vector in the first segment of the loop ($\|\mathbf{v}_{d_{1xy}}^{max}\| \leq 0.40$). It is worth remarking that the third rule at Table 3, includes negative angles for $\varphi_{d_{1xz}}$, which might seem inconsistent with the *sLBBB* definition. However, this rule stands for just 11 patients, represented by the outliers shown in Figure 5b (*sLBBB* class). This fact reminds us that anchors just explain the model predictions based on the data we have, whether the data is "good enough" or not. Anchors simply construct a predictive rule from these data and constrain its applicability with a coverage score.

4 4. Discussion

To improve the response to CRT, efforts were made to enhance the stimulation wave using biphasic configurations (4) and designs inspired by neurostimulators (37), with-

Table 3: Anchors for class *sLBBB*

anchor	precision	coverage	# samples
$[QRS_{offset} - t_{V_{max}} > 0.12]$	0.974 359	0.2237	154
$[-7.94 < \varphi_{d_1xz} \leq 63.25, QRS_{offset} - t_{V_{max}} > 0.08, t_{V_{max}} - QRS_{onset} \leq 0.04, \ \mathbf{v}_{d1xy}^{max}\ \leq 0.40]$	0.960 317	0.0507	37
$[\varphi_{d_1xz} \leq 37.37, QRS_{offset} - t_{V_{max}} > 0.08, t_{V_{max}} - QRS_{onset} \leq 0.04, \ \mathbf{v}_{d1xy}^{max}\ \leq 0.40]$	0.955 446	0.0221	11

out any success. Nowadays, a heavily explored approach for refining CRT is aimed at improving the diagnosis of LBBB, by means of a more rigid definition created empirically by Strauss (1). The strict Left Bundle Branch Block (LBBB) criteria defined by Strauss in 2011 focus on specific electrocardiographic (ECG) characteristics that predict a better response to Cardiac Resynchronization Therapy (CRT). These criteria help identify patients who are more likely to benefit from CRT, focusing on those with a true conduction block in the left bundle branch rather than other causes of prolonged QRS duration. These criteria were designed to differentiate between regular LBBB and "strict" LBBB, based on patient outcomes from the MADIT-CRT trial.

In this paper, we sought to find VCG-based variables that tell us about the pathophysiological differences, if any, underlying the Strauss criteria. To do this, we split up the QRS loop into two, according to the peak velocity. This procedure segmented the QRS loop into a first segment developed at preserved conduction, and a second one, developed at a much lower, non-specific conduction. On these two segments, variables such as the time of peak velocity referred to the QRS fiducial points ($QRS_{offset} - t_{V_{max}}$, $t_{V_{max}} - t_{QRS_{onset}}$), the angles ($\{\varphi_{d_1xz}, \varphi_{d_1xy}, \varphi_{d_1yz}\}$) and moduli ($\|\mathbf{v}_{d1xy}^{max}\|$, $\|\mathbf{v}_{d1xz}^{max}\|$, $\|\mathbf{v}_{d1yz}^{max}\|$) of the instantaneous vector on both loop segments in each plane were analyzed. From the former variables, four parameters coincided to provide the best classification performance with either an XGBoost or a Random Forest model: $QRS_{offset} - t_{V_{max}}$, $t_{V_{max}} - t_{QRS_{onset}}$, $\|\mathbf{v}_{d1xy}^{max}\|$ and φ_{d_1xz} (see Figure 6). From the vectorcardiographic point of view, a $QRS_{offset} - t_{V_{max}}$ that progressively increases from *Healthy* subjects to *sLBBB*, passing through *LBBB*, expresses in a very intuitive way the depolarization sequence in the three groups (Figure 4). That is, under normal conditions, the electrical activation of both ventricles starts from

their endocardial regions and spreads to the rest of the myocardium until depolarization is complete in the basal septal and basal posterolateral segments of the left ventricle, 80 msec after the depolarization has started. Conversely, in advanced left bundle branch blocks, the activation sequence starts in the right ventricle and must progress towards the LV through the interventricular septum (transeptal barrier) until it reaches the endocardium of said ventricle, which requires 50 msec. The impulse takes another 40-50 msec from the moment it reaches the left Purkinje system and the endocardium of the high septum and basal posterolateral segment, to complete the depolarization of said segments after 140 - 150 msec after the process has started. In the vectorcardiographic analysis of the ventricular activation sequence during LBBB, four phases are recorded, according to Perez Riera et al. (38). Precisely, phase III corresponds to the activation of the high septum and the basal posterolateral segment of the LV and is responsible for the initial apex of the R wave of the electrocardiogram in the left leads, and the nadir of the S wave in $V_1 - V_2$. This vector is processed slowly and is responsible for the delay of the mid-final forces of the ECG (50 ms) to a great extent. The last phase (IV) corresponds to the afferent branch of the QRS loop and is also of slow inscription. In the ECG, it represents the second apex of the R wave in the left leads and the second nadir of the S wave in $V_1 - V_2$. These last two phases explain the notches or slurring in these electrocardiographic leads. In the present work, the $QRS_{offset} - t_{V_{max}}$ expresses this mid-final delay after phase II of the VCG and precisely quantifies this delay. For LBBB, in which the mid-late delay (phases III-IV) is not so great, the presence of a residual condition in the left bundle branch, left ventricular hypertrophy or a combination of left anterior hemiblock of the His bundle with left ventricular hypertrophy could explain the findings for this group in the present work (39). Furthermore, the lack of statistical significance observed in the d_2 angles ($p = 0.39$), in stark contrast to the highly significant d_1 angles, suggests that the primary ventricular conduction disturbances differentiating these groups occur early in the depolarization process. The d_1 segment captures this initial, anomalous activation pathway, while the d_2 segment represents the later, slower non-specific conduction phase, which appears to be morphologically similar (in the xz plane) across all LBBB types once the main block is established.

On the other hand, the different AI tools implemented in this work support the same concept: the *LBBB* class appears to be an intermediate state between the *Healthy* and *sLBBB* conditions. This is reflected in several outcomes. First, all AI models yielded the poorest classification metrics for this class (see Table 2). Second, it showed the lowest importance values (see Figure 6). Finally, it displayed the most distant centroids in the clustering analysis (Figure 8), where the four physiological

features were grouped in an unsupervised manner and contrasted with the actual Strauss criteria classes. It is evident from Figure 8 that the features clustered similarly to the actual classes, with centroids being fairly close for all classes except for the LBBB group (class 1/cluster 1). This might suggest heterogeneity rather than a single pathophysiological entity. Future work should explore sub-phenotyping within LBBB using richer clinical covariates.

It is worth mentioning at this point, that the models tested in this piece of work produced ternary classification performances comparable to those binary classification performances compiled in the International Society for Computerized Electrocardiology in its initiative for automated LBBB detection in terms of accuracy, sensitivity and precision (2). In this way, the best accuracy reported in the 7 participants group was of 82% (with 69% sensitivity and 87% precision) (40). Moreover, with only these 4 features, we obtained better metrics than in a previous work, who utilized sets of 7 or 19 physiological features for a ternary classification scheme (17). It is crucial to construct good models when attempting to explain them. Explainable AI is useful and can be reliable only when the AI models themselves are useful and reliable (12)

4.1 Study limitations and future work

It is also important to contextualize our methodological approach. The primary goal of this study was to isolate and validate a set of novel VCG features as fundamental pathophysiological markers. To achieve this rigorously, our analysis was based on a single representative beat per patient. This approach allows for the establishment of a clear baseline, confirming with confidence that the observed discriminatory power comes directly from the proposed features by controlling for the variable of inter-beat dynamics. While this approach is complete for the stated objective, the analysis of multiple beats represents a valuable subsequent step. In fact, the study of beat-to-beat variability is the logical next step that builds upon the solid foundation we have established. Analyzing the variance of our features could reveal "meta-features" with important clinical information about the stability of the block or the severity of the pathology, representing a very promising future line of research that stems directly from our findings.

Finally, a novel approach to xAI was explored to provide a simplified set of rules to quickly classify these three classes relying on the four physiological features (see Table 3). It is important to note that Anchors do not necessarily incorporate all features from the dataset, instead focusing on a select few. This characteristic imparts a sense of hierarchical importance to the rules governing model predictions.

Moreover, consolidated Anchor rules can identify subsets of feature values that are sufficient conditions for the model to return a particular classification, providing a more granular understanding of feature interactions and their impact on model outcomes. We strongly believe that this technique might enhance the transparency and interpretability of AI models used in electrocardiology. However, this study is retrospective and draws on two existing datasets with subsampling and occasional manual delineation. Generalizability beyond these cohorts is unknown. In particular, it would be desirable to correlate the temporal features $t_{V_{\max}} - QRS_{onset}$ and $QRS_{offset} - t_{V_{\max}}$ with the activation patterns recorded during coronary mapping in an electrophysiological study, to achieve a more robust experimental validation. Thus, before any clinical implementation, prospective, multi-center validation with pre-registered analysis is required, including calibration, decision-curve analysis, and assessment of net clinical benefit.

Additionally, future work will include a systematic analysis of misclassified cases to investigate whether these errors reflect physiological overlap or outlier behavior within the *LBBB* and *sLBBB* classes. Such analysis may help refine feature boundaries and improve the physiological interpretability of the classification model.

5 5. Conclusions

This study advances the understanding of the physiological differences between regular and strict Left Bundle Branch Block by analyzing vectorcardiographic data using machine learning and explainable AI techniques, aligning with recent research trends in the International Journal of Neural Systems that emphasize explainable AI in medical applications (13; 14; 15; 16). Our work connects to the journal’s recent focus on transparent machine learning for clinical decision support across various medical domains, from EEG-based impulsivity classification to cognitive impairment assessment and neonatal seizure detection.

This explainable AI framework offers several contributions to clinical practice. Firstly, it improves diagnostic discrimination: The model identifies four key physiological features (conduction velocity-related time intervals, maximum VCG loop norm in the frontal plane, and QRS angle in the horizontal plane) that allow for robust differentiation between *LBBB*, *sLBBB*, and *Healthy* subjects, with accuracy comparable to more complex models but with greater interpretability. Secondly, simple and verifiable clinical rules are generated: Thanks to explainable artificial intelligence (XAI) techniques, clear and practically applicable rules are derived, facilitating decision-making by the clinician without depending on "black box" models,

similar to recent XAI approaches published in this journal (14; 15). The study successfully identified a reduced set of four key features that classify these groups effectively, with the latter features composing a set of three simple rules to represent the *sLBBB* class with a mean precision of 96% of all samples fulfilling the rule.

Our methodological approach builds upon established computational cardiology research (7; 8) while advancing the field through explainable AI. For future extensions of this research, we plan to explore recently developed powerful classification algorithms such as Neural Dynamic Classification (19), Finite Element Machines for fast learning (20), Dynamic Ensemble Learning Algorithms (21), and self-supervised learning approaches (22) that have shown promising results in electrophysiological signal analysis. These advanced computational techniques could further enhance the accuracy and robustness of our classification framework while maintaining clinical interpretability.

The clinical implications of this work include potential improvements in selecting CRT candidates by accurately identifying *sLBBB*, which best predicts response to CRT, thereby helping select patients with the greatest likelihood of benefit and avoiding unnecessary interventions. More precise classification may also help identify candidates for other cardiac pacing modalities, such as His bundle pacing. While this work remains a retrospective analysis requiring prospective validation, it provides a quantitative and transparent framework that could improve the diagnosis of *LBBB* and *sLBBB*, facilitating accurate selection of CRT candidates and promoting personalized, evidence-based medicine.

The submission of this work to the International Journal of Neural Systems is justified by its alignment with the journal’s recent emphasis on explainable AI in healthcare, its connection to advanced neural systems approaches for medical signal analysis, and its contribution to the growing body of research that bridges machine learning innovation with clinical applicability in cardiovascular medicine.

6 Acknowledgments

This project was funded in part by grant by grants DTS19/00175 and PDC2022-133952-100 funded by the Spanish “Ministerio de Ciencia, Innovación y Universidades” and by the European Union’s Horizon 2020 Research and Innovation Programme MSCA-SE EPISTEAM, and by Programa Iberoamericano de Ciencia/Tecnología para el Desarrollo (CYTED) (Red 225RT0169). This research has been funded by a PhD scholarship from the National Council of Science and Technology (CONICET).

References

- [1] Strauss R David Gand Selvester and Wagner G. Defining left bundle branch block in the era of cardiac resynchronization therapy. *American Journal of Cardiology* ; 107: 927–934.
- [2] Zusterzeel R, Vicente J, Ochoa-Jimenez R et al. The 43rd international society for computerized electrocardiology ecg initiative for the automated detection of strict left bundle branch block. *Journal of Electrocardiology* 2018; 51(6, Supplement): S25–S30.
- [3] Pérez-Riera AR, Barbosa-Barros R, Daminello-Raimundo R et al. Re-evaluating the electro-vectorcardiographic criteria for left bundle branch block. *Annals of Noninvasive Electrocardiology* 2019; 24(4): e12644. DOI:10.1111/anec.12644.
- [4] Ortega D, Barja L, Logarzo E et al. Nonselective his bundle pacing with a biphasic waveform: enhancing septal resynchronization. *EP Europace* 2018; 20(5): 816–822.
- [5] Ortega D, Logarzo E, Barja L et al. Novel implant technique for septal pacing. a noninvasive approach to nonselective his bundle pacing. *Journal of Electrocardiology* 2020; 63: 35–40. DOI:<https://doi.org/10.1016/j.jelectrocard.2020.09.008>.
- [6] Ansari Y, Mourad O, Qaraqe K et al. Deep learning for ecg arrhythmia detection and classification: An overview of progress for period 2017–2023. *Frontiers in Physiology* 2023; 14: 1246746. DOI:10.3389/fphys.2023.1246746.
- [7] Martis R, Acharya U and Adeli H. Current methods in electrocardiogram characterization. *Computers in Biology and Medicine* 2014; 48: 133–149. DOI: 10.1016/j.combiomed.2014.03.003.
- [8] Sankari Z and Adeli H. Heartsaver: A mobile cardiac monitoring system for auto-detection of atrial fibrillation, myocardial infarction and atrio-ventricular block. *Computers in Biology and Medicine* 2011; 41(4): 211–220.
- [9] Amezcua-Sanchez J, Valtierra-Rodriguez M, Adeli H et al. A novel wavelet transform-homogeneity model for sudden cardiac death prediction using ecg signals. *Journal of Medical Systems* 2018; 42: 176. DOI:10.1007/s10916-018-1031-5.

- [10] Murugappan M, Murugesan L, Jerritta S et al. Sudden cardiac arrest (sca) prediction using ecg morphological features. *Arabian Journal of Science and Engineering* 2021; 46(2): 947–961.
- [11] Ayano YM, Schwenker F, Dufera BD et al. Interpretable machine learning techniques in ecg-based heart disease classification: A systematic review. *Diagnostics* 2023; 13(1): 111. DOI:10.3390/diagnostics13010111.
- [12] Górriz J, Álvarez Illán I, Álvarez Marquina A et al. Computational approaches to explainable artificial intelligence: Advances in theory, applications and trends. *Information Fusion* 2023; 100: 101945. DOI:https://doi.org/10.1016/j.inffus.2023.101945.
- [13] Hüpen P, Habel U, Shymanskaya A et al. Impulsivity classification using eeg power and explainable machine learning. *International Journal of Neural Systems* 2023; 33(2): 2350006.
- [14] Jiménez-Mesa C, Arco JE, Valentí-Soler M et al. Using explainable artificial intelligence in the clock drawing test to reveal the cognitive impairment pattern. *International Journal of Neural Systems* 2023; 33(4): 2350015.
- [15] Raeisi K, Khazaei M, Tamburro G et al. A class-imbalance aware and explainable spatio-temporal graph attention network for neonatal seizure detection. *International Journal of Neural Systems* 2023; 33(9): 2350036.
- [16] Mercaldo F, Di Giammarco M, Ravelli F et al. Alzheimer’s disease evaluation through visual explainability by means of convolutional neural networks. *International Journal of Neural Systems* 2024; 34(2): 2450007.
- [17] Macas B, Garrigós J, Martínez JJ et al. An explainable machine learning system for left bundle branch block detection and classification. *Integrated Computer-Aided Engineering* 2024; 31: 43–58.
- [18] Macas Ordóñez BdC, Orellana Villavicencio DV, Suing Ochoa MA et al. Graph theory and its potential in the automatic detection of left bundle branch block. *Integrated Computer-Aided Engineering* 2025; DOI:10.1177/10692509251362613. Forthcoming.
- [19] Rafiei M and Adeli H. A new neural dynamic classification algorithm. *IEEE Transactions on Neural Networks and Learning Systems* 2017; 28(12): 3074–3083.

- [20] Pereira D, Piteri M, Souza A et al. Fema: A finite element machine for fast learning. *Neural Computing and Applications* 2020; 32(10): 6393–6404.
- [21] Alam K, Siddique N and Adeli H. A dynamic ensemble learning algorithm for neural networks. *Neural Computing with Applications* 2020; 32(10): 8675–8690.
- [22] Rafiei M, Gauthier L, Adeli H et al. Self-supervised learning for electroencephalography. *IEEE Transactions on Neural Networks and Learning Systems* 2024; 35(2): 1457–1471.
- [23] Moss AJ, Brown MW, Cannom DS et al. Multicenter automatic defibrillator implantation trial-cardiac resynchronization therapy (madit-crt): design and clinical protocol. *Annals of Noninvasive Electrocardiology* 2005; 10(4 Suppl): 34–43. DOI:10.1111/j.1542-474X.2005.00073.x.
- [24] Zheng J, Guo H and Chu H. A large scale 12-lead electrocardiogram database for arrhythmia study (version 1.0.0). <http://physionet.org/content/ecg-arrhythmia/1.0.0/>, 2022.
- [25] Edenbrandt L and Pahlm O. Vectorcardiogram synthesized from a 12-lead ecg: Superiority of the inverse dower matrix. *Journal of Electrocardiology* 1988; 21: 361–367.
- [26] Ledezma CA. WTdelineator , 2021. URL <https://github.com/caledezma/WTdelineator>.
- [27] Daniel C and Wood FS. *Fitting Equations to Data: Computer Analysis of Multifactor Data*. New York: John Wiley & Sons, 1971. ISBN 9780471193503.
- [28] Lundberg SM and Lee SI. A unified approach to interpreting model predictions. In Guyon I, Luxburg UV, Bengio S et al. (eds.) *Advances in Neural Information Processing Systems 30*. Curran Associates, Inc., 2017. pp. 4765–4774.
- [29] Menze BH, Kelm BM, Masuch R et al. A comparison of random forest and its gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinformatics* 2009; 10(1): 213. DOI: 10.1186/1471-2105-10-213.
- [30] Jović A, Brkić K and Bogunović N. A review of feature selection methods with applications. In *2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)*. pp. 1200–1205. DOI:10.1109/MIPRO.2015.7160458.

- [31] Guyon I and Elisseeff A. An introduction to variable and feature selection. *Journal of Machine Learning Research* 2003; 3: 1157–1182. URL <https://jmlr.org/papers/volume3/guyon03a/guyon03a.pdf>.
- [32] Ali M. *PyCaret: An open source, low-code machine learning library in Python*, 2020. URL <https://www.pycaret.org>. PyCaret version 1.0.
- [33] Grinsztajn L, Oyallon É and Varoquaux G. Why do tree-based models still outperform deep learning on typical tabular data? In *Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS 2022)*. p. –. URL <https://arxiv.org/abs/2207.08815>. ArXiv preprint arXiv:2207.08815.
- [34] Shwartz-Ziv R and Armon A. Tabular data: Deep learning is not all you need. *arXiv preprint* 2021; arXiv:2106.03253. URL <https://arxiv.org/abs/2106.03253>.
- [35] Borisov V, Leemann T, Seßler K et al. Deep neural networks and tabular data: A survey. *arXiv preprint* 2021; arXiv:2110.01889. URL <https://arxiv.org/abs/2110.01889>.
- [36] Ribeiro MT, Singh S and Guestrin C. Anchors: High-precision model-agnostic explanations. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32.
- [37] Ferrandez JM, Liano E, Bonomini P et al. A customizable multi-channel stimulator for cortical neuroprosthesis. In *2007 29th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. pp. 4707–4710. DOI: 10.1109/IEMBS.2007.4353390.
- [38] Pérez-Riera AR, Ferreira C, Ferreira Filho C et al. Electrovectorcardiographic diagnosis of left septal fascicular block: Anatomic and clinical considerations. *Annals of Noninvasive Electrocardiology* 2011; 16(2): 196–207.
- [39] Grant R and Dodge H. Mechanisms of qrs complex prolongation in man—left ventricular conduction disturbances. *American Journal of Medicine* 1956; 20: 834–852.
- [40] Smisek R, Viscor I, Jurak P et al. Fully automatic detection of strict left bundle branch block. *J Electrocardiol* 2018; 51: S31–S34.

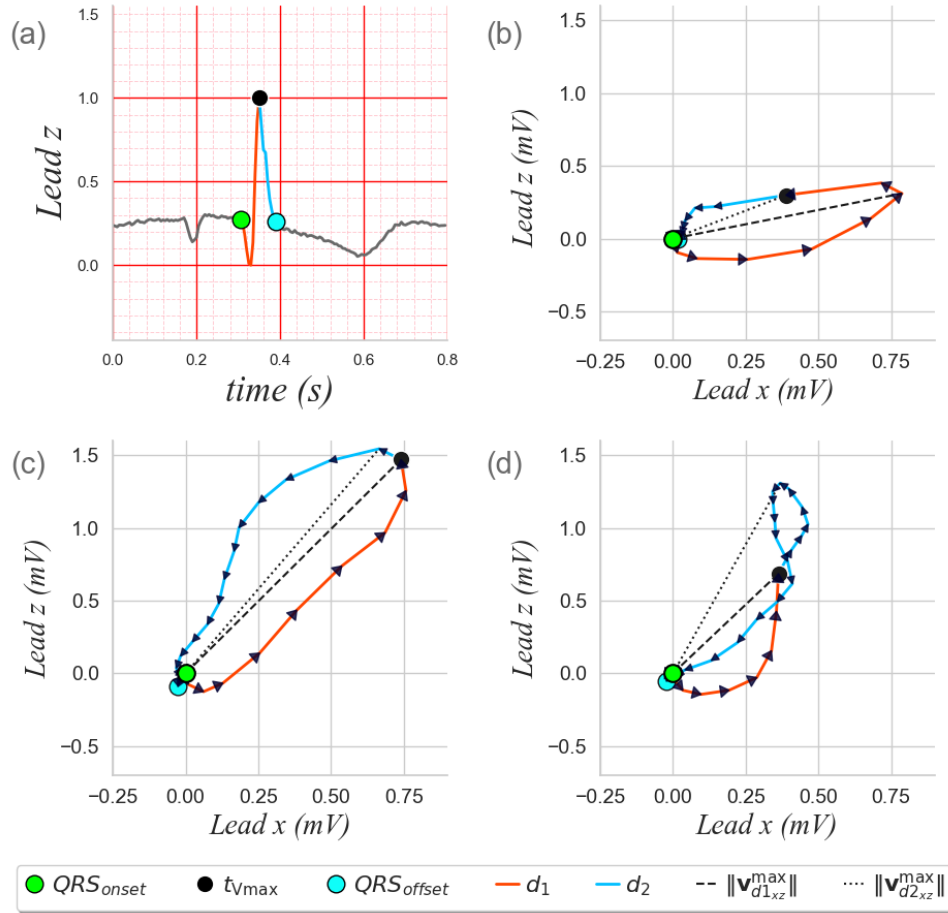


Figure 3: a) Morphological segmentation of the VCG lead z according to the three fiducial points QRS_{onset} , QRS_{offset} and t_{Vmax} . Vectorcardiographic loops segmented into d_1 and d_2 intervals for b) *Healthy*, c) *LBBB*, and d) *sLBBB* patients.

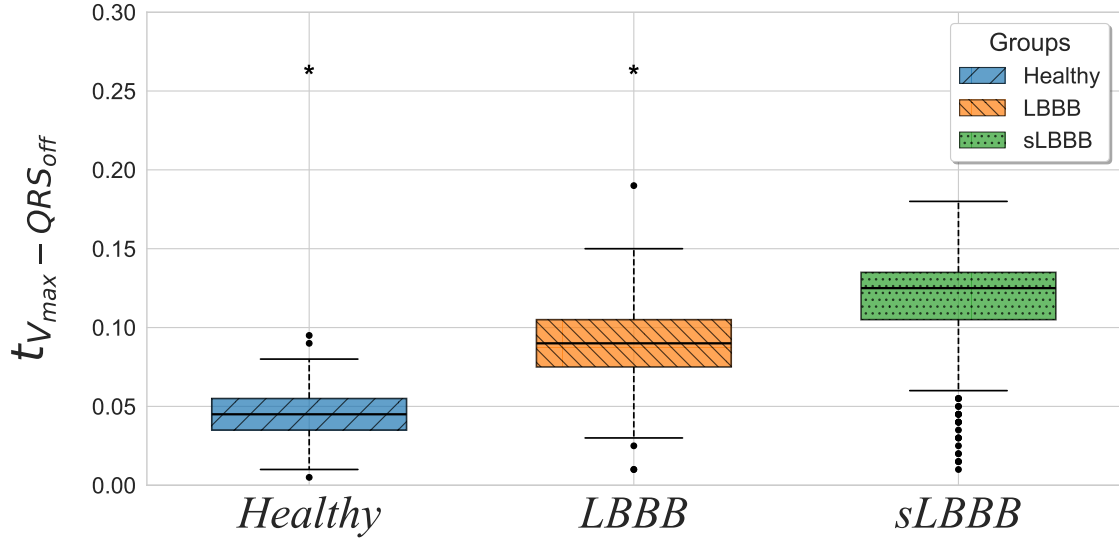


Figure 4: Boxplot representation of time to peak velocity with respect to QRS offset ($QRS_{\text{offset}} - t_{V_{\max}}$) for *Healthy*, *LBBB*, and *sLBBB* patients. * $p < 0.05$ among all classes, from Dunn's test with Bonferroni correction, following the Kruskal-Wallis test.

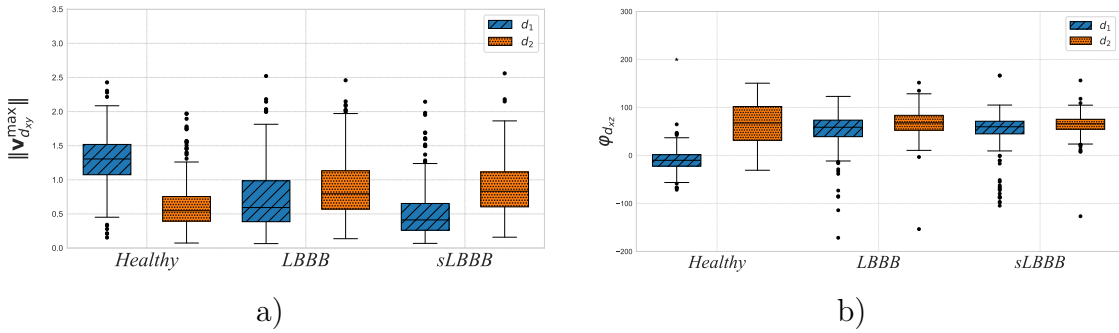


Figure 5: Boxplots for physiological features in *Healthy*, *LBBB*, and *sLBBB* classes. (a) Dunn's test with Bonferroni correction following Kruskal-Wallis. d_1 : $H = 364.84$, $p < 1.00e-95$; d_2 : $H = 93.98$, $p < 1.00e-20$. Significant differences noted for *Healthy* vs *LBBB* ($p = 7.24e-30$) and *Healthy* vs *sLBBB* ($p = 1.65e-78$) in d_1 ; *Healthy* vs *LBBB* ($p = 1.36e-12$) and *Healthy* vs *sLBBB* ($p = 5.26e-19$) in d_2 . (b) Denotes significant differences ($p < 0.05$) for *Healthy* vs *LBBB* ($p = 2.48e-59$) and *Healthy* vs *sLBBB* ($p = 1.06e-79$), but not *LBBB* vs *sLBBB* ($p = 1.00$).

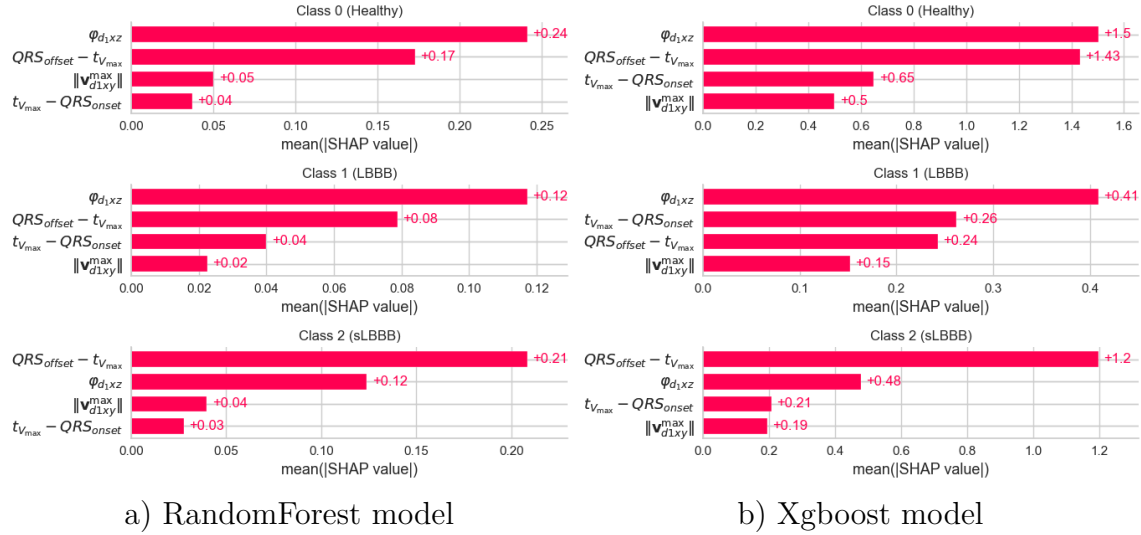


Figure 6: Feature importances by class for a) RandomForest and b) XGboost classifiers. Predictive power is represented by the length of the bars for each class (grouped and labeled in the legend). Notice the consistency in feature importance comparing both models.

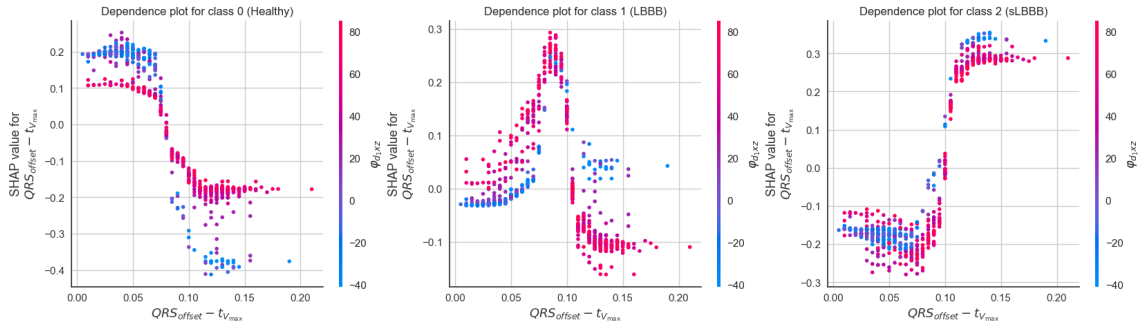


Figure 7: Dependence plots for feature $QRS_{offset} - t_{V_{max}}$ in the 4-feature Random Forest model, for each of the 3 classes. Positive SHAP values indicate that the corresponding feature values increase the model’s likelihood of predicting that class. In this case, small values of $QRS_{offset} - t_{V_{max}}$ (below 0.07) are positively associated with the *Healthy* class, moderate values (0.07 – 0.11) contribute to *LBBB*, and large values (above 0.12) contribute positively towards *sLBBB* class prediction, while decreasing the probability of the other classes.

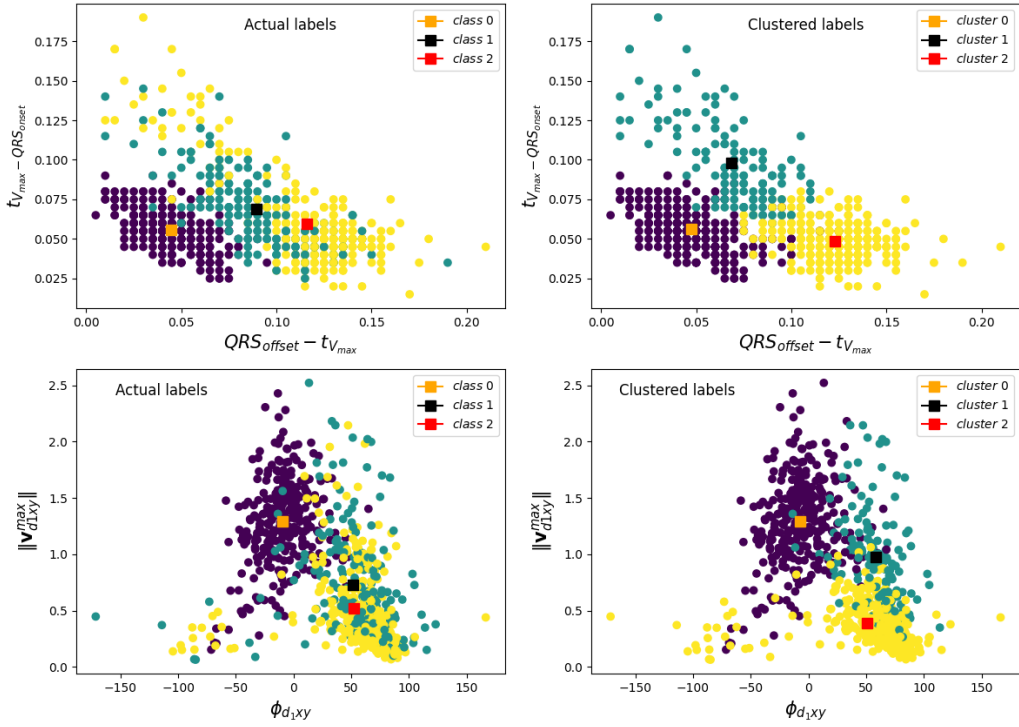


Figure 8: Spectral cluster for *Healthy*, *LBBB*, and *sLBBB* patients based on the 4 physiological features. Samples are colored by class: *Healthy* samples (class/cluster 0), *LBBB* samples (class/cluster 1), and *sLBBB* samples (class/cluster 2). Centroids are marked by solid squares. Note that both classes/clusters 0 and 2 show very similar centroids, while class/cluster 1 does not.