Evaluation of entropy methods using biomedical signals

A Thesis

submitted to the designated
by the Assembly
of the Department of Computer Science and Engineering
Examination Committee

by

Andreas Matsias

in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE IN DATA AND COMPUTER SYSTEMS ENGINEERING

WITH SPECIALIZATION
IN DATA SCIENCE AND ENGINEERING

University of Ioannina
School of Engineering
Ioannina 2026

Examining Committee:

- **Georgios Manis**, Assoc. Professor, Department of Computer Science and Engineering, University of Ioannina (Advisor)
- **Konstantinos Parsopoulos**, Professor, Department of Computer Science and Engineering, University of Ioannina
- **Lisimachos Paul Kondis**, Professor, Department of Computer Science and Engineering, University of Ioannina

Dedication

To my beloved partner for her unwavering love, patience, and encouragement throughout this journey. And to my family, whose endless support and belief in me made this work possible. This work is as much yours as it is mine.

Acknowledgements

I would like to express my sincere gratitude to my supervisor, Georgios Manis, for his continuous guidance, mentorship, and support throughout this research. His insightful feedback and expertise played a crucial role in shaping this thesis and pushing me to deliver my best work.

A special thanks goes to Panagiota, my life companion, for standing by my side with endless patience, love, and encouragement. Her unwavering belief in me, especially during the toughest times, has been my greatest source of motivation.

Finally, I am deeply thankful to my family, whose unconditional support and sacrifices have made this achievement possible. Without them, none of this would have been within reach.

Table of Contents

List of Figures vi

List of Tables x

Abstract xii

Εκτεταμένη Περίληψη xiv

1 Introduction 1

1.1 Background and Motivation 1
1.2 Problem Statement 2
1.3 Research Objectives 3
1.4 Thesis Contributions 4
1.5 Thesis Organization 4

2 Entropy Methods 6

2.1 Shannon Entropy 8
2.2 Renyi Entropy 9
2.3 Approximate Entropy 9
2.4 Sample Entropy 11
2.5 Permutation Entropy 12
2.6 Dispersion Entropy 13
2.7 Distribution Entropy 14
2.8 Bubble Entropy 15

3 Databases Description 17

3.1 Introduction 17
3.2 Fantasia Database 19

3.2.1 Data Description 20
3.3 MIT-BIH Arrhythmia Database 21
3.3.1 Data Description 21
3.3.2 Normal Sinus Rhythm Databases 21
3.4 Congestive Heart Failure Database 22
3.4.1 Data Description 22
3.4.2 Congestive Heart Failure Databases 23
3.4.3 Normal Sinus Rhythm Databases 23
3.5 Sudden Cardiac Death Holter Database 23
3.5.1 Data Description 24
3.5.2 Sudden Cardiac Death Holter Database 24
3.5.3 Normal Sinus Rhythm Database 24
3.6 Gait in Parkinson's Disease Database 25
3.6.1 Data Description 25
3.7 EMG Isometric Contractions Database 26
3.7.1 Data Description 26
3.8 APNEA HRV+SpO2 Database 27
3.8.1 Data Description 27
3.9 GAMEEMO Database 28
3.9.1 Data Description 28
3.10 EEG During Mental Arithmetic Tasks Database 29
3.10.1 Data Description 30
3.11 VOICED Database 31
3.11.1 Data Description 31
3.12 Pulse Transit Time PPG Dataset 32
3.12.1 Data Description 32

4 Methodology **34**
4.1 Signal Preprocessing 35
4.1.1 Filtering 35
4.1.2 Detrending 35
4.1.3 Normalization 36
4.2 Feature Extraction 36
4.2.1 Entropy Parameters 36

4.2.2 Implementation . . . 37
4.2.3 Feature Naming Convention . . . 38
4.3 Feature Selection . . . 38
4.3.1 Exhaustive Search Strategy . . . 38
4.3.2 Correlation-Based Pre-filtering . . . 39
4.3.3 Per-Entropy Family Evaluation . . . 39
4.4 Classification . . . 40
4.4.1 Classifier Descriptions . . . 40
4.4.2 Preprocessing Pipeline . . . 41
4.4.3 Implementation . . . 41
4.5 Cross-Validation Strategy . . . 42
4.5.1 Stratified K-Fold Cross-Validation . . . 42
4.5.2 Group K-Fold Cross-Validation . . . 43
4.5.3 Validation Protocol . . . 43
4.6 Evaluation Metrics . . . 44
4.6.1 Confusion Matrix . . . 44
4.6.2 Primary Metrics . . . 44
4.6.3 Advanced Metrics . . . 45
4.6.4 Reporting Protocol . . . 45

5 Results **47**
5.1 Overall Entropy Performance . . . 48
5.1.1 Mean Accuracy Ranking . . . 48
5.1.2 Performance Gap Analysis . . . 49
5.1.3 Consistency Analysis . . . 50
5.1.4 Pairwise Comparison . . . 50
5.2 Performance by Dataset . . . 52
5.2.1 Key Observations . . . 55
5.3 Class Separation Analysis . . . 56
5.3.1 Cohen's d Effect Size . . . 56
5.3.2 Effect Size by Entropy Method . . . 56
5.3.3 Interpretation . . . 58
5.3.4 Effect Size vs Classification Accuracy . . . 59
5.4 Classifier Performance . . . 60

5.4.1 Overall Classifier Distribution 60
5.4.2 Mean Accuracy by Classifier 61
5.4.3 Classifier Performance by Entropy Method 61
5.4.4 Best Classifier by Entropy Method 64
5.4.5 Top Entropy-Classifier Combinations 64
5.4.6 Interpretation 65
5.5 Detailed Dataset Results 66
5.5.1 Summary of Best Results 66
5.5.2 CHF vs NSR 67
5.5.3 Fantasia 70
5.5.4 ECG Arrhythmia 73
5.5.5 SCD vs Healthy 76
5.5.6 Gait 79
5.5.7 EMG 82
5.5.8 Apnea 85
5.5.9 GAMEEMO 88
5.5.10 Mental Arithmetic 91
5.5.11 PPG 94
5.5.12 Voiced 97
5.5.13 Selected Features Analysis 100
5.5.14 Key Findings 102
5.6 Chapter Summary 102
5.6.1 Entropy Method Performance 102
5.6.2 Dataset-Specific Findings 103
5.6.3 Classifier Performance 104
5.6.4 Class Separation and Effect Size Analysis 104
5.6.5 Feature Selection Observations 105

6 Conclusions **107**

6.1 Summary of Findings 107
6.2 Key Contributions 108
6.3 Limitations 109
6.4 Future Work 109
6.5 Closing Remarks 110

Bibliography **111**

List of Figures

5.1 Mean classification accuracy for each entropy measure across all datasets. Error bars indicate standard deviation. Bubble Entropy achieves the highest mean accuracy (83.6%) and the lowest coefficient of variation, indicating the most consistent relative performance. 49
5.2 Head-to-head win/loss matrix. Each cell shows the number of datasets where the row entropy outperforms (+) or underperforms (-) the column entropy. Green indicates dominance, red indicates inferiority. . . . 51
5.3 Classification accuracy by entropy measure across all datasets. Each bar represents the best accuracy achieved by any classifier for that entropy-dataset combination. 53
5.4 Best performing entropy method for each dataset with corresponding accuracy. 54
5.5 Cohen's d effect size heatmap across all entropy methods and datasets. Darker green indicates stronger class separation. Values above 0.8 represent large, clinically meaningful effect sizes. 58
5.6 Entropy × Classifier performance heatmap showing mean accuracy (%) across all datasets. Darker green indicates higher accuracy. The Bubble + SVM combination achieves the highest accuracy (82.4%). 63
5.7 CHF vs NSR: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (92.2%), highlighted with a black border. 68
5.8 CHF vs NSR: Class separation power (Cohen's d). Bubble Entropy achieves the largest effect size, explaining its superior classification performance. 69
5.9 CHF vs NSR: Confusion matrix for Bubble + SVM. The model correctly identifies 70 NSR and 37 CHF patients, with only 9 misclassifications. . 70

5.10 Fantasia: Entropy × Classifier performance heatmap for age classification. Bubble Entropy with SVM achieves the highest accuracy (90.0%). 71
5.11 Fantasia: Class separation power (Cohen's d) for age discrimination. Higher values indicate better separation between young and elderly subjects. 72
5.12 Fantasia: Confusion matrix for Bubble + SVM age classification. The model achieves balanced performance across young and elderly subjects. 73
5.13 ECG Arrhythmia: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (93.8%). 74
5.14 ECG Arrhythmia: Class separation power (Cohen's d) for arrhythmia detection. 75
5.15 ECG Arrhythmia: Confusion matrix for Bubble + SVM. High specificity (97.9%) with moderate sensitivity (81.3%) reflects the class imbalance in the dataset. 76
5.16 SCD vs Healthy: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (88.1%). 77
5.17 SCD vs Healthy: Class separation power (Cohen's d) for sudden cardiac death risk identification. 78
5.18 SCD vs Healthy: Confusion matrix for Bubble + SVM sudden cardiac death risk prediction. 79
5.19 Gait: Entropy × Classifier performance heatmap for Parkinson's detection. Dispersion Entropy with SVM achieves the highest accuracy (75.3%). 80
5.20 Gait: Class separation power (Cohen's d) for Parkinson's detection. Dispersion Entropy achieves the highest effect size, explaining its superior performance over Bubble Entropy. 81
5.21 Gait: Confusion matrix for Dispersion + SVM Parkinson's detection. Unlike other datasets, Dispersion Entropy outperforms Bubble Entropy here. 82
5.22 EMG: Entropy × Classifier performance heatmap for neuromuscular disorder classification. Dispersion Entropy with SVM achieves the highest accuracy (80.5%). 83
5.23 EMG: Class separation power (Cohen's d) for neuromuscular disorder classification. Dispersion Entropy achieves the highest effect size. 84

5.24 EMG: Confusion matrix for Dispersion + SVM neuromuscular disorder classification using deltoid muscle signals. 85
5.25 Apnea: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (91.6%) using SpO2 difference signal. 86
5.26 Apnea: Class separation power (Cohen's d) for sleep apnea detection. . 87
5.27 Apnea: Confusion matrix for Bubble + SVM sleep apnea detection using SpO2 difference signal. 88
5.28 GAMEEMO: Entropy × Classifier performance heatmap for emotion classification. Bubble Entropy with RF achieves the highest accuracy (78.6%). 89
5.29 GAMEEMO: Class separation power (Cohen's d) for emotional state classification. 90
5.30 GAMEEMO: Confusion matrix for Bubble + RF emotion classification. Random Forest is the optimal classifier for this EEG-based task. 91
5.31 Mental Arithmetic: Entropy × Classifier performance heatmap for cognitive load classification. Bubble Entropy with kNN achieves the highest accuracy (73.6%). 92
5.32 Mental Arithmetic: Class separation power (Cohen's d) for cognitive load detection. 93
5.33 Mental Arithmetic: Confusion matrix for Bubble + kNN cognitive load classification using Fz electrode. 94
5.34 PPG: Entropy × Classifier performance heatmap. Unlike other datasets, Permutation Entropy achieves the best results (95.5% with kNN). . . . 95
5.35 PPG: Class separation power (Cohen's d). Approximate Entropy achieves the highest effect size (d=3.09), yet Permutation Entropy achieves the best classification accuracy (95.5%), illustrating that the relationship between effect size and classification accuracy is not strictly linear. . . . 96
5.36 PPG: Confusion matrix for Permutation + kNN. This is the only dataset where Permutation Entropy outperforms Bubble Entropy, achieving the highest accuracy (95.5%) across all datasets. 97
5.37 Voiced: Entropy × Classifier performance heatmap for voice disorder detection. Bubble Entropy with SVM achieves the highest accuracy (76.0%). 98

5.38 Voiced: Class separation power (Cohen's d) for pathological voice detection. 99
5.39 Voiced: Confusion matrix for Bubble + SVM pathological voice detection.100

List of Tables

2.1 Summary of the eight entropy methods evaluated in this study. 8

3.1 Summary of the eleven biomedical datasets used in this study. 19
3.2 Fantasia Database summary 20
3.3 MIT-BIH Arrhythmia and Normal Sinus Rhythm databases summary . 22
3.4 CHF vs NSR databases summary 23
3.5 Sudden Cardiac Death vs Healthy dataset summary 24
3.6 Gait in Parkinson's Disease database summary 25
3.7 EMG Isometric Contractions database summary 27
3.8 APNEA HRV+SpO2 database summary 28
3.9 GAMEEMO database summary 29
3.10 EEG During Mental Arithmetic Tasks database summary 31
3.11 VOICED database summary 32
3.12 Pulse Transit Time PPG dataset summary 33

4.1 Entropy parameters and resulting feature counts. 37
4.2 Classifier configurations used in this study. 42
4.3 Confusion matrix for binary classification. 44

5.1 Overall entropy performance ranking by mean accuracy. 48
5.2 Best performing entropy and classifier for each dataset. 55
5.3 Cohen's d effect sizes by entropy method. 57
5.4 Distribution of best-performing classifiers across all entropy-dataset combinations. 60
5.5 Mean classification accuracy by classifier across all experiments. 61
5.6 Mean accuracy (%) for each entropy-classifier combination. 62
5.7 Number of dataset wins per classifier, grouped by entropy method. . . 64

5.8 Top 10 entropy-classifier combinations by mean accuracy. 65
5.9 Best classification results per dataset. 66
5.10 CHF vs NSR: Top 3 entropy-classifier combinations. 67
5.11 Fantasia: Top 3 entropy-classifier combinations. 70
5.12 ECG Arrhythmia: Top 3 entropy-classifier combinations. 73
5.13 SCD vs Healthy: Top 3 entropy-classifier combinations. 76
5.14 Gait: Top 3 entropy-classifier combinations. 79
5.15 EMG: Top 3 entropy-classifier combinations. 82
5.16 Apnea: Top 3 entropy-classifier combinations. 85
5.17 GAMEEMO: Top 3 entropy-classifier combinations. 88
5.18 Mental Arithmetic: Top 3 entropy-classifier combinations. 91
5.19 PPG: Top 3 entropy-classifier combinations. 94
5.20 Voiced: Top 3 entropy-classifier combinations. 97
5.21 Selected features for best performing entropy-classifier combination per dataset. Features are named as `entropy_parameter` (e.g., `bubben_m12` = Bubble Entropy with $m = 12$). 101

Abstract

Andreas Matsias , M.Sc. in Data and Computer Systems Engineering, Department of Computer Science and Engineering, School of Engineering, University of Ioannina, Greece, 2026 .
Evaluation of entropy methods using biomedical signals.
Advisor: Georgios Manis, Associate Professor .

Entropy-based measures have emerged as a powerful tool for characterizing the complexity and irregularity of biomedical signals, offering an alternative to traditional morphological feature extraction. However, the growing number of available entropy formulations creates a method selection challenge, as most existing studies evaluate only a limited subset of methods on specific signal types, preventing meaningful cross-method and cross-domain comparisons.

This thesis presents a comprehensive and systematic evaluation of eight entropy methods—Shannon, Renyi, Approximate, Sample, Permutation, Dispersion, Distribution, and Bubble Entropy—across eleven diverse biomedical datasets spanning cardiovascular (ECG, HRV, PPG), neural (EEG), muscular (EMG), respiratory (SpO2), gait, and voice signals. Five machine learning classifiers—k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest (RF), Gaussian Naive Bayes (GNB), and Gradient Boosting (GB)—are employed within a unified experimental framework using 5-fold cross-validation to ensure rigorous and reproducible comparisons.

The results demonstrate that Bubble Entropy, a relatively recent parameter-free method, achieves superior performance in the majority of classification tasks, winning 8 out of 11 datasets with a mean accuracy of 83.6%.

Dispersion Entropy ranks second overall, while Permutation Entropy achieves the highest single-dataset accuracy of 95.5% on PPG signals. Dispersion Entropy also proves more effective for gait and EMG classification tasks. Among classifiers,

kNN and SVM consistently outperform ensemble methods, suggesting that simpler models are better suited for entropy-based feature spaces. Additionally, Cohen's d effect size analysis provides insight into the class separation power of each entropy method, offering a physiological interpretation of discriminability beyond standard accuracy metrics.

This work constitutes the most extensive benchmark of entropy methods for biomedical signal classification to date and provides evidence-based, dataset-specific guidelines for entropy method and classifier selection.

Keywords: Entropy, Biomedical Signals, Signal Classification, Bubble Entropy, Approximate Entropy, Dispersion Entropy, Sample Entropy, Permutation Entropy, Shannon Entropy, Renyi Entropy, Distribution Entropy, Machine Learning, Feature Extraction

Εκτεταμενη Περιληψη

Ανδρέας Μάτσιας , Δ.Μ.Σ. στη Μηχανική Δεδομένων και Υπολογιστικών Συστημάτων, Τμήμα Μηχανικών Η/Υ και Πληροφορικής, Πολυτεχνική Σχολή, Πανεπιστήμιο Ιωαννίνων, 2026 .
Τίτλος Διατριβής: Αξιολόγηση μεθόδων εντροπίας με βιοϊατρικά σήματα. .
Επιβλέπων: Μανής Γεώργιος, Αναπληρωτής Καθηγητής .

Εισαγωγή και Κίνητρο

Τα βιοϊατρικά σήματα, όπως τα ηλεκτροκαρδιογραφήματα (ECG), τα ηλεκτροεγκεφαλογραφήματα (EEG), τα ηλεκτρομυογραφήματα (EMG), τα φωτοπληθυσμογραφήματα (PPG) και οι ηχογραφήσεις φωνής, αποτελούν μη επεμβατικά μέσα παρακολούθησης της φυσιολογικής κατάστασης του ανθρώπινου σώματος. Οι παραδοσιακές μέθοδοι ανάλυσης βασίζονται σε μορφολογικά χαρακτηριστικά που απαιτούν εξειδικευμένη προεπεξεργασία και είναι ευαίσθητες σε θόρυβο και τεχνουργήματα. Τα μέτρα εντροπίας προσφέρουν μια εναλλακτική προσέγγιση, ποσοτικοποιώντας την πολυπλοκότητα και την ακανονικότητα ενός σήματος χωρίς να απαιτούν υποθέσεις για τη δομή του.

Ωστόσο, ο αυξανόμενος αριθμός διαθέσιμων μεθόδων εντροπίας δημιουργεί ένα πρόβλημα επιλογής. Οι περισσότερες υπάρχουσες μελέτες αξιολογούν μεμονωμένες μεθόδους ή περιορισμένα υποσύνολα, σε διαφορετικά σύνολα δεδομένων και με διαφορετικά πρωτόκολλα, καθιστώντας δύσκολη τη σύγκριση μεταξύ τους. Η παρούσα διπλωματική εργασία αντιμετωπίζει αυτό το κενό μέσω μιας συστηματικής αξιολόγησης.

Στόχοι

Ο πρωταρχικός στόχος αυτής της εργασίας είναι η συστηματική σύγκριση μεθόδων εντροπίας για την ταξινόμηση βιοϊατρικών σημάτων. Συγκεκριμένα, η εργασία αποσκοπεί στα εξής:

- Αξιολόγηση οκτώ μεθόδων εντροπίας—Shannon, Renyi, Approximate, Sample, Permutation, Dispersion, Distribution και Bubble—σε ένα ενιαίο πειραματικό πλαίσιο.
- Εκτίμηση της απόδοσης σε έντεκα ποικίλα σύνολα βιοϊατρικών δεδομένων που καλύπτουν καρδιαγγειακά, εγκεφαλικά, μυϊκά, αναπνευστικά σήματα, σήματα βάδισης, φωνής και PPG.

Μεθοδολογία

Η πειραματική μεθοδολογία βασίζεται σε ένα ενιαίο πλαίσιο αξιολόγησης. Για κάθε σύνολο δεδομένων, εξάγονται χαρακτηριστικά εντροπίας από τα βιοϊατρικά σήματα χρησιμοποιώντας και τις οκτώ μεθόδους. Στη συνέχεια, εφαρμόζονται πέντε ταξινομητές μηχανικής μάθησης με 5-fold cross-validation για τη διασφάλιση αξιόπιστων και αναπαραγώγιμων αποτελεσμάτων. Επιπλέον, εφαρμόζεται ανάλυση μεγέθους επίδρασης Cohen's d για την ποσοτικοποίηση της ικανότητας διαχωρισμού κλάσεων κάθε μεθόδου εντροπίας.

Τα έντεκα σύνολα δεδομένων καλύπτουν ευρύ φάσμα βιοϊατρικών σημάτων: καρδιαγγειακά σήματα (CHF vs NSR, Fantasia, ECG Αρρυθμία, SCD), αναπνευστικά σήματα (Άπνοια), εγκεφαλικά σήματα (GAMEEMO, Νοητική Αριθμητική), μυϊκά σήματα (EMG), σήματα βάδισης (Parkinson), PPG και ηχογραφήσεις φωνής (VOICED).

Αποτελέσματα

Τα πειραματικά αποτελέσματα αποκαλύπτουν μια σαφή ιεραρχία απόδοσης μεταξύ των μεθόδων εντροπίας:

1. Η **Bubble Entropy** αναδείχθηκε ως η μέθοδος με την υψηλότερη απόδοση,

επικρατώντας σε 8 από τα 11 σύνολα δεδομένων (72,7%) με μέση ακρίβεια 83,6% ± 9,0%.

2. Η **Dispersion Entropy** κατατάχθηκε δεύτερη με μέση ακρίβεια 77,7% ± 8,6% και αποδείχθηκε ανώτερη της Bubble Entropy για συγκεκριμένους τύπους σημάτων, ιδιαίτερα βάδισης (ανίχνευση Parkinson) και EMG.

3. Η **Approximate Entropy** κατατάχθηκε τρίτη (77,4% ± 8,5%).

4. Η **Permutation Entropy** κατατάχθηκε τέταρτη (76,6% ± 11,5%), επιτυγχάνοντας ωστόσο την υψηλότερη ακρίβεια μεμονωμένου συνόλου δεδομένων (95,5%) στα σήματα PPG.

5. Τα παραδοσιακά μέτρα θεωρίας πληροφοριών (Shannon, Renyi) παρουσίασαν χαμηλότερη απόδοση, καταλαμβάνοντας σταθερά τις τελευταίες θέσεις της κατάταξης.

Όσον αφορά τους ταξινομητές, ο kNN πέτυχε το υψηλότερο ποσοστό νίκης (58,0%), ακολουθούμενος από τον SVM (22,7%). Αξιοσημείωτο είναι ότι οι απλοί ταξινομητές υπερτερούν σταθερά των μεθόδων συνόλου (ensemble), υποδεικνύοντας ότι τα χαρακτηριστικά εντροπίας δεν απαιτούν πολύπλοκα μοντέλα ταξινόμησης. Ο βέλτιστος συνολικός συνδυασμός ήταν Bubble Entropy με SVM, επιτυγχάνοντας μέση ακρίβεια 82,4%.

Συμπεράσματα

Η παρούσα εργασία αποτελεί την πιο εκτεταμένη συγκριτική αξιολόγηση μεθόδων εντροπίας για ταξινόμηση βιοϊατρικών σημάτων στη βιβλιογραφία. Τα κύρια συμπεράσματα είναι:

- Η Bubble Entropy αποτελεί τη συνιστώμενη μέθοδο γενικής χρήσης, λόγω της υψηλής απόδοσης και της απουσίας παραμέτρων.
- Για σήματα PPG, η Permutation Entropy είναι προτιμότερη, ενώ για σήματα βάδισης και EMG, η Dispersion Entropy αποδίδει καλύτερα.

- Οι απλοί ταξινομητές (kNN, SVM) επαρκούν για χαρακτηριστικά βασισμένα σε εντροπία, μειώνοντας τις υπολογιστικές απαιτήσεις χωρίς να επηρεάζεται η ακρίβεια.

Λέξεις-κλειδιά: Εντροπία, Βιοϊατρικά Σήματα, Ταξινόμηση Σημάτων, Bubble Entropy, Approximate Entropy, Dispersion Entropy, Sample Entropy, Permutation Entropy, Shannon Entropy, Renyi Entropy, Distribution Entropy, Μηχανική Μάθηση, Εξαγωγή Χαρακτηριστικών

Chapter 1

Introduction

1.1 Background and Motivation

1.2 Problem Statement

1.3 Research Objectives

1.4 Thesis Contributions

1.5 Thesis Organization

1.1 Background and Motivation

Biomedical signals such as electrocardiograms (ECG), electroencephalograms (EEG), electromyograms (EMG), photoplethysmograms (PPG), and voice recordings offer a non-invasive window into the physiological state of the human body. These signals encapsulate critical data regarding underlying health conditions, enabling clinicians to diagnose pathologies, monitor therapeutic progress, and predict adverse clinical events. Despite their strong diagnostic utility, the inherent morphological complexity, high dimensionality, and subject variability of these signals pose multiple challenges to the deployment of reliable automated analytical systems.

Traditional approaches to biomedical signal analysis rely on morphological features such as R-peak detection in ECG, spectral power bands in EEG, and formant frequencies in voice signals. These require precise signal segmentation and domain-specific preprocessing. While effective in controlled clinical environments, these meth-

ods are often sensitive to noise, artifacts, and recording conditions, limiting their applicability in real-world scenarios like wearable health monitoring.

Entropy-based measures offer an alternative approach to characterizing biomedical signals. Rather than extracting explicit morphological features, entropy quantifies the underlying complexity, irregularity, or unpredictability of a signal. This approach captures the intrinsic dynamics and patterns of physiological systems without requiring assumptions about signal structure or extensive preprocessing. The clinical relevance of entropy analysis is rooted in the observation that pathology often degrades physiological complexity, leaving a measurable signature on the dynamics of human signals. This degradation can manifest as reduced heart rate variability in cardiovascular disease, constrained motor patterns in neurological disorders, or altered breathing dynamics.

Since the introduction of Approximate Entropy, numerous entropy methods have been developed, each with distinct properties regarding noise robustness, computational efficiency, and parameter sensitivity. Recent methods such as Bubble Entropy, Dispersion Entropy, and Distribution Entropy have shown promising results in various biomedical applications. However, the growing number of available entropy measures creates a selection problem for researchers: which entropy method should be used for a specific biomedical classification task?

1.2 Problem Statement

Despite extensive research on individual entropy methods, a comprehensive comparative evaluation across multiple biomedical signal types and classification tasks remains lacking. Most existing studies focus on a single entropy measure or a limited subset, making it difficult to draw general conclusions about relative performance. Furthermore, comparisons are often conducted on different datasets with varying preprocessing protocols, preventing meaningful cross-study synthesis.

This fragmented landscape presents several challenges:

- **Method selection uncertainty**: Researchers lack evidence-based guidelines for choosing among entropy measures, often defaulting to methods used in prior similar studies without systematic justification.

- **Dataset-specific performance**: An entropy method that excels for one signal type may underperform for another, yet these dependencies are poorly characterized.
- **Classifier interactions**: The interplay between entropy features and machine learning classifiers is underexplored, leaving open questions about optimal entropy-classifier pairings.
- **Statistical validation**: Many comparative studies lack rigorous statistical testing, making it unclear whether observed performance differences are significant or attributable to chance.

1.3 Research Objectives

The primary objective of this thesis is to conduct a systematic and rigorous comparison of entropy methods for biomedical signal classification. Specifically, this work aims to:

1. Evaluate eight entropy methods using a unified experimental framework: Shannon, Renyi, Approximate, Sample, Permutation, Dispersion, Distribution, and Bubble Entropy.
2. Assess performance across eleven diverse biomedical datasets spanning cardiovascular (ECG, HRV, PPG), neural (EEG), muscular (EMG), respiratory (SpO2), gait, and voice signals.
3. Compare five machine learning classifiers to identify optimal entropy-classifier combinations: k-Nearest Neighbors (kNN), Support Vector Machine (SVM), Random Forest (RF), Gaussian Naive Bayes (GNB), and Gradient Boosting (GB).
4. Quantify class separation power using Cohen's d effect sizes, providing insight into why certain entropy methods outperform others.
5. Develop practical recommendations for entropy method selection based on signal type and application requirements.

1.4 Thesis Contributions

This thesis makes the following contributions to the field of biomedical signal analysis:

- **Comprehensive benchmark**: The first systematic comparison of eight entropy methods across eleven diverse biomedical datasets, representing an extensive evaluation of its kind in the literature.
- **Empirical validation of Bubble Entropy**: Demonstration that this relatively recent parameter-free method outperforms established entropy measures for the majority of biomedical classification tasks (8 of 11 datasets).
- **Dataset-specific guidelines**: Identification of exceptions where alternative entropy methods provide superior performance—Permutation Entropy for PPG signals and Dispersion Entropy for Gait and EMG signals.
- **Classifier recommendations**: Evidence that simple classifiers (kNN, SVM) outperform ensemble methods (RF, GB) for entropy-based features, enabling computational efficiency without sacrificing accuracy.
- **Effect size analysis**: Application of Cohen's d to quantify class separation power, providing physiological interpretation of entropy discriminability beyond standard accuracy metrics.

1.5 Thesis Organization

The rest of this thesis is organized into five chapters. **Chapter 2** provides theoretical background on the eight entropy methods evaluated in this study, presenting the conceptual basis, mathematical formulation, algorithm, and parameters of each measure. **Chapter 3** describes the eleven biomedical datasets used in this study, including cardiovascular signals (CHF vs NSR, Fantasia, ECG Arrhythmia, SCD vs Healthy), respiratory signals (Apnea), neural signals (GAMEEMO, Mental Arithmetic), muscular signals (EMG), gait signals, PPG signals, and voice recordings. **Chapter 4** presents the experimental framework, including feature extraction procedures, machine learning classifiers, cross-validation protocols, and statistical analysis methods. **Chapter 5** reports the experimental findings, including entropy rankings, classifier performance

analysis, and detailed results for each dataset. Finally, **Chapter 6** summarizes the key findings, discusses contributions and limitations, and proposes directions for future research.

Chapter 2

Entropy Methods

2.1 Shannon Entropy

2.2 Renyi Entropy

2.3 Approximate Entropy

2.4 Sample Entropy

2.5 Permutation Entropy

2.6 Dispersion Entropy

2.7 Distribution Entropy

2.8 Bubble Entropy

Entropy, as introduced by Shannon [1], measures the average information content of a discrete random variable. It quantifies the uncertainty associated with a probability distribution, providing a foundation for measuring the regularity of any sequential data source, including physiological time series.

Building on Shannon's formulation, a range of entropy measures has been proposed to address its limitations when applied to finite, noisy, real-world signals. Each successive formulation targets a specific shortcoming of its predecessors: Approximate Entropy [2] introduced template matching to estimate regularity directly from short time series; Sample Entropy [3] corrected its self-matching bias; Permutation Entropy [4] shifted the focus to ordinal patterns for robustness to amplitude noise; Dispersion Entropy [5] reintroduced amplitude information through symbol

mapping; Distribution Entropy [6] replaced threshold-based pattern counting with a distance-distribution approach; and Bubble Entropy [7] eliminated free parameters altogether by exploiting the sorting complexity of ordinal rankings. Meanwhile, Renyi Entropy [8] generalized Shannon's measure with a tunable order parameter that controls sensitivity to rare events.

This chapter presents the mathematical formulation, algorithmic procedure, and parameter choices for each of the eight entropy measures evaluated in this thesis:

- **Shannon Entropy.** Classical information-theoretic measure of uncertainty in a probability distribution.
- **Renyi Entropy.** Generalization of Shannon entropy with tunable sensitivity to probability distribution tails.
- **Approximate Entropy (ApEn).** Measures regularity by quantifying the likelihood that similar patterns remain similar.
- **Sample Entropy (SampEn).** Improved version of ApEn that eliminates self-matching bias.
- **Permutation Entropy (PermEn).** Captures ordinal patterns based on relative amplitude ordering.
- **Dispersion Entropy (DispEn).** Combines amplitude information with temporal patterns using symbol mapping.
- **Distribution Entropy (DistEn).** Measures complexity based on the distribution of distances between embedded vectors.
- **Bubble Entropy (BubbEn).** Parameter-free measure based on the number of swaps in bubble sort ranking.

Table 2.1 provides a concise overview of all eight measures, including their key properties and parameter requirements.

The following sections describe each entropy measure, including its theoretical background, mathematical formulation, and the specific parameters used in this study.

Table 2.1: Summary of the eight entropy methods evaluated in this study.

Method	**Abbr.**	**Core Concept**	**Parameters**	**Year**
Shannon Entropy	ShEn	Probability distribution uncertainty	Binning	1948
Renyi Entropy	RenEn	Tunable tail sensitivity	α, binning	1961
Approximate Entropy	ApEn	Template self-matching regularity	m, r, N	1991
Sample Entropy	SampEn	Bias-corrected regularity	m, r, N	2000
Permutation Entropy	PermEn	Ordinal pattern distribution	m, τ	2002
Dispersion Entropy	DispEn	Amplitude-aware symbol patterns	m, c, τ	2016
Distribution Entropy	DistEn	Embedded vector distances	m, τ, B	2015
Bubble Entropy	BubbEn	Sorting complexity (parameter-free)	m only	2017

2.1 Shannon Entropy

Shannon Entropy, introduced by Claude E. Shannon in 1948 [1], is the foundational measure of information theory. It quantifies the average amount of information, or equivalently, uncertainty, contained in a probability distribution. In signal analysis, Shannon Entropy assesses how unpredictable the amplitude values of a signal are: higher entropy indicates greater randomness or complexity, while lower entropy suggests regularity and predictability.

Given a discrete set of outcomes with probabilities $\{p_1, p_2, \ldots, p_n\}$, Shannon Entropy H is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log_b(p_i) \tag{2.1}$$

where:

- p_i represents the probability of the i-th outcome,
- b denotes the base of the logarithm (commonly $b = 2$ to express entropy in bits).

Shannon Entropy reaches its maximum when all outcomes are equally probable, reflecting maximum uncertainty. Conversely, it reaches its minimum value of zero when one outcome is certain ($p_i = 1$ for some i, and 0 for all others). Shannon Entropy has no fixed upper bound; it grows with the number of possible outcomes.

This measure forms the theoretical foundation for many entropy-based complexity metrics used in biomedical signal analysis, including Permutation Entropy, Dispersion Entropy, and Distribution Entropy, which apply Shannon Entropy to different symbolic representations of the signal.

2.2 Renyi Entropy

Renyi Entropy, introduced by Alfred Renyi in 1961 [8], generalizes Shannon Entropy by adding a parameter α that controls sensitivity to different parts of the probability distribution. This flexibility makes it useful for analyzing systems where the focus should be on either common or rare events, depending on the application.

Given a discrete probability distribution $\{p_1, p_2, \ldots, p_n\}$ and a non-negative parameter $\alpha \neq 1$, Renyi Entropy of order α is defined as:

$$H_\alpha(X) = \frac{1}{1-\alpha} \log_b \left(\sum_{i=1}^{n} p_i^\alpha \right) \tag{2.2}$$

The parameter α determines how the entropy weighs the contribution of different probabilities:

- When $\alpha > 1$, the entropy becomes more sensitive to high-probability events.
- When $\alpha < 1$, it becomes more sensitive to low-probability (rare) events.
- In the limit as $\alpha \to 1$, Renyi Entropy converges to Shannon Entropy.

In this study, Renyi Entropy with $\alpha = 2$ is used, which corresponds to the collision entropy. This choice emphasizes the most probable outcomes and provides a measure that is computationally efficient while still capturing essential distributional properties of the signal.

2.3 Approximate Entropy

Approximate Entropy (ApEn), introduced by S.M. Pincus in 1991 [2], quantifies the regularity and unpredictability of time series data. It is particularly effective for analyzing short and noisy datasets where traditional entropy methods may be unreliable.

ApEn estimates the likelihood that patterns which are similar for m consecutive points will remain similar when extended to $m+1$ points, providing a numerical measure of signal complexity.

The mathematical definition of Approximate Entropy is:

$$\text{ApEn}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \tag{2.3}$$

where:

- m is the embedding dimension, defining the length of compared vectors,
- r is the similarity tolerance (threshold),
- N is the total length of the time series.

The function $\phi^m(r)$ measures the average logarithmic frequency of similar vectors of length m:

$$\phi^m(r) = \frac{1}{N-m+1} \sum_{i=1}^{N-m+1} \log C_i^m(r) \tag{2.4}$$

Each quantity $C_i^m(r)$ represents the proportion of vectors x_j^m that are within distance r of x_i^m (including self-matches):

$$C_i^m(r) = \frac{|\{j \mid d(x_i^m, x_j^m) < r\}|}{N-m+1} \tag{2.5}$$

The vectors x_i^m are constructed by embedding the time series $u(i)$ into m-dimensional space:

$$x_i^m = [u(i), u(i+1), \ldots, u(i+m-1)] \tag{2.6}$$

The distance function $d(x_i^m, x_j^m)$ is typically defined using the maximum norm (Chebyshev distance):

$$d(x_i^m, x_j^m) = \max_{k=1,\ldots,m} |u(i+k-1) - u(j+k-1)| \tag{2.7}$$

The tolerance threshold r is commonly expressed as a fraction of the signal standard deviation:

$$\text{Threshold} = r \times \text{std}(\text{signal}) \tag{2.8}$$

Approximate Entropy is interpreted as follows:

- **Low ApEn values** indicate more regular and predictable behavior.
- **High ApEn values** suggest greater complexity or randomness.

A limitation of ApEn is the inclusion of self-matches in the counting, which introduces a bias that can affect estimation accuracy, particularly for short time series. This limitation motivated the development of Sample Entropy.

In this study, ApEn was computed with embedding dimensions $m \in \{2, 3, 4\}$ and tolerance values $r \in \{0.15, 0.20, 0.25\}$ times the signal standard deviation, yielding 9 features per signal.

2.4 Sample Entropy

Sample Entropy (SampEn), introduced by Joshua S. Richman and J. Randall Moorman in 2000 [3], was developed to address limitations of Approximate Entropy, specifically the inclusion of self-matches and sensitivity to data length. By excluding self-matches from the similarity count, SampEn provides more consistent and less biased complexity estimates, particularly for shorter time series.

Formally, SampEn is defined as the negative natural logarithm of the conditional probability that two sequences similar for m consecutive points remain similar at the next point, within tolerance r, as data length N tends to infinity:

$$\text{SampEn}(m, r) = \lim_{N \to \infty} -\log \left(\frac{\sum_{i=1}^{N-m} C_i^{m+1}(r)}{\sum_{i=1}^{N-m} C_i^{m}(r)} \right) \tag{2.9}$$

where $C_i^m(r)$ represents the count of template vectors of length m that lie within distance r from the i-th template vector, excluding self-matches.

The principal distinction between SampEn and ApEn lies in this exclusion of self-matches, which effectively reduces bias and enhances consistency of entropy estimates. While ApEn includes the comparison of each vector with itself, artificially inflating the similarity count, SampEn avoids this by requiring $j \neq i$ in all comparisons.

SampEn also demonstrates reduced dependence on time series length compared to ApEn, providing more reliable complexity estimates when data availability is limited. Both measures share the same parameters: embedding dimension m and tolerance threshold r (expressed as a fraction of signal standard deviation). However, SampEn generally yields more robust quantifications of signal complexity.

As with ApEn:

- **Low SampEn values** indicate regular, predictable signals.
- **High SampEn values** indicate complex, irregular signals.

In this study, SampEn was computed with embedding dimensions $m \in \{2, 3, 4\}$ and tolerance values $r \in \{0.15, 0.20, 0.25\}$ times the signal standard deviation, yielding 9 features per signal.

2.5 Permutation Entropy

Permutation Entropy (PE), introduced by C. Bandt and B. Pompe in 2002 [4], quantifies time series complexity by analyzing ordinal patterns rather than amplitude values. Unlike ApEn and SampEn, which depend on distance thresholds, PE focuses on the relative ordering of consecutive values, making it robust to noise and invariant to monotonic transformations of the data.

Mathematically, given an embedding dimension m, PE is defined as the Shannon entropy of the distribution of ordinal patterns of length m extracted from the time series:

$$\mathrm{PE}(m) = -\sum_{i=1}^{m!} p_i \log_b(p_i) \tag{2.10}$$

where:

- p_i denotes the probability of the i-th permutation pattern occurring in the time series,
- $m!$ represents the total number of possible ordinal patterns of length m,
- b is the base of the logarithm, typically chosen as 2 for expressing entropy in bits.

For each subsequence of m consecutive values, the ordinal pattern is determined by ranking the values from smallest to largest. For example, with $m = 3$, a subsequence $[4, 2, 7]$ would be mapped to the pattern $(2, 1, 3)$ since the second element is smallest, the first is middle, and the third is largest.

The key advantage of Permutation Entropy lies in its ability to capture intrinsic dynamical properties by examining relative ordering rather than actual values. This makes PE highly effective in noisy or non-stationary environments where amplitude-based methods may struggle.

Compared to ApEn and SampEn, PE is computationally more efficient and less sensitive to parameter choices. Its reliance on ordinal patterns provides a complementary perspective on complexity, often revealing features that amplitude-based methods might overlook.

Normalized PE divides by $\log(m!)$ to obtain values in $[0, 1]$, where 0 indicates a completely regular signal and 1 indicates maximum complexity.

In this study, PE was computed with embedding dimensions $m \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ using normalized values, yielding 9 features per signal.

2.6 Dispersion Entropy

Dispersion Entropy (DispEn), introduced by M. Rostaghi and H. Azami in 2016 [5], is a complexity measure that restores amplitude sensitivity absent in Permutation Entropy while retaining the robustness of ordinal-based methods. It quantifies irregularity by analyzing dispersion patterns of normalized embedding vectors mapped to discrete classes.

Formally, DispEn is defined as the Shannon entropy of the distribution of dispersion patterns:

$$\text{DispEn}(m, c, \tau) = -\sum_{k=1}^{c^m} p_k \log(p_k) \tag{2.11}$$

where m is the embedding dimension, c denotes the number of classes for mapping normalized values, and τ is the time delay.

The embedding vectors are constructed as:

$$\mathbf{x}_i^m = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}] \tag{2.12}$$

To reduce amplitude variations and facilitate classification into discrete categories, z-score normalization is applied:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{2.13}$$

where μ and σ represent the mean and standard deviation of the time series.

Each normalized value is then mapped to one of c classes by partitioning the range into equal intervals:

$$\tilde{z}_i = \text{floor}\left(c \times \frac{z_i - z_{\min}}{z_{\max} - z_{\min}}\right) + 1, \quad \tilde{z}_i \in \{1, 2, \ldots, c\} \tag{2.14}$$

Each normalized embedding vector is thus mapped to one of c^m possible dispersion patterns. The probability of each pattern is computed as:

$$p_k = \frac{|\{i \mid \mathbf{z}_i^m = \mathbf{x}_k\}|}{N - (m-1)\tau} \tag{2.15}$$

The strength of Dispersion Entropy lies in its ability to capture both amplitude and temporal variations by classifying normalized values into discrete categories before entropy calculation. This approach enhances sensitivity to subtle dynamical changes while maintaining computational efficiency. Unlike Shannon, Renyi, ApEn, SampEn, and Permutation Entropy, DispEn involves an explicit classification of amplitude values.

In this study, DispEn was computed with embedding dimensions $m \in \{2, 3, 4, 5\}$, number of classes $c \in \{3, 4, 5, 6, 7, 8, 9\}$, and time delay $\tau = 1$, yielding 28 features per signal.

2.7 Distribution Entropy

Distribution Entropy (DistEn), introduced by P. Li et al. in 2015 [6], quantifies complexity by analyzing the distribution of distances between embedding vectors. Unlike methods that rely on categorical or ordinal classifications, DistEn evaluates complexity through distance-based pattern analysis of the embedding space.

Given a time series $x = \{x_1, x_2, \ldots, x_N\}$, embedding vectors are constructed as:

$$\mathbf{x}_i^m = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}] \tag{2.16}$$

where m is the embedding dimension and τ is the time delay.

For each pair of embedding vectors $\mathbf{x}_i^m$ and $\mathbf{x}_j^m$, a distance metric is computed. Two commonly used distance functions are:

Euclidean Distance:

$$d_E(\mathbf{x}_i^m, \mathbf{x}_j^m) = \sqrt{\sum_{k=0}^{m-1} (x_{i+k\tau} - x_{j+k\tau})^2} \tag{2.17}$$

Chebyshev Distance:

$$d_C(\mathbf{x}_i^m, \mathbf{x}_j^m) = \max_{k=0,\ldots,m-1} |x_{i+k\tau} - x_{j+k\tau}| \tag{2.18}$$

Once distances are computed for all valid vector pairs, they are discretized into c bins. Distribution Entropy is then calculated as the Shannon entropy of the resulting histogram:

$$\text{DistEn}(m, c, \tau) = -\sum_{k=1}^{c} p_k \log p_k \tag{2.19}$$

where p_k is the empirical probability of distances falling into the k-th bin.

Comparison with Dispersion Entropy. While DispEn categorizes embedding vector values directly into discrete classes based on normalized amplitudes, DistEn focuses on the distribution of distances between vectors, capturing a more global view of the time series structure. This makes DistEn more sensitive to subtle dynamical changes and long-range correlations, whereas DispEn may be more suitable for detecting local variations and abrupt shifts. Additionally, DistEn relies solely on inter-vector distances without requiring preprocessing for normalization, potentially simplifying parameter tuning.

In this study, DistEn was computed with embedding dimensions $m \in \{2, 3, 4, 5\}$, number of bins $c \in \{3, 4, 5, 6, 7, 8\}$, and time delay $\tau = 1$, yielding 24 features per signal.

2.8 Bubble Entropy

Bubble Entropy (BubbEn) [7] is a complexity measure based on the Bubble Sort algorithm. Unlike threshold-dependent methods such as ApEn and SampEn, BubbEn quantifies signal irregularity by counting the number of element exchanges (swaps) needed to sort embedded vectors, eliminating the need for tolerance parameters.

For embedding dimension m and time delay τ, Bubble Entropy is computed as the difference between consecutive swap count entropies:

$$\text{BubbEn}(x; m, \tau) = H_m - H_{m+1} \tag{2.20}$$

where H_m represents the Shannon entropy of the swap count distribution:

$$H_m = -\sum_k p_k^{(m)} \log p_k^{(m)} \tag{2.21}$$

Here, $p_k^{(m)}$ is the probability of requiring exactly k swaps when sorting vectors of dimension m. The embedded vectors are formed as:

$$\mathbf{v}_i = [x_i, x_{i+\tau}, x_{i+2\tau}, \ldots, x_{i+(m-1)\tau}] \tag{2.22}$$

The Bubble Sort algorithm processes each vector by repeatedly comparing adjacent elements and swapping them if they are out of order. The total swap count characterizes the vector's disorder level. Across all vectors, the distribution of swap counts reflects the overall signal complexity.

Signals with consistent temporal structure produce similar swap counts across vectors, yielding concentrated distributions and low entropy. Irregular or chaotic signals generate diverse swap patterns, resulting in broader distributions and higher entropy values [9].

A key advantage of BubbEn is its minimal parameter dependency. Only the embedding dimension m and time delay τ need to be specified—no amplitude thresholds or class numbers are required. This simplicity makes BubbEn particularly practical for automated analysis across diverse signal types without extensive parameter optimization.

In this study, BubbEn was computed with $m \in \{2, 3, \ldots, 30\}$ and $\tau = 1$, yielding 29 features per signal.

Chapter 3

Databases Description

3.1 Introduction

3.2 Fantasia Database

3.3 MIT-BIH Arrhythmia Database

3.4 Congestive Heart Failure Database

3.5 Sudden Cardiac Death Holter Database

3.6 Gait in Parkinson's Disease Database

3.7 EMG Isometric Contractions Database

3.8 APNEA HRV+SpO2 Database

3.9 GAMEEMO Database

3.10 EEG During Mental Arithmetic Tasks Database

3.11 VOICED Database

3.12 Pulse Transit Time PPG Dataset

3.1 Introduction

Entropy, as already explained, is a concept that can be applied for signal analysis of signals of different origins. Studies have shown that it can be useful for assessing chaotic patterns, particularly in biomedical signals, though its applicability extends beyond this domain. It is indeed an analysis of the disorder, unpredictability, and uncertainty of a signal. Some signals can be more suited than others to entropy analysis,

showing significant differences between values from different subgroups of subjects. For example, HRV signals have shown good performance in entropy analysis, highlighting pathological conditions and the progression of diseases. Many other signals are less amenable to this complexity analysis, yielding insufficient discrimination between signals from different subject subgroups.

In this work, different types of biomedical signals were taken into consideration, in order to evaluate the entropy performance on all of them. The databases included were:

- **Fantasia Database.** HRV recordings from young and elderly healthy subjects for age-related complexity analysis.
- **MIT-BIH Arrhythmia.** ECG recordings from subjects with cardiac arrhythmias and normal sinus rhythm.
- **CHF vs NSR.** HRV data from subjects with Congestive Heart Failure and healthy controls with normal heart activity.
- **SDDB.** ECG recordings for Sudden Cardiac Death prediction, comparing high-risk patients with healthy subjects.
- **Gait in Parkinson's Disease.** Vertical ground reaction force recordings from Parkinson's disease patients and healthy controls.
- **EMG.** Electromyography recordings from healthy subjects and patients with neuromuscular disorders.
- **Apnea.** HRV and SpO2 signals from subjects with Obstructive Sleep Apnea and healthy controls.
- **GAMEEMO.** EEG recordings during gameplay for emotion recognition (negative vs positive valence).
- **Mental Arithmetic.** EEG recordings during rest and mental arithmetic tasks for cognitive state classification.
- **VOICED.** Voice recordings from healthy subjects and patients with vocal pathologies.

- **PPG.** Photoplethysmography recordings during rest and physical activity for activity state classification.

Table 3.1 provides an overview of all datasets, including signal type, number of subjects or recordings, class labels, and sampling frequency.

Table 3.1: Summary of the eleven biomedical datasets used in this study.

Dataset	Signal	N	Classes	Distribution	Fs (Hz)
Fantasia	RR intervals	40 subj.	Young vs Elderly	20 / 20	250
MIT-BIH Arrhythmia	RR intervals	120 rec.	ARR vs NSR	48 / 72	360
CHF vs NSR	RR intervals	116 rec.	CHF vs NSR	44 / 72	–
SCD vs Healthy	RR intervals	59 rec.	SCD vs Healthy	23 / 36	250/128
Gait (Parkinson's)	vGRF	166 subj.	PD vs Healthy	93 / 73	100
EMG	iEMG	241 subj.	Healthy / Myop. / Neurop.	50 / 98 / 93	32 768
Apnea (HRV)	RR intervals	83 subj.	Ctrl / Desat. / Non-d.	38 / 34 / 11	200
Apnea (SpO_2)	SpO_2	83 subj.	Ctrl / Desat. / Non-d.	38 / 34 / 11	50
GAMEEMO	EEG (14 ch.)	28 subj.	Negative vs Positive	56 / 56	128
Mental Arithmetic	EEG (23 ch.)	36 subj.	Rest vs Task	36 / 36	500
VOICED	Voice	208 subj.	Healthy vs Pathological	57 / 151	8 000
PPG	PPG (6 ch.)	22 subj.	Sitting vs Activity	22 / 44	500

The following sections provide detailed descriptions of each database, including their acquisition protocols, subject demographics, and preprocessing methods applied.

3.2 Fantasia Database

The Fantasia Database [10, 11] is an open-access dataset provided by PhysioNet that contains continuous electrocardiogram (ECG) recordings from 40 healthy subjects for studying the effects of aging on heart rate variability.

3.2.1 Data Description

Heart rate variability (HRV) reflects the beat-to-beat fluctuations in heart rate governed by the autonomic nervous system. With advancing age, the cardiovascular system undergoes significant changes, including reduced arterial compliance, decreased baroreflex sensitivity, and altered autonomic balance. These physiological changes manifest as a decline in HRV complexity, with elderly individuals typically exhibiting more regular and less adaptive cardiac rhythms compared to younger adults [10]. This age-related loss of complexity has been associated with reduced physiological reserve and increased cardiovascular risk.

The recordings were acquired while participants watched the movie Fantasia (Disney, 1940) in a supine position, a protocol designed to maintain relaxed wakefulness while remaining in sinus rhythm throughout the session. The dataset contains two balanced age groups: 20 young adults (21–34 years, mean age 25 years) and 20 elderly subjects (68–85 years, mean age 75 years). Each group has 10 male and 10 female subjects. The ECG signals were recorded at a sampling rate of 250 Hz with 12-bit resolution over approximately 120 minutes per subject.

From the raw ECG recordings, the RR intervals were extracted using the annotated R-peak locations provided in the database, yielding around 7,000 heartbeats per subject. The classification task aims to distinguish Young from Elderly subjects based on HRV complexity.

Table 3.2: Fantasia Database summary

Characteristic	**Value**
Total subjects	40
Young group	20 (21–34 years, 10M/10F)
Elderly group	20 (68–85 years, 10M/10F)
Recording duration	$\sim$120 minutes
Sampling rate	250 Hz
Resolution	12-bit
Signal analyzed	RR intervals

3.3 MIT-BIH Arrhythmia Database

This dataset combines recordings from three PhysioNet databases for arrhythmia versus normal sinus rhythm classification:

- MIT-BIH Arrhythmia Database (MITDB)
- MIT-BIH Normal Sinus Rhythm Database (NSRDB)
- Normal Sinus Rhythm RR Interval Database (NSR2DB)

3.3.1 Data Description

Cardiac arrhythmias are abnormalities in the heart's electrical conduction system that result in irregular heart rhythms. These can range from benign conditions, such as occasional premature beats, to life-threatening disorders like ventricular fibrillation. Common arrhythmias include premature ventricular contractions (PVCs), atrial fibrillation (AF), and bundle branch blocks. The ability to automatically detect and classify arrhythmias from ECG signals is clinically important for early diagnosis and treatment planning.

The MIT-BIH Arrhythmia Database [12, 11] contains 48 half-hour excerpts of two-channel ambulatory ECG recordings from 47 subjects (25 men aged 32–89 years, 22 women aged 23–89 years). The recordings were digitized at 360 Hz with 11-bit resolution over a 10 mV range. The database includes a variety of clinically significant arrhythmias, including premature ventricular contractions (PVCs), atrial premature contractions (APCs), left and right bundle branch blocks (LBBB, RBBB), and atrial fibrillation. All beat annotations were independently verified by two or more cardiologists.

3.3.2 Normal Sinus Rhythm Databases

Two additional PhysioNet databases provide normal sinus rhythm recordings as control samples:

- **NSRDB** (MIT-BIH Normal Sinus Rhythm Database) [11]: Contains 18 long-term ($\sim$24 hours) two-channel ECG recordings from subjects with no significant arrhythmias (5 men aged 26–45, 13 women aged 20–50).

- **NSR2DB** (Normal Sinus Rhythm RR Interval Database) [11]: Contains beat annotation files for 54 long-term ECG recordings (30 men aged 28.5–76, 24 women aged 58–73). The original ECG recordings are not available; only RR intervals are provided.

Table 3.3: MIT-BIH Arrhythmia and Normal Sinus Rhythm databases summary

Database	**Records**	**Duration**	**Signal**
MIT-BIH Arrhythmia	48	30 min	Raw ECG (360 Hz)
NSRDB	18	∼24 h	Raw ECG
NSR2DB	54	∼24 h	RR intervals only

3.4 Congestive Heart Failure Database

This dataset combines recordings from four PhysioNet databases to distinguish between congestive heart failure (CHF) and normal sinus rhythm (NSR) subjects:

- BIDMC Congestive Heart Failure Database (CHFDB)
- Congestive Heart Failure RR Interval Database (CHF2DB)
- MIT-BIH Normal Sinus Rhythm Database (NSRDB)
- Normal Sinus Rhythm RR Interval Database (NSR2DB)

3.4.1 Data Description

Congestive heart failure (CHF) is a chronic progressive condition in which the heart muscle is unable to pump sufficient blood to meet the body's needs. This inefficiency leads to fluid buildup in the lungs, liver, and extremities, causing symptoms such as shortness of breath, fatigue, and edema. CHF is classified by the New York Heart Association (NYHA) into four functional classes based on symptom severity, ranging from Class I (no limitation) to Class IV (severe limitation, symptoms at rest). Heart rate variability analysis has emerged as a valuable tool for CHF assessment, as patients typically exhibit reduced HRV complexity reflecting impaired autonomic regulation and diminished cardiac adaptability.

3.4.2 Congestive Heart Failure Databases

- **CHFDB** (BIDMC Congestive Heart Failure Database) [13, 11]: Contains long-term ECG recordings from 15 subjects (11 men aged 22–71, 4 women aged 54–63) with severe congestive heart failure (NYHA class 3–4).
- **CHF2DB** (Congestive Heart Failure RR Interval Database) [11]: Contains beat annotation files for 29 long-term ECG recordings from subjects aged 34–79 with congestive heart failure (NYHA classes I, II, and III). The original ECG recordings are not available; only RR intervals are provided.

3.4.3 Normal Sinus Rhythm Databases

The normal class uses the same two databases described in Subsection 3.3.2: NSRDB (18 recordings) and NSR2DB (54 recordings), totaling 72 normal subjects.

Table 3.4: CHF vs NSR databases summary

Database	**Records**	**Class**	**Signal**
CHFDB	15	CHF	Raw ECG
CHF2DB	29	CHF	RR intervals only
NSRDB	18	NSR	Raw ECG
NSR2DB	54	NSR	RR intervals only
Total	**116**	44 CHF / 72 NSR	

3.5 Sudden Cardiac Death Holter Database

This dataset combines recordings from two PhysioNet databases to distinguish between subjects who experienced sudden cardiac death (SCD) and healthy controls:

- Sudden Cardiac Death Holter Database (SDDB)
- MIT-BIH Normal Sinus Rhythm Database (NSRDB)

3.5.1 Data Description

Sudden cardiac death (SCD) is an unexpected death caused by loss of heart function, typically occurring within one hour of symptom onset. It is most commonly triggered by ventricular fibrillation (VF), a chaotic electrical activity that prevents the heart from pumping blood effectively. SCD accounts for a significant proportion of cardiovascular mortality worldwide, making early risk stratification a critical clinical goal. Analysis of heart rate variability and ECG dynamics in the hours preceding VF onset has shown promise for identifying high-risk individuals, as subtle changes in cardiac complexity may precede the fatal arrhythmia.

3.5.2 Sudden Cardiac Death Holter Database

The Sudden Cardiac Death Holter Database [14, 11] is a collection of 23 long-term Holter recordings from patients who experienced sudden cardiac death during the recordings. The recordings range from 7 to 25 hours in duration and were acquired from subjects aged 18 to 89 years. Most recordings include two ECG signals digitized at 250 Hz with 12-bit resolution. The database includes both unaudited and audited beat annotation files, along with clinical information such as cardiac history, medications, and ventricular fibrillation (VF) onset times.

For each SCD recording, a segment of up to 3 hours ending at the VF onset was extracted, with a minimum requirement of 1 hour of data.

3.5.3 Normal Sinus Rhythm Database

The healthy control class uses the MIT-BIH Normal Sinus Rhythm Database (NSRDB) described in Subsection 3.3.2. Two non-overlapping 3-hour segments were extracted from each of the 18 recordings, yielding 36 healthy samples.

Table 3.5: Sudden Cardiac Death vs Healthy dataset summary

Database	**Samples**	**Class**	**Duration**	**Sampling Rate**
SDDB	23	SCD	7–25 hours	250 Hz
NSRDB	36	Healthy	∼24 hours	128 Hz
Total	**59**	23 SCD / 36 Healthy		

3.6 Gait in Parkinson's Disease Database

The Gait in Parkinson's Disease Database [11] is a collection of multichannel recordings from force sensors beneath the feet of 93 patients with Parkinson's disease and 73 healthy control subjects, which were collected from three studies.

3.6.1 Data Description

Parkinson's disease (PD) is a chronic and progressive neurological disorder, one of the most common movement disorders, characterized by tremor, rigidity, bradykinesia, and postural instability. A disturbed gait is a common, debilitating symptom that may lead to falls and loss of functional independence.

The database contains vertical ground reaction force records of subjects as they walked at their usual, self-selected pace for approximately 2 minutes on level ground. Each foot was instrumented with 8 sensors (Ultraflex Computer Dyno Graphy, Infotronic Inc.) that measure force in Newtons as a function of time. The output of each of the 16 sensors was digitized and recorded at 100 samples per second. Additionally, two signals reflecting the sum of the 8 sensor outputs for each foot (Total_Left and Total_Right) are provided.

The database also includes demographic information and measures of disease severity using the Hoehn & Yahr staging and the Unified Parkinson's Disease Rating Scale (UPDRS).

Table 3.6: Gait in Parkinson's Disease database summary

Characteristic	**Value**
PD patients	93 (mean age: 66.3 years, 63% male)
Healthy controls	73 (mean age: 66.3 years, 55% male)
Total subjects	166
Recording duration	$\sim$2 minutes
Sensors per foot	8
Sampling rate	100 Hz
Signal analyzed	Total force (Left + Right)

3.7 EMG Isometric Contractions Database

The EMG Isometric Contractions Database [15] contains electromyographic recordings during isometric contractions of the biceps brachii and deltoid muscles, collected from patients and healthy subjects.

3.7.1 Data Description

Electromyography (EMG) is a valuable tool for exploring both the nervous and muscular system. Myopathy refers to conditions in which muscle fibers are dysfunctional, while neuropathy is a neurological disorder involving nerve degeneration, which can result in muscle weakness, cramps, stiffness, and wasting. Needle electromyography, known as intramuscular EMG (iEMG), is mainly used in clinical practice to diagnose patients with neuromuscular disorders.

This dataset was collected from 241 participants at the Neurology Department of Mustapha Pacha University Hospital of Algiers, between October 2015 and April 2019. The data includes three classes of raw invasive EMG signals recorded from two muscles (biceps brachii and deltoid) during isometric contraction:

- **Healthy**: 50 subjects (23 females, 27 males), aged 17–56 years (mean: 27.9 ± 9.5)
- **Myopathy**: 98 subjects (55 females, 43 males), aged 4–78 years (mean: 41.6 ± 16.8)
- **Neuropathy**: 93 subjects (53 females, 40 males), aged 2–81 years (mean: 46.5 ± 16.3)

The data acquisition system includes the MATRIX LIGHT 2-channel EMG EP Headbox with Integrated Electric Simulator (Micromed SPA), using disposable concentric needle electrodes (MYOLINE brand). EMG signals were acquired with amplitude in millivolts within a frequency band of 16–5000 Hz at a sampling frequency of 32,768 Hz.

Table 3.7: EMG Isometric Contractions database summary

Characteristic	Value
Total participants	241
Healthy subjects	50
Myopathy patients	98
Neuropathy patients	93
Muscles recorded	Deltoid, Biceps Brachii
Sampling rate	32,768 Hz
Frequency band	16–5000 Hz
Recording duration	∼5 seconds

For the binary classification task in this study, the Myopathy and Neuropathy classes were merged into a single "Disorder" class, resulting in a two-class problem: Healthy (50 subjects) versus Neuromuscular Disorder (191 subjects).

3.8 APNEA HRV+SpO2 Database

The APNEA HRV+SpO2 Database [16] was provided by Dr. Negrín from University Hospital (Las Palmas de Gran Canaria, Spain) and contains recordings of 83 subjects for sleep apnea classification.

3.8.1 Data Description

Obstructive Sleep Apnea (OSA) is a common sleep disorder characterized by repeated episodes of upper airway obstruction during sleep, leading to intermittent hypoxia and sleep fragmentation. OSA is associated with cardiovascular disease, hypertension, and daytime sleepiness.

The database contains ECG signals digitized at 200 Hz and SpO2 signals digitized at 50 Hz. The sleep studies and labelling process were carried out following the American Academy of Sleep Medicine (AASM) guidelines. An expert labeled every minute based on simultaneous polysomnography, marking it as apnea or non-apnea.

The subjects are divided into three groups based on the Apnea-Hypopnea Index

(AHI):

- **Control**: 38 healthy subjects with AHI < 5 (range: 0–5)
- **Desaturating OSA**: 34 patients with AHI > 25 (range: 30–106.3) who show desaturations during apneic episodes
- **Non-desaturating OSA**: 11 patients with AHI > 25 (range: 26.2–87.5) who do not always show desaturations during apneic episodes

Table 3.8: APNEA HRV+SpO2 database summary

Characteristic	**Value**
Total subjects	83
Control group	38 (AHI < 5)
Desaturating OSA	34 (AHI 30–106)
Non-desaturating OSA	11 (AHI 26–88)
ECG sampling rate	200 Hz
SpO2 sampling rate	50 Hz
Labelling	AASM guidelines

For the binary classification task in this study, the Desaturating and Non-desaturating OSA groups were merged into a single "Apnea" class, resulting in a two-class problem: Control (38 subjects) versus Apnea (45 subjects).

3.9 GAMEEMO Database

The GAMEEMO Database [17] is an electroencephalography-based dataset for emotion recognition analysis, collected using computer games as emotion elicitation stimuli.

3.9.1 Data Description

Emotion recognition from physiological signals has gained significant attention in affective computing and human-computer interaction. EEG-based emotion recognition offers a direct measure of brain activity associated with emotional states.

EEG signals were collected from 28 subjects using a wearable and portable 14-channel EMOTIV EPOC+ device at a sampling rate of 128 Hz. Subjects played 4 different computer games designed to elicit specific emotions based on Russell's Circumplex Model of Affect:

- **G1 - Boring**: Low arousal, negative valence
- **G2 - Calm**: Low arousal, positive valence
- **G3 - Horror**: High arousal, negative valence
- **G4 - Funny**: High arousal, positive valence

For binary classification, the four emotions were grouped by valence into Negative (Boring + Horror) and Positive (Calm + Funny) classes.

Each game session lasted approximately 5 minutes, resulting in 20 minutes of EEG data per subject. Subjects self-reported their experienced arousal and valence for each game using the Self-Assessment Manikin (SAM) questionnaire. Both raw and preprocessed EEG data are provided in CSV and MAT formats.

Table 3.9: GAMEEMO database summary

Characteristic	**Value**
Total subjects	28
EEG device	EMOTIV EPOC+ (14 channels)
Sampling rate	128 Hz
Games (emotions)	4 (Boring, Calm, Horror, Funny)
Session duration	$\sim$5 minutes per game
Total sessions	112 (28 subjects $\times$ 4 games)
Emotion model	Russell's Circumplex (Arousal-Valence)
Binary classes	Negative (G1+G3) / Positive (G2+G4)

3.10 EEG During Mental Arithmetic Tasks Database

The EEG During Mental Arithmetic Tasks Database [18, 11] contains EEG recordings of subjects before and during the performance of mental arithmetic tasks.

3.10.1 Data Description

Mental arithmetic tasks are commonly used to study cognitive load and working memory. During such tasks, distinct patterns of brain activity can be observed in EEG signals, particularly in frontal and parietal regions associated with executive function and numerical processing.

EEG signals were recorded monopolarly using a Neurocom EEG 23-channel system (Ukraine, XAI-MEDICA). Silver/silver chloride electrodes were placed on the scalp according to the International 10/20 scheme, with all electrodes referenced to the interconnected ear reference electrodes. A high-pass filter with a 30 Hz cut-off and a power line notch filter (50 Hz) were applied. All recordings are artifact-free EEG segments of 60 seconds duration, with Independent Component Analysis (ICA) used to eliminate artifacts from eyes, muscle activity, and cardiac pulsation.

The arithmetic task was serial subtraction of two numbers. Each trial started with the oral communication of a 4-digit number and a 2-digit number to subtract (e.g., $3141 - 42$). Participants were eligible if they had normal visual acuity, no cognitive impairment, and no psychiatric or neurological complaints.

Each subject has two recordings:

- **Background (_1)**: Resting EEG before mental arithmetic task
- **Task (_2)**: EEG during mental arithmetic performance

Subjects were divided into two groups based on performance:

- **Group G (Good)**: 24 subjects (mean operations per 4 min = 21, SD = 7.4)
- **Group B (Bad)**: 12 subjects (mean operations per 4 min = 7, SD = 3.6)

Table 3.10: EEG During Mental Arithmetic Tasks database summary

Characteristic	Value
Total subjects	36
Good performers	24
Bad performers	12
EEG system	Neurocom 23-channel
Sampling rate	500 Hz
Recording duration	60 seconds
Conditions	2 (Rest, Task)
Total recordings	72 (36 subjects $\times$ 2)

3.11 VOICED Database

The VOICED (VOice ICar fEderico II) Database [19, 11] contains clinically-verified voice samples from 208 subjects, including 150 pathological and 58 healthy voices.

3.11.1 Data Description

Voice disorders affect a significant portion of the population and can severely impact quality of life and occupational performance. Acoustic analysis of voice signals provides a non-invasive method for detecting and monitoring vocal pathologies.

Subjects involved in the study were adults aged between 18 and 70 years. Individuals under 18 or over 70 years of age, or those with conditions such as colds or upper respiratory tract infections, were excluded. Healthy voices and the presence of vocal fold disorders were clinically verified by medical experts. Medical phoniatric examinations were performed at the ambulatories of Phoniatrics and Videolaryngoscopy of the University Hospital of Naples "Federico II" and at the Institute of High Performance Computing and Networking (ICAR-CNR). The study was conducted from May 16, 2016, to May 15, 2017.

Voice recordings were made using the Vox4Health m-health system installed on a Samsung Galaxy S4 smartphone. The microphone was held approximately 20 cm from the patient, angled at 45 degrees toward the mouth. All recordings were sampled

at 8000 Hz with 32-bit resolution and filtered to remove accidental noise. Participants were instructed to produce a sustained vocalization of the vowel /a/ with constant intensity for approximately 5 seconds. Recordings were made in a quiet room ($<$ 30 dB background noise).

The database includes information such as gender, age, pathology, lifestyle habits (smoking, alcohol, coffee consumption), and the results of two medical questionnaires: the Voice Handicap Index (VHI) and Reflux Symptom Index (RSI).

Table 3.11: VOICED database summary

Characteristic	**Value**
Total subjects	208
Healthy voices	58
Pathological voices	150
Age range	18–70 years
Recording device	Mobile phone (Vox4Health)
Sampling rate	8000 Hz
Resolution	32-bit
Task	Sustained vowel /a/
Duration	$\sim$5 seconds

3.12 Pulse Transit Time PPG Dataset

The Pulse Transit Time PPG Dataset [20, 11] provides open-access, high-resolution, time-synchronized recordings from multiple sensors worn at different body locations, including photoplethysmogram (PPG), inertial sensors, pressure sensors, and ECG.

3.12.1 Data Description

Photoplethysmography (PPG) is a non-invasive optical technique used to detect volumetric changes in blood in peripheral circulation. Pulse Transit Time (PTT), the time required for pulse waves to travel in blood vessels between two sites, is an important indicator for many medical properties, with recent research focusing on its potential for continuous blood pressure monitoring.

Data were acquired from 22 healthy subjects at The University of Sydney (6 females, 16 males), with ages ranging from 20 to 53 years (mean: 28.12 years). All participants performed 3 activities in random order: sitting, optional walking, and running. The data were collected using a device similar to commercial pulse oximeters, containing commercial sensors (MAX30101) in a 3D-printed finger clip.

The dataset contains multi-site and multi-wavelength PPG signals recorded at 500 Hz from 6 channels at two finger locations:

- **Distal phalanx (fingertip)**:
  - Red (660nm)
  - Infrared (880nm)
  - Green (537nm)
- **Proximal phalanx (finger base)**:
  - Red (660nm)
  - Infrared (880nm)
  - Green (537nm)

For binary classification, the three activities were grouped into Sitting (rest) and Activity (walk + run).

Table 3.12: Pulse Transit Time PPG dataset summary

Characteristic	**Value**
Total subjects	22 (6F, 16M)
Age range	20–53 years (mean: 28.1)
Activities	3 (Sit, Walk, Run)
Total recordings	66 (22 subjects $\times$ 3)
PPG channels	6
Sampling rate	500 Hz
Recording duration	$\sim$508 seconds
Binary classes	Sitting (22) / Activity (44)

Chapter 4

Methodology

4.1 Signal Preprocessing

4.2 Feature Extraction

4.3 Feature Selection

4.4 Classification

4.5 Cross-Validation Strategy

4.6 Evaluation Metrics

This chapter presents the experimental methodology employed to evaluate entropy measures across the eleven biomedical datasets described in Chapter 3. The methodology encompasses signal preprocessing, entropy-based feature extraction, feature selection strategies, classification algorithms, cross-validation protocols, and performance evaluation metrics. All experiments were implemented in Python 3.9 using established scientific computing libraries.

The experimental pipeline follows a systematic approach: raw biomedical signals undergo preprocessing to remove artifacts and normalize amplitude ranges, entropy features are extracted using multiple parameter configurations, optimal feature subsets are identified through exhaustive search, and classification performance is evaluated using stratified cross-validation with multiple machine learning algorithms.

4.1 Signal Preprocessing

Biomedical signals acquired from physiological measurements contain various artifacts and noise components that can adversely affect entropy estimation. Preprocessing aims to enhance signal quality while preserving the intrinsic complexity characteristics that entropy measures seek to quantify. The preprocessing pipeline consists of three main stages: filtering, detrending, and normalization.

4.1.1 Filtering

Different signal types require specific filtering approaches based on their frequency content and noise characteristics:

- **ECG/HRV signals**: RR interval series extracted from ECG recordings undergo physiological filtering to remove ectopic beats and artifacts. Intervals outside the range of 0.3–2.0 seconds (corresponding to heart rates of 30–200 bpm) are excluded as physiologically implausible.
- **EMG signals**: Electromyographic recordings are processed with a low-pass filter at 3 kHz to remove high-frequency noise, a 50 Hz notch filter to eliminate power line interference, and a high-pass filter at 15 Hz to remove motion artifacts. The RMS envelope is then extracted to capture muscle activation patterns.
- **EEG signals**: Electroencephalographic data are downsampled from the original sampling rate (e.g., 500 Hz to 128 Hz) to reduce computational burden while retaining relevant frequency content below the Nyquist limit.
- **PPG signals**: Photoplethysmographic recordings are bandpass filtered between 0.5–10 Hz to isolate the cardiac pulse component and remove baseline drift and high-frequency noise, then downsampled to 100 Hz.
- **Voice signals**: Sustained vowel recordings are trimmed to extract a stable 3-second segment from the center of the phonation, avoiding onset and offset transients.

4.1.2 Detrending

Low-frequency trends unrelated to the physiological process of interest can bias entropy estimates. Linear detrending is applied to remove slow baseline drifts:

$$x_{\text{detrended}}(t) = x(t) - (\alpha + \beta t) \tag{4.1}$$

where α and β are the intercept and slope of the least-squares linear fit to the original signal $x(t)$.

4.1.3 Normalization

To ensure consistent entropy estimation across signals with different amplitude scales, mean removal is applied:

$$x_{\text{normalized}}(t) = x_{\text{detrended}}(t) - \bar{x} \tag{4.2}$$

where $\bar{x}$ is the mean of the detrended signal. For entropy measures that use amplitude-dependent thresholds (such as Approximate and Sample Entropy), the tolerance parameter r is scaled by the signal's standard deviation, making the analysis amplitude-independent.

4.2 Feature Extraction

Following preprocessing, entropy-based features are extracted from each signal using the eight entropy measures described in Chapter 2. To capture complexity at multiple scales and ensure robust characterization, each entropy measure is computed with various parameter configurations, resulting in a comprehensive feature set.

4.2.1 Entropy Parameters

The parameter configurations for each entropy measure were selected based on recommendations from the literature and preliminary experiments. Table 4.1 summarizes the parameters used in this study.

Table 4.1: Entropy parameters and resulting feature counts.

Entropy Measure	**Parameters**	**Features**
Shannon Entropy	Histogram binning (auto)	1
Renyi Entropy	$\alpha = 2$ (collision entropy)	1
Approximate Entropy	$m \in \{2, 3, 4\}$, $r \in \{0.15, 0.20, 0.25\} \times \sigma$	9
Sample Entropy	$m \in \{2, 3, 4\}$, $r \in \{0.15, 0.20, 0.25\} \times \sigma$	9
Permutation Entropy	$m \in \{2, 3, \ldots, 10\}$, $\tau = 1$	9
Dispersion Entropy	$m \in \{2, 3, 4, 5\}$, $c \in \{3, 4, \ldots, 9\}$	28
Distribution Entropy	$m \in \{2, 3, 4, 5\}$, $c \in \{3, 4, \ldots, 8\}$	24
Bubble Entropy	$m \in \{2, 3, \ldots, 30\}$	29
Total		**110**

The embedding dimension m controls the pattern length considered by each entropy measure. Higher values of m capture longer temporal dependencies but require more data points for reliable estimation. The tolerance parameter r for Approximate and Sample Entropy is expressed as a fraction of the signal's standard deviation σ, ensuring scale-invariant analysis. The number of classes c in Dispersion and Distribution Entropy determines the granularity of the symbolic representation.

4.2.2 Implementation

Entropy calculations were performed using two specialized Python libraries:

- **EntropyHub** [21]: A comprehensive entropy library providing validated implementations of Approximate Entropy, Sample Entropy, Dispersion Entropy, Distribution Entropy, and Bubble Entropy. EntropyHub follows standardized algorithmic definitions and has been validated against reference implementations.
- **Antropy** [22]: An efficient library optimized for Permutation Entropy computation, offering significant speed improvements through vectorized operations.

Shannon and Renyi entropies were computed using histogram-based probability estimation with automatic bin selection (Freedman-Diaconis rule), implemented via NumPy and SciPy functions.

4.2.3 Feature Naming Convention

Each extracted feature follows a systematic naming convention that encodes the entropy type and parameter values:

- `shannon`, `renyi`: Single-valued features
- `apen_m{m}_r{r}`: Approximate Entropy with embedding dimension m and tolerance r
- `sampen_m{m}_r{r}`: Sample Entropy with embedding dimension m and tolerance r
- `permen_m{m}`: Permutation Entropy with embedding dimension m
- `dispen_m{m}_c{c}`: Dispersion Entropy with embedding dimension m and c classes
- `disten_m{m}_c{c}`: Distribution Entropy with embedding dimension m and c classes
- `bubben_m{m}`: Bubble Entropy with embedding dimension m

This convention facilitates automated grouping of features by entropy family during analysis and interpretation of results.

4.3 Feature Selection

With 110 entropy features extracted per signal, feature selection is essential to identify the most discriminative subset while avoiding overfitting and reducing computational complexity. This study employs an exhaustive search strategy that systematically evaluates all possible feature combinations to find the optimal subset for each entropy family.

4.3.1 Exhaustive Search Strategy

Unlike filter-based or greedy wrapper methods that may converge to local optima, exhaustive search guarantees finding the globally optimal feature subset within a

specified size constraint. For each entropy family, all combinations of k features are evaluated, where k ranges from 1 to $k_{\max} = 5$:

$$\text{Combinations} = \sum_{k=1}^{k_{\max}} \binom{n}{k} \tag{4.3}$$

where n is the number of features in the entropy family.

The exhaustive search procedure operates as follows. First, features are standardized using z-score normalization to ensure equal scaling. Then, for each value of k from 1 to 5, every possible combination of k features is generated. Each combination is evaluated using 5-fold stratified cross-validation with a Support Vector Machine (SVM) classifier with radial basis function kernel as the reference model. The mean cross-validation accuracy serves as the evaluation criterion. After testing all combinations across all values of k, the feature subset achieving the highest accuracy is selected as optimal for that entropy family.

This approach eliminates the suboptimality inherent in heuristic search methods and provides reproducible, globally optimal feature subsets.

4.3.2 Correlation-Based Pre-filtering

To reduce redundancy before exhaustive search, highly correlated features are removed. For each pair of features with Pearson correlation exceeding a threshold $\rho_{\max} = 0.95$, the feature with lower variance is discarded:

$$\text{Remove } f_i \text{ if } |r(f_i, f_j)| > \rho_{\max} \text{ and } \text{Var}(f_i) < \text{Var}(f_j) \tag{4.4}$$

This pre-filtering step removes near-duplicate features while preserving the most informative variant, reducing the combinatorial search space without sacrificing discriminative information.

4.3.3 Per-Entropy Family Evaluation

Feature selection is performed independently for each of the eight entropy families. This approach enables:

1. Direct comparison of entropy families under optimal feature selection
2. Identification of the best-performing parameters within each family

3. Fair evaluation where each entropy type uses its own optimal subset

4.4 Classification

To evaluate the discriminative power of entropy features across different learning paradigms, five machine learning classifiers were employed. These classifiers represent diverse algorithmic approaches, ranging from instance-based methods to ensemble techniques, ensuring comprehensive evaluation of entropy feature utility.

4.4.1 Classifier Descriptions

k-Nearest Neighbors (kNN)

The k-Nearest Neighbors algorithm classifies samples based on the majority class among the k closest training instances in feature space. Euclidean distance is used as the similarity metric. The optimal value of k is determined through grid search over the range $k \in \{3, 5, 7, \ldots, 29\}$ using cross-validation, selecting the value that maximizes classification accuracy. Feature scaling via standardization is applied prior to classification to ensure equal contribution from all features.

Gaussian Naive Bayes (GNB)

Gaussian Naive Bayes assumes that features follow a Gaussian distribution within each class and applies Bayes' theorem with the naive assumption of feature independence. Despite this simplifying assumption, GNB often performs well in practice and provides a probabilistic baseline. The classifier estimates class-conditional means and variances from training data and computes posterior probabilities for classification.

Random Forest (RF)

Random Forest is an ensemble method that constructs multiple decision trees during training, each built on a bootstrap sample of the data with random feature subsets at each split. The final prediction is determined by majority voting across all trees. This study uses 100 trees with default hyperparameters. Random Forest is inherently robust to overfitting and provides feature importance estimates, though these are not used for feature selection in this study.

Support Vector Machine (SVM)

Support Vector Machine seeks the optimal hyperplane that maximizes the margin between classes in feature space. A radial basis function (RBF) kernel is employed to handle non-linear class boundaries:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp\left(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2\right) \tag{4.5}$$

where γ controls the kernel width. Default hyperparameters are used with probability estimates enabled for ROC-AUC computation. Feature standardization is applied to ensure proper kernel function behavior.

Gradient Boosting (GB)

Gradient Boosting builds an additive ensemble of weak learners (decision trees) sequentially, where each tree corrects the errors of the previous ensemble. The algorithm minimizes a differentiable loss function through gradient descent in function space. This study uses 100 estimators with default learning rate and tree depth parameters.

4.4.2 Preprocessing Pipeline

All classifiers except Random Forest and Gradient Boosting require feature standardization for optimal performance. A preprocessing pipeline applies z-score normalization:

$$x'_i = \frac{x_i - \mu}{\sigma} \tag{4.6}$$

where μ and σ are the mean and standard deviation computed from training data only. This standardization is performed within each cross-validation fold to prevent data leakage.

4.4.3 Implementation

All classifiers were implemented using scikit-learn [23], a widely-used Python machine learning library. Table 4.2 summarizes the classifier configurations.

Table 4.2: Classifier configurations used in this study.

Classifier	Key Parameters	Scaling
kNN	$k \in \{3, 5, \ldots, 29\}$ (grid search)	Yes
GNB	Default (Gaussian likelihood)	Yes
RF	100 trees, default depth	No
SVM	RBF kernel, default C and γ	Yes
GB	100 estimators, default learning rate	No

4.5 Cross-Validation Strategy

Cross-validation is employed to obtain reliable performance estimates while maximizing the use of available data for both training and evaluation. This study uses 5-fold cross-validation with two distinct splitting strategies depending on the dataset structure.

4.5.1 Stratified K-Fold Cross-Validation

For datasets where samples are independent observations, Stratified K-Fold cross-validation is applied. The data are partitioned into 5 equal folds, with each fold maintaining the same class proportion as the complete dataset. Each fold serves as the test set exactly once, while the remaining four folds form the training set. The final performance metrics are computed by aggregating predictions across all folds.

Stratified splitting is essential for imbalanced datasets, ensuring that each fold contains representative samples from both classes. This prevents the scenario where a fold might contain only majority class samples, which would lead to biased performance estimates.

Datasets using Stratified K-Fold include: FANTASIA, ECG Arrhythmia, ECG CHF vs NSR, ECG SCD vs Healthy, EMG (Biceps and Deltoid), APNEA (HRV and SpO2), Gait Parkinson's, and VOICED.

4.5.2 Group K-Fold Cross-Validation

For datasets containing multiple recordings per subject, Group K-Fold cross-validation is employed to prevent data leakage. This strategy ensures that all samples from a given subject appear exclusively in either the training or test set, never in both.

Subject-wise splitting is critical because samples from the same individual share subject-specific characteristics (physiological baselines, recording conditions) that classifiers can exploit. Without proper grouping, models may learn to recognize subjects rather than the underlying pathological or physiological patterns, leading to overly optimistic performance estimates that fail to generalize to new individuals.

$$\text{Group K-Fold constraint: } \forall s \in \text{Subjects},\ s \in \text{Train} \oplus s \in \text{Test} \tag{4.7}$$

Datasets using Group K-Fold include: GAMEEMO (28 subjects × 4 game sessions), Mental Arithmetic (36 subjects × 2 conditions), and PPG (22 subjects × 3 activity conditions).

4.5.3 Validation Protocol

For both cross-validation strategies, the following protocol is applied:

1. Data is split into 5 folds (stratified by class or grouped by subject)
2. For each fold iteration:
   (a) Feature scaling parameters are computed from training data only
   (b) The classifier is trained on the scaled training set
   (c) Predictions are generated for the held-out test fold
3. Predictions from all folds are aggregated for metric computation
4. Accuracy standard deviation across folds quantifies estimate stability

The random state is fixed (seed = 42) to ensure reproducibility across experiments. This allows direct comparison of entropy families under identical data partitions.

4.6 Evaluation Metrics

Classification performance is assessed using multiple complementary metrics derived from the confusion matrix. This multi-metric approach provides a comprehensive view of classifier behavior, capturing different aspects of predictive performance that a single metric might obscure.

4.6.1 Confusion Matrix

The confusion matrix forms the foundation for all evaluation metrics, summarizing predictions into four categories for binary classification:

Table 4.3: Confusion matrix for binary classification.

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

4.6.2 Primary Metrics

Accuracy

Accuracy measures the proportion of correct predictions among all samples:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{4.8}$$

While intuitive and widely used, accuracy can be misleading for imbalanced datasets where high accuracy may result simply from predicting the majority class.

Sensitivity (Recall)

Sensitivity, also known as recall or true positive rate, measures the proportion of actual positive cases correctly identified:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{4.9}$$

In clinical applications, high sensitivity is crucial to minimize missed diagnoses (false negatives), particularly for serious conditions.

Specificity

Specificity measures the proportion of actual negative cases correctly identified:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{4.10}$$

High specificity reduces false alarms (false positives), which is important to avoid unnecessary interventions or patient anxiety.

F1-Score

The F1-score is the harmonic mean of precision and recall, providing a balanced measure for imbalanced datasets:

$$\text{F1} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{4.11}$$

Weighted F1-score is used for multi-class scenarios, weighting each class by its support.

4.6.3 Advanced Metrics

Cohen's Kappa

Cohen's Kappa (κ) measures agreement between predicted and actual labels, corrected for chance agreement:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \tag{4.12}$$

where p_o is the observed agreement (accuracy) and p_e is the expected agreement by chance. Kappa values are interpreted as: $\kappa < 0.20$ (poor), 0.21–0.40 (fair), 0.41–0.60 (moderate), 0.61–0.80 (substantial), and 0.81–1.00 (near perfect).

4.6.4 Reporting Protocol

For each entropy family and classifier combination, the following metrics are reported:

- Accuracy with standard deviation across cross-validation folds
- F1-score (weighted for multi-class)

- Sensitivity and Specificity
- Cohen's Kappa
- Confusion matrix values (TP, TN, FP, FN)

The best-performing classifier for each entropy family is identified based on accuracy, and the overall best entropy-classifier combination is highlighted for each dataset.

Chapter 5

Results

5.1 Overall Entropy Performance

5.2 Performance by Dataset

5.3 Class Separation Analysis

5.4 Classifier Performance

5.5 Detailed Dataset Results

5.6 Chapter Summary

This chapter presents the experimental results of evaluating eight entropy measures across eleven biomedical datasets. The evaluation encompasses overall performance ranking, domain-specific analysis, and classifier comparison. All results are based on the exhaustive feature selection methodology and 5-fold cross-validation protocol described in Chapter 4.

The datasets span diverse biomedical signal types, including cardiac (HRV, ECG), neural (EEG), muscular (EMG), respiratory (SpO2), gait, voice, and photoplethysmographic (PPG) signals. This diversity enables assessment of the entropy method's generalizability across different physiological domains and classification tasks.

Results are organized as follows: Section 5.1 presents the overall entropy performance aggregated across all datasets. Section 5.2 analyzes performance stratified by signal type. Section 5.3 examines class separation power using effect size analysis. Section 5.4 evaluates classifier performance across entropy features. Finally, Section 5.5

summarizes key findings and identifies the best-performing entropy measures for each application domain.

5.1 Overall Entropy Performance

This section presents the aggregated classification performance of each entropy measure across all datasets. Performance is evaluated using the best accuracy achieved by any classifier (kNN, GNB, RF, SVM, or GB) for the optimally selected feature subset within each entropy family.

5.1.1 Mean Accuracy Ranking

Table 5.1 presents the mean classification accuracy for each entropy measure, computed across all 11 datasets. For datasets with multiple channels or signal variants, the best-performing channel is used as the representative result.

Table 5.1: Overall entropy performance ranking by mean accuracy.

Rank	Entropy	Mean Acc. (%)	Std Dev	Max Acc. (%)	Wins
1	Bubble	83.61	8.95	93.75	8
2	Dispersion	77.74	8.61	92.42	2
3	Approximate	77.43	8.52	92.42	0
4	Permutation	76.59	11.45	95.45	1
5	Sample	74.70	8.95	90.91	0
6	Distribution	73.42	9.86	93.94	0
7	Renyi	73.41	8.20	89.39	0
8	Shannon	72.89	8.38	92.42	0

Figure 5.1 visualizes the mean accuracy ranking with error bars representing standard deviation across datasets.

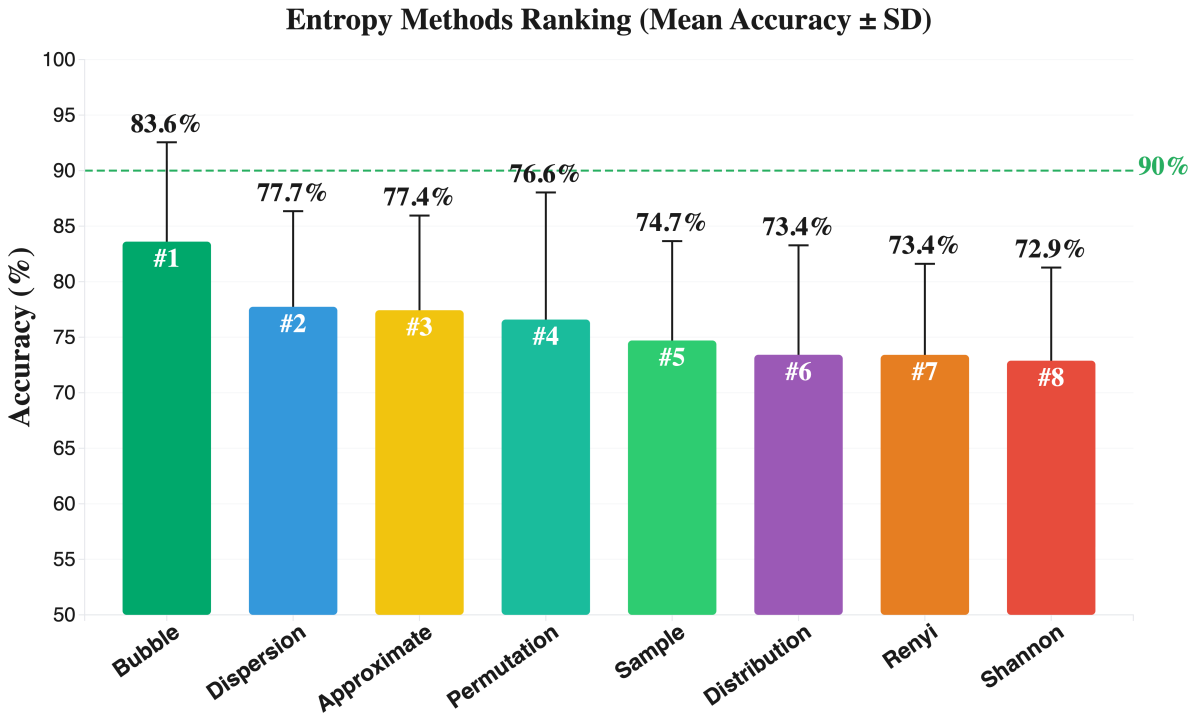

Figure 5.1: Mean classification accuracy for each entropy measure across all datasets. Error bars indicate standard deviation. Bubble Entropy achieves the highest mean accuracy (83.6%) and the lowest coefficient of variation, indicating the most consistent relative performance.

Bubble Entropy achieves the highest mean accuracy (83.6%) and, combined with a standard deviation of 9.0%, the lowest coefficient of variation (CV = 0.107) among all methods, indicating both superior accuracy and the most consistent performance across diverse signal types. The "Wins" column indicates the number of datasets where each entropy achieved the highest accuracy among all methods. Bubble Entropy dominates with 8 wins out of 11 datasets (72.7%), demonstrating its broad applicability.

5.1.2 Performance Gap Analysis

The results reveal a clear performance hierarchy among entropy measures:

- **Top tier**: Bubble Entropy stands alone with a mean accuracy approximately 6 percentage points above the second-best method.
- **Middle tier**: Dispersion Entropy (77.7%), Approximate Entropy (77.4%), Permutation Entropy (76.6%), and Sample Entropy (74.7%) form a competitive

second tier.

- **Lower tier**: Distribution, Renyi, and Shannon Entropy perform comparably, with mean accuracies around 73%.

The gap between Bubble Entropy and traditional measures (Shannon, Renyi) exceeds 10 percentage points (83.6% vs 72.9–73.4%), highlighting the advantage of complexity-based entropy measures that capture temporal dynamics rather than simple amplitude distributions.

5.1.3 Consistency Analysis

Standard deviation provides insight into method consistency across different classification tasks. In absolute terms, Renyi Entropy exhibits the lowest standard deviation (8.20%), while Permutation Entropy shows the highest (11.45%), suggesting task-dependent performance that may excel in some domains while underperforming in others. However, absolute standard deviation does not account for differences in mean accuracy, making the coefficient of variation a more appropriate measure of relative consistency.

The coefficient of variation (CV = Std/Mean) further emphasizes this pattern:

$$\text{CV}_{\text{Bubble}} = \frac{8.95}{83.61} = 0.107, \quad \text{CV}_{\text{Permutation}} = \frac{11.45}{76.59} = 0.149 \tag{5.1}$$

Lower CV values indicate more reliable performance across diverse applications, making Bubble Entropy the most dependable choice for general biomedical signal classification.

5.1.4 Pairwise Comparison

To further characterize relative performance, a head-to-head win/loss matrix was constructed. For each pair of entropy methods, the matrix counts the number of datasets where one method outperforms the other. A positive value indicates the row method wins against the column method by that many datasets.

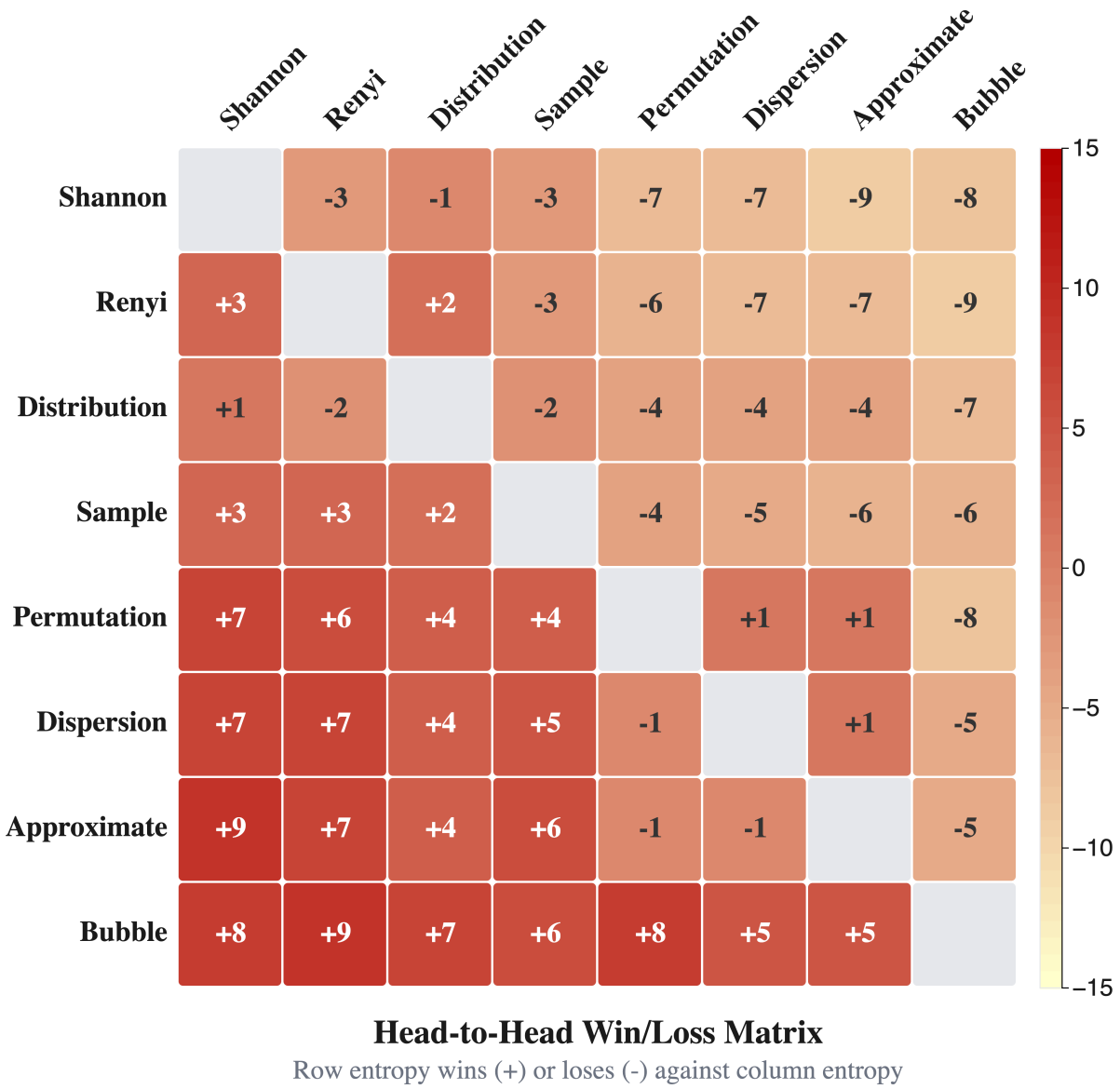

Figure 5.2: Head-to-head win/loss matrix. Each cell shows the number of datasets where the row entropy outperforms (+) or underperforms (-) the column entropy. Green indicates dominance, red indicates inferiority.

The win/loss matrix (Figure 5.2) reveals several important patterns:

- **Bubble Entropy dominance**: Bubble Entropy shows positive values against all other methods, winning by margins of +5 to +9 datasets. It achieves its largest advantages against Renyi (+9), Shannon (+8), and Permutation (+8).
- **Dispersion Entropy strength**: As the second-best performer, Dispersion Entropy wins against most methods except Bubble, with notable margins against Shannon (+7) and Renyi (+7).

- **Competitive middle tier**: Approximate, Permutation, and Sample Entropy show mixed results against each other, indicating comparable performance with task-dependent advantages.
- **Lower tier**: Distribution, Renyi, and Shannon Entropy show negative values against most other methods, confirming their position in the lower performance tier.

This pairwise analysis reinforces the ranking hierarchy while revealing the magnitude of performance differences between methods.

5.2 Performance by Dataset

This section presents classification results across the eleven datasets. For datasets with multiple channels or signal variants, results are consolidated by selecting the best-performing configuration. Detailed per-dataset analysis with comprehensive metrics is provided in Section 5.5.

Figure 5.3 presents the classification accuracy of all eight entropy measures across the eleven datasets, enabling direct comparison of entropy performance within each classification task.

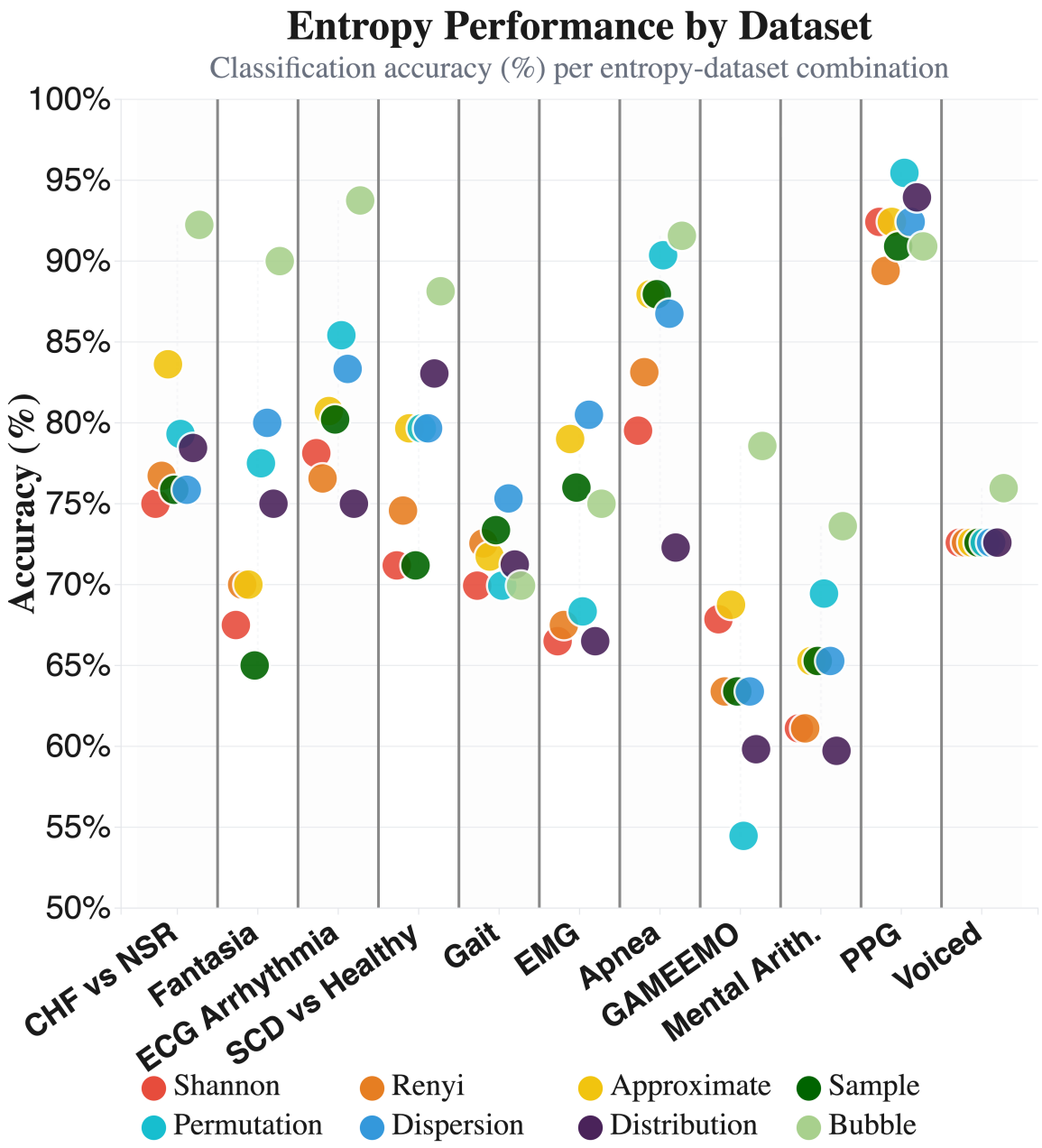

Figure 5.3: Classification accuracy by entropy measure across all datasets. Each bar represents the best accuracy achieved by any classifier for that entropy-dataset combination.

Figure 5.4 identifies the winning entropy method for each dataset, providing clear guidance for method selection.

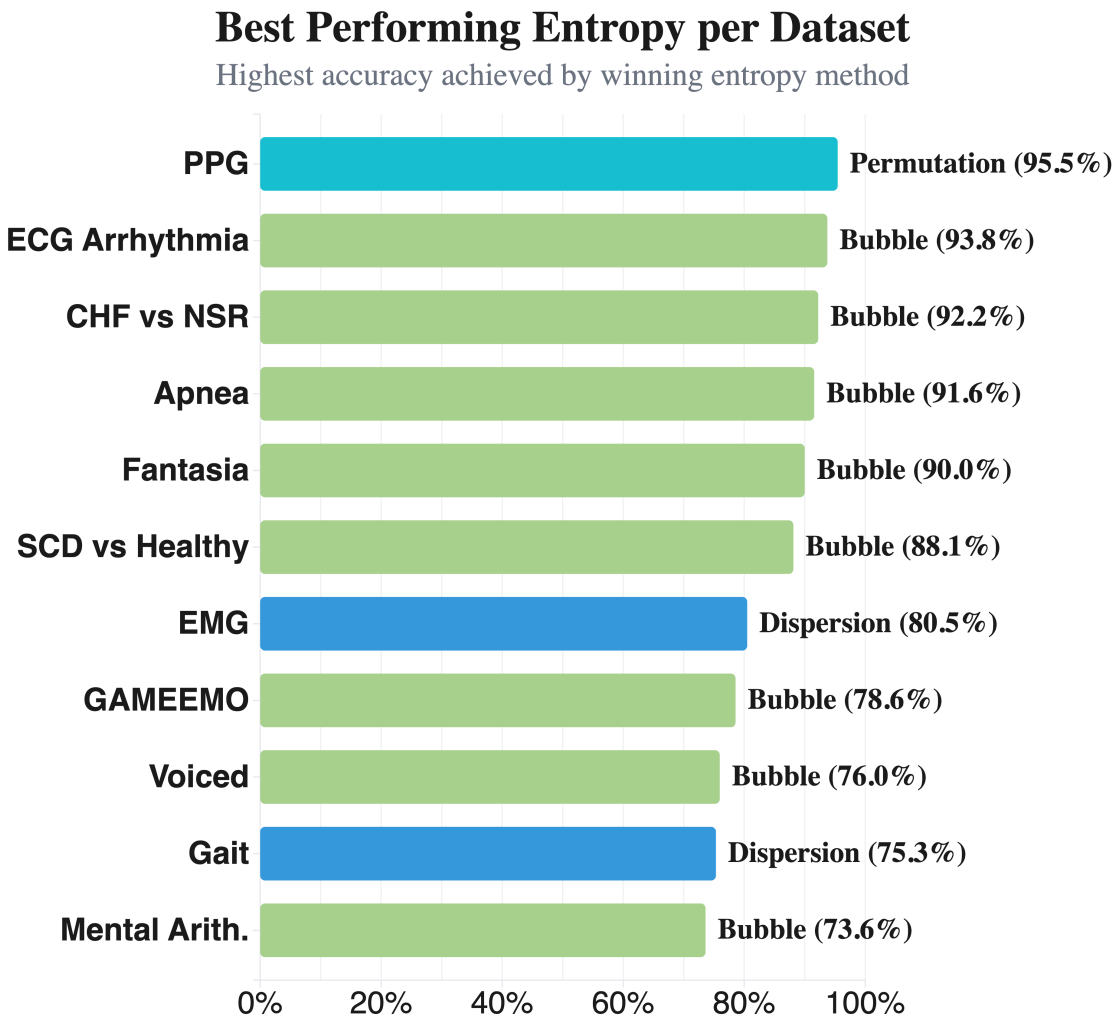

Figure 5.4: Best performing entropy method for each dataset with corresponding accuracy.

Table 5.2 summarizes the best entropy, classifier, and accuracy for each dataset.

Table 5.2: Best performing entropy and classifier for each dataset.

Dataset	Best Entropy	Best Classifier	Accuracy (%)
CHF vs NSR	Bubble	SVM	92.2
Fantasia	Bubble	SVM	90.0
ECG Arrhythmia	Bubble	SVM	93.8
SCD vs Healthy	Bubble	SVM	88.1
Gait (Parkinson)	Dispersion	SVM	75.3
EMG	Dispersion	SVM	80.5
Apnea	Bubble	SVM	91.6
GAMEEMO	Bubble	RF	78.6
Mental Arithmetic	Bubble	kNN	73.6
PPG	Permutation	kNN	95.5
Voiced	Bubble	SVM	76.0

5.2.1 Key Observations

Several patterns emerge from the dataset-wise comparison:

- **Bubble Entropy dominance**: Bubble Entropy achieves the best accuracy in 8 out of 11 datasets, confirming its broad applicability across diverse biomedical signals.
- **High-accuracy tasks**: PPG (95.5%), ECG Arrhythmia (93.8%), CHF vs NSR (92.2%), and Apnea (91.6%) exceed 90% accuracy, indicating that entropy features effectively capture discriminative patterns in these signals.
- **Challenging tasks**: GAMEEMO (78.6%), Voiced (76.0%), Gait (75.3%), and Mental Arithmetic (73.6%) show lower accuracies, reflecting the inherent difficulty of these classification tasks.
- **Motor signal preference for Dispersion**: Both Gait and EMG datasets favor Dispersion Entropy over Bubble Entropy, suggesting that symbolic dynamics approaches better capture motor control abnormalities.
- **PPG uniqueness**: PPG is the only dataset where Permutation Entropy outperforms all others, indicating that ordinal patterns are particularly discriminative

for activity classification.

- **Classifier preference**: SVM is the best classifier in 8 datasets, followed by kNN (2 datasets) and Random Forest (1 dataset).

5.3 Class Separation Analysis

Classification accuracy reflects overall performance but does not directly measure how well entropy features separate classes in feature space. This section examines class separation power using Cohen's d effect size, which quantifies the standardized difference between class distributions.

5.3.1 Cohen's d Effect Size

Cohen's d measures the separation between two distributions in terms of pooled standard deviations:

$$d = \frac{|\mu_1 - \mu_0|}{\sigma_{\text{pooled}}} \tag{5.2}$$

where μ_0 and μ_1 are the class means, and σ_{pooled} is the pooled standard deviation. Effect sizes are interpreted as: negligible ($d < 0.2$), small ($0.2 \leq d < 0.5$), medium ($0.5 \leq d < 0.8$), and large ($d \geq 0.8$).

For each entropy family, effect sizes were computed across all feature variants and datasets, then aggregated to characterize overall class separation capability.

5.3.2 Effect Size by Entropy Method

Table 5.3 presents the mean, standard deviation, and maximum Cohen's d values for each entropy measure across all datasets and parameter configurations.

Table 5.3: Cohen's d effect sizes by entropy method.

Entropy	**Mean |d|**	**Std Dev**	**Max |d|**	**Interpretation**
Bubble	1.172	0.609	2.220	Large
Approximate	1.075	0.746	3.090	Large
Sample	0.957	0.691	2.610	Large
Dispersion	0.955	0.641	2.230	Large
Permutation	0.867	0.623	2.620	Large
Renyi	0.856	0.735	2.740	Large
Distribution	0.815	0.556	2.280	Large
Shannon	0.744	0.703	2.630	Medium

Figure 5.5 visualizes the maximum Cohen's d effect size for each entropy-dataset combination, revealing where each method achieves strong class separation.

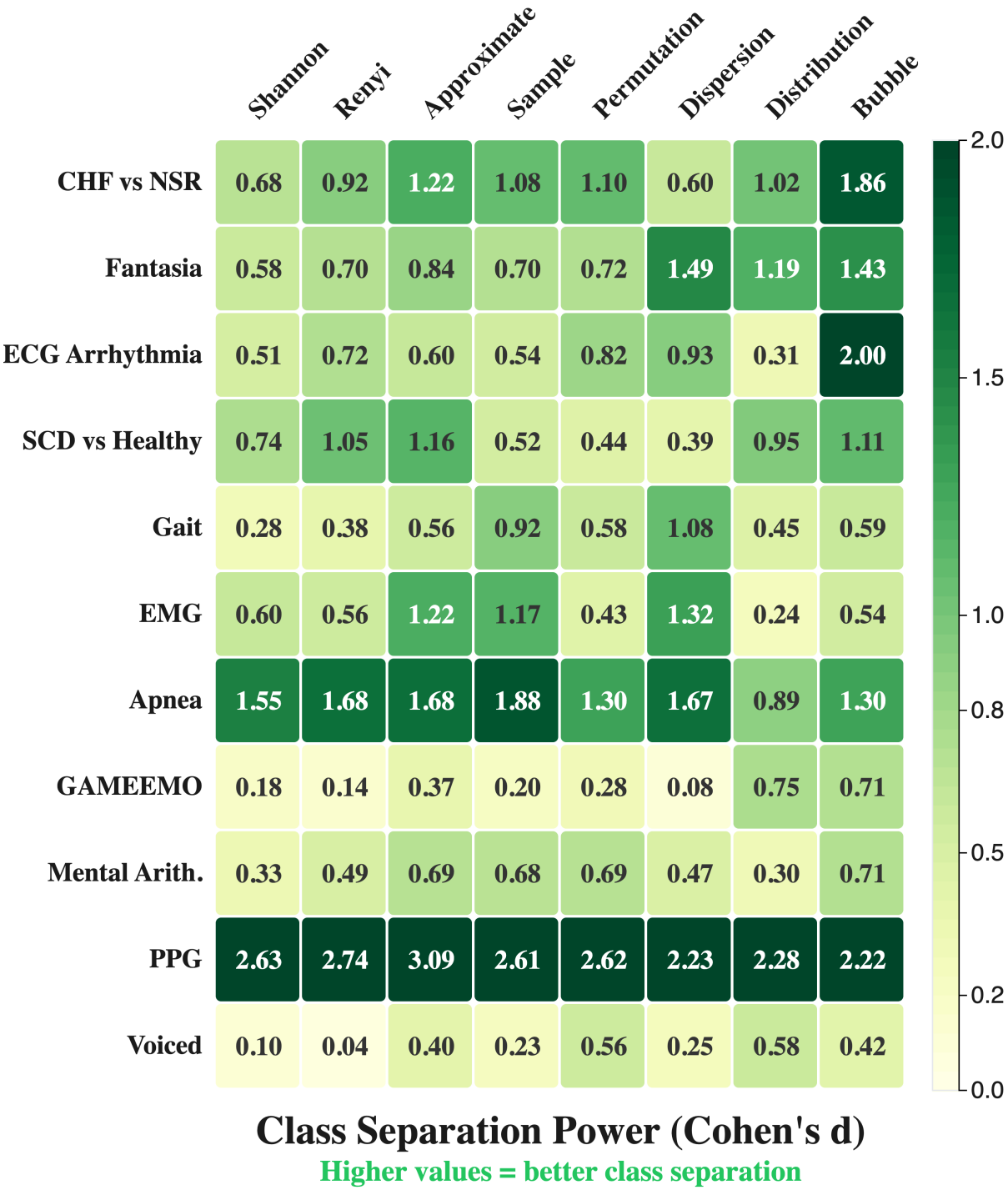

Figure 5.5: Cohen's d effect size heatmap across all entropy methods and datasets. Darker green indicates stronger class separation. Values above 0.8 represent large, clinically meaningful effect sizes.

5.3.3 Interpretation

The effect size analysis reveals several important findings:

- **Mean effect sizes**: Most entropy methods achieve mean effect sizes in the large range ($d \geq 0.8$), with values ranging from 0.744 (Shannon Entropy, the only method in the medium range) to 1.172 (Bubble Entropy). This indicates that the per-dataset maximum features generally achieve strong class separation.

- **Bubble Entropy dominance**: Bubble Entropy achieves the highest mean effect size (1.172) across all datasets, confirming that its features consistently produce the strongest class separation. Its maximum effect size (2.220) is among the highest, reflecting reliably large separation across diverse biomedical signal types.

- **Approximate Entropy maximum**: Approximate Entropy achieves the highest single maximum effect size (3.090), driven by exceptionally large class separation on the PPG dataset. However, its mean (1.075) is lower than Bubble Entropy's, indicating less consistent performance across datasets.

- **Shannon Entropy limitations**: Shannon Entropy exhibits the lowest mean effect size (0.744), the only method remaining in the medium range. This explains its consistently lower classification accuracy compared to complexity-based measures.

- **Dispersion Entropy consistency**: Dispersion Entropy shows a strong mean effect size (0.955) with moderate variability (std = 0.641), consistent with its reliable performance on motor signal datasets (EMG, Gait) where it outperforms all other methods.

5.3.4 Effect Size vs Classification Accuracy

The relationship between effect size and classification accuracy is not strictly linear. While large effect sizes generally correlate with better classification, other factors influence performance:

1. **Feature interactions**: Classifiers can exploit combinations of features with moderate individual effect sizes.

2. **Non-linear separability**: SVM with RBF kernel can achieve high accuracy even when linear separation (reflected in Cohen's d) is limited.

3. **Parameter optimization**: The exhaustive feature selection identifies optimal parameter combinations that maximize discriminative power.

Bubble Entropy's combination of the highest mean effect size, consistently large maximum effect sizes, and parameter-free computation explains its consistent top performance across diverse classification tasks.

5.4 Classifier Performance

While previous sections focused on entropy method comparison, classifier selection also impacts classification accuracy. This section analyzes the performance of five classifiers—k-Nearest Neighbors (kNN), Gaussian Naive Bayes (GNB), Random Forest (RF), Support Vector Machine (SVM), and Gradient Boosting (GB)—across entropy methods and datasets.

5.4.1 Overall Classifier Distribution

For each entropy-dataset combination, the exhaustive feature selection process identifies the best-performing classifier. Table 5.4 presents the distribution of winning classifiers across 88 entropy-dataset combinations (8 entropy methods $\times$ 11 datasets).

Table 5.4: Distribution of best-performing classifiers across all entropy-dataset combinations.

Classifier	Wins	Percentage
kNN	51	58.0%
SVM	20	22.7%
RF	10	11.4%
GNB	7	8.0%
GB	0	0.0%

The k-Nearest Neighbors classifier dominates, achieving the best performance in over half of all cases. SVM ranks second with approximately one-quarter of wins. Notably, Gradient Boosting never emerges as the best classifier for any entropy-dataset

combination, despite being a powerful ensemble method for many machine learning tasks.

5.4.2 Mean Accuracy by Classifier

Table 5.5 presents the mean accuracy achieved by each classifier across all entropy methods and datasets.

Table 5.5: Mean classification accuracy by classifier across all experiments.

Classifier	Mean Accuracy	Std Dev	Rank
kNN	75.06%	9.57%	1
SVM	74.36%	10.21%	2
GNB	71.10%	10.13%	3
RF	68.53%	12.77%	4
GB	68.22%	12.04%	5

The ranking by mean accuracy aligns with the win distribution: kNN and SVM lead, while ensemble methods (RF, GB) underperform. This suggests that entropy features create relatively simple decision boundaries that benefit from instance-based (kNN) or margin-maximizing (SVM) approaches rather than ensemble complexity.

5.4.3 Classifier Performance by Entropy Method

The interaction between entropy method and classifier choice reveals important patterns. Table 5.6 presents mean accuracy for each entropy-classifier combination.

Table 5.6: Mean accuracy (%) for each entropy-classifier combination.

Entropy	**kNN**	**GNB**	**RF**	**SVM**	**GB**
Shannon	72.0	69.3	64.4	68.7	65.0
Renyi	72.4	69.8	64.4	70.5	65.5
Approximate	76.6	72.9	66.6	75.8	65.5
Sample	74.1	71.3	66.6	74.0	66.8
Permutation	75.9	68.8	69.8	74.8	69.3
Dispersion	76.5	71.2	70.3	76.1	70.3
Distribution	72.4	68.7	67.7	72.6	67.3
Bubble	**80.5**	**76.8**	**78.4**	**82.4**	**76.1**

Figure 5.6 visualizes these entropy-classifier interactions, with darker cells indicating higher accuracy. The Bubble row stands out with consistently high values across all classifiers.

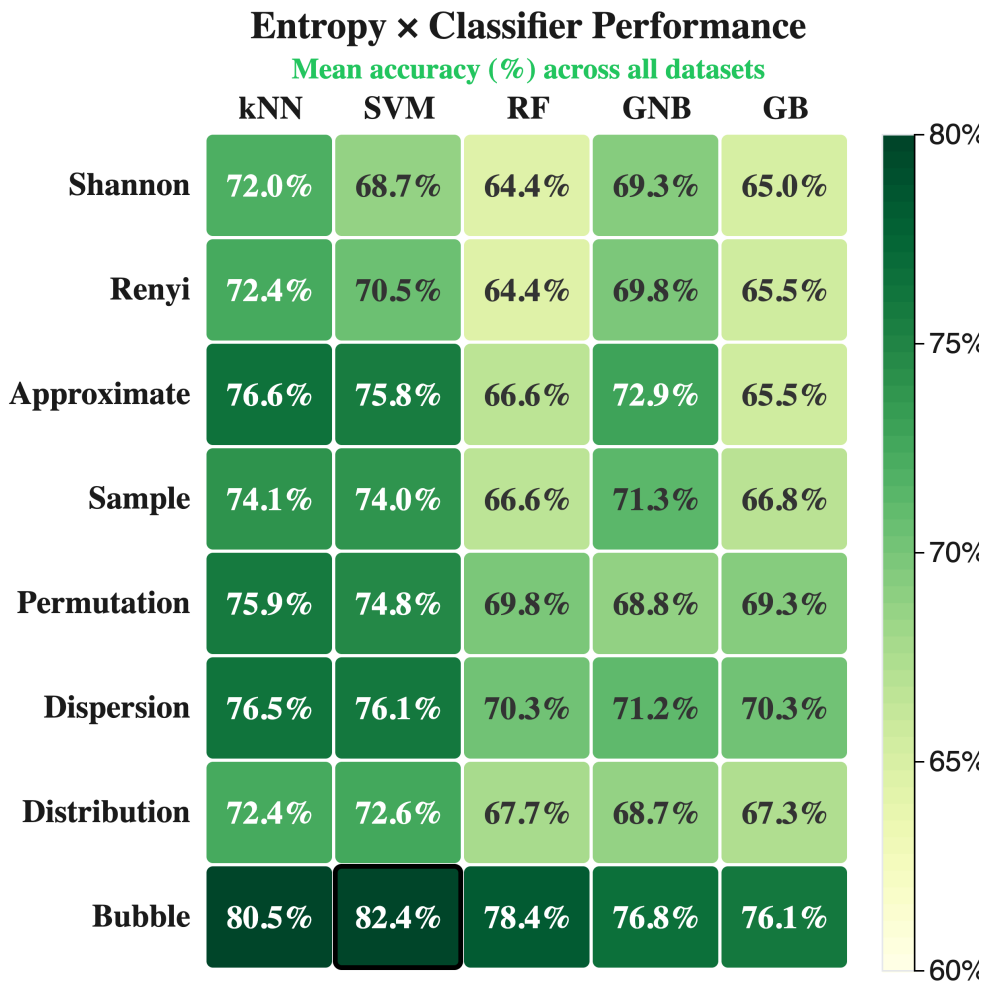

Figure 5.6: Entropy × Classifier performance heatmap showing mean accuracy (%) across all datasets. Darker green indicates higher accuracy. The Bubble + SVM combination achieves the highest accuracy (82.4%).

Key observations:

- **Bubble Entropy achieves highest accuracy with every classifier**, confirming its robustness to classifier choice.
- **SVM with Bubble Entropy** yields the single best combination (82.4%), followed closely by kNN with Bubble Entropy (80.5%).
- **Random Forest performance varies dramatically**: RF achieves only 64.4% with Renyi but 78.4% with Bubble—a 14.0 percentage point difference.
- **Traditional entropies (Shannon, Renyi) show similar classifier patterns**, with kNN slightly outperforming other classifiers.

5.4.4 Best Classifier by Entropy Method

Table 5.7 shows how many times each classifier wins for each entropy method across all 11 datasets.

Table 5.7: Number of dataset wins per classifier, grouped by entropy method.

Entropy	**kNN**	**GNB**	**RF**	**SVM**	**GB**
Shannon	7	1	2	1	0
Renyi	9	1	1	0	0
Approximate	8	1	0	2	0
Sample	7	1	0	3	0
Permutation	7	0	2	2	0
Dispersion	6	0	2	3	0
Distribution	6	2	1	2	0
Bubble	1	1	2	7	0
Total	**51**	**7**	**10**	**20**	**0**

A notable pattern emerges: **SVM becomes the dominant classifier specifically for Bubble Entropy**, winning 7 of 11 datasets (63.6%), compared to only 1 win for kNN (9.1%). This contrasts sharply with traditional entropies (Shannon, Renyi) where kNN wins 64–82% of datasets. This suggests that Bubble Entropy features benefit from SVM's ability to find optimal separating hyperplanes in high-dimensional feature spaces.

5.4.5 Top Entropy-Classifier Combinations

Table 5.8 presents the ten highest-performing entropy-classifier combinations by mean accuracy.

Table 5.8: Top 10 entropy-classifier combinations by mean accuracy.

Rank	Combination	Mean Accuracy
1	Bubble + SVM	82.36%
2	Bubble + kNN	80.46%
3	Bubble + RF	78.40%
4	Bubble + GNB	76.78%
5	Approximate + kNN	76.61%
6	Dispersion + kNN	76.46%
7	Dispersion + SVM	76.13%
8	Bubble + GB	76.06%
9	Permutation + kNN	75.92%
10	Approximate + SVM	75.76%

Bubble Entropy occupies the top four positions and five of the top eight, appearing with all five classifiers. This reinforces that Bubble Entropy's superior discriminative power transcends classifier selection, though SVM maximizes its potential.

5.4.6 Interpretation

The classifier analysis reveals several practical insights:

1. **Simple classifiers excel**: kNN and SVM outperform complex ensemble methods (RF, GB), suggesting that entropy features create well-separated clusters amenable to nearest-neighbor or hyperplane-based classification.

2. **Entropy-classifier interaction**: The optimal classifier varies by entropy method. Traditional entropies favor kNN, while Bubble Entropy benefits most from SVM, indicating that different entropy measures create different feature space geometries.

3. **Gradient Boosting ineffectiveness**: GB never achieves best performance, possibly because entropy features lack the complex, hierarchical structure that boosting methods exploit effectively.

4. **Practical recommendation**: For new biomedical classification tasks, the Bubble + SVM combination provides the highest expected accuracy, though Bubble + kNN offers a computationally simpler alternative with minimal accuracy loss.

5.5 Detailed Dataset Results

This section presents detailed classification results for each of the eleven datasets, providing insights into entropy method performance across diverse biomedical applications.

5.5.1 Summary of Best Results

Table 5.9 presents the best-performing entropy method, classifier, and accuracy for each dataset.

Table 5.9: Best classification results per dataset.

Dataset	Best Entropy	Classifier	Accuracy
PPG	Permutation	kNN	95.5%
ECG Arrhythmia	Bubble	SVM	93.8%
CHF vs NSR	Bubble	SVM	92.2%
Apnea	Bubble	SVM	91.6%
Fantasia	Bubble	SVM	90.0%
SCD vs Healthy	Bubble	SVM	88.1%
EMG	Dispersion	SVM	80.5%
GAMEEMO	Bubble	RF	78.6%
Voiced	Bubble	SVM	76.0%
Gait	Dispersion	SVM	75.3%
Mental Arithmetic	Bubble	kNN	73.6%

Key observations:

- **Bubble Entropy wins 8 of 11 datasets** (72.7%), confirming its overall dominance
- **Dispersion Entropy** excels in Gait and EMG datasets

- **Permutation Entropy** achieves the highest overall accuracy (95.5%) on PPG
- **SVM** is the best classifier for 8 of 11 datasets

5.5.2 CHF vs NSR

Classification between Congestive Heart Failure and Normal Sinus Rhythm patients.

Table 5.10: CHF vs NSR: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM	92.2%
2	Approximate + kNN	83.6%
3	Permutation + kNN	79.3%

Bubble Entropy achieves 92.2% accuracy, outperforming the second-best method by 8.6 percentage points. Figure 5.7 shows the complete entropy-classifier performance matrix.

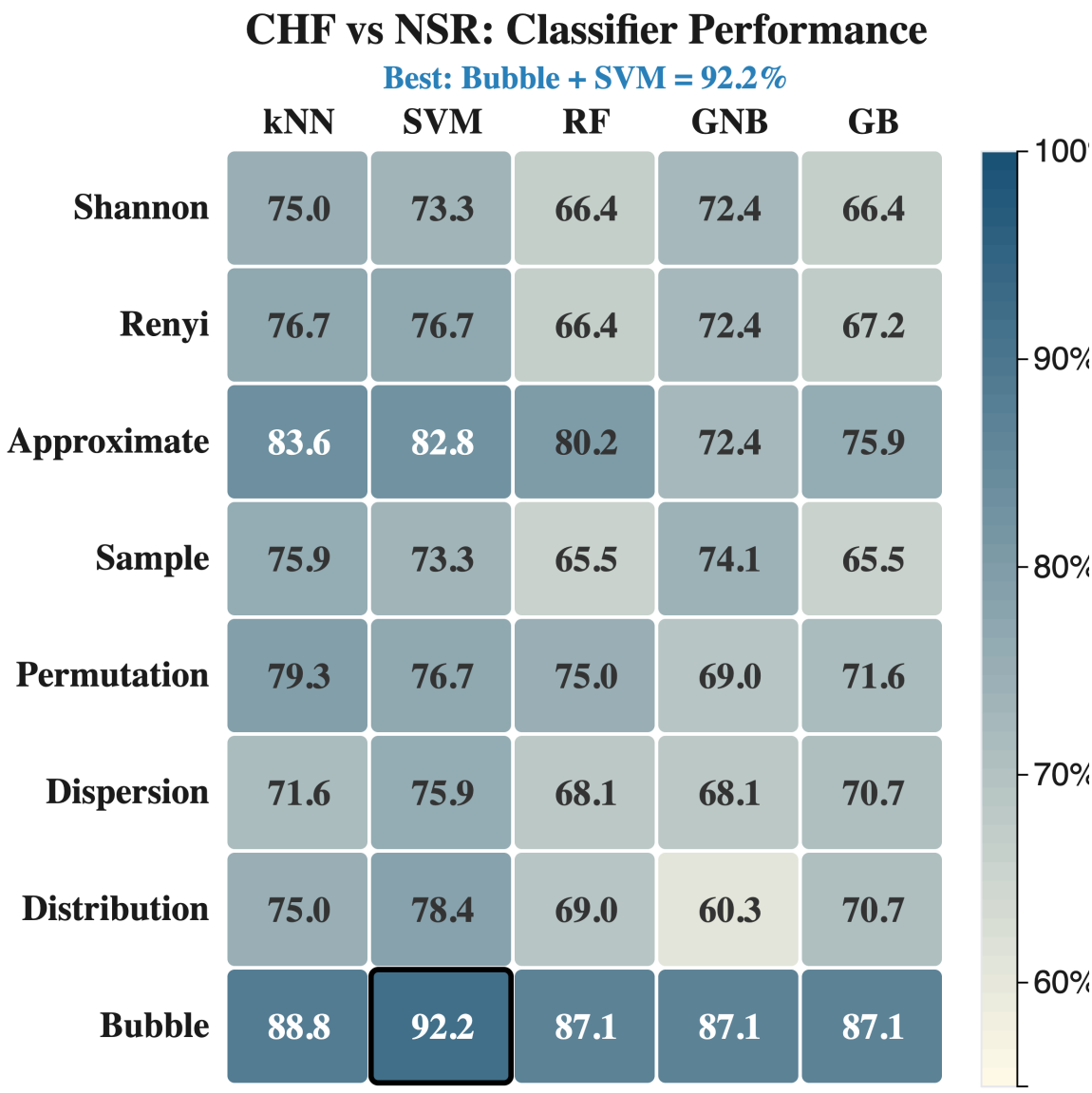

Figure 5.7: CHF vs NSR: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (92.2%), highlighted with a black border.

Figure 5.8 presents the Cohen's d effect sizes, explaining why Bubble Entropy outperforms other methods through superior class separation.

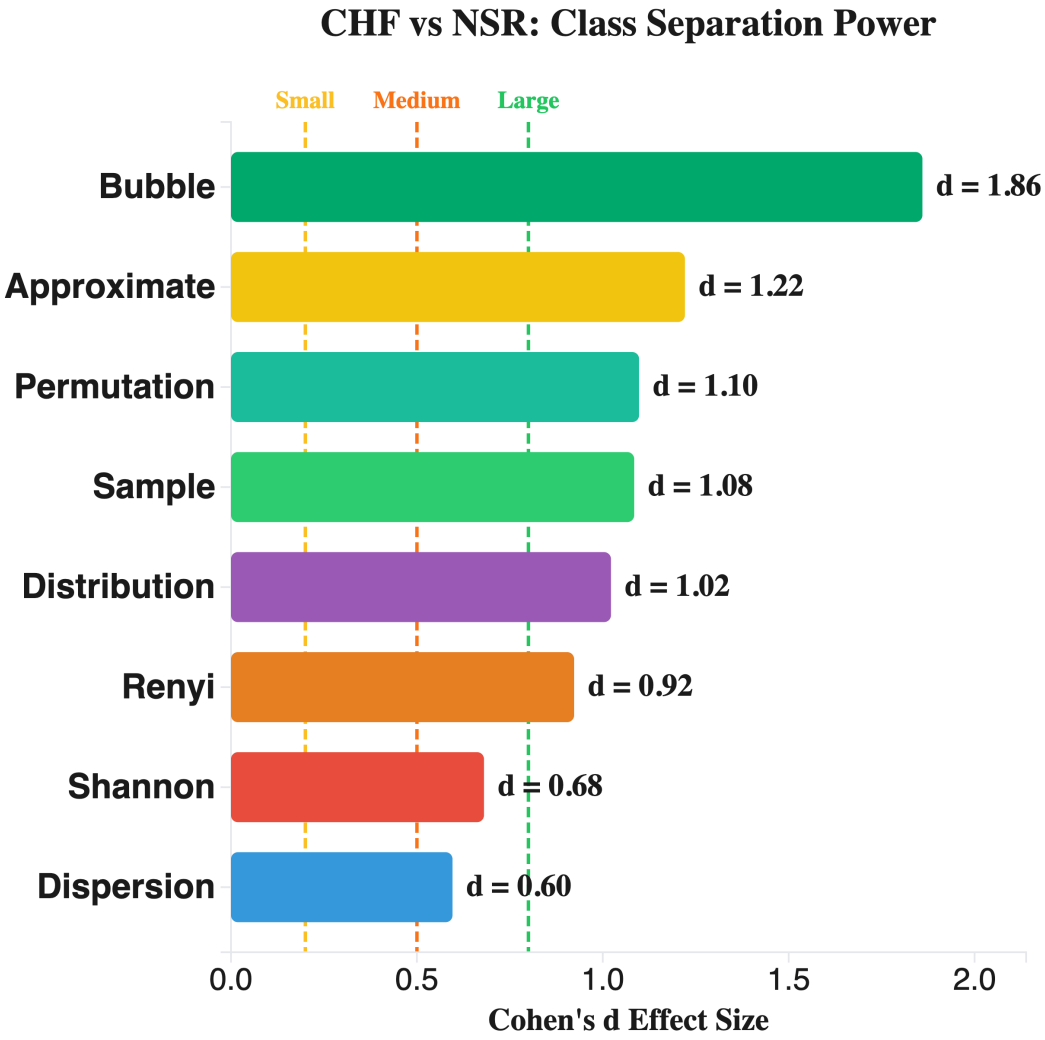

Figure 5.8: CHF vs NSR: Class separation power (Cohen's d). Bubble Entropy achieves the largest effect size, explaining its superior classification performance.

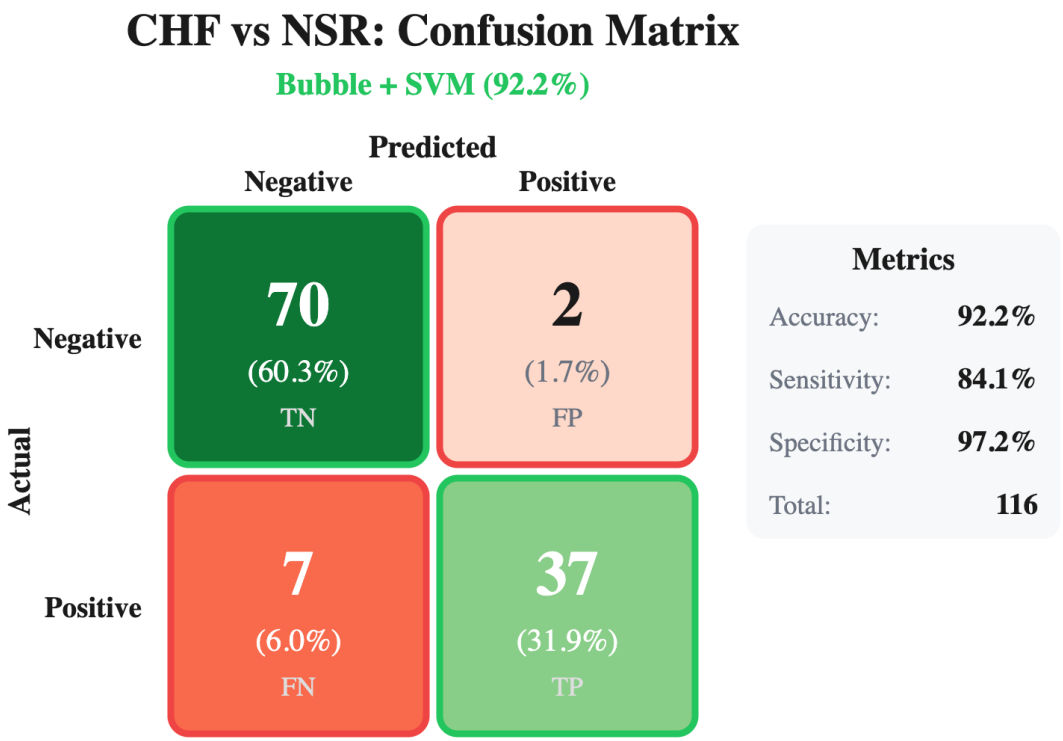

Figure 5.9: CHF vs NSR: Confusion matrix for Bubble + SVM. The model correctly identifies 70 NSR and 37 CHF patients, with only 9 misclassifications.

5.5.3 Fantasia

Distinguishing young (21–34 years) from elderly (68–85 years) subjects.

Table 5.11: Fantasia: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM	90.0%
2	Dispersion + kNN	80.0%
3	Permutation + kNN	77.5%

Bubble Entropy achieves 90% accuracy with a 10 percentage point advantage over Dispersion Entropy. Figure 5.10 shows the complete entropy-classifier performance matrix, and Figure 5.11 presents the class separation power.

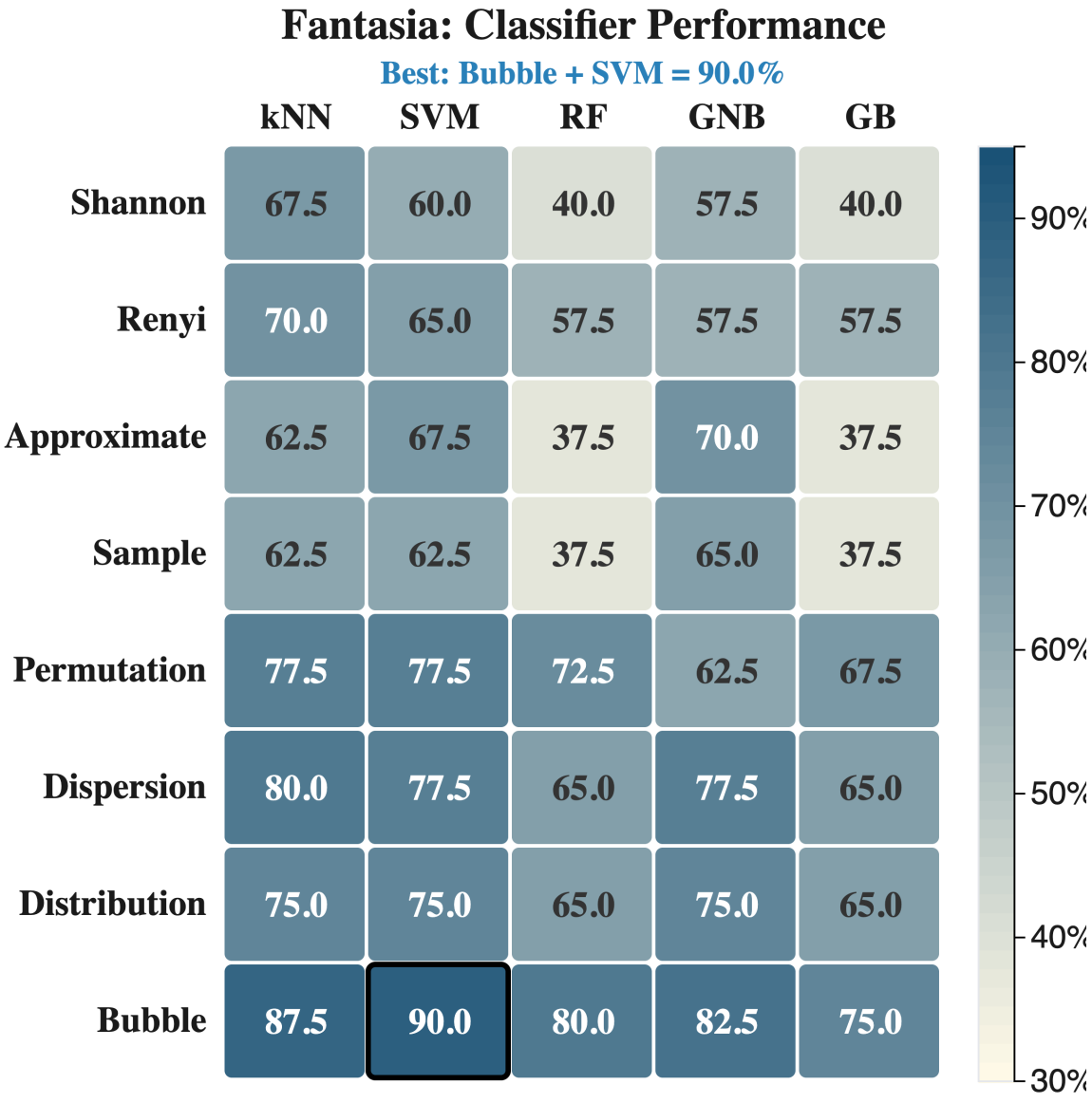

Figure 5.10: Fantasia: Entropy × Classifier performance heatmap for age classification. Bubble Entropy with SVM achieves the highest accuracy (90.0%).

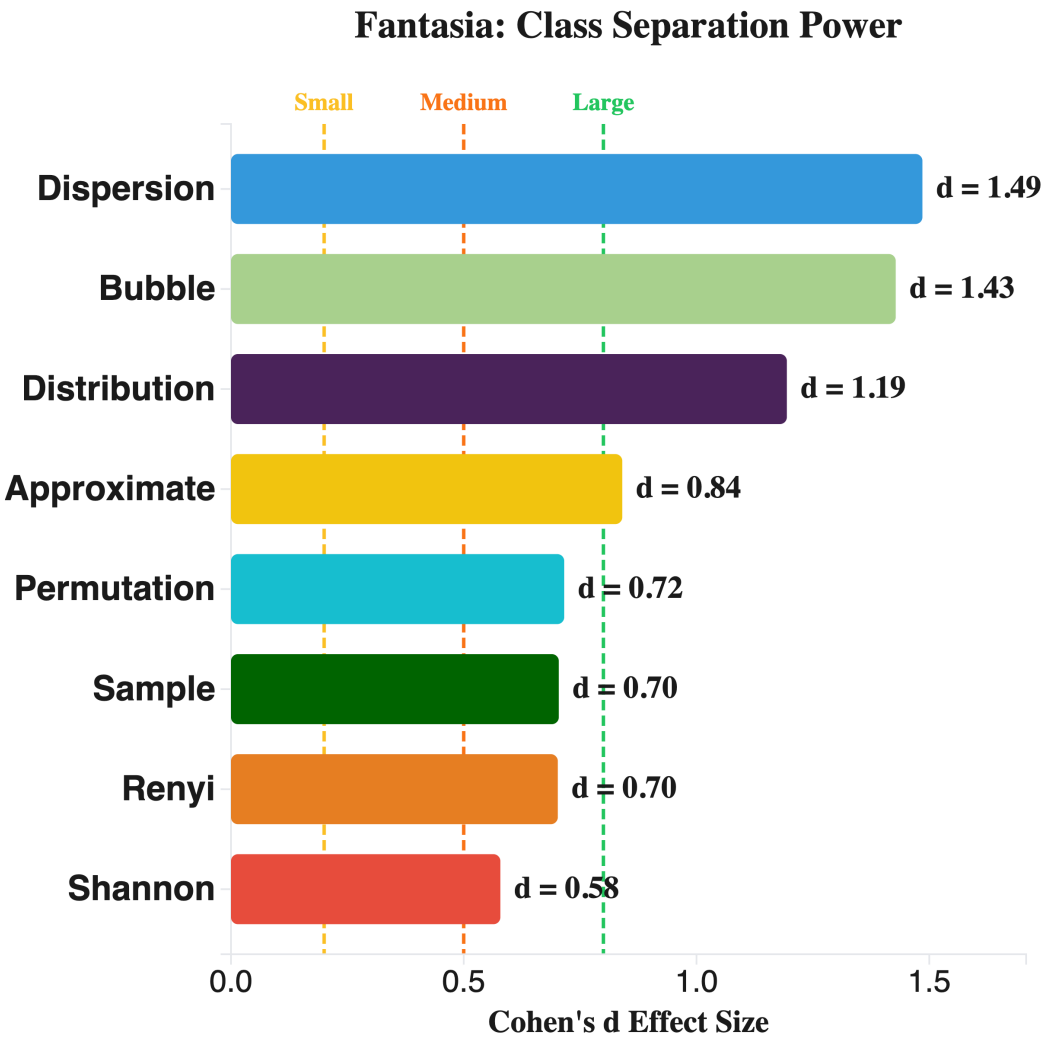

Figure 5.11: Fantasia: Class separation power (Cohen's d) for age discrimination. Higher values indicate better separation between young and elderly subjects.

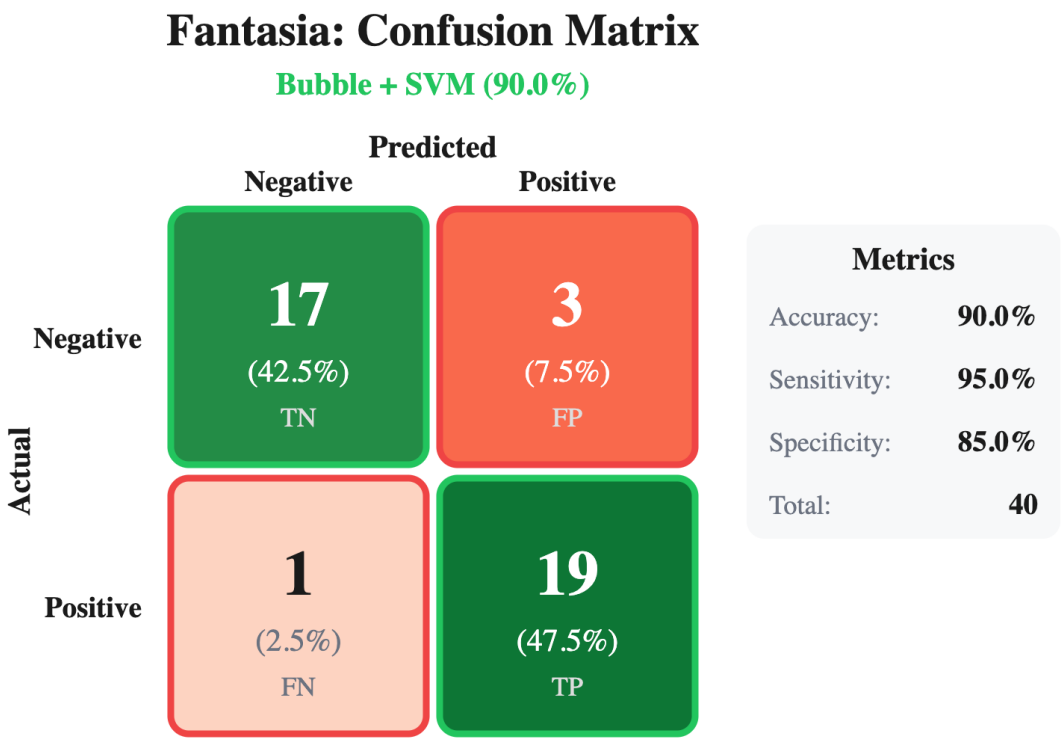

Figure 5.12: Fantasia: Confusion matrix for Bubble + SVM age classification. The model achieves balanced performance across young and elderly subjects.

5.5.4 ECG Arrhythmia

Detecting arrhythmia from MIT-BIH database ECG recordings.

Table 5.12: ECG Arrhythmia: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM	93.8%
2	Permutation + RF	85.4%
3	Dispersion + kNN	83.3%

This dataset yields the second-highest accuracy across all datasets (93.8%), with Bubble + SVM outperforming the next best method by 8.4 percentage points. Fig-

ure 5.13 shows the complete performance matrix, and Figure 5.14 presents the class separation analysis.

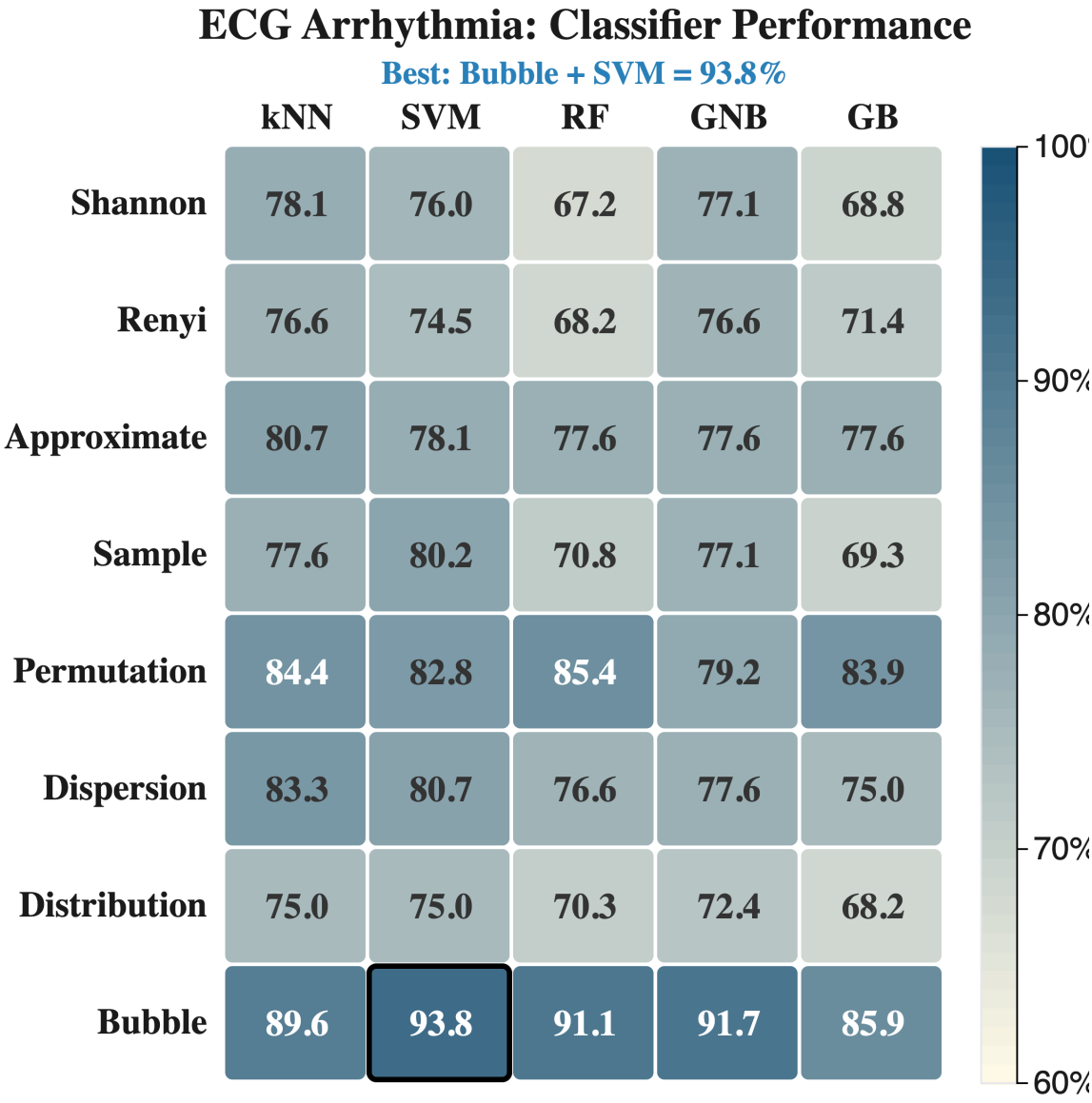

Figure 5.13: ECG Arrhythmia: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (93.8%).

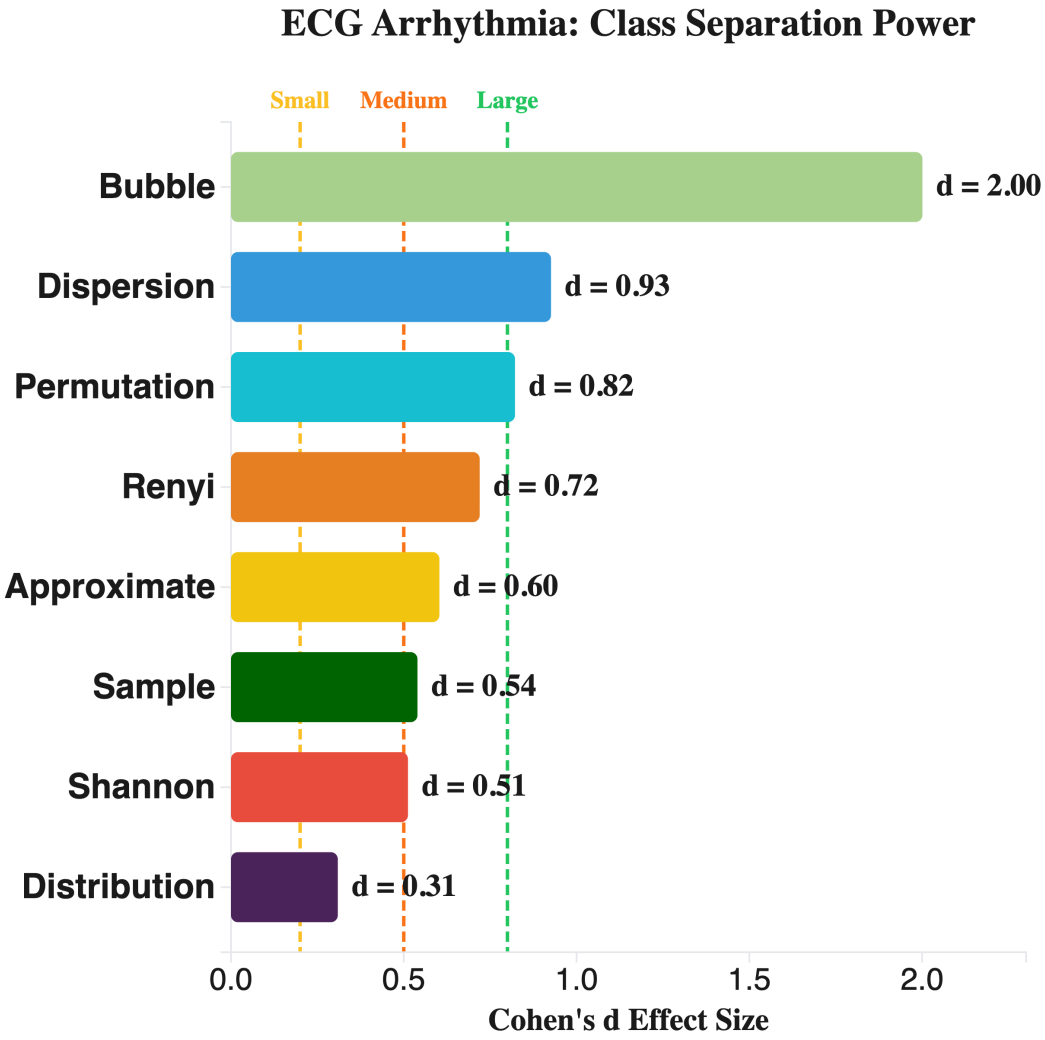

Figure 5.14: ECG Arrhythmia: Class separation power (Cohen's d) for arrhythmia detection.

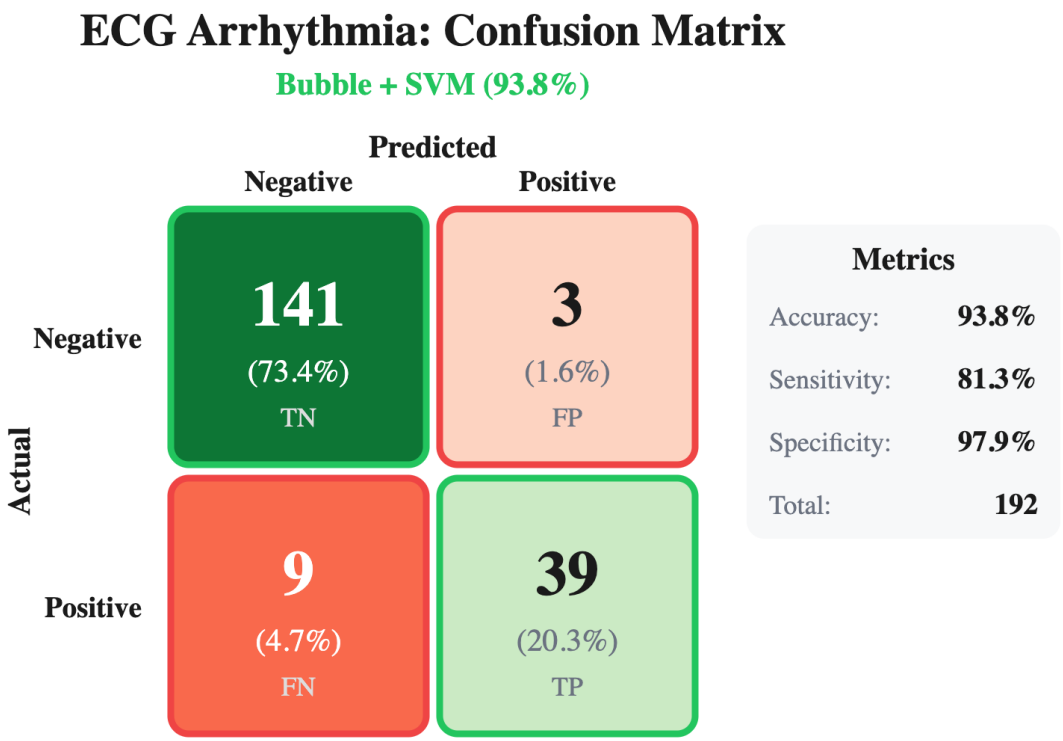

Figure 5.15: ECG Arrhythmia: Confusion matrix for Bubble + SVM. High specificity (97.9%) with moderate sensitivity (81.3%) reflects the class imbalance in the dataset.

5.5.5 SCD vs Healthy

Identifying Sudden Cardiac Death risk from HRV signals.

Table 5.13: SCD vs Healthy: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM	88.1%
2	Distribution + kNN	83.1%
3	Approximate + kNN	79.7%

Figure 5.16 shows the entropy-classifier performance matrix, and Figure 5.17 presents the class separation analysis.

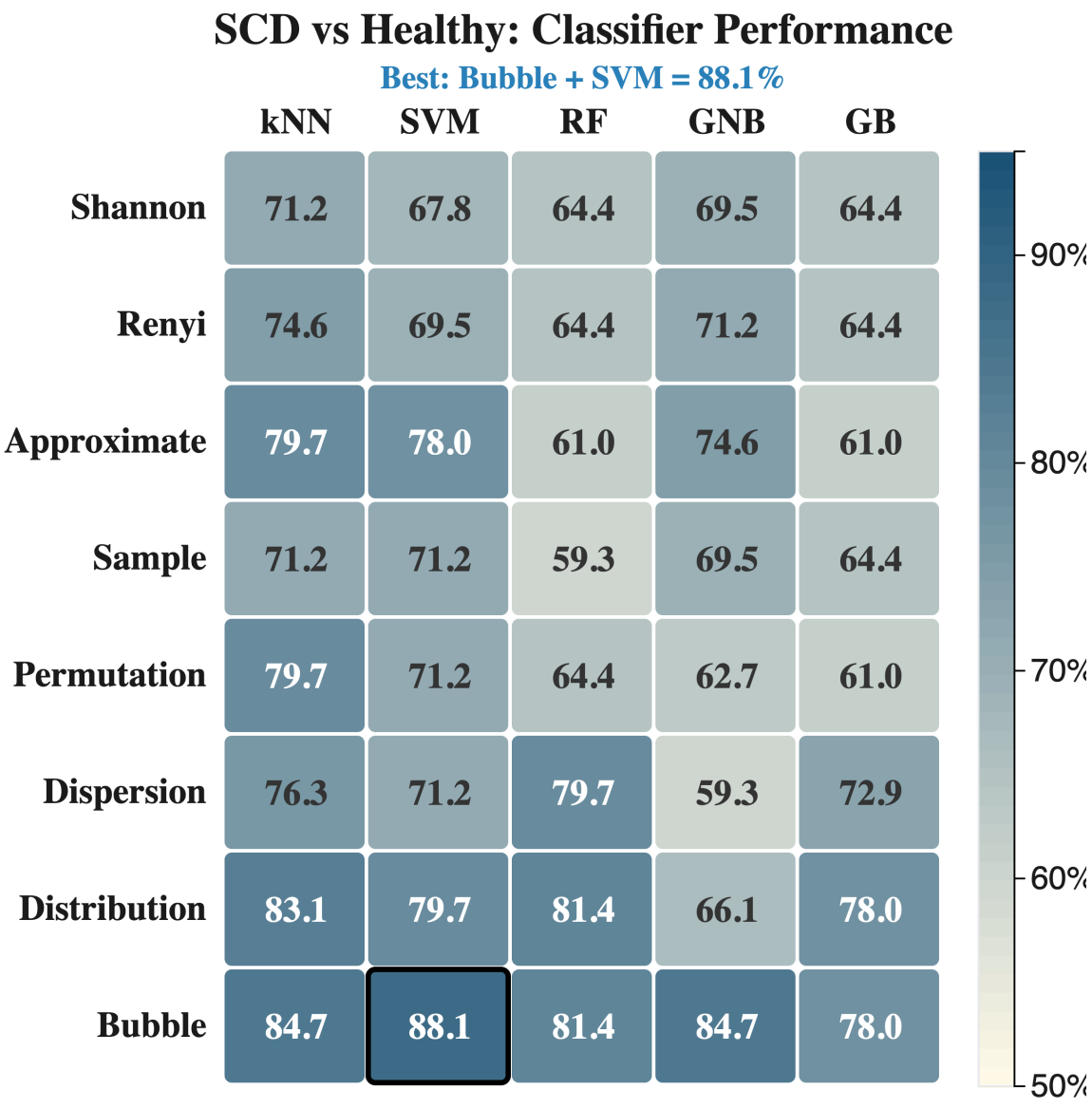

Figure 5.16: SCD vs Healthy: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (88.1%).

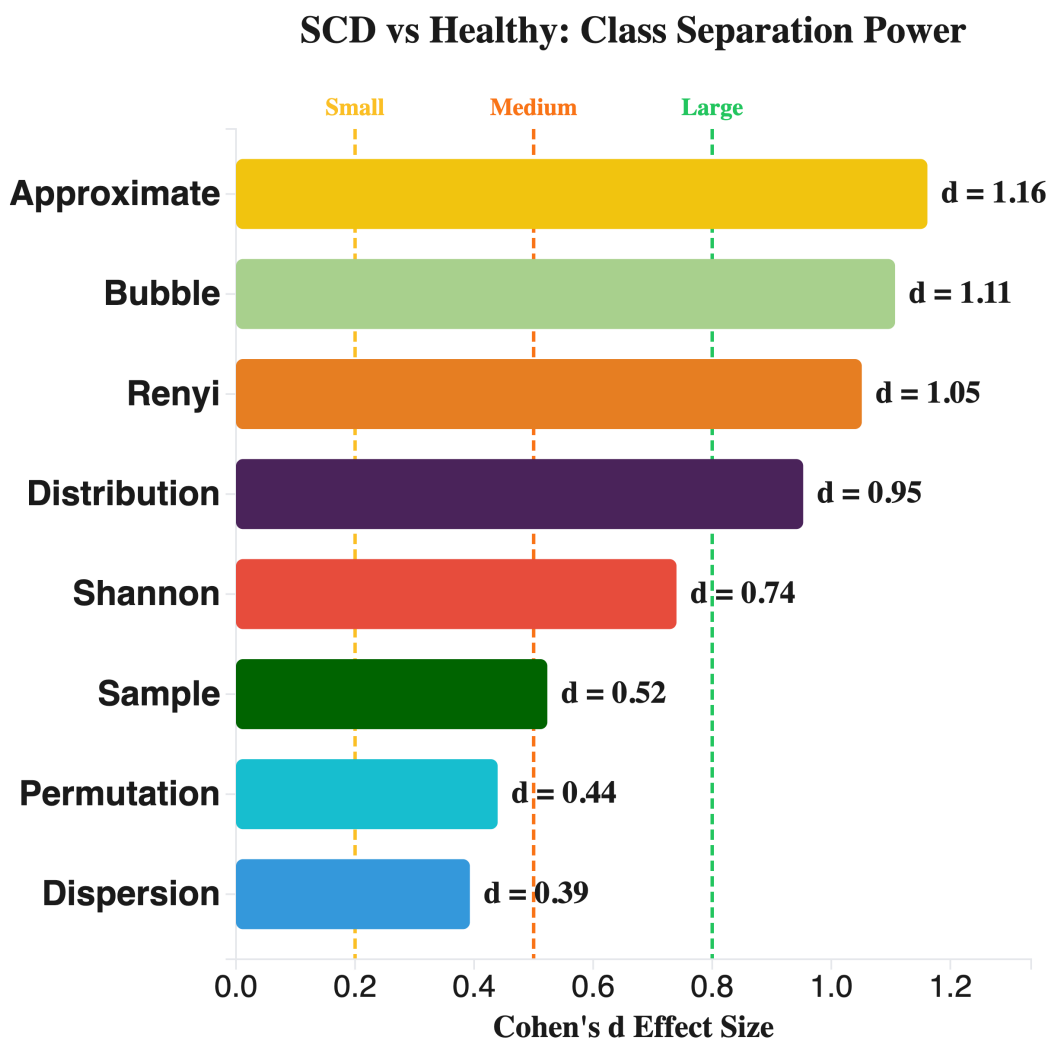

Figure 5.17: SCD vs Healthy: Class separation power (Cohen's d) for sudden cardiac death risk identification.

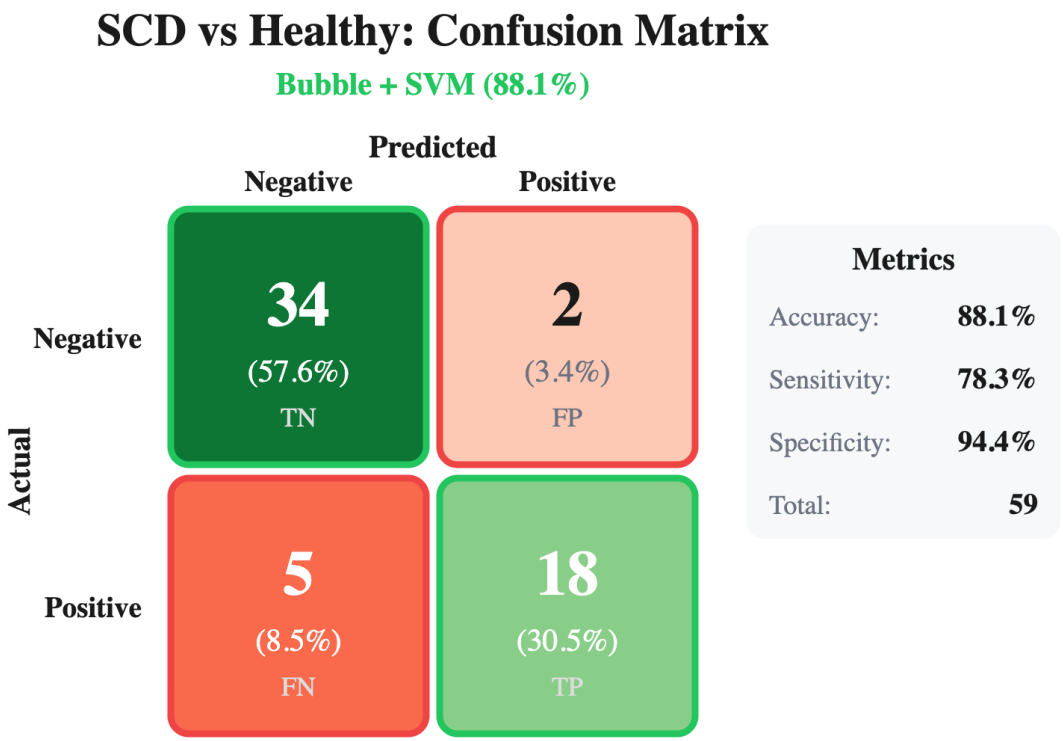

Figure 5.18: SCD vs Healthy: Confusion matrix for Bubble + SVM sudden cardiac death risk prediction.

5.5.6 Gait

Distinguishing Parkinson's disease patients from healthy controls.

Table 5.14: Gait: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Dispersion + SVM	75.3%
2	Sample + kNN	73.4%
3	Renyi + kNN	72.5%

Dispersion Entropy outperforms Bubble Entropy (69.9%) for gait analysis. Figure 5.19 shows the complete performance matrix, and Figure 5.20 reveals that Dispersion Entropy achieves superior class separation.

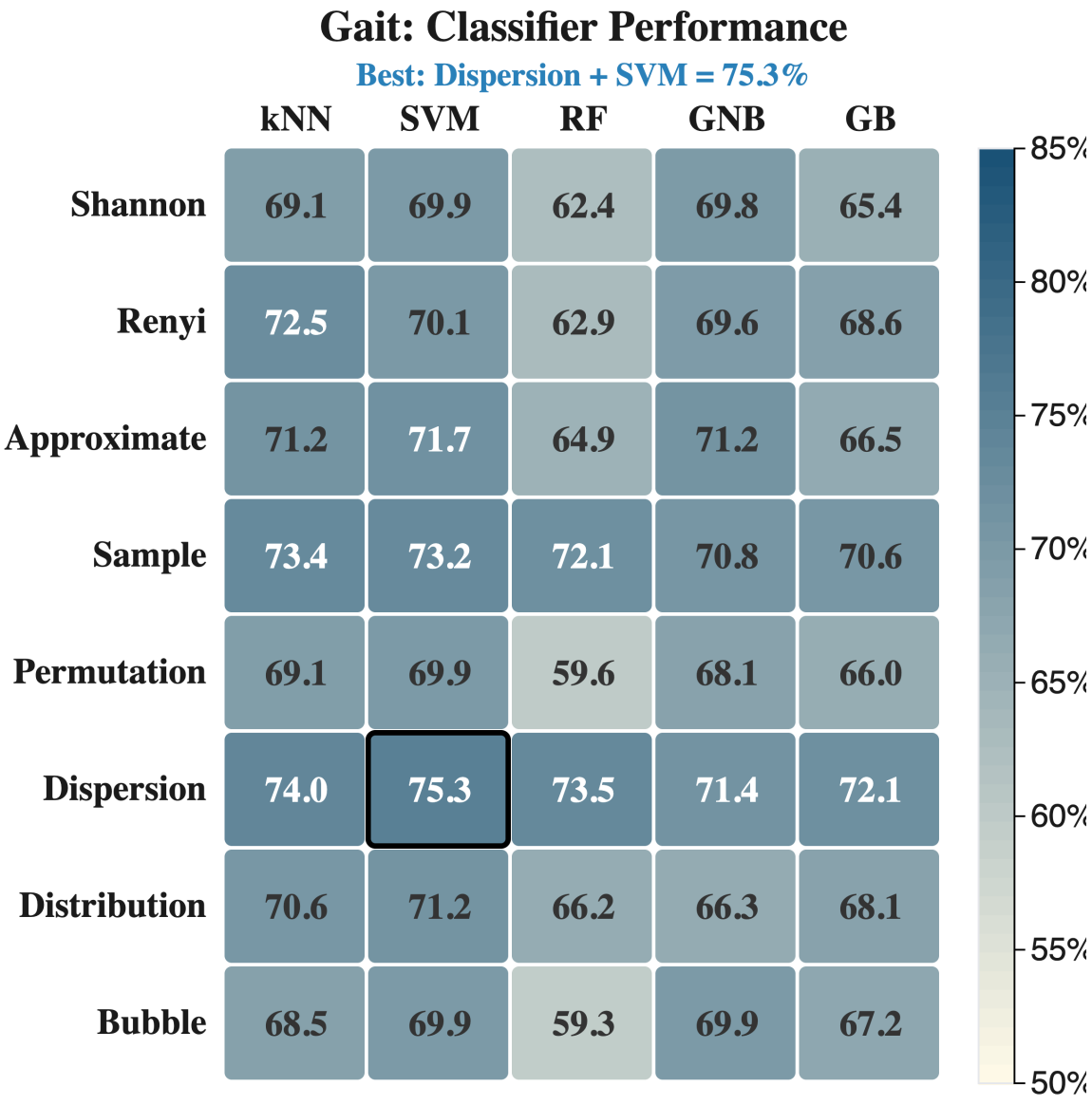

Figure 5.19: Gait: Entropy × Classifier performance heatmap for Parkinson's detection. Dispersion Entropy with SVM achieves the highest accuracy (75.3%).

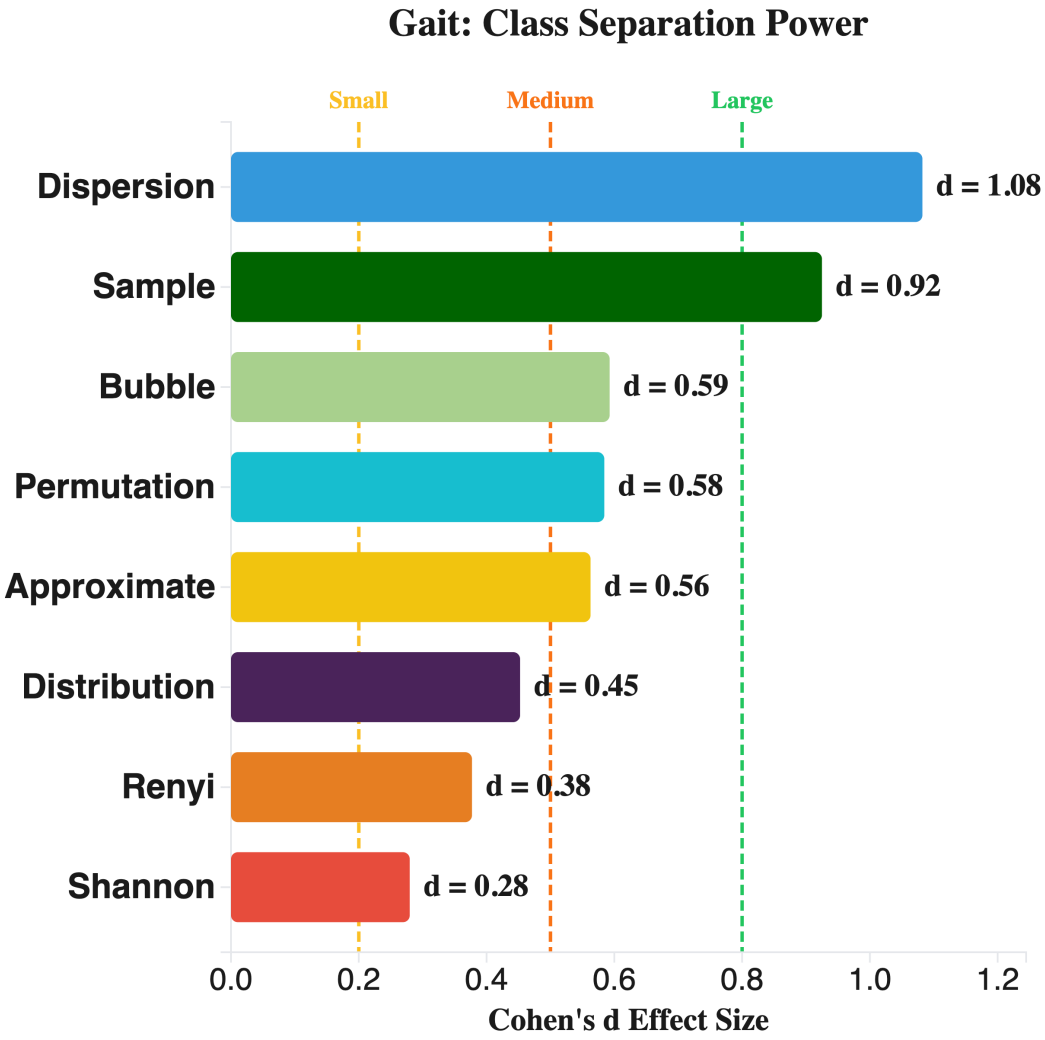

Figure 5.20: Gait: Class separation power (Cohen's d) for Parkinson's detection. Dispersion Entropy achieves the highest effect size, explaining its superior performance over Bubble Entropy.

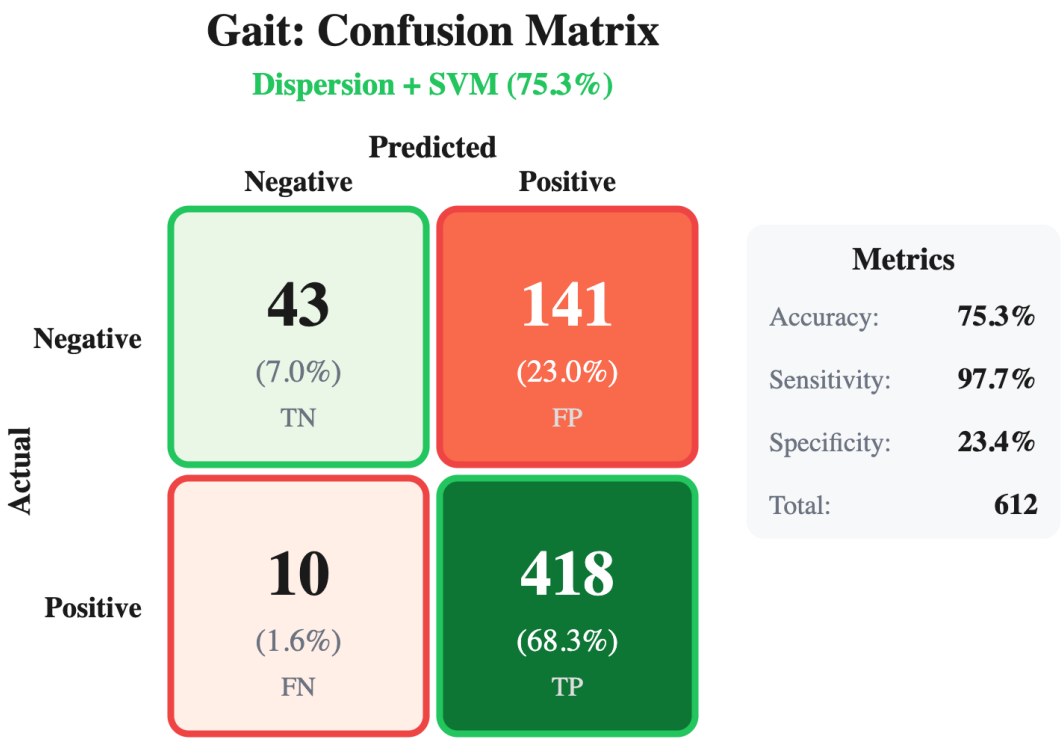

Figure 5.21: Gait: Confusion matrix for Dispersion + SVM Parkinson's detection. Unlike other datasets, Dispersion Entropy outperforms Bubble Entropy here.

5.5.7 EMG

Classifying neuromuscular disorders from EMG signals (biceps and deltoid muscles).

Table 5.15: EMG: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Dispersion + SVM (Deltoid)	80.5%
2	Approximate + SVM (Deltoid)	79.0%
3	Sample + SVM (Deltoid)	76.0%

Dispersion Entropy leads, with the deltoid muscle providing more discriminative features than biceps. Figure 5.22 shows the performance matrix, and Figure 5.23 presents the class separation analysis.

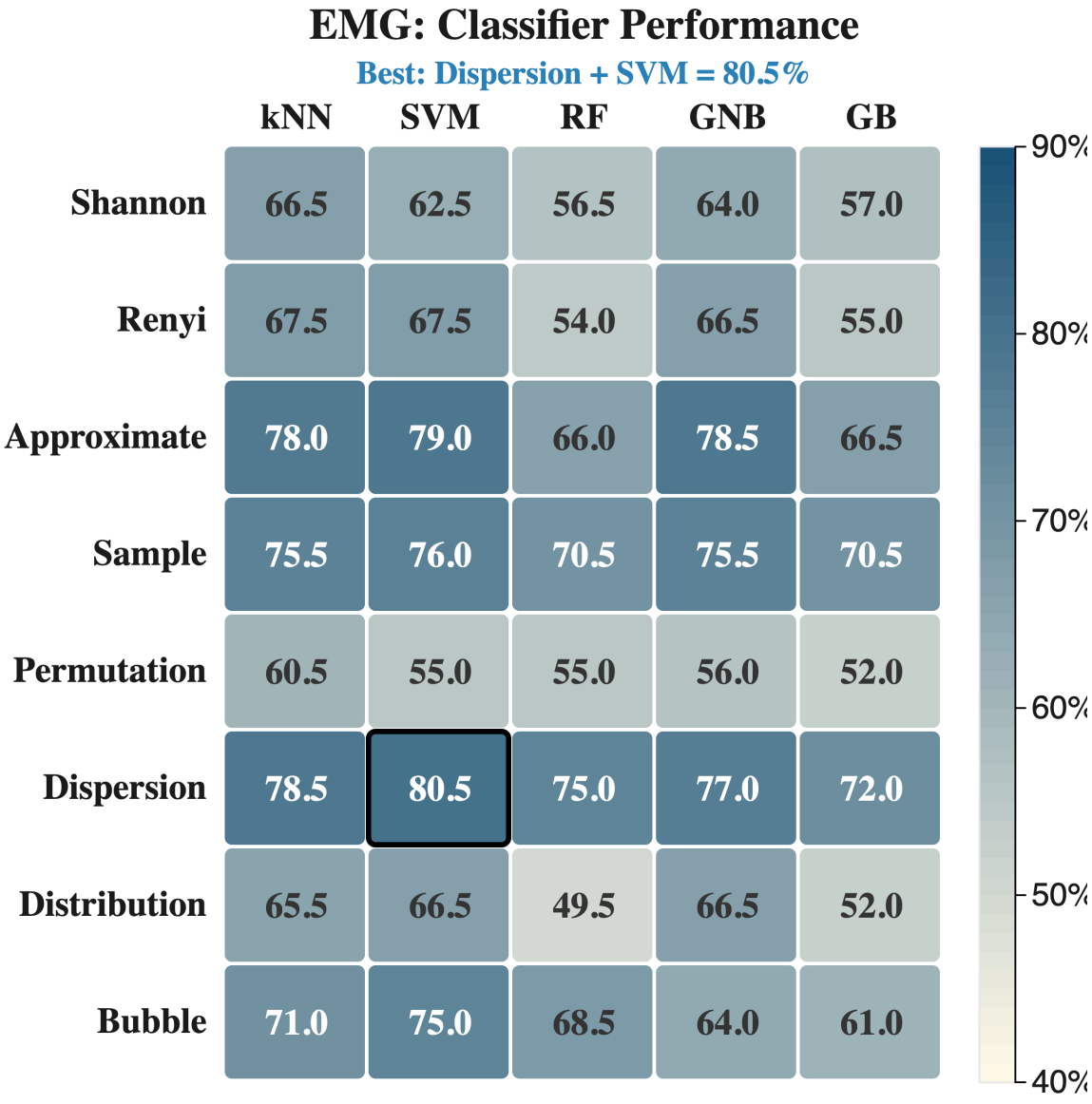

Figure 5.22: EMG: Entropy × Classifier performance heatmap for neuromuscular disorder classification. Dispersion Entropy with SVM achieves the highest accuracy (80.5%).

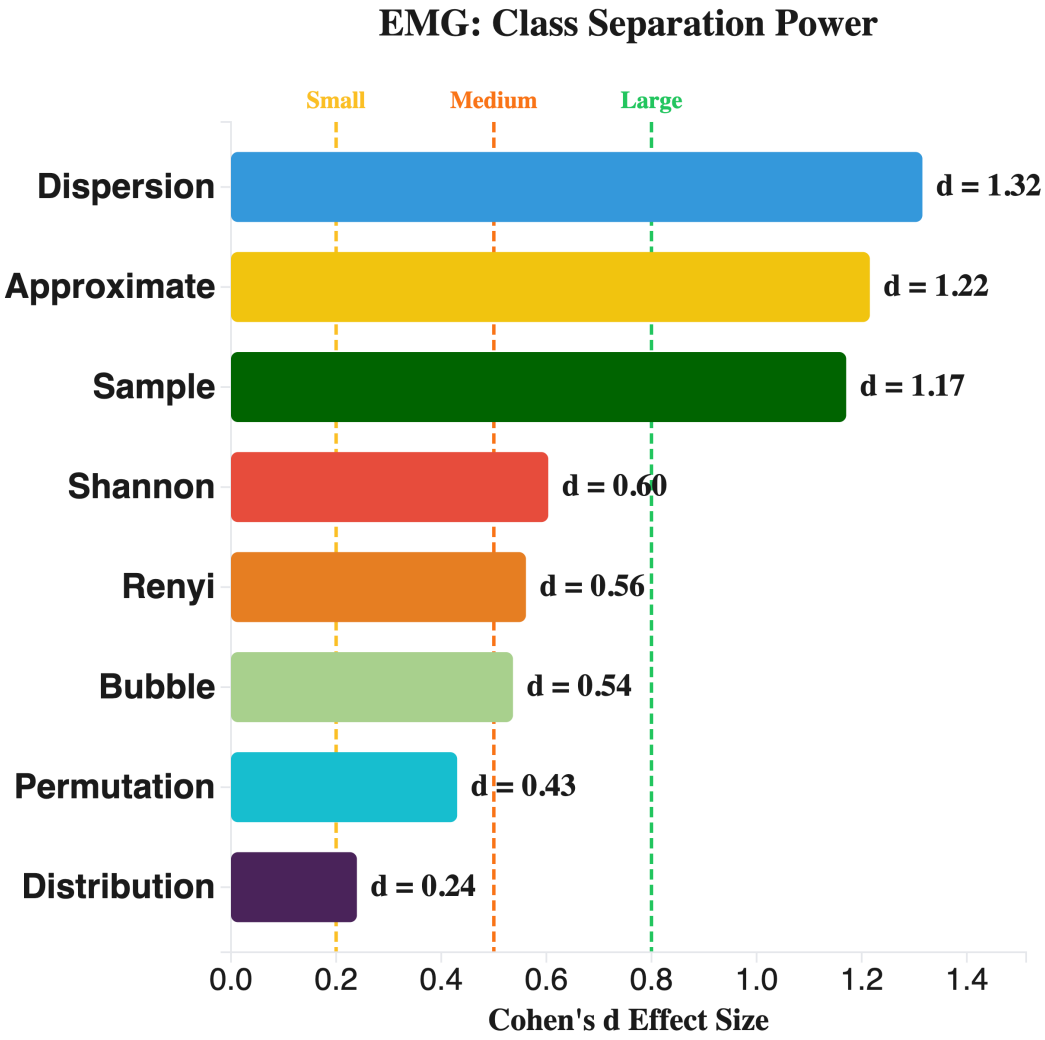

Figure 5.23: EMG: Class separation power (Cohen's d) for neuromuscular disorder classification. Dispersion Entropy achieves the highest effect size.

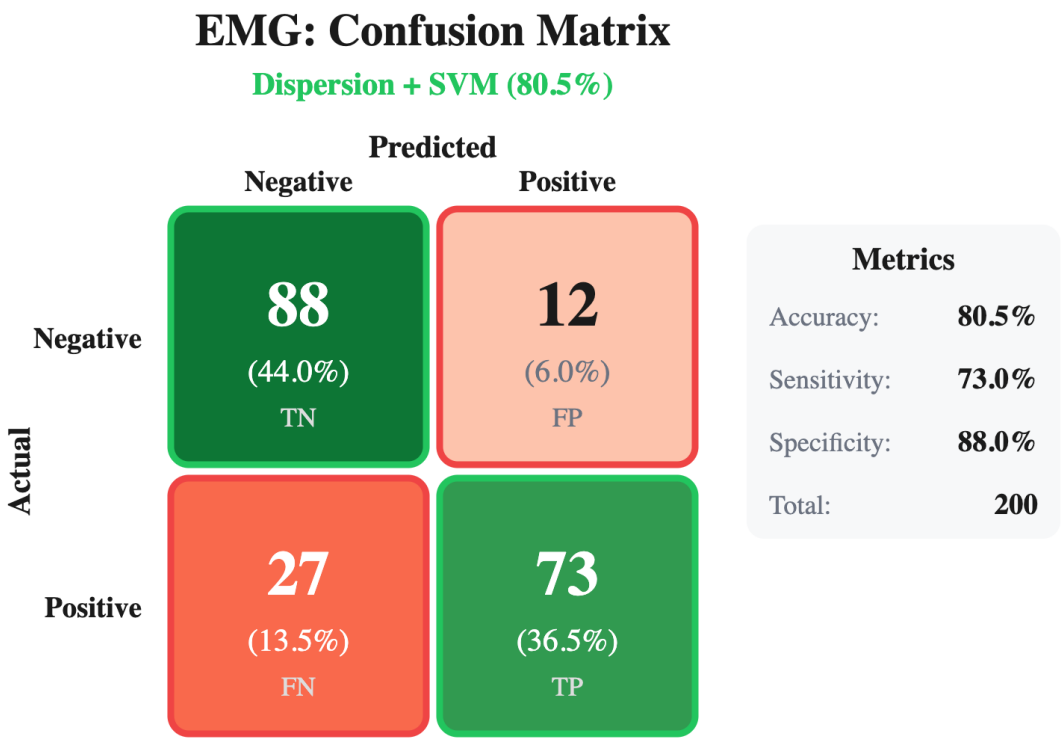

Figure 5.24: EMG: Confusion matrix for Dispersion + SVM neuromuscular disorder classification using deltoid muscle signals.

5.5.8 Apnea

Detecting sleep apnea events from HRV and SpO2 signals.

Table 5.16: Apnea: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM (SpO2 diff)	91.6%
2	Permutation + kNN (SpO2 diff)	90.4%
3	Bubble + SVM (SpO2 raw)	90.4%

The SpO2 difference signal provides the most discriminative features, with both Bubble and Permutation exceeding 90%. Figure 5.25 shows the performance matrix, and Figure 5.26 presents the class separation analysis.

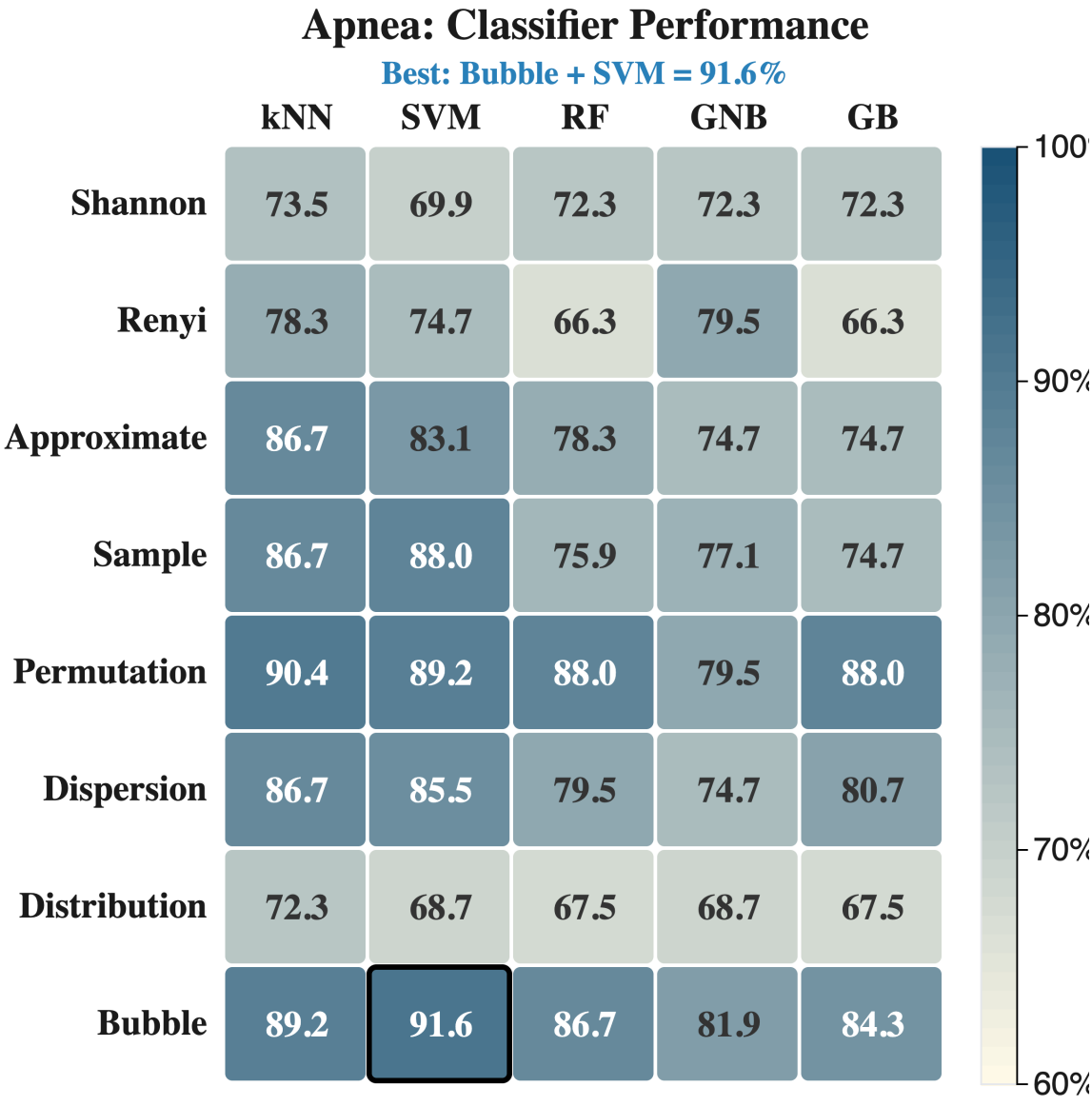

Figure 5.25: Apnea: Entropy × Classifier performance heatmap. Bubble Entropy with SVM achieves the highest accuracy (91.6%) using SpO2 difference signal.

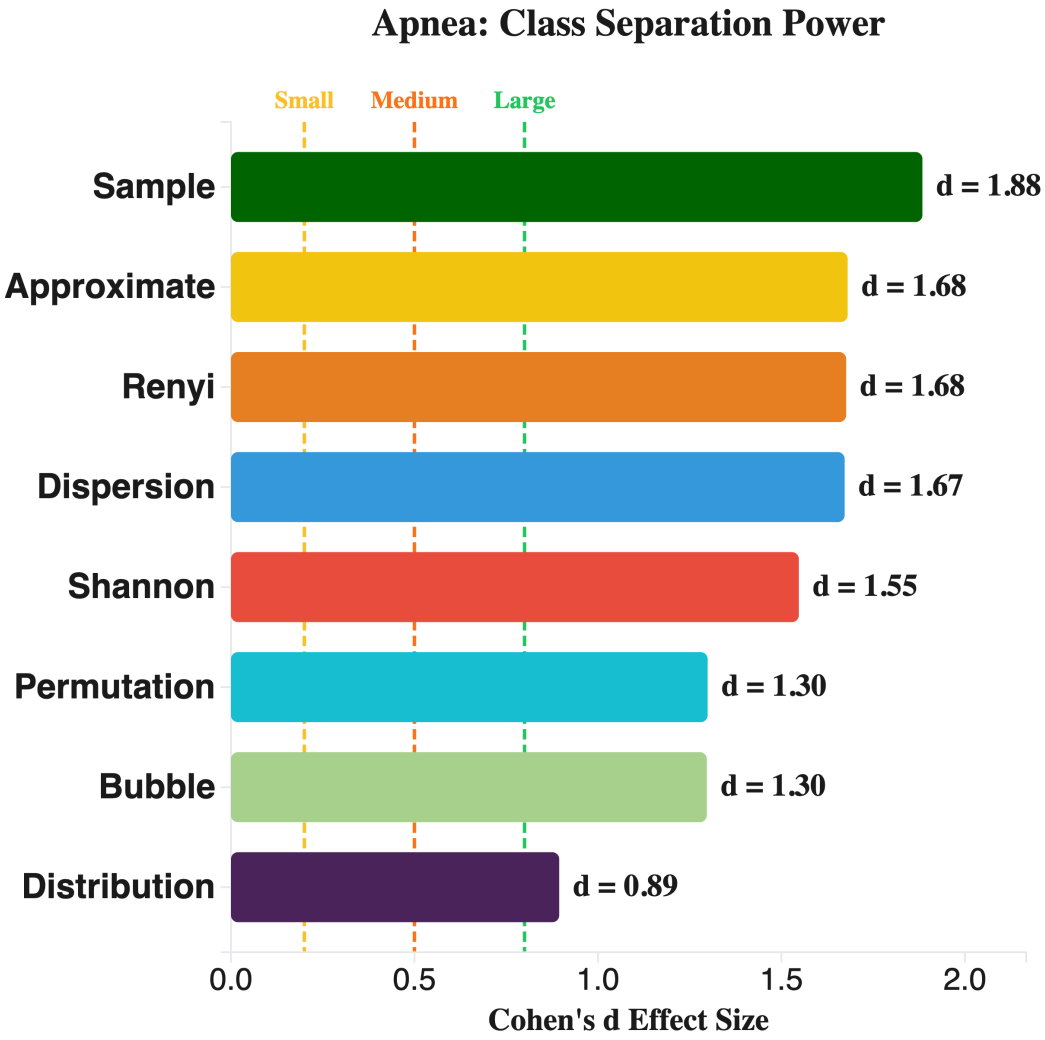

Figure 5.26: Apnea: Class separation power (Cohen's d) for sleep apnea detection.

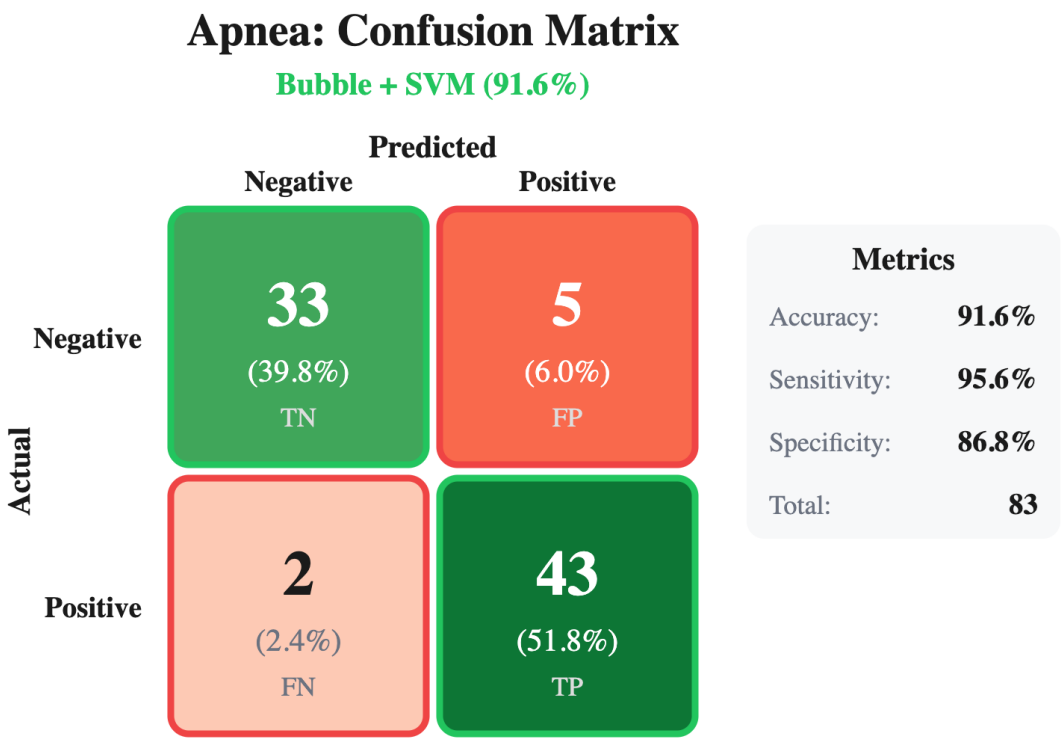

Figure 5.27: Apnea: Confusion matrix for Bubble + SVM sleep apnea detection using SpO2 difference signal.

5.5.9 GAMEEMO

Classifying emotional states from EEG signals (AF3, AF4, F4 channels).

Table 5.17: GAMEEMO: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + RF (F4)	78.6%
2	Bubble + RF (AF3)	75.0%
3	Bubble + RF (AF4)	70.5%

Random Forest emerges as the optimal classifier—the only dataset where RF outperforms both kNN and SVM. Figure 5.28 shows the performance matrix, and Figure 5.29 presents the class separation analysis.

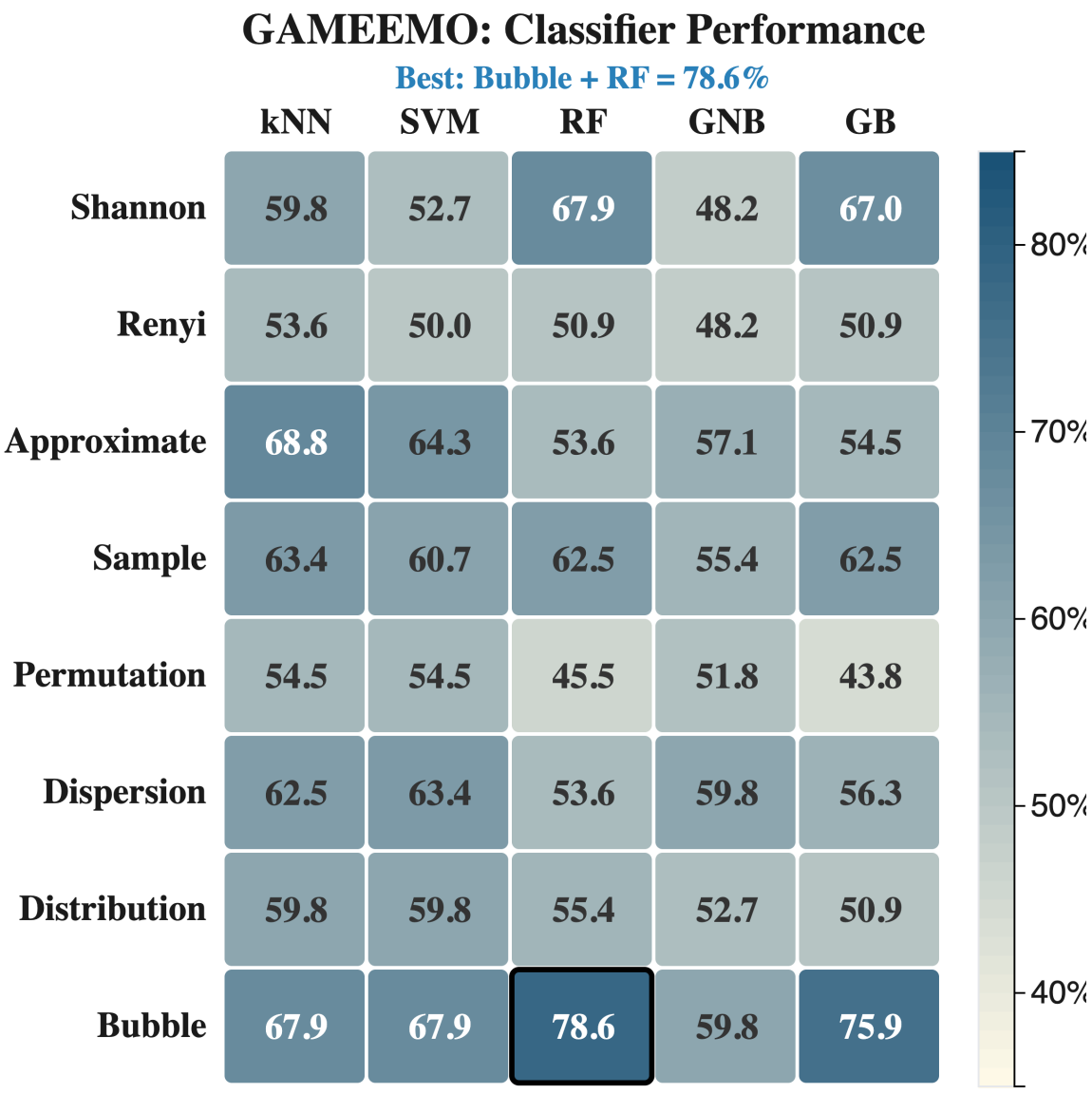

Figure 5.28: GAMEEMO: Entropy × Classifier performance heatmap for emotion classification. Bubble Entropy with RF achieves the highest accuracy (78.6%).

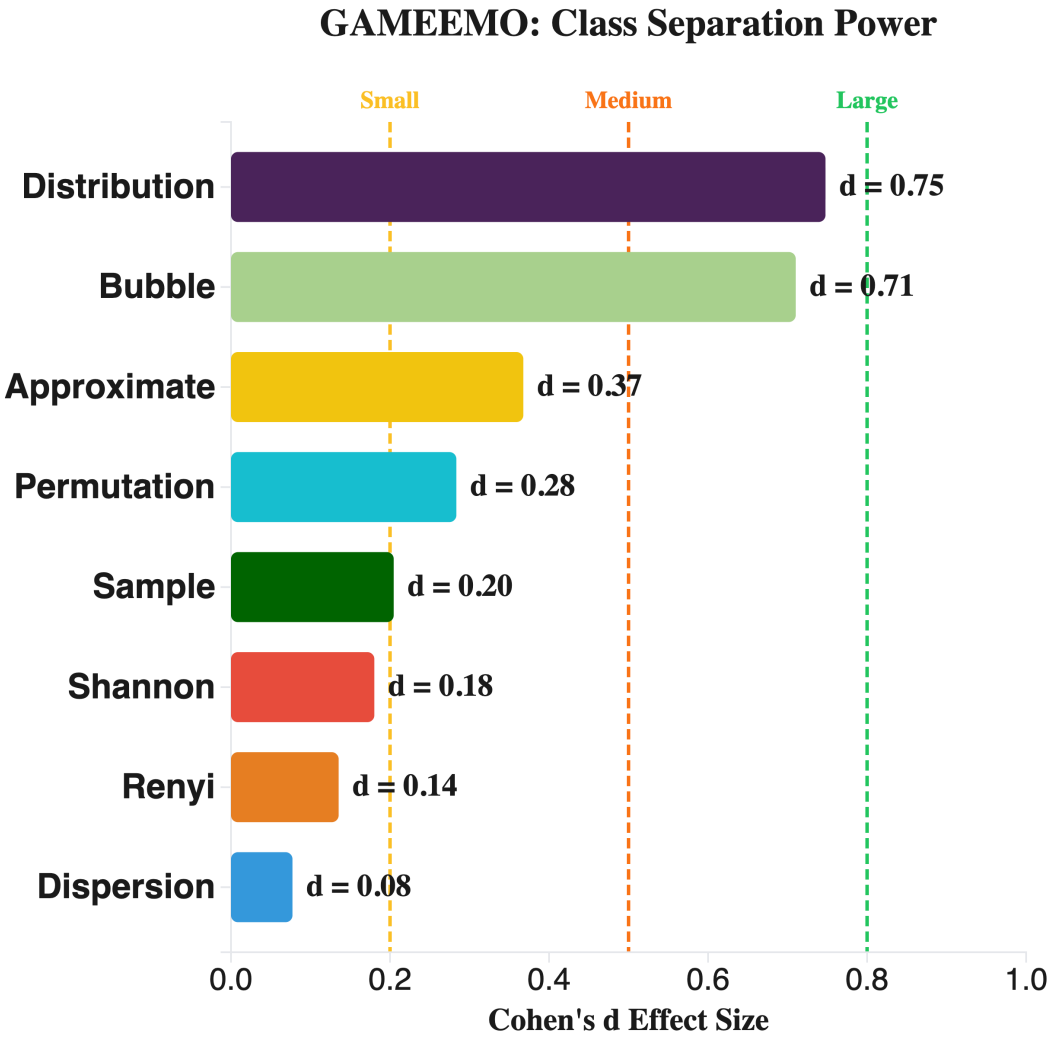

Figure 5.29: GAMEEMO: Class separation power (Cohen's d) for emotional state classification.

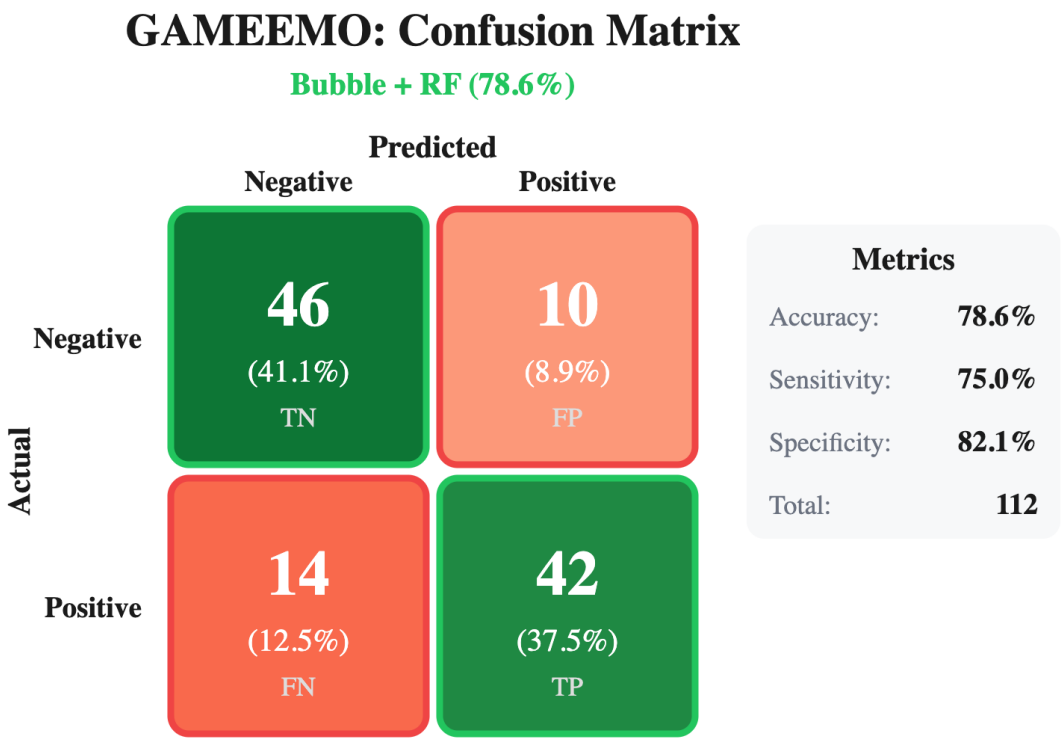

Figure 5.30: GAMEEMO: Confusion matrix for Bubble + RF emotion classification. Random Forest is the optimal classifier for this EEG-based task.

5.5.10 Mental Arithmetic

Classifying cognitive load during mental arithmetic tasks from EEG (Fz, Cz, Pz channels).

Table 5.18: Mental Arithmetic: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + kNN (Fz)	73.6%
2	Permutation + kNN (Fz)	69.4%
3	Bubble + kNN (Cz)	68.1%

The frontal electrode (Fz) provides the most informative features for cognitive load classification. Figure 5.31 shows the performance matrix, and Figure 5.32 presents

the class separation analysis.

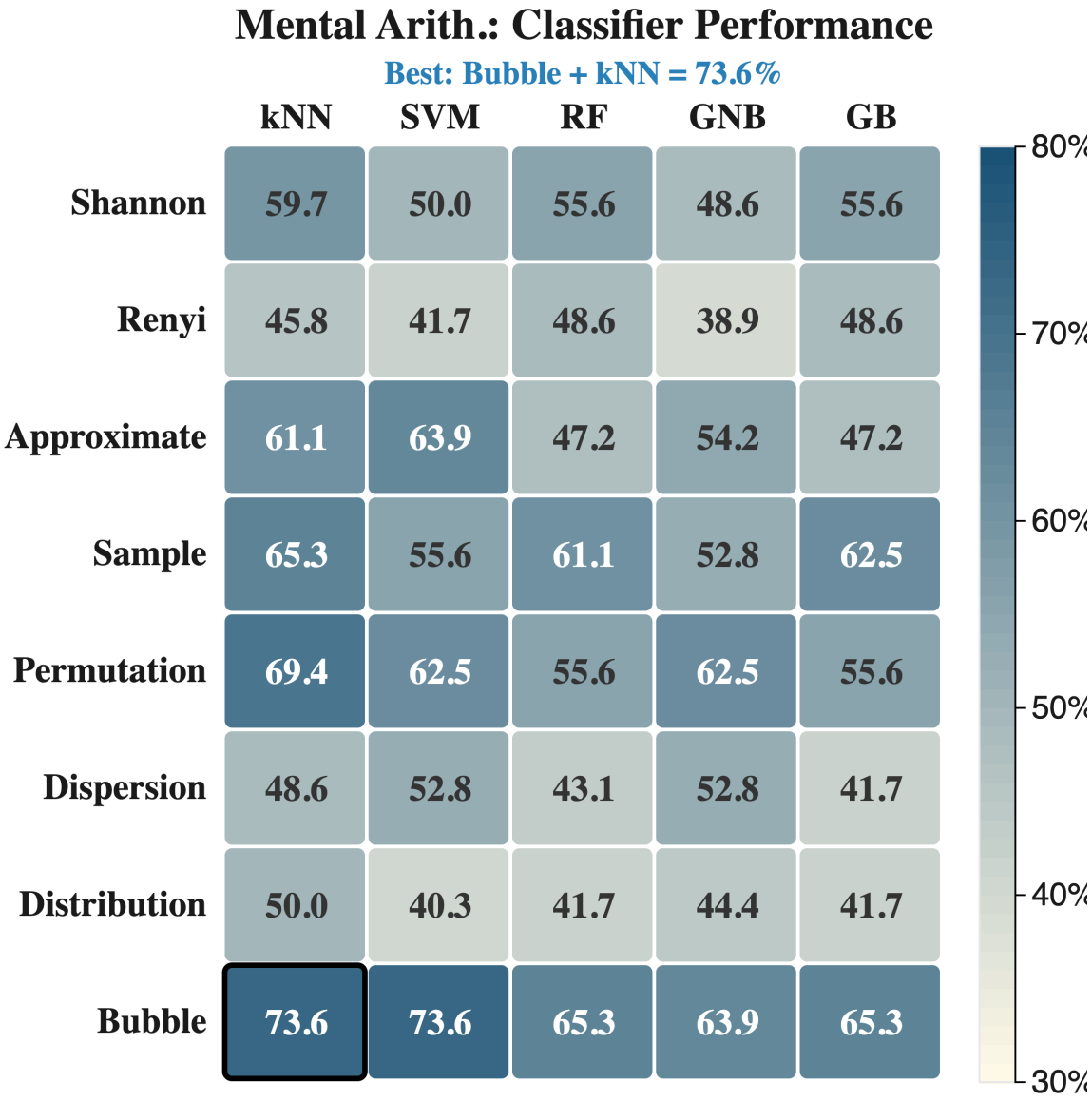

Figure 5.31: Mental Arithmetic: Entropy × Classifier performance heatmap for cognitive load classification. Bubble Entropy with kNN achieves the highest accuracy (73.6%).

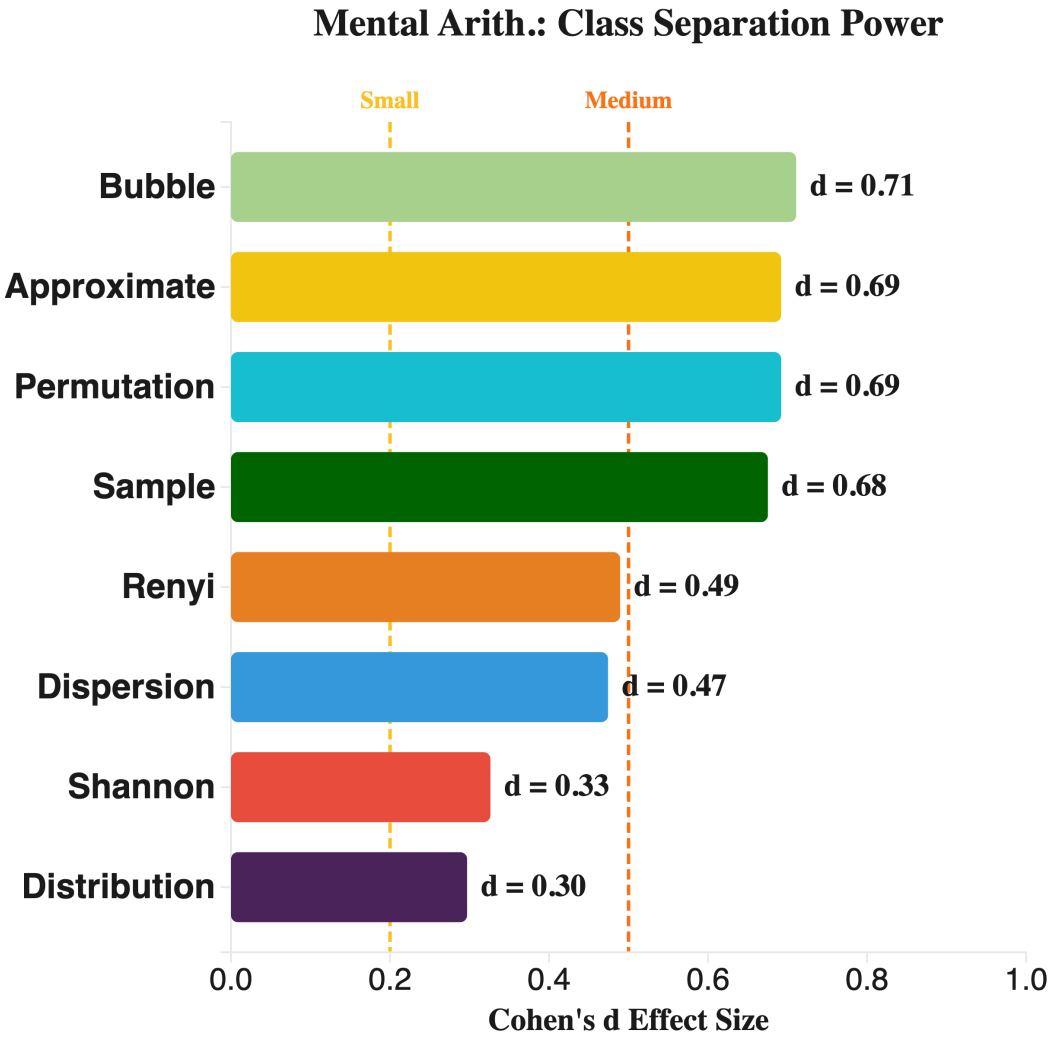

Figure 5.32: Mental Arithmetic: Class separation power (Cohen's d) for cognitive load detection.

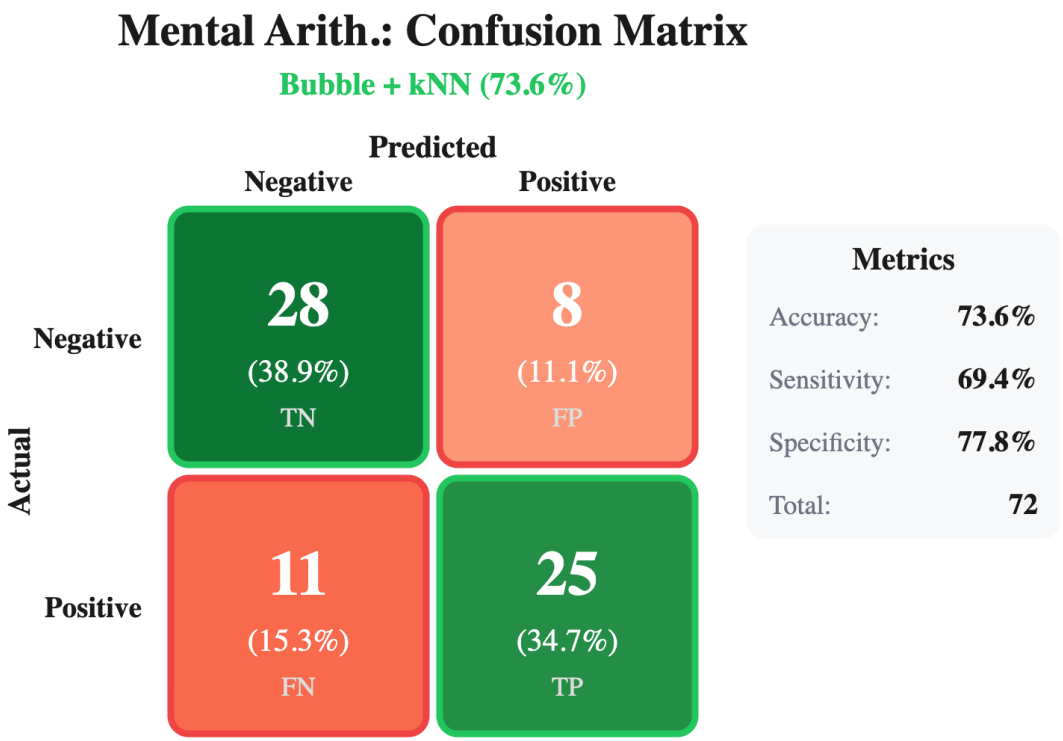

Figure 5.33: Mental Arithmetic: Confusion matrix for Bubble + kNN cognitive load classification using Fz electrode.

5.5.11 PPG

Classification using PPG waveforms across six measurement channels.

Table 5.19: PPG: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Permutation + kNN	95.5%
2	Permutation + SVM	93.9%
3	Distribution + RF	93.9%

The highest accuracy across all datasets (95.5%). PPG is the only dataset where Permutation Entropy outperforms Bubble Entropy. Figure 5.34 shows the performance matrix, and Figure 5.35 presents the class separation analysis.

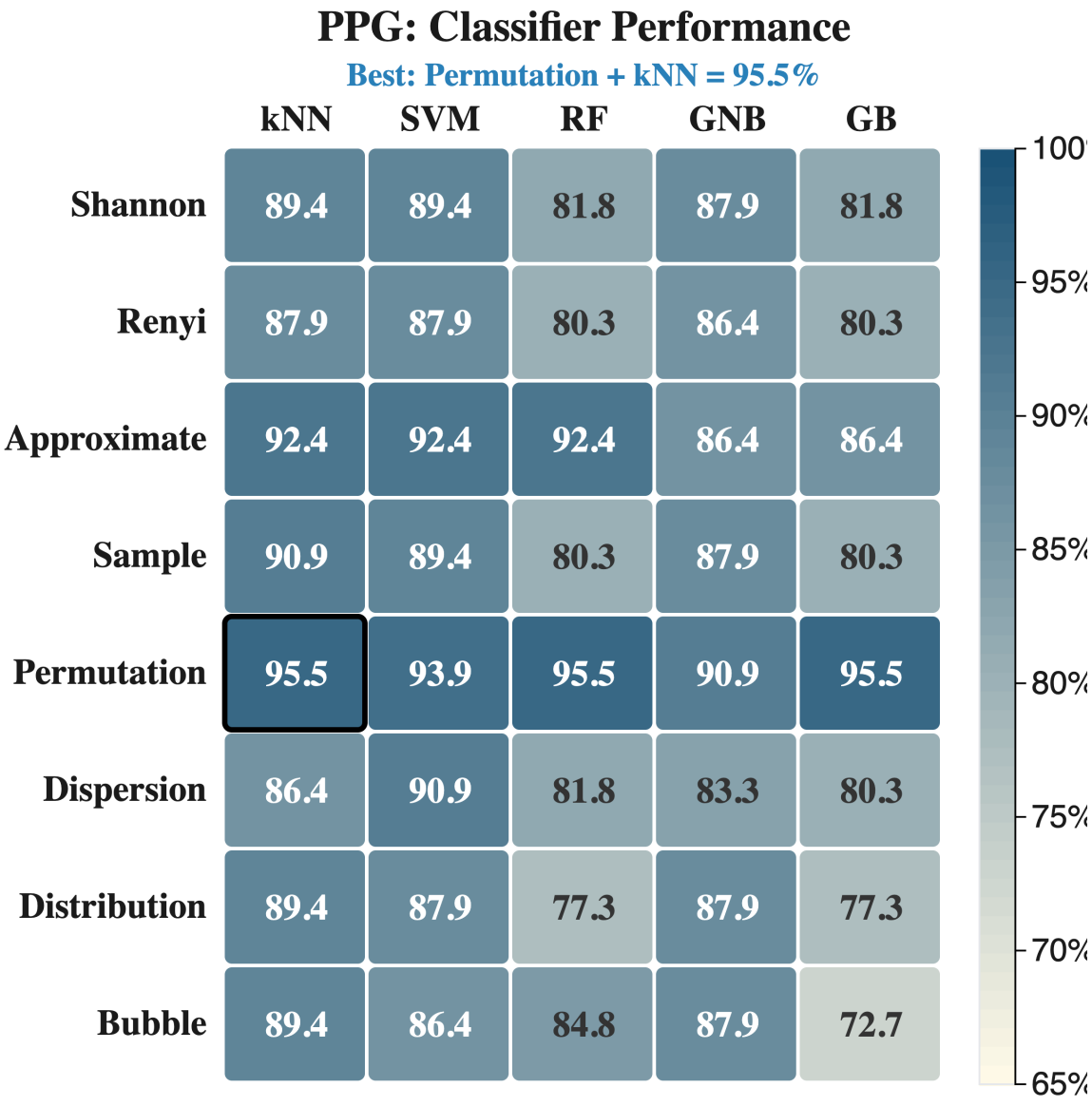

Figure 5.34: PPG: Entropy × Classifier performance heatmap. Unlike other datasets, Permutation Entropy achieves the best results (95.5% with kNN).

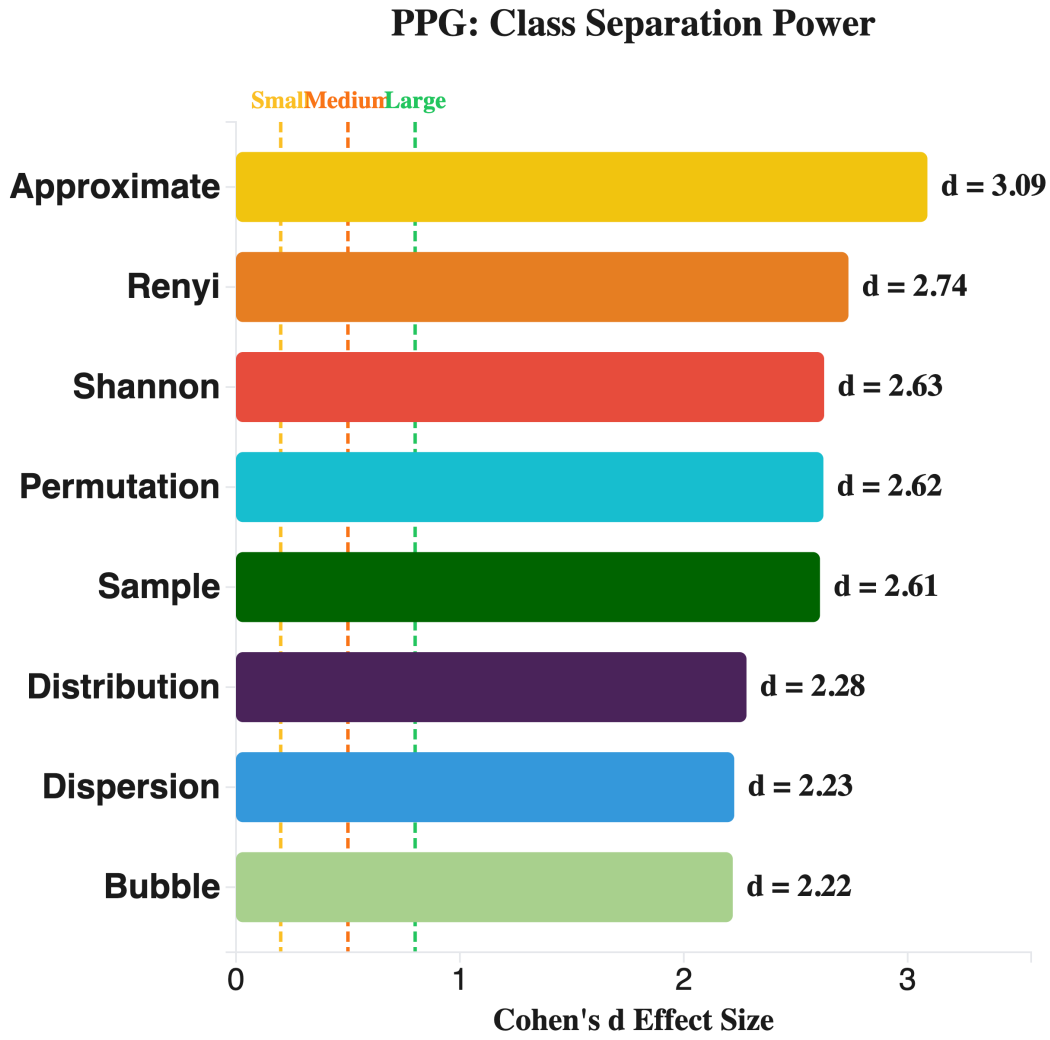

Figure 5.35: PPG: Class separation power (Cohen's d). Approximate Entropy achieves the highest effect size (d=3.09), yet Permutation Entropy achieves the best classification accuracy (95.5%), illustrating that the relationship between effect size and classification accuracy is not strictly linear.

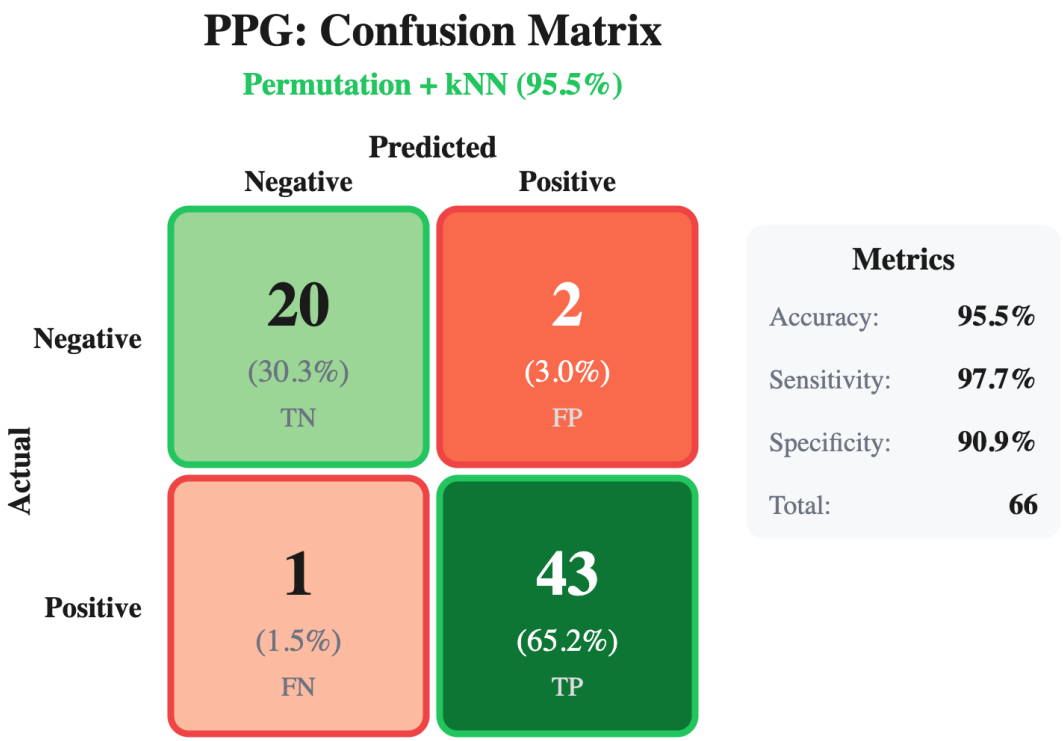

Figure 5.36: PPG: Confusion matrix for Permutation + kNN. This is the only dataset where Permutation Entropy outperforms Bubble Entropy, achieving the highest accuracy (95.5%) across all datasets.

5.5.12 Voiced

Distinguishing pathological from healthy voice samples.

Table 5.20: Voiced: Top 3 entropy-classifier combinations.

Rank	Combination	Accuracy
1	Bubble + SVM	76.0%
2	Bubble + kNN	74.0%
3	Bubble + RF	73.6%

Bubble Entropy dominates the top three positions with different classifiers, achieving up to 76.0% accuracy. Figure 5.37 shows the performance matrix, and Figure 5.38

presents the class separation analysis.

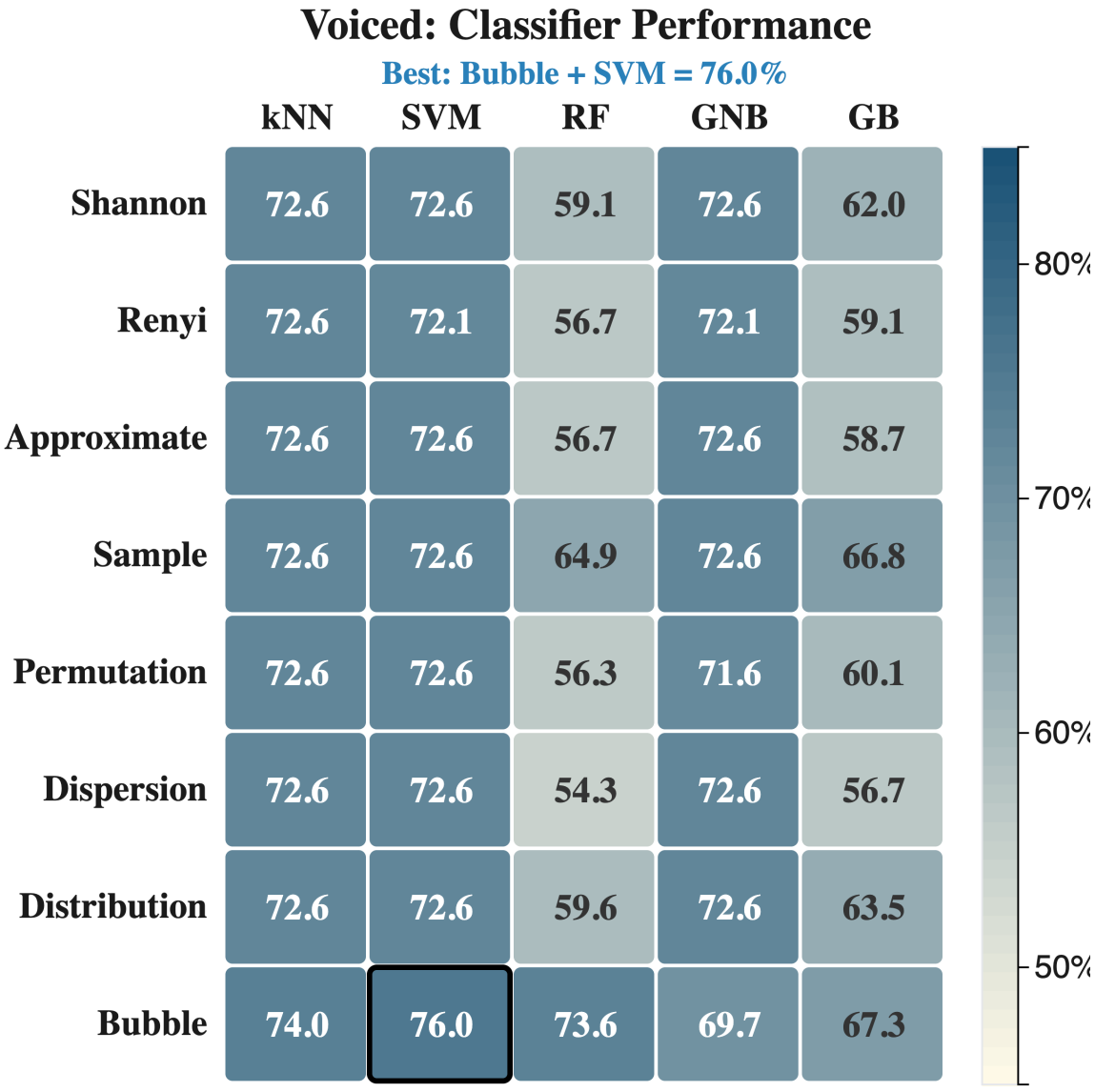

Figure 5.37: Voiced: Entropy × Classifier performance heatmap for voice disorder detection. Bubble Entropy with SVM achieves the highest accuracy (76.0%).

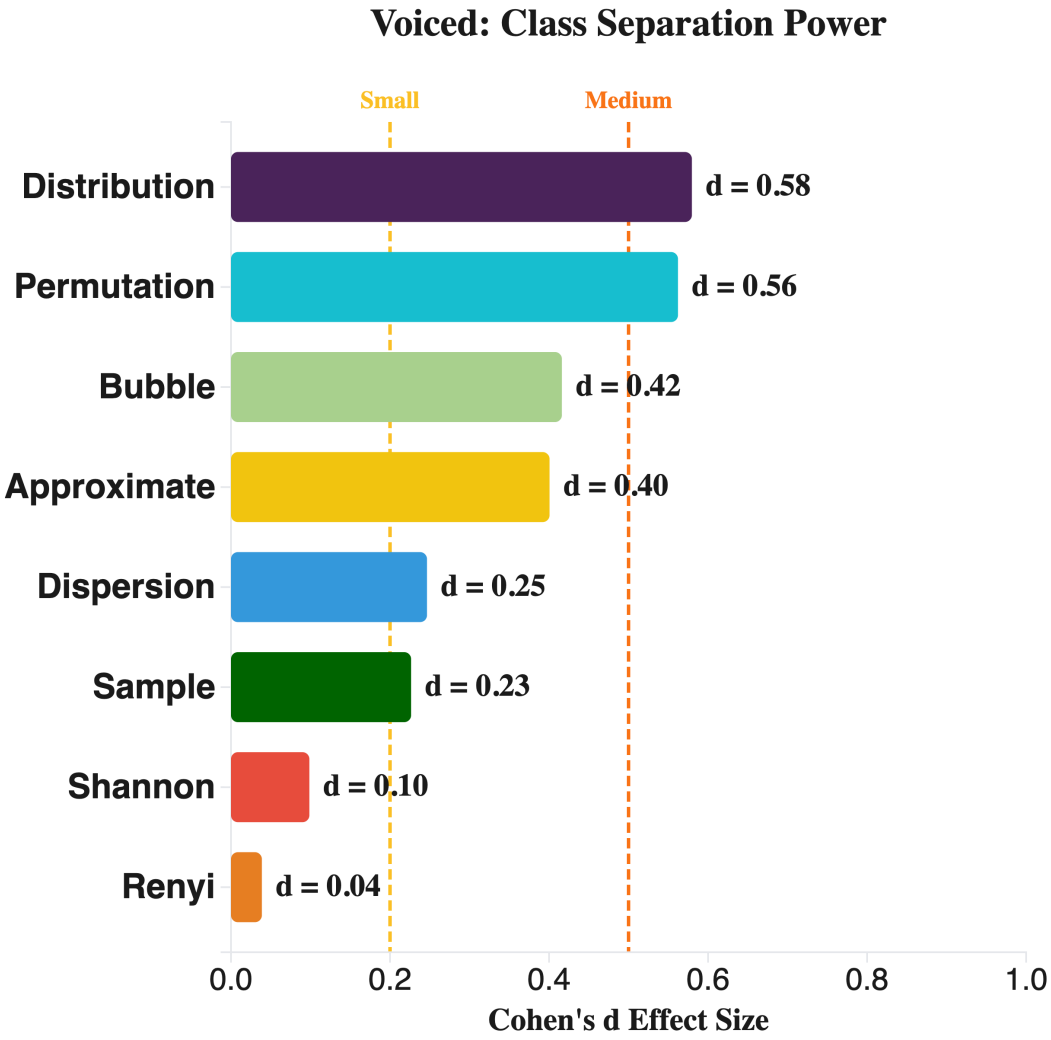

Figure 5.38: Voiced: Class separation power (Cohen's d) for pathological voice detection.

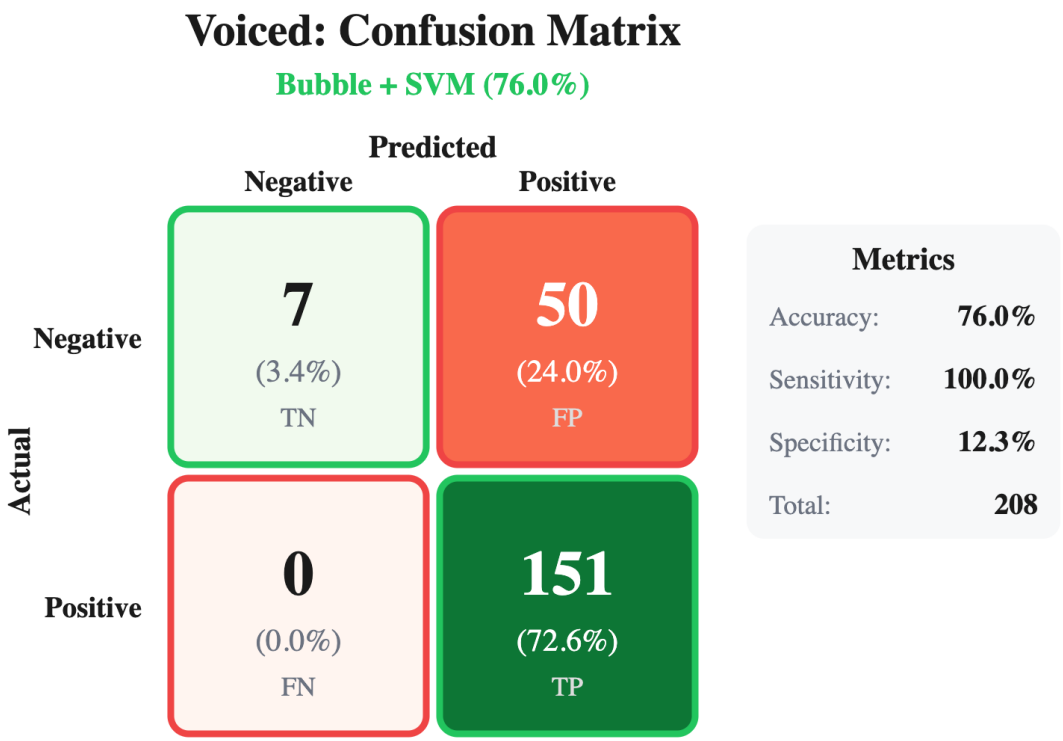

Figure 5.39: Voiced: Confusion matrix for Bubble + SVM pathological voice detection.

5.5.13 Selected Features Analysis

Table 5.21 presents the specific features selected by the exhaustive search for each dataset's best-performing entropy-classifier combination.

Table 5.21: Selected features for best performing entropy-classifier combination per dataset. Features are named as `entropy_parameter` (e.g., `bubben_m12` = Bubble Entropy with $m = 12$).

Dataset	Entropy	k	Acc	Selected Features
PPG	Permutation	1	95.5%	`permen_m2`
ECG Arrhythmia	Bubble	5	93.8%	`bubben_m2, bubben_m6, bubben_m10, bubben_m19, bubben_m24`
CHF vs NSR	Bubble	5	92.2%	`bubben_m2, bubben_m4, bubben_m12, bubben_m22, bubben_m24`
Apnea	Bubble	5	91.6%	`bubben_m2, bubben_m9, bubben_m10, bubben_m15, bubben_m19`
Fantasia	Bubble	3	90.0%	`bubben_m10, bubben_m14, bubben_m25`
SCD vs Healthy	Bubble	4	88.1%	`bubben_m2, bubben_m7, bubben_m9, bubben_m15`
EMG	Dispersion	3	80.5%	`dispen_m2_c3, dispen_m4_c3, dispen_m4_c5`
GAMEEMO	Bubble	5	78.6%	`bubben_m2, bubben_m3, bubben_m11, bubben_m14, bubben_m15`
Voiced	Bubble	5	76.0%	`bubben_m3, bubben_m6, bubben_m15, bubben_m16, bubben_m18`
Gait	Dispersion	5	75.3%	`dispen_m2_c8, dispen_m3_c3, dispen_m3_c8, dispen_m3_c9, dispen_m4_c6`
Mental Arith.	Bubble	5	73.6%	`bubben_m3, bubben_m7, bubben_m13, bubben_m25, bubben_m27`

Key observations from the selected features:

- **Bubble Entropy dominates**: 8 of 11 datasets use Bubble features, with $m = 2$ appearing most frequently (5 datasets)
- **PPG is uniquely simple**: Only requires 1 feature (`permen_m2`) for 95.5% accuracy
- **Dispersion for movement signals**: Both Gait and EMG prefer Dispersion Entropy with low m (2–4) and varied c (3–8)

- **Feature count varies**: Simpler signals (PPG) need $k = 1$, while complex signals (cardiac, neural) require $k = 5$

5.5.14 Key Findings

1. **Bubble Entropy dominance**: Wins 8 of 11 datasets with substantial margins in most cases.

2. **Dataset-specific exceptions**: Dispersion Entropy excels for Gait and EMG; Permutation Entropy achieves the highest overall accuracy on PPG.

3. **Channel selection matters**: For multi-channel datasets, specific channels consistently outperform others (deltoid for EMG, F4 for GAMEEMO, Fz for Mental Arithmetic, SpO2 diff for Apnea).

4. **Accuracy range**: Best accuracies range from 73.6% (Mental Arithmetic) to 95.5% (PPG).

5.6 Chapter Summary

This chapter presented a comprehensive evaluation of eight entropy methods across eleven biomedical datasets using five machine learning classifiers. The experimental results, supported by effect size analysis, yield several important conclusions that are summarized below.

5.6.1 Entropy Method Performance

Bubble Entropy emerged as the dominant entropy method, achieving the highest mean accuracy (83.6% $\pm$ 9.0%) across the eleven datasets. It won 8 out of 11 datasets and outperformed every other method in head-to-head pairwise comparisons. With the lowest coefficient of variation among all methods (CV = 0.107), it demonstrated not only superior accuracy but also the most consistent performance across diverse biomedical signal types.

Dispersion Entropy ranked second overall with a mean accuracy of 77.7% $\pm$ 8.6%, followed by Approximate Entropy in third place (77.4% $\pm$ 8.5%). Dispersion Entropy proved to be the superior method for motor signal classification, outperforming

Bubble Entropy in both Gait (75.3% vs 69.9%) and EMG (80.5%) datasets. Permutation Entropy placed fourth (76.6% ± 11.5%) but achieved the highest single-dataset accuracy across the entire study—95.5% on PPG signals—demonstrating that ordinal pattern analysis is particularly effective for activity classification from photoplethysmographic recordings. Traditional information-theoretic measures, Shannon and Renyi Entropy, consistently occupied the lower performance tier.

The performance hierarchy can be summarized into three tiers. The top tier consists of Bubble Entropy alone, with a mean accuracy approximately 6 percentage points above the second-best method. The middle tier includes Dispersion, Approximate, Permutation, and Sample Entropy. The lower tier contains Distribution, Renyi, and Shannon Entropy. The gap between the top and lower tiers highlights the advantage of complexity-based entropy measures that capture temporal dynamics over those that rely on simple amplitude distributions.

5.6.2 Dataset-Specific Findings

Performance varied substantially across datasets, with best accuracies ranging from 73.6% (Mental Arithmetic) to 95.5% (PPG). Cardiovascular signals proved most amenable to entropy-based classification, with ECG Arrhythmia (93.8%), CHF vs NSR (92.2%), Apnea (91.6%), and Fantasia (90.0%) all exceeding 90% accuracy. These results suggest that cardiac dynamics exhibit robust entropy signatures that reliably distinguish pathological from healthy states.

Three notable exceptions to Bubble Entropy's dominance were identified. Dispersion Entropy outperformed all other methods for both Gait (Parkinson's detection, 75.3%) and EMG (neuromuscular disorder classification, 80.5%), indicating that its symbolic dynamics approach, which maps signal amplitudes to discrete classes, better captures the motor control abnormalities characteristic of neuromuscular conditions. Permutation Entropy achieved 95.5% on PPG activity classification, the highest accuracy in the entire study, using only a single feature (`permen_m2`), which demonstrates that ordinal patterns at the lowest embedding dimension are sufficient to distinguish sitting from active states in photoplethysmographic signals.

Channel and signal variant selection proved critical for multi-channel datasets. For EMG, the deltoid muscle provided substantially more discriminative features than the biceps. For GAMEEMO emotion classification, the F4 electrode was optimal

among the tested channels. For Mental Arithmetic cognitive load classification, the frontal Fz electrode yielded the best results. For Apnea detection, the SpO2 difference signal outperformed the raw SpO2 signal. These findings highlight that entropy-based classification is sensitive to channel selection, and the choice of measurement site can be as important as the choice of entropy method.

5.6.3 Classifier Performance

Among the five classifiers evaluated, Support Vector Machine (SVM) proved to be the most effective, achieving the best performance in 8 out of 11 datasets when paired with the best entropy method. k-Nearest Neighbors (kNN) was optimal for 2 datasets (PPG and Mental Arithmetic), while Random Forest (RF) excelled only for the GAMEEMO emotion classification task. Gaussian Naive Bayes and Gradient Boosting did not achieve the best result on any dataset.

A particularly noteworthy finding is the entropy-dependent classifier preference. When considering all entropy-dataset combinations (8 entropy methods $\times$ 11 datasets = 88 combinations), kNN achieved the highest win rate (58.0%), followed by SVM (22.7%). However, for Bubble Entropy specifically, SVM became the preferred classifier, yielding the best overall combination—Bubble Entropy with SVM—which achieved the highest mean accuracy (82.4%). Bubble Entropy occupied the top four positions and five of the top eight in the entropy-classifier combination ranking, appearing with all five classifiers, which confirms that its superior discriminative power is robust to classifier choice.

The consistent underperformance of ensemble methods (Random Forest, Gradient Boosting) suggests that entropy features create relatively simple, well-separated decision boundaries in feature space. These boundaries are effectively exploited by instance-based (kNN) and margin-maximizing (SVM) approaches. The added complexity of ensemble methods provides no benefit here and may even lead to overfitting given the small-to-moderate sample sizes typical of biomedical datasets.

5.6.4 Class Separation and Effect Size Analysis

Cohen's d effect size analysis provided insight into why certain entropy methods outperform others. Most methods exhibited mean effect sizes in the large range ($d \geq 0.8$), with Bubble Entropy achieving the highest mean effect size (1.172) across all

datasets. The maximum effect sizes were substantial for all methods: Approximate Entropy achieved the highest single maximum (3.090), followed by Renyi Entropy (2.740) and Shannon Entropy (2.630). Shannon Entropy had the lowest mean effect size (0.744), the only method remaining in the medium range, consistent with its lower classification accuracy.

This pattern reveals that Bubble Entropy's classification advantage stems from consistently strong class separation across datasets, as reflected in its highest mean effect size. The combination of reliably large effect sizes across diverse signal types enables classifiers—particularly SVM—to find highly discriminative decision boundaries. This also explains the effectiveness of the exhaustive feature selection approach: by testing all feature combinations, the methodology identifies these high-separation features that might be missed by filter-based selection methods.

The relationship between effect size and classification accuracy is not strictly linear, as classifiers can exploit combinations of features with moderate individual effect sizes, and non-linear classifiers like SVM with RBF kernel can achieve high accuracy even when linear separation is limited. Nevertheless, the effect size analysis provides a valuable physiological interpretation of entropy discriminability, complementing the accuracy-based evaluation with a measure of how well each method captures the underlying differences between pathological and healthy signal dynamics.

5.6.5 Feature Selection Observations

The exhaustive feature selection revealed informative patterns about optimal feature subsets. The number of selected features varied from $k = 1$ (PPG with Permutation Entropy) to $k = 5$ (most Bubble Entropy datasets), suggesting that simpler classification tasks require fewer features while complex tasks benefit from richer feature representations. For Bubble Entropy, the feature $m = 2$ appeared most frequently across datasets (5 of 8 winning datasets), indicating that the lowest non-trivial embedding dimension captures fundamental complexity differences. However, higher-dimensional features ($m = 10$ to $m = 27$) were also consistently selected, suggesting that Bubble Entropy captures complementary information at different temporal scales.

For Dispersion Entropy in motor signal datasets, the selected features used low embedding dimensions ($m = 2$ to $m = 4$) with varied class counts ($c = 3$ to $c = 9$), indicating that the interplay between temporal ordering and amplitude discretization

is more informative than either factor alone. The diversity of selected parameters across datasets reinforces the value of the exhaustive search approach over fixed-parameter evaluation.

Chapter 6

Conclusions

6.1 Summary of Findings

6.2 Key Contributions

6.3 Limitations

6.4 Future Work

6.5 Closing Remarks

This chapter summarizes the key findings of this thesis, highlights the main contributions, discusses limitations, and proposes directions for future research.

6.1 Summary of Findings

This thesis presents a systematic comparison of eight entropy methods across eleven diverse biomedical datasets, evaluated using five machine learning classifiers within a unified experimental framework. The central finding is that Bubble Entropy, a relatively recent, parameter-free method, consistently outperforms all other entropy measures, achieving 8 wins across 11 datasets with a mean accuracy of 83.6% $\pm$ 9.0%.

However, no single method proved to be universally dominant. Dispersion Entropy proved superior for motor signal classification—specifically, Gait and EMG

datasets—while Permutation Entropy achieved the highest single-dataset accuracy in the entire study (95.5% on PPG) using only a single feature. These exceptions reveal that the optimal entropy method depends on the underlying physiological dynamics of the signal being analyzed.

On the classifier side, simpler models outperformed complex ensembles. kNN and SVM consistently achieved the best results, while Gradient Boosting never won a single entropy-dataset combination. The optimal classifier depended on the entropy method: traditional measures favored kNN, whereas Bubble Entropy paired best with SVM. The effect size analysis using Cohen's d provided a complementary perspective, showing that Bubble Entropy achieves the highest mean class separation among all methods, which explains its consistently superior classification performance.

6.2 Key Contributions

This thesis contributes to the field of biomedical signal analysis in the following ways:

- **Comprehensive benchmark**: Systematic comparison of eight entropy methods across eleven datasets spanning seven distinct signal types.
- **Empirical validation of Bubble Entropy**: Demonstration that Bubble Entropy outperforms established measures for the majority of classification tasks, while identifying specific exceptions where Permutation Entropy (PPG) and Dispersion Entropy (Gait, EMG) are preferable.
- **Classifier recommendations**: Evidence that simple classifiers such as kNN and SVM are sufficient for entropy-based feature spaces, reducing computational requirements without sacrificing accuracy.
- **Effect size analysis**: Application of Cohen's d to quantify class separation power across entropy methods, providing a physiological interpretation of discriminability that complements standard accuracy-based evaluation.

6.3 Limitations

Several limitations should be considered when interpreting these results. Some datasets contained a limited number of subjects (e.g., Fantasia with 40), which may affect the generalizability of the reported rankings despite the use of cross-validation. All entropy methods were evaluated using default or commonly recommended parameters; systematic parameter optimization could yield different performance hierarchies. The evaluation was restricted to binary classification tasks, and multi-class scenarios may exhibit different entropy method preferences. Furthermore, this study evaluated each entropy method independently rather than exploring combinations of multiple entropy families, which could potentially improve classification performance. Finally, the comparison focused on traditional machine learning classifiers, and the integration of entropy features with deep learning architectures remains an open question.

6.4 Future Work

Based on the findings and limitations of this thesis, several directions for future research are proposed:

1. **Multi-scale entropy analysis**: Investigate whether combining entropy values across multiple scales improves classification performance.

2. **Parameter optimization**: Develop automated methods for entropy parameter selection tailored to specific signal characteristics.

3. **Feature fusion**: Explore combinations of multiple entropy methods to leverage complementary information.

4. **Deep learning integration**: Evaluate entropy features as inputs to deep neural networks, potentially capturing non-linear relationships.

5. **Real-time implementation**: Develop computationally efficient implementations for wearable and point-of-care devices.

6. **Clinical validation**: Conduct prospective clinical studies to validate the diagnostic utility of entropy-based biomarkers.

7. **Extended signal types**: Apply the evaluation framework to additional biomedical signals.

6.5 Closing Remarks

This thesis demonstrated that entropy-based features provide effective biomarkers for biomedical signal classification, with Bubble Entropy emerging as the recommended default choice. However, the finding that no single method universally dominates underscores the importance of signal-specific evaluation. The comprehensive benchmark, effect size analysis, and visualization tools developed in this work provide a foundation for researchers and practitioners to make informed decisions when applying entropy methods to biomedical signal analysis.

Bibliography

[1] C. E. Shannon, "A mathematical theory of communication," *Bell System Technical Journal*, vol. 27, no. 3, pp. 379–423, 1948.

[2] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, no. 6, pp. 2297–2301, 1991.

[3] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[4] C. Bandt and B. Pompe, "Permutation entropy: A natural complexity measure for time series," *Physical Review Letters*, vol. 88, no. 17, p. 174102, 2002.

[5] M. Rostaghi and H. Azami, "Dispersion entropy: A complexity measure for time series," *IEEE Signal Processing Letters*, vol. 23, no. 5, pp. 610–614, 2016.

[6] P. Li, C. Liu, K. Li, D. Zheng, C. Liu, and Y. Hou, "Assessing the complexity of short-term heartbeat interval series by distribution entropy," *Medical & Biological Engineering & Computing*, vol. 53, no. 1, pp. 77–87, 2015.

[7] G. Manis, M. Aktaruzzaman, and R. Sassi, "Bubble entropy: An entropy almost free of parameters," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 11, pp. 2711–2718, 2017.

[8] A. Rényi, "On measures of entropy and information," ., vol. 1, pp. 547–561, 1961.

[9] H. Xie, W. He, H. Liu, and H. Zhao, "Improved bubble entropy and its application to biomedical signal analysis," *Biomedical Signal Processing and Control*, vol. 52, pp. 181–188, 2019.

[10] N. Iyengar, C.-K. Peng, R. Morin, A. L. Goldberger, and L. A. Lipsitz, "Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics," *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology*, vol. 271, pp. R1078–R1084, 1996.

[11] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, rRID:SCR_007345.

[12] G. B. Moody and R. G. Mark, "The impact of the MIT-BIH arrhythmia database," *IEEE Engineering in Medicine and Biology Magazine*, vol. 20, no. 3, pp. 45–50, 2001.

[13] D. S. Baim, W. S. Colucci, E. S. Monrad, H. S. Smith, R. F. Wright, A. Lanoue, D. F. Gauthier, B. J. Ransil, W. Grossman, and E. Braunwald, "Survival of patients with severe congestive heart failure treated with oral milrinone," *Journal of the American College of Cardiology*, vol. 7, no. 3, pp. 661–670, 1986.

[14] S. D. Greenwald, "Development and analysis of a ventricular fibrillation detector," Master's thesis, MIT Dept. of Electrical Engineering and Computer Science, 1986.

[15] A. Messekher, S. Mekaoui, M. Kedir-Talha, M. I. Kediha, and F. Mostefaoui, "Electromyographic data recordings during isometric contractions of the biceps brachii and deltoid collected from patients and healthy subjects," 2023.

[16] G. Juliá-Serdá, J. Navarro-Esteva, and A. G. Ravelo-García, "APNEA HRV+SPO2 DATASET," 2023.

[17] T. B. Alakus, M. Gonen, and I. Turkoglu, "Database for an emotion recognition system based on EEG signals and various computer games – GAMEEMO," *Biomedical Signal Processing and Control*, vol. 60, p. 101951, 2020.

[18] I. Zyma, S. Tukaev, I. Seleznov, K. Kiyono, A. Popov, M. Chernykh, and O. Shpenkov, "Electroencephalograms during mental arithmetic task performance," *Data*, vol. 4, no. 1, p. 14, 2019.

[19] U. Cesari, G. De Pietro, E. Marciano, C. Niri, G. Sannino, and L. Verde, "A new database of healthy and pathological voices," *Computers & Electrical Engineering*, vol. 68, pp. 310–321, 2018.

[20] P. Mehrgardt, M. Khushi, S. Poon, and A. Withana, "Pulse transit time PPG dataset," 2022.

[21] M. W. Flood and B. Grimm, "EntropyHub: An open-source toolkit for entropic time series analysis," *PLoS ONE*, vol. 16, no. 11, p. e0259448, 2021.

[22] R. Vallat, "Antropy: Entropy and complexity of (eeg) time-series in Python," 2021, gitHub repository. [Online]. Available: https://github.com/raphaelvallat/antropy

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg *et al.*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

Short Biography

Andreas Matsias was born in Preveza, Greece, in 1997. He completed his undergraduate studies in the Department of Computer Science & Engineering at the University of Ioannina in 2021, with a focus primarily on Software Engineering. He subsequently completed his Master's degree in the Department of Computer Science and Engineering at the University of Ioannina in 2026, specializing in Data Science and Engineering. His research interests include data analysis, machine learning, and explainability in machine learning models.

From November 2022 to October 2025, he worked as a full-stack engineer at Deloitte, where he was involved in the design, development, and delivery of full-stack software solutions for enterprise clients. Since October 2025, he has been working as a Software Engineer at Plum, a fintech company, contributing to the development of intelligent financial products and services.