FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Predicting Cerebral Vasospasm in ICU Patients with aSAH

Diogo Fernando Macedo Lopes



Master in Bioengineering – Biomedical Engineering

Supervisor: Prof. Nuno Feixa Rodrigues

Co-supervisor: Prof. António Coelho

September 17, 2026

© Diogo Fernando Macedo Lopes, 2026

Resumo

A hemorragia subaracnoideia aneurismática (aSAH) é uma emergência neurovascular associada a elevada mortalidade e morbilidade, sendo o vasospasmo cerebral uma complicação relevante durante a evolução clínica após a hemorragia. A previsão precoce deste evento poderá beneficiar da integração de informação clínica com dados de neuromonitorização, embora permaneça pouco claro se a informação fisiológica longitudinal fornece valor preditivo adicional relativamente às variáveis clínicas convencionais. Esta dissertação investigou a previsão do diagnóstico binário de vasospasmo cerebral atribuído na prática clínica, baseado predominantemente em vigilância sonológica por Doppler transcraniano (TCD), incluindo critérios de velocidade arterial cerebral e índice de Lindegaard.

O estudo retrospetivo, realizado num único centro, foi estruturado em quatro fases analíticas com validação interna ao nível do doente, através de 50 repetições de validação cruzada com cinco folds. A primeira fase avaliou 13 preditores clínicos de admissão ou precoces em 220 episódios de admissão correspondentes a 219 doentes, tendo o XGBoost obtido a maior AUROC média, de 0.720. As fases seguintes utilizaram uma coorte monitorizada de 45 doentes, 23 com vasospasmo e 22 sem vasospasmo, considerando as primeiras 24 horas de cada registo de monitorização elegível. A representação exclusivamente fisiológica agregada mais favorável atingiu uma AUROC média de 0.656. Quando a ordem temporal minuto a minuto foi preservada, a rede neuronal convolucional primária atingiu uma AUROC de 0.617, enquanto a análise de sensibilidade com máscaras de observação atingiu 0.654. Na fase de fusão tardia, o comparador clínico estático correspondente atingiu uma AUROC média de 0.799, enquanto a fusão primária entre informação clínica e temporal atingiu 0.760, correspondendo a uma diferença média emparelhada de -0.039.

A informação clínica de admissão e precoce constituiu, assim, a fonte de sinal preditivo mais consistente nas análises realizadas. A fisiologia das primeiras 24 horas apresentou alguma capacidade discriminativa isoladamente, mas a preservação explícita da ordem temporal minuto a minuto não proporcionou uma vantagem clara sobre a representação exclusivamente fisiológica agregada mais favorável. Nem a adição de resumos fisiológicos à informação estática nem a integração de previsões temporais por fusão tardia produziram um incremento médio positivo de AUROC relativamente aos respetivos comparadores estáticos emparelhados. Estes resultados são específicos das coortes, representações e procedimentos de modelação avaliados e não implicam ausência de relevância biológica ou clínica da monitorização fisiológica. O principal contributo deste trabalho consiste na avaliação estruturada do valor incremental da fisiologia neurocrítica para a previsão de vasospasmo, identificando como prioridades futuras a validação multicêntrica, uma caracterização temporal mais precisa do desfecho e a exploração de representações fisiológicas de maior resolução.

Palavras-chave Hemorragia subaracnoide aneurismática; Vasospasmo cerebral; Aprendizagem automática; Predição clínica; Monitorização neurocrítica; Modelação de séries temporais; Integração de dados clínicos e fisiológicos.

Abstract

Aneurysmal subarachnoid haemorrhage (aSAH) is a neurovascular emergency associated with substantial mortality and morbidity, with cerebral vasospasm representing an important complication during the clinical course following haemorrhage. Early prediction may benefit from integrating clinical information with neuromonitoring data, although it remains unclear whether longitudinal physiological information provides predictive value beyond conventional clinical variables. This dissertation investigated prediction of the clinician-assigned binary diagnosis of cerebral vasospasm, based predominantly on sonological surveillance using transcranial Doppler (TCD), including cerebral arterial velocity and Lindegaard-ratio criteria.

The single-centre retrospective study was organised into four analytical phases with patient-level internal validation using 50 repetitions of five-fold cross-validation. Phase 1 evaluated 13 admission or early clinical predictors across 220 admission episodes from 219 patients, with XGBoost achieving the highest mean AUROC of 0.720. The subsequent phases used a monitored cohort of 45 patients, including 23 with vasospasm and 22 without vasospasm, and considered the first 24 hours of each eligible monitoring record. The strongest aggregated physiology-only representation achieved a mean AUROC of 0.656. When minute-level temporal order was preserved, the primary convolutional neural network achieved an AUROC of 0.617, while the observation-mask sensitivity analysis reached 0.654. In the late-fusion analysis, the matched static clinical comparator achieved a mean AUROC of 0.799, whereas the primary static–temporal fusion achieved 0.760, corresponding to a mean paired difference of -0.039.

Admission and early clinical information therefore provided the most consistent predictive signal across the evaluated analyses. First-24-hour physiology showed some standalone discrimination, but explicit preservation of minute-level temporal order did not provide a clear advantage over the strongest aggregated physiology-only representation. Neither adding physiological summaries to static information nor integrating temporal predictions through late fusion produced a positive mean AUROC increment over their respective matched static comparators. These findings are specific to the evaluated cohorts, representations, and modelling procedures and do not imply that physiological monitoring lacks biological or clinical relevance. The principal contribution of this work is a structured assessment of the incremental predictive value of neurocritical physiology, while identifying multicentre validation, more precise outcome timing, and higher-resolution physiological representations as priorities for future investigation.

Keywords Aneurysmal subarachnoid haemorrhage; Cerebral vasospasm; Machine learning; Clinical prediction; Neurocritical monitoring; Time-series modelling; Clinical and physiological data integration.

UN Sustainable Development Goals

The United Nations Sustainable Development Goals (SDGs), established within the 2030 Agenda for Sustainable Development, provide a global framework for addressing major societal, health, technological, and environmental challenges. This dissertation was assessed in relation to the SDGs most directly connected to its objectives, methodology, and potential future impact. The strongest alignment was identified with SDG 3 (Good Health and Well-Being) and SDG 9 (Industry, Innovation and Infrastructure), particularly Targets 3.4 and 9.5.

The contribution described below is research-oriented. The dissertation does not demonstrate a reduction in morbidity or mortality, nor does it validate a deployable clinical prediction system. Instead, it evaluates the predictive information available from routinely collected clinical and neurophysiological data and develops a methodological framework for their integration and internal validation in cerebral vasospasm prediction after aSAH.

The indicators reported in Table 1 are project-level measures of the scope, methodological performance, and predictive evaluation achieved in this dissertation. They should not be interpreted as direct measurements of population-level progress towards the corresponding United Nations SDG indicators.

Table 1: Alignment of the dissertation with the United Nations Sustainable Development Goals.

SDG	Goal and specific target	Contribution	Performance indicators and metrics
3	**Good Health and Well-Being** **Target 3.4:** Reduce premature mortality from non-communicable diseases through prevention and treatment and promote health and well-being.	The dissertation investigates whether admission and early clinical information and first-24-hour neurocritical physiology can support earlier risk stratification of subsequent clinician-assigned cerebral vasospasm after aSAH. Its contribution is indirect and methodological: it evaluates which routinely collected information sources provide predictive signal and whether physiological monitoring adds dependable predictive value beyond static clinical information.	**Study population:** 220 admission episodes from 219 patients in Phase 1; 45 patients in the matched monitored cohort. **Prediction metrics:** AUROC, average precision, Brier score, sensitivity and specificity. **Results:** highest Phase 1 mean AUROC = 0.720; matched Phase 4 static AUROC = 0.799; primary fusion AUROC = 0.760, with ΔAUROC = -0.039.
9	**Industry, Innovation and Infrastructure** **Target 9.5:** Enhance scientific research and technological capabilities.	The dissertation develops an integrated biomedical data-analysis framework combining retrospective clinical information with processed ICM+ neurophysiological monitoring. The framework spans static clinical modelling, aggregated physiological representations, minute-level temporal modelling, and static–temporal late fusion, while maintaining patient-level validation and explicit temporal eligibility.	**Data integration:** 13 clinical predictors and 11 physiological signals. **Temporal representation:** 1,440 one-minute observations per signal over the first 24 hours. **Validation:** 50 repetitions of five-fold outer cross-validation with patient-level separation and fold-contained preprocessing.

Acknowledgements

First and foremost, I would like to thank Professor Ana Maria Mendonça for, against all odds, making this dissertation possible.

I would also like to express my sincere gratitude to Professor Nuno Rodrigues, my supervisor, for his immense patience.

To Prof Celeste Dias and Prof Elisabete Monteiro, without whom this work would simply not have been possible.

To my family, thank you for supporting me in every possible way throughout this journey.

To my friends, thank you for softening every obstacle with laughter, smiles, and the comfort of never having to face it all alone.

To the love of my life, thank you for holding me up when all I wanted to do was fall.

To my day-one, thank you for all the conversations that somehow became impossible to end.

My heartfelt thanks to Laica – the one and only.

And, finally, to you, reading this.

Diogo Lopes

Contents

1 Introduction **1**
- 1.1 Clinical Context and Motivation . . . 1
  - 1.1.1 National and Institutional Context . . . 3
- 1.2 Research Questions and Hypotheses . . . 3
- 1.3 Objectives . . . 4
  - 1.3.1 General Objective . . . 4
  - 1.3.2 Specific Objectives . . . 4
  - 1.3.3 Main Contributions . . . 5
- 1.4 Dissertation Structure . . . 5

2 Theoretical Background **7**
- 2.1 Pathophysiology of Aneurysmal Subarachnoid Haemorrhage . . . 7
  - 2.1.1 Intracranial Aneurysm Formation and Rupture . . . 8
  - 2.1.2 Blood Products and Cerebrovascular Dysfunction . . . 9
  - 2.1.3 Early and Delayed Secondary Brain Injury . . . 10
  - 2.1.4 Aneurysm Securing and Early Management . . . 11
- 2.2 Cerebral Vasospasm . . . 11
  - 2.2.1 Clinical, Sonographic, and Angiographic Assessment . . . 12
  - 2.2.2 Transcranial Doppler and the Lindegaard Ratio . . . 13
  - 2.2.3 Relationship Between Vasospasm and Delayed Cerebral Ischaemia . . . 14
  - 2.2.4 Prevention and Treatment of Vasospasm-Related Complications . . . 14
- 2.3 Cerebral Autoregulation . . . 15
  - 2.3.1 The Lassen Curve . . . 16
- 2.4 Physiological Monitoring Signals and Derived Indices . . . 17
  - 2.4.1 Intracranial Pressure . . . 17
  - 2.4.2 Arterial Blood Pressure and Mean Arterial Pressure . . . 18
  - 2.4.3 Cerebral Perfusion Pressure . . . 19
  - 2.4.4 Cerebrovascular Resistance and Cerebral Blood Flow . . . 19
  - 2.4.5 ICP Pulse Amplitude, RAP, and RAC . . . 21
  - 2.4.6 Pressure Reactivity Index . . . 21
  - 2.4.7 Autoregulation-Guided Cerebral Perfusion Pressure . . . 22
- 2.5 Machine Learning and the Temporal Dimension of Neurocritical Monitoring . . . 23
  - 2.5.1 The “Dengler Paradox” and the Information Bottleneck . . . 23

3 Literature Review **25**
- 3.1 Literature Expectation vs. Reality . . . 25
  - 3.1.1 The Dominance of Static Modelling . . . 25
  - 3.1.2 The Scarcity of Dynamic and Autoregulatory Data . . . 25

3.2 Landscape Analysis and Research Positioning . . . 26
3.3 Systematic Review Methodology . . . 27
3.3.1 Search Strategy and Information Sources . . . 27
3.3.2 Eligibility Criteria . . . 28
3.3.3 Study Selection and Data Extraction . . . 29
3.3.4 Quality and Certainty Assessment Strategy . . . 31
3.4 Results & Discussion . . . 32
3.4.1 Risk of Bias and Quality of Evidence . . . 34
3.4.2 Temporal Resolution and Representation . . . 36
3.4.3 Outcome-Dependent Performance . . . 38
3.4.4 Input Modality Analysis . . . 41
3.5 Synthesis of Illustrative Modelling Approaches . . . 42
3.5.1 Longitudinal Monitoring Across Centres: Zarrin et al. (2024) . . . 42
3.5.2 Postoperative Biomarkers: Yang et al. (2025) . . . 43
3.5.3 Treatment-Related Information: Kim et al. (2023) . . . 43
3.6 Clinical Utility and Methodological Implications . . . 43
3.6.1 The Dengler Paradox as a Conceptual Interpretation . . . 44
3.6.2 Static and Longitudinal Representations: Reported Trade-Offs . . . 44
3.6.3 Potential Rule-Out Value and Its Limits . . . 45
3.7 Gap Analysis: Temporal Representation of Early Physiological Monitoring . . . 45

4 Methodology **47**
4.1 Study Design and Analytical Framework . . . 47
4.2 Clinical Setting, Data Sources and Governance . . . 49
4.3 Outcome Definition and Temporal Framework . . . 50
4.4 Study Populations and Eligibility . . . 51
4.4.1 Static Clinical Cohort . . . 51
4.4.2 Monitored Landmark Cohort . . . 52
4.4.3 Descriptive Cohort Characteristics . . . 54
4.5 Predictors and Physiological Signals . . . 54
4.5.1 Admission and Early Clinical Predictors . . . 54
4.5.2 Physiological Monitoring and Signal Derivation . . . 56
4.5.3 Common First-24-Hour Minute Grid . . . 57
4.5.4 Aggregated Representations . . . 58
4.5.5 Temporally Resolved Representation . . . 59
4.6 Phase-Specific Modelling . . . 60
4.6.1 Phase 1: Static Clinical Modelling . . . 60
4.6.2 Phase 2: Aggregated Physiological Modelling . . . 61
4.6.3 Phase 3: Temporally Resolved Physiological Modelling . . . 61
4.6.4 Phase 4: Static–Temporal Late Fusion . . . 63
4.7 Validation, Model Selection and Calibration . . . 64
4.8 Evaluation Metrics and Statistical Interpretation . . . 67

5 Results **69**
5.1 Cohort and Physiological Data Availability . . . 69
5.2 Phase 1: Static Clinical Prediction . . . 70
5.2.1 Predictive Performance . . . 70
5.2.2 Exploratory Held-Out Permutation Importance . . . 71
5.2.3 Exploratory PCoA Recoding Sensitivity . . . 71

5.3 Phase 2: Aggregated Physiological Prediction . . . 73
5.3.1 Representation and Model Performance . . . 73
5.3.2 Incremental Value of Aggregated Physiology . . . 73
5.4 Phase 3: Temporally Resolved Physiological Prediction . . . 75
5.4.1 Primary, Sensitivity and Comparator Results . . . 75
5.4.2 Paired and Probability-Based Findings . . . 75
5.5 Phase 4: Static–Temporal Late Fusion . . . 76
5.5.1 Static and Fusion Performance . . . 76
5.5.2 Paired Incremental Value . . . 77
5.5.3 Cross-Fitting Diagnostics . . . 77
5.6 Secondary Operating-Point and Calibration Results . . . 78
5.7 Summary of Main Results . . . 79

6 Discussion **80**
6.1 Principal Findings . . . 80
6.2 Static Clinical Information as a Predictive Baseline . . . 81
6.3 What Was Learned from First-24-Hour Physiology . . . 82
6.4 Incremental Physiological Value and Static–Temporal Fusion . . . 83
6.5 Methodological Strengths and Limitations . . . 84
6.6 Clinical, Methodological and Generalisability Implications . . . 85

7 Conclusions and Future Work **88**
7.1 Conclusions . . . 88
7.2 Future Work . . . 89

References **91**

A Characteristics of Studies Included in Literature Review **98**

B Predictor and Feature Definitions **105**
B.1 Physiological Signals and Derived Features . . . 105
B.2 Static Clinical Predictors . . . 107

C Expanded Results **108**
C.1 Phase 1: Static Clinical Prediction . . . 108
C.2 Phase 2: Aggregated Physiological Prediction . . . 109
C.3 Phase 3: Temporally Resolved Physiological Prediction . . . 110
C.4 Phase 4: Static–Temporal Late Fusion . . . 111
C.5 Secondary Operating-Point and Calibration Results . . . 112
C.6 Phase 4 Cross-Fitting Diagnostics . . . 113

List of Figures

2.1 **Anatomy of a Subarachnoid Haemorrhage.** The diagram illustrates an aneurysm rupture releasing blood into the subarachnoid space. *Illustration by Sophia Smith for Healthline (2012).* 8

2.2 **Circle of Willis**. Diagram of the arterial circulation at the base of the brain (inferior view). Green and blue labels identify the arterial territories grouped with the anterior and posterior circulations, respectively, in the study's aneurysm-location coding. *Adapted from Rhcastilhos (2007), itself based on Gray's Anatomy, via Wikimedia Commons.* 9

2.3 **The Lassen Curve of Cerebral Autoregulation.** Conceptual relationship between MAP and CBF. Within the autoregulatory range, compensatory changes in cerebrovascular resistance attenuate the effect of pressure fluctuations on CBF. Below the lower limit of autoregulation, cerebral perfusion becomes increasingly pressure-dependent and hypoperfusion may develop; above the upper limit, excessive perfusion may contribute to vascular and tissue injury. Adapted from Lidington et al. (2021). 16

2.4 **ICP Pulse Morphology.** (A) Normal ICP waveform with dominant P1; (B) abnormal ICP waveform with dominant P2. *Adapted from Dai et al. (2020).* 18

2.5 **Mechanism of the PRx.** With preserved pressure reactivity, increases in arterial pressure trigger compensatory vasoconstriction, limiting the increase in cerebral blood volume and producing an inverse or weak relationship between arterial pressure and ICP. When pressure reactivity is impaired, cerebral vessels respond passively and changes in arterial pressure are transmitted more directly to cerebral blood volume and ICP. In this illustration, CVR denotes cerebrovascular reactivity rather than cerebrovascular resistance. *Adapted from Schmidt et al. (2016).* . . 22

3.1 PRISMA 2020 flow diagram for the systematic review. 30

4.1 Overview of the four-phase analytical framework. 48

5.1 Phase 1 AUROC by static model family. Points show the mean across 50 repetitions and whiskers show one standard deviation across repetitions. The whiskers describe split variability and are not confidence intervals. 70

5.2 Exploratory held-out permutation importance for the Phase 1 XGBoost model. Bars show the mean change in outer-test AUROC after permutation of each predictor; whiskers show the standard deviation across the 250 outer-fold estimates. Negative and near-zero values indicate that permutation did not produce a consistent decrease in held-out discrimination. The estimates are model-specific predictive associations rather than causal effects. 72

5.3 Mean Phase 2 AUROC for each model–representation combination across 50 repetitions. The minimal representations were primary; the extended representations were secondary. 73
5.4 Repetition-level paired AUROC change after adding minimal physiological summaries to the static representation. Points show the 50 matched repetitions, boxes show the interquartile range, horizontal black lines show the median and diamonds show the mean. The dashed line denotes no change. 74
5.5 Phase 3 temporal-model results across the 50 matched repetitions. Panel A shows repetition-level AUROC for the primary value-only CNN, the value-plus-mask sensitivity and ROCKET–Ridge. Panel B shows the paired AUROC change for the mask sensitivity relative to the value-only CNN. Boxes show the interquartile range, black lines the median and diamonds the mean; the dashed line denotes no paired change. 75
5.6 Repetition-level paired AUROC difference between each Phase 4 fusion analysis and the matched fixed-C static logistic-regression comparator. Points show the 50 matched repetitions, boxes show the interquartile range, black lines the median and diamonds the mean. The dashed line denotes no incremental change. 78

List of Tables

1 Alignment of the dissertation with the United Nations Sustainable Development Goals. iii

3.1 Simplified summary of included studies grouped by modelling approach. Outcome labels are descriptive shorthand for study-specific endpoints and do not imply uniform diagnostic criteria across studies. Angiographic, TCD-defined, symptomatic, treatment-defined and DCI endpoints are distinguished where supported by the source reports; qualifiers are retained when they materially change the positive outcome definition. Additional endpoint qualifiers, validation details and source-reporting limitations are provided in Table A.1.
Abbreviations: ACV, Angiographic Vasospasm; DCI, Delayed Cerebral Ischaemia; SCV, Symptomatic Vasospasm; CVRV, Cerebral Vasospasm Requiring Verapamil; IA, Intra-Arterial; Tx, Treatment; ACC, Accuracy; AUC, Area Under the Curve; F1, F1-score; NPV, Negative Predictive Value; PPV, Positive Predictive Value; TPR, True Positive Rate (Sensitivity); TNR, True Negative Rate (Specificity); MSE, Mean Squared Error.
The modelling groups are descriptive: fixed patient-level representations may include information obtained after admission. Zarrin's rule-out result concerns CVRV rather than generic vasospasm; see Section 3.6. 32
3.2 Consolidated PROBAST Assessment: Risk of Bias and Applicability Concerns across the 19 included studies. Risk of Bias assesses internal validity, while Applicability evaluates relevance to the target clinical question. 35

4.1 Overview of the four-phase analytical framework. 49
4.2 Construction, characteristics, and physiological data adequacy of the monitored cohort used in Phases 2–4. 53
4.3 Descriptive characteristics of the Phase 1 analytical cohort and the final monitored cohort. 54
4.4 Final clinical and physiological inputs used across the modelling phases. 54
4.5 Summary statistics used to construct the Phase 2 physiological representations. . 58
4.6 Phase 1 model families and candidate configurations. 60
4.7 Phase 3 temporal model specifications and analytical roles. 62
4.8 Components of the Phase 4 late-fusion procedure. 64
4.9 Phase-specific internal-validation, selection and calibration procedures. 66

5.1 Phase 1 static-model discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values. 70

5.2 Paired Phase 1 AUROC differences between XGBoost and each comparator. Positive values favour XGBoost. Percentile ranges are empirical across the 50 matched repetitions. . . . 71
5.3 Exploratory Phase 1 sensitivity analysis with PCoA coded as posterior circulation. Paired differences are alternative coding minus the primary anterior-circulation coding. . . . 72
5.4 Matched Phase 2 AUROC differences after adding minimal physiological summaries to the static representation. Differences are combined minus static; positive values favour the combined representation. . . . 74
5.5 Phase 3 temporal-model discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values. ROCKET–Ridge AUROC is based on its raw discrimination score. . . . 75
5.6 Paired Phase 3 AUROC comparisons within the temporal analyses. Differences are analysis A minus analysis B; positive values favour analysis A. . . . 76
5.7 Phase 4 static and late-fusion discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values. 77
5.8 Paired Phase 4 AUROC comparisons across matched repetitions. Differences are mean within-repetition differences for analysis A minus analysis B; positive values favour analysis A. Percentile ranges are empirical across the 50 paired AUROC differences. . . . 77

A.1 Comprehensive summary of included studies. Outcome labels are descriptive shorthand for study-specific endpoints and do not imply uniform diagnostic criteria across studies. ACV is used only for imaging-defined arterial narrowing; TCD-defined, symptomatic, treatment-defined and DCI endpoints are retained separately, and missing operational details are not inferred. **Abbreviations:** ACV, Angiographic Vasospasm; DCI, Delayed Cerebral Ischaemia; SCV, Symptomatic Vasospasm; CVRV, Cerebral Vasospasm Requiring Verapamil; IA, Intra-Arterial; Tx, Treatment; ACC, Accuracy; TPR, True Positive Rate (Sensitivity); TNR, True Negative Rate (Specificity); PPV, Positive Predictive Value; NPV, Negative Predictive Value; F1, F1-score; MSE, Mean Squared Error; AUC, Area Under the Curve. **Models:** MLP, Multilayer Perceptron; RF, Random Forest; LR, Logistic Regression; SVC, Support Vector Classifier; GBM, Gradient Boosting Machine; BM, Baseline Model; PBD, Post-Bleed Day. . . . 99

B.1 Principal physiological signals and derived indices relevant to the neurocritical monitoring framework considered in this dissertation. . . . 105
B.2 Admission and early clinical predictors included in the final static analytical representation. . . . 107

C.1 Detailed Phase 1 static-model performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. AP is average precision. . . . 108
C.2 Detailed Phase 2 performance by representation and model family. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. Extended representations were secondary analyses. . . . 109

C.3 Detailed matched Phase 2 differences after adding minimal physiological summaries to the static representation. Differences are combined minus static; positive AUROC and AP differences and negative Brier differences favour the combined representation. 109
C.4 Detailed Phase 3 temporal-model performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. ROCKET–Ridge AUROC and AP are based on its raw discrimination score, whereas Brier score uses calibrated probabilities. . . 110
C.5 Detailed paired Phase 3 comparisons. Differences are analysis A minus analysis B. Positive AUROC and AP differences and negative Brier differences favour analysis A. 110
C.6 Detailed Phase 4 static, temporal and late-fusion performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. The temporal component rows report standalone Phase 3 performance from the outer-test probabilities reused in Phase 4. 111
C.7 Detailed paired Phase 4 comparisons across 50 matched repetitions. Differences are mean within-repetition differences for analysis A minus analysis B. Positive AUROC and AP differences and negative Brier differences favour analysis A. Percentile ranges describe the empirical 2.5th–97.5th percentiles of paired AUROC differences; the percentage column reports repetitions with strictly higher AUROC for analysis A. 111
C.8 Secondary screening operating-point and calibration summaries for the analysis-defining models. Values are mean ± SD across 50 repetitions. **P1 XGB** denotes Phase 1 XGBoost; **P2 Static LR**, Phase 2 static logistic regression; **P2 Min RF**, Phase 2 minimal-physiology random forest; **P2 Static+Min RF**, Phase 2 static-plus-minimal-physiology random forest; **P3 CNN**, the Phase 3 value-only CNN; **P3 Mask CNN**, the Phase 3 value-plus-mask CNN; **P3 ROCKET**, Phase 3 ROCKET–Ridge; **P4 Static LR**, the Phase 4 fixed-C static logistic regression; **P4 Fusion**, the primary static-LR plus value-only-CNN fusion; and **P4 Mask Fusion**, the static-LR plus value-plus-mask-CNN sensitivity fusion. **PPV** and **NPV** denote positive and negative predictive value. 112
C.9 Phase 4 cross-fitting diagnostics. Values are mean ± SD across the 250 outer-training fits arising from 50 repetitions of five-fold outer cross-validation, without intermediate repetition-level averaging. Static–temporal correlation is Pearson correlation between static decision scores and logits of clipped CNN probabilities; fusion coefficients are expressed on the standardised input scale. These diagnostics describe variation across overlapping training sets and are not independent replications or substitutes for outer-test performance. 113

List of Acronyms

ABP	Arterial Blood Pressure
ACA	Anterior Cerebral Artery
ACoA	Anterior Communicating Artery
ACV	Angiographic Vasospasm
AMP	Intracranial Pressure Pulse Amplitude
ANN	Artificial Neural Network
AP	Average Precision
aSAH	Aneurysmal Subarachnoid Haemorrhage
AUC	Area Under the Curve
AUROC	Area Under the Receiver Operating Characteristic Curve
CA	Cerebral Autoregulation
CBF	Cerebral Blood Flow
CHUSJ	Centro Hospitalar Universitário de São João
CNN	Convolutional Neural Network
CPP	Cerebral Perfusion Pressure
CPP$_{opt}$	Optimal Cerebral Perfusion Pressure
CSF	Cerebrospinal Fluid
CT	Computed Tomography
CTA	Computed Tomography Angiography
CV	Cerebral Vasospasm
CVR	Cerebrovascular Resistance
DBP	Diastolic Blood Pressure
DCI	Delayed Cerebral Ischaemia
DOI	Digital Object Identifier
DSA	Digital Subtraction Angiography
EBI	Early Brain Injury
EHR	Electronic Health Record
ET-1	Endothelin-1
EtCO$_2$	End-Tidal Carbon Dioxide
GCS	Glasgow Coma Scale

GRADE	Grading of Recommendations Assessment, Development and Evaluation
HIF	Hypoxia-Inducible Factor
HR	Heart Rate
ICA	Internal Carotid Artery
ICM+	Intensive Care Monitor Plus
ICP	Intracranial Pressure
ICPmax	Maximum Intracranial Pressure
ICPmin	Minimum Intracranial Pressure
ICU	Intensive Care Unit
LASSO	Least Absolute Shrinkage and Selection Operator
LLA	Lower Limit of Autoregulation
LR	Logistic Regression
LSTM	Long Short-Term Memory
MAP	Mean Arterial Pressure
MCA	Middle Cerebral Artery
mCPP	Mean Cerebral Perfusion Pressure
MFV	Mean Flow Velocity
ML	Machine Learning
MLP	Multilayer Perceptron
NO	Nitric Oxide
NPV	Negative Predictive Value
PCA	Posterior Cerebral Artery
PCoA	Posterior Communicating Artery
PICA	Posterior Inferior Cerebellar Artery
PPV	Positive Predictive Value
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PROBAST	Prediction Model Risk Of Bias ASsessment Tool
PRx	Pressure Reactivity Index
RAC	RAC Index
RAP	RAP Index
RF	Random Forest
ROCKET	RandOm Convolutional KErnel Transform
RQ	Research Question
SAH	Subarachnoid Haemorrhage
SBP	Systolic Blood Pressure
SCV	Symptomatic Vasospasm
SNS	National Health Service

SVM	Support Vector Machine
TCD	Transcranial Doppler
ULSSJ	Unidade Local de Saúde São João
VEGF	Vascular Endothelial Growth Factor
WBC	White Blood Cell
XGBoost	eXtreme Gradient Boosting

Chapter 1

Introduction

This chapter introduces the clinical and methodological problem addressed by this dissertation: prediction of cerebral vasospasm following aneurysmal subarachnoid haemorrhage (aSAH). It first establishes the clinical relevance of aSAH, cerebral vasospasm, and neurocritical surveillance, before motivating the use of routinely collected physiological monitoring as a potential source of predictive information beyond conventional clinical variables. The research questions, hypotheses, objectives, and principal contributions are then defined, followed by an overview of the dissertation structure.

1.1 Clinical Context and Motivation

Aneurysmal subarachnoid haemorrhage (aSAH) is an acute neurovascular emergency caused by rupture of an intracranial aneurysm and subsequent extravasation of arterial blood into the subarachnoid space. Although its worldwide incidence has declined over recent decades, from an estimated 10.2 per 100,000 person-years in 1980 to 6.1 in 2010, aSAH remains associated with substantial mortality and long-term neurological and cognitive morbidity (Etminan et al. 2019; Hoh et al. 2023). Its impact is particularly important because affected patients may require prolonged intensive care and can experience persistent disability or death.

The initial haemorrhage produces an abrupt mechanical and haemodynamic insult, including increased intracranial pressure, reduced cerebral perfusion, global cerebral ischaemia, disruption of cerebrovascular regulation, inflammation, and other components of early brain injury (Sehba et al. 2012; Hoh et al. 2023). Patients surviving the initial event remain vulnerable to secondary complications developing over the subsequent days. Among these, cerebral vasospasm is characterised by pathological narrowing of cerebral arteries and may contribute to impaired cerebral perfusion and delayed cerebral ischaemia (DCI) (Kassell et al. 1985; Hoh et al. 2023).

Cerebral vasospasm and DCI, however, are not interchangeable entities. DCI is a clinical and functional ischaemic syndrome with multifactorial pathophysiology involving large-vessel narrowing alongside mechanisms such as microcirculatory dysfunction, microthrombosis, neuroinflammation, impaired autoregulation, and cortical spreading depolarisations (Vergouwen et

al. 2010; Macdonald 2014; Hoh et al. 2023). Substantial arterial narrowing may occur without subsequent cerebral infarction, while delayed ischaemic injury may also develop without severe proximal vasospasm. Clear specification of the predicted outcome is therefore essential when developing and comparing prognostic models.

The modelling target in this dissertation is the clinician-assigned binary diagnosis of cerebral vasospasm used in the available clinical dataset. At the participating centre, surveillance is based primarily on transcranial Doppler (TCD), a non-invasive and repeatable bedside technique used to monitor cerebral arterial blood-flow velocities following aSAH (Hoh et al. 2023). Sonological findings are interpreted within the broader clinical surveillance pathway, and angiographic assessment may additionally be used where clinically indicated. The present work therefore predicts the recorded clinician-assigned vasospasm outcome rather than independently re-adjudicating angiographic narrowing, DCI, or functional outcome.

The delayed and evolving nature of post-aSAH complications creates an opportunity for physiological monitoring. Selected neurocritical patients undergo repeated or continuous acquisition of signals describing intracranial and systemic haemodynamics, cerebral perfusion, compensatory reserve, and cerebrovascular pressure reactivity. These include intracranial pressure (ICP), arterial blood pressure (ABP), cerebral perfusion pressure (CPP), ICP pulse amplitude, and derived indices such as the pressure reactivity index (PRx) (Dai et al. 2020; Smielewski et al. 2024). Unlike a single admission measurement, physiological monitoring describes an evolving state and may therefore contain information concerning subsequent cerebrovascular deterioration.

Predictive modelling following aSAH has nevertheless relied predominantly on fixed patient-level representations composed of demographics, neurological and radiological grades, aneurysm characteristics, imaging, laboratory measurements, and other clinical variables. The systematic review conducted in this dissertation identified comparatively few studies explicitly using longitudinal physiological monitoring, and those studies adopted substantially different representations, including daily or cumulative statistical summaries and sequence-preserving models (Roederer et al. 2014; Zarrin et al. 2024; Sen et al. 2025). The reviewed evidence also provided limited direct assessment of whether routinely collected physiological information adds predictive value beyond static clinical information.

This distinction between information source and representation motivates the experimental design of the present study. Static admission and early clinical information is first evaluated in a broader clinical cohort to establish a conventional predictive reference. Within the matched monitored cohort, first-24-hour physiology is then represented through patient-level statistical summaries, followed by minute-level sequence-preserving modelling of the same physiological observation period. Finally, static and temporal predictions are combined through late fusion to assess whether the physiological component provides information complementary to the static clinical baseline. The resulting four-phase framework evaluates progressively richer representations while preserving explicit matched comparisons where incremental physiological value is assessed.

The purpose of this dissertation is therefore not to develop a deployable bedside alarm or replace established surveillance such as TCD. Instead, it evaluates whether routinely collected early

clinical and physiological information can support prediction of subsequent cerebral vasospasm, whether preserving the temporal organisation of physiological monitoring improves its predictive representation, and whether physiological information provides incremental value beyond static clinical variables. Both positive and negative findings are relevant to this objective because they identify which information sources and representations warrant further investigation before prospective clinical translation.

1.1.1 National and Institutional Context

In Portugal, intensive care delivery is organised within the national *Rede de Referenciação de Medicina Intensiva*, which coordinates specialised intensive-care resources across the National Health Service (SNS) (Paiva et al. 2016). The availability of specialised infrastructure, human resources, and advanced monitoring capabilities nevertheless varies between institutions (Penedo et al. 2016), making neurocritical monitoring particularly concentrated in tertiary referral centres.

The present study was conducted in collaboration with the Neurocritical Care Unit of Unidade Local de Saúde São João (ULSSJ), formerly Centro Hospitalar Universitário de São João (CHUSJ), in Porto, Portugal. The clinical research environment combines routinely collected clinical information, serial cerebrovascular surveillance, and multimodal physiological monitoring using Intensive Care Monitor Plus (ICM+). These resources enabled retrospective integration of clinical and physiological information for the predictive analyses developed in this dissertation.

1.2 Research Questions and Hypotheses

The experimental framework was designed around a central predictive problem: although admission and early clinical information can provide a conventional basis for cerebral vasospasm prediction after aSAH, it remains unclear whether routinely collected physiological monitoring from the first 24 hours provides additional predictive information, particularly when its temporal structure is explicitly preserved.

The problem statement is therefore:

Problem Statement: The incremental predictive value of first-24-hour neuromonitoring beyond admission and early static clinical information, and the potential benefit of preserving its minute-level temporal structure, remain uncertain for the prediction of a subsequent clinician-assigned diagnosis of cerebral vasospasm after aSAH.

This problem statement is decomposed into three specific research questions and corresponding hypotheses:

RQ1: How can admission and early static clinical information be used to discriminate between patients who subsequently do and do not develop cerebral vasospasm?

H1: Admission and early static clinical information will provide meaningful but incomplete discrimination of subsequent cerebral vasospasm.

RQ2: How does the predictive performance of first-24-hour physiological information differ when represented as patient-level summary features versus minute-level temporal sequences?

H2: Preserving minute-level temporal structure will improve physiological prediction relative to patient-level summary representations, consistent with clinically relevant information being present in the temporal evolution of the monitored signals.

RQ3: Does first-24-hour physiological information provide incremental predictive value beyond static clinical information within the matched neuromonitoring cohort?

H3: First-24-hour physiological information will provide predictive information complementary to static clinical variables, such that its addition will improve discrimination relative to matched static-only prediction.

1.3 Objectives

The objectives of this dissertation follow the progression from characterising the existing evidence to evaluating static clinical information, aggregated physiology, temporally resolved physiology, and their incremental predictive value.

1.3.1 General Objective

The general objective of this dissertation is to determine whether admission and early clinical information and routinely collected first-24-hour neurophysiological monitoring can support prediction of subsequent clinician-assigned cerebral vasospasm after aSAH, with particular emphasis on the representation and incremental predictive value of physiological information.

1.3.2 Specific Objectives

To address the general objective and research questions, the following specific objectives were established:

1. **Characterise the current evidence base** through a systematic literature review of machine-learning approaches to cerebral vasospasm and selected related prediction tasks after aSAH, with particular attention to input modality, temporal representation, outcome definition, validation strategy, and methodological quality.

2. **Construct the analytical cohorts and temporal framework** by integrating retrospective clinical information with ICM+ neuromonitoring, assessing physiological data availability, and defining eligible pre-event first-24-hour observation windows for the monitored analyses.

3. **Establish a static clinical predictive reference** by evaluating admission and early clinical information across complementary machine-learning model families in the broader eligible clinical cohort.

4. **Evaluate aggregated physiological information and its incremental value** by representing first-24-hour monitoring through patient-level summary features and comparing physiology-only and static-plus-physiology prediction with matched static models in the monitored cohort.

5. **Evaluate temporally resolved physiological prediction** using models that preserve the minute-level ordering of the monitored signals and comparing their performance descriptively with aggregated representations of the same physiological observation period.

6. **Evaluate static–temporal integration** by combining static clinical and temporal predictions through late fusion and determining whether temporal physiology improves prediction relative to its contemporaneous matched static comparator.

1.3.3 Main Contributions

The principal contribution of this dissertation is the matched empirical evaluation of whether first-24-hour physiological monitoring provides predictive value beyond static clinical information. By distinguishing standalone physiological discrimination from incremental predictive value, the study evaluates whether information associated with the outcome also improves an existing clinical prediction.

A second contribution is the comparison of aggregated and minute-level sequence-preserving physiological representations within the same monitored cohort. This provides an empirical assessment of the temporal-representation gap identified in the Literature Review without assuming that preserving temporal structure must improve performance.

A further methodological contribution is the construction of an internally consistent analytical framework linking static clinical, aggregated physiological, temporal, and late-fusion prediction. The reconstruction and documentation of the relevant ICM+ signal-processing pathway additionally established the provenance and temporal interpretation of the processed physiological trends used in the modelling analyses.

1.4 Dissertation Structure

The remainder of this dissertation is organised into six further chapters. Chapter 2 presents the theoretical foundations of aSAH, cerebral vasospasm, cerebral autoregulation, and the physiological signals and derived indices relevant to neurocritical monitoring. Chapter 3 presents the systematic literature review and critically examines the existing prediction literature according to outcome definition, input modality, temporal representation, and validation methodology. Chapter 4 describes the retrospective study design, clinical and physiological datasets, cohort construction, temporal framework, four analytical phases, preprocessing procedures, and internal-validation strategy. Chapter 5 reports the results of the static, aggregated physiological, temporal, and fusion analyses. Chapter 6 interprets these findings in relation to the research questions and existing literature, while considering their methodological and clinical implications. Finally, Chapter 7

summarises the principal conclusions, explicitly addresses the research questions and hypotheses, and outlines priorities for future research.

Chapter 2

Theoretical Background

This chapter establishes the anatomical and pathophysiological foundations of the acute cerebral injury explored in this dissertation. Aneurysmal Subarachnoid Haemorrhage (aSAH) is a complex disease that evolves from the formation and rupture of an intracranial aneurysm to the acute and delayed consequences of blood flooding the subarachnoid space. Particular attention is given to delayed cerebral ischaemia (DCI), a severe complication often associated with the development of cerebral vasospasm. Cerebral vasospasm is a sustained, abnormal narrowing of cerebral arteries, characterised by specific pathophysiological mechanisms and assessed through clinical, sonographic and angiographic approaches, and warrants an individualised management strategy. Subsequently, the principles of cerebral autoregulation required to maintain stable cerebral blood flow are introduced, followed by the physiological signals and derived indices used in neuromonitoring to evaluate cerebrovascular function. Finally, the chapter establishes the conceptual motivation for moving from static clinical information towards temporally resolved physiological modelling.

2.1 Pathophysiology of Aneurysmal Subarachnoid Haemorrhage

Subarachnoid Haemorrhage (SAH) is defined as bleeding into the subarachnoid space, the cerebrospinal fluid-filled compartment located between the arachnoid and pia mater surrounding the brain. SAH may result from traumatic or non-traumatic causes. Among spontaneous non-traumatic cases, rupture of an intracranial aneurysm constitutes the predominant aetiology, accounting for approximately 85% of cases, while the remainder include perimesencephalic and other non-aneurysmal haemorrhages, vascular lesions, and systemic causes (Sehba et al. 2012; Mohan et al. 2019).

A subset of non-traumatic SAH cases remains angiographically negative, meaning that no causative aneurysm or other vascular lesion is demonstrated despite vascular imaging. These cases represent a distinct diagnostic category and should not be considered synonymous with definitively demonstrated aneurysmal rupture, although they may initially follow overlapping diagnostic and neurocritical care pathways (Mohan et al. 2019).

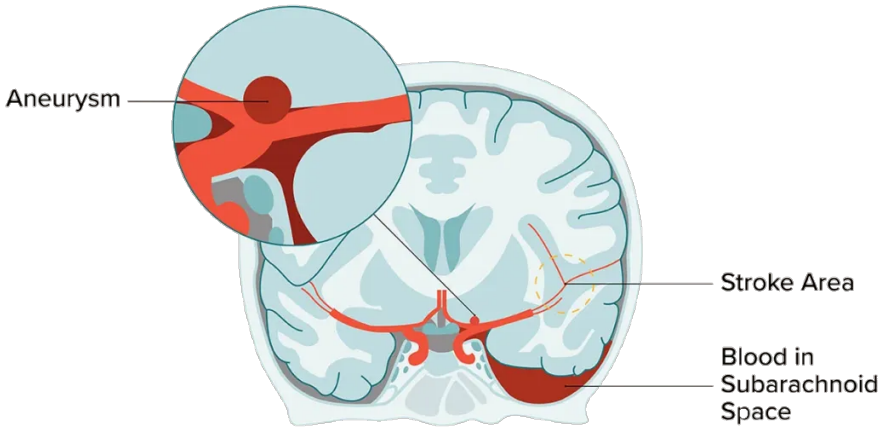


Figure 2.1: **Anatomy of a Subarachnoid Haemorrhage.** The diagram illustrates an aneurysm rupture releasing blood into the subarachnoid space. *Illustration by Sophia Smith for Healthline (2012).*

2.1.1 Intracranial Aneurysm Formation and Rupture

Intracranial aneurysms are pathological dilatations of the cerebral arterial wall, most commonly occurring at arterial bifurcations within or near the Circle of Willis, visible in Figure 2.2. Saccular aneurysms are the predominant morphology associated with spontaneous SAH. Their development reflects a complex interaction between haemodynamic wall stress, vascular-wall degeneration, and individual risk factors such as hypertension, smoking, age, sex, and genetic susceptibility (Etminan and Rinkel 2016).

Aneurysmal Subarachnoid Haemorrhage (aSAH) occurs when the aneurysmal wall ruptures, permitting high-pressure arterial blood to escape directly into the surrounding subarachnoid space (Figure 2.1). Unlike the contained circulation within an intact vessel, this newly created communication exposes the cerebrospinal fluid (CSF) compartment and the external surfaces of cerebral vessels to oxygenated arterial blood and its subsequent degradation products.

The immediate physical consequences of rupture are profound. Rapid accumulation of arterial blood increases the volume contained within the rigid cranial vault while simultaneously displacing CSF. Consequently, ICP rises abruptly, reducing the pressure gradient available to perfuse cerebral tissue. In severe haemorrhage, ICP may transiently approach systemic arterial pressure, producing a marked fall in Cerebral Perfusion Pressure (CPP) and a corresponding reduction in Cerebral Blood Flow (CBF) (Sehba et al. 2012). Acute cerebral vasoconstriction may additionally occur shortly after rupture and contribute to the reduction in cerebral perfusion independently of changes in ICP and CPP (Sehba et al. 2012; Kolias et al. 2009).

The rapid elevation of ICP may itself contribute to haemorrhage arrest. As the pressure difference across the ruptured aneurysm decreases, arterial outflow slows and local clot formation becomes possible (Sehba et al. 2012). However, the same process may transiently compromise cerebral circulation and initiate global cerebral ischaemia. Furthermore, subarachnoid and intraventricular blood may obstruct normal CSF circulation and absorption, increasing resistance

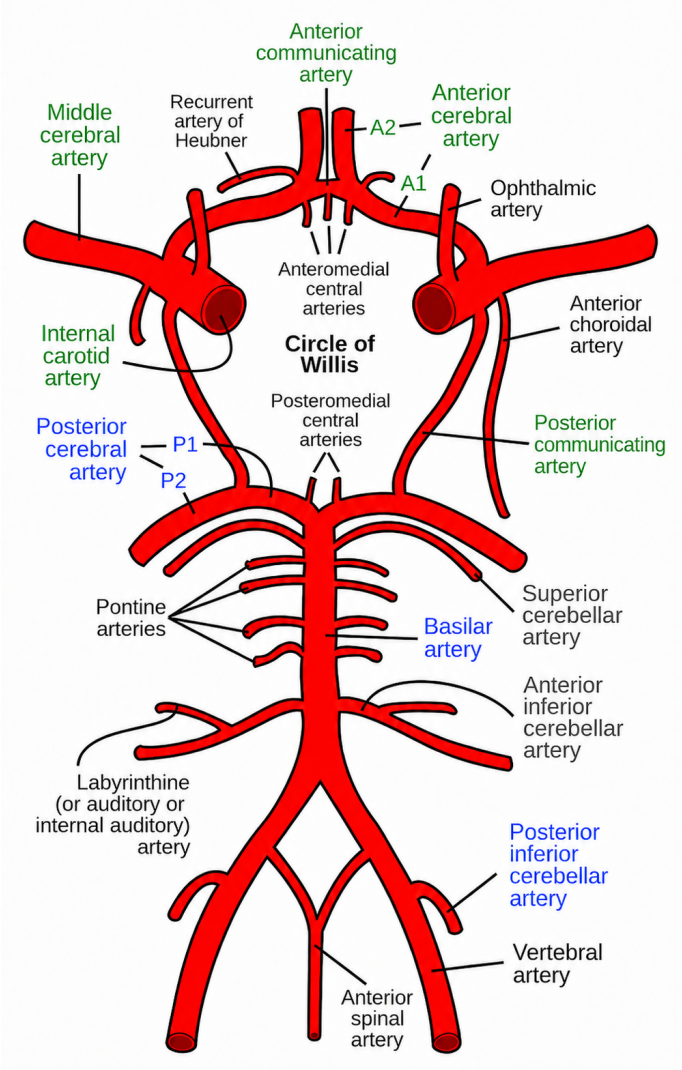


Figure 2.2: **Circle of Willis**. Diagram of the arterial circulation at the base of the brain (inferior view). Green and blue labels identify the arterial territories grouped with the anterior and posterior circulations, respectively, in the study's aneurysm-location coding. *Adapted from Rhcastilhos (2007), itself based on Gray's Anatomy, via Wikimedia Commons.*

to CSF outflow and contributing to acute or delayed hydrocephalus (Sehba et al. 2012; Hoh et al. 2023).

Therefore, aneurysm rupture should not be understood solely as the deposition of blood visible on imaging. It represents an abrupt mechanical and haemodynamic insult involving changes in intracranial volume, pressure, cerebral perfusion, vascular tone, and CSF dynamics.

2.1.2 Blood Products and Cerebrovascular Dysfunction

The presence of blood within the subarachnoid space also creates an abnormal biochemical environment around the cerebral arteries. Blood components that are physiologically contained within the vascular compartment become exposed to the adventitial surfaces of cerebral vessels

and CSF environment. As erythrocytes within the subarachnoid clot degrade over the following days, haemoglobin and other breakdown products accumulate within the CSF and surrounding tissues.

Oxyhaemoglobin has historically been identified as an important contributor to post-SAH vasoconstriction. Several interacting mechanisms have been proposed. Extracellular haemoglobin can bind and reduce the bioavailability of Nitric Oxide (NO), an endogenous mediator of vascular smooth-muscle relaxation. Oxyhaemoglobin may additionally promote the generation of reactive oxygen species, lipid peroxidation, endothelial dysfunction, and increased production or activity of vasoconstrictor mediators such as Endothelin-1 (ET-1) (Sobey and Faraci 1998; Kolias et al. 2009).

Under physiological conditions, endothelial NO contributes to the maintenance of cerebral vascular tone by activating soluble guanylate cyclase in vascular smooth muscle and promoting relaxation. Following SAH, this vasodilatory pathway may become impaired through endothelial injury, NO scavenging by extracellular haemoglobin, oxidative stress, and altered downstream signalling. Conversely, ET-1 is a potent and sustained vasoconstrictor whose concentration may increase following SAH. The resulting imbalance between vasodilatory and vasoconstrictive mechanisms provides one mechanistic explanation for the prolonged arterial narrowing observed during delayed cerebral vasospasm (Sobey and Faraci 1998; Kolias et al. 2009).

Importantly, these mechanisms are not independent. Oxidative stress, inflammation, endothelial dysfunction, smooth-muscle calcium signalling, altered potassium-channel activity, and vasoactive mediators interact within a progressively altered cerebrovascular environment. Consequently, cerebral vasospasm is currently regarded as a multifactorial vascular response rather than the consequence of a single spasmogenic substance (Kolias et al. 2009; Macdonald 2014).

2.1.3 Early and Delayed Secondary Brain Injury

The pathophysiological consequences of aSAH therefore evolve over time. The initial aneurysmal rupture produces a primary mechanical and haemodynamic insult, followed by a period of Early Brain Injury (EBI) encompassing approximately the first 72 hours. This early phase may involve cerebral oedema, blood–brain barrier disruption, microcirculatory dysfunction, inflammatory activation, oxidative stress, impaired cerebral autoregulation, and persistent abnormalities in cerebral perfusion (Sehba et al. 2012).

Reduced cerebral perfusion does not necessarily result immediately in irreversible tissue injury. Cerebral vessels initially attempt to maintain CBF through autoregulatory changes in vascular resistance. However, when perfusion pressure falls beyond the available autoregulatory reserve, cerebral tissue becomes increasingly hypoperfused. Persistent or sufficiently severe hypoperfusion compromises oxygen and substrate delivery, increasing the risk of cellular energy failure and cerebral ischaemia. The transition is continuous rather than defined by a universal CBF threshold and depends on factors including duration, regional metabolic demand, collateral circulation, and the patient's underlying cerebrovascular state (Baron 2001; Brassard et al. 2021).

During the following days, additional secondary complications may emerge. These include delayed cerebral vasospasm, microcirculatory dysfunction, microthrombosis, cortical spreading depolarisations, persistent impairment of autoregulation, inflammation, and Delayed Cerebral Ischaemia (DCI) (Macdonald 2014). Although historically much of this delayed deterioration was attributed directly to large-vessel vasospasm, contemporary understanding recognises that these processes interact but are not equivalent.

2.1.4 Aneurysm Securing and Early Management

Once aSAH has been diagnosed, an immediate therapeutic priority is prevention of aneurysm rebleeding by securing the ruptured lesion. Current guidelines recommend treatment as early as reasonably feasible, preferably within the first 24 hours when clinically possible (Hoh et al. 2023).

Two principal approaches are used. **Microsurgical clipping** involves direct surgical exposure of the aneurysm and placement of a clip across its neck, mechanically excluding the aneurysmal sac from the cerebral circulation while preserving the parent vessel. **Endovascular coiling** instead accesses the aneurysm through the arterial system and fills the aneurysmal sac with detachable coils, promoting localized thrombosis and preventing further blood entry.

Selection between clipping and endovascular treatment depends on aneurysm location and morphology, vascular anatomy, patient characteristics, haemorrhage severity, and local expertise. More complex endovascular techniques, including balloon-assisted or stent-assisted coiling and flow-diverting devices, may be considered for selected aneurysm morphologies but introduce additional considerations such as antiplatelet therapy in the setting of acute intracranial haemorrhage (Hoh et al. 2023).

Securing the aneurysm addresses the source of the initial haemorrhage and substantially reduces the risk of rebleeding. It does not, however, eliminate the secondary physiological consequences already initiated by the presence of blood within the subarachnoid space. Patients therefore remain vulnerable to delayed complications, most notably cerebral vasospasm.

2.2 Cerebral Vasospasm

Cerebral Vasospasm (CV) refers to pathological narrowing of cerebral arteries following SAH. In contrast to the acute vasoconstrictive response that may occur immediately following aneurysm rupture, delayed cerebral vasospasm generally emerges several days after the ictus, commonly within approximately days 3–14, with maximal incidence during the first and second post-haemorrhage weeks (Kassell et al. 1985; Hoh et al. 2023).

The delayed time course is consistent with the progressive degradation of subarachnoid blood and the development of endothelial, inflammatory, oxidative, and smooth-muscle abnormalities described in Section 2.1.2. Vasospasm most commonly affects the proximal large arteries of the Circle of Willis, reducing luminal diameter and increasing resistance to cerebral blood flow, as discussed further in Section 2.4.4.

However, the term vasospasm has historically been applied to several related but non-equivalent phenomena, including vessel narrowing visible on angiography, elevated cerebral blood-flow velocities on Transcranial Doppler (TCD), and delayed neurological deterioration. Clear distinction between these definitions is therefore essential.

2.2.1 Clinical, Sonographic, and Angiographic Assessment

Three complementary domains are particularly relevant to the assessment of vasospasm:

- **Clinical assessment:** Neurological deterioration may present as a new focal deficit or a reduction in the level of consciousness. Its assessment can be substantially limited in patients who are sedated, mechanically ventilated, or have severely impaired consciousness. Historical terms such as "clinical vasospasm" and "symptomatic vasospasm" were used inconsistently and sometimes combined neurological deterioration with vascular evidence. A multidisciplinary consensus recommended describing deterioration attributable to cerebral ischaemia, after exclusion of other causes, as "clinical deterioration caused by delayed cerebral ischaemia", while reserving "vasospasm" for radiologically demonstrated arterial narrowing (Vergouwen et al. 2010).

- **Sonographic assessment:** Sonographic vasospasm is identified using TCD measurements of cerebral arterial blood-flow velocity. Progressive luminal narrowing generally produces an increase in measured flow velocity, although absolute velocity must be interpreted alongside systemic haemodynamics and comparative vascular indices to distinguish vasospasm from hyperaemia. Because TCD is non-invasive, repeatable, and can be performed at the bedside, it is particularly suited to serial surveillance of neurocritical patients (Hoh et al. 2023; Fujimoto et al. 2026).

- **Angiographic assessment:** Angiographic vasospasm describes anatomical narrowing of the cerebral arterial lumen demonstrated using vascular imaging. Digital Subtraction Angiography (DSA) provides high spatial and temporal resolution and remains the reference anatomical technique, while Computed Tomography Angiography (CTA) provides a less invasive method of visualising the cerebral arterial tree. Angiography provides direct information regarding vessel geometry and the distribution of arterial narrowing but is less suited than bedside TCD to frequent serial monitoring (Hoh et al. 2023).

These assessment domains provide complementary but non-equivalent information and do not constitute mutually exclusive outcome categories. A patient may exhibit marked arterial narrowing without corresponding clinical deterioration, while another may develop cerebral ischaemia despite only modest large-vessel narrowing. Similarly, an isolated elevation in Doppler velocity does not automatically imply vasospasm because hyperdynamic circulation may produce increased velocities without a corresponding reduction in arterial diameter (Vergouwen et al. 2010; Macdonald 2014).

2.2.2 Transcranial Doppler and the Lindegaard Ratio

Transcranial Doppler ultrasonography provides a non-invasive method of estimating blood-flow velocity within the major basal cerebral arteries. A low-frequency ultrasound beam is transmitted through an acoustic cranial window and reflected by moving erythrocytes. The frequency shift between emitted and reflected ultrasound is proportional to the velocity component of the blood moving along the ultrasound beam:

$$f_D = \frac{2 f_0 v \cos\theta}{c}, \tag{2.1}$$

where f_D is the Doppler frequency shift, f_0 the transmitted ultrasound frequency, v the blood velocity, θ the insonation angle, and c the propagation velocity of ultrasound through tissue.

The resulting spectral waveform allows estimation of peak systolic, end-diastolic, and mean flow velocities. For monitoring cerebral vasospasm, Mean Flow Velocity (MFV) is particularly relevant because narrowing of a vessel at approximately preserved volumetric flow increases local blood velocity. In the Middle Cerebral Artery (MCA), mean velocities below approximately 120 cm/s are generally considered unlikely to represent clinically important vasospasm, whereas velocities approaching or exceeding 200 cm/s strongly support severe vasospasm (Lindegaard et al. 1989; Fujimoto et al. 2026).

Absolute velocity alone is nevertheless insufficient because increased global cerebral flow, such as during hyperaemia, can also elevate MCA velocity. This limitation motivates the use of the **Lindegaard Ratio**, defined as:

$$LR = \frac{MFV_{MCA}}{MFV_{ICA}}, \tag{2.2}$$

where MFV_{MCA} is the mean flow velocity in the MCA and MFV_{ICA} is the mean velocity in the ipsilateral extracranial Internal Carotid Artery (ICA).

When cerebral flow is globally increased, both MCA and ICA velocities increase and the ratio generally remains low. In contrast, focal narrowing of the MCA disproportionately elevates its velocity, producing a higher Lindegaard ratio. Ratios below approximately 3 are therefore more compatible with hyperaemia, values between approximately 3 and 6 support increasing degrees of vasospasm, and values above 6 are strongly suggestive of severe vasospasm (Lindegaard et al. 1989; Fujimoto et al. 2026).

The interpretation used clinically at ULSSJ follows this same principle while combining both a magnitude component, based on absolute arterial velocities, and a comparative component, based on velocity ratios between vascular territories. The resulting classification allows distinction between hyperaemia and different degrees of sonographic vasospasm. Vessel-specific velocity criteria are additionally applied to arteries other than the MCA.

Serial change is also clinically relevant. A rapid rise in cerebral arterial velocity may indicate evolving vasospasm even before the highest absolute thresholds are reached. Consequently,

repeated TCD measurements provide information not only on the presence and severity of vasospasm at a particular examination, but also on its trajectory over time (Grosset et al. 1992).

Despite its practical advantages, TCD has important limitations. Flow velocity is an indirect measure of vessel calibre rather than a direct measurement of arterial diameter or tissue perfusion. Measurements are operator-dependent, adequate acoustic windows may not be obtainable in all patients, and diagnostic performance differs between vascular territories. These limitations explain why angiographic and perfusion imaging remain important complementary investigations when sonographic findings and the clinical picture are discordant (Hoh et al. 2023; Fujimoto et al. 2026).

2.2.3 Relationship Between Vasospasm and Delayed Cerebral Ischaemia

Large-vessel vasospasm can reduce cerebral perfusion and therefore represents an important mechanism contributing to delayed cerebral injury. Nevertheless, vasospasm and DCI are not synonymous. DCI is a clinical and functional syndrome resulting from insufficient cerebral perfusion and may develop through several interacting mechanisms, including macrovascular narrowing, microvascular dysfunction, microthrombosis, cortical spreading depolarisations, neuroinflammation, and impaired cerebral autoregulation (Vergouwen et al. 2010; Macdonald 2014).

This distinction is supported by the imperfect correspondence between angiographic vasospasm and neurological deterioration. Some patients tolerate substantial arterial narrowing without developing cerebral infarction, whereas delayed ischaemic injury may occur in territories without severe proximal vasospasm. Therefore, vasospasm should be interpreted as an important and potentially treatable cerebrovascular complication of aSAH, rather than as a complete mechanistic explanation for all delayed neurological deterioration (Vergouwen et al. 2010; Macdonald 2014).

2.2.4 Prevention and Treatment of Vasospasm-Related Complications

Management following aSAH combines pharmacological prevention, maintenance of adequate cerebral perfusion, close neurological and physiological surveillance, and endovascular rescue strategies when necessary.

Nimodipine, a dihydropyridine calcium-channel antagonist, remains a central pharmacological therapy following aSAH. Early enteral administration improves neurological outcome and reduces the risk of DCI; importantly, its clinical benefit does not appear to derive simply from preventing angiographic vasospasm, emphasising that reduction of vessel narrowing and improvement in functional outcome are not equivalent therapeutic endpoints (Hoh et al. 2023).

Haemodynamic management aims to preserve adequate cerebral perfusion while avoiding complications caused by excessive fluid administration. Historical “Triple-H” therapy combined hypertension, hypervolaemia, and haemodilution. Contemporary management no longer supports prophylactic hypervolaemia or haemodilution because these approaches have not demonstrated clinical benefit and may increase cardiopulmonary complications. Instead, maintenance of euvolaemia is recommended (Hoh et al. 2023).

In patients who develop symptomatic cerebral ischaemia associated with vasospasm, controlled elevation of systemic blood pressure may be used to augment CPP and improve blood flow through compromised vascular territories. Vasopressors, including noradrenaline, may be used according to the patient's haemodynamic state and institutional protocol. This strategy attempts to compensate for reduced perfusion rather than directly reversing arterial smooth-muscle contraction. Prophylactic haemodynamic augmentation in otherwise stable patients is not recommended (Hoh et al. 2023).

When neurological deterioration or severe vasospasm persists despite medical management, endovascular therapy may be considered. Mechanical balloon angioplasty can produce durable dilation of accessible proximal vessels, whereas intra-arterial vasodilators can reach more distal vascular territories. Agents including verapamil, nicardipine, nimodipine, and milrinone have been used intra-arterially according to institutional practice and clinical circumstances. The comparative clinical benefit of these endovascular rescue strategies remains uncertain (Hoh et al. 2023).

Haemoglobin concentration represents another physiologically relevant component of cerebral oxygen delivery because reduced arterial oxygen-carrying capacity may exacerbate vulnerability to cerebral hypoperfusion. Red-cell transfusion, however, should not be regarded as a direct treatment for vasospasm. Recent randomised evidence has not established a universally superior transfusion threshold in aSAH; in the SAHARA trial, a liberal transfusion strategy did not significantly reduce unfavourable neurological outcome compared with a more restrictive strategy (English et al. 2025). Consequently, haemoglobin management remains part of broader neurocritical supportive care rather than a specific vasospasm therapy.

The molecular pathways described in Section 2.1.2 have also motivated pharmacological attempts to directly modify vasoconstrictive signalling. For example, endothelin-receptor antagonism can reduce angiographic vasospasm but has not consistently improved functional outcome, once again demonstrating the incomplete correspondence between large-vessel narrowing and neurological outcome (Al-Salihi et al. 2024).

2.3 Cerebral Autoregulation

Cerebral Autoregulation (CA) describes the intrinsic ability of the cerebral circulation to maintain relatively stable CBF despite changes in systemic perfusion pressure (Czosnyka et al. 2009). This mechanism protects cerebral tissue from both inadequate perfusion during reductions in arterial pressure and excessive blood flow during hypertension.

Autoregulation is achieved through interacting myogenic, metabolic, neurogenic, and endothelial mechanisms that continuously adjust cerebrovascular resistance. Cerebral arterioles constrict as perfusion pressure increases and dilate as pressure decreases, thereby buffering the transmission of systemic pressure fluctuations to the cerebral microcirculation.

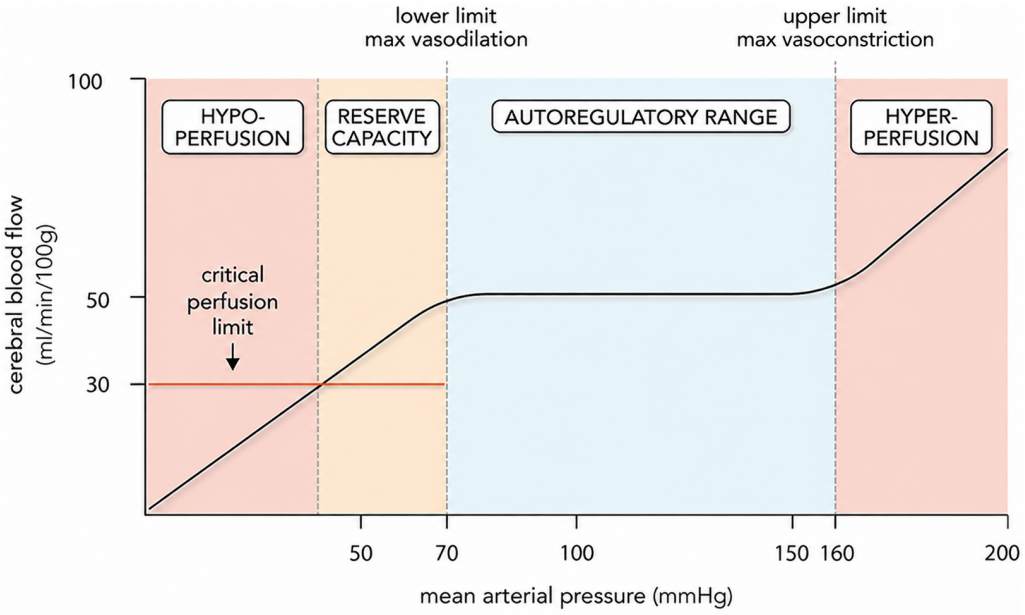


Figure 2.3: **The Lassen Curve of Cerebral Autoregulation.** Conceptual relationship between MAP and CBF. Within the autoregulatory range, compensatory changes in cerebrovascular resistance attenuate the effect of pressure fluctuations on CBF. Below the lower limit of autoregulation, cerebral perfusion becomes increasingly pressure-dependent and hypoperfusion may develop; above the upper limit, excessive perfusion may contribute to vascular and tissue injury. Adapted from Lidington et al. (2021).

2.3.1 The Lassen Curve

The classical representation of this relationship is the Lassen curve (Figure 2.3), which depicts a broad autoregulatory region between a Lower Limit of Autoregulation (LLA) and an upper limit. Within this range, changes in vascular resistance attenuate the effect of changes in perfusion pressure on CBF.

The traditional flat plateau should, however, be interpreted as a conceptual population-level representation rather than a universal pressure range shared by all individuals. Autoregulatory limits vary between patients and over time, and cerebral blood flow within the autoregulatory region is not perfectly independent of perfusion pressure (Brassard et al. 2021).

When perfusion pressure falls towards and below the individual's LLA, compensatory vasodilation progressively approaches its maximum capacity. Further reductions in pressure then produce increasingly pressure-dependent reductions in CBF, resulting in cerebral hypoperfusion and, if sufficiently severe or prolonged, ischaemic tissue injury. Conversely, pressure above the upper autoregulatory limit can lead to excessive cerebral blood flow, increased vascular permeability, and cerebral oedema (Brassard et al. 2021; Lidington et al. 2021).

Following aSAH, autoregulation may be impaired by several components of the disease process, including endothelial dysfunction, early brain injury, altered vascular smooth-muscle reactivity, inflammation, cerebral vasospasm, and microcirculatory disturbances (Budohoski et al. 2013;

Lidington et al. 2021). The cerebral circulation may therefore become increasingly pressure-passive, meaning that changes in systemic pressure are transmitted more directly to cerebral blood flow.

The dynamic and patient-specific nature of this impairment motivates continuous assessment of cerebrovascular reactivity rather than reliance exclusively on fixed population pressure thresholds.

2.4 Physiological Monitoring Signals and Derived Indices

Neurocritical monitoring integrates directly measured physiological variables with indices derived from their temporal relationships. Together, these signals provide complementary information regarding intracranial pressure, systemic haemodynamics, cerebral perfusion, intracranial compensatory reserve, and cerebrovascular pressure reactivity.

Table B.1 summarises the principal physiological variables and derived indices relevant to this dissertation.

2.4.1 Intracranial Pressure

Intracranial Pressure (ICP) is the pressure within the cranial vault. Its physiological behaviour is classically interpreted through the Monro–Kellie doctrine, according to which the adult cranium behaves approximately as a fixed-volume compartment containing three principal components: brain tissue, cerebral blood, and CSF (Czosnyka and Pickard 2004).

An increase in the volume of one component does not necessarily produce an immediate equivalent rise in ICP. Initially, compensatory mechanisms can accommodate additional intracranial volume, principally through displacement of CSF into the spinal compartment and reduction of intracranial venous blood volume. This capacity constitutes the intracranial compensatory reserve and may be expressed by the RAP index.

As additional volume is introduced and compensatory reserve becomes progressively exhausted, intracranial compliance decreases. The intracranial pressure–volume relationship therefore becomes increasingly steep: relatively small further increases in volume may then produce disproportionately large increases in ICP (Czosnyka and Pickard 2004). This non-linear relationship is clinically more informative than the simplified statement that any increase in intracranial volume necessarily causes an immediate increase in pressure.

Beyond its mean value, ICP is a pulsatile physiological signal. At adequate sampling resolution, a cardiac-induced ICP pulse typically contains three identifiable sub-peaks:

- **P1 (Percussion Wave):** The initial peak, predominantly reflecting transmission of arterial pulsation into the intracranial compartment.
- **P2 (Tidal Wave):** A second peak whose relative amplitude is associated with intracranial compliance and compensatory reserve.

- **P3 (Dicrotic Wave):** A later component associated with the dicrotic phase of the arterial waveform and aortic valve closure.

Under conditions of preserved intracranial compliance, P1 is typically the dominant component of the ICP pulse, resulting in the classical relationship $P1 > P2 > P3$. As intracranial compliance decreases, the same pulsatile increase in cerebral blood volume produces a greater pressure response, causing the tidal wave (P2) to become progressively more prominent relative to P1. P2 may eventually equal or exceed P1, a waveform pattern commonly associated with reduced compensatory reserve. In more advanced states of intracranial hypertension and impaired compliance, the individual pulse components may become less clearly distinguishable and the waveform may assume a broader or more rounded morphology (Dai et al. 2020; Kazimierska et al. 2023). Figure 2.4 illustrates the characteristic transition from a P1-dominant waveform under preserved compliance to a P2-dominant waveform associated with reduced compliance.

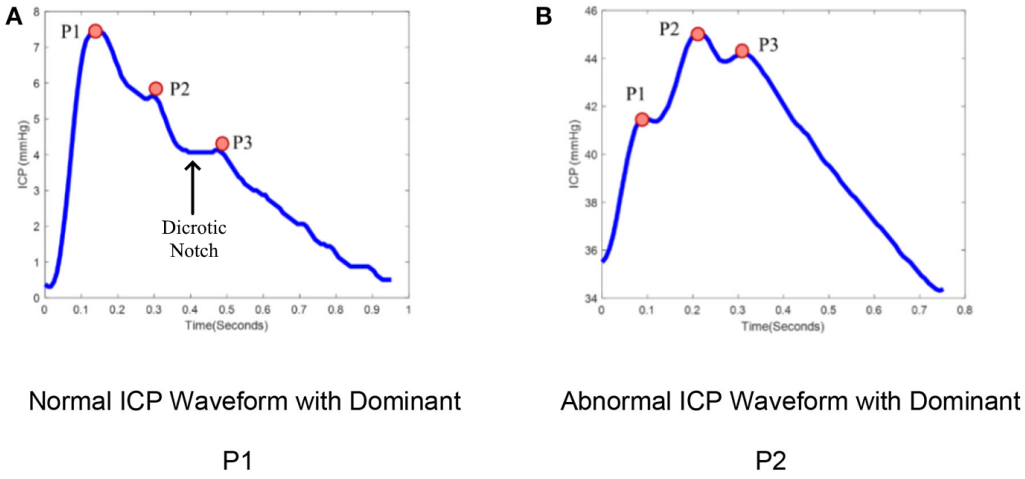


Figure 2.4: **ICP Pulse Morphology.** (A) Normal ICP waveform with dominant P1; (B) abnormal ICP waveform with dominant P2. *Adapted from Dai et al. (2020).*

2.4.2 Arterial Blood Pressure and Mean Arterial Pressure

Arterial blood pressure (ABP) represents the pressure generated within the systemic arterial circulation and constitutes the principal driving input for cerebral perfusion. The arterial waveform varies throughout each cardiac cycle between the SBP, corresponding approximately to peak ventricular ejection, and the DBP, representing the minimum arterial pressure before the subsequent systolic upstroke.

For haemodynamic and autoregulatory analysis, the MAP provides a representation of the average arterial driving pressure across the cardiac cycle. At normal heart rates it is commonly approximated as:

$$MAP \approx DBP + \frac{1}{3}(SBP - DBP). \tag{2.3}$$

The exact time-averaged MAP, however, can also be calculated directly from continuously sampled arterial waveforms rather than from this approximation.

When cerebral autoregulation is intact, changes in MAP are partially buffered by adaptive changes in cerebrovascular resistance. When autoregulatory capacity is impaired, cerebral blood flow becomes increasingly pressure-dependent and systemic hypotension may therefore translate more directly into cerebral hypoperfusion (Lidington et al. 2021; Czosnyka et al. 2009).

2.4.3 Cerebral Perfusion Pressure

Cerebral Perfusion Pressure (CPP) represents the effective pressure gradient available to drive cerebral blood flow. In conventional neurocritical monitoring it is approximated as:

$$CPP = MAP - ICP. \tag{2.4}$$

This relationship illustrates the interaction between systemic and intracranial haemodynamics: cerebral perfusion may decrease because systemic arterial pressure falls, because intracranial pressure rises, or through a combination of both.

Targets of approximately 60–70 mmHg are commonly cited in severe traumatic brain injury guidelines and neurocritical care literature (Carney et al. 2017; Hawryluk et al. 2022; Mathur et al. 2025). These values should not, however, be interpreted as a universally optimal CPP range for every neurocritical pathology or every patient with aSAH. The pressure required to maintain adequate cerebral blood flow depends on the patient's current autoregulatory state and may vary substantially over time (Lidington et al. 2021; Svedung Wettervik et al. 2021; Megjhani et al. 2023).

2.4.4 Cerebrovascular Resistance and Cerebral Blood Flow

As mentioned, cerebral perfusion pressure describes the pressure available to drive blood through the cerebral vascular bed, but it does not determine cerebral blood flow by itself. A simplified relationship between CPP, Cerebrovascular Resistance (CVR), and CBF can be written as (Fantini et al. 2016; Svedung Wettervik et al. 2023):

$$\mathrm{CBF} = \frac{\mathrm{CPP}}{\mathrm{CVR}}. \tag{2.5}$$

Equivalently,

$$\mathrm{CVR} = \frac{\mathrm{CPP}}{\mathrm{CBF}}. \tag{2.6}$$

In this formulation, CVR describes the resistance to blood flow across the cerebral vascular bed rather than the behaviour of a single artery. Small arteries and arterioles contribute substantially to this resistance and can constrict or dilate as part of cerebral autoregulation (Fantini et

al. 2016; Lidington et al. 2021). As a result, changes in CPP do not necessarily produce proportional changes in CBF. Within the autoregulatory range, changes in vessel diameter help limit the effect of fluctuations in perfusion pressure on cerebral blood flow.

The relevance of vessel diameter can also be illustrated using the Poiseuille relationship for idealised laminar flow through a cylindrical vessel:

$$R = \frac{8\eta L}{\pi r^4}, \tag{2.7}$$

where R represents vascular resistance, η blood viscosity, L vessel length, and r vessel radius (Fantini et al. 2016). Cerebral vessels do not behave as rigid cylindrical tubes, and cerebral blood flow under physiological conditions does not satisfy all assumptions of this model. The relationship is therefore useful here mainly to illustrate that relatively small changes in vessel radius can have a large effect on resistance.

This is particularly relevant to cerebral vasospasm. Narrowing of a cerebral artery can increase resistance along the affected vascular pathway and, if compensatory mechanisms are insufficient, reduce blood flow to distal territories (Hoh et al. 2023; Lidington et al. 2021; Svedung Wettervik et al. 2023). A relatively preserved CPP therefore does not necessarily imply that cerebral perfusion is also preserved, since flow can decrease when vascular resistance increases.

It is also important to distinguish cerebral blood-flow velocity from cerebral blood flow. TCD measures blood-flow velocity within the major cerebral arteries accessible through the cranial acoustic windows, where large-vessel vasospasm can be detected (Hoh et al. 2023; Fujimoto et al. 2026). However, total CVR also depends on the smaller resistance vessels and arterioles of the cerebral microcirculation, which are not directly assessed by TCD (Fantini et al. 2016; Lidington et al. 2021). CT perfusion can provide semiquantitative information on cerebral blood flow and indirectly assess microcirculatory perfusion (Abdulazim et al. 2023).

For an individual vessel, volumetric flow Q is related to mean blood velocity $\bar{v}$ and its cross-sectional area A by:

$$Q = \bar{v}A. \tag{2.8}$$

Changes in Doppler-measured velocity can therefore reflect changes in volumetric flow only when vessel cross-sectional area remains sufficiently stable (Jarrett et al. 2020). This assumption becomes particularly important during vasospasm because vessel diameter is no longer constant. A narrowed vessel may show an increased local blood velocity even when volumetric flow is reduced, because the decrease in cross-sectional area can be proportionally greater than the reduction in flow. For this reason, together with the limitations discussed in Section 2.2.2, TCD is useful for assessing cerebral arterial haemodynamics and identifying sonographic vasospasm, but velocity should not be interpreted as a direct measurement of tissue perfusion or absolute CBF (Jarrett et al. 2020; Grosset et al. 1992).

2.4.5 ICP Pulse Amplitude, RAP, and RAC

The pulsatile component of ICP can be quantified through its pulse amplitude (AMP). Rather than describing only mean intracranial pressure, AMP reflects the pressure response produced by the transient increase in intracranial arterial blood volume associated with each cardiac cycle. As intracranial compliance decreases, a comparable pulsatile volume change may produce a larger pressure response, increasing AMP (Kazimierska et al. 2023; Hawryluk et al. 2022).

The **RAP index** was developed as a continuous descriptor of intracranial compensatory reserve. Its name derives from the correlation coefficient (R) between ICP pulse amplitude (A) and mean intracranial pressure (P). It is therefore calculated as a moving correlation between AMP and ICP (Czosnyka and Pickard 2004).

When compensatory reserve is preserved, changes in mean ICP may occur with relatively limited coupling to pulse amplitude and RAP remains closer to zero. As reserve becomes exhausted and the pressure–volume relationship becomes steeper, AMP increasingly co-varies with mean ICP, causing RAP to approach positive values near $+1$. At very high intracranial pressures, however, vascular collapse and severely impaired cerebral perfusion can cause RAP to decrease again. Consequently, RAP cannot be interpreted as a simple monotonic scale in which lower values invariably indicate healthier physiology (Czosnyka and Pickard 2004).

The **RAC index** applies a related correlation-based approach and is calculated as a moving correlation between AMP and CPP. It has been investigated as an additional ICP-derived marker of cerebrovascular reactivity and the position of cerebral perfusion relative to autoregulatory limits (Zeiler et al. 2018).

2.4.6 Pressure Reactivity Index

The Pressure Reactivity Index (PRx) is one of the most extensively investigated continuous indices of cerebrovascular pressure reactivity in neurocritical care (M. Czosnyka et al. 1998; Czosnyka et al. 2009; Marek Czosnyka et al. 2017). It assesses whether slow spontaneous changes in systemic arterial pressure evoke an opposing vascular response or are instead passively transmitted to the intracranial vascular compartment.

In its conventional implementation, PRx is calculated as a moving Pearson correlation between slow fluctuations in ABP/MAP and ICP. Common implementations use thirty consecutive 10-second averaged values, corresponding to a 5-minute moving correlation window. The relevant slow-wave components broadly occupy a frequency range of approximately 0.005–0.05 Hz (Czosnyka et al. 2009; Marek Czosnyka et al. 2017).

Its basic physiological interpretation can be summarised as follows:

- **Negative or near-zero PRx:** Generally consistent with preserved cerebrovascular pressure reactivity, because changes in systemic pressure are opposed by adaptive changes in cerebrovascular tone.

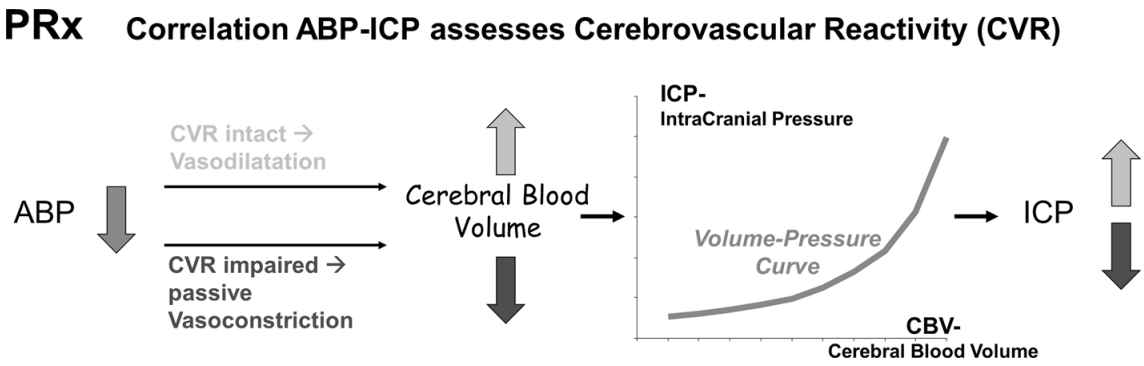


Figure 2.5: **Mechanism of the PRx.** With preserved pressure reactivity, increases in arterial pressure trigger compensatory vasoconstriction, limiting the increase in cerebral blood volume and producing an inverse or weak relationship between arterial pressure and ICP. When pressure reactivity is impaired, cerebral vessels respond passively and changes in arterial pressure are transmitted more directly to cerebral blood volume and ICP. In this illustration, CVR denotes cerebrovascular reactivity rather than cerebrovascular resistance. *Adapted from Schmidt et al. (2016).*

- **Increasingly positive PRx:** Generally indicates progressively impaired pressure reactivity, with systemic pressure fluctuations being transmitted more directly to the intracranial vascular compartment.

This interpretation should not be reduced to a strict binary threshold. PRx is a continuous index whose prognostic and physiological meaning depends on pathology, monitoring conditions, time, and the threshold selected for a particular clinical or research application. Values around +0.2 to +0.3 are frequently used to describe impaired pressure reactivity, but no single cut-off universally defines autoregulatory failure across all neurocritical populations (Sorrentino et al. 2012; Marek Czosnyka et al. 2017).

In patients with aSAH, impaired pressure reactivity has been associated with unfavourable neurological outcomes and with the evolving pathophysiology of secondary cerebral injury (Lidington et al. 2021; Budohoski et al. 2013; Svedung Wettervik et al. 2021). However, current evidence does not establish an increase in PRx as a specific or reliable precursor of subsequent sonographic vasospasm. Its relevance to vasospasm prediction should therefore be considered a physiological hypothesis to be evaluated rather than an established clinical relationship.

2.4.7 Autoregulation-Guided Cerebral Perfusion Pressure

The interaction between CPP and PRx has motivated attempts to define patient-specific perfusion targets rather than relying exclusively on fixed population thresholds.

When PRx is plotted against CPP over an appropriate monitoring period, a U-shaped relationship may emerge. Pressure reactivity is poorest at very low and very high perfusion pressures and more favourable within an intermediate range. The CPP associated with the minimum of the fitted PRx–CPP relationship has been termed the optimal cerebral perfusion pressure CPP_{opt} (Aries et al. 2012; Czosnyka et al. 2009; Marek Czosnyka et al. 2017).

Conceptually, CPP_{opt} represents the perfusion pressure at which cerebrovascular pressure reactivity appears most preserved for an individual patient during a particular monitoring period. Because autoregulatory state changes over time, CPP_{opt} is itself dynamic rather than a fixed patient characteristic.

Observational studies have associated deviation below or above CPP_{opt} with unfavourable outcomes in neurocritical populations, motivating interest in autoregulation-guided haemodynamic management (Aries et al. 2012; Beqiri et al. 2019). However, much of the foundational evidence originates from traumatic brain injury, and the optimal interpretation of PRx and CPP_{opt} may differ in aSAH. Recent aSAH-specific studies support the physiological relevance of individualised autoregulatory targets but do not yet establish CPP_{opt}-guided management as a universal standard of care (Svedung Wettervik et al. 2021; Megjhani et al. 2023).

Accordingly, CPP_{opt} is best understood as an important example of how continuous physiological monitoring can transform raw pressure measurements into dynamic, patient-specific descriptors of cerebrovascular function.

2.5 Machine Learning and the Temporal Dimension of Neurocritical Monitoring

Clinical risk stratification following aSAH has traditionally relied heavily on variables obtained at or shortly after admission, including age, neurological examination, haemorrhage burden, and grading systems such as the Hunt and Hess and Modified Fisher scales. These variables provide clinically useful summaries of initial disease severity but represent predominantly static observations of a pathophysiological process that continues to evolve over subsequent hours and days.

For example, Rabinstein et al. (2004) demonstrated that although Fisher grades were associated with angiographic vasospasm, they showed limited specificity for subsequent cerebral infarction. More broadly, such clinical scales necessarily compress complex anatomical and physiological information into a small number of categories and therefore cannot directly represent the evolving haemodynamic state of the patient.

2.5.1 The "Dengler Paradox" and the Information Bottleneck

Machine learning has long been proposed as a means of modelling the complex and potentially non-linear relationships underlying clinical outcomes. Early work by Toledo et al. (2009), for example, demonstrated the feasibility of applying machine-learning techniques to outcome prediction following subarachnoid haemorrhage. However, the predictive potential of increasingly sophisticated algorithms remains constrained by the information contained in their inputs.

This limitation was illustrated by Dengler et al. (2021), who compared multiple machine-learning approaches with established clinico-radiological scores for functional outcome prediction after aSAH. Increasing algorithmic sophistication produced no substantial improvement over

conventional scoring systems: the best-performing machine-learning model achieved an AUC of approximately 0.78 compared with 0.76 for the strongest established clinical score.

In this dissertation, this observation motivates the term *"Dengler Paradox"*: increasing algorithmic complexity does not necessarily overcome an information bottleneck imposed by limited input information. The term is used here as a conceptual description rather than as established terminology within the wider literature.

This interpretation does not imply an intrinsic limitation of machine learning. Instead, it suggests that meaningful gains in predictive performance may require enrichment of the input representation itself. When the underlying biological process evolves dynamically, static admission variables cannot explicitly represent the trajectory through which physiology changes over time.

Neurocritical monitoring provides a potential means of addressing this limitation. Physiological signals such as ICP, ABP, CPP, PRx, and related indices are repeatedly measured throughout the patient's ICU course and therefore encode temporal information that is absent from conventional admission scores. Importantly, preserving temporal order is conceptually distinct from simply adding more variables: statistical summaries may retain information about magnitude and variability while discarding the sequence and evolution through which those values occurred.

The distinction between static, aggregated longitudinal, and temporally resolved modelling therefore forms an important methodological axis for predictive neurocritical care. Chapter 3 examines how the existing machine-learning literature has approached this problem and evaluates the extent to which current cerebral vasospasm prediction models exploit the temporal information available in physiological monitoring.

Chapter 3

Literature Review

This chapter synthesises machine learning (ML) approaches to the prediction of cerebral vasospasm and selected studies of Delayed Cerebral Ischaemia (DCI) following subarachnoid haemorrhage. Cerebral vasospasm is the primary focus of this dissertation; DCI is considered as a distinct, adjacent prediction problem in the aSAH literature. The review examines input information, temporal representation, outcome definitions and validation, providing the rationale for evaluating physiological monitoring alongside static clinical information. The review was conducted in accordance with the PRISMA 2020 guidelines (Page et al. 2021).

3.1 Literature Expectation vs. Reality

This systematic review was initially motivated by an interest in ML methods using continuous physiological monitoring, including ICP, ABP and CPP, and derived autoregulatory indices such as PRx, to predict cerebral vasospasm after aSAH. The review covered publications from 2011 to 2025. The identified evidence placed greater emphasis on clinical, laboratory and imaging features than on temporally resolved physiological monitoring.

3.1.1 The Dominance of Static Modelling

Most included studies used fixed patient-level feature representations, commonly based on clinical scores and grades, aneurysm morphology, laboratory biomarkers or imaging radiomics. Many of these inputs were available at admission, although some studies incorporated postoperative measurements or information collected over the early hospital course. These approaches generally did not preserve the sequence of physiological observations.

3.1.2 The Scarcity of Dynamic and Autoregulatory Data

Within the included corpus, three studies explicitly analysed longitudinal physiological monitoring, incorporating signals such as ICP, MAP or CPP alongside other clinical or laboratory information:

- **Zarrin et al. (2024):** Combined routinely collected intensive-care data, including ICP and laboratory measurements such as white blood cell count, platelet count, glucose, chloride and partial pressure of oxygen (PO_2), using cumulative feature summaries.
- **Sen et al. (2025):** Combined hourly physiological measurements with daily TCD measurements and serum sodium information in a sequential model.
- **Roederer et al. (2014):** Integrated routinely recorded physiological and intermittent laboratory measurements through individual-day and cumulative summaries.

These studies illustrate different ways of representing longitudinal information: statistical aggregation and preservation of temporal order. No included study reported predictive modelling of pulse-wave morphology or the use of derived autoregulatory indices such as PRx. This observation identifies a limitation of the reviewed evidence, rather than establishing their absence from all related research. It motivates evaluation of temporally resolved monitoring and derived physiological indices without assuming that raw waveform modelling is required.

3.2 Landscape Analysis and Research Positioning

Palermo et al. (2025) published a systematic review and meta-analysis of ML models for vasospasm prediction, with substantial overlap between their included studies and this chapter's corpus. Their review included both clinically and angiographically detected vasospasm and explicitly distinguished vasospasm from DCI. Studies focused solely on DCI were eligible for their synthesis only if they also provided specific vasospasm results or predictive modelling.

Palermo et al. (2025) reported a mean sensitivity of 97.6% for models in their deep-learning group. This is a result of their synthesis across heterogeneous studies and does not establish that a model family is intrinsically superior across settings. Their review also discussed longitudinal approaches such as Zarrin et al. (2024) and Roederer et al. (2014). The present review provides a complementary analysis of how input information and its temporal representation relate to the reported prediction tasks.

The positioning of this review rests on three aspects:

1. **Population and Outcome Specification:** This review is adult-focused and excluded a dedicated paediatric cohort, whereas Palermo et al. (2025) included the paediatric study by Skoch et al. (2017). However, the retained study by Zarrin et al. (2024) includes mixed-age and mixed-aetiology SAH cohorts and is an exception to the adult aSAH eligibility framework. Angiographic and sonographic vasospasm, historical study-defined symptomatic vasospasm (SCV), and DCI are distinguished throughout the interpretation.
2. **Adjacent DCI Evidence:** Selected studies predicting DCI remain in the corpus because they address an overlapping post-aSAH prediction problem within a sparse vasospasm-oriented literature. Examples include Kim et al. (2023), Azzam et al. (2024) and Alexopoulos et al. (2022). Their inclusion broadens the evidence considered without treating

DCI as a form of vasospasm or interpreting treatment associations and overlapping time courses as proof of causal equivalence.

3. **Multi-Level Stratification:** While Palermo et al. (2025) performed a meta-analysis, this chapter organises its 19 included studies by temporal representation, input modality and target outcome. These comparisons informed the rationale for the subsequent experimental design.

Static and Longitudinal Information The static/dynamic distinction is useful as a broad organising framework, but input timing, temporal representation and prediction updating are separate properties. Three representations are particularly relevant:

1. **Baseline Information:** Clinical, laboratory or imaging variables collected at a defined early time point and used to produce a patient-level risk estimate.

2. **Aggregated Longitudinal Information:** Repeated observations summarised through measures such as means, extrema, percentiles or variability. These features retain aspects of magnitude and variability while discarding the order of observations within the summarised period.

3. **Sequence-Preserving Information:** Repeated observations represented in temporal order, allowing a model to use the evolution of the measurements as well as their values.

The tables distinguish longitudinal monitoring approaches from predominantly fixed patient-level representations. These are descriptive groups: a fixed feature vector may include information collected after admission, and longitudinal inputs do not automatically imply continuously updated predictions or demonstrated clinical utility. For example, the laboratory information summarised for Savarraj et al. (2021) spans days 0–3, while Yang et al. (2025) used biomarkers collected after surgery. The distinction between aggregation and sequence preservation is therefore considered explicitly in the narrative synthesis.

3.3 Systematic Review Methodology

This systematic review was conducted following PRISMA 2020 guidelines (Page et al. 2021).

3.3.1 Search Strategy and Information Sources

A comprehensive literature search was executed to identify primary research applying machine learning techniques for the prediction of cerebral vasospasm following aneurysmal subarachnoid haemorrhage (aSAH). The systematic database searches were conducted between 27 November and 4 December 2025, covering publications from 2011 to 2025. Screening, study selection and data extraction were subsequently completed in January 2026. Systematic queries were performed

across five major electronic databases: PubMed, Scopus, IEEE Xplore, PubMed Central (PMC), and the Cochrane Library.

The specific search query employed a combination of disease, outcome, and intervention descriptors, explicitly filtering for predictive tasks rather than diagnostic or segmentation-based research:

```
("aneurysmal subarachnoid haemorrhage" OR "subarachnoid haemorrhage"
OR aSAH) AND ("cerebral vasospasm" OR vasospasm) AND ("machine
learning" OR "deep learning" OR "neural network*" OR "artificial
intelligence") AND (predict* OR infer*)
```

In addition to automated indexing, a manual search strategy was implemented to ensure the capture of non-indexed seminal works and grey literature. This included backward citation chaining from the reference lists of all included studies and relevant literature reviews, alongside targeted web-based searching of indexed literature.

The search was vasospasm-oriented. The inclusion of selected DCI studies should therefore not be interpreted as a separate comprehensive search of all DCI prediction research.

3.3.2 Eligibility Criteria

The review used the following inclusion and exclusion framework to define its clinical and methodological scope. Exceptions retained in the final corpus are identified below.

Inclusion Criteria

- **Population:** Research involving adult cohorts ($\geq$18 years) with a confirmed diagnosis of aSAH.
- **Methodology:** Original research focusing on the development or validation of machine learning models for prognostic tasks.
- **Outcome:** Prediction of Cerebral Vasospasm (CV) or Delayed Cerebral Ischaemia (DCI), with the endpoint defined separately for each study.
- **Publication Type:** Peer-reviewed primary research reports, with full-text availability as the intended requirement. Conference papers were distinguished from conference abstracts; the two abstract-based inclusions are described below.

Exclusion Criteria

- **Ineligible Task:** Studies focused on diagnostic detection (classification of current state) or segmentation rather than forecasting future events.

- **Ineligible Endpoints:** Prediction of global functional scores, mortality, or composite secondary brain injuries without a specific vasospasm component, together with the historically applied restriction on DCI scope described below.
- **Ineligible Data Source:** Animal models, synthetic simulations and biomarker sources outside the monitoring-related clinical scope applied during screening. This did not require all included studies to use continuous physiological monitoring; clinical-score, imaging and laboratory studies were also included.
- **Ineligible Publication Type:** Narrative reviews, editorials, and conference abstracts lacking comprehensive methodological reporting.

Scope and Retained Exceptions The documented selection account considered DCI studies within a vasospasm-oriented scope and excluded reports whose DCI definitions were explicitly framed independently of vasospasm. This historical restriction limits the breadth of the DCI evidence captured. It is not a clinical definition of DCI: as described in Section 2.2.3, vasospasm is one contributor to a distinct, multifactorial syndrome and is neither necessary nor sufficient for its occurrence. Historical SCV terminology is retained when reporting a study's endpoint, without treating it as interchangeable with DCI.

Two departures from the stated framework are relevant to interpretation. First, Zarrin et al. (2024) was retained despite including children and traumatic as well as non-traumatic SAH; the corpus is therefore adult-focused rather than exclusively adult aSAH. Second, Sadri et al. (2024) and Lintas et al. (2024) were retained on the basis of detailed structured abstracts when full texts were unavailable. Their extraction and appraisal were limited to the information available in those reports. These exceptions describe the retained corpus without changing the historical study-selection decisions.

3.3.3 Study Selection and Data Extraction

Title and abstract screening, full-text assessment, study selection and data extraction were conducted by the author using the predefined eligibility criteria and structured extraction framework.

The initial database search yielded a total of 89 records (Scopus: 46; PubMed: 30; PMC: 11; IEEE Xplore: 2; Cochrane Library: 0).

Automated deduplication based on DOI and title matching excluded 36 records, with one additional residual duplicate removed manually.

The remaining 52 unique records underwent a two-stage screening process: (1) Title and Abstract Screening excluded 18 records that did not meet primary research criteria, such as conference abstracts, editorials, and narrative reviews; (2) Full-Text Assessment considered the remaining 34 reports, with the source-availability exceptions described below, and resulted in the exclusion of 16 studies for the following reasons:

- **Ineligible Outcome** ($n = 8$)**:** Global functional outcomes, composite endpoints outside the review's scope, or DCI definitions explicitly framed independently of vasospasm under the historical scope restriction.
- **Ineligible Methodology** ($n = 5$)**:** Research focusing on diagnostic segmentation or event detection rather than prognostic prediction.
- **Ineligible Data Source** ($n = 2$)**:** A synthetic simulation and a salivary viral DNA analysis, excluded under the data-source scope applied during screening.
- **Wrong Population** ($n = 1$)**:** Use of a paediatric cohort (<18 years).

Complementary backward citation chaining and web-based searching identified five additional records. Of these, one unique study met the eligibility criteria (Savarraj et al. 2021) and was added to the synthesis. The final selection comprised 19 studies ($N = 19$).

The flow diagram illustrating this identification and screening process is displayed in Figure 3.1. The diagram summarises report selection; the availability of full-text evidence is qualified below.

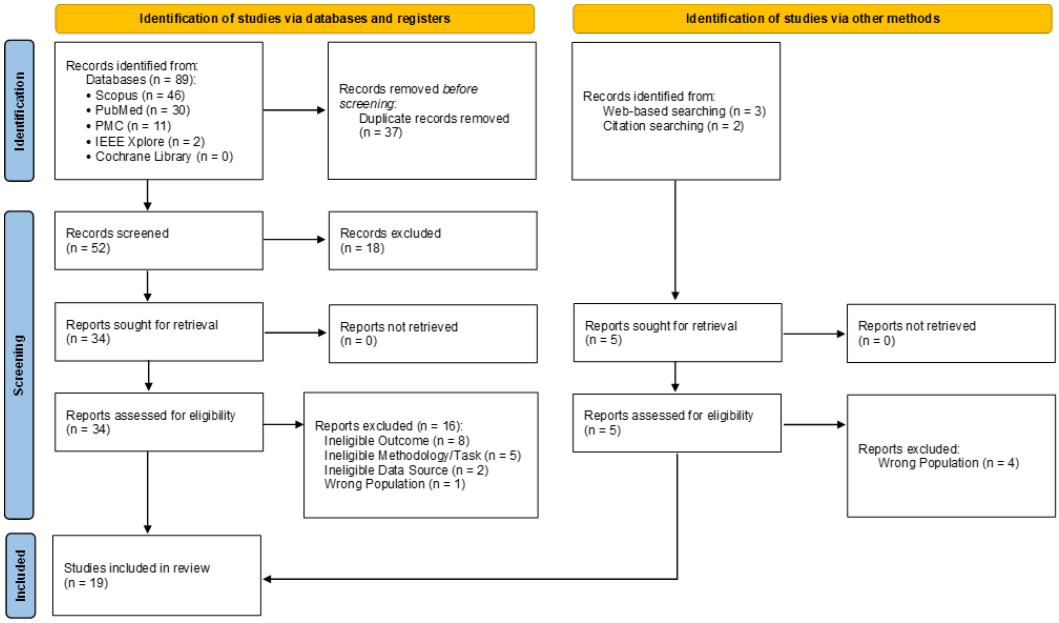


Figure 3.1: **PRISMA 2020 flow diagram for the systematic review.** The diagram illustrates the selection process from electronic databases and supplementary manual searches. *Adapted from Page et al. (2021).* Two included studies were assessed using detailed structured abstracts because their full texts were unavailable; these source-availability exceptions are described in Section 3.3.3.

Data were extracted using a structured collection sheet. Full-text articles were used where available. In two instances where full-text retrieval was precluded (Sadri et al. 2024; Lintas et al. 2024), the structured abstracts supplied the methodological information and quantitative metrics used for extraction. These were exceptions to the intended full-text requirement, and details

not reported in the abstracts could not be assessed. The extraction protocol captured the following specific domains:

1. **Category:** Classification of the study approach, presented as predominantly fixed patient-level representations or longitudinal monitoring approaches, with temporal aggregation distinguished from sequence preservation in the synthesis.
2. **Study Characteristics:** Authors, year, design (retrospective/prospective), and cohort sample size (*N*).
3. **Target:** Specific endpoint definition, including ACV, TCD-defined vasospasm, study-defined SCV, DCI or treatment-requiring vasospasm.
4. **Input Predictors:** Granular listing of features utilised, including clinical scores, radiomics, biomarkers, or physiological signals.
5. **Models:** Machine Learning architectures employed.
6. **Validation Strategies:** Internal (e.g. k-fold cross-validation) or external validation protocols, distinguishing evaluation of a fixed model from separate model training at another centre where reported.
7. **Performance Metrics:** Quantitative reporting of predictive performance (e.g. AUC, Sensitivity, Accuracy).

3.3.4 Quality and Certainty Assessment Strategy

The PROBAST assessments and qualitative appraisal of the evidence were performed by the author. The PROBAST framework (Wolff et al. 2019) was used to assess risk of bias and applicability to the review question. Risk of bias covers four domains: Participants, Predictors, Outcome and Analysis. Applicability concerns address Participants, Predictors and Outcome. Given the heterogeneity of the included studies, the review used a qualitative assessment rather than formal quantitative assessment of reporting bias or GRADE certainty grading:

1. **Reporting Bias:** Where full texts were available, abstract claims were compared with results tables to identify potential selective reporting. The assessment also considered whether metrics such as AUC, Sensitivity, F1-score and Negative Predictive Value were reported alongside overall Accuracy. For abstract-based inclusions, this assessment was limited by the available reporting.
2. **Confidence in the Evidence:** Confidence was assessed qualitatively using the overall PROBAST profile, the reported validation strategies and sample adequacy. Events-per-variable ratios were considered in relation to analytical stability and overfitting risk, rather than as a direct measure of statistical precision. The resulting judgement was a narrative assessment, not a formal GRADE rating.

3.4 Results & Discussion

The systematic selection process yielded 19 studies of ML prediction across vasospasm-related endpoints and separately identified DCI outcomes. Their reported performance reflects differences in populations, inputs, targets and validation, as well as the modelling methods used.

The synthesis is organised along three axes: **Temporal Representation** (baseline information, longitudinal summaries and ordered sequences), **Outcome Definition** (angiographic or sonographic vasospasm, study-defined SCV, DCI and treatment-related endpoints), and **Input Modality** (e.g. biomarkers, physiological signals and clinical scores). Comparisons across studies are descriptive and do not isolate the effect of an algorithm or input modality.

A simplified summary of the included studies is presented in Table 3.1.

Table 3.1: Simplified summary of included studies grouped by modelling approach. Outcome labels are descriptive shorthand for study-specific endpoints and do not imply uniform diagnostic criteria across studies. Angiographic, TCD-defined, symptomatic, treatment-defined and DCI endpoints are distinguished where supported by the source reports; qualifiers are retained when they materially change the positive outcome definition. Additional endpoint qualifiers, validation details and source-reporting limitations are provided in Table A.1.
Abbreviations: ACV, Angiographic Vasospasm; DCI, Delayed Cerebral Ischaemia; SCV, Symptomatic Vasospasm; CVRV, Cerebral Vasospasm Requiring Verapamil; IA, Intra-Arterial; Tx, Treatment; ACC, Accuracy; AUC, Area Under the Curve; F1, F1-score; NPV, Negative Predictive Value; PPV, Positive Predictive Value; TPR, True Positive Rate (Sensitivity); TNR, True Negative Rate (Specificity); MSE, Mean Squared Error.
The modelling groups are descriptive: fixed patient-level representations may include information obtained after admission. Zarrin's rule-out result concerns CVRV rather than generic vasospasm; see Section 3.6.

Author (Year)	**Target Outcome**	**Reported Model**	**Performance**
Longitudinal Monitoring Approaches			
Sen et al. (2025)	ACV	Admission Logistic Reg.	ACC: 0.736 F1: 0.566 TPR: 0.683 PPV: 0.483
Roederer et al. (2014)	ACV (IA-treated)	Naive Bayes	AUC: 0.708 MSE: 0.26 TPR: 0.600 PPV: 0.677
Zarrin et al. (2024)	CVRV	LightGBM	Binary AUC: 0.88 Rule-out Sens: 1.00 (selected threshold)

Continued on next page

Table 3.1 continued from previous page

Author (Year)	Target Outcome	Reported Model	Performance
Predominantly Fixed Patient-Level Representations			
Sadri et al. (2024)	Vasospasm (requiring IA Tx)	SVM	AUC: 0.86 ACC: 0.77 PPV: 0.78 NPV: 0.76
Capoglu et al. (2019)	ACV	Logistic Reg.	AUC: 0.93 ACC: 0.85
Colip et al. (2021)	ACV	LASSO	AUC: 0.74 OR: 3.4
Gollwitzer et al. (2025)	Moderate/severe ACV	Logistic Reg.	AUC: 0.723 F1: 0.675 TPR: 0.660 TNR: 0.652 PPV: 0.700 NPV: 0.620
Kim et al. (2021)	ACV	RF Regressor	AUC: 0.88 ACC: 0.85
Scherer et al. (2020)	ACV	Univariate LR	AUC: 0.74
Urbanos et al. (2025)	Vasospasm (study-defined)*	Ensemble	AUC: 0.66 ACC: 0.66 TPR: 0.66 TNR: 0.62
Yang et al. (2025)	TCD-defined vasospasm	Random Forest	AUC: 0.966 TPR: 0.917 TNR: 0.912
Dumont et al. (2011)	SCV	ANN	AUC: 0.960
Dumont (2016)	SCV	ANN (External Validation)	NPV: 1.00 TPR: 1.00 TNR: 0.84 PPV: 0.67
Lintas et al. (2024)	SCV	SVM	ACC: 0.86

Continued on next page

Table 3.1 continued from previous page

Author (Year)	Target Outcome	Reported Model	Performance
Alexopoulos et al. (2022)	DCI (clinical & infarction)	XGBoost	ACC: 0.843 F1: 0.684 NPV: 0.977 TPR: 0.901 TNR: 0.831 PPV: 0.527
Azzam et al. (2024)	DCI (clinical)	MLP Neural Net	ACC: 0.93 F1: 0.85 AUC: 0.83
Kim et al. (2023)	DCI (clinical OR infarction)	Logistic Reg.	AUC: 0.95 ACC: 0.92 TPR: 0.80 TNR: 0.95
Savarraj et al. (2021)	DCI	ANN	ACC: 0.80 AUC: 0.75 TPR: 0.82 TNR: 0.72
Palmisciano et al. (2023)	DCI (clinical & infarction)	XGBoost (External Validation)	ACC: 0.648 NPV: 0.798 F1: 0.288 TPR: 0.328 TNR: 0.737 PPV: 0.257

* For Urbanos et al. (2025), the Methods describe clinical and radiological criteria for vasospasm, whereas the Results distinguish clinical and radiological cases; the precise relationship between these descriptions in outcome labelling is therefore treated as unclear. Additional endpoint details and source limitations are reported in Table A.1.

3.4.1 Risk of Bias and Quality of Evidence

The recorded PROBAST assessment (Table 3.2) classified 18 of the 19 studies as having High Risk of Bias. Applicability Concern was High in 16 studies and Unclear in three. All 19 studies therefore had a High rating in at least one of these two overall assessments. Risk of bias and applicability are considered separately because they address different questions.

3.4.1.1 Internal Validity: Sources of Risk of Bias

Zarrin et al. (2024) was the only study assigned Low Risk of Bias across all four domains in the recorded assessment. This judgement is distinct from the interpretation of its cross-centre evaluation, discussed in Section 3.4.2.

Table 3.2: Consolidated PROBAST Assessment: Risk of Bias and Applicability Concerns across the 19 included studies. Risk of Bias assesses internal validity, while Applicability evaluates relevance to the target clinical question.

	Risk of Bias (Internal Validity)					Applicability Concern			
Study	**Part.**	**Pred.**	**Out.**	**Anal.**	**Overall**	**Part.**	**Pred.**	**Out.**	**Overall**
Low Overall Risk of Bias									
Zarrin et al. (2024)	+	+	+	+	**Low**	-	+	+	**High***
Participant-Domain Concerns									
Palmisciano et al. (2023)	-	+	+	+	High	-	+	+	High
Urbanos et al. (2025)	-	+	+	+	High	-	+	+	High
Sen et al. (2025)	-	+	+	+	High	-	+	+	High
Savarraj et al. (2021)	-	+	+	+	High	-	+	+	High
Kim et al. (2023)	-	+	+	+	High	-	+	+	High
Alexopoulos et al. (2022)	-	+	+	+	High	-	+	+	High
Analysis-Domain Concerns									
Gollwitzer et al. (2025)	+	+	+	-	High	?	+	+	Unclear
Colip et al. (2021)	+	+	+	-	High	?	+	+	Unclear
Kim et al. (2021)	+	+	+	-	High	?	+	+	Unclear
Multiple-Domain Concerns									
Roederer et al. (2014)	-	+	+	-	High	-	+	+	High
Azzam et al. (2024)	-	?	+	-	High	-	+	+	High
Yang et al. (2025)	-	+	-	-	High	-	-	-	High
Scherer et al. (2020)	-	+	+	-	High	-	+	+	High
Dumont (2016)	-	+	+	-	High	-	+	+	High
Dumont et al. (2011)	-	+	+	-	High	-	+	+	High
Lintas et al. (2024)	-	+	?	-	High	-	+	?	High
Sadri et al. (2024)	-	+	+	-	High	-	-	+	High
Capoglu et al. (2019)	-	?	+	-	High	-	-	+	High

Legend: (+) Low Risk/Concern; (-) High Risk/Concern; (?) Unclear Risk/Concern.
* Zarrin was assessed as having Low Risk of Bias but High Applicability Concern due to population mismatch. Its cohorts included children and traumatic as well as non-traumatic SAH.
The Analysis domain is not assessed for Applicability under PROBAST.

The remaining assessments identified concerns in the following domains:

- **Selection Bias (Participants Domain):** Studies including Palmisciano et al. (2023) and Urbanos et al. (2025) were assigned High Risk of Bias in the Participants domain. Relevant concerns included exclusion criteria that could affect sample representativeness, such as exclusion of patients who died early or lacked particular imaging. These choices can introduce selection or survivor bias; retrospective or single-centre design alone does not establish its magnitude.

- **Analytical Limitations (Analysis Domain):** Studies including Roederer et al. (2014) and

Dumont (2016) received High Risk of Bias judgements in the Analysis domain. The review identified concerns about sample adequacy and the extent of validation. These concerns warrant caution about optimistic performance estimates, rather than establishing that overfitting occurred in every study.

3.4.1.2 Applicability and Transportability

High Applicability Concern was common in the recorded assessment, but it should not be interpreted as direct evidence that a model failed external validation.

- **Population Mismatch:** Zarrin et al. (2024) included mixed-age cohorts and both traumatic and non-traumatic SAH. These populations differ from the adult aSAH focus of this dissertation. Their findings remain relevant, but their applicability to the narrower target population requires qualification.
- **Institutional Context:** For single-centre studies such as Sen et al. (2025) and Kim et al. (2023), local patient characteristics and clinical protocols may affect performance elsewhere. Limited evaluation in other settings leaves transportability uncertain; it does not prove institutional overfitting.

Overall confidence in the evidence was judged low in the qualitative appraisal. Reported predictive performance must be interpreted alongside the recorded risk of bias, applicability concerns and validation design. The studies provide evidence about predictive potential, with limited evidence of clinical benefit from deploying these models.

3.4.2 Temporal Resolution and Representation

Most studies used predominantly fixed patient-level representations, with their input modalities discussed in Section 3.4.4. Three studies explicitly analysed longitudinal physiological monitoring. Their approaches illustrate differences between daily aggregation, cumulative aggregation and sequence-preserving modelling.

3.4.2.1 Daily and Early-Period Statistical Summaries

Roederer et al. (2014) represented routinely recorded physiological measurements, including ICP, MAP and HR, together with laboratory information through statistical summaries rather than ordered sequences. The authors calculated six statistics: Mean, Median, Minimum, Maximum, Standard Deviation and Coefficient of Variation. These were calculated for each of the first three individual days of illness and for the cumulative first-three-day period.

- **Representation:** The Coefficient of Variation (σ/μ) describes relative variability alongside measures of magnitude.

- **Model Comparison:** The Naive Bayes classifier trained on automated monitoring features achieved a higher reported AUC than Logistic Regression using that feature set (0.708 vs 0.544). Naive Bayes and Logistic Regression were selected as a generative/discriminative pair. Their results support a study-specific comparison, without establishing an intrinsic advantage of one model class for physiological data.

Models were evaluated using ten-fold cross-validation and feature-subset ablation. The report does not clearly establish whether outcome-based feature selection and mean imputation were repeated within each training fold.

3.4.2.2 Cumulative Aggregation

The largest multicentre study in the synthesis, Zarrin et al. (2024), processed 172 routinely collected intensive-care variables. Its prospective prediction setup used cumulative data from the start of ICU admission, with models evaluated using the first 4 hours and 1, 3, 5, 7 and 10 days of data. LightGBM (Light Gradient Boosting Machine) was compared with Logistic Regression.

For each variable, feature vectors summarised the data available by the prediction time. ICP and MAP measurements were represented by percentiles from the 5th to the 100th in five-percentage-point increments. Laboratory measurements were summarised using the minimum, maximum, median, mean and number of measurements. This representation retains distributional information but not the order of observations within each cumulative period.

At the University of California, Los Angeles (UCLA), binary prospective LightGBM performance ranged from an AUC of 0.87–0.88 using the first 4 hours of ICU data to 0.68–0.71 using 10 days, depending on whether the conservative or institutional predictor set was used. Patients who had already reached an endpoint were excluded at subsequent prediction times, producing progressively smaller prediction cohorts. The performance difference therefore does not, by itself, establish physiological signal decay or an optimal observation duration. Institutional and conservative predictor sets differed by an average of 0.00–0.03 AUC points, with no statistically significant differences reported ($p > 0.05$).

The prospective trinary task distinguished no Cerebral Vasospasm Requiring Verapamil (CVRV), CVRV within three days and CVRV after three days of prediction. Across the assessed time points, reported AUCs were 0.73–0.83 for events within three days and 0.50–0.88 for later events. These ranges describe different horizons and do not establish a consistent ranking between them. LightGBM achieved higher reported AUCs than Logistic Regression at all assessed time points.

The authors described their Vanderbilt University Medical Center (VUMC) analysis as external validation. However, operationally, the UCLA modelling code was transferred to VUMC and a separate model was locally trained and tested using the conservative predictor set and the prospective trinary task. Performance averaged 0.01 AUC points below the UCLA results. The analysis therefore supports cross-centre reproducibility of the modelling approach, while differing from external validation based on transport of an unchanged UCLA-fitted model.

3.4.2.3 Sequence-Preserving Modelling

Sen et al. (2025) used Long Short-Term Memory (LSTM) networks to represent the temporal order of monitoring data ($t_1 \rightarrow t_2 \rightarrow \cdots \rightarrow t_n$).

- **Representation:** The recurrent model retains information from earlier time steps, enabling the use of sequential patterns rather than relying solely on summary statistics.
- **Input Modality:** The LSTM combined hourly physiological measurements (ICP, CPP and MAP) and ventriculostomy output with daily TCD measurements, including velocities and Lindegaard ratios, serum sodium, and additional clinical and radiographic information.

Fixed-duration moving windows generated multiple examples per patient, with all windows from each patient retained in the same partition of an 80/20 training–test split. The primary task predicted angiographic vasospasm at any subsequent time after the observation window.

Among the three longitudinal monitoring studies, this was the sequence-preserving approach. Its comparison with the admission-based Logistic Regression (LR) Baseline Model illustrates a trade-off between reported metrics. The LSTM achieved higher accuracy (0.786 vs 0.736) and precision (0.683 vs 0.483), while LR achieved higher sensitivity (0.683 vs 0.277) and F1-score (0.566 vs 0.394). Accuracy alone is insufficient to interpret this comparison, particularly with an imbalanced outcome. The baseline identified a greater proportion of positive cases, but these results do not establish clinical superiority or show that either architecture inherently captures physiological trajectories better.

The TCD-only ablation increased precision from 0.683 to 0.824 but reduced recall from 0.277 to 0.139 and F1-score from 0.394 to 0.237, indicating a different operating trade-off rather than an overall improvement.

3.4.3 Outcome-Dependent Performance

The clinical, sonographic and angiographic domains described in Section 2.2.1 concern complementary forms of assessment, whereas the outcome labels used in this review identify study-specific prediction endpoints. An endpoint may combine clinical and vascular findings or impose a particular severity or treatment requirement. Historical SCV terminology is retained where used by the source studies, while DCI is considered a distinct, adjacent prediction outcome. The labels summarised in Tables 3.1 and A.1 therefore do not imply uniform diagnostic criteria within each group. Differences in populations, predictor information and validation additionally limit direct performance comparisons.

3.4.3.1 Angiographic and Sonographic Vasospasm

Angiographic vasospasm (ACV) describes vessel narrowing identified by vascular imaging. TCD-defined vasospasm is a sonographic endpoint based on blood-flow velocity criteria. These related assessments are distinguished here rather than treated as interchangeable measurements.

Anatomical and Biological Features Capoglu et al. (2019) reported an AUC of 0.93 using sparse dictionary learning to represent 3D vessel geometry from admission digital subtraction angiography (DSA). The representation captured covariance of local vascular structures beyond simple diameter measurements. The small cohort ($N = 20$) limits the strength and generalisability of this proof-of-concept result and warrants caution about optimistic performance estimates.

Yang et al. (2025) reported an AUC of 0.966 in 175 patients for TCD-defined vasospasm. The outcome was assessed by serial TCD, using middle cerebral artery velocity criteria, rather than angiographic confirmation. The predictors combined GCS with serum HIF-1α, VEGF and ET-1 measured 24 hours after surgery. This supports discussion of postoperative biomarker-based prediction, but does not establish that biomarkers caused the high discrimination or demonstrate the value of temporally resolved physiological modelling.

Clinical Scores and Additional Features Scherer et al. (2020) ($N = 125$) reported an AUC of 0.74 for the SAH/CSF volumetric ratio, compared with 0.50 for the Fisher grade. In Colip et al. (2021) ($N = 195$), a LASSO model combining clinical and CT/CTA features, including arterial calcification and undulation, achieved a c-index of 0.74, significantly higher than the modified-Fisher-only model (0.61; $p < 0.001$). This comparison concerns the combined model, not the contribution of arterial undulation alone.

Gollwitzer et al. (2025) predicted moderate or severe angiographic vasospasm, with mild narrowing grouped with absent vasospasm. The reported AUC was 0.723 for Logistic Regression and 0.720 for a Support Vector Machine using clinical information. Kim et al. (2021) reported an AUC of 0.88 with a Random Forest Regressor using inputs including aneurysm characteristics and clinical grades. These results show variation in the performance reported for clinical-feature models; differences between studies cannot establish that non-linearity alone accounts for the higher value or that a universal performance ceiling exists.

Treatment Requirement as an Endpoint Some studies incorporated treatment into the outcome definition rather than predicting vascular narrowing alone. Roederer et al. (2014) required angiographically identified vasospasm followed by endovascular treatment. In Zarrin et al. (2024), CVRV was operationalised by first verapamil administration, and the institutional binary model achieved an AUC of 0.88 using the first 4 hours of ICU data at UCLA. Sadri et al. (2024) reported an AUC of 0.86 for vasospasm requiring intra-arterial treatment using an SVM and imaging-derived features, but the available abstract did not specify the diagnostic confirmation criteria. These endpoints concern treatment-linked events within the respective clinical pathways, and their confirmation may depend on institutional intervention criteria as well as disease severity.

3.4.3.2 Study-Defined Symptomatic Vasospasm and DCI

Historical SCV and DCI studies are discussed together for organisation, but their endpoints remain distinct. As described in Sections 2.2.1 and 2.2.3, neurological deterioration associated with vascular evidence of spasm is not equivalent to all DCI, which can arise through several interacting

mechanisms. Differences in operational definitions and assessment also limit comparisons within these groups.

Symptomatic Vasospasm (SCV): Predictor Timing and Outcome Definition Three included studies used the designation SCV for their prediction endpoint. Dumont et al. (2011) developed an Artificial Neural Network (ANN) using 91 retrospective patients and evaluated it in 22 prospective patients. In the 22-patient prospective test cohort, the ANN achieved an AUC of 0.960, compared with 0.933 and 0.897 for the Adams and Qureshi regression-based models, respectively.

Elevated TCD on or before day five after ictus was included as a predictor, defined by a Lindegaard ratio > 3 or, when ratios were unavailable, velocity > 120 cm/s. The SCV endpoint required neurological change with corresponding vascular evidence of spasm after exclusion of alternative causes. TCD velocity > 180 cm/s could provide vascular evidence when angiography was not performed. This overlap warrants attention to predictor timing and outcome assessment. However, using an earlier measurement of a related physiological quantity does not by itself establish leakage, and the available evidence does not quantify any inflation of performance attributable to this overlap.

Dumont (2016) subsequently evaluated the previously developed model prospectively at a different institution in 25 patients, with predictor collection described at admission. The study retained SCV as the target and explicitly described it as the same endpoint used in the earlier publication: delayed focal neurological deterioration accompanied by corresponding radiographic evidence of vasospasm. However, the operational description of that radiographic evidence differed from the earlier report. In the 2016 study, vasospasm was defined by a Lindegaard ratio > 6 in patients without angiographic treatment, or by angiographically confirmed vessel narrowing in patients undergoing intervention. These criteria should therefore not be assumed to be numerically identical to the TCD thresholds reported in the 2011 study.

Lintas et al. (2024) reported accuracy ranging from 0.61 to 0.86 using a Support Vector Machine (SVM) with clinical grades, including Hunt & Hess, Fisher and the BNI scale, and other recorded features. This result describes another fixed-feature approach to SCV prediction. Its comparison with the Dumont studies cannot isolate the contribution of TCD because the cohorts and assessment designs differ.

Delayed Cerebral Ischaemia (DCI): Performance Across Cohorts Five included studies predicted study-defined DCI. In Alexopoulos et al. (2022) and its external validation by Palmisciano et al. (2023), a positive outcome required neurological deterioration together with new cerebral infarction on CT or MRI. Alexopoulos et al. (2022) reported an XGBoost model with accuracy of 0.843 and F1-score of 0.684. Its training used ROSE resampling to address class imbalance.

In the external evaluation by Palmisciano et al. (2023), accuracy was 0.648, F1-score was 0.288 and sensitivity was 0.328, corresponding to a false-negative rate of approximately 67%. This deterioration supports caution about transportability. It does not identify oversampling, overfitting

or any particular predictor as its cause; differences in patient characteristics, clinical management and other aspects of the evaluation may contribute.

Kim et al. (2023) reported mean AUCs of 0.95 ± 0.04 for Logistic Regression and 0.94 ± 0.03 for XGBoost. These estimates indicate comparable discrimination with a numerically higher mean for Logistic Regression. Cilostazol use was included among the predictors, but the retrospective treatment comparison and predictive explanations do not establish that this feature caused the high AUC or that its model contribution is a causal treatment effect. The paper uses the term delayed cerebral infarction while defining its DCI endpoint to include new infarction or otherwise unexplained neurological deterioration; this study-specific definition is relevant to interpretation.

Using a clinical DCI definition without requiring demonstrated cerebral infarction, Azzam et al. (2024) reported an AUC of 0.85 for Random Forest (RF) and accuracy of 0.93 with F1-score of 0.85 for the multilayer perceptron (MLP) model. Clinical vasospasm was an important predictor, and the study reported its occurrence 48 hours before DCI. This is a study-specific temporal association, not evidence that all DCI follows vasospasm or that the two outcomes are interchangeable.

Synthesis Across Distinct Clinical Endpoints The SCV work (Dumont et al. 2011) and DCI studies (Kim et al. 2023; Alexopoulos et al. 2022) illustrate the use of early clinical information, physiological measurements and treatment-related variables for distinct clinical prediction tasks. Their findings do not establish a shared performance ceiling for SCV and DCI, or show that one type of predictor is inherently more transportable. The external evaluations by Palmisciano et al. (2023) and Dumont (2016) differ in endpoint, population and sample size. Together, they support attention to outcome definition, predictor availability and evaluation in other settings rather than a universal ranking of input modalities.

3.4.4 Input Modality Analysis

Across the 19 studies, predictors ranged from readily available clinical scores to imaging-derived features, serum biomarkers and physiological monitoring. These modalities provide different kinds of information, but their heterogeneous evaluations do not establish a universal hierarchy of predictive utility.

Prevalent Predictors: The Clinical Baseline Common features included the **Fisher Grade**, **Hunt & Hess (H&H) scale** and **Glasgow Coma Scale (GCS)**. These variables featured in studies including Alexopoulos et al. (2022), Gollwitzer et al. (2025), Lintas et al. (2024) and Palmisciano et al. (2023). Their routine availability in electronic health records (EHR) makes them practical baseline predictors. Their reported performance varied, and fixed clinical grades do not explicitly describe subsequent physiological trajectories.

Additional Predictors: Radiomics and Biomarkers A subset of studies used more detailed anatomical or biological features:

- **Advanced Radiomics:** Capoglu et al. (2019) represented the covariance of local 3D vessel geometry from DSA. Sadri et al. (2024) and Urbanos et al. (2025) used high-dimensional features derived from admission CT, including representations of haemorrhage or brain tissue. These features describe aspects of imaging information beyond categorical clinical grades.
- **Specific Biomarkers:** Yang et al. (2025) used serum **HIF-1α**, **VEGF** and **ET-1**, measured 24 hours after surgery, alongside GCS to predict TCD-defined vasospasm. These inputs differ from routinely collected laboratory markers, such as WBC, used in studies including Savarraj et al. (2021).

Treatment-Related and Physiological Predictors Kim et al. (2023) included **cilostazol use** in predictive modelling of its study-defined DCI endpoint, reporting an AUC of 0.95 for Logistic Regression. This represents treatment-related predictive information in an observational dataset, not a demonstrated causal contribution to model performance. Dumont et al. (2011) used **TCD elevation** among the predictors of SCV. Its relevance must be interpreted alongside the timing and outcome-assessment overlap discussed in Section 3.4.3.2; the study does not establish it as the single strongest predictor across the reviewed literature.

Categorical Scores and Quantitative Imaging Scherer et al. (2020) reported an AUC of 0.50 for Fisher grade and 0.74 for the SAH/CSF volumetric ratio. Colip et al. (2021) reported a significantly higher c-index for a combined clinical and CT/CTA model than for modified Fisher grade alone (0.74 vs 0.61; $p < 0.001$). These findings support the value of evaluating more detailed representations alongside categorical grades in those settings. They do not establish that categorisation invariably removes all useful prognostic information.

3.5 Synthesis of Illustrative Modelling Approaches

Three studies illustrate complementary aspects of the reviewed evidence: the scale of longitudinal monitoring, postoperative biomarker information and treatment-related predictors. They are considered as methodological examples with different strengths and limitations, rather than ranked as directly comparable models. The sample-size and outcome-assessment concerns discussed in Section 3.4.3 remain relevant when interpreting high reported performance.

3.5.1 Longitudinal Monitoring Across Centres: Zarrin et al. (2024)

Zarrin et al. (2024) studied 3,404 patients across two centres (1,750 at UCLA and 1,654 at VUMC), providing a large-scale example of routinely collected ICU data used for prediction of verapamil requirement.

- **Strength:** Cumulative aggregation produced an AUC of 0.88 using the first 4 hours of ICU data at UCLA. The approach was also evaluated through separate model training and testing at VUMC.
- **Limitation:** The second-centre analysis evaluated reproducibility of the modelling approach rather than transport of an unchanged model. Treatment requirement may reflect local intervention criteria, and the mixed-age, mixed-aetiology cohorts differ from the adult aSAH focus of this dissertation.

3.5.2 Postoperative Biomarkers: Yang et al. (2025)

Yang et al. (2025) reported an AUC of 0.966 ($N = 175$) for TCD-defined vasospasm using a Random Forest model combining GCS with serum biomarkers, including HIF-1α, VEGF and ET-1, collected 24 hours after surgery.

- **Strength:** The study provides an example of combining clinical assessment with postoperative biological information for sonographic vasospasm prediction.
- **Limitation:** Its single-centre design and sample size limit generalisation. The reported result does not isolate the contribution of biomarkers relative to model choice or establish that similar performance would be obtained with other endpoints or observation periods.

3.5.3 Treatment-Related Information: Kim et al. (2023)

Kim et al. (2023) reported mean AUCs of 0.95 ± 0.04 for Logistic Regression and 0.94 ± 0.03 for XGBoost in 321 patients. The comparison supports retaining a simple baseline when evaluating more complex models.

- **Strength:** The study examined treatment-related information, including cilostazol use, alongside other predictors of its study-defined DCI endpoint.
- **Limitation:** Associations in this retrospective dataset and model explanations should not be interpreted as causal treatment effects. Predictor relevance and model performance may differ in centres using other pharmacological protocols.

3.6 Clinical Utility and Methodological Implications

Predictive performance is relevant to potential clinical use, but discrimination alone does not establish patient benefit or safe changes in monitoring. This section considers the role of input information, the distinction between fixed and longitudinal representations, and the interpretation of rule-out metrics.

3.6.1 The Dengler Paradox as a Conceptual Interpretation

Early work such as Toledo et al. (2009) explored ML for outcome prediction after SAH. As explained in Section 2.5.1, the term "Dengler Paradox" is used in this dissertation as a conceptual label for the possibility that increasing algorithmic complexity may offer limited benefit when input information is constrained. It is not established terminology or a universal law confirmed by this review.

The reviewed comparisons provide mixed evidence. Gollwitzer et al. (2025) reported similar AUCs for Logistic Regression and a Support Vector Machine, while Lintas et al. (2024) reported accuracy, rather than an AUC that could support a shared discrimination threshold. Other studies, including Kim et al. (2021), reported higher discrimination using clinical-feature models. These observations do not support an invariant AUC ceiling.

- Yang et al. (2025) reported high discrimination with a combination of clinical and postoperative biomarker information, but did not establish that biomarkers rather than model choice caused that result.
- Capoglu et al. (2019) reported high discrimination using 3D vascular representations in a small cohort, supporting further evaluation of detailed anatomical inputs rather than establishing a broadly validated advantage.

The defensible implication is to evaluate input representations alongside modelling methods and retain appropriate clinical baselines. These findings motivated the subsequent experimental design without determining whether temporal information would improve performance.

3.6.2 Static and Longitudinal Representations: Reported Trade-Offs

The reviewed studies address different prediction tasks and do not establish universal accuracy or clinical-utility winners.

Fixed Patient-Level Representations Some of the highest reported AUCs came from fixed feature representations, including Yang et al. (2025) and Kim et al. (2023). Their inputs include postoperative or treatment-related information, so these results should not all be described as admission-only prediction. Such models provide patient-level risk estimates, but their discrimination does not establish an appropriate clinical intervention or resource-allocation policy.

Longitudinal Monitoring Representations Zarrin et al. (2024) evaluated predictions at several times from ICU admission, while Sen et al. (2025) preserved temporal order through sequential modelling. These approaches illustrate ways to use information accumulated during hospitalisation. However, longitudinal data, sequence preservation and repeated prediction updating are distinct design choices. The reported studies do not establish that any one choice necessarily improves clinical outcomes.

In particular, the AUC of 0.88 reported by Zarrin et al. (2024) used the first 4 hours of ICU data. The changing cohorts at later prediction times preclude interpreting the performance trend as direct evidence of physiological signal decay. The result motivates consideration of early monitoring information without validating a specific observation duration or the hypothesis of this dissertation in advance.

3.6.3 Potential Rule-Out Value and Its Limits

False negatives are important when assessing a proposed monitoring aid, but false positives can also lead to additional investigations or interventions. Sensitivity and Negative Predictive Value (NPV) must therefore be interpreted with specificity, Positive Predictive Value (PPV), the decision threshold and the predicted endpoint. NPV also depends on outcome prevalence in the evaluated population.

Several studies reported results relevant to potential rule-out use:

- **Zarrin et al. (2024):** At a threshold with no observed false negatives, the model ruled out verapamil requirement in approximately 8% of patients who did not require it. This limited rule-out proportion is essential context for the reported sensitivity of 100%. The endpoint was verapamil requirement, not all subsequent vasospasm, and zero observed false negatives does not guarantee future safety.
- **Dumont (2016):** In the 25-patient prospective external cohort, all 16 negative predictions were correct (NPV 100%). Six patients developed SCV, and all six were identified; three other patients had false-positive predictions (PPV 67%). Thus, the model did not identify every patient without SCV as low risk, and the small sample limits generalisation.
- **Alexopoulos et al. (2022):** The model reported NPV of 97.7% and PPV of 52.7% for DCI. These values concern absence or occurrence of that specific endpoint. The lower performance in the external evaluation by Palmisciano et al. (2023) further limits claims of reliable rule-out use across settings.

These findings support investigation of rule-out applications under clearly defined conditions. They do not establish that the reviewed models can safely replace clinical monitoring or provide a reliable bedside safety system.

3.7 Gap Analysis: Temporal Representation of Early Physiological Monitoring

Within the included corpus, relatively few studies explicitly analysed longitudinal neurocritical monitoring, and they used different representations of the available information:

1. **Sequence-Preserving Representation:** Sen et al. (2025) used LSTM networks with hourly physiological measurements and daily TCD data, preserving their temporal order at the represented resolution.

2. **Cumulative Statistical Summaries:** Zarrin et al. (2024) summarised observations accumulated from ICU admission to successive prediction times. Distributional features retained magnitude and variability while discarding within-window temporal order.

3. **Daily and Early-Period Summaries:** Roederer et al. (2014) used individual-day and cumulative first-three-day statistics to describe physiological measurements without modelling their full sequence.

The central gap is the limited evaluation of temporally resolved physiological monitoring and derived indices, particularly the distinction between summarising an early monitoring period and preserving its temporal structure. The reviewed evidence also leaves scope to assess whether physiological information adds predictive value beyond static clinical variables. Neither the value of derived indices nor the benefit of sequence preservation can be assumed from the absence of these comparisons in the corpus.

Early Monitoring as an Observation Period Zarrin et al. (2024) reported predictive information using the first 4 hours of ICU data. This provides a rationale for investigating early observations, but differences in performance across changing prediction cohorts do not establish an optimal duration. In the present dissertation, the first 24 hours define the physiological observation period; they should not be interpreted as a fixed 24-hour prediction lead time.

Rationale for the Subsequent Study The review informed the evaluation of four linked sources or representations of information:

1. **Static Clinical Information:** Patient-level clinical variables used to establish a baseline for vasospasm prediction.

2. **Aggregated First-24-Hour Physiology:** Patient-level summaries of physiological monitoring during the first 24 hours.

3. **Temporally Preserved First-24-Hour Physiology:** The same monitoring period represented at minute-level resolution while preserving temporal order.

4. **Static–Temporal Integration:** Combination of static clinical and temporally represented physiological information to assess whether the physiological component provides additional predictive information.

This comparison addresses how the representation of early physiological monitoring relates to prediction of cerebral vasospasm. It provides a rationale for the experimental design without presupposing that temporal preservation or integration improves performance.

Chapter 4

Methodology

This chapter describes the retrospective prediction-model development and internal-validation study conducted to assess whether routinely collected clinical and neurophysiological information could support the prediction of cerebral vasospasm after aSAH. The methodology followed the incremental structure introduced in Chapter 1: static admission and early clinical information was evaluated first, followed by first-24-hour physiology represented as patient-level summaries, the same physiology represented as minute-level time series, and, finally, late fusion of static and temporal predictions. This progression was designed to separate the contribution of information source and temporal representation; it was not intended to establish that any representation was intrinsically preferable.

4.1 Study Design and Analytical Framework

The study was a single-centre, retrospective model-development study with internal validation. The primary methodological question was whether a clinician-assigned vasospasm outcome could be predicted from admission or early clinical information and, within patients with eligible physiological monitoring, whether information from the first 24 hours of ICM+ recording added predictive information beyond the static clinical representation. Four linked analytical phases were defined, as summarised visually in Figure 4.1 and in greater detail in Table 4.1.

Phase 1 intentionally used the broader eligible clinical cohort rather than restricting static modelling to patients with available physiological monitoring. Eligibility for a static clinical model did not require an ICM+ recording, and restricting this phase to the monitored subset would therefore have discarded otherwise usable clinical observations solely because a separate data modality was unavailable. The broader cohort also provided substantially more observations for development and internal validation of the static models. Phase 1 consequently characterised static clinical prediction in the broader available clinical population and treated the admission episode as the unit of analysis.

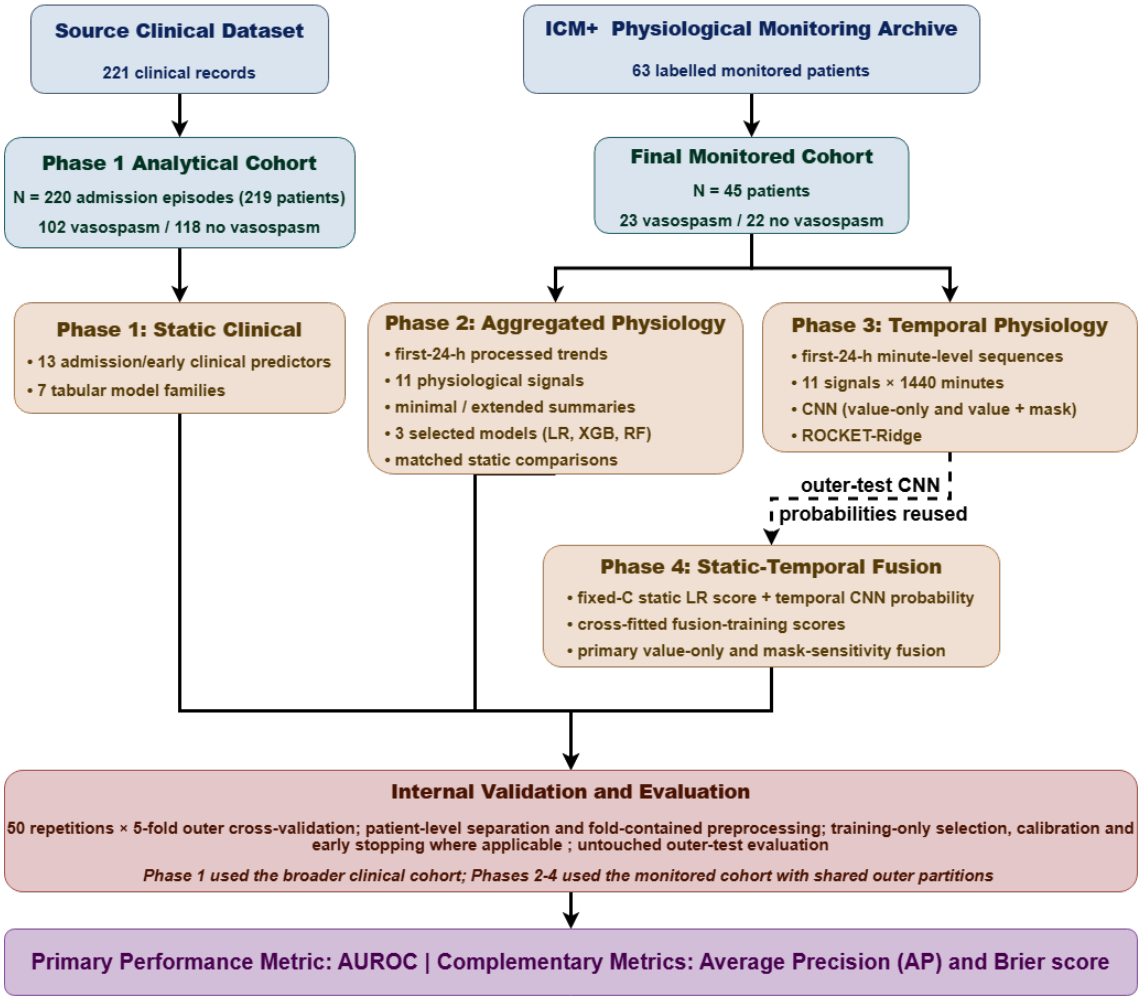


Figure 4.1: **Overview of the four-phase analytical framework.** The flowchart summarises the source clinical and physiological datasets, construction of the Phase 1 and monitored analytical cohorts, the four modelling phases, and the internal-validation and evaluation framework. Phase 1 used the broader clinical cohort, whereas Phases 2–4 used the same 45-patient monitored cohort with shared outer partitions. Phase 4 combined a fixed-C static logistic-regression score with temporal CNN probabilities, with the corresponding Phase 3 outer-test probabilities reused in the late-fusion analysis.

Table 4.1: Overview of the four-phase analytical framework.

Phase	Purpose	Cohort and unit	Input representation
1	Establish a static clinical reference in the broader eligible clinical cohort	220 admission episodes from 219 patients	Thirteen admission/early clinical predictors
2	Assess aggregated physiology and its incremental contribution within the monitored cohort	45 monitored patients	Static predictors, first-24-hour physiological summaries, and their combinations
3	Assess physiology while preserving temporal order	The same 45 monitored patients	Eleven signals on a common 1,440-minute grid, with an observation-mask sensitivity analysis
4	Assess static–temporal integration	The same 45 monitored patients	Two-feature late-fusion stack with cross-fitted static and temporal training scores and reused Phase 3 outer-test CNN probabilities

Phases 2–4 instead used the selected monitored cohort and treated the patient as the unit of analysis. Consequently, Phase 1 and the later phases address related but non-interchangeable populations, and differences in performance between them were not interpreted as direct evidence of the relative value of static and physiological information. Direct assessment of the incremental contribution of physiology was performed in Phase 2, where static-only, physiology-only, and combined models were evaluated within the same 45-patient monitored cohort. The same 45 patients, target labels, and outer-fold assignments were used throughout Phases 2–4 so that comparisons within the monitored cohort were made under matched partitions.

4.2 Clinical Setting, Data Sources and Governance

The study was conducted in collaboration with the Neurocritical Care Unit of Unidade Local de Saúde São João (ULSSJ) in Porto, Portugal.

The clinical research team supplied a curated clinical dataset covering baseline characteristics, SAH presentation and severity, in-hospital complications and management, neuromonitoring, vasospasm diagnosis and timing, and longitudinal clinical outcomes. For the present work, only variables relevant to baseline and early clinical prediction, together with vasospasm status, timing, and key event dates, were retained.

The source dataset comprised 221 clinical records from patients with confirmed or presumed aneurysmal subarachnoid haemorrhage (SAH) admitted between 01/01/2018 and 31/12/2022. This included non-traumatic, angiography-negative SAH cases that had been clinically classified by the research team as presumed aneurysmal SAH despite the absence of angiographic confirmation. Traumatic SAH and haemorrhages associated with arteriovenous malformations were excluded.

Clinical input was also used to clarify the interpretation of the outcome and of the routinely collected variables. Physiological recordings originated from the neurocritical-care monitoring archive and were stored, inspected and exported through ICM+. Only monitoring records that could be matched to a valid clinical outcome were eligible for supervised modelling.

The study protocol was submitted to and approved by the Ethics Committee of Unidade Local de Saúde São João. The retrieval, storage and subsequent handling of the clinical and physiological data used in this dissertation were performed within the scope of that approval and according to the corresponding institutional data-governance procedures. Data used for analysis were pseudonymised through study identifiers, and no direct patient identifiers were used as model predictors.

4.3 Outcome Definition and Temporal Framework

The modelling target was the clinician-assigned binary diagnosis of cerebral vasospasm. At ULSSJ, vasospasm surveillance is based primarily on sonological assessment using TCD, with predefined velocity-based criteria and Lindegaard-ratio interpretation. The local surveillance routine also varies according to the patient's clinical and sonological status, with more frequent examinations when vasospasm is suspected or present. Angiographic imaging may subsequently be used to further confirm or characterise the diagnosis where clinically indicated. The detailed institution-specific TCD thresholds and surveillance criteria are not reproduced in this dissertation. Although the clinical assessment can distinguish different degrees of vasospasm severity and hyperaemia, the dataset used in the present study provided a frozen binary outcome, Vasospasm or No Vasospasm, and no attempt was made to reconstruct or re-adjudicate severity categories from the underlying examinations.

The clinical registry and physiological monitoring data were anchored to different time origins. Clinical records were available from hospital admission, whereas intensive ICM+ monitoring did not necessarily begin at admission. In practice, this monitoring was initiated when the patient's neurocritical condition warranted intensive physiological surveillance, such that the beginning of continuous monitoring could occur later in the clinical course. Consequently, hospital admission and ICM+ monitoring initiation were treated as distinct temporal reference points.

Phase 1 was based exclusively on static admission/early-clinical information. It therefore did not use a duration-based physiological observation window; its predictor information was anchored to the admission baseline.

For Phases 2–4, let T_0 denote the start of ICM+ monitoring. The physiological observation window was defined as the half-open interval

$$[T_0,\ T_0 + 24\ \text{h}),$$

with the prediction landmark located at $T_0 + 24$ h. Phase 2 combined the static admission/early-clinical covariates with physiological features aggregated over this 24-hour monitoring interval.

It should therefore be understood as combining a baseline clinical snapshot with a 24-hour physiological observation window, rather than as having a single homogeneous observation window covering all predictors. Phase 3 used the same 24-hour physiological interval while preserving its minute-level temporal structure, and Phase 4 combined the static clinical representation with the temporally represented physiological information.

The prediction period began after the 24-hour landmark. Because no fixed post-landmark prediction horizon was imposed, the target was defined as any subsequent vasospasm recorded in the clinical dataset.

For patients with vasospasm, the clinical dataset recorded event timing only as an integer day d counted from the aSAH event. Let D_0 denote midnight at the beginning of the recorded aSAH event date. The possible interval corresponding to the recorded vasospasm day was therefore represented conservatively as

$$[D_0 + (d-1) \text{ days}, D_0 + d \text{ days}).$$

Because the exact time of vasospasm onset within the recorded day was unavailable, positive patients were retained in Phases 2–4 only when the end of the 24-hour physiological observation window occurred no later than the beginning of the assigned vasospasm-day interval. Thus, the complete physiological predictor window was required to precede the earliest possible time compatible with the recorded event day. Patients whose monitoring began too late to satisfy this condition, including cases in which the first 24 hours of monitoring could overlap the assigned vasospasm day, were excluded. This conservative temporal eligibility rule was used to reduce the possibility of incorporating peri-event or post-event physiology into the predictors.

Patients labelled No Vasospasm could not be subjected to an equivalent event-overlap criterion because no vasospasm event time existed for temporal alignment. Their clinician-provided negative outcome was therefore retained without an additional event-based timing exclusion. No additional retrospective timing cut-off was imposed on the negative group. The recorded vasospasm timing was used exclusively to establish pre-event eligibility among positive cases and was never included as a model predictor.

4.4 Study Populations and Eligibility

4.4.1 Static Clinical Cohort

Of the 221 labelled clinical entries described above, one was found during identifier and episode validation to duplicate an admission episode already represented for the same patient and was therefore not treated as a separate analytical observation.

The resulting Phase 1 cohort comprised 220 distinct admission episodes from 219 patients, including 102 positive and 118 negative episodes. Incomplete predictor values were handled within

cross-validation rather than by complete-case deletion. One patient contributed two genuinely distinct admission episodes; these remained separate units of analysis but were assigned to the same patient group in every training and test split.

4.4.2 Monitored Landmark Cohort

The physiological archive and the curated clinical dataset did not contain identical populations. Matching monitoring records to an available valid binary clinical outcome produced an initial labelled monitored cohort of 63 patients, including 41 positive and 22 negative cases. Coverage-based screening was incorporated into monitored-cohort construction, followed by verification of data adequacy for the final analytical signal set. Eligibility criteria were applied sequentially, with each patient counted only at the first criterion not met. After cohort construction, class balance, physiological signal coverage, total available monitoring duration, and the temporal separation between the extracted observation window and the assigned vasospasm-day interval were summarised descriptively (Table 4.2).

Overall, 18 of the 63 initially matched patients (28.6%) were excluded, reducing the monitored cohort from 41 to 23 positive cases while all 22 negative cases were retained. Of these exclusions, 15 positive patients were removed specifically because monitoring began after the assigned vasospasm-day interval ($n = 8$) or because the first 24 monitoring hours could overlap that interval ($n = 7$). Most exclusions therefore resulted from the temporal eligibility rules applied to positive cases. Among the retained positive patients, all extracted 24-hour observation windows ended before the earliest boundary compatible with the assigned vasospasm day, confirming that the intended pre-event temporal eligibility condition was satisfied. All 45 retained patients also met the required first-24-hour coverage threshold for each of the eleven final core physiological signals.

The final monitored cohort therefore comprised 45 patients. Among the retained negative patients, ICM+ monitoring began between 0.64 and 4.40 days after the recorded admission/event date (median 1.69 days), and no negative patient began monitoring seven or more days later. Because the clinical date was recorded at day rather than clock-time resolution, these intervals are descriptive rather than exact admission-to-monitoring durations. The same 45 patients and outcome labels were subsequently retained throughout Phases 2–4, with no further exclusion or relabelling.

Table 4.2: Construction, characteristics, and physiological data adequacy of the monitored cohort used in Phases 2–4.

Panel A. Sequential cohort construction

Selection step	**Excluded, *n***	**Remaining, *n***
Initial labelled monitored cohort	–	63
Monitoring file could not be processed	0	63
Total monitoring span below 24 hours	0	63
Insufficient core-signal coverage during the first 24 hours	2	61
Monitoring started before the event or admission, after ICU discharge, or more than 30 days after the event	0	61
Positive outcome without usable event date or vasospasm timing	1	60
Monitoring started after the assigned vasospasm-day interval	8	52
First 24 monitoring hours could overlap the assigned vasospasm-day interval	7	45
Final monitored cohort	**18 total**	**45**

Panel B. Final-cohort characteristics and temporal data adequacy

Metric	**Estimate**	**Denominator, *n***
Retained from the initial monitored cohort, *n* (%)	45 (71.4%)	63
Patients meeting the $\geq$ 50% coverage requirement for all eleven final core signals, *n* (%)	45 (100%)	45
Patient-level minimum core-signal coverage, median [IQR]; range, %	96.0 [92.9–97.1]; 52.1–100.0	45
Total source-monitoring span, median [IQR]; range, hours	313.6 [198.4–437.3]; 40.4–855.5	45
Positive patients with a definitely pre-vasospasm observation window, *n* (%)	23 (100%)	23
Minimum separation between observation-window end and assigned vasospasm-day boundary, median [IQR]; range, hours	75.6 [33.9–135.5]; 3.8–246.5	23

Notes: Exclusion criteria were applied sequentially, with each patient counted only at the first applicable step. The final analytical physiological set comprised eleven physiological signals. All 45 retained patients were verified to have at least 50% first-24-hour coverage for each of these eleven signals. Coverage was calculated over 1,440 one-minute bins, and the patient-level minimum reported in Panel B represents the lowest coverage across the eleven final signals. Vasospasm timing was available at daily resolution. Temporal separation therefore measures the interval from the end of the extracted 24-hour window to the earliest boundary of the assigned vasospasm day and represents a conservative lower bound.

4.4.3 Descriptive Cohort Characteristics

After cohort definition, age, sex, and outcome distributions were summarised descriptively for the broader Phase 1 analytical cohort and the final monitored cohort (Table 4.3). Because the admission episode was the unit of analysis in Phase 1, its descriptive statistics are reported at episode level; the monitored cohort is described at patient level. The single patient contributing two distinct Phase 1 admission episodes therefore contributes two observations to the Phase 1 descriptive summaries.

Table 4.3: Descriptive characteristics of the Phase 1 analytical cohort and the final monitored cohort.

Characteristic	**Phase 1 cohort ($N = 220$ admissions)**	**Monitored cohort ($N = 45$ patients)**
Analytical observations, *n*	220 admissions	45 patients
Unique patients represented, *n*	219	45
Age, mean ± SD, years	58.3 ± 13.4	58.1 ± 14.8
Age, median [IQR], years	58 [49–68]	60 [45–68]
Age range, years	23–89	23–89
Female, *n* (%)	136 (61.8%)	33 (73.3%)
Male, *n* (%)	84 (38.2%)	12 (26.7%)
Vasospasm, *n* (%)	102 (46.4%)	23 (51.1%)
No Vasospasm, *n* (%)	118 (53.6%)	22 (48.9%)

4.5 Predictors and Physiological Signals

Table 4.4 summarises the final analytical inputs, their representation, and their use across the four modelling phases.

Table 4.4: Final clinical and physiological inputs used across the modelling phases.

Input block	**Variables / signals**	**Phases**
Clinical predictors	Age; admission haemoglobin; GCS; Fisher scale; Hunt–Hess scale; pre-event modified Rankin Scale; sex; family history of aneurysm; active smoking; hypertension; ischaemia on admission CT; anterior-circulation location; posterior-circulation location	1, 2, 4
Physiological signals	ICP; mCPP; ABP; AMP; HR; RAP; PRx; RAC; ICPmax; ICPmin; $EtCO_2$	2, 3, 4

4.5.1 Admission and Early Clinical Predictors

The source clinical dataset provided 12 admission or early clinical variables considered for static modelling: age, admission haemoglobin, GCS at hospital admission, Fisher scale, Hunt–Hess

scale, pre-event modified Rankin Scale, sex, family history of aneurysm, active smoking, hypertension, ischaemia on admission CT, and aneurysm location. The location variable was subsequently recoded into two binary circulation indicators, resulting in a final static block of 13 predictors. The final clinical predictors were therefore age, admission haemoglobin, GCS at hospital admission, Fisher scale, Hunt–Hess scale, pre-event modified Rankin Scale, sex, family history of aneurysm, active smoking, hypertension, ischaemia on admission CT, anterior-circulation location, and posterior-circulation location, as listed compactly in Table 4.4. Detailed definitions and units or coding for these predictors are provided in Table B.2.

Source categories were recoded deterministically before cross-validation, and designated sentinel or invalid values were converted to missing. After this cleaning, ischaemia on admission CT was missing in 13 of 220 episodes and admission haemoglobin was missing in one episode. One of the patients had a recorded Fisher-Scale value of 5, which lay outside the valid 1–4 range used for this variable and was therefore recoded as missing; this occurred in a different episode from the missing admission-haemoglobin value. The remaining final predictors were complete. No data-dependent imputation, scaling, class balancing or feature selection was applied globally.

The original aneurysm-location field distinguished eight anatomical locations together with angiography-negative cases. Rather than expanding this variable into up to nine sparse location-specific indicators through one-hot encoding, locations were grouped into two clinically interpretable binary indicators representing the anterior and posterior cerebral circulations. Although the broader Phase 1 cohort comprised 220 admission episodes, this representation was selected in anticipation of its subsequent use in the substantially smaller monitored cohort ($N = 45$), particularly in Phase 4. The same encoding was retained across static-containing phases to maintain an identical static predictor definition across modelling stages.

The anterior-circulation indicator comprised the Internal Carotid Artery (ICA), Middle Cerebral Artery (MCA), Anterior Cerebral Artery (ACA), Anterior Communicating Artery (ACoA), and Posterior Communicating Artery (PCoA). The posterior-circulation indicator comprised the Basilar Artery, Posterior Cerebral Artery (PCA), and Posterior Inferior Cerebellar Artery (PICA). Angiography-negative cases had neither location indicator set. The relevant vascular territories and their circulation grouping are illustrated in the earlier Figure 2.2.

The PCoA lies anatomically at the interface between the anterior and posterior circulations. Most aneurysms clinically designated as PCoA aneurysms arise at the internal-carotid–posterior-communicating-artery junction, whereas "true" aneurysms arising from the PCoA itself are comparatively rare (He et al. 2011). Morphological studies have also associated PCoA aneurysm formation with geometric characteristics of the adjacent internal carotid artery and the ICA–PCoA bifurcation (Hao et al. 2022). However, classification is not completely uniform across the literature: some natural-history studies have grouped PCoA aneurysms with posterior-circulation lesions because of their observed rupture-risk profile, despite their more conventional anatomical grouping with the anterior circulation (Clarke et al. 2005). In the primary analysis, PCoA was therefore classified with the anterior circulation on anatomical grounds. Because of this recognised classification ambiguity, a post-hoc Phase 1 sensitivity analysis examined the alternative

coding of PCoA as posterior circulation while leaving the remaining cohort, predictors, validation partitions, preprocessing, and modelling procedures unchanged. This sensitivity analysis did not determine the primary anatomical grouping and was conducted only as an exploratory robustness check.

4.5.2 Physiological Monitoring and Signal Derivation

Physiological monitoring data were acquired and processed using ICM+, a configurable multimodal monitoring platform capable of storing high-resolution physiological waveforms together with lower-frequency processed trends and derived indices (Smielewski et al. 2024). In the present archive, the high-frequency waveforms were acquired at 250 Hz, whereas the exported trends used for modelling were available at nominal one-minute resolution. The present study did not use the high-frequency waveforms directly; subsequent alignment of the exported trends to a common first-24-hour grid is described in Section 4.5.3.

The final physiological input set comprised eleven signals: ICP, mean cerebral perfusion pressure (mCPP), ABP, ICP pulse amplitude (AMP), heart rate (HR), RAP, PRx, RAC, maximum ICP (ICPmax), minimum ICP (ICPmin), and end-tidal carbon dioxide ($EtCO_2$), as summarised in Table 4.4. Because ICM+ allows the formulas, calculation-window durations, update intervals and optional filtering of derived variables to be configured by the user, the processing associated with these trends could not be inferred reliably from their signal names or from generic software documentation alone.

The historical calculation profile was investigated using ICM+-specific HDF5 packages generated from selected monitoring recordings. One representative recording underwent detailed programmatic audit of the embedded `signalscfg` and `analysiscfg` configuration fields, including direct numerical reconstruction of PRx. Additional recordings from the monitored cohort were inspected to verify consistency of the principal signal definitions, calculation periods and update intervals. No material differences were identified in the processing characteristics relevant to the present analyses. The detailed numerical reconstruction was performed only for the representative recording.

Within the inspected profile, the recovered ABP and ICP signal definitions specified no additional ICM+ digital filter (`Filter=None`). Mean ICP and mean ABP were calculated over 10-s windows and updated every 10 s, while AMP was derived over the same temporal scale. Cerebral perfusion pressure was calculated from the corresponding mean pressure trends as

$$\mathrm{mCPP} = \mathrm{Mean}(\mathrm{ABP}) - \mathrm{Mean}(\mathrm{ICP}).$$

The principal correlation-based indices were subsequently defined as

$$\begin{aligned} \mathrm{PRx} &= \mathrm{Correl}(\mathrm{ABP}, \mathrm{ICP}), \\ \mathrm{RAP} &= \mathrm{Correl}(\mathrm{AMP}, \mathrm{ICP}), \\ \mathrm{RAC} &= \mathrm{Correl}(\mathrm{CPP}, \mathrm{AMP}), \end{aligned}$$

calculated from 10-s trends using 300-s moving Pearson correlation windows updated every 60 s. PRx, RAP and RAC therefore represent overlapping summaries of recent physiological history rather than instantaneous independent measurements.

The recovered profile was additionally examined for operations that could introduce temporal look-ahead. No zero-phase filtering, centred-window operation, scripted time shift or other explicit future-looking operation was identified in the relevant processing pathway. To assess temporal alignment directly, PRx was independently reconstructed from the archived pressure signals using trailing windows corresponding to the recovered configuration. Across 729 comparable values, reconstructed and stored PRx showed strong agreement ($r = 0.978$; MAE $= 0.021$), whereas shifting the reconstruction by one minute in either direction reduced the correlation to approximately 0.76. This supported interpretation of the stored PRx timestamp as marking the end of a trailing calculation window and provided no evidence of look-forward processing in the inspected profile.

The audit established the principal derivation and temporal structure of mICP, mABP, mCPP, AMP, PRx, RAP and RAC. It did not independently reconstruct every upstream implementation detail of HR, ICPmax, ICPmin or $EtCO_2$, nor could it determine device-level conditioning, manual artefact handling or other processing occurring before data entered ICM+. These variables were therefore treated as exported ICM+ trends without assigning undocumented processing steps.

Accordingly, the physiological modelling archive should be interpreted as a set of processed ICM+ trends available at nominal one-minute resolution, rather than as direct samples of the original high-frequency waveforms. Subsequent first-24-hour alignment, preservation of missing observations and fold-contained preprocessing were applied to these exported values as described in Sections 4.5.3–4.5.5. The physiological indices themselves were not recalculated by the modelling pipeline.

4.5.3 Common First-24-Hour Minute Grid

Physiological records were curated independently for each patient. For each signal, the available alias with the greatest number of numeric observations was selected. Timestamps were parsed and sorted, and the first 24 chronological hours were selected using elapsed time rather than row count. Exact duplicate timestamps were collapsed signal by signal using the median. Distinct observations were then mapped to the nearest relative minute from 0 to 1439; collisions within a minute were again collapsed by the median. Every patient was reindexed to the same 1,440-minute grid, and minutes without an observation remained missing.

A preliminary timestamp-cadence audit assessed the median and 95th-percentile intervals between consecutive unique timestamps in each monitoring file. Across the final 45-patient cohort, both intervals were approximately 60 seconds in every file when rounded to one decimal place. This confirmed that the exported ICM+ records had a nominal temporal resolution of one minute and supported their organisation on a common minute grid. These intervals describe the cadence of the exported trends rather than the original acquisition frequency of the bedside waveforms.

The preceding operations depended only on each patient's timestamps and values and could therefore be completed before cross-validation without sharing population information. No unrestricted interpolation, population imputation, scaling, clipping, winsorisation or outcome-dependent cleaning was applied during construction of the canonical grid. Minutes without an observation remained explicitly missing so that their subsequent treatment could be performed within the relevant modelling and training split.

Nine patients had at least one missing run of 120 minutes or longer in one or more signals, and the longest observed gap was 689 minutes. These gaps remained explicit in the canonical temporal data and were not bridged during construction of the common minute grid.

4.5.4 Aggregated Representations

Phase 2 converted the observed minute bins into patient-level summary features. Statistics were calculated from observed values only. Two physiological representations were constructed: a primary minimal representation containing 20 features and a broader secondary representation containing 76 features. The summary statistics assigned to each signal group are shown in Table 4.5.

Table 4.5: Summary statistics used to construct the Phase 2 physiological representations.

Signals	**Minimal representation**	**Extended representation**	**Features**
ICP, mCPP, ABP, AMP, HR, $EtCO_2$	Median; IQR	Mean; SD; median; minimum; maximum; Q25; Q75; slope	12 / 48
RAP, PRx, RAC	Median; IQR	Mean; SD; median; Q25; Q75; slope	6 / 18
ICPmax	Q75	Mean; SD; median; Q75; maximum	1 / 5
ICPmin	Q25	Mean; SD; median; Q25; minimum	1 / 5
Total	**20 features**	**76 features**	**20 / 76**

For the extended representation, slope was estimated by ordinary least squares over the observed minute-level values. The five evaluated Phase 2 input representations were therefore static only (13 features), minimal physiology (20), static plus minimal physiology (33), extended physiology (76), and static plus extended physiology (89).

The minimal representation was designated as the primary physiological representation as a deliberately parsimonious summary of the first 24 hours, rather than because its individual statistics were assumed to be intrinsically superior to alternatives such as the mean or standard deviation. For the main physiological trends, median and IQR provided one measure of central tendency and one of dispersion while limiting redundancy between closely related summary statistics and reducing the influence of isolated extreme values. For ICPmax and ICPmin, which already represent upper and lower ICP extremes within the exported trends, a single directional quantile was retained (Q75 and Q25, respectively) to preserve information about relatively high or low values without expanding these variables into several correlated summaries.

The extended representation was retained as a secondary exploratory analysis to assess whether additional distributional and temporal information, including the mean, standard deviation, extrema and slope, provided useful predictive information beyond this compact description. The distinction between the two representations therefore reflected a trade-off between parsimony and information breadth rather than an assumption that the minimal statistics were universally preferable. This was particularly relevant given the 45 independent patients available for Phase 2 and the increase from 20 physiological features in the minimal representation to 76 in the extended representation.

After feature construction, any remaining feature-level missingness was handled within the relevant training fold. Continuous and ordinal static variables and physiological features were median-imputed, whereas binary static variables were imputed using the most frequent training-fold value. For logistic regression, continuous, ordinal and physiological features were additionally standardised using parameters estimated from the training fold. Random forest and XGBoost used the imputed values without standardisation. No feature-selection step was applied.

4.5.5 Temporally Resolved Representation

Phase 3 retained the minute-level ordering of the physiological data in a tensor of shape $45 \times 11 \times 1{,}440$, corresponding to patient, signal and minute. The canonical tensor retained missing values together with a Boolean mask identifying cells that had been observed before interpolation or imputation. Overall, 93.09% of tensor cells were observed. Deterministic patient-local timestamp curation preceded cross-validation, whereas every population-dependent transformation was estimated separately within the relevant training split.

For each inner or outer fitting split, only bounded internal gaps of exactly one minute were linearly interpolated. This restricted interpolation was used only to bridge isolated missing minutes between observed neighbouring values; leading and trailing gaps, and internal gaps longer than one minute, were not interpolated. Remaining missing values were filled using the per-signal median estimated exclusively from the training patients. Values were then centred by those training medians and divided by the corresponding training IQR. If an IQR was degenerate, the training standard deviation and then 1.0 were used as fallbacks. No patient-wise normalisation, clipping, winsorisation or outcome-responsive cleaning was applied.

For the observation-mask sensitivity analysis, the eleven binary observation masks were appended to the eleven processed value channels, resulting in 22 input channels. Each mask indicated whether the corresponding minute-level cell had originally been observed; it therefore remained zero for values later obtained through either interpolation or median imputation. The mask did not distinguish between these two forms of missing-value handling. This representation was evaluated separately because patterns of signal availability may contain information about acquisition and care processes in addition to physiological signal availability.

4.6 Phase-Specific Modelling

4.6.1 Phase 1: Static Clinical Modelling

Phase 1 used the 13-variable static block to establish a broad benchmark of low-dimensional clinical prediction across seven complementary model families. L2-penalised logistic regression, random forest, XGBoost and LightGBM underwent restrained inner-loop candidate selection, while a multilayer perceptron, Gaussian naive Bayes and a linear support vector classifier were retained as fixed secondary benchmarks. Together, these models represented regularised linear, non-linear tree-ensemble, neural, probabilistic and margin-based approaches, allowing the predictive information in the same clinical representation to be assessed without assuming a single functional form. The model specifications are summarised in Table 4.6.

Table 4.6: Phase 1 model families and candidate configurations.

Model family	**Selection status**	**Configurations**
Logistic regression	Inner-loop selection	Unweighted or balanced L2-penalised fitting, crossed with $C \in \{1.0, 0.3\}$, giving four candidates.
Random forest	Inner-loop selection	Both structural settings used 100 trees and maximum depth 5. The first used minimum split size 5, minimum leaf size 1 and $\sqrt{p}$ features per split; the second used minimum split size 2, minimum leaf size 10 and $0.5p$ features. Each was evaluated with balanced and unweighted fitting.
XGBoost	Inner-loop selection	Either 100 depth-3 trees or 150 depth-2 trees, with learning rate 0.05, minimum child weight 1, full row and column sampling, and L2 regularisation of 1. Each configuration was evaluated with balanced and unweighted fitting.
LightGBM	Inner-loop selection	The first configuration used 100 trees, seven leaves, maximum depth 3, learning rate 0.05, minimum child size 20, full row and column sampling, and no L2 penalty. The second used 200 trees, five leaves, maximum depth 3, learning rate 0.03, minimum child size 10, row and column sampling fractions of 0.8, and L2 regularisation of 5. Each was evaluated with balanced and unweighted fitting.
Multilayer perceptron	Fixed secondary benchmark	One hidden layer with 16 ReLU units, Adam optimisation, learning rate 0.001 and L2 penalty 0.05.
Gaussian naive Bayes	Fixed secondary benchmark	Empirical class priors and variance smoothing of 10^{-8}.
Linear support vector classifier	Fixed secondary benchmark	Unweighted linear kernel with $C = 1$.

Candidate sets were deliberately restricted because, despite being larger than the monitored cohort, the 220-admission Phase 1 sample remained modest for extensive hyperparameter optimisation. This allowed alternative model structures to be compared while limiting the additional selection variance and overfitting risk associated with a broad search.

No data-dependent imputation was performed globally before cross-validation. Within each fitting fold, numeric and ordinal predictors were median-imputed using the corresponding training data, while binary predictors were imputed with the most frequent training-fold category. Numeric predictors were subsequently standardised for logistic regression, the multilayer perceptron and the linear support vector classifier using training-fold parameters. Tree models and Gaussian

naive Bayes used the imputed values without standardisation. No feature-selection procedure was applied.

After model fitting, held-out permutation importance was calculated as an exploratory description of model-specific predictive association. Within each outer-test fold, each predictor was permuted five times and importance was expressed as the decrease in outer-test AUROC relative to the corresponding unpermuted predictions. Permutation importance was not used for preprocessing, candidate selection, calibration, threshold selection or designation of a primary model.

4.6.2 Phase 2: Aggregated Physiological Modelling

Phase 2 shifted the analytical focus from broad model-family benchmarking to comparison of alternative representations within the same 45 monitored patients. The five representations defined in Section 4.5.4—static only, minimal physiology, static plus minimal physiology, extended physiology, and static plus extended physiology—were therefore evaluated using three complementary tabular approaches: logistic regression as a regularised linear model, random forest as a bagged non-linear ensemble, and XGBoost as a boosted tree ensemble. This retained distinct modelling assumptions while limiting model-selection complexity in the substantially smaller monitored cohort.

Logistic regression compared two unweighted L2-penalised candidates with $C \in \{1.0, 0.3\}$. Random forest and XGBoost used the two structural configurations reported in Table 4.6, without the balanced variants; class weighting was not used because the monitored cohort was nearly balanced. The same model families and restricted candidate spaces were applied to every representation so that differences between representations were not confounded by different tuning spaces. Candidate selection, calibration and threshold construction followed the nested procedure in Section 4.7.

Aggregated-feature construction, treatment of unobserved minute bins, training-fold imputation and model-specific standardisation followed Section 4.5.4, with no feature selection. The minimal physiological representation constituted the primary physiology-only analysis, and static plus minimal physiology the primary incremental analysis against the matched static representation. The extended representations remained secondary and exploratory because their 76 physiological features, or 89 features when combined with the static block, were comparatively high-dimensional relative to the 45 independent patients.

4.6.3 Phase 3: Temporally Resolved Physiological Modelling

Phase 3 used the canonical temporal representation and fold-contained missing-data treatment defined in Section 4.5.5. Three analyses were evaluated: a primary value-only 1D-CNN, a prespecified value-plus-mask CNN sensitivity analysis, and a ROCKET–Ridge comparator. Their roles and input representations are summarised in Table 4.7.

Table 4.7: Phase 3 temporal model specifications and analytical roles.

Model	**Input**	**Role**	**Principal configuration**
Value-only 1D-CNN	$11 \times 1{,}440$ processed values	Primary temporal analysis	Three convolutional layers with 16, 32 and 32 filters; kernel widths 15, 9 and 5; max-pooling by 4 after the first two layers; global average pooling, dropout 0.30 and one output logit.
Value-plus-mask 1D-CNN	$22 \times 1{,}440$: eleven values and eleven observation masks	Prespecified sensitivity analysis	Architecture and optimisation identical to the value-only CNN. The added channels represented original observation status rather than an additional physiological measurement.
ROCKET–Ridge	$11 \times 1{,}440$ processed values	Non-neural temporal comparator	10,000 random convolutional kernels producing 20,000 transformed features, followed by sparse-compatible standardisation and a Ridge classifier.

The 1D-CNN and ROCKET–Ridge represented two distinct strategies for extracting information from the same temporally ordered physiological input. The CNN was selected as the primary temporal model because convolutional filters can learn local and multi-scale patterns directly from aligned multichannel sequences while sharing parameters across the 1,440-minute window. A compact convolutional architecture also limited model complexity given the 45 independent patients available for temporal modelling. ROCKET–Ridge provided a non-neural alternative: a large collection of fixed random convolutional kernels generated the temporal representation, while only the downstream linear classifier was learned from the outcome. The comparison therefore contrasted supervised learning of temporal filters with a rich but fixed convolutional transformation.

Both CNN variants were trained using unweighted binary cross-entropy and AdamW, with learning rate 0.001, weight decay 0.001 and batch size 8. Within each outer training set, five inner models used early stopping on validation binary cross-entropy, with a maximum of 150 epochs, patience of 20 epochs and a minimum improvement of 0.0001. The rounded median of the five selected epoch counts determined the training duration of a fresh final network fitted to the complete outer training set. Native sigmoid outputs served as both ranking scores and probabilities, and no post-hoc probability calibration was fitted.

The value-plus-mask sensitivity used the same architecture and optimisation procedure but appended eleven binary channels indicating whether each minute-level value had originally been observed. After interpolation or median imputation, the value-only CNN could not directly distinguish an observed value from one introduced by missing-data handling; the masks therefore retained this provenance without altering the physiological values themselves. Interpolated and median-imputed values were both marked as originally unobserved. This representation remained a sensitivity analysis because observation patterns could encode information about signal acquisition and care processes in addition to physiology.

For ROCKET–Ridge, the transform and subsequent sparse-compatible standardisation were

fitted within the current training split. Four Ridge penalties were considered:

$$\alpha \in \{0.01p,\ 0.1p,\ p,\ 10p\}, \qquad p = 20{,}000,$$

where p was the transformed feature count. Candidate selection used inner out-of-fold raw-score AUROC, with calibrated Brier score as the tie-breaker. The raw Ridge score remained the discrimination endpoint. Probability calibration was constrained to be non-decreasing; if the fitted sigmoid slope was negative, calibrated probabilities were replaced by the outer-training prevalence rather than allowing calibration to reverse the score ranking.

4.6.4 Phase 4: Static–Temporal Late Fusion

Phase 4 tested whether the temporal CNN predictions contained information incremental to the static clinical representation in the same 45 patients. It was designed as a focused multimodal integration test rather than a search over alternative temporal architectures or optimisation settings. The primary analysis combined a fixed static logistic-regression score with a value-only CNN score; the prespecified sensitivity analysis used the value-plus-mask CNN. The CNN architectures, input representations and optimisation settings were retained from Phase 3, while fusion-training scores were generated using the nested procedure described below. ROCKET–Ridge retained its Phase 3 role as a temporal comparator and was not included as an additional fusion candidate. The fitting and validation protocol for the reported Phase 4 analyses was fixed before those results were examined.

Score-level late fusion was chosen instead of concatenating the original static variables with the full temporal tensor and fitting a new end-to-end multimodal model. With only 45 independent patients, the latter would have introduced a substantially larger joint modelling problem and an additional layer of model selection. Late fusion instead reduced each modality to a separately generated score and directly tested whether the temporal prediction provided complementary information beyond the static prediction, while retaining the modality-specific model specifications and preprocessing definitions established in the preceding phases.

The matched standalone comparator was a new fixed-C Phase 4 static logistic regression generated within the same outer-fold procedure; the separately tuned Phase 2 static estimate was not used as the comparator. The resulting model was therefore a two-score late-fusion stacked ensemble rather than an end-to-end neural hybrid. Its components are summarised in Table 4.8.

Within each 36-patient outer-training set, five stratified base folds were used to generate the temporal fusion-training features. For each fold, 7–8 patients formed the base holdout and the remaining 28–29 formed the base-training subset. Five further stratified folds were formed exclusively within the base-training subset to select the CNN training duration, following the early-stopping rule described in Section 4.6.3. Each nested CNN was fitted on its nested training subset, with checkpoint selection based only on the corresponding nested validation subset. The rounded median of the five selected epoch counts determined the training duration of a freshly initialised CNN fitted to the complete base-training subset. This network then predicted the base holdout

Table 4.8: Components of the Phase 4 late-fusion procedure.

Component	**Implementation**	**Output or role**
Static branch	Six numeric or ordinal variables were training-fold median-imputed and standardised; seven binary variables received training-fold most-frequent imputation. The classifier was unweighted L2 logistic regression with fixed $C = 1$ and the `liblinear` solver.	Decision score for fusion; native sigmoid probability for the standalone static comparator.
Temporal branch	Phase 3 CNN specifications retained for the primary value-only and value-plus-mask sensitivity analyses. Fusion-training probabilities generated using five base folds, each with five nested epoch-selection folds and a fresh fixed-duration fit. Population-dependent preprocessing fitted within the corresponding training subsets. Phase 3 outer-test probabilities reused unchanged; no post-hoc recalibration.	CNN probability, clipped to $[10^{-6}, 1 - 10^{-6}]$ and transformed to a logit.
Fusion model	The static decision score and temporal logit were standardised within the fusion training data and combined using unweighted L2 logistic regression with fixed $C = 1$ and the `liblinear` solver.	One fused probability. No interactions, hidden layers, alternative temporal models or additional post-hoc calibration.
Meta-level cross-fit	A second-level five-fold cross-fit within each outer training set generated meta-model out-of-fold probabilities.	Training-only probabilities for screening-threshold selection.

once, without using that holdout for early stopping. Population-dependent preprocessing was fitted separately within the training subset associated with each CNN fit.

Repeating this procedure across the five base folds produced one temporal probability for each of the 36 outer-training patients. Each patient's temporal score was generated with that patient excluded from gradient-based weight updates, preprocessing fitting and epoch selection. These probabilities were paired with static decision scores obtained by five-fold cross-fitting of the fixed static logistic-regression pipeline within the same outer-training set. The fusion combiner was fitted using these 36 paired scores and their outcomes.

For the temporal input at outer-test evaluation, the corresponding saved Phase 3 outer-test probabilities were reused unchanged. These probabilities had been generated by CNNs fitted to all 36 outer-training patients, with training duration selected entirely within that set as described in Section 4.6.3. Only the Phase 3 outer-test probabilities were reused, and its inner-validation probabilities were not used as fusion-training features. The fitted fusion combiner was applied once to the paired static and temporal scores of the nine untouched outer-test patients. The validation properties and qualifications of this procedure are described in Section 4.7.

4.7 Validation, Model Selection and Calibration

All reported model evaluation was internal. Phase 1 used 50 deterministic repetitions of five-fold stratified, patient-grouped outer cross-validation, with five grouped inner folds within every outer training set. The two episodes belonging to the repeated patient were always assigned to the same

group, preventing that patient from appearing in both training and test data. All Phase 1 models used the same saved outer assignments.

Phases 2–4 used 50 fixed repetitions of stratified five-fold outer cross-validation with the same 45 patients, labels and fold assignments throughout. Each outer fold contained approximately 36 training patients and nine untouched test patients, with five stratified inner folds formed from the outer training set as required for model selection, early stopping, calibration or threshold construction. Because each patient contributed one monitored record, patient-level separation was implicit. Phase 4 additionally used nested CNN epoch selection within each base-training subset, as described in Section 4.6.4.

Repeated five-fold evaluation was chosen because a single train–test split would make performance strongly dependent on one allocation of a limited cohort, particularly in Phases 2–4. Repetition characterised variation across multiple plausible partitions while five-fold splitting retained most patients for fitting in each iteration. Stratification maintained approximately comparable outcome proportions across the small outer-test sets; in Phase 1, patient grouping additionally prevented leakage between the two admissions contributed by the same patient.

The nested structure separated model development from performance estimation. All fold-dependent preprocessing, candidate selection, calibration, early stopping and threshold construction were confined to the corresponding outer-training data, leaving the outer-test fold untouched until final evaluation. This reduced optimism from data-dependent model development and was particularly important because the study ranged from low-dimensional static models to higher-dimensional aggregated and temporal representations.

Shared outer assignments also enabled matched comparisons. All Phase 1 model families were evaluated on the same patient-grouped partitions, while Phases 2–4 retained identical patients and folds across representations and models. Differences could therefore be calculated within the same repetition rather than being confounded by different train–test allocations. Phase 1 remained only descriptively comparable with Phases 2–4 because its cohort and unit of analysis differed. The phase-specific functions of the inner procedures and probability handling are summarised in Table 4.9.

For the tuned tabular models in Phases 1 and 2, each candidate generated inner out-of-fold raw scores. A separate sigmoid calibrator was fitted to each candidate's scores and outcomes, and its calibrated Brier score was calculated on those same inner out-of-fold observations, without a further cross-fit of the calibration layer. Candidate selection maximised raw-score inner out-of-fold AUROC, with exact ties resolved by lower calibrated Brier score and then the lexicographically greatest candidate name. The selected candidate's already fitted calibrator was retained without refitting. The underlying model was then refitted to the complete outer-training set, and the retained calibrator was applied unchanged to its outer-test scores.

The three fixed Phase 1 benchmarks followed the same sigmoid-calibration and threshold-construction procedures using five grouped inner folds, but each had a single predefined configuration and underwent no candidate comparison. For all Phase 1 models, reported outer AUROC and AP were calculated from the calibrated probabilities.

Table 4.9: Phase-specific internal-validation, selection and calibration procedures.

Analysis	**Outer evaluation**	**Inner procedure**	**Probability handling**
Phase 1	50×5 stratified, patient-grouped folds	Candidate selection for tuned models only; sigmoid calibration and threshold construction for all models using five grouped inner folds	Candidate-specific inner out-of-fold sigmoid calibration; selected or fixed configuration's calibrator retained
Phase 2	50×5 fixed stratified folds in the 45-patient cohort	Candidate selection, sigmoid calibration and threshold construction using five stratified inner folds	Candidate-specific inner out-of-fold sigmoid calibration; selected candidate's calibrator retained
Phase 3 CNNs	Same fixed outer folds as Phase 2	Early stopping and threshold construction using five inner folds	Native sigmoid probability; no post-hoc calibration
Phase 3 ROCKET–Ridge	Same fixed outer folds as Phase 2	Ridge-penalty selection, calibration and threshold construction using inner out-of-fold scores	Non-decreasing sigmoid calibration with an outer-training-prevalence fallback
Phase 4 fusion	Same fixed outer folds as Phases 2–3	Five base folds with five nested CNN epoch-selection folds per base-training subset; rounded-median fixed-duration refits; second-level five-fold cross-fit for meta-model out-of-fold probabilities and threshold construction	Native logistic output of the fusion combiner; no additional recalibration

Where threshold-dependent metrics were calculated, a screening-oriented threshold was selected from inner out-of-fold probabilities by maximising specificity subject to sensitivity of at least 0.85. Ties favoured higher sensitivity and then the higher threshold. The selected threshold was applied unchanged to the corresponding outer-test probabilities and represented an analytical operating point rather than a clinical diagnostic or treatment cut-off.

All population-dependent preprocessing parameters, including imputation and standardisation statistics, were estimated within the relevant training split and applied unchanged to its validation or test data. Deterministic patient-local minute-grid construction preceded these splits as described in Section 4.5.3. Hyperparameter selection, epoch selection, calibration and threshold selection also excluded the outer-test fold.

After all five outer folds had been evaluated, their predictions were concatenated into one complete out-of-fold prediction vector for the analytical cohort. Performance metrics were calculated once on this vector for each repetition rather than averaged across the five small outer-test folds.

Two phase-specific qualifications apply. For the Phase 3 CNNs, each inner validation fold, including its outcomes, was used to select the early-stopping checkpoint and supplied predictions from that checkpoint for inner threshold construction. This dual role may make the inner threshold estimate optimistic, although no outer-test outcome was used for checkpoint or threshold selection.

For Phase 4, the second-level meta-cross-fit reused the assembled base-model out-of-fold feature set rather than regenerating the base features within each meta-training split. Although each patient's own temporal feature was generated without that patient contributing to CNN fitting or epoch selection, a meta-validation patient's outcome could still have influenced the base mod-

els generating other patients' feature rows used to fit its meta-level predictor. The base models generating fusion-training features were fitted on 28–29 patients, whereas the models used for outer-test prediction were fitted on all 36 outer-training patients. The nested CNN fits used for epoch selection involved still smaller training subsets of 22–24 patients. Training-side diagnostics and threshold estimates were therefore interpreted with caution because of these dependencies and training-size differences, rather than as substitutes for outer-test performance. No outer-test outcome entered preprocessing, epoch selection, temporal-model fitting, fusion fitting or threshold selection. The procedure is consequently described as cross-fitted stacking with an untouched outer-test fold, rather than fully independent end-to-end nested stacking.

4.8 Evaluation Metrics and Statistical Interpretation

The primary performance measure was the area under the receiver operating characteristic curve (AUROC), which quantifies threshold-independent ranking discrimination and can be interpreted as the probability that a randomly selected patient with vasospasm receives a higher score than a randomly selected patient without vasospasm. In Chapter 3, the term AUC is retained when reporting the terminology used by source studies; in this dissertation's own analyses, the more explicit term AUROC is used for the area under the receiver operating characteristic curve. AUROC was selected as the primary metric because it could be applied consistently across the different model families and representations without requiring a common operating threshold.

Average precision (AP) and the Brier score provided complementary information. AP was calculated using `average_precision_score` and summarises the precision–recall relationship as a recall-weighted mean of precision without trapezoidal interpolation. It therefore emphasises retrieval of the positive class while accounting for false positives as recall increases and was interpreted relative to the positive-outcome prevalence of each analytical cohort. The Brier score measured the mean squared difference between predicted probabilities and observed outcomes, providing a measure of overall probabilistic accuracy sensitive to both discrimination and calibration. Together, AUROC, AP and Brier score characterised ranking discrimination, positive-class prediction and probabilistic accuracy without dependence on a single threshold.

Sensitivity, specificity, positive predictive value and negative predictive value were retained as secondary operating-point measures at the screening-oriented threshold defined in Section 4.7. They were calculated once per repetition from the complete outer out-of-fold classifications, retaining each observation's fold-specific threshold. For reporting, the five training-selected thresholds were averaged within each repetition, weighted by the number of outer-test observations assigned to each fold. This summary threshold was not reapplied to the concatenated predictions.

Calibration intercept and slope were estimated jointly using an unpenalised logistic regression of the observed outcome on the logit of predicted probability, with probabilities clipped to $[10^{-6}, 1-10^{-6}]$. This diagnostic regression was fitted once per repetition to the complete concatenated outer out-of-fold probability vector. The intercept was estimated together with the slope, rather than with the slope fixed at one. These coefficients were used only as descriptive calibration

diagnostics and were not applied to modify the evaluated predictions. This assessment was distinct from the inner out-of-fold fitting of probability calibrators where applicable. Operating-point and calibration assessments remained secondary given the limited cohort sizes and lack of independent external validation.

For each model, the 50 repetition-level performance values were summarised by their mean and sample standard deviation, with empirical 2.5th and 97.5th percentiles reported where percentile ranges were used. Paired within-repetition differences were calculated when analyses shared the same patients and partitions. These summaries describe variation across repeated data splits and were not treated as confidence intervals, independent replications, probabilities or hypothesis tests.

Phase 4 training-side cross-fitting diagnostics were calculated from the static out-of-fold decision scores, the temporal probabilities generated by the nested procedure in Section 4.6.4, and the fusion meta-model out-of-fold probabilities. Each training-side AUROC was calculated on the corresponding complete 36-patient score vector. The diagnostics were summarised by their mean and sample standard deviation across the 250 outer-training fits arising from 50 repetitions of five-fold cross-validation, without intermediate repetition-level averaging. Static–temporal correlation was Pearson correlation between static decision scores and logits of clipped CNN probabilities. Fusion coefficients were expressed on the standardised two-score input scale. These diagnostics described variation across overlapping training sets and were not treated as independent replications or substitutes for outer-test performance.

Phase 2 incremental comparisons subtracted static-only performance from the corresponding static-plus-minimal-physiology performance within the same model family and repetition. Phase 3 paired the primary value-only CNN with the value-plus-mask sensitivity analysis; comparisons with Phase 2 were descriptive despite sharing patients and outer folds because the algorithms and representations differed. Phase 4 paired each fusion analysis with its contemporaneously generated fixed-C static comparator. Comparisons involving Phase 1 were descriptive only because its cohort and unit of analysis differed.

Exploratory permutation importance was retained as a reported feature-level analysis only in Phase 1, where it characterised which admission and early clinical predictors most consistently contributed to static prediction. It was interpreted as the model-specific change in held-out discrimination after feature permutation, not as evidence of a causal risk factor, biological mechanism or biomarker.

No directly comparable feature-level importance analysis was retained for reporting in Phases 2–4. Their objective was instead to assess the predictive contribution of physiological information, temporal representation and multimodal integration. Differences in representation, the 45-patient monitored cohort and the higher dimensionality of the later inputs also made cross-phase feature rankings difficult to interpret reliably. The Phase 3 observation-mask sensitivity was not treated as feature importance, and the two Phase 4 fusion coefficients were not interpreted as importance estimates for individual clinical or physiological variables.

Chapter 5

Results

This chapter reports the internal-validation results in the four analytical phases defined in Chapter 4. Unless otherwise stated, performance estimates are reported as the mean ± standard deviation across 50 repetitions of five-fold outer cross-validation. Empirical 2.5th–97.5th percentile ranges describe variation across those repetitions and are not confidence intervals. Average precision (AP) is interpreted relative to the observed outcome prevalence. Comparisons involving Phase 1 are descriptive because that phase used a broader cohort and an admission-episode unit of analysis; incremental claims concerning physiology are based on matched comparisons within the 45-patient monitored cohort.

5.1 Cohort and Physiological Data Availability

Phase 1 comprised 220 admission episodes from 219 patients, including 102 vasospasm-positive and 118 vasospasm-negative episodes. The positive-class prevalence was 46.4%, providing an AP reference of 0.464. The initial labelled monitored cohort comprised 63 patients, including 41 positive and 22 negative cases; after sequential eligibility criteria, 45 patients remained, comprising 23 positive and 22 negative cases. The resulting positive-class prevalence for Phases 2–4 was 51.1%, providing an AP reference of 0.511. Of the 18 monitored-cohort exclusions, 15 arose from the two temporal criteria that prevented the first 24 monitoring hours from occurring after, or potentially overlapping, the assigned vasospasm-day interval. The complete sequential flow and baseline characteristics of the analytical cohorts are reported in Tables 4.2 and 4.3, respectively.

All 45 retained patients met the first-24-hour coverage requirement for each of the eleven final physiological signals. Across the canonical patient–signal–minute tensor, 93.09% of cells were observed and 49,226 cells remained missing before fold-contained interpolation and imputation. Nine patients had at least one missing run of 120 minutes or longer, with a maximum missing run of 689 minutes.

5.2 Phase 1: Static Clinical Prediction

5.2.1 Predictive Performance

Table 5.1 summarises the seven static model families according to the primary discrimination metric. XGBoost achieved the highest mean AUROC at 0.720 ± 0.015, followed closely by LightGBM, random forest, logistic regression and the linear SVC. Gaussian naive Bayes and the multilayer perceptron produced lower mean AUROCs of 0.660 ± 0.019 and 0.642 ± 0.029, respectively. Complementary AP and Brier-score results for all seven models are reported in Table C.1.

Table 5.1: Phase 1 static-model discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values.

Model	AUROC, mean ± SD	2.5th–97.5th percentiles
XGBoost	**0.720 ± 0.015**	0.690–0.745
LightGBM	0.713 ± 0.016	0.678–0.739
Random forest	0.708 ± 0.015	0.683–0.732
Logistic regression	0.703 ± 0.013	0.681–0.725
Linear SVC	0.702 ± 0.015	0.677–0.730
Gaussian naive Bayes	0.660 ± 0.019	0.617–0.691
Multilayer perceptron	0.642 ± 0.029	0.585–0.691

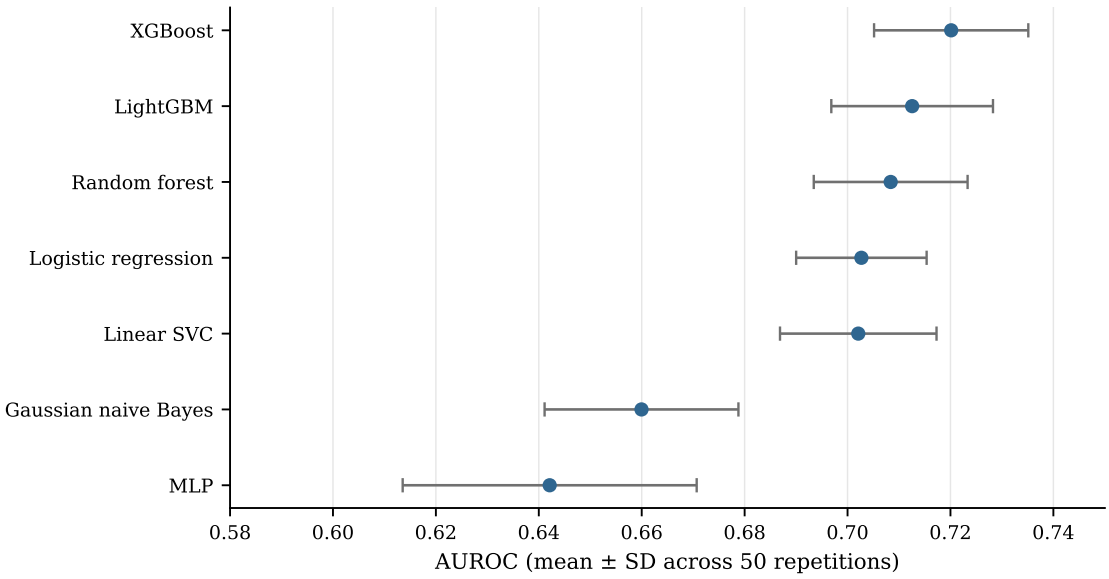


Figure 5.1: Phase 1 AUROC by static model family. Points show the mean across 50 repetitions and whiskers show one standard deviation across repetitions. The whiskers describe split variability and are not confidence intervals.

Using XGBoost, which achieved the highest mean AUROC, as a descriptive reference, the paired AUROC differences are reported in Table 5.2. XGBoost was higher than LightGBM, random forest and logistic regression by mean AUROC differences of 0.008, 0.012 and 0.017, re-

spectively. The empirical percentile ranges for each of these three paired differences crossed zero, as did that for the linear SVC. The ranking therefore identified XGBoost as the model with the highest mean AUROC without establishing formal superiority over the nearest alternatives.

Table 5.2: Paired Phase 1 AUROC differences between XGBoost and each comparator. Positive values favour XGBoost. Percentile ranges are empirical across the 50 matched repetitions.

Comparator	Mean difference ± SD	2.5th–97.5th percentiles	XGBoost higher, %
LightGBM	0.008 ± 0.013	−0.011 to 0.038	72
Random forest	0.012 ± 0.012	−0.008 to 0.035	86
Logistic regression	0.017 ± 0.017	−0.015 to 0.044	82
Linear SVC	0.018 ± 0.019	−0.017 to 0.053	78
Gaussian naive Bayes	0.060 ± 0.023	0.017 to 0.106	100
Multilayer perceptron	0.078 ± 0.033	0.020 to 0.135	100

5.2.2 Exploratory Held-Out Permutation Importance

Held-out permutation importance was retained as an exploratory description of model-specific predictive association in Phase 1. Age and anterior-circulation location occupied the first two ranks in every model family, with their order reversed only for Gaussian naive Bayes. Admission haemoglobin ranked third in six of the seven families. Within XGBoost, the mean decreases in outer-test AUROC were 0.121 for age, 0.073 for anterior-circulation location and 0.053 for admission haemoglobin; the remaining mean importances were close to zero (Figure 5.2). These quantities do not identify causal factors, independent risk factors or biomarkers.

No directly comparable feature-level importance analysis was retained for reporting in Phases 2–4. Those phases used substantially different representations and model families, and their principal objective was to assess the predictive contribution of physiological information, temporal representation and multimodal integration rather than to rank individual variables. The 45-patient cohort and higher dimensionality of the later representations provided an additional reason not to promote potentially unstable feature-level rankings. The Phase 3 observation-mask sensitivity and Phase 4 fusion coefficients were not treated as feature importance.

5.2.3 Exploratory PCoA Recoding Sensitivity

The exploratory sensitivity analysis that reassigned PCoA from the anterior to the posterior circulation used the same cohort and validation partitions as the primary 13-predictor Phase 1 analysis. Mean AUROC decreased under the alternative coding for all seven model families (Table 5.3). XGBoost retained the highest mean AUROC at 0.700 ± 0.015. The empirical percentile range of the paired change crossed zero for six families and remained below zero for Gaussian naive Bayes. This exploratory result did not determine or validate the primary anatomical grouping.

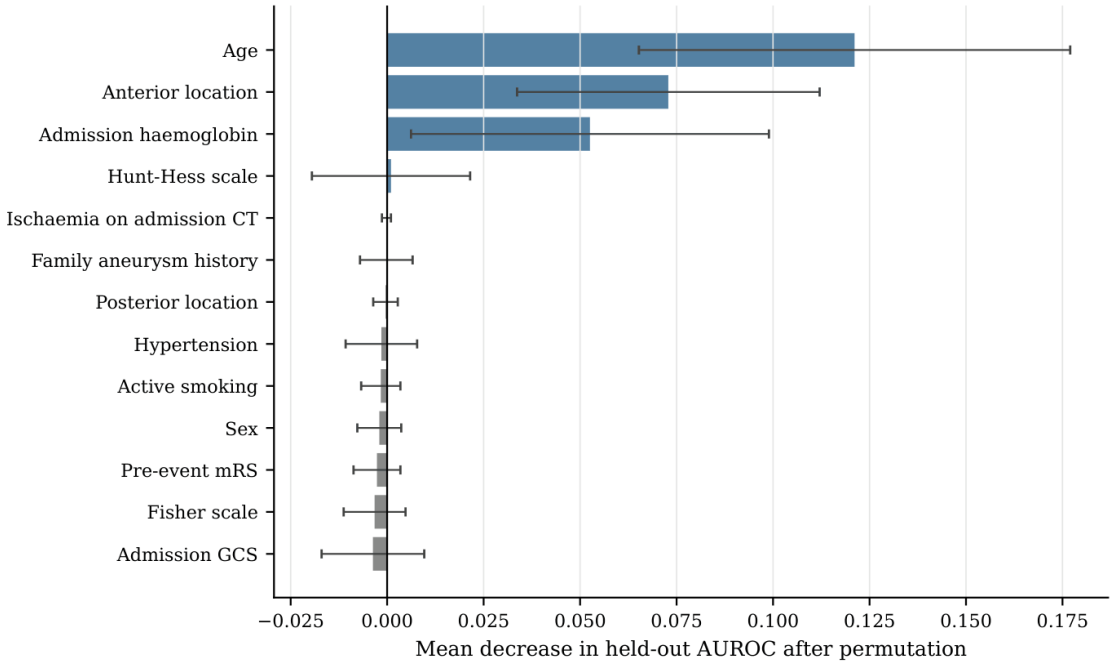


Figure 5.2: Exploratory held-out permutation importance for the Phase 1 XGBoost model. Bars show the mean change in outer-test AUROC after permutation of each predictor; whiskers show the standard deviation across the 250 outer-fold estimates. Negative and near-zero values indicate that permutation did not produce a consistent decrease in held-out discrimination. The estimates are model-specific predictive associations rather than causal effects.

Table 5.3: Exploratory Phase 1 sensitivity analysis with PCoA coded as posterior circulation. Paired differences are alternative coding minus the primary anterior-circulation coding.

Model	**Alternative AUROC**	**Mean paired difference**	**2.5th–97.5th percentiles**
XGBoost	0.700 ± 0.015	-0.020	-0.0404 to 0.0002
Random forest	0.695 ± 0.018	-0.014	-0.0377 to 0.0107
Linear SVC	0.693 ± 0.015	-0.009	-0.0242 to 0.0052
Logistic regression	0.693 ± 0.015	-0.010	-0.0275 to 0.0008
LightGBM	0.692 ± 0.018	-0.021	-0.0323 to 0.0016
Multilayer perceptron	0.632 ± 0.027	-0.010	-0.0518 to 0.0296
Gaussian naive Bayes	0.617 ± 0.032	-0.043	-0.0976 to -0.0084

5.3 Phase 2: Aggregated Physiological Prediction

5.3.1 Representation and Model Performance

The minimal physiological representation and its combination with the static block constituted the primary physiological and incremental analyses, respectively, with static-only performance serving as the matched reference. The extended representations remained secondary analyses. Mean AUROC across all 15 model–representation combinations is displayed in Figure 5.3, with the complete AUROC, AP and Brier-score summaries reported in Table C.2. Static-only logistic regression achieved the highest mean AUROC across the 15 combinations at 0.786 ± 0.022. Among physiology-only analyses, minimal-physiology random forest achieved the highest mean AUROC at 0.656 ± 0.070. The strongest physiology-bearing combined model was static-plus-minimal random forest, with AUROC 0.711 ± 0.057.

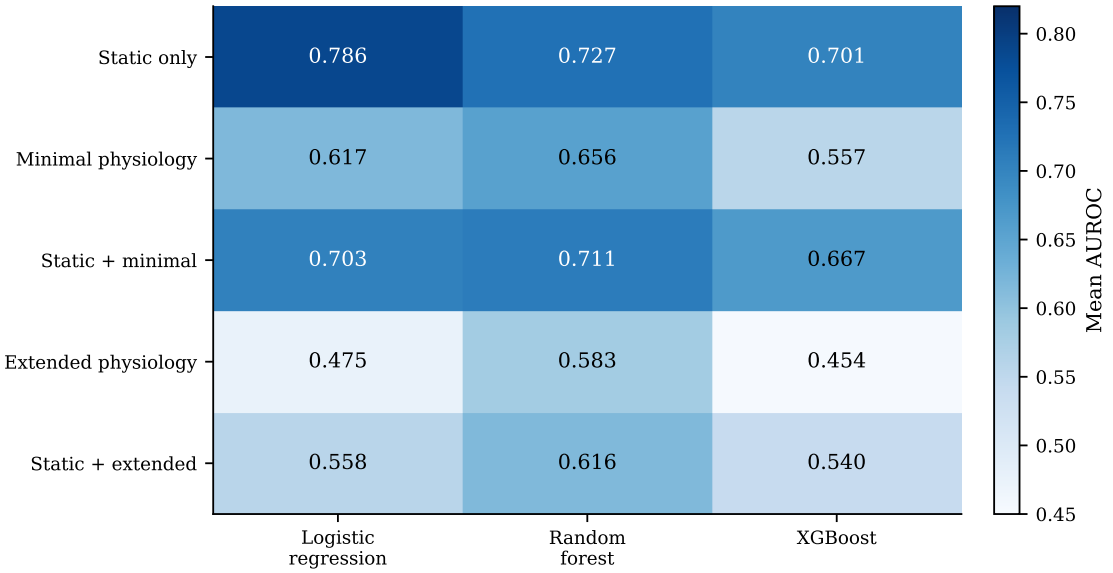


Figure 5.3: Mean Phase 2 AUROC for each model–representation combination across 50 repetitions. The minimal representations were primary; the extended representations were secondary.

Across all three model families, the extended physiology-only representation produced a lower mean AUROC than the corresponding minimal representation, and the same pattern was observed for static-plus-extended relative to static-plus-minimal physiology. The highest extended physiology-only and static-plus-extended mean AUROCs were 0.583 and 0.616, respectively, both obtained with random forest.

5.3.2 Incremental Value of Aggregated Physiology

The matched incremental comparison subtracted static-only performance from static-plus-minimal-physiology performance within each model family and repetition. Mean AUROC differences were

negative for logistic regression, random forest and XGBoost (Table 5.4). The combined representation was higher than its matched static baseline in 2%, 48% and 36% of repetitions, respectively. Complementary AP and Brier-score differences were also unfavourable on average and are reported in Table C.3. The tested minimal summaries therefore did not produce a positive mean AUROC increment over the matched static representation.

Table 5.4: Matched Phase 2 AUROC differences after adding minimal physiological summaries to the static representation. Differences are combined minus static; positive values favour the combined representation.

Model	**ΔAUROC, mean ± SD**	**2.5th–97.5th percentiles**	**Combined higher, %**
Logistic regression	-0.082 ± 0.058	-0.215 to -0.004	2
Random forest	-0.016 ± 0.061	-0.132 to 0.071	48
XGBoost	-0.033 ± 0.068	-0.157 to 0.070	36

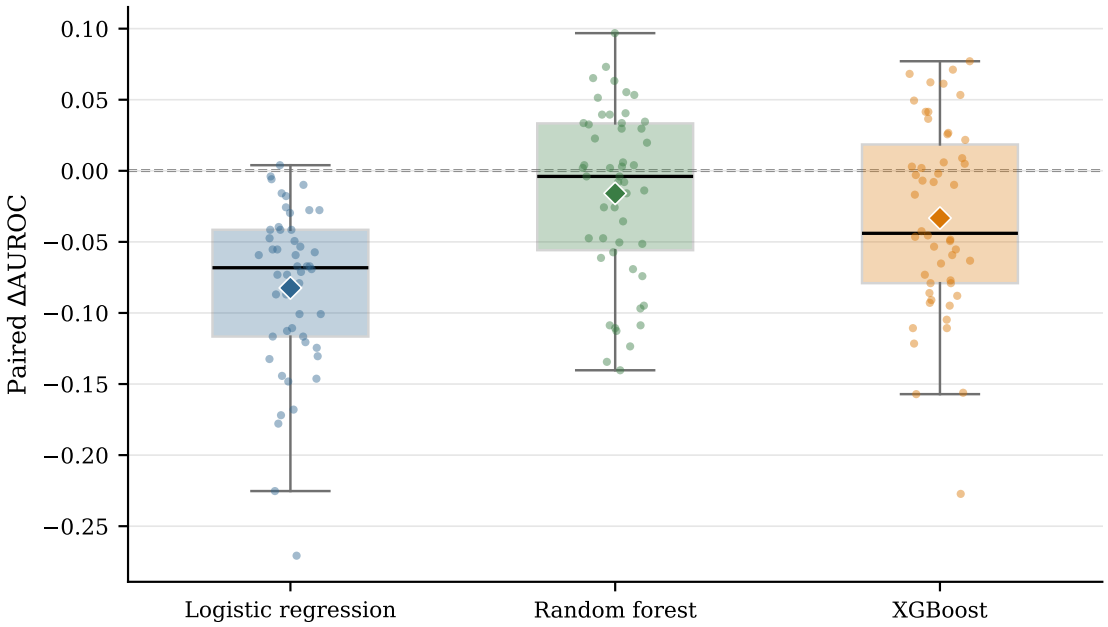


Figure 5.4: Repetition-level paired AUROC change after adding minimal physiological summaries to the static representation. Points show the 50 matched repetitions, boxes show the interquartile range, horizontal black lines show the median and diamonds show the mean. The dashed line denotes no change.

Phase 3 subsequently evaluated the same first-24-hour physiological information while preserving its minute-level temporal order.

5.4 Phase 3: Temporally Resolved Physiological Prediction

5.4.1 Primary, Sensitivity and Comparator Results

The primary value-only CNN achieved mean AUROC 0.617 ± 0.051. The prespecified value-plus-mask sensitivity achieved a numerically higher mean AUROC of 0.654 ± 0.062 but retained its sensitivity status, while ROCKET–Ridge achieved raw-score AUROC 0.553 ± 0.072. Table 5.5 summarises the primary discrimination results, while complementary AP and Brier-score values are reported in Table C.4.

Table 5.5: Phase 3 temporal-model discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values. ROCKET–Ridge AUROC is based on its raw discrimination score.

Analysis	Role	AUROC, mean ± SD	2.5th–97.5th percentiles
Value-only CNN	Primary	0.617 ± 0.051	0.510–0.694
Value-plus-mask CNN	Sensitivity	0.654 ± 0.062	0.525–0.767
ROCKET–Ridge	Comparator	0.553 ± 0.072	0.423–0.660

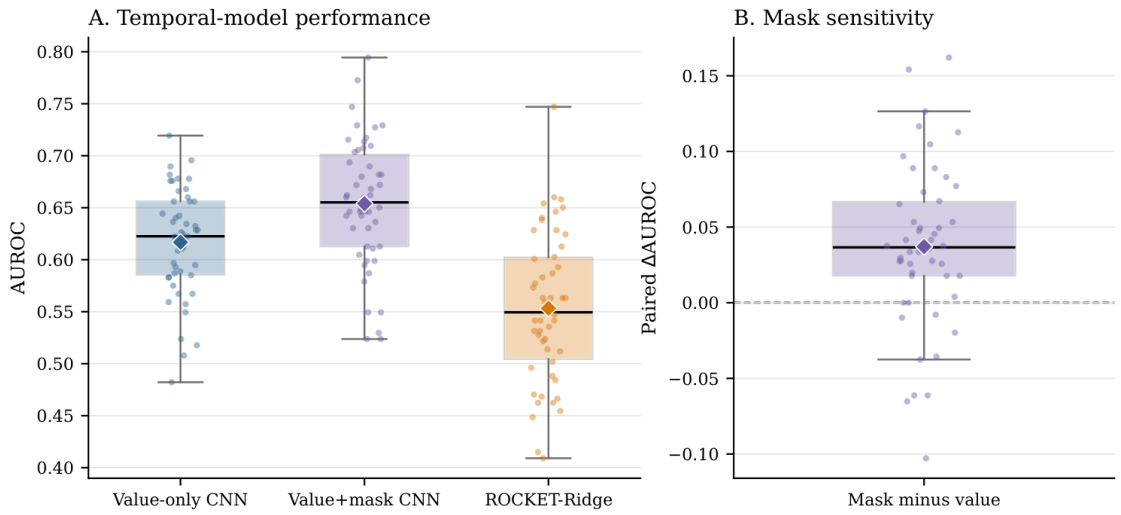


Figure 5.5: Phase 3 temporal-model results across the 50 matched repetitions. Panel A shows repetition-level AUROC for the primary value-only CNN, the value-plus-mask sensitivity and ROCKET–Ridge. Panel B shows the paired AUROC change for the mask sensitivity relative to the value-only CNN. Boxes show the interquartile range, black lines the median and diamonds the mean; the dashed line denotes no paired change.

5.4.2 Paired and Probability-Based Findings

Within Phase 3, the value-plus-mask sensitivity exceeded the primary value-only CNN by a mean paired AUROC difference of 0.037 and was higher in 78% of repetitions. Its empirical paired

percentile range extended from −0.064 to 0.148, while ROCKET–Ridge was lower than the value-only CNN by a mean paired AUROC difference of 0.064 (Table 5.6).

Table 5.6: Paired Phase 3 AUROC comparisons within the temporal analyses. Differences are analysis A minus analysis B; positive values favour analysis A.

Comparison: A minus B	**ΔAUROC**	**2.5th–97.5th percentiles**	**A higher, %**
Value-plus-mask CNN minus value-only CNN	+0.037	−0.064 to 0.148	78
ROCKET–Ridge minus value-only CNN	−0.064	−0.224 to 0.062	20

For descriptive cross-phase context, the value-only CNN was lower than the strongest Phase 2 physiology-only analysis, minimal-physiology random forest, by a mean paired AUROC difference of 0.039 and was higher in 32% of repetitions. The corresponding comparison for the value-plus-mask sensitivity is reported in Table C.5. These comparisons did not constitute equivalence or non-inferiority analyses, and they do not isolate temporal ordering because model family, dimensionality, training and probability handling also differed between Phases 2 and 3. The complete paired Phase 3 results, including these cross-phase comparisons and complementary AP and Brier-score differences, are reported in Table C.5.

All three temporal analyses had mean Brier scores above the approximately 0.250 prevalence-only benchmark. For ROCKET–Ridge, monotone-calibrated probability AUROC was 0.502 ± 0.073 and probability AP was 0.561 ± 0.061, compared with raw-score AUROC 0.553 ± 0.072 and AP 0.614 ± 0.058. The outer-training-prevalence calibration fallback was used in 36 of the 250 outer folds. These probability-based findings remained secondary to the raw-score discrimination endpoint.

Phase 4 then tested whether temporal CNN prediction provided incremental information when combined with a matched static branch, using the late-fusion procedure described in Section 4.6.4.

5.5 Phase 4: Static–Temporal Late Fusion

5.5.1 Static and Fusion Performance

The contemporaneously generated fixed-C static logistic-regression comparator achieved mean AUROC 0.799 ± 0.019. The primary fusion of this static branch with the value-only CNN achieved AUROC 0.760 ± 0.034, while the mask-sensitivity fusion achieved 0.771 ± 0.028. Table 5.7 focuses on the matched static and fusion analyses; the standalone temporal-component results were reported in Phase 3, and their outer-test probabilities were reused unchanged in Phase 4. These component results are included alongside complementary AP and Brier-score values in Table C.6. The Phase 4 static estimate is not interchangeable with the Phase 2 static logistic-regression estimate because the two analyses used different locked fitting and probability-handling procedures; all Phase 4 incremental comparisons therefore use the contemporaneous fixed-C comparator.

Table 5.7: Phase 4 static and late-fusion discrimination. Values are mean ± SD across 50 repetitions; percentile ranges are empirical across repetition-level AUROC values.

Analysis	Role	AUROC, mean ± SD	2.5th–97.5th percentiles
Phase 4 fixed-C static LR	Comparator	0.799 ± 0.019	0.767–0.828
Static LR + value-only CNN	Primary fusion	0.760 ± 0.034	0.671–0.814
Static LR + value-plus-mask CNN	Sensitivity fusion	0.771 ± 0.028	0.713–0.810

5.5.2 Paired Incremental Value

Relative to the matched Phase 4 static comparator, the primary fusion had a mean paired AUROC difference of -0.039 and was higher in 10% of repetitions (Table 5.8; Figure 5.6). The mask-sensitivity fusion had a mean paired AUROC difference of -0.028 relative to the same static comparator and was higher in 12% of repetitions. The mask fusion was numerically higher than the primary fusion by a mean paired AUROC difference of 0.011, with an empirical percentile range that crossed zero. Both fusions also had lower mean AP and higher mean Brier scores than the static comparator; the corresponding paired differences are reported in Table C.7.

Table 5.8: Paired Phase 4 AUROC comparisons across matched repetitions. Differences are mean within-repetition differences for analysis A minus analysis B; positive values favour analysis A. Percentile ranges are empirical across the 50 paired AUROC differences.

Comparison: A minus B	ΔAUROC	2.5th–97.5th percentiles	A higher, %
Primary fusion minus Phase 4 static LR	-0.039	-0.114 to 0.008	10
Mask-sensitivity fusion minus Phase 4 static LR	-0.028	-0.099 to 0.014	12
Mask-sensitivity fusion minus primary fusion	$+0.011$	-0.049 to 0.071	66

5.5.3 Cross-Fitting Diagnostics

Mean temporal inner out-of-fold AUROC was 0.601 for the primary analysis and 0.635 for the mask sensitivity; the corresponding meta-model out-of-fold AUROCs were 0.731 and 0.739. These training-side diagnostics were not substituted for the untouched outer-test comparisons. Differences in training-set size and aggregation, together with the residual cross-row dependence described in Section 4.7, preclude interpreting the difference between internal and outer mean AUROCs as a matched estimate of bias. Static–temporal score correlations and fusion coefficients were likewise treated as training-side descriptors rather than feature-importance estimates or evidence of incremental outer-test value. The diagnostics in Table C.9 were summarised directly across the 250 outer-training fits, without intermediate repetition-level averaging.

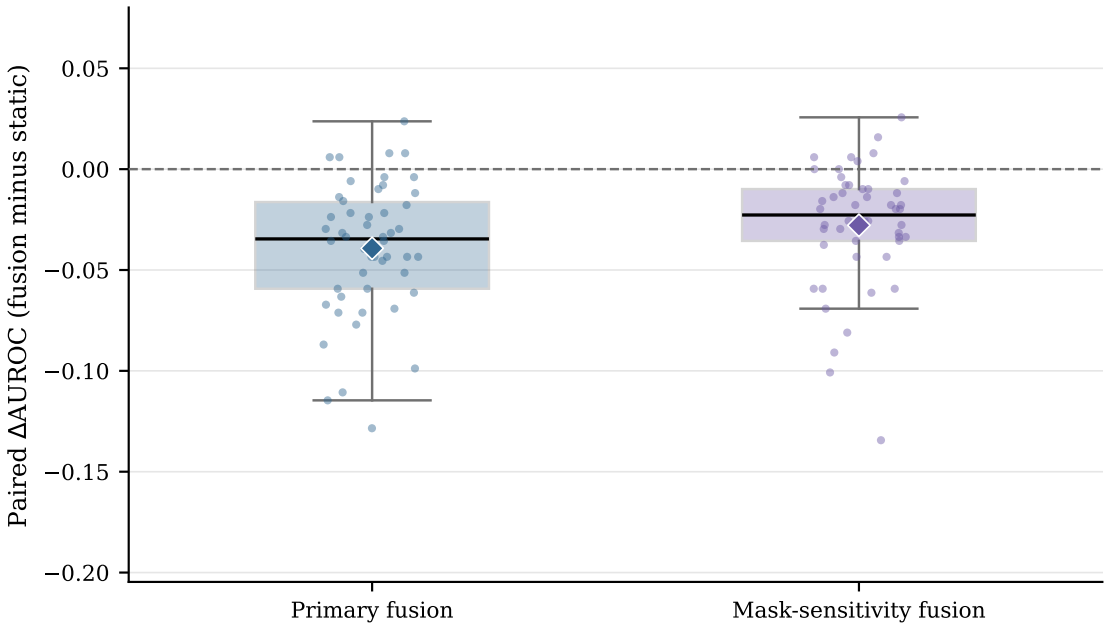


Figure 5.6: Repetition-level paired AUROC difference between each Phase 4 fusion analysis and the matched fixed-C static logistic-regression comparator. Points show the 50 matched repetitions, boxes show the interquartile range, black lines the median and diamonds the mean. The dashed line denotes no incremental change.

5.6 Secondary Operating-Point and Calibration Results

The screening thresholds were constructed within the corresponding outer-training procedures by requiring inner out-of-fold sensitivity of at least 0.85 before maximising specificity. This inner criterion did not guarantee the same sensitivity on untouched outer-test data: across the analysis-defining models in Table C.8, mean outer sensitivity ranged from 0.731 to 0.873. In the monitored cohort, mean specificity was lower than mean sensitivity for every analysis-defining model, with the lowest mean specificity observed for the Phase 3 ROCKET–Ridge comparator. Calibration slopes for the Phase 3 temporal analyses were also well below one. These threshold-dependent and calibration findings were secondary analytical descriptions rather than externally validated clinical cut-offs. Because thresholds were selected independently within each outer-training procedure, they do not represent a single deployable cut-off or directly comparable fixed thresholds across models or phases. Calibration intercept and slope were estimated jointly from each repetition's complete outer out-of-fold probability vector; these diagnostic fits did not modify the evaluated predictions (Section 4.8). The complete screening operating-point and calibration summaries are reported in Table C.8.

5.7 Summary of Main Results

Across the matched 45-patient analyses, static-only models produced the highest mean AUROCs in Phases 2 and 4. Adding aggregated physiological summaries in Phase 2 and incorporating temporal predictions through late fusion in Phase 4 did not produce a positive mean paired AUROC increment over their respective matched static comparators. Phase 3 evaluated temporal physiology without a same-phase static comparator, so its cross-phase comparisons were descriptive rather than direct incremental tests. These findings are restricted to the evaluated representations, model families and internally validated cohort and do not rely on a direct numerical comparison with the broader Phase 1 cohort.

Chapter 6

Discussion

6.1 Principal Findings

The central finding of this study was that incorporating first-24-hour physiological information, including preservation of minute-level temporal order, did not by itself translate into dependable incremental predictive value beyond static clinical information. Admission and early clinical variables provided moderate, although incomplete, discrimination of subsequent clinician-assigned cerebral vasospasm in the broader clinical cohort. Within the monitored cohort, aggregated physiology also showed standalone discriminatory information, but the evaluated temporal pipelines did not clearly improve on the strongest aggregate physiology-only analysis. Neither adding aggregated physiology to static predictors nor combining static and temporal predictions through late fusion produced a positive mean paired AUROC increment over the corresponding matched static comparator. The numerically more favourable value-plus-mask CNN suggested that observation-status information may contribute to prediction, while remaining a sensitivity analysis rather than the primary temporal finding.

These results distinguish standalone physiological discrimination, temporal representation and incremental value over a clinical baseline. The absence of a positive mean increment in the evaluated comparisons does not establish redundancy between clinical and physiological information or exclude useful physiological information under other representations or study conditions.

The scope of this interpretation is defined by the outcome. The study predicted the binary vasospasm diagnosis assigned in routine clinical care, with surveillance based primarily on transcranial Doppler (TCD), rather than independently adjudicated angiographic narrowing, delayed cerebral ischaemia (DCI), or functional outcome. As discussed in Chapter 2, these outcomes are related but not interchangeable (Vergouwen et al. 2010; Macdonald 2014). The findings therefore concern subsequent clinician-assigned vasospasm within the evaluated clinical setting rather than the broader spectrum of secondary cerebrovascular injury after aSAH.

6.2 Static Clinical Information as a Predictive Baseline

Phase 1 established that admission and early clinical information contained a consistent internally validated discriminatory signal. Across 220 admission episodes from 219 patients, XGBoost achieved the highest mean AUROC of 0.720, while logistic regression achieved 0.703 and several other model families produced similar values. The modest separation between the highest-performing models identifies a useful static modelling reference without establishing a clear advantage for one algorithm. More broadly, greater algorithmic flexibility did not consistently improve discrimination over logistic regression within the same 13-predictor representation, although this does not imply that predictor relationships are intrinsically linear or that static prediction has reached a universal performance ceiling.

This pattern is consistent with selected findings from the Literature Review. Gollwitzer et al. (2025) reported similar discrimination for logistic regression and a support vector machine (SVM) using clinical information, while Kim et al. (2023) reported comparable discrimination for logistic regression and XGBoost in a study-defined DCI prediction task. These examples support the value of retaining comparatively simple clinical models as substantive references when evaluating more complex approaches. However, differences in endpoints, populations, predictor availability and validation prevent direct numerical ranking across studies, while higher discrimination reported elsewhere, including by Kim et al. (2021), argues against treating the present Phase 1 result as a general ceiling for static prediction.

The exploratory permutation analysis provided additional information about model reliance within the clinical representation. Age and anterior-circulation location occupied the first two importance ranks across all seven model families, while admission haemoglobin ranked third in six. Their recurrence suggests consistent predictive contribution within the fitted models, but permutation importance does not establish an independent risk factor, direction of association or causal mechanism. These results should therefore be interpreted as properties of the predictive models rather than as a ranking of biological or clinical importance.

The prominence of age is consistent with previous clinical evidence identifying younger age as being associated with greater susceptibility to cerebral vasospasm after aSAH (Rumalla et al. 2021; Gollwitzer et al. 2025). Aneurysm location has likewise been associated with vasospasm risk in previous studies, although the direction and strength of territory-specific associations have varied across cohorts (Qureshi et al. 2000; Gollwitzer et al. 2025). The importance of anterior-circulation location in the present models should therefore be interpreted as predictive rather than as evidence of a uniform anatomical risk effect.

PCoA was classified with the anterior circulation in the primary analysis on anatomical grounds supported by the literature. Because this classification is not uniform across the literature, posterior-circulation coding was evaluated as an exploratory sensitivity analysis. The alternative coding reduced mean AUROC across all seven models, a pattern directionally consistent with the primary coding choice but which did not determine or validate the anatomical grouping (Table 5.3).

Phase 1 characterises what the available admission and early clinical information could achieve

in the broader clinical cohort, but it is not the numerical baseline for assessing physiological increment in the monitored patients. The later static estimates arose from a smaller selected cohort and, in Phase 4, a different fitting procedure. Differences between these estimates should therefore not be interpreted as changes in static predictive ability across phases.

6.3 What Was Learned from First-24-Hour Physiology

A compact summary of first-24-hour monitoring retained some information about subsequent recorded vasospasm. The primary 20-feature minimal representation relied principally on medians and interquartile ranges, with directional quantiles for the exported ICP extrema, and its strongest physiology-only model achieved a mean AUROC of 0.656. The physiological analyses were designed primarily to compare representations rather than to attribute performance to individual signals.

This finding is conceptually consistent with the aggregated monitoring approach of Roederer et al. (2014), in which repeated physiological observations were similarly converted into patient-level descriptors. Differences in observation period, accompanying laboratory information, endpoint and validation prevent direct performance comparison, but both studies show that aggregation can retain predictive information from monitoring. Within the present cohort, the more parsimonious representation was also consistently more favourable than the extended alternative. Expanding from 20 to 76 physiological features, or to 89 when combined with static predictors, did not improve mean AUROC across the three model families. The result supports parsimony in this small cohort without establishing the minimal representation as universally optimal.

The physiological rationale for the monitored signals nevertheless remains relevant. Systemic and intracranial pressures, perfusion, compensatory reserve and cerebrovascular reactivity are all related to the evolving consequences of aSAH (Czosnyka and Pickard 2004; Budohoski et al. 2013; Lidington et al. 2021). They are, however, broad dimensions of neurocritical physiology rather than established specific precursors of the clinician-assigned vasospasm outcome. For example, evidence relating impaired pressure reactivity to secondary cerebral injury does not establish PRx as a reliable precursor of subsequent vasospasm. The signals may therefore contain clinically meaningful physiological variation without that variation necessarily aligning strongly with the prediction target evaluated here.

Preserving minute-level temporal order did not clearly improve physiological prediction under the evaluated pipelines. The primary value-only CNN achieved a mean AUROC of 0.617 compared descriptively with 0.656 for the minimal-physiology random forest, while ROCKET–Ridge achieved a raw-score AUROC of 0.553. The aggregate–temporal analysis compared complete predictive representations rather than a one-factor ablation of temporal ordering: patients, observation windows and outer partitions were shared, while each representation used an appropriate modelling pipeline. The result therefore does not imply that temporal order itself was detrimental; it shows that explicit preservation of the minute-level sequence did not improve prediction under the evaluated temporal approaches. Moreover, the aggregated inputs were not entirely devoid

of temporal information because PRx, RAP and RAC already summarised recent physiological relationships, while the extended representation also included slopes.

This interpretation is consistent with the broader lesson from Sen et al. (2025), whose sequence-based model combined hourly physiology, daily TCD measurements and serum sodium but did not dominate logistic regression across all reported metrics. Differences in architecture, temporal resolution, modalities and outcome definition prevent treating the two studies as direct replications, but neither supports the assumption that preserving temporal structure must automatically outperform simpler representations.

The value-plus-mask sensitivity produced a higher mean CNN AUROC of 0.654, a paired difference of $+0.037$ relative to the value-only model. The masks explicitly distinguished observed from subsequently filled values, suggesting that observation status may itself carry useful predictive information. However, such information can reflect signal availability, acquisition processes or clinical workflow as well as underlying physiology. The sensitivity analysis therefore supports cautious attention to how physiological data are observed without establishing missingness as a biological biomarker or replacing the value-only CNN as the primary temporal analysis.

6.4 Incremental Physiological Value and Static–Temporal Fusion

The most direct evidence concerning incremental physiological value comes from comparisons made within the same patients and evaluation partitions. In Phase 2, adding minimal physiology produced negative mean paired AUROC changes for logistic regression, random forest and XGBoost: -0.082, -0.016 and -0.033, respectively. Complementary AP and Brier-score differences were likewise unfavourable on average, with the random-forest comparison closest to a neutral effect. Across the evaluated model families, the tested physiological summaries therefore did not produce a positive mean predictive increment over matched static information.

Standalone physiological discrimination is compatible with this result. Physiology may contain outcome-associated information that overlaps with variables already represented clinically, or complementary information may be difficult to estimate reliably from the available patients and representations. The present analyses cannot distinguish these possibilities. The larger ICU study by Zarrin et al. (2024) demonstrates that longitudinal physiological and laboratory information can support prediction in a substantially larger and differently defined setting, but differences in endpoint, observation window, predictor content and validation mean that its findings do not imply that a positive increment should necessarily have emerged here.

Phase 4 tested complementarity at the prediction level by combining a static score with a temporal CNN score. The primary fusion achieved a mean AUROC of 0.760 compared with 0.799 for its matched static LR comparator, corresponding to a mean paired difference of -0.039. The mask-sensitivity fusion achieved a mean AUROC of 0.771, with a mean paired difference of -0.028 relative to the same static comparator. Thus, the numerically stronger mask-based temporal prediction did not translate into a positive mean discrimination increment beyond static

prediction. Both fusions also had lower mean AP and higher mean Brier scores than the static comparator (Table C.7).

Taken together, these findings refine the conceptual "Dengler Paradox" introduced earlier in this dissertation. The term is used here as a thesis-specific description of the possibility that greater algorithmic sophistication cannot necessarily overcome an information bottleneck in the available inputs. The limited separation between several Phase 1 model families is compatible with this idea and with the motivation drawn from Dengler et al. (2021). The later phases add an important qualification: increasing input richness through physiological measurements or minute-level temporal structure also did not automatically generate dependable incremental prediction. Richer data improve prediction only when their additional signal can be represented, estimated reliably and translated into better performance on held-out patients.

6.5 Methodological Strengths and Limitations

Several design choices strengthen the internal validity of the comparisons. Patient-level grouping prevented the same individual from contributing to both training and test data, while Phases 2–4 used the same monitored patients, labels and outer partitions for paired analyses. Data-dependent preprocessing and model-development procedures were conducted without access to the corresponding outer-test outcomes, and the distinction between primary, secondary and sensitivity analyses preserved the intended analytical hierarchy. Performance was calculated from complete outer out-of-fold prediction vectors rather than averages of metrics from individual small test folds.

The internal training procedures nevertheless require qualification. For the standalone Phase 3 CNN models, the same inner validation outcomes selected the early-stopping checkpoint and informed threshold construction from that checkpoint's predictions. This dual role could make the internal threshold assessment optimistic, although the outer-test fold was excluded from both operations.

In Phase 4, each patient's temporal fusion-training score was generated without that patient contributing to CNN parameter fitting, population-dependent preprocessing or epoch selection. However, the second-level meta-cross-fit reused the assembled base-model out-of-fold feature matrix rather than regenerating its features within each meta-training split. A meta-validation patient's outcome could therefore influence the base models generating other patients' feature rows used to fit the meta-model. Fusion-training features also came from base models fitted to 28–29 patients, whereas outer-test scores came from models fitted to 36. These dependencies and training-size differences limit interpretation of internal diagnostics and threshold estimates. The outer-test outcomes remained excluded from model development, but the internal stacking procedure was not fully independent end to end (Section 4.7).

Repeated cross-validation also reduced dependence on one favourable partition, but the repeated estimates reused the same patient outcomes and are therefore not independent replications. Their variability describes sensitivity to data partitioning rather than inferential confidence intervals. More importantly, the physiological analyses ultimately contained 45 independent monitored

patients, irrespective of the large number of measurements within each temporal sequence. This limited sample size constrains the precision of model comparison and the amount of independent information available for learning complex representations.

The temporal eligibility procedure provided an important safeguard against using physiology recorded at or after vasospasm. Because event timing was available by day, positive patients were retained only when the complete 24-hour predictor window ended no later than the earliest possible boundary of the recorded event day. This strengthened the interpretation of the physiological analyses as pre-event prediction, while selecting a monitored cohort in which cases beginning monitoring too close to the event were less likely to remain eligible.

Monitoring initiation also did not represent a common biological time origin, and the target included any subsequently recorded vasospasm rather than an event within a fixed post-landmark horizon. The same relative minute could therefore correspond to different stages of disease evolution, and positive outcomes could occur at different temporal distances from the observation window. This heterogeneity may make common precursor patterns more difficult to identify, although the study does not establish temporal misalignment as the cause of the observed performance. The first 24 hours should consequently be understood as a standardised observation window rather than as a fixed biological period or prediction lead time.

Physiological completeness was high overall, with 93.09% of tensor cells originally observed, but some recordings contained substantial gaps. Fold-contained interpolation and imputation preserved validation separation, although missing trajectories cannot be reconstructed from imputed values. The temporal models also used processed ICM+ trends at nominal one-minute resolution rather than the archived high-frequency waveforms. They therefore evaluated the evolution of processed trends and derived indices, not beat-to-beat dynamics, pulse morphology or other sub-minute phenomena. Whether finer-scale physiological representations contain additional predictive information for this endpoint was not tested.

Finally, the outcome reflected clinician-assigned vasospasm under routine institutional surveillance, based primarily on TCD, with event timing available at daily resolution. This should be considered when comparing the findings with angiographic, symptomatic, treatment-defined or otherwise standardised outcomes. The single-centre design and absence of independent external validation further restrict conclusions about transportability. The reported performance should therefore be interpreted as internally validated evidence for the evaluated cohorts and analytical pipelines rather than as validation of a clinical prediction tool.

6.6 Clinical, Methodological and Generalisability Implications

The evaluated models were more informative as tools for discrimination than as clinically reliable risk estimators. No model was externally validated or established as a clinical risk calculator, and probability quality was particularly limited for the standalone Phase 3 temporal models, whose mean Brier scores exceeded the approximately 0.250 prevalence-only reference and whose calibration slopes were well below one. The Phase 4 fusions had mean calibration intercepts closer

to zero and mean slopes closer to one than the matched static comparator, but their higher mean Brier scores indicated poorer overall probabilistic accuracy relative to that comparator (Tables C.6 and C.8). More favourable values for these calibration diagnostics therefore did not establish better overall probability estimates or clinical reliability.

Similarly, training-selected screening thresholds did not establish a clinically validated operating point: a target sensitivity of at least 0.85 during training translated into mean outer-test sensitivities ranging from 0.731 to 0.873 across the analysis-defining models. At their respective selected operating points, the Phase 4 fusions had slightly higher mean sensitivity but lower mean specificity than the matched static comparator (Table C.8). From a potential screening perspective, this trade-off is clinically relevant because greater sensitivity reduces missed subsequent vasospasm cases at the cost of more false-positive alerts, although no clinical operating threshold was established in this study. These results support interpretation of AUROC and threshold analyses as predictive evaluation rather than evidence for deployment.

This distinction is consistent with the Literature Review. Zarrin et al. (2024) illustrated the practical trade-off between high sensitivity and limited specificity for a treatment-related endpoint, while the high sensitivity reported across the heterogeneous deep-learning studies reviewed by Palermo et al. (2025) does not provide a transferable threshold for the present cohort. Prevalence-dependent measures are also especially sensitive to the selected monitored population, in which vasospasm prevalence was 51.1%. The present study therefore does not establish that model-guided changes in surveillance or treatment would improve clinical outcomes.

Generalisability depends on population, monitoring practice and outcome definition as well as algorithm choice. The study was conducted at one centre, and the monitored cohort reflected local indications for intensive monitoring, signal availability and conservative temporal eligibility. Differences in these factors could alter both predictor distributions and their relationship with the target elsewhere, particularly for observation-mask information that may partly encode acquisition or workflow. This concern is consistent with the broader literature, where the review judged 18 of 19 included studies to have high risk of bias and overall confidence in the evidence to be low. The principal contribution of the present study therefore lies in its internally consistent evaluation of static, aggregated physiological, temporal and fused representations, including matched tests of incremental physiological value.

The findings place greater emphasis on the quality and alignment of predictive information than on model complexity alone. The absence of dependable physiological increment does not diminish the biological or clinical relevance of neuromonitoring, nor does it establish that temporal modelling in general is ineffective. The most important next evidential steps are larger independent multicentre cohorts, prospectively standardised and documented outcome ascertainment, and more precise event timing. Finer physiological representations may also test information not captured by the processed minute-level trends, but increasing model complexity within the same limited evidence base would not resolve the principal uncertainties identified here. For eventual clinical application, any predictive approach would additionally require reproducible performance, reliable probability estimates and evidence that its predictions meaningfully support

clinical decision-making.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This dissertation investigated whether admission and early clinical information and routinely collected physiology from the first 24 hours of neuromonitoring could support prediction of subsequent clinician-assigned cerebral vasospasm after aSAH. In answer to the overarching research question, admission and early clinical information supported moderate but incomplete discrimination, while first-24-hour physiology contained some standalone predictive signal but did not provide a consistent positive increment beyond matched static information. Separately, explicitly preserving minute-level temporal structure did not provide a clear predictive advantage over the aggregated physiological representation under the evaluated pipelines. These findings concern the binary vasospasm diagnosis assigned in routine clinical care, with surveillance based primarily on TCD, and should not be extended to independently adjudicated angiographic vasospasm, DCI, or functional outcome.

Phase 1 directly addressed RQ1 by showing that admission and early static clinical information provided meaningful but incomplete discrimination of subsequent vasospasm. Across 220 admission episodes from 219 patients, XGBoost achieved the highest mean AUROC of 0.720, while logistic regression achieved 0.703 and several other model families produced similar discrimination. Greater algorithmic flexibility did not consistently improve discrimination over logistic regression within the same 13-predictor representation. These findings therefore support **H1**: static clinical information provided a useful predictive baseline, although discrimination remained incomplete.

The comparison between aggregated physiology in Phase 2 and sequence-preserving modelling in Phase 3 addressed RQ2. First-24-hour physiology retained predictive information when represented through compact patient-level summaries, with the strongest minimal-physiology model achieving a mean AUROC of 0.656, whereas the primary value-only CNN achieved 0.617 and the value-plus-mask sensitivity reached 0.654. The minimal physiological representation was also consistently more favourable than the extended alternative across the evaluated Phase 2 model families. Because these analyses compared complete aggregate and temporal modelling pipelines

rather than an isolated manipulation of temporal ordering, they do not establish that temporal structure itself was uninformative. They do show that explicit preservation of the minute-level sequence did not improve physiological prediction under the evaluated temporal approaches. Accordingly, **H2 was not supported by the observed results**.

Finally, RQ3 was addressed through the matched incremental analyses in Phases 2 and 4. In Phase 2, adding the primary aggregated physiological representation produced negative mean paired AUROC changes across logistic regression, random forest and XGBoost. In Phase 4, the primary late-fusion model achieved a paired mean AUROC difference of -0.039 relative to its matched static comparator, while the mask-sensitivity fusion produced a difference of -0.028. Both fusion analyses also had lower mean AP and higher mean Brier scores than the matched static comparator. The physiological information therefore did not translate into reliable additional predictive performance once static clinical prediction was already available. Within the evaluated cohort and representations, **H3 was therefore not supported**.

The broader Phase 1 cohort and the selected 45-patient monitored cohort remain distinct analytical populations, so their numerical results are not interchangeable. These conclusions therefore remain specific to the internally validated cohorts, representations and prediction framework evaluated here.

The main scientific contribution of this work is an internally consistent evaluation of static, aggregated physiological, temporal and fused prediction strategies, distinguishing standalone physiological discrimination from incremental predictive value beyond a matched clinical baseline. The findings show that richer data and more complex representations did not automatically translate into dependable incremental prediction under the evaluated conditions.

The evaluated models do not yet support clinical deployment, given the absence of external validation, the limitations in probability estimation and the lack of clinically validated operating thresholds. Nevertheless, this work contributes a methodologically grounded step towards the future integration of machine-learning decision support into neurocritical care, for which predictive performance must ultimately be accompanied by demonstrated clinical utility.

7.2 Future Work

The immediate priority should be to strengthen the evidence base and data quality rather than simply expand the set of model architectures. Larger independent cohorts assembled across multiple centres are required to enable external validation and determine whether predictive performance is reproducible across patient populations, monitoring practices and clinical settings. Outcome ascertainment should also be prospectively standardised and more completely documented, with clearer diagnostic criteria and more precise characterisation of when events occur.

Future studies could also investigate more granular outcome definitions rather than a single binary vasospasm label. Where sufficiently reliable clinical, sonographic or angiographic information is available, ordinal or multiclass formulations could distinguish different degrees of vasospasm severity. Such an extension would require clearly defined and consistently applied

severity criteria and sufficient numbers of patients within each class to avoid replacing one source of heterogeneity with another.

More precise event timing would permit physiological observation windows to be aligned with biologically meaningful landmarks while ensuring that the complete predictor interval remains pre-event. Where possible, future studies should also define clear post-landmark prediction horizons instead of grouping all subsequently recorded events into a single outcome. These changes would reduce heterogeneity in the temporal relationship between the physiological record and the event being predicted.

Higher-resolution modelling could evaluate physiological information not represented by the processed one-minute trends used here. A necessary enabling step would be the development and validation of procedures for physiological artefact detection and the reconstruction of short corrupted or missing segments in continuous ICP and ABP waveforms. By preserving temporal continuity, such procedures could make longer and more complete high-frequency observation periods available to vasospasm-prediction architectures. This should be treated as a data-quality strategy whose effect on prediction requires empirical evaluation, not as an intervention already shown to improve predictive performance.

ICP waveform morphology provides one specific extension of this high-frequency approach. During methodological development, existing P1–P3 peak-detection approaches were explored on archived 250-Hz ICM+ waveforms. A small manually annotated subset of approximately 50 pulses with relatively clear morphology provided preliminary evidence that automated peak identification was feasible, although detection was not consistently correct. Identification became less reliable in more complex morphologies, particularly when individual peaks were poorly separated or merged, making the correspondence between detected maxima and physiological P1–P3 components increasingly ambiguous. These annotations were used only for exploratory assessment and were not treated as a formal ground-truth dataset.

Reliable quantitative validation would require a substantially larger expert-annotated local reference set, including difficult and morphologically ambiguous waveforms, with appropriate adjudication when individual components cannot be identified confidently. Future work could establish such a dataset, validate and, where necessary, adapt the detector to the local recording domain, and subsequently derive features including P1, P2 and P3 amplitudes, amplitude ratios, latencies and their temporal evolution. This staged approach is consistent with evidence that supervised ICP peak detection requires labelled landmark ground truth and that continuous tracking across successive pulses can improve robustness to noise and temporary artefacts (Wei et al. 2024). Neither successful artefact handling nor validated morphology extraction should be assumed to improve vasospasm prediction without direct evaluation.

Continuous or repeatedly updated risk estimation could also be investigated alongside improvements in data quality and temporal alignment. Any subsequent clinical application would nevertheless require external validation, reproducible discrimination, reliable probability estimation and clinically meaningful evaluation.

References

Abdulazim, Amr, Marina Heilig, Gabriel Rinkel, and Nima Etminan. 2023. "Diagnosis of Delayed Cerebral Ischemia in Patients with Aneurysmal Subarachnoid Hemorrhage and Triggers for Intervention." *Neurocritical Care* 39 (2): 311–319. https://doi.org/10.1007/s12028-023-01812-3.

Alexopoulos, Georgios, Justin Zhang, Ioannis Karampelas, Maheen Khan, et al. 2022. "Applied Forecasting for Delayed Cerebral Ischemia Prediction Post Subarachnoid Hemorrhage: Methodological Fallacies." *Informatics in Medicine Unlocked* 28:100817. https://doi.org/10.1016/j.imu.2021.100817.

Aries, Marcel J. H., Marek Czosnyka, Karol P. Budohoski, Luzius A. Steiner, Andrea Lavinio, Angelos G. Kolias, et al. 2012. "Continuous Determination of Optimal Cerebral Perfusion Pressure in Traumatic Brain Injury." *Critical Care Medicine* 40 (8): 2456–2463. https://doi.org/10.1097/CCM.0b013e3182514eb6.

Azzam, Ahmed Y., Dhrumil Vaishnav, Muhammed Amir Essibayi, et al. 2024. "Prediction of Delayed Cerebral Ischemia Followed Aneurysmal Subarachnoid Hemorrhage. A Machine-Learning Based Study." *Journal of Stroke and Cerebrovascular Diseases* 33 (4). https://doi.org/10.1016/j.jstrokecerebrovasdis.2023.107553.

Baron, Jean-Claude. 2001. "Perfusion Thresholds in Human Cerebral Ischemia: Historical Perspective and Therapeutic Implications." *Cerebrovascular Diseases* 11 (Suppl. 1): 2–8. https://doi.org/10.1159/000049119.

Beqiri, Erta, Peter Smielewski, Chiara Robba, et al. 2019. "Feasibility of Individualised Severe Traumatic Brain Injury Management Using an Automated Assessment of Optimal Cerebral Perfusion Pressure." *BMJ Open* 9 (9): e030727. https://doi.org/10.1136/bmjopen-2019-030727.

Brassard, Patrice, Lawrence Labrecque, Jonathan D. Smirl, Michael M. Tymko, Hannah G. Caldwell, Ryan L. Hoiland, et al. 2021. "Losing the Dogmatic View of Cerebral Autoregulation." *Physiological Reports* 9 (15): e14982. https://doi.org/10.14814/phy2.14982.

Budohoski, Karol P., Marek Czosnyka, Peter Smielewski, et al. 2013. "Cerebral Autoregulation after Subarachnoid Hemorrhage: Comparison of Three Methods." *Journal of Cerebral Blood Flow & Metabolism* 33 (3): 449–456. https://doi.org/10.1038/jcbfm.2012.189.

Capoglu, Seymanur, Jude P. Savarraj, Sunil A. Sheth, H. Alex Choi, and Luca Giancardo. 2019. "Representation Learning of 3D Brain Angiograms, an Application for Cerebral Vasospasm Prediction." In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC),* 3394–3398. https://doi.org/10.1109/EMBC.2019.8857815.

Carney, Nancy, Annette M Totten, Cindy O'Reilly, Jamie S Ullman, Gregory WJ Hawryluk, Michael J Bell, Susan L Bratton, Randall Chesnut, Odette A Harris, Niranjan Kissoon, et al. 2017. "Guidelines for the management of severe traumatic brain injury, fourth edition." *Neurosurgery* 80 (1): 6–15. https://doi.org/10.1227/NEU.0000000000001432.

Clarke, G., A. D. Mendelow, and P. Mitchell. 2005. "Predicting the Risk of Rupture of Intracranial Aneurysms Based on Anatomical Location." *Acta Neurochirurgica* 147:259–263. https://doi.org/10.1007/s00701-004-0473-3.

Colip, Charles G., Sean Wo, Daniel S. Hippe, et al. 2021. "Computed Tomography Angiography Findings Predictive of Post-Intervention Vasospasm in Patients with Aneurysmal Subarachnoid Hemorrhage." *British Journal of Radiology* 94 (1121): 20200893. https://doi.org/10.1259/bjr.20200893.

Czosnyka, M., and J. D. Pickard. 2004. "Monitoring and Interpretation of Intracranial Pressure." *Journal of Neurology, Neurosurgery & Psychiatry* 75 (6): 813–821.

Czosnyka, M., P. Smielewski, P. Kirkpatrick, S. Piechnik, R. Laing, and J. D. Pickard. 1998. "Continuous Monitoring of Cerebrovascular Pressure-Reactivity in Head Injury." *Acta Neurochirurgica Supplementum* 71:74–77.

Czosnyka, Marek, Ken Brady, Matthias Reinhard, et al. 2009. "Monitoring of Cerebrovascular Autoregulation: Facts, Myths, and Missing Links." *Neurocritical Care* 10 (3): 373–386. https://doi.org/10.1007/s12028-008-9175-7.

Czosnyka, Marek, Zofia Czosnyka, and Peter Smielewski. 2017. "Pressure Reactivity Index: Journey through the Past 20 Years." *Acta Neurochirurgica* 159 (11): 2063–2065. https://doi.org/10.1007/s00701-017-3310-1.

Dai, Honghao, Xiaodong Jia, Laura Pahren, Jay Lee, and Brandon Foreman. 2020. "Intracranial Pressure Monitoring Signals After Traumatic Brain Injury: A Narrative Overview and Conceptual Data Science Framework." *Frontiers in Neurology* 11:959. https://doi.org/10.3389/fneur.2020.00959.

Dengler, Nora Franziska, Vince Istvan Madai, Meike Unteroberdörster, et al. 2021. "Outcome Prediction in Aneurysmal Subarachnoid Hemorrhage: A Comparison of Machine Learning Methods and Established Clinico-Radiological Scores." *Neurosurgical Review* 44 (5): 2837–2846. https://doi.org/10.1007/s10143-020-01453-6.

Dumont, Travis M. 2016. "Prospective Assessment of a Symptomatic Cerebral Vasospasm Predictive Neural Network Model." *World Neurosurgery* 94:126–130. https://doi.org/10.1016/j.wneu.2016.06.110.

Dumont, Travis M., Anand I. Rughani, and Bruce I. Tranmer. 2011. "Prediction of Symptomatic Cerebral Vasospasm after Aneurysmal Subarachnoid Hemorrhage with an Artificial Neural Network." *World Neurosurgery* 75 (1): 57–63. https://doi.org/10.1016/j.wneu.2010.07.007.

English, Shane W., Anthony Delaney, Dean A. Fergusson, et al. 2025. "Liberal or Restrictive Transfusion Strategy in Aneurysmal Subarachnoid Hemorrhage." *New England Journal of Medicine* 392 (11): 1079–1088. https://doi.org/10.1056/NEJMoa2410962.

Etminan, Nima, Han-Sol Chang, Katharina Hackenberg, et al. 2019. "Worldwide Incidence of Aneurysmal Subarachnoid Hemorrhage According to Region, Time Period, Blood Pressure, and Smoking Prevalence." *JAMA Neurology* 76 (5): 588–597. https://doi.org/10.1001/jamaneurol.2019.0006.

Etminan, Nima, and Gabriel J. E. Rinkel. 2016. "Unruptured Intracranial Aneurysms: Development, Rupture and Preventive Management." *Nature Reviews Neurology* 12 (12): 699–713. https://doi.org/10.1038/nrneurol.2016.150.

Fantini, Sergio, Angelo Sassaroli, Konstantinos T. Tgavalekos, and Julie Kornbluth. 2016. "Cerebral blood flow and autoregulation: current measurement techniques and prospects for noninvasive optical methods." *Neurophotonics* 3 (3): 031411. https://doi.org/10.1117/1.NPh.3.3.031411.

Fujimoto, Yoshihisa, Kohei Yamada, and Takuo Yoshida. 2026. "Diagnostic Accuracy of Transcranial Doppler for Cerebral Vasospasm in Aneurysmal Subarachnoid Hemorrhage: A Systematic Review and Meta-Analysis." *Critical Care* 30 (1): 101. https://doi.org/10.1186/s13054-026-05849-6.

Gollwitzer, Maria, Vanessa Mazanec, Markus Steindl, et al. 2025. "Early Prediction of Cerebral Vasospasm After Aneurysmal Subarachnoid Hemorrhage Using a Machine Learning Model and Interactive Web Application." *Brain Sciences* 15 (11). https://doi.org/10.3390/brainsci15111187.

Grosset, D. G., J. Straiton, M. du Trevou, and R. Bullock. 1992. "Prediction of Symptomatic Vasospasm after Subarachnoid Hemorrhage by Rapidly Increasing Transcranial Doppler Velocity and Cerebral Blood Flow Changes." *Stroke* 23 (5): 674–679. https://doi.org/10.1161/01.STR.23.5.674.

Hao, Weili, Hong Hao, Chun-Feng Ren, Xiangling Wang, and Bulang Gao. 2022. "Associations Between Posterior Communicating Artery Aneurysms and Morphological Characteristics of Surrounding Arteries." *Frontiers in Neurology* 13:874466. https://doi.org/10.3389/fneur.2022.874466.

Hawryluk, Gregory W. J., Giuseppe Citerio, Peter Hutchinson, et al. 2022. "Intracranial Pressure: Current Perspectives on Physiology and Monitoring." *Intensive Care Medicine* 48 (10): 1471–1481. https://doi.org/10.1007/s00134-022-06786-y.

He, Wenzhuan, Chirag D. Gandhi, John Quinn, Reza Karimi, and Charles J. Prestigiacomo. 2011. "True Aneurysms of the Posterior Communicating Artery: A Systematic Review and Meta-Analysis of Individual Patient Data." *World Neurosurgery* 75 (1): 64–72. https://doi.org/10.1016/j.wneu.2010.09.012.

Healthline. 2012. "Subarachnoid Hemorrhage: Symptoms, Causes & Risk Factors." Illustration by Sophia Smith. https://www.healthline.com/health/subarachnoid-hemorrhage.

Hoh, Brian L., Nerissa U. Ko, Sepideh Amin-Hanjani, et al. 2023. "2023 Guideline for the Management of Patients With Aneurysmal Subarachnoid Hemorrhage." *Stroke* 54 (7): e314–e370. https://doi.org/10.1161/STR.0000000000000436.

Jarrett, C. L., K. L. Shields, R. M. Broxterman, J. R. Hydren, S. H. Park, J. R. Gifford, and R. S. Richardson. 2020. "Imaging transcranial Doppler ultrasound to measure middle cerebral artery blood flow: the importance of measuring vessel diameter." *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 319 (1): R33–R42. https://doi.org/10.1152/ajpregu.00025.2020.

Kassell, N. F., T. Sasaki, A. R. Colohan, and G. Nazar. 1985. "Cerebral Vasospasm Following Aneurysmal Subarachnoid Hemorrhage." *Stroke* 16 (4): 562–572. https://doi.org/10.1161/01.str.16.4.562.

Kazimierska, Agnieszka, Romain Manet, Alexandra Vallet, Eric Schmidt, Zofia Czosnyka, Marek Czosnyka, and Magdalena Kasprowicz. 2023. "Analysis of Intracranial Pressure Pulse Waveform in Studies on Cerebrospinal Compliance: A Narrative Review." *Physiological Measurement* 44 (10): 10TR01. https://doi.org/10.1088/1361-6579/ad0020.

Kim, Kwang Hyeon, Hae-Won Koo, Byung-Jou Lee, and Min-Jae Sohn. 2021. "Analysis of risk factors correlated with angiographic vasospasm in patients with aneurysmal subarachnoid hemorrhage using explainable predictive modeling." *Journal of Clinical Neuroscience* 91:334–342.

Kim, Kwang Hyeon, Byung-Jou Lee, and Hae-Won Koo. 2023. "Effect of Cilostazol on Delayed Cerebral Infarction in Aneurysmal Subarachnoid Hemorrhage Using Explainable Predictive Modeling." *Bioengineering* 10 (7). https://doi.org/10.3390/bioengineering10070797.

Kolias, Angelos G., Jon Sen, and Antonio Belli. 2009. "Pathogenesis of Cerebral Vasospasm Following Aneurysmal Subarachnoid Hemorrhage: Putative Mechanisms and Novel Approaches." *Journal of Neuroscience Research* 87 (1): 1–11. https://doi.org/10.1002/jnr.21823.

Lidington, Darcy, Hoyee Wan, and Steffen-Sebastian Bolz. 2021. "Cerebral Autoregulation in Subarachnoid Hemorrhage." *Frontiers in Neurology* 12. https://doi.org/10.3389/fneur.2021.688362.

Lindegaard, K. F., H. Nornes, S. J. Bakke, W. Sorteberg, and P. Nakstad. 1989. "Cerebral Vasospasm Diagnosis by Means of Angiography and Blood Velocity Measurements." *Acta Neurochirurgica* 100 (1–2): 12–24. https://doi.org/10.1007/BF01405268.

Lintas, Konstantinos, Stefan Rohde, Anna Mpoukouvala, et al. 2024. "Artificial Intelligence Prediction Model of Occurrence of Cerebral Vasospasms Based on Machine Learning." Published online, *Journal of Neurological Surgery Part A: Central European Neurosurgery,* https://www.thieme-connect.de/products/ejournals/abstract/10.1055/a-2402-6136.

Macdonald, R. Loch. 2014. "Delayed Neurological Deterioration after Subarachnoid Haemorrhage." *Nature Reviews Neurology* 10 (1): 44–58. https://doi.org/10.1038/nrneurol.2013.246.

Mathur, Rohan, Lin Cheng, Josiah Lim, et al. 2025. "Evolving Concepts in Intracranial Pressure Monitoring - from Traditional Monitoring to Precision Medicine." *Neurotherapeutics* 22 (1): e00507.

Megjhani, Murad, Miriam Weiss, Jenna Ford, Kalijah Terilli, Nick Kastenholz, Daniel Nametz, et al. 2023. "Optimal Cerebral Perfusion Pressure and Brain Tissue Oxygen in Aneurysmal Subarachnoid Hemorrhage." *Stroke* 54 (1): 189–197. https://doi.org/10.1161/STROKEAHA.122.040339.

Mohan, M., A. I. Islim, F. T. Rasul, O. Rominiyi, R. M. deSouza, M. T. C. Poon, et al. 2019. "Subarachnoid Haemorrhage with Negative Initial Neurovascular Imaging: A Systematic Review and Meta-Analysis." *Acta Neurochirurgica* 161 (10): 2013–2026. https://doi.org/10.1007/s00701-019-04025-w.

Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, Cynthia D. Mulrow, Larissa Shamseer, et al. 2021. "The PRISMA 2020 Statement: An Updated Guideline for Reporting Systematic Reviews." The definitive PRISMA 2020 guideline, *BMJ* 372:n71. https://doi.org/10.1136/bmj.n71.

Paiva, José Artur, Antero Fernandes, Cristina Granja, et al. 2016. *Rede de Referenciação de Medicina Intensiva.* Technical Report. Serviço Nacional de Saúde (SNS).

Palermo, Matteo, Sonia D'Arrigo, Alessandro Olivi, and Carmelo Lucio Sturiale. 2025. "Machine learning models for predicting vasospasm following ruptured intracranial aneurysms: a systematic review and meta-analysis." *Acta Neurochirurgica* 167 (1): 314. https://doi.org/10.1007/s00701-025-06725-y.

Palmisciano, Paolo, Samer S. Hoz, Mark D. Johnson, et al. 2023. "External Validation of an Extreme Gradient Boosting Model for Prediction of Delayed Cerebral Ischemia After Aneurysmal Subarachnoid Hemorrhage." *World Neurosurgery* 175:e108–e114. https://doi.org/10.1016/j.wneu.2023.03.036.

Penedo, Jorge Manuel Virtudes dos Santos, António Augusto Batista Ribeiro, Helena do Amparo Romão de Castro Lopes, Jorge Manuel Perição Costa Pimentel, José Afonso Gonçalves Pereira da Silva Pedrosa, Rui Alberto Marques Vasconcelos e Sá, and Rui Paulo Jinó Moreno. 2016. *Avaliação da Situação Nacional das Unidades de Cuidados Intensivos: Relatório Final.* Technical Report. 459 p. Serviço Nacional de Saúde (SNS), May. https://www.sns.gov.pt/wp-content/uploads/2016/05/Avalia%C3%A7%C3%A3o-nacional-dasitua%C3%A7%C3%A3o-das-unidades-de-cuidados-intensivos.pdf.

Qureshi, Adnan I., Grace Y. Sung, Alexander Y. Razumovsky, Kevin Lane, Robert N. Straw, and John A. Ulatowski. 2000. "Early Identification of Patients at Risk for Symptomatic Vasospasm after Aneurysmal Subarachnoid Hemorrhage." *Critical Care Medicine* 28 (4): 984–990. https://doi.org/10.1097/00003246-200004000-00012.

Rabinstein, Alejandro A., Jonathan A. Friedman, Stephen D. Weigand, et al. 2004. "Predictors of Cerebral Infarction in Aneurysmal Subarachnoid Hemorrhage." *Stroke* 35 (8): 1862–1866. https://doi.org/10.1161/01.STR.0000133132.76983.8e.

Rhcastilhos. 2007. *Circle of Willis.* Wikimedia Commons. Schematic representation of the circle of Willis, derived from Gray519.png; public domain. https://commons.wikimedia.org/wiki/File:Circle_of_Willis_en.svg.

Roederer, Alexander, John H. Holmes, Michelle J. Smith, Insup Lee, and Soojin Park. 2014. "Prediction of Significant Vasospasm in Aneurysmal Subarachnoid Hemorrhage Using Automated Data." *Neurocritical Care* 21 (3): 444–450. https://doi.org/10.1007/s12028-014-9976-9.

Rumalla, Karthik, Michael Lin, Luying Ding, Matthew Gaddis, Steven L. Giannotta, Frank J. Attenello, and William J. Mack. 2021. "Risk Factors for Cerebral Vasospasm in Aneurysmal Subarachnoid Hemorrhage: A Population-Based Study of 8346 Patients." *World Neurosurgery* 145:e233–e241. https://doi.org/10.1016/j.wneu.2020.10.008.

Sadri, Youkabed, Neman Abdoli, Ke Zhang, et al. 2024. "Developing a CT-image-based Clinical Marker to Predict Vasospasm for Aneurysmal Subarachnoid Hemorrhage Patients." In *Biophotonics and Immune Responses XIX,* 12843:13–18. SPIE. https://doi.org/10.1117/12.3005015.

Al-Salihi, Mohammed Maan, Ram Saha, Ahmed Abd Elazim, Amer Helal, Maryam Sabah Al-Jebur, Yezan Al-Salihi, and Ali Ayyad. 2024. "The Effectiveness and Safety of Clazosentan in Treating Aneurysmal Subarachnoid Hemorrhage: A Systematic Review and Meta-Analysis." *Journal of Clinical Neuroscience* 126:173–181. https://doi.org/10.1016/j.jocn.2024.06.019.

Savarraj, Jude P., Georgene W. Hergenroeder, Liang Zhu, et al. 2021. "Machine learning to predict delayed cerebral ischemia and outcomes in subarachnoid hemorrhage." *Neurology* 96:e553–e562. https://doi.org/10.1212/WNL.0000000000011211.

Scherer, Moritz, Jin-On Jung, Jonas Cordes, et al. 2020. "Association of Cerebrospinal Fluid Volume with Cerebral Vasospasm After Aneurysmal Subarachnoid Hemorrhage: A Retrospective Volumetric Analysis." *Neurocritical Care* 33 (1): 152–164. https://doi.org/10.1007/s12028-019-00878-2.

Schmidt, Bernhard, Matthias Reinhard, Vesna Lezaic, et al. 2016. "Autoregulation Monitoring and Outcome Prediction in Neurocritical Care Patients: Does One Index Fit All?" *Journal of Clinical Monitoring and Computing* 30 (3): 367–375.

Sehba, Fatima A., Jack Hou, Ryszard M. Pluta, and John H. Zhang. 2012. "The Importance of Early Brain Injury after Subarachnoid Hemorrhage." *Progress in Neurobiology* 97 (1): 14–37. https://doi.org/10.1016/j.pneurobio.2012.02.003.

Sen, Rajeev D., Margaret C. McGrath, Varadaraya S. Shenoy, et al. 2025. "A Dynamic Machine Learning Model to Predict Angiographic Vasospasm After Aneurysmal Subarachnoid Hemorrhage." Published online, *Neurosurgery,* https://doi.org/10.1227/neu.0000000000003639.

Skoch, Jesse, Rana Tahir, Todd Abruzzo, Joseph M. Taylor, Mario Zuccarello, and Sudhakar Vadivelu. 2017. "Predicting symptomatic cerebral vasospasm after aneurysmal subarachnoid hemorrhage with an artificial neural network in a pediatric population." *Child's Nervous System* 33 (12): 2153–2157. https://doi.org/10.1007/s00381-017-3573-0.

Smielewski, P., E. Beqiri, C. Mataczynski, M. Placek, A. Kazimierska, P.J. Hutchinson, M. Czosnyka, and M. Kasprowicz. 2024. "Advanced Neuromonitoring Powered by ICM+ and Its Place in the Brand New AI World, Reflections at the 20th Anniversary Boundary." *Brain & Spine* 4:102835. https://doi.org/10.1016/j.bas.2024.102835. https://pmc.ncbi.nlm.nih.gov/articles/PMC11278591/.

Sobey, Christopher G., and Frank M. Faraci. 1998. "Subarachnoid Haemorrhage: What Happens to the Cerebral Arteries?" *Clinical and Experimental Pharmacology and Physiology* 25 (11): 867–876. https://doi.org/10.1111/j.1440-1681.1998.tb02337.x.

Sorrentino, E., J. Diedler, M. Kasprowicz, K. P. Budohoski, C. Haubrich, P. Smielewski, et al. 2012. "Critical Thresholds for Cerebrovascular Reactivity after Traumatic Brain Injury." *Neurocritical Care* 16 (2): 258–266. https://doi.org/10.1007/s12028-011-9630-8.

Svedung Wettervik, Teodor, Henrik Engquist, Timothy Howells, et al. 2023. "Higher intracranial pressure variability is associated with lower cerebrovascular resistance in aneurysmal subarachnoid hemorrhage." *Journal of Clinical Monitoring and Computing* 37:319–326. https://doi.org/10.1007/s10877-022-00894-2.

Svedung Wettervik, Teodor, Timothy Howells, Anders Lewén, Elisabeth Ronne-Engström, and Per Enblad. 2021. "Temporal Dynamics of ICP, CPP, PRx, and CPPopt in High-Grade Aneurysmal Subarachnoid Hemorrhage and the Relation to Clinical Outcome." *Neurocritical Care* 34 (2): 390–402. https://doi.org/10.1007/s12028-020-01162-4.

Toledo, Paula de, Pablo M. Rios, Agapito Ledezma, et al. 2009. "Predicting the Outcome of Patients with Subarachnoid Hemorrhage Using Machine Learning Techniques." *IEEE Transactions on Information Technology in Biomedicine* 13 (5): 794–801. https://doi.org/10.1109/TITB.2009.2020434.

Urbanos, Gemma, Ana M. Castaño-León, Mónica Maldonado-Luna, et al. 2025. "Comprehensive Predictive Modeling in Subarachnoid Hemorrhage: Integrating Radiomics and Clinical Variables." *Neurosurgical Review* 48 (1): 528. https://doi.org/10.1007/s10143-025-03679-8.

Vergouwen, Mervyn D. I., Marinus Vermeulen, Jan van Gijn, et al. 2010. "Definition of Delayed Cerebral Ischemia after Aneurysmal Subarachnoid Hemorrhage as an Outcome Event in Clinical Trials and Observational Studies." *Stroke* 41 (10): 2391–2395. https://doi.org/10.1161/STROKEAHA.110.589275.

Wei, Miaomiao, Solventa Krakauskaite, Sreya Subramanian, and Fabien Scalzo. 2024. "Peak detection in intracranial pressure signal waveforms: a comparative study." *BioMedical Engineering OnLine* 23 (1): 61. https://doi.org/10.1186/s12938-024-01245-9.

Wolff, Robert F., Karel G. M. Moons, Richard D. Riley, Penny F. Whiting, Marie Westwood, Gary S. Collins, Johannes B. Reitsma, Jos Kleijnen, and Susan Mallett. 2019. "PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies." The definitive PROBAST guideline, *Annals of Internal Medicine* 170 (1): 51–58. https://doi.org/10.7326/M18-1376.

Yang, Chenghao, Fei Luo, Xi Chen, Yunliang Deng, and Meirui Lan. 2025. "Association between Multiple Serum Biomarkers and Cerebral Vasospasm in Patients with Aneurysmal Subarachnoid Hemorrhage." *Frontiers in Neurology* 16. https://doi.org/10.3389/fneur.2025.1592390.

Zarrin, David A., Abhinav Suri, Karen McCarthy, Bilwaj Gaonkar, et al. 2024. "Machine Learning Predicts Cerebral Vasospasm in Patients with Subarachnoid Haemorrhage." *eBioMedicine* 105:105206. https://doi.org/10.1016/j.ebiom.2024.105206.

Zeiler, Frederick A., Joseph Donnelly, David K. Menon, Peter Smielewski, Peter J. A. Hutchinson, and Marek Czosnyka. 2018. "A Description of a New Continuous Physiological Index in Traumatic Brain Injury Using the Correlation between Pulse Amplitude of Intracranial Pressure and Cerebral Perfusion Pressure." *Journal of Neurotrauma* 35 (7): 963–974. https://doi.org/10.1089/neu.2017.5241.

Appendix A

Characteristics of Studies Included in Literature Review

Table A.1: Comprehensive summary of included studies. Outcome labels are descriptive shorthand for study-specific endpoints and do not imply uniform diagnostic criteria across studies. ACV is used only for imaging-defined arterial narrowing; TCD-defined, symptomatic, treatment-defined and DCI endpoints are retained separately, and missing operational details are not inferred. **Abbreviations:** ACV, Angiographic Vasospasm; DCI, Delayed Cerebral Ischaemia; SCV, Symptomatic Vasospasm; CVRV, Cerebral Vasospasm Requiring Verapamil; IA, Intra-Arterial; Tx, Treatment; ACC, Accuracy; TPR, True Positive Rate (Sensitivity); TNR, True Negative Rate (Specificity); PPV, Positive Predictive Value; NPV, Negative Predictive Value; F1, F1-score; MSE, Mean Squared Error; AUC, Area Under the Curve. **Models:** MLP, Multilayer Perceptron; RF, Random Forest; LR, Logistic Regression; SVC, Support Vector Classifier; GBM, Gradient Boosting Machine; BM, Baseline Model; PBD, Post-Bleed Day.

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Longitudinal Monitoring Approaches						
Sen et al. (2025)	Retro. / Single-C $N = 424$	ACV	Daily TCD (max. velocity, Lindegaard ratio), hourly ICP, MAP, CPP and EVD output, serum sodium, and additional clinical/radiographic information.	**Admission LR (BM)** *vs LSTM*	**Internal:** Patient-separated 80/20 holdout of moving-window samples; TCD feature-ablation analyses.	**ACC:** 0.736 *(vs 0.786)* **PPV:** 0.483 *(vs 0.683)* **TPR:** 0.683 *(vs 0.277)* **F1:** 0.566 *(vs 0.394)*

Continued on next page

Table A.1 continued from previous page

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Roederer et al. (2014)	Retro. / Single-C $N = 81$	ACV (IA-treated)	Selected daily and cumulative summaries of routinely recorded monitoring and laboratory data (PBD 0–2).	**Naive Bayes** *vs Logistic Reg.*	**Internal:** 10-fold cross-validation with feature-subset ablation. Fold-wise repetition of outcome-based feature selection and population-mean imputation was not clearly reported.	**AUC:** 0.708 *(vs 0.544)* **PPV:** 0.677 *(vs 0.465)* **TPR:** 0.600 *(vs 0.571)* **MSE:** 0.26 *(vs 0.433)*
Zarrin et al. (2024)	Retro. / Multi-C $N = 3404$ (UCLA: 1750 VUMC: 1654)	CVRV	172 routinely collected ICU variables represented through cumulative summaries. Institutional and conservative predictor sets were evaluated at UCLA; VUMC used the conservative set.	**LightGBM** *vs Logistic Reg.*	**UCLA:** Stratified 5-fold cross-validation with SMOTE. **VUMC:** Authors described the analysis as external validation, but a separate model was locally trained and tested using the conservative prospective-trinary workflow.	**Binary AUC (4 h–10 d):** Inst. 0.88–0.71; Cons. 0.87–0.68 **Trinary AUC (UCLA):** 0.68–0.88 / 0.73–0.83 / 0.50–0.88 *(No CVRV / within 3 d / after 3 d)* **VUMC:** mean ΔAUC ≈ -0.01 vs UCLA **Rule-out Sens:** 1.00
Predominantly Fixed Patient-Level Representations						

Continued on next page

Table A.1 continued from previous page

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Sadri et al. (2024)	Retro. / Single-C $N = 52$	Vasospasm (requiring IA Tx)	Admission CT: 1032 radiomic + 768 ViT features (top 30 selected).	**SVM**	**Internal:** 4-fold cross-validation.	**AUC:** 0.86 **ACC:** 0.77 **PPV:** 0.78 **NPV:** 0.76
Capoglu et al. (2019)	Retro. / Single-C $N = 20$	ACV	3D DSA vessel geometry (covariance of local vasculature windows).	**Logistic Reg.** *vs Clinical LR*	**Internal:** Leave-One-Subject-Out (LOSO) cross-validation.	**AUC:** 0.93 *(vs 0.65)* **ACC:** 0.85
Colip et al. (2021)	Retro. / Single-C $N = 195$	ACV ($> 30\%$ narrowing)	CTA morphology (arterial undulation, calcification, narrowing), mFisher.	**LASSO** *vs mFisher Only*	**Internal:** Bootstrap optimism-adjustment; cross-validation used for LASSO tuning.	**AUC:** 0.74 *(vs 0.61)* **OR:** 3.4 for Irreg.
Gollwitzer et al. (2025)	Retro. / Single-C $N = 345$	Moderate/ Severe ACV	EVD Req., H&H, Fisher, aneurysm presence, age.	**Logistic Reg.** *vs Linear SVC*	**Internal:** Repeated nested cross-validation (10×10), with SMOTE and inverse class weights.	**AUC:** 0.723 *(vs 0.720)* **TPR:** 0.660 *(vs 0.642)* **TNR:** 0.652 *(vs 0.661)* **PPV:** 0.700 *(vs 0.699)* **NPV:** 0.620 *(vs 0.609)* **F1:** 0.675 *(vs 0.665)*
Kim et al. (2021)	Retro. / Single-C $N = 343$	ACV ($\geq 50\%$ narrowing)	Aneurysm size/location, age, GCS, Fisher, HTN, antiplatelets.	**RF Regressor** (+ SHAP)	**Internal:** Holdout split (80/20).	**AUC:** 0.88 **ACC:** 0.85 **TPR:** 0.78 **TNR:** 0.77

Continued on next page

Table A.1 continued from previous page

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Scherer et al. (2020)	Retro. / Single-C $N = 125$	ACV ($> 34\%$ narrowing)	CT volumetrics: SAH volume, CSF volume, SAH/CSF ratio.	**Univariate Binary LR** *(Predictor comparison)*	**Internal:** Retrospective statistical analysis; Boruta feature selection.	**AUC (by Predictor):** **SAH/CSF Ratio:** 0.74 **SAH Vol:** 0.73 **CSF Vol:** 0.56 **Fisher:** 0.50
Urbanos et al. (2025)	Retro. / Single-C $N = 444$	Vasospasm (study-defined)	Clinical (20 vars) + radiomics (brain tissue: GM/WM).	**Ensemble** (ExtraRF, RF, XGB) *vs LR (ictWFNS)*	**Internal:** 5-fold stratified cross-validation ($N = 328$) followed by an independent test set ($N = 41$).	**Test AUC:** 0.66 *(vs 0.596)* **Test ACC:** 0.66 **Test TPR:** 0.66 **Test TNR:** 0.62
Yang et al. (2025)	Retro. / Single-C $N = 175$	TCD-defined vasospasm	Serum biomarkers (HIF-1α, VEGF, ET-1) + GCS.	**Random Forest** *vs Logistic Reg.*	**Internal:** Holdout (70/30) + 5-fold cross-validation for tuning; MICE imputation.	**AUC:** 0.966 *(vs 0.955)* **TPR:** 0.917 *(vs 0.905)* **TNR:** 0.912 *(vs 0.912)*
Dumont et al. (2011)	Single-C Retro. develop.: $N = 91$ Prosp. eval.: $N = 22$	SCV	Demographics, clinical (GCS), TCD elevation, clot thickness, aneurysm location, EVD.	**ANN** *vs Adams & Qureshi MLR*	**Internal:** Retrospective development ($N = 91$) + prospective temporal holdout at the same centre ($N = 22$).	**AUC:** 0.960 *(vs 0.933 & 0.897)*

Continued on next page

Table A.1 continued from previous page

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Dumont (2016)	Prosp. / Single-C $N = 25$	SCV	Demographics, clinical (GCS), TCD elevation, clot thickness, aneurysm location, EVD.	**ANN**	**External:** Prospective geographical validation of the previously developed Dumont et al. (2011) ANN at a different hospital; no local retraining reported.	**TPR:** 1.00 **TNR:** 0.84 **PPV:** 0.67 **NPV:** 1.00
Lintas et al. (2024)	Retro. / Single-C $N = 87$	SCV	H&H, Fisher Grade, BNI Scale, intervention length.	**SVM**	**Internal:** Holdout (65/22) and cross-validation.	**ACC:** 0.61–0.86 (No AUC reported)
Alexopoulos et al. (2022)	Retro. / Single-C $N = 213$	DCI (clinical & infarction)	Clinical grades (H&H, Fisher, mFisher), EVD, aneurysm/Tx type.	**XGBoost** *vs Logistic Reg.*	**Internal:** Holdout (70/30) + repeated 10-fold cross-validation; ROSE oversampling applied.	**ACC:** 0.843 *(vs 0.648)* **TPR:** 0.901 *(vs 0.663)* **TNR:** 0.831 *(vs 0.643)* **PPV:** 0.527 *(vs 0.371)* **NPV:** 0.977 *(vs 0.858)* **F1:** 0.684 *(vs 0.475)*
Azzam et al. (2024)	Retro. / Single-C $N = 325$	DCI (clinical)	Clinical vasospasm, RDW, MCV, BUN, nimodipine, VP shunt, comorbidities.	**MLP Neural Net** *vs Logistic Reg. [RF had highest reported AUC: 0.85]*	**Internal:** Holdout (80/20) + 10-fold cross-validation; SMOTE and ENN augmentation.	**ACC:** 0.93 *(vs 0.90)* **AUC:** 0.83 *(vs 0.82)* **TPR:** 0.87 *(vs 0.82)* **PPV:** 0.84 *(vs 0.83)* **F1:** 0.85 *(vs 0.82)*

Continued on next page

Table A.1 continued from previous page

Study	Design / N	Target	Input Predictors	Model / Comparator	Validation Strategy	Performance
Kim et al. (2023)	Retro. / Single-C $N = 321$	DCI (clinical OR infarction)	Cilostazol use, aneurysm size (> 10 mm), age, sex.	**Logistic Reg.** *vs XGBoost*	**Internal:** 5-fold cross-validation for generalisation assessment.	**AUC:** 0.95 *(vs 0.94)* **ACC:** 0.92 *(vs 0.91)* **TPR:** 0.80 *(vs 0.70)* **TNR:** 0.95 *(vs 0.95)*
Savarraj et al. (2021)	Retro. / Single-C $N = 399$	DCI	Sodium, WBC, neutrophils, Hgb, platelets, creatinine (Days 0–3).	**ANN** *vs Logistic Reg.*	**Internal:** Stratified holdout (80/20) + stratified 10-fold cross-validation.	**AUC:** 0.75 *(vs 0.56)* **ACC:** 0.80 *(vs 0.55)* **TPR:** 0.82 *(vs 0.79)* **TNR:** 0.72 *(vs 0.25)*
Palmisciano et al. (2023)	Retro. / Single-C $N = 267$	DCI (clinical & infarction)	Clinical grades (H&H, Fisher, mFisher), EVD, aneurysm/Tx type.	**XGBoost**	**External:** Validation of the Alexopoulos et al. (2022) model in an independent retrospective cohort.	**ACC:** 0.648 **TPR:** 0.328 **TNR:** 0.737 **PPV:** 0.257 **NPV:** 0.798 **F1:** 0.288

Sen et al. (2025): Performance values follow the detailed Results and Figure 2 of the source article; the abstract reports inconsistent values and comparator attribution.

Zarrin et al. (2024): The authors describe the VUMC analysis as external validation, but a separate local model was trained and tested there; trinary metrics correspond to No CVRV, CVRV within three days, and CVRV after three days, respectively.

Urbanos et al. (2025): The Methods describe clinical and radiological criteria for vasospasm, whereas the Results distinguish clinical and radiological cases; the exact relationship between these descriptions in outcome labelling is therefore treated as unclear.

Kim et al. (2021): The source reports accuracy of 0.85, sensitivity of 0.78 and specificity of 0.77. Sensitivity and specificity were evaluated at an ROC-derived operating point, while the operating threshold corresponding to the reported accuracy was not specified; these metrics should therefore not be assumed to derive from the same confusion matrix.

Source limitations: Sadri et al. (2024) and Lintas et al. (2024) were abstract-based inclusions. The detailed operational DCI criteria used by Savarraj et al. (2021) were not available in the reviewed material. Missing criteria were not inferred.

Appendix B

Predictor and Feature Definitions

B.1 Physiological Signals and Derived Features

Table B.1: Principal physiological signals and derived indices relevant to the neurocritical monitoring framework considered in this dissertation.

Variables	Description	Units
Intracranial Pressure (ICP)	Pressure within the cranial vault, reflecting the interaction between brain tissue, cerebral blood, and cerebrospinal fluid volumes.	mmHg
Mean Cerebral Perfusion Pressure (mCPP)	Mean pressure gradient available to drive cerebral blood flow, conventionally derived from mean arterial pressure and intracranial pressure.	mmHg
Arterial Blood Pressure (ABP)	Systemic arterial pressure measured from the arterial circulation, providing the principal haemodynamic driving input for cerebral perfusion.	mmHg
ICP Pulse Amplitude (AMP)	Amplitude of the cardiac component of the ICP waveform, reflecting the pulsatile intracranial pressure response to changes in cerebral arterial blood volume during each cardiac cycle.	mmHg
Heart Rate (HR)	Frequency of cardiac cycles, providing a measure of systemic cardiac rate.	beats/min
RAP Index	Moving Pearson correlation between ICP pulse amplitude (AMP) and mean ICP, used as a continuous indicator of intracranial compensatory reserve.	Dimensionless

Continued on next page

Table B.1 continued from previous page

Variables	Brief Description	Units
RAC Index	Moving Pearson correlation between ICP pulse amplitude (AMP) and cerebral perfusion pressure (CPP), investigated as an additional index of cerebrovascular reactivity.	Dimensionless
Pressure Reactivity Index (PRx)	Moving correlation between slow fluctuations in arterial pressure and ICP, used to assess cerebrovascular pressure reactivity and autoregulatory function.	Dimensionless
Maximum Intracranial Pressure (ICPmax)	Maximum ICP value recorded within the corresponding processed monitoring interval.	mmHg
Minimum Intracranial Pressure (ICPmin)	Minimum ICP value recorded within the corresponding processed monitoring interval.	mmHg
End-Tidal Carbon Dioxide ($EtCO_2$)	Partial pressure of carbon dioxide measured at the end of expiration, used as a non-invasive surrogate of arterial CO_2 and an indicator of ventilation that may influence cerebral vascular tone and blood flow.	mmHg

B.2 Static Clinical Predictors

Table B.2: Admission and early clinical predictors included in the final static analytical representation.

Variables	**Description**	**Units / Coding**
Age	Patient age at the time of the index clinical event.	years
Admission Haemoglobin	Haemoglobin concentration recorded at hospital admission, representing baseline blood oxygen-carrying capacity.	g/dL
Glasgow Coma Scale (GCS) at Hospital Admission	Clinical assessment of level of consciousness at hospital admission using the Glasgow Coma Scale.	Ordinal score
Fisher Scale	Radiological grading of subarachnoid blood burden on admission imaging.	Ordinal score
Hunt–Hess Scale	Clinical grading of subarachnoid haemorrhage severity based on neurological condition at presentation.	Ordinal score
Pre-event Modified Rankin Scale	Functional-status score describing the patient's degree of disability before the index haemorrhagic event.	Ordinal score
Sex	Recorded biological sex of the patient.	Binary
Family History of Aneurysm	Indicator of a documented family history of intracranial aneurysm.	Binary
Active Smoking	Indicator of active smoking status recorded in the clinical dataset.	Binary
Hypertension	Indicator of a recorded history of arterial hypertension.	Binary
Ischaemia on Admission CT	Indicator of ischaemic changes identified on the admission computed tomography examination.	Binary
Anterior-circulation Location	Binary indicator derived from the original aneurysm-location field. It was set for aneurysms classified in the ICA, MCA, ACA, ACoA, or PCoA territories.	Binary
Posterior-circulation Location	Binary indicator derived from the original aneurysm-location field. It was set for aneurysms classified in the Basilar Artery, PCA, or PICA territories.	Binary

Appendix C

Expanded Results

C.1 Phase 1: Static Clinical Prediction

Table C.1: Detailed Phase 1 static-model performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. AP is average precision.

Model	AUROC	AUROC percentiles	AP	Brier
XGBoost	**0.720 ± 0.015**	0.690–0.745	0.659 ± 0.025	**0.214 ± 0.005**
LightGBM	0.713 ± 0.016	0.678–0.739	0.662 ± 0.024	0.216 ± 0.005
Random forest	0.708 ± 0.015	0.683–0.732	**0.665 ± 0.024**	0.217 ± 0.005
Logistic regression	0.703 ± 0.013	0.681–0.725	0.662 ± 0.022	0.218 ± 0.004
Linear SVC	0.702 ± 0.015	0.677–0.730	0.659 ± 0.026	0.219 ± 0.005
Gaussian naive Bayes	0.660 ± 0.019	0.617–0.691	0.611 ± 0.028	0.239 ± 0.003
Multilayer perceptron	0.642 ± 0.029	0.585–0.691	0.586 ± 0.032	0.236 ± 0.006

C.2 Phase 2: Aggregated Physiological Prediction

Table C.2: Detailed Phase 2 performance by representation and model family. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. Extended representations were secondary analyses.

Representation	Model	AUROC	AUROC percentiles	AP	Brier
Static only	Logistic regression	**0.786 ± 0.022**	0.748–0.824	0.767 ± 0.027	0.196 ± 0.010
	Random forest	0.727 ± 0.041	0.658–0.799	0.765 ± 0.054	0.221 ± 0.022
	XGBoost	0.701 ± 0.043	0.608–0.770	0.734 ± 0.045	0.227 ± 0.015
Minimal physiology	Logistic regression	0.617 ± 0.079	0.430–0.725	0.597 ± 0.055	0.254 ± 0.018
	Random forest	**0.656 ± 0.070**	0.531–0.774	0.679 ± 0.075	0.240 ± 0.027
	XGBoost	0.557 ± 0.075	0.404–0.677	0.587 ± 0.060	0.258 ± 0.019
Static + minimal	Logistic regression	0.703 ± 0.054	0.572–0.781	0.643 ± 0.040	0.238 ± 0.018
	Random forest	**0.711 ± 0.057**	0.601–0.817	0.733 ± 0.060	0.223 ± 0.023
	XGBoost	0.667 ± 0.061	0.543–0.754	0.693 ± 0.070	0.234 ± 0.019
Extended physiology	Logistic regression	0.475 ± 0.072	0.332–0.606	0.505 ± 0.047	0.267 ± 0.017
	Random forest	0.583 ± 0.070	0.407–0.678	0.612 ± 0.065	0.268 ± 0.033
	XGBoost	0.454 ± 0.075	0.307–0.590	0.516 ± 0.057	0.273 ± 0.021
Static + extended	Logistic regression	0.558 ± 0.064	0.421–0.669	0.563 ± 0.049	0.258 ± 0.017
	Random forest	0.616 ± 0.060	0.488–0.715	0.644 ± 0.064	0.257 ± 0.023
	XGBoost	0.540 ± 0.076	0.388–0.661	0.580 ± 0.070	0.261 ± 0.020

Table C.3: Detailed matched Phase 2 differences after adding minimal physiological summaries to the static representation. Differences are combined minus static; positive AUROC and AP differences and negative Brier differences favour the combined representation.

Model	ΔAUROC, mean ± SD	AUROC percentiles	Combined higher, %	ΔAP	ΔBrier
Logistic regression	−0.082 ± 0.058	−0.215 to −0.004	2	−0.124	+0.042
Random forest	−0.016 ± 0.061	−0.132 to 0.071	48	−0.031	+0.002
XGBoost	−0.033 ± 0.068	−0.157 to 0.070	36	−0.041	+0.007

C.3 Phase 3: Temporally Resolved Physiological Prediction

Table C.4: Detailed Phase 3 temporal-model performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. ROCKET–Ridge AUROC and AP are based on its raw discrimination score, whereas Brier score uses calibrated probabilities.

Analysis	**Role**	**AUROC**	**AUROC percentiles**	**AP**	**Brier**
Value-only CNN	Primary	0.617 ± 0.051	0.510–0.694	0.579 ± 0.040	0.268 ± 0.025
Value-plus-mask CNN	Sensitivity	0.654 ± 0.062	0.525–0.767	0.621 ± 0.049	0.259 ± 0.030
ROCKET–Ridge	Comparator	0.553 ± 0.072	0.423–0.660	0.614 ± 0.058	0.265 ± 0.017

Table C.5: Detailed paired Phase 3 comparisons. Differences are analysis A minus analysis B. Positive AUROC and AP differences and negative Brier differences favour analysis A.

Comparison: A minus B	**ΔAUROC**	**AUROC percentiles**	**A higher, %**	**ΔAP / ΔBrier**
Value-plus-mask CNN minus value-only CNN	$+0.037$	-0.064 to 0.148	78	$+0.042/-0.010$
ROCKET–Ridge minus value-only CNN	-0.064	-0.224 to 0.062	20	$+0.036/-0.003$
Value-only CNN minus Phase 2 minimal-physiology RF	-0.039	-0.191 to 0.082	32	$-0.100/+0.028$
Value-plus-mask CNN minus Phase 2 minimal-physiology RF	-0.002	-0.131 to 0.159	38	$-0.058/+0.018$

C.4 Phase 4: Static–Temporal Late Fusion

Table C.6: Detailed Phase 4 static, temporal and late-fusion performance. Values are mean ± SD across 50 repetitions; the percentile column gives the empirical 2.5th–97.5th percentile range of repetition-level AUROC. The temporal component rows report standalone Phase 3 performance from the outer-test probabilities reused in Phase 4.

Analysis	**Role**	**AUROC**	**AUROC percentiles**	**AP**	**Brier**
Phase 4 fixed-*C* static LR	Comparator	0.799 ± 0.019	0.767–0.828	0.777 ± 0.027	0.182 ± 0.011
Phase 3 value-only CNN	Component	0.617 ± 0.051	0.510–0.694	0.579 ± 0.040	0.268 ± 0.025
Static LR + value-only CNN	Primary fusion	0.760 ± 0.034	0.671–0.814	0.732 ± 0.040	0.207 ± 0.014
Phase 3 value-plus-mask CNN	Component	0.654 ± 0.062	0.525–0.767	0.621 ± 0.049	0.259 ± 0.030
Static LR + value-plus-mask CNN	Sensitivity fusion	0.771 ± 0.028	0.713–0.810	0.750 ± 0.035	0.201 ± 0.012

Table C.7: Detailed paired Phase 4 comparisons across 50 matched repetitions. Differences are mean within-repetition differences for analysis A minus analysis B. Positive AUROC and AP differences and negative Brier differences favour analysis A. Percentile ranges describe the empirical 2.5th–97.5th percentiles of paired AUROC differences; the percentage column reports repetitions with strictly higher AUROC for analysis A.

Comparison: A minus B	**ΔAUROC**	**AUROC percentiles**	**A higher, %**	**ΔAP**	**ΔBrier**
Primary fusion minus Phase 4 static LR	−0.039	−0.114 to 0.008	10	−0.045	+0.025
Mask-sensitivity fusion minus Phase 4 static LR	−0.028	−0.099 to 0.014	12	−0.027	+0.019
Mask-sensitivity fusion minus primary fusion	+0.011	−0.049 to 0.071	66	+0.018	−0.005

C.5 Secondary Operating-Point and Calibration Results

Table C.8: Secondary screening operating-point and calibration summaries for the analysis-defining models. Values are mean ± SD across 50 repetitions. **P1 XGB** denotes Phase 1 XGBoost; **P2 Static LR**, Phase 2 static logistic regression; **P2 Min RF**, Phase 2 minimal-physiology random forest; **P2 Static+Min RF**, Phase 2 static-plus-minimal-physiology random forest; **P3 CNN**, the Phase 3 value-only CNN; **P3 Mask CNN**, the Phase 3 value-plus-mask CNN; **P3 ROCKET**, Phase 3 ROCKET–Ridge; **P4 Static LR**, the Phase 4 fixed-C static logistic regression; **P4 Fusion**, the primary static-LR plus value-only-CNN fusion; and **P4 Mask Fusion**, the static-LR plus value-plus-mask-CNN sensitivity fusion. **PPV** and **NPV** denote positive and negative predictive value.

Metric	P1 XGB	P2 Static LR	P2 Min RF	P2 Static+Min RF	P3 CNN	P3 Mask CNN	P3 ROCKET	P4 Static LR	P4 Fusion	P4 Mask Fusion
Threshold	0.340 ± 0.008	0.389 ± 0.027	0.365 ± 0.026	0.365 ± 0.025	0.440 ± 0.029	0.446 ± 0.028	0.435 ± 0.015	0.330 ± 0.035	0.363 ± 0.043	0.361 ± 0.038
Sensitivity	0.845 ± 0.026	0.857 ± 0.051	0.830 ± 0.083	0.843 ± 0.071	0.745 ± 0.126	0.731 ± 0.101	0.838 ± 0.060	0.868 ± 0.041	0.872 ± 0.044	0.873 ± 0.046
Specificity	0.424 ± 0.038	0.561 ± 0.069	0.321 ± 0.098	0.396 ± 0.104	0.460 ± 0.126	0.527 ± 0.118	0.167 ± 0.093	0.535 ± 0.073	0.438 ± 0.091	0.447 ± 0.091
PPV	0.559 ± 0.015	0.673 ± 0.032	0.562 ± 0.037	0.596 ± 0.043	0.593 ± 0.047	0.621 ± 0.048	0.514 ± 0.031	0.663 ± 0.034	0.621 ± 0.038	0.625 ± 0.040
NPV	0.760 ± 0.029	0.794 ± 0.053	0.651 ± 0.113	0.711 ± 0.091	0.647 ± 0.092	0.661 ± 0.073	0.465 ± 0.179	0.796 ± 0.050	0.768 ± 0.064	0.771 ± 0.069
Calibration intercept	-0.005 ± 0.035	0.015 ± 0.055	0.018 ± 0.060	-0.006 ± 0.071	0.046 ± 0.041	0.037 ± 0.079	0.045 ± 0.038	-0.110 ± 0.048	-0.014 ± 0.062	-0.004 ± 0.070
Calibration slope	0.964 ± 0.127	0.682 ± 0.259	0.612 ± 0.518	0.732 ± 0.401	0.182 ± 0.216	0.246 ± 0.155	0.090 ± 0.582	0.625 ± 0.073	0.715 ± 0.220	0.783 ± 0.215

Note: Within each repetition, the reported threshold is the mean of the five training-selected thresholds weighted by outer-test set size. Sensitivity, specificity, PPV and NPV use the complete outer out-of-fold classifications with each observation's fold-specific threshold. Calibration intercept and slope are estimated jointly from the complete outer out-of-fold probability vector; this diagnostic fit does not modify predictions.

C.6 Phase 4 Cross-Fitting Diagnostics

Table C.9: Phase 4 cross-fitting diagnostics. Values are mean ± SD across the 250 outer-training fits arising from 50 repetitions of five-fold outer cross-validation, without intermediate repetition-level averaging. Static–temporal correlation is Pearson correlation between static decision scores and logits of clipped CNN probabilities; fusion coefficients are expressed on the standardised input scale. These diagnostics describe variation across overlapping training sets and are not independent replications or substitutes for outer-test performance.

Diagnostic	**Primary fusion**	**Mask-sensitivity fusion**
Static inner OOF AUROC	0.786 ± 0.059	0.786 ± 0.059
Temporal inner OOF AUROC	0.601 ± 0.102	0.635 ± 0.095
Meta-model OOF AUROC	0.731 ± 0.076	0.739 ± 0.077
Static–temporal correlation	0.139 ± 0.154	0.172 ± 0.140
Static coefficient	0.964 ± 0.335	—
Temporal coefficient	0.116 ± 0.353	—

Note: Dashes indicate mask-sensitivity coefficients not displayed in this table.