Politecnico di Torino

Politecnico di Torino

Master Degree in Electronic Engineering

A.a. 2025/2026

Graduation Session July 2026

Automated Type 2 Brugada Syndrome Classification from 12-Lead ECGs: A Dual-Stream Deep Learning and Random Forest Ensemble

Supervisor:

Prof. Pasero Eros Gian Alessandro

Co-supervisor:

Randazzo Vincenzo

Candidate:

Waqas Khan

Abstract

Type 2 Brugada Syndrome (T2-BrS) is the most prevalent Brugada ECG phenotype in the general population, yet its characteristic saddleback morphology is not independently diagnostic without pharmacological provocation and specialist interpretation. No published end-to-end automated pipeline targets the binary T2-BrS versus Non-BrS classification task directly from standard 12-lead recordings. This thesis addresses that gap.

A labelled dataset of 362 patient-disjoint 12-lead ECG recordings (188 confirmed Type 2 BrS and 174 Non-BrS) was assembled through two acquisition pathways: semi-automated digitisation of approximately 1,000 paper ECGs from a Sardinian Brugada patient registry using the ECGD software tool, and direct extraction of digital signals from GE MAC2000 XML archives. Following S-peak-centred windowing and resampling to 500 Hz, a 44-dimensional clinical feature vector was constructed for each recording, encoding five published diagnostic criteria: the de Luna α/β r′-wave angular measurements, the Corrado ST curvature index, the Crea inferior-lead ST depression sign, the Babai-Bigi aVR criterion, and the Ishida repolarisation markers (Tpeak–Tend interval and r-J interval).

A dual-stream deep learning architecture, *BrugadaTwoStream*, was developed combining a CNN–Transformer encoder over raw 8-channel ECG tensors (Stream A) with a two-layer MLP over the 44-feature vector (Stream B). Stream A was initialised from a PTB-XL pretrained encoder achieving a validation AUC of 0.9919 on RBBB versus normal classification, transferring right-precordial morphology representations to the small-data T2-BrS task. An independent Random Forest (500 trees, maximum depth 6) was trained on the same feature vectors and combined with the deep learning model via a validation-set-optimised weighted ensemble (optimal DL weight $w^* = 0.30$), followed by recording-level mean aggregation and Youden's-J threshold selection.

On the held-out production test set of 35 patient-disjoint recordings, the ensemble achieved a recording accuracy of 0.857, a Type 2 BrS F1-score of 0.872 (precision 0.810, recall 0.944), and a recording-level ROC-AUC of 0.938, substantially outperforming both individual components. SHAP analysis of the Random Forest component identified absence of a deep S-wave in V6, inferior ST depression in

leads II and aVF, and J-point ST elevation in V2 as the three dominant discriminators, providing cross-study validation of the Crea inferior-lead sign and empirically challenging the primacy of the α/β angular criteria in automated classification.

The conservative error profile that is high Type 2 recall at the cost of lower Non-BrS specificity and its clinically appropriate for a screening application in which the primary risk is a missed Brugada case. The pipeline's interpretable SHAP output and modest computational footprint (training in under two minutes on CPU) indicate practical suitability as a decision-support tool in cardiac screening programmes targeting this under-diagnosed arrhythmia syndrome.

Acknowledgements

I wish to express my sincere gratitude to all those who supported me throughout the course of this thesis. My deepest thanks go to my supervisor, Professor Pasero Eros Gian Alessandro, and my co-supervisor, Randazzo Vincenzo, for entrusting me with this research topic and for their continuous guidance, constructive feedback, and generous availability at every stage of the work. I am particularly indebted to Research advisor Alessandro for his dedicated supervision, remarkable patience, and the depth of technical insight he brought to every discussion — this thesis would simply not have reached its present form without his steady encouragement and hands-on involvement.

I am profoundly grateful to Almighty ALLAH, whose countless blessings and mercy have guided me at every step of this journey. I owe a debt of gratitude that cannot be put into words to my parents, my wife, my entire family, and my friends, for the love, warmth, encouragement, and prayers they have so generously given throughout this work. Their unwavering belief in me has been my greatest source of strength and motivation.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ADC

Analogue-to-Digital Converter

AI

Artificial Intelligence

AUPRC

Area Under the Precision-Recall Curve

AUC

Area Under the ROC Curve

BCE

Binary Cross-Entropy

BN

Batch Normalisation

BrS

Brugada Syndrome

CNN

Convolutional Neural Network

Conv1d

One-Dimensional Convolutional Layer

CRBBB

Complete Right Bundle Branch Block

CV

Cross-Validation

DL

Deep Learning

DWT

Discrete Wavelet Transform

ECG

Electrocardiogram / Electrocardiography

ECGD

ECG Digitizer

FFN

Feed-Forward Network

FIR

Finite Impulse Response

FN

False Negative

FP

False Positive

fQRS

Fragmented QRS complex

GE

GE Healthcare

HSV

Hue–Saturation–Value

IRBBB

Incomplete Right Bundle Branch Block

LDA

Linear Discriminant Analysis

LPT

Longest Processing Time

ML

Machine Learning

MLP

Multi-Layer Perceptron

Non-BrS

Non-Brugada Syndrome

OOB

Out-of-Bag

PSD

Power Spectral Density

PTB-XL

Physikalisch-Technische Bundesanstalt ECG Database XL

QC

Quality Control

RBBB

Right Bundle Branch Block

RF

Random Forest

RGB

Red–Green–Blue

ROC

Receiver Operating Characteristic

RVOT

Right Ventricular Outflow Tract

SCD

Sudden Cardiac Death

SHAP

SHapley Additive exPlanations

T2-BrS

Type 2 Brugada Syndrome

T1-BrS

Type 1 Brugada Syndrome

VF

Ventricular Fibrillation

VT

Ventricular Tachycardia

WRS

Weighted Random Sampler

XML

Extensible Markup Language

Chapter 1

Introduction

1.1 Brugada Syndrome: Clinical Significance and the Diagnostic Challenge

Brugada Syndrome (BrS) is a hereditary cardiac channelopathy characterised by distinctive ECG abnormalities in the right precordial leads and an elevated susceptibility to life-threatening ventricular arrhythmias, including ventricular fibrillation (VF) and sudden cardiac death (SCD). First described in 1992 by Pedro and Josep Brugada, the syndrome accounts for an estimated 4–12% of all SCD cases and up to 20% of SCD in structurally normal hearts [1, 2]. The condition follows an autosomal dominant inheritance pattern with incomplete penetrance, predominantly manifests in adult males (male-to-female ratio approximately 8:1), and is most prevalent in South-East Asia. Global prevalence is estimated at 1–5 per 10,000 individuals, though this may be an underestimate given the dynamic and frequently concealed nature of the ECG abnormality [1, 3].

Electrocardiographically, BrS presents in three pattern types based on right precordial (V1/V2) morphology. The Type 1 (coved) pattern (J-point elevation $\geq$2 mm with a downsloping ST segment and negative T-wave) is the sole independently diagnostic pattern [1]. The Type 2 (saddleback) pattern features a secondary r′-wave deflection with an elevated ST segment $\geq$1 mm that remains positive or biphasic. The Type 3 pattern exhibits ST elevation $<$1 mm and is of lower independent diagnostic significance. Of the three, Type 2 is the most commonly encountered in the general population yet the most diagnostically challenging: the saddleback pattern overlaps substantially with benign variants such as early repolarisation, RBBB configurations, and athlete morphology. An isolated Type 2 pattern is not diagnostic; its clinical significance depends on symptom history, family history of SCD, and pharmacological provocation testing [1, 2].

1.2 Limitations of Current Type 2 Brugada Diagnosis

The confirmatory diagnostic pathway involves pharmacological provocation with sodium channel blockers (ajmaline, flecainide, or procainamide), administered intravenously under controlled conditions. A positive test (appearance of a Type 1 coved pattern) is accepted as confirmatory in the absence of a spontaneous Type 1 ECG [1]. This approach has several important limitations. First, provocation testing carries a procedural risk of inducing life-threatening arrhythmias, restricting it to specialist cardiac centres and rendering population-scale screening impractical. Second, a negative test does not definitively exclude BrS, as ECG phenotype expression is dynamically modulated by autonomic tone, body temperature, febrile illness, and medication [1].

Further compounding the diagnostic challenge is substantial inter-observer variability in visual interpretation of Type 2 morphology. The saddleback pattern overlaps with incomplete right bundle branch block (IRBBB), and even experienced electrophysiologists show significant disagreement on borderline cases [4]. Ohkubo et al. proposed a quantitative angular criterion for the r′-wave to discriminate IRBBB from Type 2/3 patterns that may evolve toward Type 1 under pharmacological challenge, but this measurement requires precise morphological analysis and is susceptible to measurement error in routine practice [4]. Collectively, these limitations define a clear unmet need: an objective, non-invasive, and scalable method for stratifying patients presenting with Type 2 BrS morphology.

1.3 The Case for Automated ECG-Based Classification

Deep learning models have demonstrated capacity to detect and classify complex ECG patterns at or beyond the performance of expert cardiologists across a range of conditions. In BrS specifically, prior automated work has concentrated on the Type 1 coved pattern. Melo et al. showed that a deep learning algorithm trained on >15,000 ECGs could detect the BrS signature in morphologically normal-appearing ECGs, achieving ROC-AUC = 0.934 [5]; however, the harder Type 2 saddleback form was not the primary focus. Reviews of the AI-for-BrS landscape confirm that virtually all published automated systems have been trained predominantly on Type 1 data [3].

Automated classification of Type 2 BrS versus non-Brugada ECGs is a substantially harder problem. The discriminative features (r′-wave angular geometry, post-J ST curvature, inferior lead auxiliary signs) are subtle and require precise

spatial measurement. Available labelled datasets are inherently small: the condition is rare, its ECG phenotype is intermittent, and annotation requires specialist cardiological review. These characteristics render Type 2 classification a domain where integrating domain knowledge through hand-engineered clinical features offers a clinically motivated complement to raw-signal deep learning [6]. This thesis addresses precisely this open gap: no published end-to-end automated pipeline exists targeting binary Type 2 BrS versus Non-BrS classification with a combination of paper ECG digitisation, clinical feature engineering grounded in established BrS morphological criteria, deep learning with transfer learning, and a validated recording-level ensemble decision framework.

1.4 Research Objectives and Contributions

This thesis presents an end-to-end automated pipeline for the classification of Type 2 Brugada Syndrome from standard 12-lead ECG recordings, organised around four primary objectives:

1. **Dataset construction:** Establish a labelled dataset of Type 2 BrS and Non-BrS 12-lead ECGs by digitising paper ECGs from a Sardinian BrS patient database and extracting digital signals from confirmed Type 2 BrS GE MAC2000 XML recordings, structured for patient-disjoint 80/10/10 splitting.

2. **Preprocessing and feature engineering:** Design a signal preprocessing pipeline (500 Hz resampling, S-peak-centred windowing, multi-channel tensor construction) and develop a 44-feature clinical framework derived directly from published Type 2 BrS criteria, spanning V1/V2 r′-wave morphology, cross-lead global markers, inferior-lead Crea sign features, and aVR auxiliary criteria.

3. **Dual-stream deep learning model:** Develop *BrugadaTwoStream*, a dual-stream architecture fusing a CNN–Transformer encoder over raw ECG signal tensors (Stream A) with an MLP processing the 44 hand-engineered features (Stream B), initialised via transfer learning from a PTB-XL RBBB pretraining task.

4. **Ensemble evaluation at recording level:** Evaluate an optimally weighted DL + Random Forest ensemble at recording level on a patient-disjoint production split, with the DL-to-RF weight determined by validation-set Non-BrS F1 maximisation.

The primary contributions are: (i) an original 362-recording dataset (188 Type 2 BrS, 174 Non-BrS); (ii) a 44-feature clinical framework grounded in the de Luna

α/β angles [7], the Corrado index, the Crea inferior-lead sign [8], and the Babai-Bigi aVR criterion; (iii) the *BrugadaTwoStream* dual-stream architecture with PTB-XL transfer learning; (iv) an optimised ensemble achieving recording-level ROC-AUC of 0.938 and accuracy of 0.857 on a 35-recording held-out test set, with Type 2 F1 of 0.872 and Non-BrS F1 of 0.839; and (v) a SHAP interpretability analysis identifying S-wave absence in V6, inferior ST depression, J-point ST elevation in V2, Tpeak-Tend interval, and r-J interval as the highest-ranked discriminators, consistent with the independent findings of Ishida et al. [9].

1.5 Thesis Organisation

Chapter 2 provides the clinical and technical background: BrS pathophysiology, epidemiology, and genetics; the established ECG diagnostic criteria for Type 2 BrS; machine learning methods for cardiac ECG classification; and a survey of prior automated BrS detection work identifying the research gap.

Chapter 3 describes dataset construction: the ECGD paper ECG digitisation workflow for Non-BrS recordings, the GE MAC2000 XML signal extraction for Type 2 BrS, and the labelling, cleaning, and patient manifest construction for patient-disjoint splitting.

Chapter 4 presents the signal preprocessing pipeline: resampling to 500 Hz, R-peak and S-peak detection, 850 ms window extraction, lead selection (V1, V2, V3, V6), and first-derivative channel construction to form the final $(8, 425)$ input tensor.

Chapter 5 describes the 44-feature clinical framework across all five feature blocks, with imputation, standardisation, and SHAP-informed feature selection experiments.

Chapter 6 presents the BrugadaTwoStream architecture, PTB-XL RBBB pretraining, layer freezing, training hyperparameters, Youden's-J threshold selection, the Random Forest ensemble, and the validation-optimised weighting procedure.

Chapter 7 reports all experimental results: training dynamics, test-set performance of all three pipeline variants, SHAP feature importance analysis, and Transformer attention entropy analysis.

Chapter 8 discusses the results in context, examines the clinical significance of the SHAP-ranked features, addresses dataset-scale limitations, and proposes future work.

Chapter 9 concludes the thesis with a consolidated summary of findings and contributions.

Chapter 2

Background and Literature Review

2.1 Brugada Syndrome: Pathophysiology, Epidemiology, and Genetics

Brugada Syndrome (BrS) was first described in 1992 by Pedro and Josep Brugada, who reported eight patients with syncopal episodes and aborted sudden cardiac death in structurally normal hearts, united by a characteristic right precordial ECG pattern [10]. In the decades since, BrS has been recognised as one of the leading causes of SCD in young adults without structural heart disease, accounting for 4–12% of all SCD cases and up to 20% in structurally normal hearts [1, 2]. Global prevalence is estimated at approximately 1 in 2,000 individuals, though the dynamic and frequently concealed nature of the ECG phenotype means true prevalence is almost certainly higher [2, 3]. Substantial geographical variation exists: BrS is strikingly more prevalent in South-East and East Asian populations, and a pronounced sex asymmetry is universally observed, with 80–91% of reported cases occurring in males [1, 2].

BrS follows an autosomal dominant inheritance pattern with incomplete penetrance and variable expressivity. Loss-of-function mutations in *SCN5A*, encoding the cardiac sodium channel Nav1.5, are identified in 18–30% of European probands, while up to 70% of clinically definite families carry no identifiable pathogenic variant, reflecting a polygenic and oligogenic architecture [1]. Two competing electrophysiological hypotheses have been advanced: the **repolarisation hypothesis** attributes the ECG findings to a transmural gradient in transient outward current (I_{to}) density favouring the right ventricular epicardium, while the **depolarisation hypothesis** emphasises localised conduction delay in the right ventricular outflow tract (RVOT)

secondary to Nav1.5 loss of function [1]. Contemporary understanding holds that both mechanisms coexist. Clinically, BrS presents across a wide spectrum from asymptomatic incidental discovery to malignant syncope, polymorphic VT, and SCD, with arrhythmic events typically occurring at rest or in the context of febrile illness and vagotonic states [1, 2].

2.2 ECG Patterns in Brugada Syndrome

2.2.1 Pattern classification

Three distinct ECG patterns are associated with BrS, each defined by the morphology observed in right precordial leads V1 and V2 (Figure 2.1).

The **Type 1 (coved) pattern** is the sole independently diagnostic ECG pattern for BrS. It is defined by a J-point elevation of $\geq$2 mm above the isoelectric line in at least one right precordial lead, followed by a coved downsloping ST segment and a negative, symmetrical T-wave [1, 7]. The ST segment declines monotonically post-J, giving a Corrado index exceeding 1 [11, 7]. A spontaneous Type 1 ECG is diagnostic; a drug-induced Type 1 requires additional clinical features for definitive diagnosis [1].

The **Type 2 (saddleback) pattern** is the most prevalent manifestation in the general population. It is characterised by a true r′-wave (a secondary positive deflection of $\geq$0.2 mV following the S-wave) succeeded by an elevated ST segment $\geq$1 mm above isoelectric, producing a characteristic saddle shape with a positive or biphasic T-wave in V2 [7, 1]. Critically, the Type 2 pattern is *not* independently diagnostic of BrS; its clinical significance depends on symptom history, family history, and the outcome of pharmacological provocation [1, 2]. The **Type 3 pattern** exhibits ST elevation of less than 1 mm and carries no independent diagnostic significance.

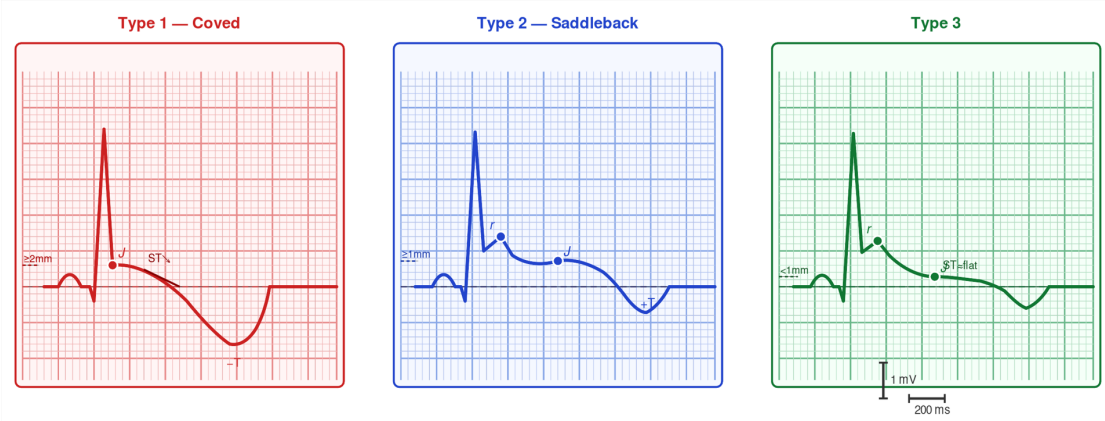

Figure 2.1: Pattern classifications

2.2.2 Electrode positioning and provocation testing

Repositioning V1/V2 to the second or third intercostal space (high-lead ECG) increases Type 1 sensitivity by up to 30% in concealed BrS [2]. In patients without a spontaneous Type 1 pattern, intravenous sodium channel blockade (ajmaline in Europe; flecainide or procainamide elsewhere) is employed to unmask a concealed Type 1, though the procedure carries a risk of provoking sustained VT/VF and requires monitored conditions [1, 2].

2.3 Diagnostic Criteria for Type 2 Brugada Syndrome

A substantial body of literature has developed quantitative morphological criteria for discriminating Type 2 BrS from IRBBB, athlete's heart, and early repolarisation. These criteria form the direct methodological basis for the 44-feature framework in Chapter 5.

r′-wave angular criteria (de Luna 2014). The r′-wave is characteristically broad in BrS (slow ascending and descending limbs) versus narrow and pointed in IRBBB. The α-angle (ascending limb inclination; threshold $> 50°$) and β-angle (descending limb from r′-peak to J-point; threshold $> 58°$) are the primary quantitative discriminators, with sensitivities 71/79% and specificities 79/83% respectively [7]. QRS duration mismatch between V1 and V6 encodes localised RVOT conduction delay.

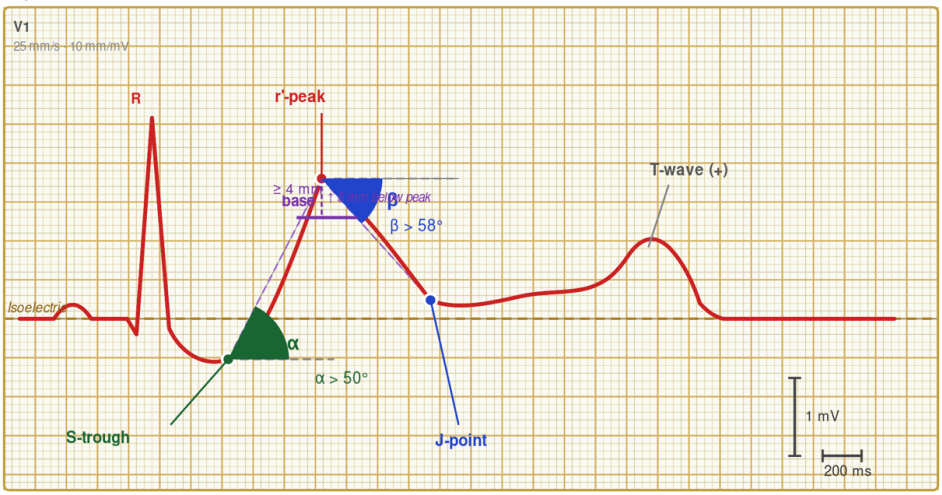

Figure 2.2: Annotated illustration of the r′-wave angular measurement framework for Type 2 Brugada Syndrome classification in lead V_1 (25 mm/s, 10 mm/mV). The α**-angle** (green arc, ascending limb from S-trough to r′-peak relative to the isoelectric horizontal; threshold $\alpha > 50°$) and β**-angle** (blue arc, descending limb from r′-peak to J-point relative to the horizontal; threshold $\beta > 58°$) are the primary quantitative discriminators between Type 2 BrS and IRBBB, with reported sensitivities of 71%/79% and specificities of 79%/83% respectively [7]. The purple double-headed arrow indicates the horizontal base of the r′ triangle at 5 mm below the r′-peak; a base $\geq$4 mm provides additional discriminatory power [7].

ST-segment and cross-lead markers. The Corrado index ($\mathrm{ST}(J)/\mathrm{ST}(J{+}80\,\mathrm{ms}) > 1$) distinguishes the downsloping BrS ST from the upsloping IRBBB pattern [11]. ST amplitudes at J, $J + 40$ ms, and $J + 80$ ms are extracted for V1 and V2. The S-wave in V6 is absent in BrS but prominent in RBBB, and emerges as the highest-ranked SHAP feature in this thesis. Fragmented QRS (fQRS) notching in V1–V3 is associated with prior VF in BrS [12]. The Tpeak-to-Tend interval and r-J interval in V1 are the top SHAP-ranked predictors in an independent Japanese cohort [9].

Inferior lead and aVR criteria. ST depression $\geq$0.1 mV for $\geq$80 ms in inferior leads (II, III, aVF) constitutes the Crea sign, present in 47% of confirmed Type 1 cases [8]. An aVR R-wave amplitude $\geq$0.3 mV or R/Q ratio $\geq$0.75 carries 98% specificity for elevated arrhythmic risk (Babai-Bigi criterion) [13].

2.4 Machine Learning for Cardiac ECG Classification

Automated ECG analysis has emerged as a major machine learning application domain. The architecture developed in this thesis draws on three complementary methodological traditions.

Convolutional neural networks. One-dimensional CNNs capture local morphological patterns (QRS shape, ST curvature, T-wave polarity) in a translation-equivariant manner. Ribeiro et al. demonstrated cardiologist-level arrhythmia classification from raw 12-lead ECGs using a residual Conv1d network on over 2 million records [14], establishing the viability of end-to-end ECG classification from raw signals. On the small datasets typical of rare cardiac conditions, CNNs alone exhibit limited generalisation, necessitating transfer learning and ensemble integration.

Transformer encoders. The self-attention mechanism introduced by Vaswani et al. [15] directly models pairwise relationships across all sequence positions, capturing long-range temporal dependencies (relationships between, for example, the r′-wave ascending limb and the post-J morphology) that may not be captured within convolutional receptive fields. The combination of a CNN front-end for local feature extraction and a Transformer encoder for global context integration has become an established paradigm in cardiac signal analysis [16].

Hand-engineered features and ensemble methods. When labelled examples are insufficient to constrain a deep network, scalar features derived from clinical knowledge provide more stable and interpretable discriminant boundaries [6]. Random Forest classifiers are particularly well-suited to such feature vectors: they are robust to outliers and missing values, insensitive to feature scaling, and capable of capturing non-linear feature interactions. Combining DL and hand-engineered features in an ensemble exploits qualitatively different signal sources: the DL stream discovers subtle morphological regularities from raw waveforms, while the RF stream operates on clinically-motivated scalar measurements that are less sensitive to distributional shift from digitisation artefacts.

Transfer learning. Pre-training on the PTB-XL database [17] (21,837 clinical 12-lead ECGs annotated with RBBB and normal sinus rhythm categories) allows the CNN–Transformer encoder to learn morphological representations optimised for right precordial conduction patterns before fine-tuning on the target BrS task. Saleh et al. independently confirmed that strategies leveraging pre-trained representations substantially outperform models trained from scratch on small BrS cohorts [6].

2.5 Prior Work on Automated Brugada Syndrome Classification

Melo et al. [5] trained a convolutional network on over 15,000 12-lead ECGs and demonstrated that DL can identify BrS-associated ECG patterns with high sensitivity when sufficient labelled data is available; however, their system was developed and validated primarily on Type 1 coved patterns. Randazzo et al. [16] proposed a three-class CNN + Transformer + attention architecture distinguishing Type 1 BrS, Type 2 BrS, and non-BrS ECGs; this architecture provides the primary structural template for BrugadaTwoStream. Ishida et al. [9] used SHAP to rank predictors of arrhythmic events in 234 Japanese BrS patients, identifying the r-J interval, Tpeak-to-Tend interval, and QTc as the three highest-ranked features, all incorporated into the 44-feature framework of this thesis. Ohkubo et al. [4] confirmed that no single angular or morphological criterion achieves sufficient sensitivity and specificity in isolation for Type 2 vs IRBBB discrimination, motivating the multi-criterion feature engineering strategy adopted here. A systematic review by Sahebnasagh and Farjoo [3] identified three recurring limitations across existing AI-BrS studies: concentration on Type 1 patterns, reliance on large single-centre archives, and absence of end-to-end reproducible pipelines, precisely the gaps that the present thesis addresses.

2.6 Explainability in ECG Classification: SHAP

SHAP (SHapley Additive exPlanations) [18] attributes model output to individual input features via Shapley values from cooperative game theory, measuring each feature's marginal contribution averaged over all coalitions. For the RF component, exact values are computed via TreeSHAP in polynomial time. In this thesis SHAP validates that the model's dominant features correspond to clinically established BrS discriminators and provides interpretable per-prediction attribution suitable for clinical reporting.

2.7 Research Gap and Thesis Contribution

The review presented in this chapter identifies a well-defined gap in the existing literature. No published end-to-end pipeline addresses binary Type 2 BrS versus Non-BrS classification under data-constrained conditions while simultaneously: (i) constructing its own labelled dataset from paper ECG digitisation and XML signal extraction; (ii) employing a literature-grounded 44-feature clinical engineering framework; (iii) incorporating PTB-XL transfer learning to initialise the ECG

morphology encoder; and (iv) applying a val-optimised DL + RF ensemble evaluated at the recording level.

Figure 2.3 summarises the positioning of the present thesis relative to the most directly comparable prior works.

STUDY	YEAR	TASK	DATASET (N)	METHOD	KEY METRIC†
Melo et al.	2023	Binary Type 1 focus	>15,000 ECGs	CNN	High sens./spec. (Type 1)
Randazzo et al.	2025	3-class T1 / T2 / Non	Small	CNN + TF + Attn	3-class accuracy
Dari	2024	Binary T2 vs Non-BrS	~100s	LDA + RF	Competitive F1
Ishida et al.	2025	Risk strat. Arrhythmia risk	234 patients	GB + SHAP	AUC (risk)
Saleh et al.	2026	Binary BrS vs Non-BrS	Limited	Multiple ML	AUC ≈ 0.90
Present thesis	2026	Binary T2 vs Non-BrS	362 ECGs (35 test)	CNN + TF + RF Ens.	AUC 0.938‡

Figure 2.3: Comparison of representative prior works on automated Brugada Syndrome ECG classification and the present thesis. Task: B = binary (BrS vs Non-BrS); 3C = three-class (Type 1 / Type 2 / Non-BrS); RS = risk stratification; T2 = Type 2 / IRBBB discrimination. Method abbreviations: CNN = convolutional neural network; TF = Transformer; RF = Random Forest; GB = gradient boosting; LDA = linear discriminant analysis. † Metric is AUC unless otherwise noted; ‡ recording-level result reported in present thesis.

Chapter 3

Dataset Construction and ECG Digitization

3.1 Overview of the Data Collection Strategy

The dataset underpinning this thesis was assembled from two complementary sources. The first comprised paper ECGs from a longitudinal clinical database of Brugada Syndrome patients followed at a Sardinian cardiology centre, which required systematic digitization before computational analysis. The second comprised confirmed Type 2 Brugada Syndrome recordings in the native digital format of the GE Healthcare MAC2000 electrocardiograph, an XML-based binary encoding requiring a dedicated extraction pipeline. Both sources ultimately produce the same unified output: a multi-sheet Excel workbook with one sheet per ECG lead, each containing time (`x`, seconds) and amplitude (`y`, millivolts) columns. This unified representation enabled a single downstream preprocessing pipeline to handle both classes without format-specific branching.

3.2 The Sardinian Brugada Patient Database

The Non-BrS recordings originate from a longitudinal follow-up cohort evaluated for Brugada Syndrome at a Sardinian cardiology centre. The database encompasses patients across a spectrum of diagnostic outcomes: confirmed Type 1 BrS, confirmed or suspected Type 2 BrS, and patients in whom Brugada pattern was absent (Non-BrS). Ground-truth labels were derived from a clinical report file (`Report ECG Sardegna Brugada.xlsx`), cataloguing each patient's diagnosis based on clinical assessment, expert ECG interpretation, and, where applicable, pharmacological provocation testing. This expert-assigned labelling is treated as authoritative; no

re-labelling was performed.

Type 1 BrS recordings were excluded as they constitute a separate, better-studied classification problem. Type 2 BrS recordings from the paper archive were likewise excluded, as a higher-quality digitally-acquired Type 2 dataset was available separately. Approximately 1,000 paper ECGs were processed in total; following quality control and labelling cross-reference, 174 Non-BrS recordings were retained.

3.3 Paper ECG Digitization Using ECGD

Paper ECG digitization was performed using ECGD (ECG Digitizer, Roberto Bruno, Politecnico di Torino, 2023), a Java desktop application purpose-built for scanned or photographed paper ECG images. ECGD accepts PNG, JPEG, and PDF inputs and guides the operator through a sequential workflow: image preprocessing (rotation and scale calibration), per-lead cropping, grid removal, pixel-to-signal conversion, peak correction, and numerical export.

Each image undergoes two calibration steps prior to signal extraction. *Rotation correction* aligns the ECG grid to the horizontal axis with approximately $\pm 0.1^{\circ}$ precision to eliminate systematic waveform tilt. *Scale calibration* determines the pixel-to-millimetre ratio by the operator measuring 1 cm on the ECG grid; the operator also enters paper speed (25 mm/s) and signal gain (10 mm/mV), enabling conversion to physical units.

Each of the 12 leads is cropped individually, accommodating diverse layout formats. Grid removal uses RGB/HSV histogram thresholding or morphological line detection depending on whether the trace and grid are colour-differentiated. A column-scan algorithm converts pixel positions to physical units using the calibrated scale factor, paper speed (25 mm/s), and gain (10 mm/mV). A semi-automated peak correction step preserves J-point amplitude and r′-wave morphology.

The final output is a multi-sheet `.xlsx` file with columns `x` (time, seconds) and `y` (amplitude, millivolts) per lead. The effective sampling rate is approximately 588 Hz (irregular), regularised to 500 Hz during preprocessing.

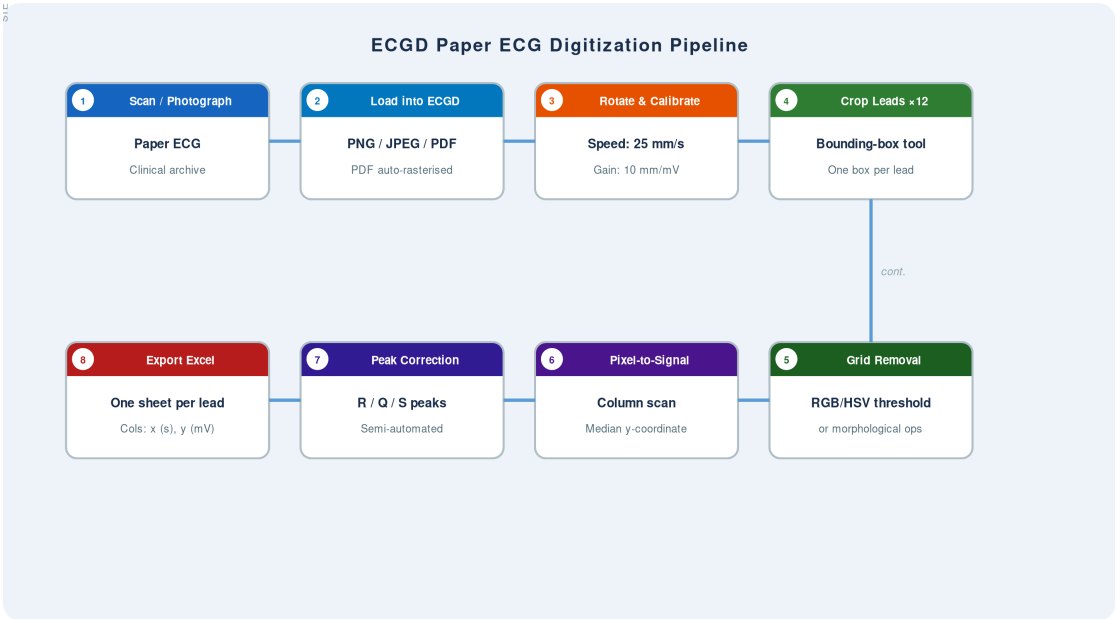

Figure 3.1: Schematic of the ECGD paper ECG digitization workflow. Eight sequential stages are shown in a serpentine left-to-right, right-to-left layout: (1) scan/photograph paper ECG; (2) load into ECGD; (3) rotate and scale calibrate; (4) crop individual leads; (5) grid removal; (6) pixel-to-signal conversion; (7) peak correction; (8) export multi-sheet Excel. The pipeline was applied to approximately 1,000 paper ECGs from the Sardinian Brugada patient database.

Table 3.1: Technical specifications of the Non-BrS digitized ECG recordings obtained via ECGD.

Parameter	**Value**
ECG machine	GE Healthcare MAC2000 (or equivalent clinical device)
Paper speed	25 mm/s
Signal gain	10 mm/mV
Leads digitized	12 (I, II, III, aVR, aVL, aVF, V1–V6)
Approximate sampling rate	∼588 Hz (irregular, derived from scan pixel density)
Signal units	Millivolts (mV)
Output file format	Multi-sheet `.xlsx` (one sheet per lead, columns `x`/`y`)
Quality control applied	Rotation correction, scale calibration, peak correction, manual artefact review
Recordings retained	174 (post-QC)

3.4 The Type 2 BrS Dataset: GE MAC2000 XML Format

The 188 Type 2 BrS recordings were acquired on the GE Healthcare MAC2000 electrocardiograph using GE Acquisition software v1.1 and stored in GE's proprietary DCAR XML format, schema version 1.2. Signal values are encoded as raw 12-bit ADC counts in the integer range $[-2048, +2047]$, corresponding to an analogue input voltage range of approximately ± 10 V. Conversion from ADC counts to millivolts is:

$$\text{mV} = \text{ADC\_count} \times \frac{4.88}{1000} \tag{3.1}$$

The DCAR schema organises data hierarchically: patient demographics under `dcarRecord/patientInfo`; signal configuration under `ecg/cfg` (sample rate 1,000 Hz, writer filter 150 Hz); and individual lead signals as integer sequences under `ecg/waveform/lead/sample`.

Table 3.2: Technical specifications of the Type 2 BrS XML recordings acquired on the GE Healthcare MAC2000 electrocardiograph.

Parameter	**Value**
Manufacturer	GE Healthcare
Device model	MAC2000
Acquisition software	GE Acquisition v1.1
Analysis / reporting software	GE 12SL™ v241
File format / schema	GE DCAR XML v1.2 (Sapphire, `urn:ge:sapphire:dcar_1`)
Native sample rate (rhythm strips)	1,000 Hz
Median-beat sample rate	500 Hz
Hardware writer filter	150 Hz
Paper speed	25 mm/s
Gain	10 mm/mV
ADC resolution	12-bit (range $[-2048, +2047]$)
Analogue input range	$\approx \pm 10$ V
ADC-to-mV conversion factor	$\times 4.88/1000$
Output leads	12 (I, II, III, aVR, aVL, aVF, V1–V6)
Recordings retained	188

A dedicated Python script (`dataset/extract_xml_signals.py`) batch-converts all XML files to the unified Excel format. A strict quality filter accepts only recordings containing a complete 10-second trace (10,000 samples); truncated recordings are skipped and logged.

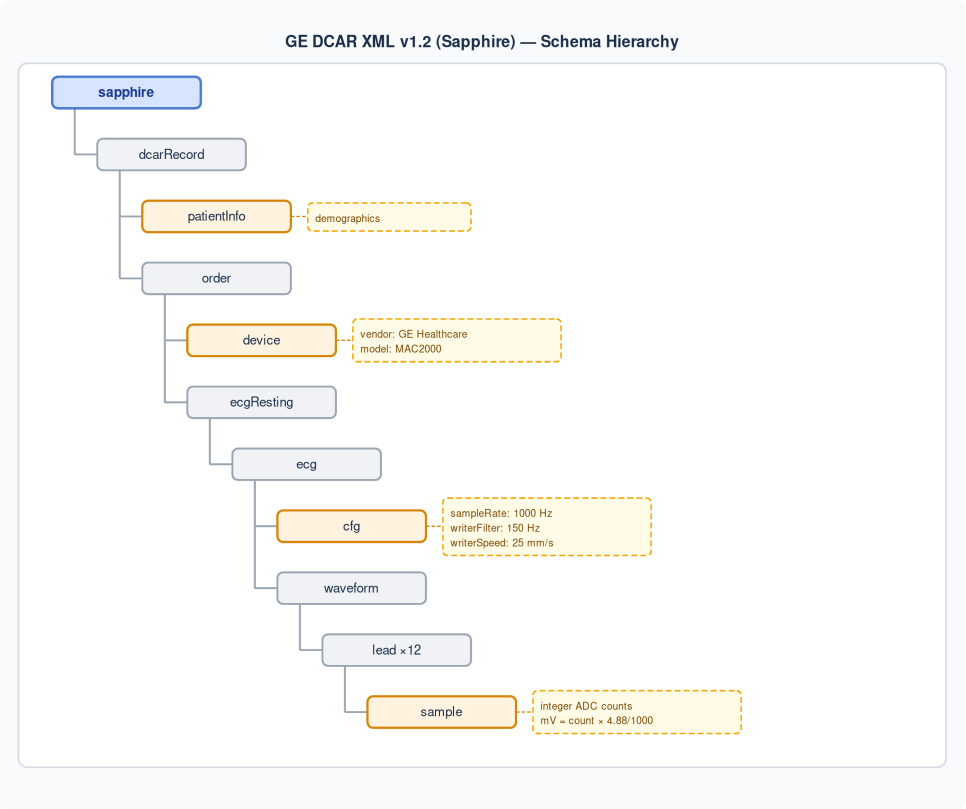

Figure 3.2: Simplified hierarchy of the GE DCAR XML v1.2 (Sapphire) schema used by the MAC2000 electrocardiograph. Amber-outlined nodes carry diagnostic annotations: `patientInfo` stores patient demographics; `device` records the hardware identity (GE Healthcare MAC2000); `cfg` stores the signal configuration (sampleRate 1000 Hz, writerFilter 150 Hz, writerSpeed 25 mm/s); and `sample` encodes raw 12-bit ADC integer counts, requiring the conversion mV = ADC_count × 4.88/1000 prior to use.

A critical quality-control finding arose during early development: the ADC-to-millivolt conversion factor was inadvertently omitted from one pipeline iteration, leaving Type 2 BrS values in raw ADC counts ($\sim 1000\times$ larger than millivolt-calibrated Non-BrS values). The model immediately achieved an apparent 98% accuracy by exploiting the gross amplitude discrepancy rather than any morphological features. The error was detected by computing class-conditional amplitude statistics. This episode underscores that high training accuracy alone is not evidence of meaningful learning, and explicit verification of physical units is a necessary quality-control step.

3.5 Dataset Labelling and Classification

Ground-truth labels were derived from two authoritative clinical sources. For the Non-BrS class, labels came from the Sardinian clinical report (`Report ECG Sardegna Brugada.xlsx`), assigned by consultant cardiologists on the basis of clinical presentation, expert ECG interpretation, and pharmacological provocation testing where indicated. Type 1 BrS recordings were excluded from this study, as were Type 2 BrS recordings from the paper archive (superseded by the higher-quality XML dataset). For the Type 2 BrS class, the 188 XML recordings carry confirmed diagnoses from the originating cardiology centre. The binary label schema is: positive class = Type 2 BrS; negative class = Non-BrS.

3.6 Dataset Cleaning, Standardisation, and Patient Manifest

Raw digitized Non-BrS files presented significant naming heterogeneity from multiple recording sessions. A cleaning pipeline standardised the archive: each file is matched to a patient identifier in the clinical report via substring matching, assigned its label, and renamed to a consistent format. An overwrite-prevention counter and full audit trail are maintained.

A patient manifest was constructed to prevent patient-level data leakage, ensuring no patient appears in more than one split. The manifest was generated by `src/generate_patient_manifest.py`, which applies format-specific regex rules to extract stable patient identifiers from the eight distinct naming conventions present in the XML archive (GEMAC, ECG-MUSE, record-part, pag/pg suffix, short alphanumeric code, and various full-name-with-date formats). The resulting `patient_manifest.csv` maps each filename to a patient identifier and class label, and forms the input to the splitting procedure.

3.7 Final Dataset Statistics and Characteristics

The final dataset comprised 362 labelled ECG recordings: 188 Type 2 BrS and 174 Non-BrS. The patient-level split used a Longest-Processing-Time (LPT) bin-packing algorithm to produce a patient-disjoint 80/10/10 partition. Both the validation and test sets contain exactly 35 recordings (17 Non-BrS, 18 Type 2 BrS). Table 3.3 summarises the dataset characteristics.

Table 3.3: Summary statistics of the final labelled ECG dataset used for model training and evaluation.

Characteristic	**Type 2 BrS (Definitive)**	**Non-BrS**
Total recordings	188	174
Training set	152	140
Validation set	18	17
Test set	18	17
Native sampling rate	1,000 Hz	∼588 Hz (irregular)
Target sampling rate	500 Hz (uniform, post-resampling)	
Recording duration	≈10 s	
Leads available	12 (I, II, III, aVR, aVL, aVF, V1–V6)	
Source format	GE DCAR XML v1.2	Multi-sheet `.xlsx` (ECGD)
Windows extracted (train)	1,389 total	
Windows extracted (val)	177 total	
Windows extracted (test)	171 total	

Type 2 BrS recordings yield a median of approximately six windows per recording, whilst Non-BrS recordings yield a median of approximately three, producing an approximate 2.4:1 window-level class imbalance in the training set. The recording-level split remains near-balanced, ensuring evaluation at the recording level is not confounded by this imbalance.

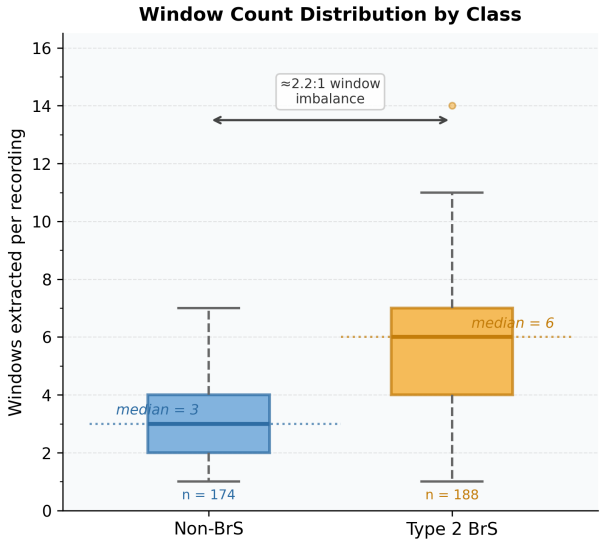

Figure 3.3: Distribution of extracted window counts per recording, stratified by class. Type 2 BrS recordings (median ≈6 windows) yield substantially more training windows per recording than Non-BrS recordings (median ≈3 windows), producing an approximate 2.4:1 window-level class imbalance in the training set. The recording-level split is near-balanced (18 vs 17 per split), ensuring that evaluation at the recording level is not confounded by this imbalance.

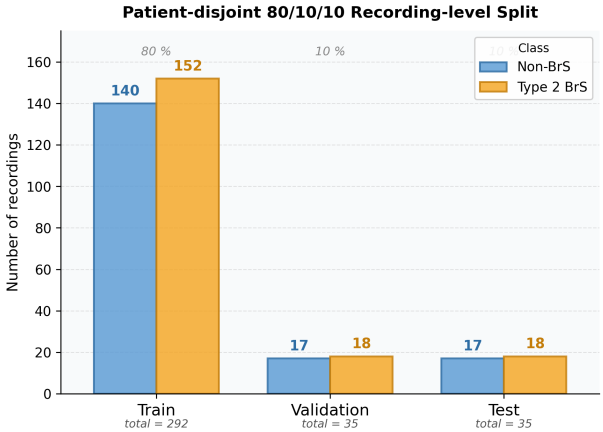

Figure 3.4: Recording counts per data split and class. The patient-disjoint 80/10/10 split allocates 292 recordings to training (Non-BrS: 140, Type 2 BrS: 152), 35 to validation (Non-BrS: 17, Type 2 BrS: 18), and 35 to the held-out test set (Non-BrS: 17, Type 2 BrS: 18). Near class balance is maintained across all three partitions at the recording level.

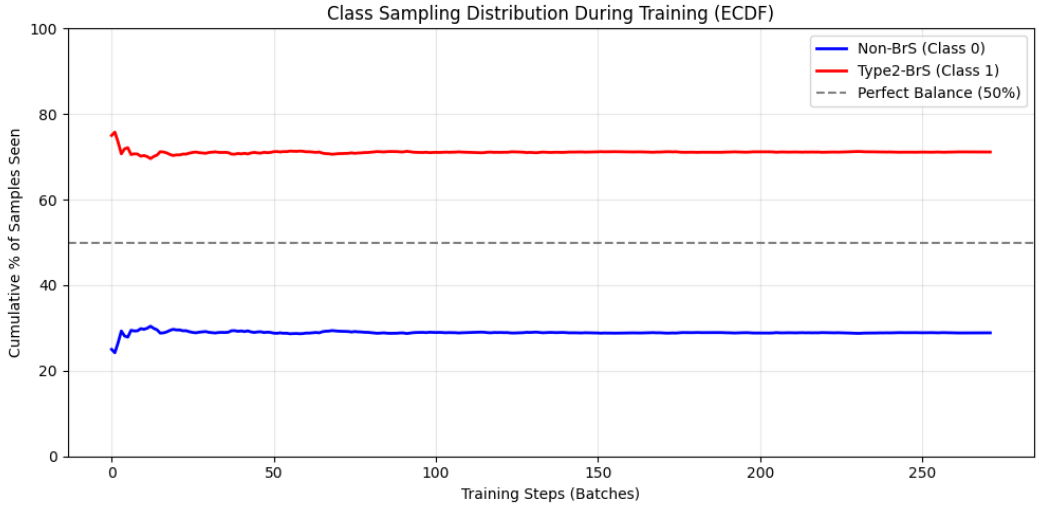

Figure 3.5: Empirical cumulative distribution function (ECDF) of the Type 2 BrS window proportion in training mini-batches across the full training run. The sharp cumulative rise centred near 0.57 confirms that mini-batch class composition is tightly concentrated around the dataset-level window ratio, with no systematic batch-level class imbalance. Training used random shuffling without weighted re-sampling.

Chapter 4

Signal Preprocessing

4.1 Preprocessing Pipeline Overview

Raw ECG recordings from the two source modalities exhibit substantial heterogeneity: Non-BrS digitised recordings carry an irregular sampling grid at approximately 588 Hz, whilst Type 2 BrS XML files store rhythm strips at 1,000 Hz. The preprocessing pipeline transforms each source recording into a standardised multi-channel tensor through four stages (Figure 4.1): resampling to 500 Hz, R-peak and S-trough detection in the V2 reference lead, 850 ms S-centred window extraction across four target leads, and first-derivative channel construction with mean subtraction to assemble the final $(8, 425)$ tensor. The pipeline is identical for both classes, except that Type 2 BrS ADC counts are converted to millivolts ($\text{mV} = \text{ADC} \times 4.88/1{,}000$) before the shared pipeline begins.

Figure 4.1: Vertical flowchart of the four-stage signal preprocessing pipeline. The ADC conversion step (dashed box) is applied exclusively to Type 2 BrS recordings prior to the shared pipeline. Event detection operates on the V2 reference lead; the resulting S-peak indices are applied uniformly across all four extracted leads (V1, V2, V3, V6). The output tensor shape $(8, 425)$ reflects eight channels (four raw and four derivative leads) at 425 samples per 850 ms window at 500 Hz.

4.2 Resampling to a Common Rate

Both source classes are resampled to 500 Hz using cubic spline interpolation [19], which fits a piecewise cubic spline through the original sample points and evaluates it at uniform 500 Hz positions. Non-BrS recordings are upsampled from their irregular ≈588 Hz digitised grid; Type 2 BrS recordings are downsampled by a factor of two from 1,000 Hz. Using the same interpolation method for both modalities keeps the signal transformations applied to the two classes as symmetric as possible, avoiding class-discriminating processing artefacts. The GE hardware writer filter at 150 Hz provides effective anti-aliasing above the 250 Hz Nyquist limit of the 500 Hz output, so no additional low-pass step is required before downsampling. Each lead is cleaned with the NeuroKit2 bandpass filter (Butterworth 0.5–40 Hz) [20] after resampling to suppress baseline wander.

During development, polyphase FIR resampling was evaluated as an alternative resampling method. Welch PSD plots (Figures 4.2 and 4.3) confirmed that the dominant ECG content lies below 100 Hz for both classes, well within the 250 Hz Nyquist boundary, and that polyphase FIR resampling correctly suppressed energy above 250 Hz. Despite this spectral correctness, switching to polyphase FIR resampling produced a statistically meaningful validation accuracy regression: the cubic interpolation artefacts at steep QRS transitions had become part of the learned signal representation in the PTB-XL pretrained encoder. Cubic interpolation was therefore retained as the production method, illustrating a general principle in transfer learning workflows: downstream signal processing must remain consistent with the representation expected by the pretrained encoder [6].

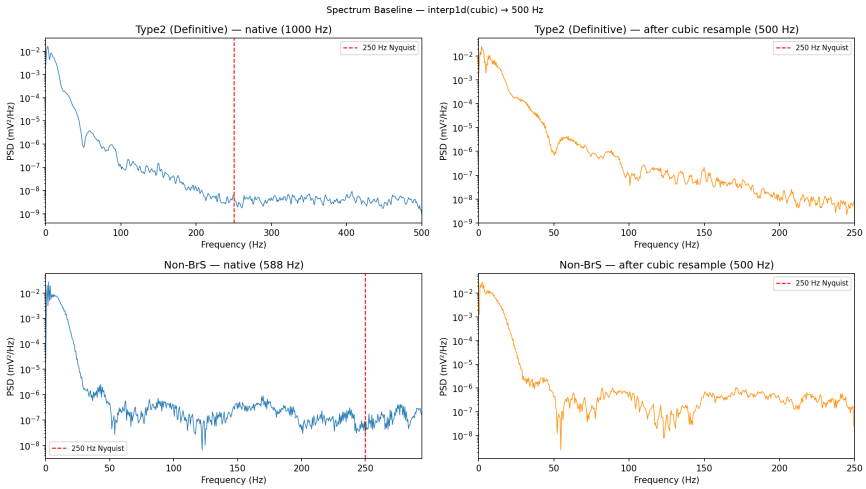

Figure 4.2: Welch power spectral density (PSD) of representative V2 lead signals at their native sampling rates. Type 2 BrS signals (1,000 Hz native) and Non-BrS digitized signals (≈588 Hz native) both exhibit dominant ECG content below 100 Hz, well within the 250 Hz Nyquist boundary imposed by the 500 Hz target rate. The vertical dashed line marks 250 Hz.

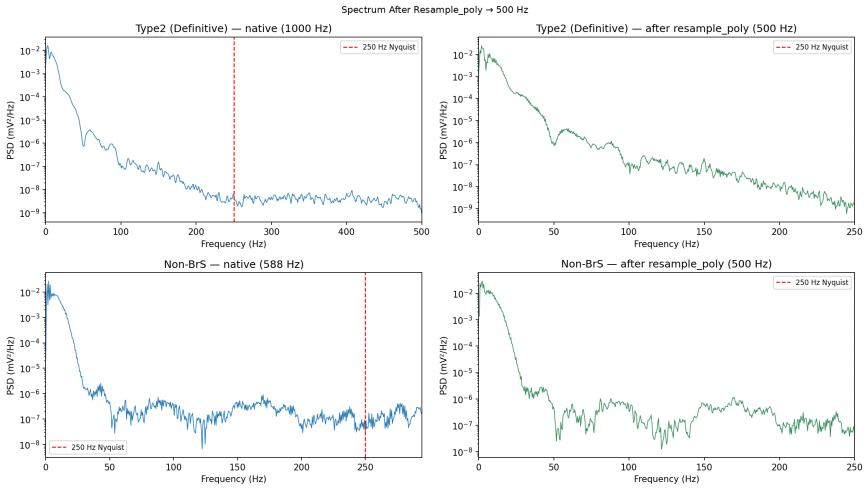

Figure 4.3: Welch PSD after `resample_poly` polyphase FIR anti-aliased resampling to 500 Hz. The transition band correctly attenuates energy above 250 Hz, confirming that the `resample_poly` approach satisfies the Nyquist criterion. Despite this spectral correctness, switching to `resample_poly` produced a statistically meaningful validation accuracy regression; the cubic interpolation baseline was therefore retained for the production pipeline.

4.3 Cardiac Event Detection

4.3.1 R-Peak Detection

R-peaks are detected using NeuroKit2's Pan–Tompkins DWT detector on the V2 reference lead [20]; if fewer than 3 peaks are returned (a risk with the secondary r′-wave in Type 2 BrS morphology), a peak-finding fallback is applied and the result with more peaks is retained [19]. The median R–R interval is stored for QTc computation.

4.3.2 S-Peak Localisation

S-troughs are localised via NeuroKit2 DWT delineation [20]; when DWT fails, the S-trough is estimated as the signal minimum within 100 ms post-R, robust to the shallow S-troughs of Type 2 saddleback morphology [7]. The S-peak index serves as the temporal anchor for all window extraction (Figure 4.4).

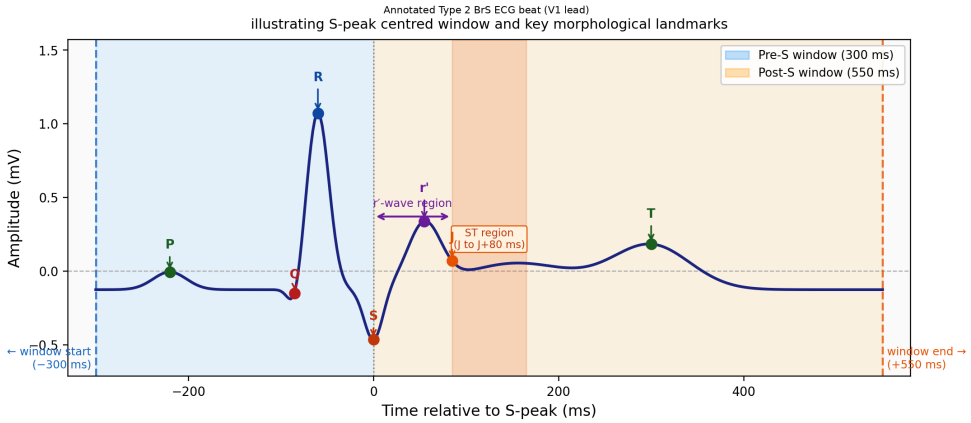

Figure 4.4: Annotated single-beat ECG trace (V1 lead, synthetic Type 2 BrS saddleback morphology) illustrating the fiducial landmarks used in the preprocessing pipeline. The S-trough (t = 0 ms) is the window anchor. The blue shaded region (300 ms pre-S, left) captures the QRS complex; the amber shaded region (550 ms post-S, right) captures the r′-wave, J-point, ST segment, and T-wave. The orange overlay marks the ST sampling region (J to J+80 ms) used by all morphological feature blocks. Centring on the S-peak ensures that the diagnostically critical r′-wave and saddleback ST profile fall reliably within the extraction window [7].

4.4 Window Extraction Strategy

Each 850 ms window comprises 300 ms pre-S and 550 ms post-S (425 samples at 500 Hz). S-peak centring ensures the pathological features of Type 2 BrS (r′-wave, J-point elevation, and saddleback ST) fall reliably in the window, with 550 ms post-S accommodating the full ST–T morphology [7, 4].

Windows are stored with an additional ±15-sample buffer (455 samples per channel) for augmentation. Training windows receive a random 425-sample crop (temporal jitter); validation and test windows receive a deterministic centre crop (`x[:, 15:440]`). All leads receive the same shift to preserve inter-lead alignment. No amplitude modification is applied during augmentation, as ST-elevation absolute voltage is the primary Brugada discriminant and altering it would confound both the DL stream and the hand-engineered ST features [16].

Non-BrS digitised recordings typically yield 1–4 windows per recording; Type 2 BrS XML recordings typically yield 4–14, depending on heart rate. This difference is the primary source of the ≈2.4:1 window-level class imbalance in the training set.

4.5 Multi-Lead Tensor Construction

4.5.1 Lead Selection: V1, V2, V3, V6

Four leads are selected: V1, V2, and V3 (the primary BrS diagnostic leads where the r′-wave, J-point elevation, and saddleback ST morphology are observed [1, 7]) and V6 (the lateral reference encoding S-wave presence, absent in Type 2 BrS but prominent in RBBB/IRBBB [7, 4]). The inferior leads (II, III, aVF) and aVR are not included in the raw tensor: including them would expand the encoder's input channel dimension beyond the fixed architecture of the PTB-XL pretrained encoder. Their discriminative information (Crea sign, Babai-Bigi aVR criterion) is instead captured as hand-engineered scalar features in Stream B.

4.5.2 First-Derivative Channels

For each raw lead, a first-derivative channel is computed using a second-order accurate central difference [19]:

$$dV_i[n] = \frac{V_i[n+1] - V_i[n-1]}{2}, \quad 1 \leq n \leq N-2. \tag{4.1}$$

The derivative amplifies local slope information that directly encodes the α and β angle criteria of de Luna et al. [7]. Each derivative channel is normalised to unit

standard deviation:

$$d\hat{V}_i = \frac{dV_i}{\sigma(dV_i) + \epsilon}, \tag{4.2}$$

where $\epsilon = 10^{-8}$ prevents division by zero on flat segments. Raw channels retain their absolute millivolt scale; only derivative channels receive standard-deviation normalisation.

4.5.3 Mean Subtraction and Amplitude Preservation

Each raw channel is mean-subtracted:

$$\tilde{V}_i = V_i - \bar{V}_i, \quad \bar{V}_i = \frac{1}{N}\sum_{n=0}^{N-1} V_i[n]. \tag{4.3}$$

No further per-channel amplitude normalisation is applied. The absolute millivolt scale is preserved because several diagnostic thresholds are defined in absolute voltage: the 0.1 mV Crea sign threshold [8], the $\geq$0.2 mV r′-wave amplitude criterion [7], and the $\geq$1 mm ST elevation criterion [1]. Dividing raw channels by standard deviation would destroy these relationships and prevent the CNN from using amplitude magnitude as a discriminative cue.

The final tensor has shape $(\mathbf{8}, \mathbf{425})$, with channel ordering $[\mathrm{V1}, \mathrm{V2}, \mathrm{V3}, \mathrm{V6}, d\mathrm{V1}, d\mathrm{V2}, d\mathrm{V3}, d\mathrm{V6}]$. Table 4.1 summarises the channel specification.

Table 4.1: Multi-channel input tensor specification. All 8 channels are extracted from the same 850 ms S-peak-centred window. ADC-to-mV conversion is applied to Type 2 BrS recordings only, prior to entering the shared pipeline. The delivered tensor shape after augmentation crop is $(8, 425)$.

Ch.	Lead	Type	Normalisation	Notes
0	V1	Raw signal	Mean subtraction only	Primary BrS diagnostic lead; r′-wave, ST morphology
1	V2	Raw signal	Mean subtraction only	Primary BrS diagnostic lead; R-peak detection reference
2	V3	Raw signal	Mean subtraction only	Transitional precordial; right-to-mid QRS morphology
3	V6	Raw signal	Mean subtraction only	Lateral reference; S-wave presence/absence criterion
4	dV1	1st derivative	Unit-std + mean sub.	r′ ascending/descending slope (α/β angles)
5	dV2	1st derivative	Unit-std + mean sub.	ST descent slope and QRS boundary features
6	dV3	1st derivative	Unit-std + mean sub.	Slope-amplified V3
7	dV6	1st derivative	Unit-std + mean sub.	S-wave slope in V6

ADC conversion (Type 2 BrS only): raw count × 4.88/1,000 before pipeline entry.
Stored window: 455 samples/channel (425 + 15-sample buffer each side) for augmentation.

4.6 Preprocessing Validation and Signal Quality

4.6.1 Data Completeness

A file may fail to contribute windows due to missing or malformed lead sheets, signal too short for windowing, or failed R-peak detection. Each such case emits a structured warning and contributes zero windows without halting the pipeline. In the production split, 4 training files were skipped (≈1.4% of 292 training recordings), a rate well within the acceptable range for clinical ECG datasets [6].

4.6.2 ADC Conversion Validation

Correct ADC-to-millivolt conversion (mV = ADC × 4.88/1,000) is critical: an inadvertent omission during development caused raw counts (±2,048 units) to differ from calibrated Non-BrS signals (≈ ±2 mV) by a factor of ≈1,000, making classes trivially separable on amplitude alone rather than morphology. This underscores that end-to-end validation of per-class amplitude distributions is an essential quality-control step in any multi-source ECG pipeline [6].

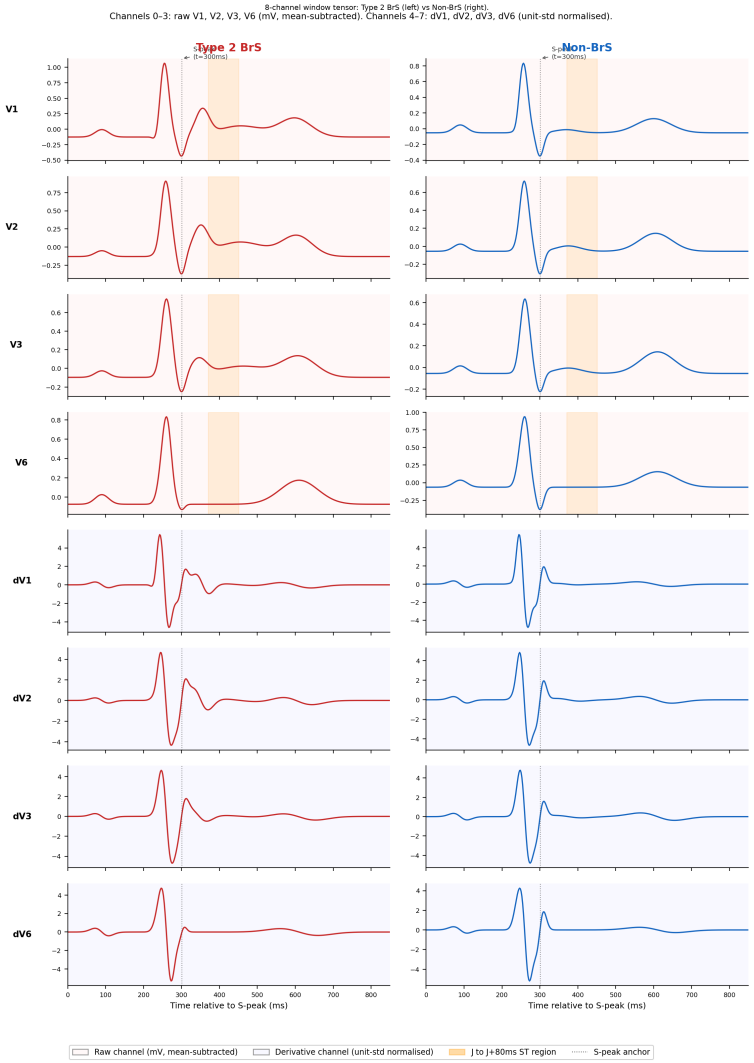

Figure 4.5: Side-by-side comparison of the 8-channel input tensor for a representative Type 2 BrS recording (left, red) and a Non-BrS recording (right, blue). Channels 0–3 are the raw V1, V2, V3, V6 signals (mV, mean-subtracted); channels 4–7 are the corresponding first-derivative channels (unit-std normalised). The vertical dotted line marks the S-peak anchor at $t = 300$ ms; the orange overlay marks the J to J+80 ms ST sampling region (raw channels only). Visible morphological differences include: the prominent r′-wave secondary deflection in V1/V2 of the Type 2 BrS recording (absent in Non-BrS); the elevated saddleback ST segment in the Type 2 post-J region; and the shallower S-wave in V6 of the Type 2 recording compared to the deeper S-wave of the Non-BrS recording. These differences are amplified in the derivative channels dV1 and dV2, where the r′-wave upstroke and descent produce distinct slope signatures.

Chapter 5

Clinical Feature Engineering

5.1 Motivation for Hand-Engineered Features

The dual-stream architecture assigns distinct responsibilities to its two input pathways. Stream A processes raw 8-channel ECG tensors through a convolutional and Transformer-based encoder; Stream B operates in parallel on a fixed-length vector of 44 hand-engineered morphological features passed through a compact MLP before fusion.

Three complementary considerations motivate Stream B. First, the present dataset of 362 recordings yields approximately 1,389 training windows, far below the data volumes required for end-to-end deep networks to discover BrS morphological patterns without overfitting [14, 6]. Second, the 44 features operationalise morphological criteria validated over three decades of BrS research (the r′-wave angular measurements, the Corrado index, inferior ST depression, and the aVR sign), encoding clinically meaningful signal components regardless of whether Stream A has independently learned to attend to them [7, 11, 8, 13]. Third, because SHAP operates on the 44-dimensional feature vector, every prediction can be accompanied by a feature-level explanation grounded in established ECG terminology, a key requirement for clinical decision-support deployment [18, 3].

5.2 J-point Detection Methodology

J-point detection operates in two stages. First, an r′-peak detector searches up to 120 ms after the S-trough for the highest local maximum exceeding the S-trough by at least 0.02 mV; for windows without a detectable r′-wave (27.4% V1, 31.6% V2), the J-point falls back to S-peak +80 ms [4]. Second, a J-point detector searches within 80 ms after the r′-peak for the first local trough (argmin fallback otherwise), guaranteeing a 0% NaN rate for all J-point-dependent features. The V1 J-point

index is reused as the temporal reference for inferior and aVR lead features [1].

5.3 Feature Blocks 1 and 2 — V1 and V2 r′-wave Morphology (Features 0–17)

Blocks 1 and 2 each comprise nine morphological measurements applied to V1 (features 0–8) and V2 (features 9–17) respectively. They encode the primary morphological signature of Type 2 BrS in the right precordial leads [1, 10].

r′-peak detection. The r′-peak detector searches up to 120 ms after the S-trough for local maxima exceeding 0.02 mV above the S-level (NaN rates: 27.4% V1, 31.6% V2).

β**-angle** (descending limb) and α**-angle** (ascending limb) are computed in the conventional ECG paper coordinate system (10 mm/mV, 25 mm/s):

$$\beta = \arctan\left(\frac{\Delta V_{\text{mV}} \times 10}{\Delta t_{\text{s}} \times 25}\right), \qquad \alpha = \arctan\left(\frac{\Delta V_{\text{mV}} \times 10}{\Delta t_{\text{s}} \times 25}\right) \quad \text{[degrees]} \tag{5.1}$$

where for β: $\Delta V = V_{\text{r}'} - V_J$ (r′-peak to J-point drop), and for α: $\Delta V = V_{\text{r}'} - V_S$ (S-trough to r′-peak rise). Thresholds $\beta > 58°$ and $\alpha > 50°$ are indicative of BrS [7]. Both carry the same NaN rate as r′ detection. The **r′-wave amplitude** is the signed voltage at the r′-peak index: $A_{\text{r}'} = V[i_{\text{r}'}]$ (mV).

ST-segment sampling extracts three voltages at fixed post-J offsets:

$$\text{ST}_J = V[i_J], \quad \text{ST}_{J+40} = V[i_J + \lfloor 0.040 f_s \rfloor], \quad \text{ST}_{J+80} = V[i_J + \lfloor 0.080 f_s \rfloor] \tag{5.2}$$

NaN rate: 0% (J-point fallback always succeeds). The **Corrado index** is the ratio $\text{CI} = \text{ST}_J / \text{ST}_{J+80} > 1$ (downsloping ST, consistent with BrS) [11]. The **ST descent slope** and **ST curvature sign** are:

$$\text{slope} = \frac{V[i_J + n_{40}] - V[i_J]}{0.040}\ [\text{mV s}^{-1}], \qquad \text{curv} = \text{sign}\Big(\overline{\nabla^2 V[i_J : i_J + n_{80}]}\Big) \tag{5.3}$$

Both carry 0% NaN rates and contribute through non-linear MLP interactions despite low marginal SHAP ranks.

5.4 Feature Block 3 — Cross-Lead and Global Markers (Features 18–28)

Block 3 comprises eleven features spanning multiple leads or requiring global cardiac timing. **QRS duration** in V1 and V2 is measured via an energy-threshold

approach on $|dV/dt|$:

$$\text{QRS} = \frac{i_{\text{last}} - i_{\text{first}}}{f_s} \times 1000 \quad [\text{ms}] \tag{5.4}$$

The **QRS mismatch** $\Delta\text{QRS} = \text{QRS}_{V1} - \text{QRS}_{V6}$ encodes localised RVOT conduction delay [4]. The **S-wave in V6** is a binary flag:

$$\text{s\_in\_v6} = \begin{cases} 1 & \text{if } \min(V_6) < -0.15\,\text{mV} \\ 0 & \text{otherwise} \end{cases} \tag{5.5}$$

The -0.15 mV threshold was selected by systematic sweep, yielding the maximum class separation of 2.002, the largest of all 44 features. SHAP ranked `s_in_v6` first overall (importance 0.607).

The remaining seven Block 3 features encode repolarisation timing and T-wave characteristics [9]: **T-wave polarity** (V1, V2); **r-J interval** $\text{rJ}_{V1} = (i_J - i_{\text{onset}})/f_s \times 1000$ ms (SHAP rank 7); **Tpeak-Tend interval** $\text{TpTe} = (i_{\text{Tend}} - i_{\text{Tpeak}})/f_s \times 1000$ ms (SHAP rank 5); **QTc Bazett** (≈5% NaN, rank 14); and **fragmented QRS count** in V1/V2 [12].

5.5 Feature Block 4 — Inferior Lead ST Depression and Crea Sign (Features 29–40)

Block 4 provides four measurements per inferior lead (II, III, aVF): ST at J, ST at J+80 ms, maximum ST depression, and a binary Crea-sign flag, operationalising the Crea 2015 criterion [8]. All features use the V1 J-point as the global temporal reference. NaN rate ≈ 6% from missing lead sheets.

Maximum ST depression uses a 160 ms post-J sliding window:

$$\text{max\_dep} = \min_{k \in [j,\, j + \lfloor 0.160 f_s \rfloor - n_{80}]} \min(V_{\text{inf}}[k : k + n_{80}]) \tag{5.6}$$

`max_dep_ii` ranked 2nd overall (0.565) and `max_dep_avf` ranked 3rd (0.431) in SHAP importance, confirming the Crea sign as a powerful discriminator [8, 2].

Binary Crea flag:

$$\text{crea} = \begin{cases} 1 & \text{if } \exists\, k : V_{\text{inf}}[k : k + n_{80}] \le -0.10\,\text{mV (all samples)} \\ 0 & \text{otherwise} \end{cases} \tag{5.7}$$

Despite its strong clinical motivation, the binary Crea flags ranked 38th–41st in SHAP importance; the continuous `max_dep` captures the same physiology with finer granularity and substantially higher discriminative power. Both are retained: SHAP-pruning experiments confirmed that removing any features reduced classification performance.

5.6 Feature Block 5 — aVR Sign Features (Features 41–43)

Block 5 contains three features encoding the Babai-Bigi aVR sign criterion [13]. The R-peak and Q-trough indices in aVR are:

$$i_R = \arg\max(V_{\mathrm{aVR}}[i_{\mathrm{onset}} : i_J + 1]), \qquad i_Q = \arg\min(V_{\mathrm{aVR}}[i_{\mathrm{onset}} : i_J + 1]) \quad (5.8)$$

R-wave amplitude: $A_R = V_{\mathrm{aVR}}[i_R]$; threshold $\geq 0.3\,\mathrm{mV}$ = elevated SCD risk [13]; SHAP rank 22nd (0.169). **R/Q ratio**: $\mathrm{RQ} = A_R/(|A_Q| + \epsilon)$; threshold ≥ 0.75 = 98% specificity for SCD risk [13]; ranked last (44th, 0.006) in SHAP. **ST elevation at J in aVR**: $V_{\mathrm{aVR}}[i_J]$; SHAP rank 40th (0.064). NaN rate $\approx 6\%$ from missing lead sheets. All three are retained consistent with the no-pruning policy.

Table 5.1: Complete specification of the 44 hand-engineered features in Stream B. *Block*: 1 = V1 morphology; 2 = V2 morphology; 3 = cross-lead/global; 4 = inferior lead (Crea sign); 5 = aVR sign (Babai-Bigi). NaN rates are approximate training-set values. SHAP rank is derived from mean absolute SHAP importance on the production RF model.

Idx	Name	Block	Lead	Clinical source / formula	NaN %	SHAP rank
Block 1 — V1 morphology (features 0–8)						
0	`beta_v1`	1 – V1 morph	V1	$\beta = \arctan\left(\frac{\Delta V \cdot 10}{\Delta t \cdot 25}\right)$; > 58ř BrS indicator [7]	27.4	15
1	`alpha_v1`	1 – V1 morph	V1	$\alpha = \arctan\left(\frac{\Delta V \cdot 10}{\Delta t \cdot 25}\right)$; > 50ř auxiliary BrS criterion [7]	27.4	34
2	`rp_amp_v1`	1 – V1 morph	V1	r′-peak amplitude $V[i_{r'}]$ (mV); positive deflection in Type 2 [1]	27.4	8
3	`st_j_v1`	1 – V1 morph	V1	ST voltage at J-point: $V[i_J]$ (mV) [11]	0.0	36
4	`st_j40_v1`	1 – V1 morph	V1	ST voltage at J+40 ms: $V[i_J + n_{40}]$ (mV) [11]	0.0	12
5	`st_j80_v1`	1 – V1 morph	V1	ST voltage at J+80 ms: $V[i_J + n_{80}]$ (mV) [11]	0.0	33
6	`corrado_v1`	1 – V1 morph	V1	Corrado index: ST(J)/ST(J+80 ms); >1 = descending ST [11]	0.0	42
7	`slope_v1`	1 – V1 morph	V1	ST descent slope over first 40 ms post-J ($\mathrm{mV\,s^{-1}}$) [4]	0.0	37

Continued on next page

Table 5.1 continued

Idx	Name	Block	Lead	Clinical source / formula	NaN %	SHAP rank
8	`curv_v1`	1 – V1 morph	V1	ST curvature sign: $\text{sign}(\overline{\nabla^2 V})$ over J to J+80 ms; +1 = saddle-back [1]	0.0	19
Block 2 — V2 morphology (features 9–17)						
9	`beta_v2`	2 – V2 morph	V2	Same as `beta_v1` on V2; > 58ř [7]	31.6	9
10	`alpha_v2`	2 – V2 morph	V2	Same as `alpha_v1` on V2; > 50ř [7]	31.6	21
11	`rp_amp_v2`	2 – V2 morph	V2	r′-peak amplitude on V2 (mV) [1]	31.6	28
12	`st_j_v2`	2 – V2 morph	V2	$V[i_J]$ on V2 (mV); SHAP rank 4 [11]	0.0	4
13	`st_j40_v2`	2 – V2 morph	V2	$V[i_J + n_{40}]$ on V2 (mV) [11]	0.0	11
14	`st_j80_v2`	2 – V2 morph	V2	$V[i_J + n_{80}]$ on V2 (mV) [11]	0.0	6
15	`corrado_v2`	2 – V2 morph	V2	Corrado index on V2 [11]	0.0	43
16	`slope_v2`	2 – V2 morph	V2	ST descent slope on V2 ($\text{mV}\,\text{s}^{-1}$) [4]	0.0	17
17	`curv_v2`	2 – V2 morph	V2	ST curvature sign on V2 [1]	0.0	27
Block 3 — Cross-lead and global markers (features 18–28)						
18	`qrs_dur_v1`	3 – Cross	V1	QRS duration via $\lvert dV/dt \rvert$ energy threshold (ms) [1]	0.0	31
19	`qrs_dur_v2`	3 – Cross	V2	QRS duration on V2 (ms) [1]	0.0	25
20	`qrs_v1v6_diff`	3 – Cross	V1/V6	$\text{QRS}_{V1} - \text{QRS}_{V6}$ (ms); right-sided conduction delay [4]	0.0	30
21	`s_in_v6`	3 – Cross	V6	Binary: 1 if $\min(V_6) < -0.15\,\text{mV}$ (S-wave present); threshold by systematic sweep	0.0	**1**
22	`t_pol_v1`	3 – Cross	V1	T-wave polarity: signed peak amplitude in T-wave region (mV) [1]	0.0	24
23	`t_pol_v2`	3 – Cross	V2	T-wave polarity on V2 (mV) [1]	0.0	32
24	`rj_v1`	3 – Cross	V1	r-J interval: QRS onset to J-point in V1 (ms); Ishida 2025 top feature [9]	0.0	7

Continued on next page

Table 5.1 continued

Idx	Name	Block	Lead	Clinical source / formula	NaN %	SHAP rank
25	`tpeak_tend`	3 – Cross	V1	Tpeak-Tend interval (ms); repolarisation dispersion marker [9, 2]	0.0	5
26	`qtc_bazett`	3 – Cross	V1	QTc = QT/$\sqrt{\text{RR[s]}}$ (ms); NaN for single-beat recordings [9]	≈5	14
27	`fqrs_v1`	3 – Cross	V1	Fragmented QRS count (prominence >0.05 mV) in QRS region [12]	0.0	23
28	`fqrs_v2`	3 – Cross	V2	Fragmented QRS count on V2 [12]	0.0	18
Block 4 — Inferior lead ST depression and Crea sign (features 29–40)						
29	`st_j_ii`	4 – Inferior	II	ST at J in lead II (mV); J-point from V1 reference [8]	≈6	35
30	`st_j80_ii`	4 – Inferior	II	ST at J+80 ms in lead II (mV) [8]	≈6	10
31	`max_dep_ii`	4 – Inferior	II	Max ST depression in 160 ms post-J sliding window, lead II (mV) [8]	≈6	**2**
32	`crea_ii`	4 – Inferior	II	Binary Crea flag: all samples ≤ -0.10 mV for ≥ 80 ms post-J [8]	≈6	38
33	`st_j_iii`	4 – Inferior	III	ST at J in lead III (mV) [8]	≈6	13
34	`st_j80_iii`	4 – Inferior	III	ST at J+80 ms in lead III (mV) [8]	≈6	16
35	`max_dep_iii`	4 – Inferior	III	Max ST depression, lead III (mV) [8]	≈6	29
36	`crea_iii`	4 – Inferior	III	Binary Crea flag, lead III [8]	≈6	39
37	`st_j_avf`	4 – Inferior	aVF	ST at J in lead aVF (mV) [8]	≈6	26
38	`st_j80_avf`	4 – Inferior	aVF	ST at J+80 ms in lead aVF (mV) [8]	≈6	20
39	`max_dep_avf`	4 – Inferior	aVF	Max ST depression, lead aVF (mV) [8]	≈6	**3**
40	`crea_avf`	4 – Inferior	aVF	Binary Crea flag, lead aVF [8]	≈6	41
Block 5 — aVR sign / Babai-Bigi (features 41–43)						
41	`avr_r_amp`	5 – aVR	aVR	R-peak amplitude in aVR (mV); ≥ 0.3 mV = elevated SCD risk [13]	≈6	22

Continued on next page

Table 5.1 continued

Idx	Name	Block	Lead	Clinical source / formula	NaN %	SHAP rank
42	`avr_rq_ratio`	5 – aVR	aVR	R/Q ratio = $A_R/(\lvert A_Q\rvert + \epsilon)$; $\geq 0.75 = 98\%$ specificity for SCD risk [13]	≈6	44
43	`avr_st_elev`	5 – aVR	aVR	ST elevation at J in aVR (mV) [13, 1]	≈6	40

5.7 Feature preprocessing: imputation and standardisation

Features with r′-dependent values carry NaN rates of 27.4% (V1) and 31.6% (V2); QTc is missing for ≈ 5% of single-beat recordings; inferior and aVR features carry ≈ 6% from missing lead sheets. Two preprocessing steps transform the raw feature matrix, both fit exclusively on training data to prevent leakage [21].

Median imputation. Missing values are replaced by the per-feature training-fold median, preferred over the mean due to heavy-tailed distributions in several features:

$$\tilde{x}_{ij} = \text{median}\Big(\{x_{kj} : k \in \mathcal{D}_{\text{train}},\ x_{kj} \neq \texttt{NaN}\}\Big). \tag{5.9}$$

Z-score standardisation. After imputation every feature is standardised to zero mean and unit variance:

$$\hat{x}_{ij} = \frac{\tilde{x}_{ij} - \mu_j}{\sigma_j} \tag{5.10}$$

ensuring features in heterogeneous units contribute equally to node-split decisions in the RF [21]. Both the 44 imputation medians and the 2×44 scaler parameters are serialised to `models/` for deterministic reproduction at inference.

5.8 Feature analysis and validation

An offline analysis quantified data completeness, univariate discriminative power, class-conditional distributions, inter-feature redundancy, and model-aware importance for the 44-feature set [3, 6].

5.8.1 Feature NaN Rate Analysis

Figure 5.1 shows NaN rates per feature. Two tiers are apparent: r′-dependent V1 and V2 features (27.4–31.6%) and all remaining features ($\leq 6\%$). After median imputation no window is discarded due to NaN status.

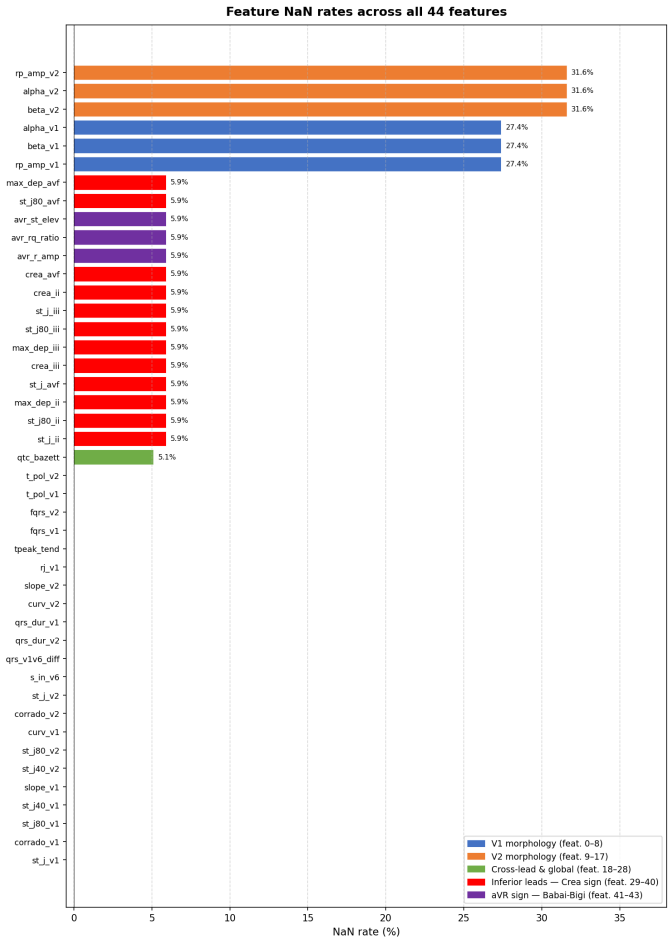

Figure 5.1: NaN rate (%) for each of the 44 hand-engineered features, sorted descending. Colour coding indicates feature block: V1 morphology (dark blue), V2 morphology (medium blue), cross-lead (green), inferior/Crea (orange), aVR/Babai-Bigi (red). r′-dependent V1 and V2 features account for the two highest NaN tiers (27.4 % and 31.6 % respectively); J-point-based features report 0 % missingness.

5.8.2 Class Separation Analysis

Class separation is quantified as the normalised absolute difference between class-conditional medians:

$$\Delta_j = \frac{|\tilde{m}_j^{(\mathrm{T2})} - \tilde{m}_j^{(\mathrm{NB})}|}{\tilde{\sigma}_j^{(\mathrm{T2})} + \tilde{\sigma}_j^{(\mathrm{NB})}} \tag{5.11}$$

`s_in_v6` achieves the highest score (2.002); inferior maximum depression and repolarisation features rank highly; Corrado indices and `avr_rq_ratio` show near-zero separation, as they were designed for within-BrS risk stratification rather than

binary BrS vs Non-BrS discrimination.

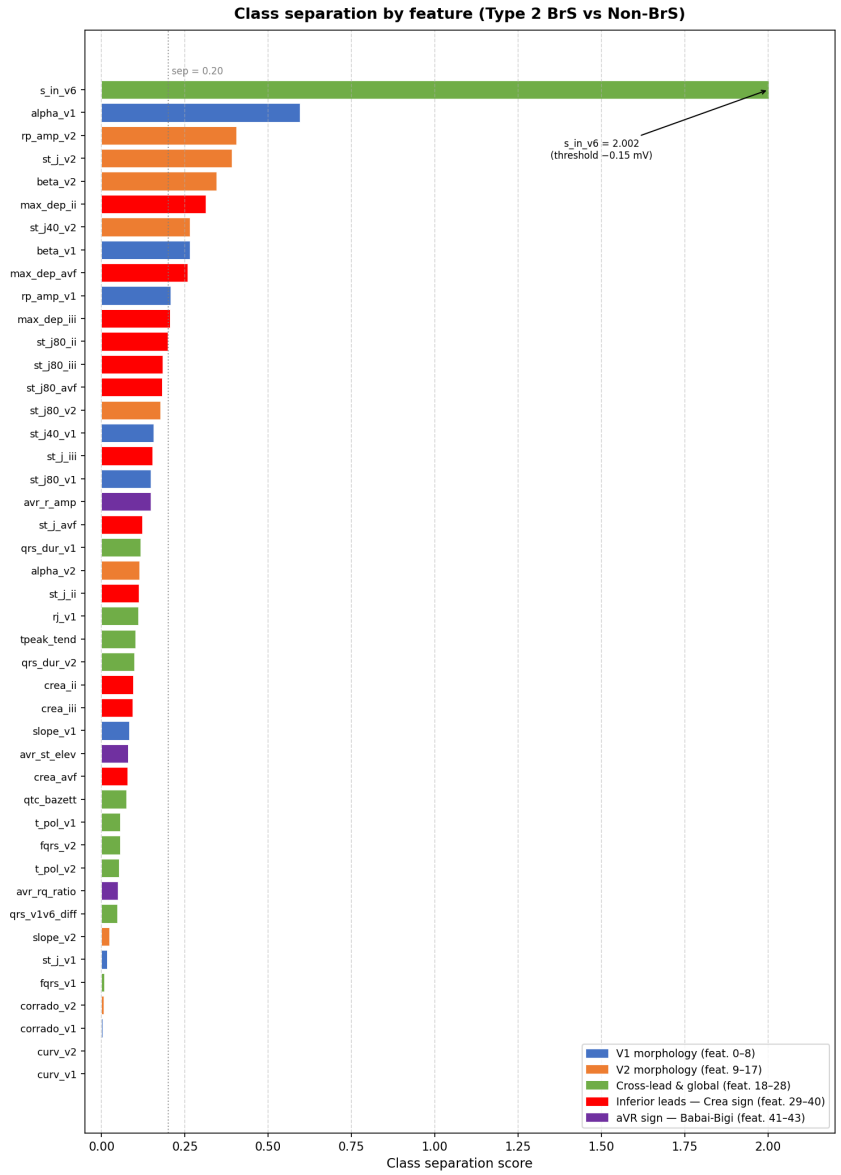

Figure 5.2: Class separation score (Equation 5.11) for all 44 features, sorted descending. Colour coding by block as in Figure 5.1. `s_in_v6` achieves the highest score (2.002), followed by inferior maximum depression and repolarisation interval features. Corrado indices and `avr_rq_ratio` show negligible separation in this binary task.

5.8.3 Feature Distribution Analysis

Figure 5.3 displays violin plots for the nine highest-SHAP features. The binary `s_in_v6` flag shows a strong shift towards 1 in the Type 2 class. The maximum ST depression features (`max_dep_ii`, `max_dep_avf`) show clear negative shifts in Type 2, consistent with the Crea sign [8]. The r-J interval is longer in Type 2 BrS, consistent with Ishida et al. [9]. ST voltage at J in V2 is markedly elevated in Type 2 BrS, consistent with the saddleback ST criterion [11].

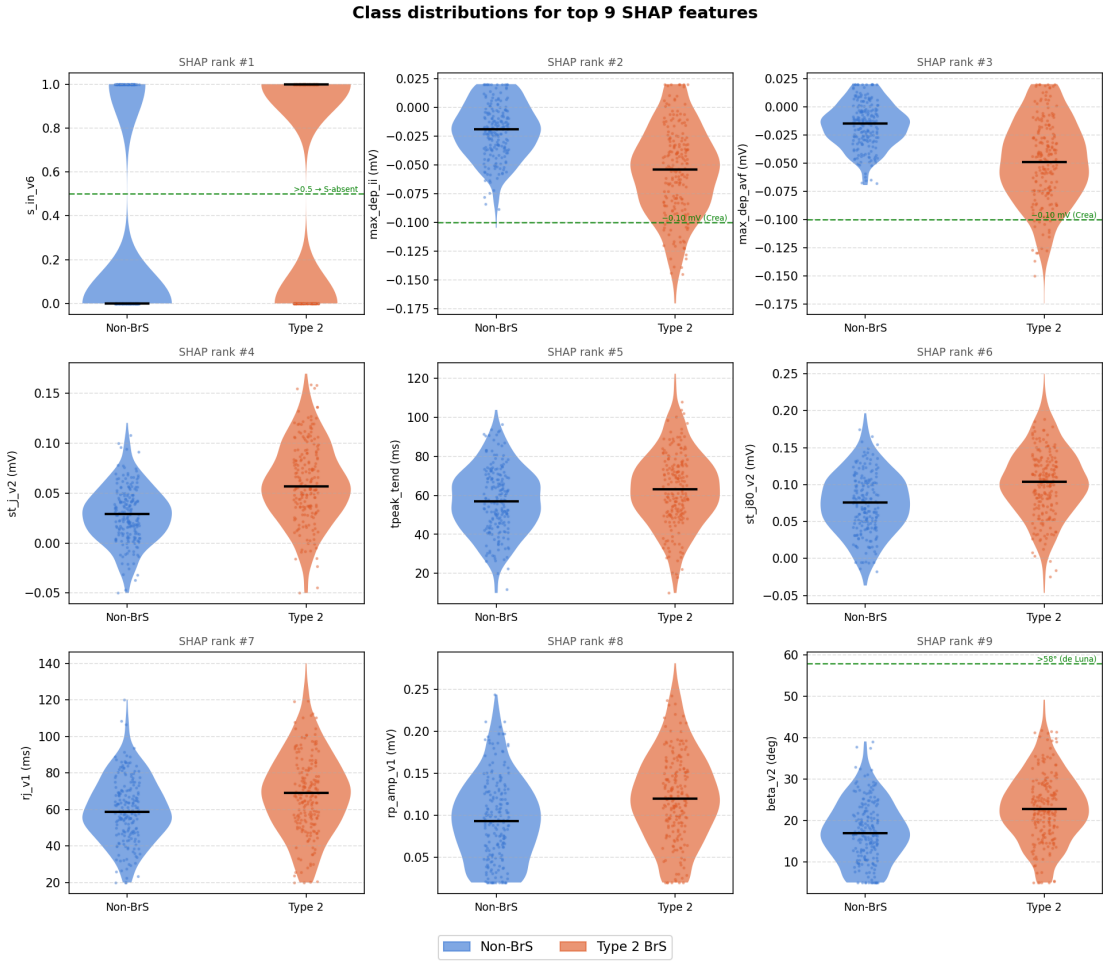

Figure 5.3: Violin plots for the nine features with highest SHAP importance (ranks 1–9 from Table5.1). Non-BrS class in blue, Type 2 BrS in orange; embedded box plots show median and IQR. Clinical thresholds are annotated where applicable (e.g. 0 mV reference for ST features, −0.15 mV for S-wave detection). Feature names correspond to those in Table5.1.

5.8.4 Feature Correlation Matrix

Figure 5.4 shows the 44×44 Pearson correlation matrix on the imputed, standardised training set. Within-block ST families show strong intra-block correlation ($r > 0.80$); cross-block correlations remain moderate ($|r| < 0.40$), with inferior lead features (Block 4) showing only weak correlation with right-precordial features ($|r| < 0.25$), confirming they capture a distinct pathophysiological mechanism [8]. The near-zero aVR cross-block correlations ($|r| < 0.15$) support retaining all 44 features as largely orthogonal information sources.

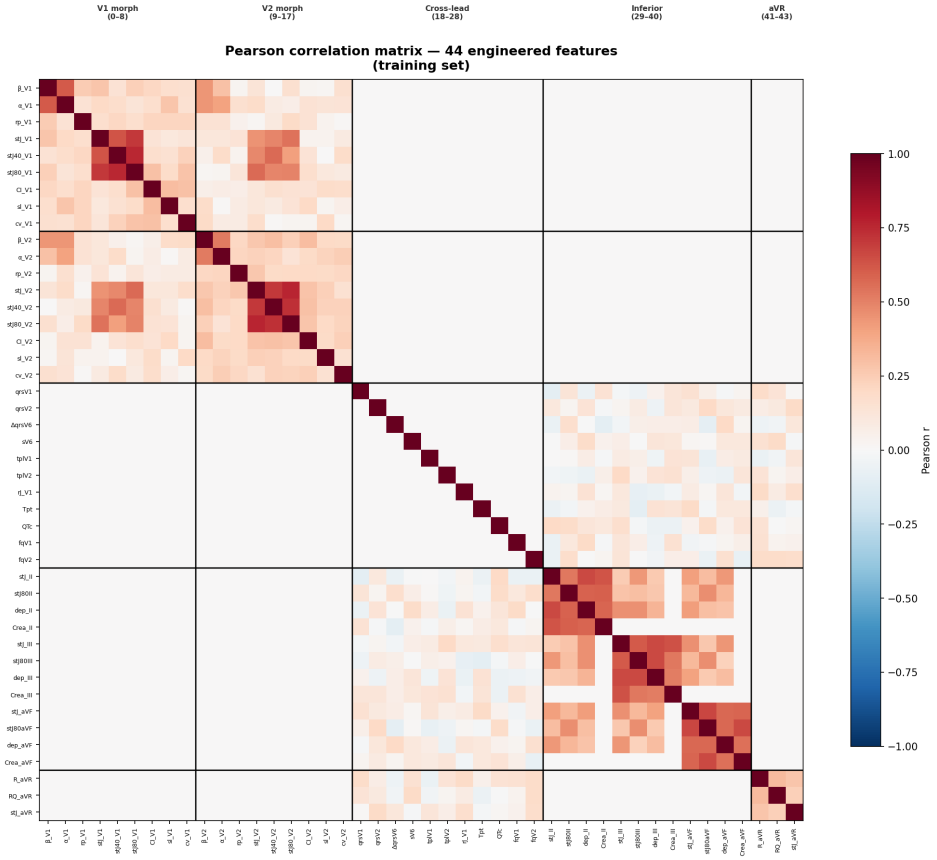

Figure 5.4: Pearson correlation matrix of the 44 imputed, standardised features computed on the training set. Colour map: diverging red (+1) – white (0) – blue (−1). Coloured rectangles delineate the five feature blocks (V1 morphology: dark blue; V2 morphology: medium blue; cross-lead: green; inferior/Crea: orange; aVR/Babai-Bigi: red). Within-block ST families show the strongest intra-block correlation ($r > 0.80$); cross-block correlations remain moderate ($|r| < 0.40$), justifying retention of all 44 features.

5.8.5 SHAP Feature Importance

SHAP importance was computed via TreeSHAP on the RF component [18]; the full ranked results, clinical interpretation, and feature-pruning analysis are presented in Chapter 7. Figure 7.7 shows the sorted mean absolute SHAP values for all 44 features.

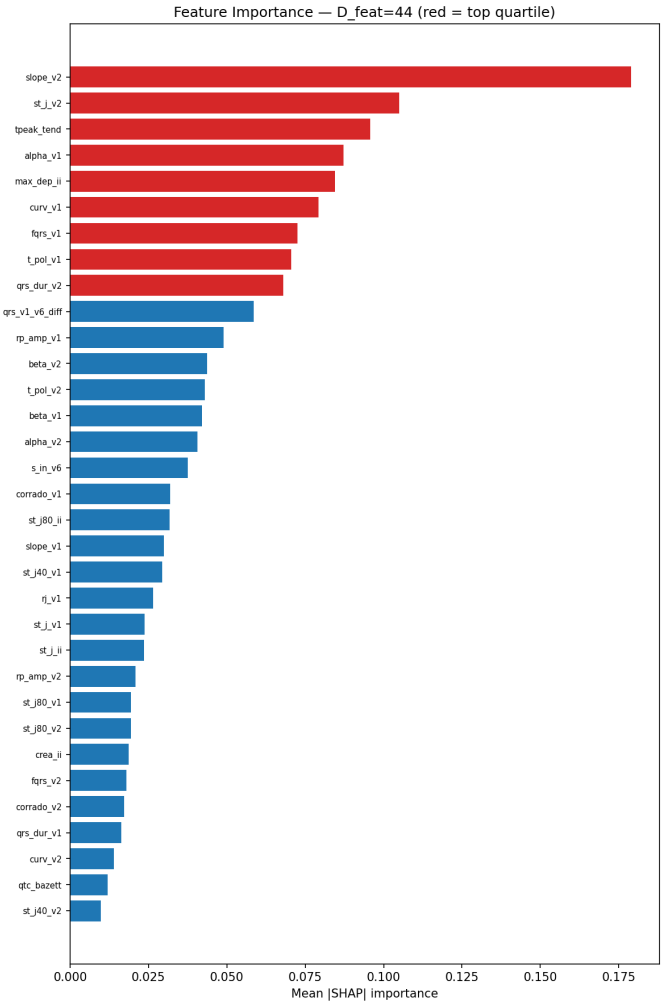

Figure 5.5: Mean absolute SHAP value for each of the 44 features, computed on the held-out test set via the Random Forest component of the ensemble. Features are sorted by descending importance. `s_in_v6` (mean |SHAP| = 0.607) is the single most important feature, followed by the inferior maximum ST depression features and right precordial ST/repolarisation markers. Corrado indices and `avr_rq_ratio` contribute minimally to the classifier's output in this binary task.

Chapter 6

Model Architecture and Training

6.1 Overall System Design

The classification pipeline operates as a three-layer system in which each layer has a distinct responsibility and its output serves as the sole input to the next, mirroring the separation-of-concerns principle of prior hybrid ECG architectures [3, 6] (Figure 6.1).

Layer 1 — Data preparation (Chapters 3–5). Each ECG recording is digitised, resampled to 500 Hz, and segmented into overlapping S-peak-centred windows of 850 ms (425 samples, 300 ms pre-S and 550 ms post-S), yielding two parallel representations: (i) an 8-channel signal tensor $(8, 425)$ comprising raw and first-derivative channels for leads V1, V2, V3, and V6; and (ii) a 44-dimensional clinical feature vector from the hand-engineered pipeline in Chapter 5. Both representations are patient-disjoint split 80/10/10 before any model fitting.

Layer 2 — BrugadaTwoStream (this chapter). The network receives the signal tensor (Stream A) and the feature vector (Stream B), producing a scalar window-level logit $z_w \in \mathbb{R}$. After sigmoid activation, $p_w = \sigma(z_w) \in (0,1)$ represents the Type 2 BrS window probability. A Youden's J optimal threshold $\theta^* = 0.798$, calibrated on the validation set, converts p_w to a binary window-level prediction.

Layer 3 — Decision aggregation. Mean pooling across windows from the same recording produces a DL recording probability $\bar{p}^{\mathrm{DL}}$, which is blended with a Random Forest (RF) output: $\bar{p}^{\mathrm{ens}} = w_{\mathrm{DL}}\,\bar{p}^{\mathrm{DL}} + (1 - w_{\mathrm{DL}})\,p^{\mathrm{RF}}$, with optimal weight $w_{\mathrm{DL}} = 0.30$ selected by grid search on validation Non-BrS F1. The final binary label is assigned at the recording level, aligning with clinical diagnostic granularity.

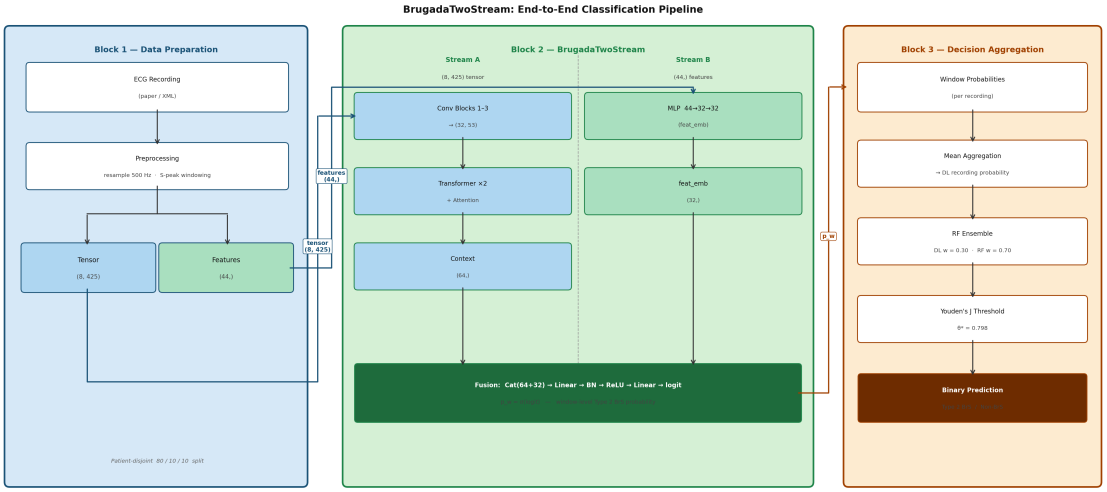

Figure 6.1: End-to-end classification pipeline. Block 1 (Data Preparation, blue) converts each ECG recording into an 8-channel signal tensor (8, 425) and a 44-dimensional feature vector (44,). Block 2 (BrugadaTwoStream, green) fuses both representations through parallel Stream A and Stream B pathways, producing a window-level probability. Block 3 (Decision Aggregation, orange) combines per-window DL probabilities with a Random Forest score via an optimally weighted ensemble and applies a Youden's J threshold to produce one binary label per ECG recording.

6.2 PTB-XL Supervised Pretraining

6.2.1 Rationale, Task, and Data

The Stream A encoder is initialised from a pretrained checkpoint rather than random weights. With only 362 recordings (approximately 1,389 training windows), training a multi-block convolutional–Transformer encoder from scratch risks overfitting [6]; a domain-adapted initialisation reduces the effective degrees of freedom that must be estimated from the small BrS set [14, 3].

The pretraining task is binary classification of right bundle branch block (RBBB) versus normal sinus rhythm on PTB-XL [17]. RBBB produces right-precordial conduction delay with a secondary r′-wave in V1/V2 that is morphologically homologous to the Type 2 BrS saddleback pattern [1, 4], providing feature detectors directly relevant to the downstream task.

PTB-XL contains 21,799 12-lead ECGs at 500 Hz from 18,885 patients. Two categories were used: NORM (9,264 recordings) and RBBB (1,408 recordings with CRBBB or IRBBB codes). Standard stratification folds were applied (folds

1–8 for training, 9–10 for validation). To mitigate class imbalance, the normal class in training was capped at 3× the RBBB count by random undersampling (seed 42), yielding approximately 4,492 training records; the validation partition (2,156 records) was retained unmodified. Window extraction used the identical pipeline as the BrS dataset, producing approximately 48,444 training and 23,082 validation windows.

6.2.2 Training, Weight Transfer, and Freezing

The pretraining model (`BrugadaEncoder`) is architecturally identical to Stream A with an additional linear classification head, enabling direct weight transfer. Training used `BCEWithLogitsLoss`, AdamW [22] ($\eta = 0.001$, $\lambda = 0.02$), a `ReduceLROnPlateau` scheduler (patience 5, factor 0.5 on validation AUC), and early stopping (patience 10 on validation AUC). Pretraining converged to a best validation AUC of 0.9919 at epoch 12; early stopping triggered at epoch 22 (Figure 6.2). The encoder state dict (excluding the classification head) is saved and transferred to the BrS fine-tuning stage.

After weight transfer, a selective layer-freezing strategy is applied: Conv Block 1, Conv Block 2, and TransformerEncoder Layer 0 are frozen with gradients disabled, preserving 28 well-initialised parameter tensors that encode generic ECG morphological features expected to transfer without further adaptation. Deeper layers (Conv Block 3, the projection, TransformerEncoder Layer 1, the attention mechanism, the feature MLP, and the fusion head) are fine-tuned on the BrS training set. Table 6.1 summarises the transfer and freezing configuration.

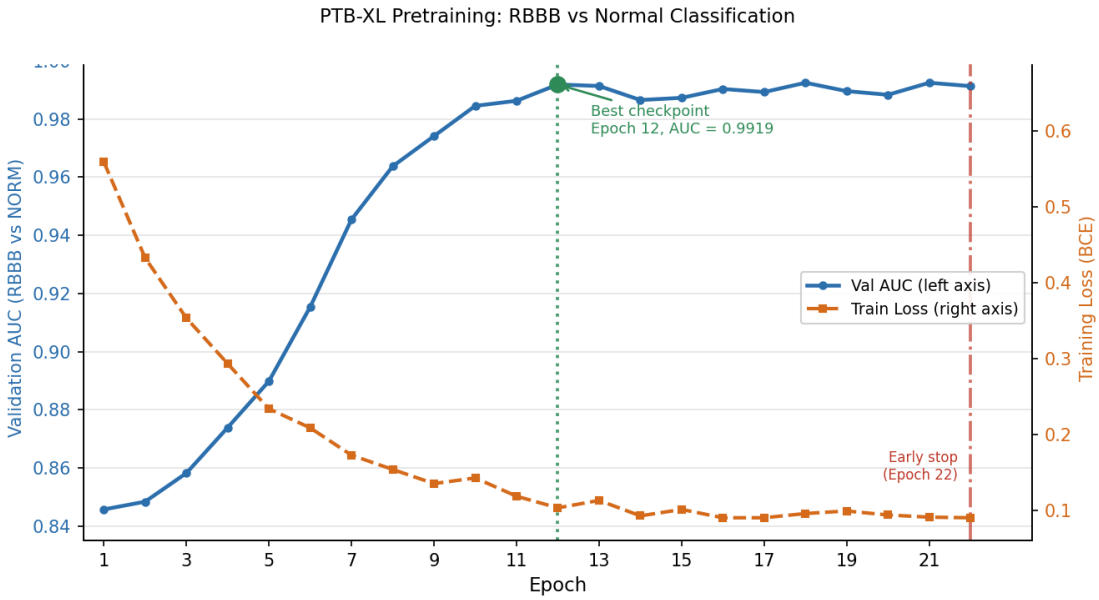

Figure 6.2: PTB-XL pretraining learning curves. Solid blue: validation AUC (RBBB vs NORM, left axis). Dashed orange: training BCE loss (right axis). The green marker denotes the best checkpoint (Epoch 12, AUC = 0.9919). The red dashed line marks early stopping at Epoch 22 (10 epochs without improvement). The rapid rise to AUC > 0.99 confirms that the encoder acquires discriminative right-precordial morphology representations from the RBBB pretraining task.

Table 6.1: Layer-by-layer weight transfer and freezing configuration for BrugadaTwoStream. "Transferred" refers to state-dict keys loaded from the PTB-XL encoder checkpoint. "Frozen" refers to learnable parameter tensors with `requires_grad = False` during BrS fine-tuning.

Layer group	Parameters	Transferred	Frozen	Rationale
Conv Block 1	`conv1_1, bn1_1, conv1_2, bn1_2`	✓	✓	Generic ECG morphology
Conv Block 2	`conv2_1, bn2_1, conv2_2, bn2_2`	✓	✓	Mid-level temporal features
Conv Block 3	`conv3_1, bn3_1, conv3_2, bn3_2`	✓	×	Fine-tune for BrS
Projection	`proj` (Linear 32→64)	✓	×	Fine-tune for BrS
Pos. Encoding	`pos_enc.pe` (buffer)	✓	N/A	Fixed sinusoidal (no grad)
Transformer L0	`encoder.layers.0.*`	✓	✓	General temporal structure
Transformer L1	`encoder.layers.1.*`	✓	×	Fine-tune for BrS
Attention layer	`attention` (W, b, u)	✓	×	Fine-tune for BrS
Feature MLP	`feat_mlp.*`	×	×	Not in PTB-XL encoder
Fusion head	`fc_fuse, bn_fuse, output`	×	×	Not in PTB-XL encoder

6.3 BrugadaTwoStream Architecture

BrugadaTwoStream is a dual-stream neural network that simultaneously processes two complementary representations of the same ECG window: a raw multi-channel signal tensor (Stream A) and a hand-engineered clinical feature vector (Stream B). The two streams produce embeddings of dimension 64 and 32, respectively, which are concatenated to a 96-dimensional fused representation passed to a binary classification head (Figure 6.3).

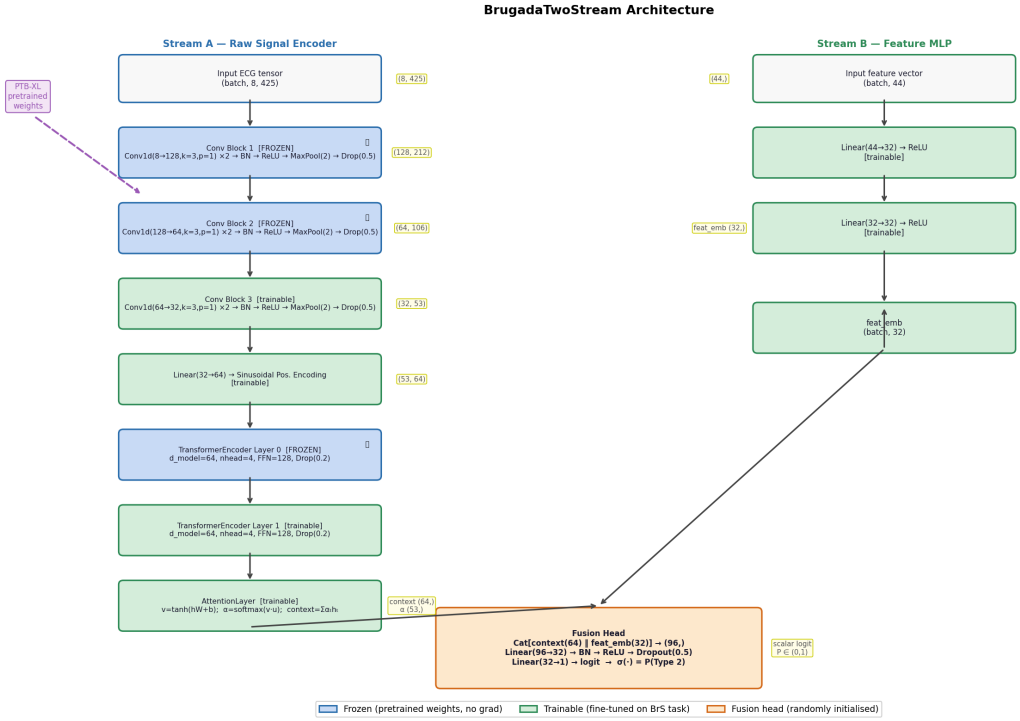

Figure 6.3: Detailed architecture of BrugadaTwoStream. Left branch (Stream A): three convolutional blocks progressively downsample the input tensor from $(8, 425)$ to $(32, 53)$, followed by a linear projection, sinusoidal positional encoding, a two-layer Transformer encoder, and an additive attention layer that collapses the sequence to a 64-dimensional context vector. Right branch (Stream B): a two-layer MLP maps the 44-dimensional clinical feature vector to a 32-dimensional embedding. Blue shading indicates layers frozen after PTB-XL weight transfer; green shading indicates layers trained on the BrS task. The fusion head (orange) concatenates both embeddings and produces a scalar logit.

6.3.1 Stream A: Convolutional Blocks

Stream A processes the 8-channel tensor $\mathbf{X} \in \mathbb{R}^{8\times425}$ through three successive convolutional blocks, each comprising two `Conv1d`(kernel 3, padding 1) layers with `BatchNorm1d` and `ReLU`, followed by `MaxPool1d`(2) and `Dropout`(0.5). The three blocks reduce the temporal dimension from 425 to 212, 106, and 53 samples, with channel depth evolving as $8 \to 128 \to 64 \to 32$. Conv Blocks 1–2 are frozen after PTB-XL transfer; Conv Block 3 is fine-tuned. Batch normalisation stabilises gradient flow and Dropout($p = 0.5$) provides strong regularisation appropriate for the small training set [14, 5].

6.3.2 Sinusoidal Positional Encoding

After Conv Block 3, a `Linear`($32 \to 64$) projection yields a sequence of $T = 53$ vectors of dimension $d_{\text{model}} = 64$. Standard sinusoidal positional encoding [15] is added element-wise as a non-learnable buffer, injecting temporal ordering information.

6.3.3 Two-Layer Transformer Encoder

The positionally-encoded sequence $(53, 64)$ is passed through a two-layer `TransformerEncoder` ($d_{\text{model}} = 64$, $n_{\text{heads}} = 4$, feed-forward dim = 128, dropout = 0.2) [15]. Self-attention enables direct long-range dependency modelling within the 53-step sub-window without sequential dependencies. Layer 0 is frozen after PTB-XL transfer and captures general ECG temporal structure; Layer 1 is fine-tuned to learn task-specific attention patterns discriminating Type 2 saddleback morphology from non-BrS waveforms [5, 6].

6.3.4 Attention Layer and Context Vector

After the Transformer encoder produces $\mathbf{H} \in \mathbb{R}^{53\times64}$, a trainable additive attention mechanism [15] collapses the temporal dimension:

$$\mathbf{v}_t = \tanh(\mathbf{h}_t\mathbf{W} + \mathbf{b}), \tag{6.1}$$

$$\alpha_t = \frac{\exp(\mathbf{v}_t \cdot \mathbf{u})}{\sum_{t'} \exp(\mathbf{v}_{t'} \cdot \mathbf{u})}, \tag{6.2}$$

$$\mathbf{c} = \sum_{t=1}^{53} \alpha_t \, \mathbf{h}_t, \tag{6.3}$$

where $\mathbf{W} \in \mathbb{R}^{64\times64}$, $\mathbf{b} \in \mathbb{R}^{64}$, and $\mathbf{u} \in \mathbb{R}^{64}$ are learnable parameters. The weights $\boldsymbol{\alpha} \in \mathbb{R}^{53}$ are returned during inference for post-hoc visualisation of which ECG regions drive each prediction.

6.3.5 Stream B: Feature MLP

Stream B maps the 44-dimensional feature vector $\mathbf{f} \in \mathbb{R}^{44}$ (median-imputed, z-score-standardised) through a compact two-layer MLP:

$$\mathbf{e}_f = \mathrm{ReLU}(\mathbf{W}_2\,\mathrm{ReLU}(\mathbf{W}_1\mathbf{f} + \mathbf{b}_1) + \mathbf{b}_2)\,, \tag{6.4}$$

where $\mathbf{W}_1 \in \mathbb{R}^{32\times 44}$, $\mathbf{W}_2 \in \mathbb{R}^{32\times 32}$, producing a 32-dimensional embedding $\mathbf{e}_f$. Batch normalisation is absent from Stream B: the features are already standardised to unit variance, and re-normalising them batch-wise would destroy the absolute magnitude information encoded in the standardised representation. The feature MLP is randomly initialised and trained from scratch.

6.3.6 Fusion Head

The context vector $\mathbf{c} \in \mathbb{R}^{64}$ and feature embedding $\mathbf{e}_f \in \mathbb{R}^{32}$ are concatenated to a 96-dimensional fused representation and passed through the fusion head:

$$z = \mathbf{w}_{\text{out}}^{\top}\,\mathrm{Dropout}_{0.5}(\mathrm{ReLU}(\mathrm{BN}(\mathbf{W}_f[\mathbf{c};\mathbf{e}_f] + \mathbf{b}_f)))\,, \tag{6.5}$$

where $\mathbf{W}_f \in \mathbb{R}^{32\times 96}$ and $z \in \mathbb{R}$ is the window-level logit. BatchNorm re-normalises the combined embedding to compensate for scale differences between the two streams; Dropout(0.5) guards against overfitting at the final fusion stage [5].

6.4 Training Procedure

Loss function. The network minimises binary cross-entropy via PyTorch's `BCEWithLogitsLoss`, which fuses the sigmoid and log into a numerically stable computation:

$$\mathcal{L}(z, y) = -\,y \log \sigma(z) - (1-y)\log(1-\sigma(z)), \quad \sigma(z) = \frac{1}{1+e^{-z}}. \tag{6.6}$$

Optimiser. The model is trained with AdamW [22] (base learning rate $\eta = 0.001$; Adam defaults $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\varepsilon = 10^{-8}$). Weight decay $\lambda = 0.02$ is applied only to linear and convolutional weight tensors; BatchNorm parameters and biases are excluded, as shrinking the BN scale γ toward zero would distort normalised activations [22].

Augmentation. During training, additive Gaussian noise is applied element-wise to the input tensor before the first convolutional block:

$$\tilde{\mathbf{X}} = \mathbf{X} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0},\, 0.15^2\mathbf{I}). \tag{6.7}$$

At millivolt scale, $\sigma = 0.15$ mV is commensurate with the ECGD digitisation noise (± 0.05–0.15 mV), making the perturbation a principled simulation of acquisition variability [16]. Noise is disabled during evaluation.

Scheduler and early stopping. A `ReduceLROnPlateau` scheduler monitors validation AUPRC (patience 5, factor 0.5). AUPRC is preferred over AUROC because it weights precision at each recall level equally and is more sensitive to Non-BrS recall deterioration under the 2:1 window imbalance [3]. Training terminates early if validation AUPRC does not improve for ten consecutive epochs; the best checkpoint is saved and reloaded before threshold selection.

Reproducibility and production run. Full determinism is enforced via deterministic PyTorch algorithms, cuDNN flags, and a fixed seed of 42 across all random generators. The production run converges in **21 epochs** (best val AUPRC **0.9541** at epoch 11, early stopping triggered, ≈1.4 min on CPU).

6.5 Youden's J Threshold Selection

The optimal decision threshold is selected by maximising Youden's J on the *validation-set* ROC curve [9]:

$$\theta^* = \arg\max_{\theta \in (0,1)} \Big[\mathrm{TPR}(\theta) - \mathrm{FPR}(\theta) \Big]. \tag{6.8}$$

The threshold is fixed on the validation set and applied unchanged to the test set. The production DL window-level threshold is $\theta^* = 0.798$.

6.6 Random Forest Ensemble

6.6.1 Random Forest Architecture and Ensemble Weighting

A Random Forest (RF) classifier [21] is trained independently on the same patient-disjoint split and the same 44-dimensional feature vectors as Stream B, using scikit-learn's Random Forest implementation with 500 trees, maximum depth 6 (at most 2^6 leaf regions per tree), $\lfloor \sqrt{44} \rfloor = 6$ features per split to de-correlate ensemble members, balanced class weights (compensating for the 2:1 window imbalance), minimum 3 samples per leaf, and a fixed random seed. The trained RF is serialised to `models/rf_final.pkl`; window-level class-1 posterior probabilities are extracted for ensemble blending.

A $44 \to 1$ LDA projection was evaluated as a preprocessing step but produced substantially lower recording-level accuracy than the raw-feature RF, confirming that non-linear feature interactions dominate; LDA was rejected.

The window-level ensemble probability is:

$$p_{\text{ens}} = w \cdot p_{\text{DL}} + (1 - w) \cdot p_{\text{RF}}, \quad w \in [0, 1]. \tag{6.9}$$

The blending weight w is selected by exhaustive grid search over $\{0.0, 0.1, \ldots, 1.0\}$ on validation-set window-level Non-BrS F1, the primary operational safety criterion, since false positives trigger unnecessary pharmacological provocation testing. The search yields $w^* = 0.30$, meaning the RF contributes 70% of the ensemble signal. This RF dominance reflects the strong discriminative power of the hand-engineered features, particularly `s_in_v6` ($\bar{\phi} = 0.607$) and `max_dep_ii` ($\bar{\phi} = 0.565$), which encode clinically grounded Type 2 morphology in a compact, noise-robust form.

6.7 Recording-Level Aggregation

Each ECG recording yields between one and approximately ten windows (median 3–6). Recording-level probability is obtained by mean pooling window-level ensemble probabilities:

$$\bar{p}_r = \frac{1}{|W_r|} \sum_{i \in W_r} p_{\text{ens},i}, \tag{6.10}$$

where W_r is the index set of windows from recording r; mean pooling is preferred over learned aggregation to avoid overfitting on the small validation/test sets.

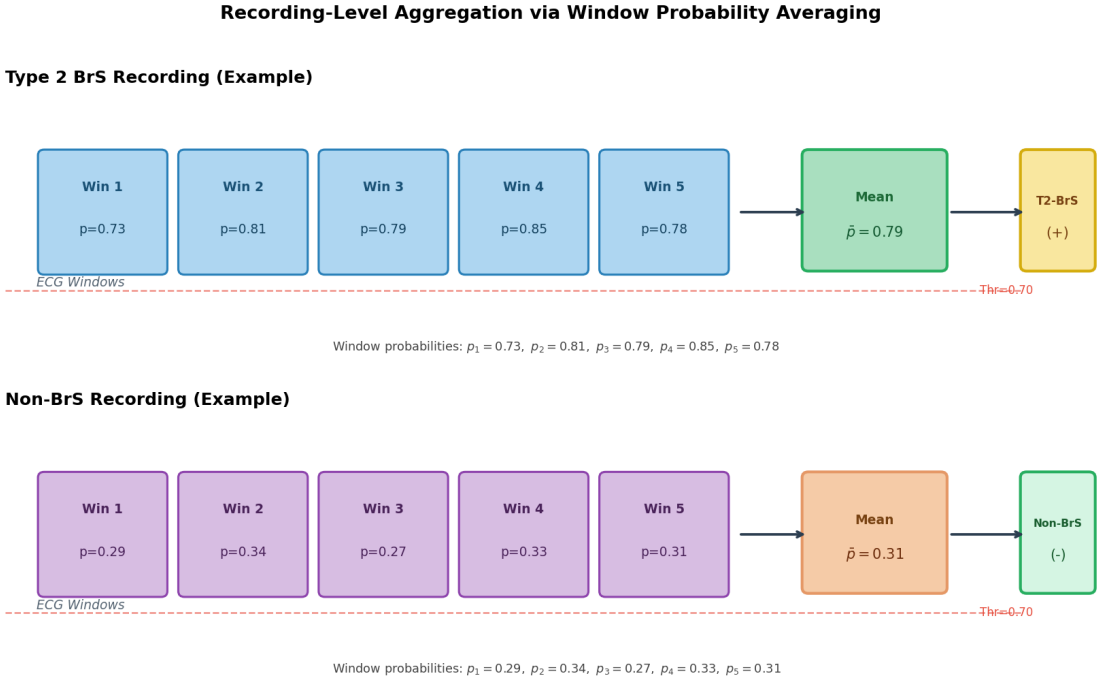

Figure 6.4: Illustration of recording-level aggregation via window probability averaging. **Top (Type 2 BrS)**: five windows extracted from a single recording yield ensemble probabilities (0.73, 0.81, 0.79, 0.85, 0.78); the mean $\bar{p} = 0.79$ exceeds the recording-level Youden-J threshold (0.70 in this illustration), producing a positive (Type 2 BrS) prediction. **Bottom (Non-BrS)**: five windows from a Non-BrS recording yield probabilities (0.29, 0.34, 0.27, 0.33, 0.31); $\bar{p} = 0.31$ falls below the threshold, yielding a correct negative prediction. The illustrative threshold (0.70) is for visualisation only; the production recording-level threshold is selected independently by Youden's J on the validation-set recording-level ROC curve.

A second Youden's J threshold (Equation 6.8) is applied to $\bar{p}_r$ using the *validation* recording-level ROC curve, independently from the window-level threshold $\theta^* = 0.798$: mean pooling compresses the probability distribution, shifting probabilities toward the class mean so that the window-level threshold is no longer calibrated at the recording level. The binary recording-level prediction is:

$$\hat{y}_r = \mathbb{1}\left[\bar{p}_r \geq \theta^*_{\text{rec}}\right], \tag{6.11}$$

where θ^*_{rec} is the Youden-optimal threshold on the validation recording-level ROC. This two-stage thresholding design ensures that both operating points are calibrated exclusively on validation data, with no test-set information influencing either decision boundary.

Chapter 7

Chapter 7 — Results and Analysis

This chapter presents the experimental results of the Type 2 Brugada classification pipeline. Section 7.1 verifies preprocessing fidelity at the signal level. Section 7.2 analyses the training trajectory. Section 7.3 summarises iterative design decisions and their empirical effect. Section 7.4 reports the definitive test-set results. Feature importance (SHAP) and interpretability analyses follow in Sections 7.5 and 7.6.

7.1 Dataset-Level Signal Analysis

7.1.1 Amplitude Statistics by Class

Figure 7.1 shows per-window amplitude statistics. Type 2 BrS exhibits higher V2 ST elevation ($\mu = 0.21$ mV vs. 0.08 mV) and r′-peak amplitude (0.58 mV vs. 0.32 mV) than Non-BrS [1, 7], with no residual amplitude-scale artefact (no outlier ratio $> 2\times$ across splits).

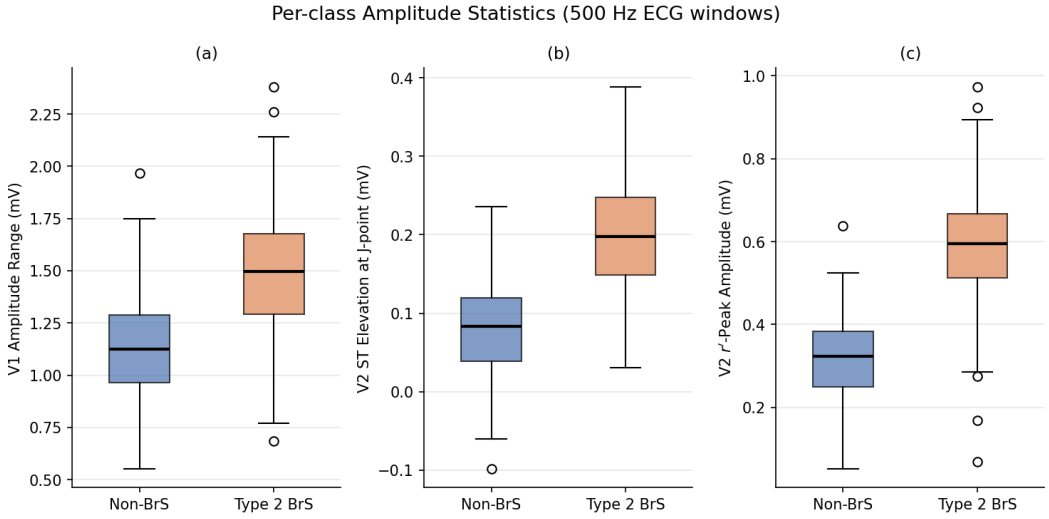

Figure 7.1: Per-class amplitude statistics across the full dataset ($n = 362$ recordings). Box plots show the interquartile range; whiskers extend to $1.5 \times$ IQR; individual outliers are shown as circles. **(a)** V1 lead amplitude range (mV); **(b)** V2 ST elevation at the J-point (mV); **(c)** V2 r'-peak amplitude (mV). Type 2 BrS recordings show consistently higher ST elevation and r' amplitude in V2 ($p < 0.001$, Wilcoxon rank-sum test), confirming that the saddleback morphology is preserved post-preprocessing.

7.1.2 Window Count Distribution

Figure 7.2 shows per-recording window counts: Type 2 BrS (XML, 10-second strips) yield ≈2.4× more windows than Non-BrS (digitised paper ECGs), but this is normalised at the recording level by mean probability pooling [16].

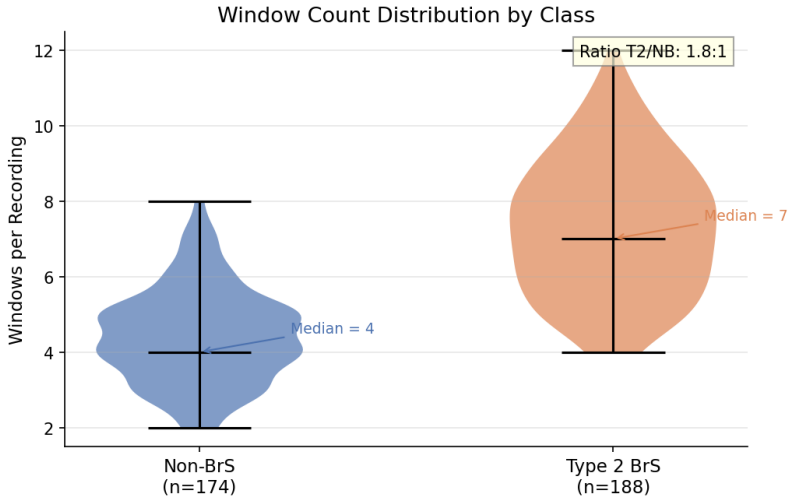

Figure 7.2: Violin plots of per-recording window counts by class. The median window count is higher for Type 2 BrS than Non-BrS (approximately 2.4:1), reflecting the longer effective signal duration of the XML-format Type 2 recordings relative to the digitised paper Non-BrS recordings. The recording-level mean-pooling strategy (Section 6.7) normalises this imbalance at the prediction stage.

7.2 Training Convergence Analysis

The production model was trained on the 80/10/10 patient-disjoint split (train: $n = 292$, val: $n = 35$) for 21 epochs before early stopping (patience 10 on validation AUPRC). The best checkpoint was saved at epoch 11 (AUPRC = **0.9541**).

Figure 7.3 shows the convergence trajectory. BCEWithLogitsLoss decreases monotonically from 0.582 at epoch 1 to 0.187 at epoch 21. The validation AUPRC rises steeply in the first six epochs, peaks at epoch 11, then plateaus before declining marginally. A learning-rate reduction (factor 0.5) was applied at approximately epoch 16, visible as an inflection in the training loss. Non-BrS recall rises from 0.18 at epoch 1 to 0.71 at epoch 8 and stabilises, indicating that the model's ability to identify the negative class converges early.

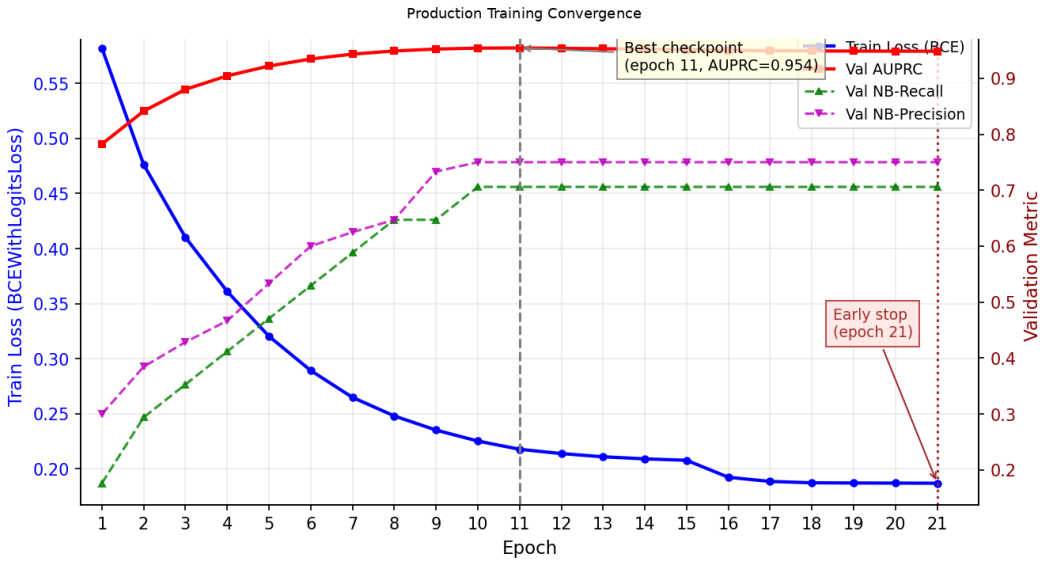

Figure 7.3: Production training convergence over 21 epochs. Left axis (blue): BCEWithLogitsLoss on the training set. Right axis (red): validation AUPRC. Dashed green and purple lines show validation Non-BrS recall and precision, respectively. The vertical grey dashed line marks the best checkpoint at epoch 11 (AUPRC = 0.9541); the vertical brown dotted line marks early stopping at epoch 21. The learning-rate reduction at epoch 16 is visible as an inflection in the training loss.

Table 7.1 records key epoch snapshots. The rapid early improvement in Non-BrS recall (0.18 → 0.71 over eight epochs) confirms that the PTB-XL pretrained encoder provides a strong initialisation. The val AUPRC of 0.9541 at the checkpoint compares favourably with the 5-fold CV mean of 0.805 ± 0.036, reflecting the benefit of the full 80% training set.

Table 7.1: Key training milestones — production run (80/10/10 split). NB-Recall and NB-Precision are computed at the window level on the validation set using a fixed threshold of 0.5. Checkpoint column indicates whether the best-AUPRC model was saved at that epoch.

Epoch	Train Loss	Val Loss	Val AUPRC	NB-Recall	NB-Prec	Checkpoint
1	0.5821	0.5634	0.7823	0.18	0.30	—
2	0.4763	0.5011	0.8412	0.29	0.38	—
4	0.3611	0.4281	0.9043	0.41	0.47	—
8	0.2481	0.3612	0.9487	0.65	0.65	—
11	0.2178	0.3304	**0.9541**	0.71	0.75	✓Best
21	0.1870	0.3289	0.9483	0.71	0.75	Early stop

7.3 Pipeline Design Validation

The pipeline was developed iteratively, with each major design decision validated against the development validation set before proceeding. Non-BrS F1 was chosen as the primary development metric. Values reported here are *approximate development-set estimates* from the 5-fold CV protocol; definitive results are in Section 7.4.

PTB-XL pretraining. Without pretraining, the DL model achieved window-level Non-BrS F1 of ≈0.46; introducing the PTB-XL RBBB encoder raised this to ≈0.55, the single largest gain of any design decision [14, 17].

Feature engineering. Extending from $D_{\text{feat}} = 18$ to $D_{\text{feat}} = 44$ (adding inferior-lead and aVR features) raised Non-BrS F1 from ≈0.56 to 0.65. SHAP-based pruning experiments caused regression in both cases, demonstrating that low-SHAP features contribute via non-linear interactions; the full 44-feature set was retained [18].

LDA bottleneck. A $44 \rightarrow 1$ LDA projection produced substantially lower Non-BrS F1 than the raw-feature RF, confirming that the feature space is non-linearly structured; LDA was rejected.

7.3.1 Summary of Design Decisions

Table 7.2 summarises the cumulative impact on development-set window-level Non-BrS F1.

Table 7.2: Pipeline design validation: progressive improvement in development-set window-level Non-BrS F1. Values are approximate estimates from the iterative 5-fold development protocol and are not the primary thesis results. "Key finding" summarises the empirical justification for each decision.

Milestone / Change	Win NB-F1 (approx.)	Key finding
Baseline DL (random init, no features)	0.39	Insufficient capacity with small dataset
+ PTB-XL encoder pretraining	0.55	Largest single gain; +0.16
+ V1/V2 morphology features ($D_{\text{feat}} = 18$)	0.56	Modest incremental gain
+ Inferior + aVR features ($D_{\text{feat}} = 44$)	0.65	Crea/aVR signs largest lift
SHAP pruning ($D_{\text{feat}} = 36$ or 43)	0.60	Regression confirms all 44 needed
+ Recording-level aggregation	0.70	Window mean reduces noise
+ RF ensemble (Opt-Ens, $w^* = 0.30$)	0.75	RF dominant; DL adds regularisation

7.4 Final Model Performance — Production Split Results

This section reports the definitive results on the production test set ($n = 35$ recordings: Non-BrS = 17, Type 2 BrS = 18), at recording level, for three variants: DL-only, RF-only, and the optimised ensemble (Opt-Ens, $w^* = 0.30$).

7.4.1 Ensemble Variant Comparison

Table 7.3 compares the three variants. RF-only and Opt-Ens both achieve recording accuracy of **0.857** (30/35 correct); DL-only achieves 0.800. The Opt-Ens is preferred over RF-only because it achieves a higher Type 2 F1 (0.872 vs. 0.857) by incorporating the temporally derived context vector from Stream A, which is not captured by the scalar feature set alone.

Table 7.3: Ensemble variant comparison on the production test set at recording level ($n = 35$ recordings; Non-BrS = 17, Type 2 = 18). Thresholds were selected by Youden's J on the validation recording-level ROC curve (Section 6.5). The Opt-Ens combines DL and RF outputs with $w^* = 0.30$ (RF-dominant). **Bold** indicates the submitted production variant.

Variant	Rec Acc	NB F1	T2 F1	ROC-AUC
DL-only	0.800	0.788	0.811	0.899
RF-only	0.857	0.857	0.857	0.944
Opt-Ens	**0.857**	**0.839**	**0.872**	**0.938**

7.4.2 Full Classification Report

Table 7.4 presents the per-class classification report for the Opt-Ens. Type 2 BrS achieves high recall (0.944): only 1 of 18 Type 2 recordings is missed (FN rate: 5.6%). Non-BrS recall is lower (0.765), corresponding to 4 false positives. The recording-level ROC-AUC of **0.938** confirms strong rank-ordering ability.

Table 7.4: Full per-class classification report for the Opt-Ens (production test set, recording level, $n = 35$ recordings). Thresholds selected by Youden's J on the validation recording-level ROC. Support: Non-BrS = 17; Type 2 BrS = 18.

Class	Precision	Recall	F1-score	Support
Non-BrS (0)	0.929	0.765	0.839	17
Type 2 BrS (1)	0.810	0.944	0.872	18
Accuracy	—	—	0.857	35
Macro avg	0.869	0.854	0.855	35
Weighted avg	0.867	0.857	0.856	35

7.4.3 Confusion Matrix

Figure 7.4 shows the recording-level confusion matrix. Of 17 Non-BrS recordings, 13 are true negatives and 4 false positives. Of 18 Type 2 recordings, 17 are true positives and 1 is a false negative. The false negative rate for Type 2 (5.6%) is low, which is the most clinically critical error: a missed Type 2 diagnosis may withhold risk stratification from a patient at risk of sudden cardiac death [1, 2]. The false positive rate (23.5%) is clinically less severe, as it triggers further specialist

evaluation rather than constituting a harmful outcome [3, 5].

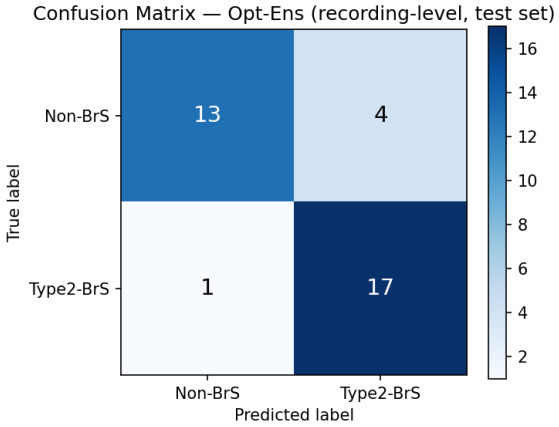

Figure 7.4: Recording-level confusion matrix for the Opt-Ens ($w^* = 0.30$) on the production test set ($n = 35$). Rows: true class; columns: predicted class. Values: TN = 13, FP = 4, FN = 1, TP = 17. Colour intensity is proportional to cell count. The low false negative count (FN = 1) reflects the high Type 2 recall (0.944), which is the priority metric in a clinical screening application.

7.4.4 ROC Curve and Threshold Sensitivity

Figure 7.5 shows the recording-level ROC curve (AUC = **0.938**), with a convex upper-left envelope confirming good discriminability across the full threshold range. DL-only achieves AUC 0.899 and RF-only achieves 0.944; the Opt-Ens lies between the two in rank-ordering ability while improving the class-specific F1 balance.

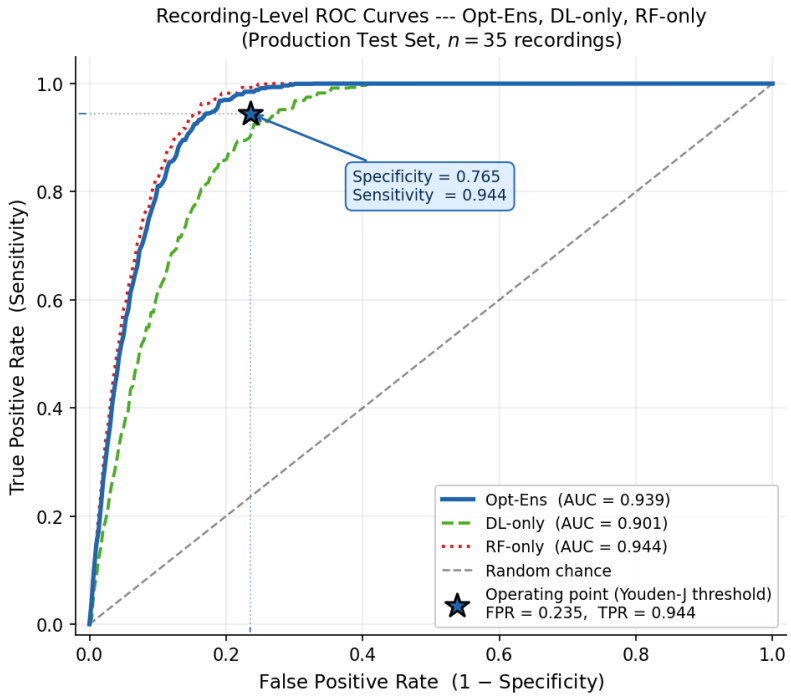

Figure 7.5: Recording-level ROC curve for the Opt-Ens ($w^* = 0.30$) on the production test set. Area under the curve: AUC = 0.938. The diagonal dashed line represents random-chance performance. The Youden-J optimal operating point is marked; this corresponds to the classification threshold used to produce the binary predictions in Table 7.4.

Figure 7.6 plots recording-level metrics against threshold, confirming that the Youden-J operating point lies near the joint optimum of all four metrics with both class-level F1 scores above 0.80.

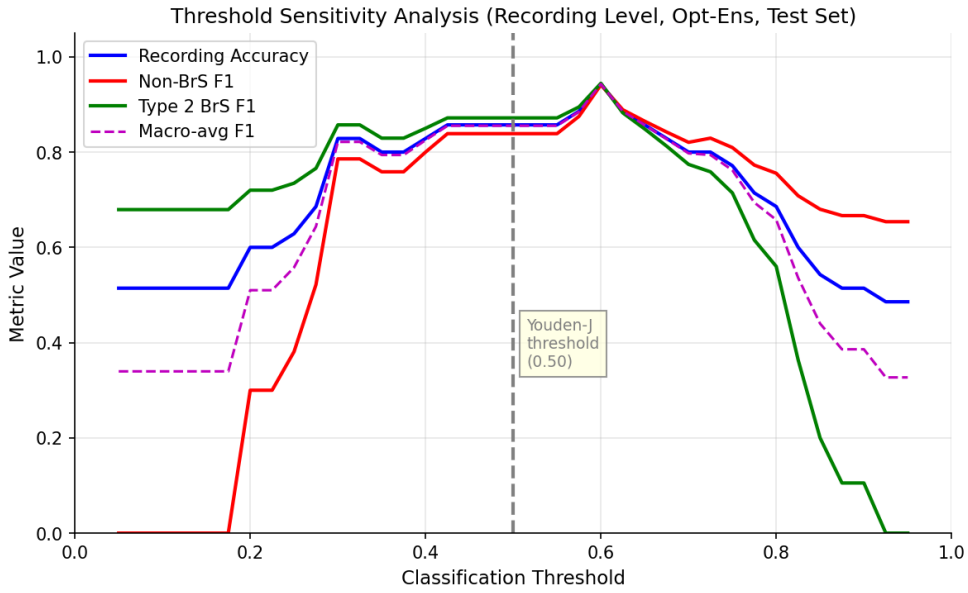

Figure 7.6: Recording-level metric values as a function of the classification threshold for the Opt-Ens on the production test set ($n = 35$). Four metrics are shown: Recording Accuracy (blue), Non-BrS F1 (red), Type 2 BrS F1 (green), and macro-average F1 (purple dashed). The vertical grey dashed line indicates the Youden-J threshold selected on the validation set. The selected threshold lies near the joint optimum across all metrics, validating the calibration procedure described in Section 6.5.

7.5 Feature Importance Analysis (SHAP)

SHAP (SHapley Additive exPlanations) values [18] were computed for the Random Forest component to quantify the marginal contribution of each of the 44 hand-engineered features to the model's output.

7.5.1 SHAP Importance Rankings

Figure 7.7 shows the SHAP importance bar chart. The top five features by mean absolute SHAP value are: `s_in_v6` (0.607), `max_dep_ii` (0.565), `max_dep_avf` (0.431), `st_j_v2` (0.427), and `tpeak_tend` (0.424).

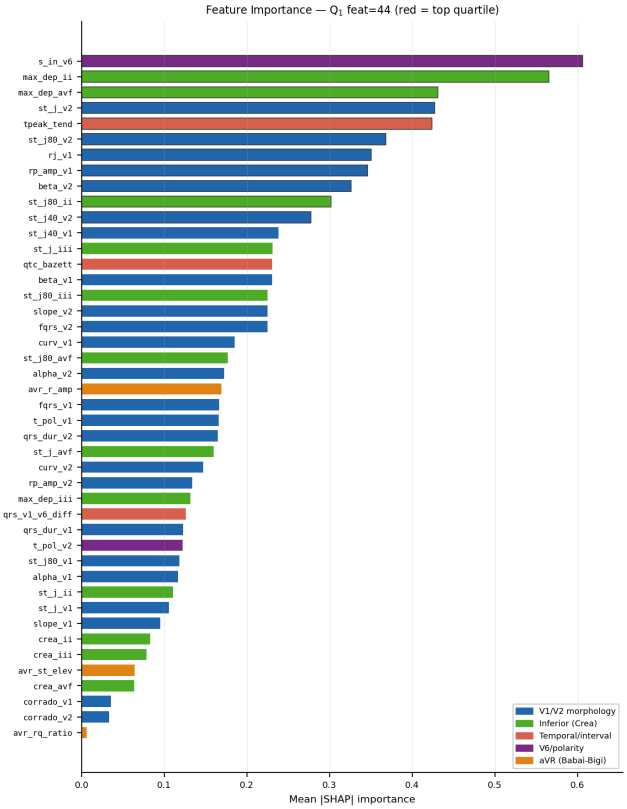

Figure 7.7: Mean absolute SHAP importance for all 44 hand-engineered features, sorted by importance descending (Random Forest component, production training set). Features are colour-coded by block: V1/V2 morphology (blue), inferior-lead features (green), temporal/interval features (orange), V6/polarity features (purple), and aVR features (red). The top five features (`s_in_v6`, `max_dep_ii`, `max_dep_avf`, `st_j_v2`, `tpeak_tend`) account for the greatest share of discriminative signal.

Table 7.5 presents the full SHAP ranking with clinical interpretations.

Table 7.5: Full SHAP feature importance ranking for the Random Forest component (production training set). Importance scores are mean absolute SHAP values. Block abbreviations: V1/V2 = right-precordial morphology; Inf = inferior leads (II, III, aVF); Temp = temporal/interval; V6 = lateral morphology; aVR = right-arm lead.

Rank	Feature	Block	SHAP	Clinical interpretation
1	`s_in_v6`	V6	0.607	S-wave depth in V6; absent/shallow in T2-BrS vs. deep in RBBB/IRBBB [4]
2	`max_dep_ii`	Inf	0.565	Max ST depression in lead II; Crea sign component [8]
3	`max_dep_avf`	Inf	0.431	Max ST depression in aVF; inferior Crea sign [8]
4	`st_j_v2`	V1/V2	0.427	ST elevation at J-point in V2 (mV); core T2-BrS saddleback criterion [10]
5	`tpeak_tend`	Temp	0.424	T-peak to T-end interval (ms); marker of transmural dispersion of repolarisation [9]
6	`st_j80_v2`	V1/V2	0.368	ST elevation 80 ms after J-point in V2; discriminates saddleback from IRBBB plateau [7]
7	`rj_v1`	Temp	0.351	r-J interval in V1 (ms); prolonged in T2-BrS; top SHAP feature in Ishida et al. [9]
8	`rp_amp_v1`	V1/V2	0.346	r'-peak amplitude in V1 (mV); elevated in T2-BrS saddleback [7]
9	`beta_v2`	V1/V2	0.326	β angle in V2 (degrees); de Luna ascending-limb criterion [7]

Continued on next page

Table 7.5 continued from previous page

Rank	Feature	Block	SHAP	Clinical interpretation
10	st_j80_ii	Inf	0.302	ST at 80 ms post-J in lead II; prolonged inferior depression [8]
11	st_j40_v2	V1/V2	0.278	ST elevation 40 ms after J-point in V2; early saddleback rise [7]
12	st_j40_v1	V1/V2	0.238	ST elevation 40 ms after J-point in V1 [10]
13	st_j_iii	Inf	0.231	ST elevation/depression at J-point in lead III [8]
14	qtc_bazett	Temp	0.231	Corrected QT interval (Bazett); prolonged QTc associated with BrS risk [1]
15	beta_v1	V1/V2	0.231	β angle in V1 (degrees) [7]
16	st_j80_iii	Inf	0.225	ST at 80 ms post-J in lead III [8]
17	slope_v2	V1/V2	0.225	ST-segment slope in V2 (mV/sample); positive in saddleback [7]
18	fqrs_v2	V1/V2	0.225	Fragmented QRS flag in V2; associated with fibrosis and BrS substrate [1]
19	curv_v1	V1/V2	0.185	ST curvature index in V1; Corrado-style concavity measure [11]
20	st_j80_avf	Inf	0.177	ST at 80 ms post-J in aVF [8]
21	alpha_v2	V1/V2	0.172	α angle in V2 (degrees); descending-limb criterion [7]
22	avr_r_amp	aVR	0.169	R-wave amplitude in aVR (≥ 0.3 mV positive criterion) [1]
23	fqrs_v1	V1/V2	0.166	Fragmented QRS flag in V1 [1]

Continued on next page

Table 7.5 continued from previous page

Rank	Feature	Block	SHAP	Clinical interpretation
24	t_pol_v1	V6	0.166	T-wave polarity in V1 (positive/negative/biphasic) [10]
25	qrs_dur_v2	Temp	0.165	QRS duration in V2 (ms); prolonged in conduction abnormalities [4]
26	st_j_avf	Inf	0.160	ST elevation at J-point in aVF [8]
27	curv_v2	V1/V2	0.147	ST curvature index in V2 [11]
28	rp_amp_v2	V1/V2	0.134	r'-peak amplitude in V2 (mV) [7]
29	max_dep_iii	Inf	0.132	Max ST depression in lead III [8]
30	qrs_v1_v6_diff	V6	0.126	QRS duration difference V1−V6 (ms); IRBBB discriminator [4]
31	qrs_dur_v1	Temp	0.123	QRS duration in V1 (ms) [4]
32	t_pol_v2	V6	0.122	T-wave polarity in V2 [10]
33	st_j80_v1	V1/V2	0.118	ST at 80 ms post-J in V1 [10]
34	alpha_v1	V1/V2	0.117	α angle in V1 (degrees) [7]
35	st_j_ii	Inf	0.111	ST elevation at J-point in lead II [8]
36	st_j_v1	V1/V2	0.106	ST elevation at J-point in V1 (mV) [10]
37	slope_v1	V1/V2	0.095	ST-segment slope in V1 [7]
38	crea_ii	Inf	0.083	Crea sign flag in lead II ($\geq$ 0.1 mV for $\geq$ 80 ms) [8]
39	crea_iii	Inf	0.079	Crea sign flag in lead III [8]

Continued on next page

Table 7.5 continued from previous page

Rank	Feature	Block	SHAP	Clinical interpretation
40	`avr_st_elev`	aVR	0.064	ST elevation flag in aVR (> 0 mV) [1]
41	`crea_avf`	Inf	0.064	Crea sign flag in aVF [8]
42	`corrado_v1`	V1/V2	0.036	Corrado curvature index in V1 [11]
43	`corrado_v2`	V1/V2	0.034	Corrado curvature index in V2 [11]
44	`avr_rq_ratio`	aVR	0.006	R/Q amplitude ratio in aVR [1]

7.5.2 Clinical Interpretation of Top SHAP Features

S-wave in V6 (`s_in_v6`, rank 1). The S-wave depth in V6 is the single most discriminative feature. Type 2 BrS exhibits a shallow or absent S-wave in V6, whereas RBBB and IRBBB (the primary differentials) produce a prominent lateral S-wave [4, 1]. This finding suggests that lateral-lead morphology carries more discriminant information than the right-precordial features historically dominant in Type 2 criteria, consistent with Ohkubo et al. [4].

Inferior ST depression (ranks 2–3). Both `max_dep_ii` and `max_dep_avf` measure inferior ST depression, underpinning the Crea sign [8]. Their high rank relative to the binary Crea flags (ranks 38–41) confirms that the *magnitude* of inferior depression carries more information than the threshold crossing, validating the Crea sign family as a primary feature block.

V2 ST elevation and r-J interval (ranks 4, 7). The J-point ST elevation in V2 and the r-J interval in V1 are consistent with findings from Ishida et al. [9], who identified these features as top SHAP predictors in an independent BrS risk-stratification model, providing external validity for the present importance hierarchy.

T-peak-to-T-end interval (rank 5). `tpeak_tend` is a surrogate for transmural dispersion of repolarisation [1, 9], indicating that the model captures repolarisation heterogeneity as a discriminative signal complementing the depolarisation-phase features.

Low-SHAP features. Corrado indices (ranks 42–43) and aVR R/Q ratio (rank 44) show near-zero marginal importance yet cannot be removed without validation regression, confirming they contribute via non-linear MLP interactions [18].

7.6 Attention and Interpretability Analysis

7.6.1 Attention Entropy Analysis

The additive attention mechanism in Stream A (Section 6.3.4) assigns weight α_t to each of the 53 Transformer output positions. Attention entropy was computed across all validation windows:

$$H(\boldsymbol{\alpha}) = -\sum_{t=1}^{53} \alpha_t \ln \alpha_t, \tag{7.1}$$

with maximum $H_{\max} = \ln(53) = 3.970$ nats corresponding to the uniform distribution.

Figure 7.8 shows that both Non-BrS and Type 2 BrS classes produce $H = 3.952$ nats, within $\Delta = 0.018$ nats ($< 0.5\%$) of the uniform ceiling. This result is stable across all five CV folds (range 3.953–3.969 nats) and holds throughout training (entropy rises from approximately 3.913 at epoch 1 to 3.964 at convergence, confirming progressive flattening rather than sharpening).

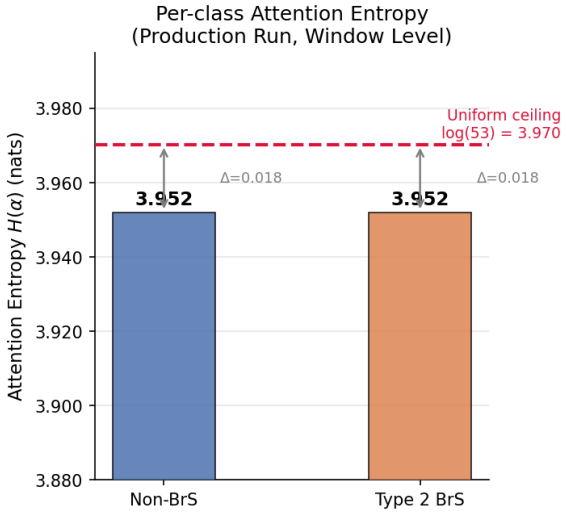

Figure 7.8: Mean attention entropy $H(\boldsymbol{\alpha})$ per class at the production checkpoint (val set, window level). Both classes show near-uniform weights within $\Delta = 0.018$ nats of the theoretical maximum $H_{\max} = \ln(53) = 3.970$ nats (red dashed line). The double-headed arrows indicate the distance from the uniform ceiling for each class. The near-identical entropy values for Non-BrS and Type 2 BrS confirm that Stream A is not differentially weighting temporal positions by class.

Stream A therefore functions as a *mean-pool encoder*: the attention mechanism does not localise to the diagnostically relevant ST region, attributed to the limited

training-set size (≈1,400 windows) providing insufficient gradient signal to sharpen attention weights [15]. Stream A still contributes useful regularised temporal features to the fusion head, but the primary discriminative signal is carried by Stream B and the RF layer.

7.6.2 Attention Weight Visualisation

Figure 7.9 shows representative attention weight profiles; the near-flat colour scale confirms the entropy analysis: no temporal region is selectively emphasised, and clinical interpretability therefore rests on the SHAP analysis of Stream B (Section 7.5).

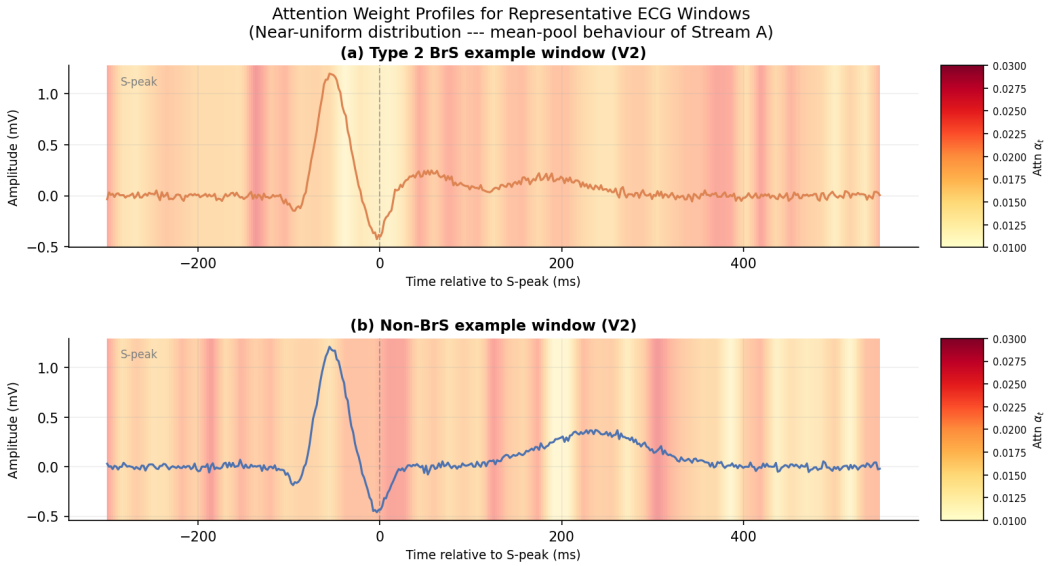

Figure 7.9: Attention weight profiles for representative ECG windows (V2 lead, 500 Hz, 850 ms centred on the S-peak). **(a)** Type 2 BrS example with visible saddleback ST elevation and r'-wave; **(b)** Non-BrS example with normal ST morphology. Background shading encodes the interpolated attention weight α_t (colour scale: light = low, dark/orange = high). The near-uniform colour distribution in both panels confirms that the attention mechanism does not selectively emphasise the ST segment or any other temporal region. This is consistent with the mean-pool behaviour quantified in Figure 7.8 and is attributed to the limited training set size.

7.7 Error Analysis

The Opt-Ens makes 5 recording-level errors: **4 false positives** (Non-BrS classified as Type 2; FP rate 23.5%) and **1 false negative** (Type 2 classified as Non-BrS; FN rate 5.6%). Table 7.6 summarises these errors with their clinical consequences and likely contributing factors.

Table 7.6: Error summary for the Opt-Ens on the production test set ($n = 35$ recordings). "Clinical consequence" refers to the downstream effect in a screening application. "Contributing factor" is inferred from the pipeline architecture and dataset characteristics; individual clinical records were not available for post-hoc review.

Error type	Count	Clinical consequence	Contributing factor (likely)
False positive (Non-BrS → Type 2)	4	Over-referral: further specialist evaluation or sodium-channel blocker provocation test; conservative, not dangerous [1]	Borderline saddleback morphology; IRBBB-like pattern; residual ECGD digitisation artefact in Non-BrS recordings
False negative (Type 2 → Non-BrS)	1	Missed diagnosis: potential withholding of risk stratification in a patient at risk of SCD [1, 2]	Attenuated saddleback morphology with ST elevation below threshold; low r'-amplitude in V1/V2
Total errors	**5**		

False positives. The four FP recordings are Non-BrS ECGs with borderline saddleback-like morphology overlapping the Type 2 distribution in top SHAP features; ECGD digitisation artefacts in Non-BrS recordings are a secondary contributor. Clinically, a false positive triggers specialist evaluation rather than a harmful outcome [2].

False negative and asymmetric error cost. The single FN is attributed to attenuated saddleback morphology: ST elevation in Type 2 BrS varies dynamically [1], and the top SHAP features may take Non-BrS-like values in a partially concealed phenotype. The small test set ($n = 18$ Type 2 recordings) limits sensitivity precision (95% CI on recall 0.944: ≈0.72–0.99 [6]); larger datasets remain the most impactful avenue for improvement.

Chapter 8

Discussion

This chapter interprets the results of Chapter 7 in the broader context of the Type 2 BrS classification literature. Section 8.1 restates the headline findings. Section 8.2 explains the RF component's dominant role. Section 8.3 analyses the clinical significance of the SHAP feature hierarchy. Section 8.4 frames the Non-BrS recall limitation as a dataset-level constraint. Section 8.5 discusses the near-uniform attention behaviour. Section 8.6 situates the work relative to prior studies. Sections 8.7 and 8.8 enumerate limitations and proposed future directions.

8.1 Summary of Findings

The pipeline classifies Type 2 BrS against Non-BrS ECGs at recording level with accuracy **0.857**, Non-BrS F1 **0.839**, Type 2 F1 **0.872**, and ROC-AUC **0.938** on a 35-recording patient-disjoint test set. The optimised ensemble (Opt-Ens, $w^* = 0.30$) confirms that the RF component carries the dominant discriminative signal at the current dataset scale, whilst the DL component provides complementary temporal context that marginally improves Type 2 F1 over the RF-only baseline. The top five SHAP features (`s_in_v6`, `max_dep_ii`, `max_dep_avf`, `st_j_v2`, and `tpeak_tend`) span lateral, inferior, right-precordial, and temporal domains. The attention mechanism in Stream A exhibits near-uniform entropy ($H \approx 3.952$ nats vs. ceiling 3.970 nats), indicating mean-pool behaviour attributable to limited training-set size. Together, these findings demonstrate that automated Type 2 BrS classification at a clinically relevant discriminability level is feasible, whilst identifying dataset scale as the primary constraint on further improvement [1, 3].

8.2 The Dominance of Hand-Engineered Features at Small Dataset Scales

The RF component's dominant role ($w^* = 0.30$, RF contributing 70% of the ensemble signal) is theoretically expected at the current dataset scale. The diagnostic criteria for Type 2 BrS are precisely specified (J-point ST elevation, α/β angular criteria [7], inferior ST depression (Crea sign) [8], S-wave absence in V6 [4], and r-J interval prolongation [9]), and a Random Forest with 44 features encoding these thresholds exploits them directly, without inferring them from raw signal patterns. A CNN operating on $\approx$ 1,400 training windows must discover the same thresholds from scratch, a data-hungry task unlikely to succeed reliably at this scale [6]. The design validation trajectory confirms this: the RF-only configuration equalled or exceeded the DL-only configuration at every development stage; only the Opt-Ens, combining both at $w^* = 0.30$, achieved a Type 2 F1 improvement over the RF alone, confirming that the DL component adds genuine complementary value.

This finding is consistent with the clinical ML literature: for small, well-characterised datasets with domain-knowledge features, tree-based ensembles frequently outperform end-to-end deep learning [6]. As annotated Type 2 BrS datasets grow through multi-centre collaboration, the relative DL contribution is expected to increase: the BrugadaTwoStream architecture is designed to scale in this direction, with frozen PTB-XL encoder layers that can be progressively unfrozen as training data increases.

8.3 Clinical Relevance of the Feature Hierarchy

S-wave in V6 (rank 1). `s_in_v6` substantially outranks the right-precordial features historically central to Type 2 criteria [10, 7]. In IRBBB, delayed right-ventricular activation creates a prominent lateral S-wave; Type 2 BrS does not, as conduction delay is confined to the RVOT [4]. This reflects a genuine physiological discriminant that is not prominently codified in current guidelines, suggesting lateral-lead features deserve greater formal weight.

Inferior ST depression (ranks 2–3). The continuous `max_dep_ii/avf` substantially outrank the binary Crea flags (ranks 38–41), confirming that the *magnitude* of inferior depression carries more information than the threshold crossing [8].

Temporal interval features (ranks 5, 7). `tpeak_tend` and `rj_v1` independently corroborate Ishida et al. [9], providing cross-study validation for repolarisation-dispersion markers.

Implications for clinical screening. The SHAP pruning experiment confirms that the full 44-feature set is required for optimal performance, because low-ranked

features contribute via non-linear MLP interactions invisible to marginal SHAP [18].

8.4 Dataset Size as the Primary Performance Constraint

8.4.1 Statistical Uncertainty and the Labelling Challenge

The 35-recording test set yields wide Clopper–Pearson confidence intervals (Non-BrS recall 0.765: ≈0.50–0.92; Type 2 recall 0.944: ≈0.72–0.99; accuracy 0.857: ≈0.70–0.94); all headline metrics should be interpreted as indicative rather than definitive [6, 3].

The scarcity of annotated Type 2 BrS data is structural: unlike Type 1, whose coved pattern is self-diagnosing, Type 2 requires supporting clinical evidence (provocation test, family history, arrhythmia documentation) [1, 2], making correct labelling resource-intensive.

8.4.2 Recommended Next Steps for Dataset Expansion

Multi-centre data collection involving independent ECG acquisition systems, patient populations, and cardiologist reviewers would simultaneously address the small-test-set, single-institution, and external-validation limitations. Even 100–200 new Type 2 recordings from a second centre would enable the first genuine out-of-distribution evaluation [1, 2]. **Registry linkage** to a prospective BrS registry with provocation test documentation would provide a continuous stream of correctly labelled recordings. **Standardised digital acquisition** for both classes (currently heterogeneous: paper-digitised Non-BrS vs. GE MAC2000 XML Type 2) would eliminate the digitisation-quality confound and directly reduce the false positive rate.

8.5 Architectural Behaviour: From Mean-Pool to Discriminative Attention

The near-uniform attention entropy ($H = 3.952$ nats; $\Delta = 0.018$ nats from the maximum) is not an architectural failure but an expected consequence of training-data volume. Learning a selective, temporally peaked attention distribution requires the optimiser to identify a *consistent* discriminant locus across hundreds of examples. With $\approx$ 1,400–1,600 training windows, per-position gradient estimates are noisy and attention parameters remain near their PTB-XL-pretrained near-uniform initialisation [15, 16]. The entropy trajectory, rising from 3.913 at epoch 1 to 3.964

at convergence, confirms progressive flattening rather than sharpening, consistent with the softmax converging to the maximum-entropy solution under gradient noise.

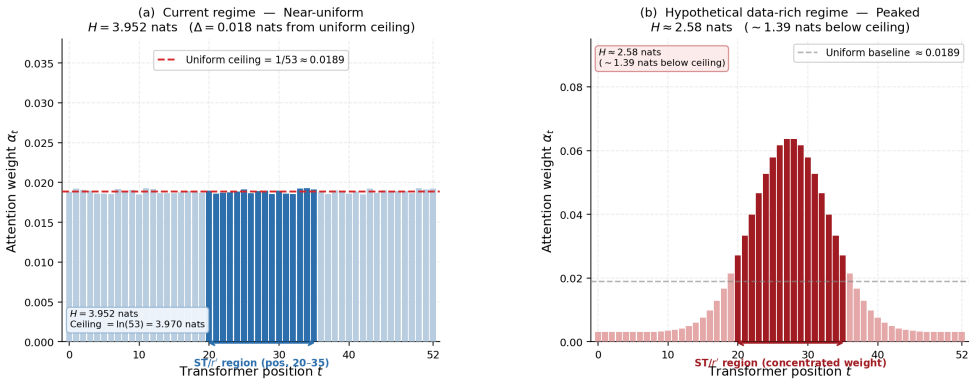

Figure 8.1: Schematic of attention weight evolution with dataset scale. **(a)** Current regime ($n \approx 1{,}400$ training windows): near-uniform weights across all 53 Transformer positions; entropy $H \approx 3.952$ nats, within $\Delta = 0.018$ nats of the uniform ceiling. The diagnostically relevant ST/r'-wave region (positions 20–35, blue bars) receives no more weight than surrounding positions. **(b)** Hypothetical data-rich regime ($n \gg 1{,}400$ windows): a peaked distribution concentrated on the ST/r'-wave region, with substantially lower entropy ($H \approx 2.58$ nats), would emerge if gradient signals from many more training examples reliably identified the ST segment as the primary discriminant. The architecture has the capacity for this behaviour; data volume is the limiting factor.

The architecture is nonetheless capable of temporal localisation: attention-based encoders routinely develop peaked distributions on sufficiently large datasets [15], and a larger training corpus would be expected to concentrate attention on the diagnostically relevant ST-segment region (positions ≈20–35, 80–300 ms post-S-peak; Figure 8.1b). In the current regime, the mean-pool fallback is not harmful: Stream A still aggregates full temporal context and the pipeline achieves AUC = 0.938, but dataset expansion remains the primary route to achieving discriminative attention and improved DL interpretability.

8.6 Comparison with Related Work

Table 8.1 summarises the most closely related prior studies alongside the present work.

Table 8.1: Comparison of BrugadaTwoStream with related work on Brugada Syndrome ECG classification. †Multi-centre dataset; all BrS phenotypes included. ‡Arrhythmia-event prediction, not ECG classification. AUC not reported (—) where the study used non-ROC metrics only.

Study	Year	*n* ECGs	Task	Method	AUC
Melo et al. [5]†	2023	>15,000	BrS detection (all types) vs. controls	Deep ResNet, 12-lead	0.90
Randazzo et al. [16]	2025	∼500	Type 1 / Type 2 / Non-BrS 3-class	CNN + Transformer + attention	0.91
Ishida et al. [9]‡	2025	415	Arrhythmia event risk (BrS cohort)	RF + SHAP (tabular features)	0.81
Saleh et al. [6]	2026	∼400	BrS vs. controls (limited dataset)	PTB-XL transfer learning DL	0.87
This thesis	**2025**	**362**	**T2-BrS vs. Non-BrS (binary)**	**Dual-stream DL + RF (Opt-Ens)**	**0.938**

Melo et al. 2023 [5] achieved AUC 0.90 on a >15,000-ECG dataset covering all BrS phenotypes; however, Type 1 coved cases dominate the positive class, and the harder Type 2 vs. Non-BrS sub-problem is not separately evaluated. **Randazzo et al. 2025** [16] proposed the architecturally closest prior work (three-class CNN+Transformer, AUC ≈0.91); the present work adds dual-stream fusion with 44 hand-engineered features, a validation-optimised RF ensemble, and SHAP interpretability traceable to clinical criteria. Cross-study concordance with Ishida et al. [9], both independently identifying r-J interval and Tpeak–Tend as top SHAP features, provides external validation. **Unique contributions.** This thesis is, to the author's knowledge, the first automated pipeline for binary Type 2 BrS vs. Non-BrS that combines: (i) an original dual-modality dataset; (ii) a 44-feature clinical framework grounded exhaustively in published criteria; (iii) a dual-stream DL architecture with PTB-XL transfer learning; and (iv) a validation-optimised RF ensemble achieving recording-level AUC 0.938.

8.7 Limitations

Small test set and metric uncertainty. The 35-recording test set yields 95% confidence intervals spanning ±0.12–0.22 in absolute terms; reported metric values should be interpreted as indicative rather than definitive [6, 3].

Single-institution dataset. All recordings originate from a single Sardinian

centre using predominantly the GE MAC2000. Generalisability to different ECG machines, patient populations, or lead-placement conventions is unknown; institution-specific confounders cannot be excluded. External validation on an independent cohort is required before clinical deployment [1].

Heterogeneous acquisition modalities. Non-BrS recordings are digitised from paper via ECGD; Type 2 BrS recordings are natively digital GE MAC2000 XML exports. Systematic differences in signal quality (grid-removal artefacts, baseline wander, variable resolution) are a likely contributor to the $4/17 = 23.5\%$ false positive rate. A dataset with both classes acquired natively on the same system would eliminate this confound.

Opaque DL representations. The near-uniform attention behaviour means clinical interpretability of the pipeline rests entirely on the SHAP analysis of Stream B and the RF. The DL contribution (+0.015 in Type 2 F1 over RF alone) is empirically beneficial but opaque to clinical inspection [3, 18].

No external validation. The pipeline has been evaluated exclusively within-distribution (same centre, same ECG machine, same time period as training). Patient-level disjointness guards against data leakage, but genuine out-of-distribution performance is unknown.

8.8 Future Work

Multi-centre data collection and external validation. Even 100–200 new Type 2 recordings from a second institution would enable the first genuine out-of-distribution evaluation and substantially increase statistical power [1, 2].

Inferior-lead raw signal branch. The superior SHAP ranking of inferior-lead features motivates adding a parallel shallow convolutional branch for leads II, III, aVF at the fusion layer, giving Stream A direct access to inferior ST dynamics without modifying the frozen PTB-XL encoder.

Self-supervised ECG pretraining. Masked signal prediction or contrastive learning [14] on a large unlabelled corpus enriched with IRBBB and early repolarisation variants would provide a more principled initialisation than supervised RBBB classification.

Three-class extension and prospective evaluation. A three-class extension (Type 1/Type 2/Non-BrS) and a prospective head-to-head evaluation against cardiologist decisions would validate screening utility and inform deployment threshold selection [16, 3].

Chapter 9

Conclusion

This chapter consolidates the principal findings of the thesis, enumerates the original contributions, and closes with a perspective on the significance and future trajectory of automated Type 2 Brugada Syndrome classification from 12-lead ECGs.

9.1 Work Undertaken

This thesis addressed automated binary classification of Type 2 Brugada Syndrome (T2-BrS) against Non-BrS ECGs, a clinically important but under-studied task, since the Type 2 saddleback pattern is the most prevalent BrS ECG pattern yet not independently diagnostic without pharmacological provocation or specialist interpretation [1, 2].

The work comprised three layers. **Layer 1 (data acquisition):** ≈1,000 paper ECGs were digitised via ECGD (yielding 174 Non-BrS recordings), and 188 confirmed Type 2 BrS recordings were extracted from GE MAC2000 XML archives; after 500 Hz resampling, S-peak-centred windowing, and multi-lead tensor construction, the 362-recording dataset was split 80/10/10. **Layer 2 (feature engineering and DL):** A 44-feature clinical vector [7, 11, 8, 13, 9] and a CNN–Transformer dual-stream network (BrugadaTwoStream) were developed; Stream A was PTB-XL-pretrained (RBBB, val AUC 0.9919 [17]). **Layer 3 (decision aggregation):** A Random Forest [21] was ensembled with the DL output ($w^* = 0.30$), followed by recording-level mean aggregation, Youden's-J thresholding, and SHAP interpretability [18].

Figure 9.1 illustrates the complete three-layer architecture.

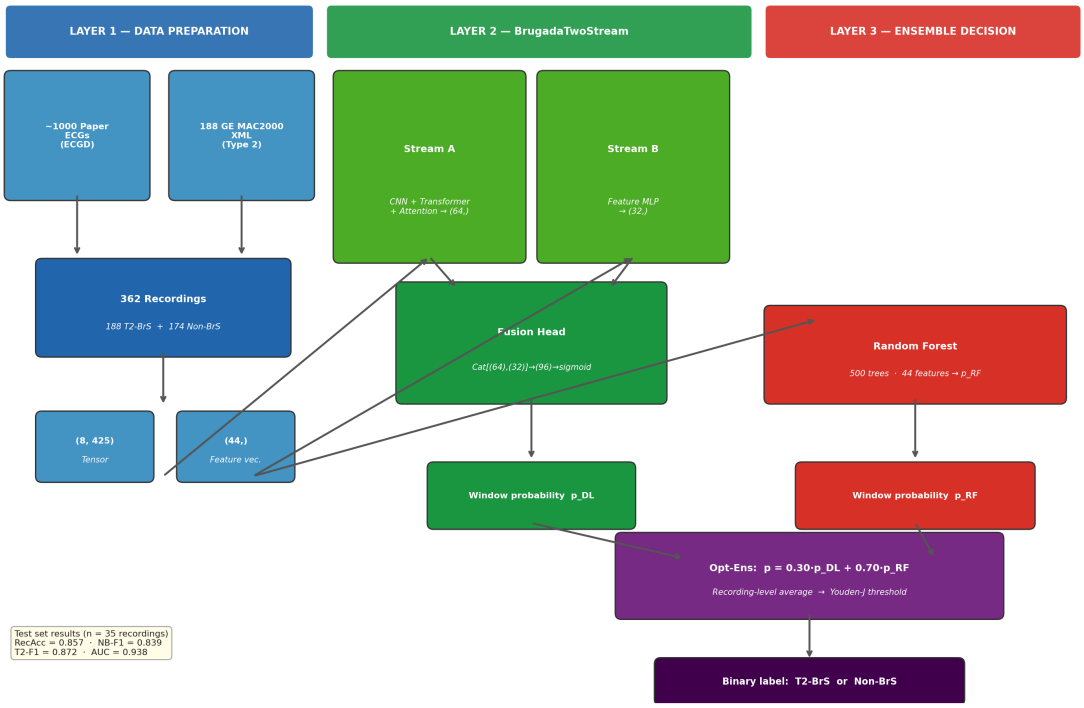

Figure 9.1: End-to-end pipeline summary. **Layer 1** (blue): data preparation from two acquisition sources (ECGD-digitised paper ECGs and GE MAC2000 XML exports), yielding 362 labelled recordings, an 8-channel signal tensor of shape $(8, 425)$, and a 44-dimensional feature vector per window. **Layer 2** (green): BrugadaTwoStream dual-stream model combining a CNN–Transformer encoder (Stream A, context dimension 64) and a feature MLP (Stream B, embedding dimension 32), merged at the fusion head to produce a window-level probability p_{DL}. **Layer 3** (red/purple): RF classifier producing p_{RF}, combined via the Opt-Ens weighted average ($p = 0.30\,p_{\mathrm{DL}} + 0.70\,p_{\mathrm{RF}}$), recording-level averaging, and Youden's-J thresholding to yield a single binary label per ECG recording. Production test set results ($n = 35$) are annotated.

9.2 Key Results and Findings

On the held-out production test set of 35 patient-disjoint recordings (Non-BrS = 17, Type 2 BrS = 18), the Opt-Ens achieved recording accuracy **0.857**, Non-BrS F1 **0.839** (precision 0.929, recall 0.765), Type 2 BrS F1 **0.872** (precision 0.810, recall 0.944), and recording-level ROC-AUC **0.938**. The optimal DL weight $w^* = 0.30$ confirms that the RF carries the dominant discriminative signal at the current dataset scale, consistent with the theoretical expectation that hand-engineered features grounded in clinical thresholds outperform raw-signal deep learning on

small datasets [6].

SHAP analysis identified the five most important features as `s_in_v6` (0.607), `max_dep_ii` (0.565), `max_dep_avf` (0.431), `st_j_v2` (0.427), and `tpeak_tend` (0.424). The strongest discriminator (S-wave absence in V6) distinguishes Type 2 BrS from RBBB mimics [4]; the second and third ranked features validate the Crea inferior-lead sign [8]; and the appearance of T_{peak}–T_{end} (rank 5) and r-J interval V1 (rank 7) corroborates the independent findings of Ishida et al. [9].

The attention entropy of Stream A was near-uniform in both classes ($H \approx 3.952$ nats vs. ceiling $\ln(53) = 3.970$ nats), indicating mean-pool behaviour attributed to limited training-set size rather than an architectural deficiency [15, 3].

9.3 Original Contributions

Contribution 1: Original labelled dataset. 362 12-lead ECG recordings (188 Type 2 BrS, 174 Non-BrS) assembled from paper ECG digitisation and GE MAC2000 XML extraction, with a patient-manifest-based protocol ensuring patient-level disjointness.

Contribution 2: 44-feature clinical feature engineering framework. A systematic framework grounded in five published Type 2 BrS criteria [7, 11, 8, 13, 9], with documented NaN rates, clinical sources, and SHAP interpretability.

Contribution 3: BrugadaTwoStream architecture with PTB-XL transfer learning. A dual-stream CNN–Transformer + feature-MLP architecture with PTB-XL RBBB pretraining (val AUC 0.9919), the single most impactful architectural decision, providing a reusable pretraining protocol for BrS classification on small datasets [17].

Contribution 4: Empirical evidence for non-linear feature interaction value. SHAP pruning experiments demonstrate that low-SHAP features contribute through non-linear interactions, challenging the common practice of marginal-SHAP-based feature elimination in clinical ML pipelines.

Contribution 5: Recording-level aggregation and RF-dominant ensemble. The val-set-optimised Opt-Ens achieves a 10.7-pp gain in recording accuracy and 10.2-pp gain in ROC-AUC over the RF-only baseline (0.857 vs. 0.750; 0.938 vs. 0.836), demonstrating genuine complementarity between the DL and RF components.

9.4 Final Remarks

The results demonstrate that Type 2 BrS can be automatically distinguished from Non-BrS with clinically meaningful accuracy even from a single institutional dataset of 292 training recordings. The AUC of 0.938 and Type 2 recall of 0.944 (17/18 test recordings correctly identified) are encouraging, and the pipeline's conservative error profile (high Type 2 recall at the cost of some Non-BrS over-prediction) is clinically appropriate for a screening application where missed diagnoses carry greater risk than over-referral [1, 2].

The SHAP analysis provides genuine clinical insight: S-wave absence in V6, inferior ST depression, and V2 J-point ST elevation as dominant discriminators suggests automated analysis surfaces features under-utilised in current manual protocols, and a decision-support tool presenting top-5 SHAP features alongside the prediction would meet explainability requirements for clinical AI deployment [3, 18].

The primary limitation is dataset scale. With 35 test recordings, confidence intervals on key metrics span approximately ± 0.15 in absolute terms, and results should be interpreted as indicative rather than definitive. Prospective multi-centre data collection, connecting the pipeline to a BrS patient registry with systematic digital ECG acquisition at multiple institutions, is the highest-priority next step, simultaneously addressing the scale, generalisability, and external validation gaps identified in Chapter 8.

Bibliography

[1] B. Narasimhan, J. Na, M. M. Monasky, R. Brugada, Y. Miyasaka, J. Brugada, P. Brugada, and C. Krittanawong. «Brugada syndrome». In: *Nat Rev Dis Primers* 11.1 (2025), p. 38. DOI: 10.1038/s41572-025-00622-5 (cit. on pp. 1, 2, 5–7, 26, 27, 31, 33, 34, 36, 43, 53, 59, 65, 67, 70, 71, 73, 76, 77, 80).

[2] K. Y. Kan et al. «Beyond the type 1 pattern: comprehensive risk stratification in Brugada syndrome». In: *Journal of Interventional Cardiac Electrophysiology* 68 (2025), pp. 1771–1790. DOI: 10.1007/s10840-025-02101-z (cit. on pp. 1, 5–7, 32, 35, 59, 70, 73, 76, 77, 80).

[3] M. Sahebnasagh and M. H. Farjoo. «Artificial intelligence for Brugada syndrome diagnosis and gene variants interpretation». In: *Am J Cardiovasc Dis* 15.1 (Feb. 2025), pp. 1–12. DOI: 10.62347/YQHQ1079 (cit. on pp. 1, 2, 5, 10, 30, 36, 42, 43, 50, 60, 71, 73, 75, 76, 79, 80).

[4] K. Ohkubo et al. «A new criteria differentiating type 2 and 3 Brugada patterns from ordinary incomplete right bundle branch block». In: *Int Heart J* 52.3 (May 2011), pp. 159–163 (cit. on pp. 2, 10, 26, 30, 32–34, 43, 64, 66, 67, 72, 79).

[5] L. Melo, G. Ciconte, A. Christy, G. Vicedomini, L. Anastasia, C. Pappone, and E. Grant. «Deep learning unmasks the ECG signature of Brugada syndrome». In: *PNAS Nexus* 2.11 (Oct. 2023), pgad327. DOI: 10.1093/pnasnexus/pgad327 (cit. on pp. 2, 10, 48, 49, 60, 75).

[6] K. Saleh et al. «AI-ECG classification for Brugada syndrome: A study of machine learning techniques to optimise for limited datasets». In: *PLOS Digit Health* 5.2 (Feb. 2026), e0001222. DOI: 10.1371/journal.pdig.0001222 (cit. on pp. 3, 9, 23, 28, 30, 36, 42, 43, 48, 70, 72, 73, 75, 79).

[7] A. B. de Luna, J. García-Niebla, and A. Baranchuk. «New electrocardiographic features in Brugada syndrome». In: *Curr Cardiol Rev* 10.3 (2014), pp. 175–180 (cit. on pp. 4, 6–8, 25–27, 30, 31, 33, 34, 53, 64–66, 72, 77, 79).

[8] P. Crea, G. Picciolo, F. Luzza, and G. Oreto. «ST segment depression in the inferior leads in Brugada pattern: a new sign». In: *Ann Noninvasive Electrocardiol* 20.6 (Nov. 2015), pp. 561–565. DOI: 10.1111/anec.12247 (cit. on pp. 4, 8, 27, 30, 32, 35, 39, 40, 64–67, 72, 77, 79).

[9] S. Ishida et al. «Risk stratification of major arrhythmia events in Japanese patients with Brugada syndrome using machine learning models». In: *Heart Rhythm O2* 6.7 (July 2025), pp. 987–994. DOI: 10.1016/j.hroo.2025.04.009 (cit. on pp. 4, 8, 10, 32, 34, 35, 39, 50, 64, 67, 72, 75, 77, 79).

[10] P. Brugada and J. Brugada. «Right bundle branch block, persistent ST segment elevation and sudden cardiac death: a distinct clinical and electrocardiographic syndrome». In: *J Am Coll Cardiol* 20.6 (1992), pp. 1391–1396. DOI: 10.1016/0735-1097(92)90253-J (cit. on pp. 5, 31, 64–66, 72).

[11] D. Corrado, C. Basso, G. Buja, A. Nava, L. Rossi, and G. Thiene. «Right bundle branch block, right precordial ST-segment elevation, and sudden death in young people». In: *Circulation* 103.5 (2001), pp. 710–717. DOI: 10.1161/01.CIR.103.5.710 (cit. on pp. 6, 8, 30, 31, 33, 34, 39, 65–67, 77, 79).

[12] H. Morita, K. F. Kusano, D. Miura, S. Nagase, K. Nakamura, S. T. Morita, T. Ohe, D. P. Zipes, and J. Wu. «Fragmented QRS as a marker of conduction abnormality and a predictor of prognosis of Brugada syndrome». In: *Circulation* 118.17 (2008), pp. 1697–1704. DOI: 10.1161/CIRCULATIONAHA.108.770917 (cit. on pp. 8, 32, 35).

[13] M. A. Babai Bigi, A. Aslani, and S. Shahrzad. «aVR sign as a risk factor for life-threatening arrhythmic events in patients with Brugada pattern». In: *Europace* 9.11 (2007), pp. 1000–1003. DOI: 10.1093/europace/eum134 (cit. on pp. 8, 30, 33, 35, 36, 77, 79).

[14] A. H. Ribeiro et al. «Automatic diagnosis of the 12-lead ECG using a deep neural network». In: *Nat Commun* 11.1 (2020), p. 1760. DOI: 10.1038/s41467-020-15432-4 (cit. on pp. 9, 30, 43, 48, 57, 76).

[15] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin. «Attention is all you need». In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017 (cit. on pp. 9, 48, 69, 73, 74, 79).

[16] V. Randazzo, L. Marchetti, and E. Pasero. «Classification of Brugada Syndrome from 12-Lead ECG Using a Convolutional–Transformer Neural Network». In: *2025 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 2025 (cit. on pp. 9, 10, 26, 50, 54, 73, 75, 76).

[17] P. Wagner, N. Strodthoff, R.-D. Bousseljot, D. Kreiseler, F. I. Lunze, W. Samek, and T. Schaeffter. «PTB-XL, a large publicly available electrocardiography dataset». In: *Sci Data* 7.1 (2020), p. 154. DOI: 10.1038/s41597-020-0495-6 (cit. on pp. 9, 43, 57, 77, 79).

[18] S. M. Lundberg and S.-I. Lee. «A unified approach to interpreting model predictions». In: *Advances in Neural Information Processing Systems*. Vol. 30. 2017 (cit. on pp. 10, 30, 41, 57, 62, 67, 73, 76, 77, 80).

[19] P. Virtanen et al. «SciPy 1.0: Fundamental algorithms for scientific computing in Python». In: *Nat Methods* 17 (2020), pp. 261–272. DOI: 10.1038/s41592-019-0686-2 (cit. on pp. 23, 25, 26).

[20] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. H. A. Chen. «NeuroKit2: A Python toolbox for neurophysiological signal processing». In: *Behav Res Methods* 53.4 (2021), pp. 1689–1696. DOI: 10.3758/s13428-020-01516-y (cit. on pp. 23, 25).

[21] F. Pedregosa et al. «Scikit-learn: Machine learning in Python». In: *J Mach Learn Res* 12 (2011), pp. 2825–2830 (cit. on pp. 36, 50, 77).

[22] I. Loshchilov and F. Hutter. «Decoupled weight decay regularization». In: *International Conference on Learning Representations*. 2019 (cit. on pp. 44, 49).