

Enhancing Microsleep Prediction Performance Through Temporal Modeling, Multimodal Fusion, and Label-Efficient Learning

Maram Awad Abdelrahman Mohamed, B.Eng.,
M.Sc.

A dissertation submitted in fulfilment of the requirements for
the award of Doctor of Philosophy (Ph.D.)

to the



Dublin City University

School of Computing

Supervised by

Prof. Noel E. O'Connor, Prof. Tomas Ward

August 2026

Declaration

I hereby certify that this material, which I now submit for assessment on the programme of study leading to the award of Doctor of Philosophy (Ph.D.), is entirely my own work, and that I have exercised reasonable care to ensure that the work is original and have conformed to the regulations on the use and declaration of Generative Artificial Intelligence (Gen AI), and does not, to the best of my knowledge, breach any law of copyright, and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of this thesis.

I further declare that all research questions, study design, experimental methodology, data processing, statistical analyses, model development, interpretation of results, discussion, and conclusions presented in this thesis are my own original work. Any use of Generative Artificial Intelligence tools was limited to approved supportive tasks as declared in the following Generative AI Statement and did not contribute to the intellectual content, scientific findings, or scholarly contributions of this research.

Signed (Typed Name): *Maram A. Mohamed*

Student ID: 22268933

Date: August 27, 2026

Generative AI Statement

Use of Generative AI in this Thesis

In accordance with Dublin City University's regulations on the use and declaration of Generative Artificial Intelligence (Gen AI) tools and the associated guidance for postgraduate research students, I declare the following use of Gen AI tools during the course of this PhD research and the preparation of this thesis.

Tools

The following Gen AI tools were used during different stages of the PhD research and thesis preparation:

- Google Gemini (various versions), used between March 2024 and January 2026.
- ChatGPT (OpenAI), used between June 2023 and May 2026.
- Claude (Anthropic), used between January 2025 and May 2026.

Outputs

Two distinct phases of use are declared:

Research Support Phase (June 2023 – January 2026)

During the research and experimentation stages of the PhD, Gen AI tools were used primarily for technical and administrative support, including:

- Assistance with Python programming, debugging, and code optimisation.
- Troubleshooting software implementation issues and resolving technical errors.
- Assistance with LaTeX formatting and document preparation.
- Identification of potentially relevant literature for subsequent independent verification and review.
- General technical discussions related to machine learning, deep learning, signal processing, data analysis, and software implementation.

Thesis Preparation Phase (February 2026 – May 2026)

During the writing and preparation of the thesis, Gen AI tools were used in a limited supportive capacity for:

- Language refinement, grammar checking, and improvement of sentence clarity.
- Assistance with LaTeX formatting, troubleshooting, and document organisation.
- Generation of three methodological illustration figures based entirely on author-provided specifications, descriptions, and scientific content.

Prompts

Typical prompts requested assistance with:

- Correcting LaTeX errors and improving document formatting.
- Identifying potential coding issues and suggesting debugging strategies.
- Improving the academic style of selected sentences and paragraphs.
- Creating methodological diagrams from detailed author-provided descriptions and specifications.

Iterations

The tools were used on multiple occasions throughout the PhD. Outputs generated by the tools were critically reviewed, modified where necessary, independently verified, and either accepted or rejected by the author. No AI-generated output was incorporated into the thesis without human review and validation. All technical suggestions, and generated materials were assessed independently before use.

Critical Reflection

The use of Gen AI tools was restricted to technical, organisational, and formatting support activities. These tools did not contribute to the conception of the research questions, experimental design, data collection, feature engineering, model development, statistical analyses, interpretation of results, scientific discussion, or conclusions.

All scientific contributions, intellectual content, methodological decisions, analyses, interpretations, and conclusions presented in this thesis are the author's own original work. Any information obtained through Gen AI tools was independently verified before use. Full responsibility for the content of this thesis remains with the author.

Dedication

Dedicated with gratitude and faith

First and foremost, I dedicate this thesis to Allah, the Most Gracious and the Most Merciful, for His endless kindness, protection, mercy, and guidance throughout every step of this journey. Alhamdulillah for granting me strength during moments of weakness, hope during uncertainty, and opportunities beyond what I could have imagined. Being far away from my family and home was never easy, but by His mercy and guidance, I was able to continue and reach this milestone.

I dedicate this work to my beloved family, my first line of support and the foundation behind every achievement in my life. To my parents, whose sacrifices, encouragement, and unwavering belief in me gave me the courage to pursue education abroad and strive for a better future. Their support and prayers have always been my greatest source of strength.

To my siblings, for their endless love, encouragement, and support. Thank you for always believing in me, motivating me, and seeing me as someone worthy of admiration. Your love has carried me through many difficult moments.

Finally, and most especially, I dedicate this thesis to the souls of the martyrs who lost their lives during the current war in Sudan while defending our country and our people. Their sacrifices can never be repaid or fully expressed in words. They gave their souls for the sake of Allah and so that we may one day return safely to a homeland we can once again call home. May Allah grant them the highest levels of Jannah, and may Sudan one day know peace, safety, and justice again.

Acknowledgements

“In the name of Allah, the Most Gracious, the Most Merciful.”

Alhamdulillah, all praise and thanks are due to Allah for His endless mercy, protection, guidance, and blessings throughout this journey. Without His will and support, this achievement would never have been possible.

I would like to express my deepest gratitude to all those who supported me throughout my PhD journey. I am sincerely thankful to Dublin City University and the Insight Research Ireland Centre for Data Analytics for providing a welcoming, supportive, and inspiring research environment. I am also grateful to Science Foundation Ireland (SFI) for funding and supporting this research journey and to Ireland for granting me the opportunity to pursue my studies and grow both academically and personally.

I would like to extend my heartfelt appreciation to all my supervisors who contributed to my PhD journey from its very beginning until this stage, including supervisors from academia and industry at Xperi and Tobii. Every one of them contributed in different ways to the researcher and person I am today, and I remain deeply grateful for their time, expertise, and guidance.

In particular, I would like to express my deepest gratitude and appreciation to Prof. Noel O’Connor for his exceptional support, generosity, guidance, and belief in me throughout some of the most difficult stages of my PhD journey. His continuous encouragement, understanding, and willingness to help truly saved my PhD on multiple occasions, and I will always remain grateful for his mentorship and kindness.

I would also like to sincerely thank all co-supervisors, mentors, and senior researchers who supported and guided me throughout this work. I am especially grateful to Dr. Peter Redmond for his continuous support, generosity with his time, insightful guidance, and readiness to help whenever needed.

My sincere appreciation also goes to my colleagues and friends in the lab for creating a supportive, encouraging, and positive environment that made this long journey easier and more enjoyable. I am equally grateful to my friends back home and around the world who supported me remotely, listened to me during difficult moments, and helped me preserve my mental well-being throughout this journey.

A very special thanks goes to Dr. Amro for his endless support, guidance, understanding, and kindness, and for the important role he played in supporting my mental health during challenging periods.

Finally, I would like to express my deepest love and gratitude to my family. To my parents, for their unconditional love, sacrifices, prayers, and constant encouragement; to my siblings for their endless support and belief in me; and to my uncle and cousin for their care and encouragement throughout the years. Their love and support gave me strength during the hardest moments of this journey.

To everyone who supported me directly or indirectly, thank you. This achievement belongs not only to me, but also to all those who stood beside me along the way.

List of Publications

Journal Publications

- Submitted / Under Review

1. **Maram A. Mohamed**, Prateek Mathur, Peter Redmond, Tomas Ward, and Noel E. O'Connor. "Mutual-Information-Guided EEG–EOG Fusion for Interpretable Microsleep Prediction During the Maintenance of Wakefulness Test." Under Review at *IEEE Access*, 2026.
2. **Maram A. Mohamed**, Noel E. O'Connor, and Hossein Javidnia. "Deep Learning in Driver Safety: A Literature Review on Detecting Microsleep Episodes Using EEG Signals." Under Review at *African Journal of Engineering and Technology (Open Science)*, 2025.

Conference Publications

- Papers (published)

1. **Maram A. Mohamed**, Mayug Maniparambil, Peter Redmond, Tomas Ward, and Noel E. O'Connor. "Attention-Based Temporal Resolution Enhancement for Microsleep Forecasting Across Multiple Horizons." *33rd International Conference on Artificial Intelligence and Cognitive Science (AICS 2025)*, Ireland, 2025.
2. **Maram A. Mohamed**, Mayug Maniparambil, Peter Redmond, Tomas Ward, and Noel E. O'Connor. "Comprehensive Evaluation of EEG Features for Microsleep Detection: Statistical Validation and Methodological

Improvements.” *33rd International Conference on Artificial Intelligence and Cognitive Science (AICS 2025)*, Ireland, 2025.

- Accepted

1. **Maram A. Mohamed**, Prateek Mathur, Peter Redmond, Tomas Ward, and Noel E. O’Connor. “Label-Efficient Multimodal Microsleep Forecasting via Interpretable Disagreement-Driven Active Learning.” Accepted to *MSB EMERGE 2026 Workshop at MICCAI*, 2026.

Contents

List of Publications	IX
List of Figures	XVI
List of Tables	XXVI
Abstract	XXX
1 Introduction	1
1.1 Background and Motivation	1
1.2 Problem Statement	3
1.3 Research Aim	5
1.4 Research Questions	5
1.5 Research Contributions	6
1.5.1 Contribution 1: Physiologically Meaningful Feature Represen- tation	7
1.5.2 Contribution 2: Sequence-Sensitive Temporal Forecasting . . .	7
1.5.3 Contribution 3: Multimodal Physiological Fusion and Direc- tionality Analysis	7
1.5.4 Contribution 4: Label-Efficient Multimodal Learning	7
1.5.5 Contribution 5: Integrated Perspective on Microsleep Fore- casting	8
1.6 Significance of the Research	8
1.7 Relationship Between the Research Questions	9

1.8	Thesis Structure	9
1.9	Chapter Summary	11
2	Foundations and Related Work	12
2.1	Introduction	12
2.2	Microsleeps and Vigilance Degradation	14
2.3	EEG and EOG Signals in Microsleep Analysis	15
2.4	Neurophysiological Basis of Microsleeps	18
2.5	EEG-Based Microsleep Detection and Forecasting	20
2.5.1	Traditional EEG-Based Detection Methods	20
2.5.2	Deep Learning for Microsleep Detection and Forecasting	21
2.6	Multimodal Microsleep Detection and Prediction	24
2.7	Active Learning and Label-Efficient EEG Analysis	26
2.7.1	Motivation for Active Learning in EEG	26
2.7.2	Active Learning in EEG and Brain–Computer Interfaces	27
2.7.3	Multimodal Learning and Active Learning	27
2.7.4	Active Learning for Fatigue and Microsleep Monitoring	28
2.8	Current Challenges in Microsleep Forecasting	29
2.9	Research Gaps and Thesis Motivation	31
2.10	Summary	33
3	Dataset and Evaluation Framework	35
3.1	Introduction	35
3.2	Datasets	36
3.2.1	Primary Dataset: Maintenance of Wakefulness Test (MWT)	36
3.2.2	Bern Scoring Criteria and Label Definition	37
3.2.3	Additional Datasets	40
3.3	Signal Preprocessing Pipeline	41
3.3.1	Filtering	42
3.3.2	Artifact Handling	42

3.3.3	Epoching Strategy	43
3.4	Feature Construction	44
3.4.1	Handcrafted Features (RQ1)	44
3.4.2	Time-Series Input for Temporal Models (RQ2)	45
3.4.3	Multimodal Representation (RQ3)	45
3.4.4	Active Learning Sampling Pool (RQ4)	45
3.5	Experimental Protocol	46
3.5.1	Train–Validation–Test Split	46
3.5.2	Class Imbalance Handling	47
3.5.3	Evaluation Metrics	47
3.6	Statistical Analysis	48
3.7	Implementation Details	49
3.8	Reproducibility Statement	50
3.9	Chapter Summary	50
4	Feature Representation for Microsleep Prediction	52
4.1	Experimental Design and Reproducibility Details	54
4.2	Systematic Evaluation of EEG Feature Families	55
4.2.1	Time-domain descriptors	55
4.2.2	Frequency-domain descriptors	56
4.2.3	Bispectrum descriptors	56
4.2.4	Nonlinear dynamics descriptors	57
4.2.5	Time-frequency descriptors	57
4.2.6	Intensity-weighted spectral descriptors	58
4.3	Classifier Benchmarking	58
4.3.1	Feature-family performance under representative classifiers . .	62
4.3.2	Model performance across feature families	63
4.4	Statistical Reporting	64
4.5	Identification of Robust Predictive Indicators	66
4.6	Interpretation of Static Feature Results	68

4.7	Limitations of Static Feature Modelling	69
4.8	RQ1 Findings on EEG Indicators of Microsleep	70
5	Temporal Modelling for Microsleep Forecasting	72
5.1	Introduction	72
5.2	Research Question and Chapter Scope	74
5.3	Related Work	74
5.3.1	Feature Extraction for Microsleep Analysis	75
5.3.2	Forecasting Models and the Role of Temporal Dependence	76
5.4	Methodology	76
5.4.1	Datasets	76
5.4.2	Problem Formulation	77
5.4.3	Baseline Replication: CSP + SBL	78
5.4.4	GRU Baseline	79
5.4.5	Proposed Attention-Augmented GRU	79
5.4.6	Feature Representation	80
5.4.7	Additional Comparator: EEG-TCNet Adaptation	81
5.4.8	Loss Function	81
5.4.9	Evaluation Metrics	82
5.4.10	Implementation Details	82
5.5	Results	83
5.5.1	GRU Baseline versus Attention-Augmented GRU	83
5.5.2	Comparison with the CSP+SBL Forecasting Baseline	84
5.5.3	Evaluation on the Sleepy Driver Dataset	84
5.5.4	Illustrative EEG and Feature Trends	85
5.5.5	Ablation Study	87
5.5.6	Attention Visualisation	87
5.5.7	Latency Analysis	88
5.5.8	Comparison with EEG-TCNet	88
5.6	Discussion	89

5.6.1	Temporal Degradation and the Value of Causal Attention . . .	89
5.6.2	Interpretation of the Ablation Results	90
5.6.3	Cross-Dataset Behaviour	91
5.6.4	Practical Implications	91
5.7	Chapter Summary	91
6	Multimodal EEG–EOG Fusion for Forecasting	93
6.1	Introduction	93
6.2	Related Work	94
6.3	Methodology	96
6.3.1	Dataset	96
6.3.2	Feature Extraction	97
6.3.3	Fusion Strategies	99
6.3.4	Classifier	101
6.3.5	Statistical Analysis	101
6.3.6	Experimental Design	102
6.4	Results	103
6.4.1	Stage 1: Unimodal Baselines	103
6.4.2	Stage 2: MI-Enhanced Unimodal Models	104
6.4.3	Stage 3: Fusion Architectures	104
6.4.4	Feature Importance Analysis	105
6.4.5	Stage 4: Transfer Entropy Augmentation	107
6.4.6	Stage 5: Statistical Validation	108
6.5	Discussion	109
6.5.1	Value of Multimodal Neural–Ocular Fusion	109
6.5.2	Interpretation of MI and TE Findings	111
6.5.3	Methodological Implications	112
6.6	Chapter Summary	112

7	Active Learning for Label-Efficient Forecasting	114
7.1	Introduction	114
7.2	Research Question and Chapter Scope	115
7.3	Methodology	116
7.3.1	Baseline Model	116
7.3.2	Dataset and Forecasting Protocol	116
7.3.3	Feature Representation	117
7.3.4	Active Learning Framework	117
7.3.5	Standard Query Strategies	118
7.3.6	Proposed Multimodal Disagreement Active Learning	119
7.3.7	Ablation Design	120
7.3.8	Evaluation Metrics and Statistical Analysis	120
7.4	Results	120
7.4.1	Learning Curves Including the Proposed Method	120
7.4.2	Label-Efficiency Including the Proposed Method	122
7.4.3	Final Performance Including the Proposed Method	123
7.4.4	Benchmark Strategy Behaviour	124
7.4.5	Statistical Validation	126
7.4.6	Ablation Study of the MD-AL Query Score	132
7.5	Discussion	133
7.6	Chapter Summary	136
8	Cross-Study Synthesis and Design Insights	137
8.1	Introduction	137
8.2	Restatement of the Thesis Aim	138
8.3	Synthesis of Findings Across the Research Questions	139
8.3.1	RQ1: Feature Representation as the Foundation of Prediction	139
8.3.2	RQ2: Temporal Modelling Improves Forecasting Beyond Static Detection	139

8.3.3	RQ3: Multimodal Fusion Adds Complementary Physiological Information	140
8.3.4	RQ4: Label Efficiency Can Be Improved Without Sacrificing Performance	141
8.4	Cross-Cutting Themes	142
8.4.1	Microsleep Prediction is a Progressive Modelling Problem . . .	142
8.4.2	Interpretability Matters Alongside Accuracy	142
8.4.3	Multimodality Improves Both Prediction and Data Efficiency .	143
8.4.4	The Main Constraint is Not Only Model Capacity	143
8.5	Overall Contributions of the Thesis	143
8.6	Concluding Remarks	144
9	Conclusions and Future Work	146
9.1	Thesis-Wide Contributions	146
9.2	Cross-RQ Synthesis	147
9.3	Future Work	148
9.3.1	Larger and More Diverse Multimodal Datasets	148
9.3.2	Dynamic Physiological Interaction Modelling	149
9.3.3	Advanced Multimodal Learning	149
9.3.4	Real-Time and Embedded Deployment	150
9.4	Closing Statement	150
	Bibliography	150

List of Figures

1.1	Illustrative EEG frequency bands and their broad associations with vigilance and sleep. Beta activity is commonly associated with active wakefulness, alpha activity with relaxed wakefulness, theta activity with drowsiness and light sleep, and delta activity with deeper sleep. Microsleep does not constitute a separate EEG frequency band; instead, it is commonly associated with a transition involving increased theta activity, reduced or poorly defined alpha activity, and reduced beta activity relative to alert wakefulness. The displayed waveforms are schematic and are intended to illustrate differences in frequency rather than represent recorded EEG segments.	2
3.1	Simplified RQ1 pipeline for microsleep detection using hand-crafted EEG features. After preprocessing and one-second windowing, descriptive EEG features are extracted and used for classifier-based prediction.	36

3.2	Example of microsleep-related events evaluated according to the Bern criteria. The blue vertical lines delimit a microsleep episode candidate (MSEc), and the red vertical lines delimit a microsleep episode (MSE). E1 and E2 show the EOG activity, O1 and O2 show the occipital EEG signals, and the lower panels present the corresponding O1 and O2 time–frequency profiles. During the MSEc interval, alpha activity becomes partially attenuated, but not all mandatory criteria for an MSE are satisfied. During the MSE interval, the occipital EEG demonstrates the principal electrophysiological characteristics of the Bern criteria: predominant irregular theta activity, poorly defined or absent alpha activity relative to wakefulness, and a simultaneous reduction in beta activity. The EOG channels provide complementary evidence of changes in eye-movement behaviour. The event boundaries and classifications were determined from the expert annotations supplied with the dataset.	39
4.1	Nemenyi post-hoc test heatmaps showing pairwise comparisons for accuracy, sensitivity, and specificity across classifiers. Lower values indicate significant differences between classifier pairs.	66
4.2	Top 20 features ranked by importance for the Random Forest classifier under the RQ1 configuration.	67
4.3	Confusion matrices for different classifiers using nonlinear features. . .	68
5.1	Illustration of the microsleep forecasting problem. A sequence of EEG feature vectors $\mathbf{X} = [x_1, \dots, x_T]$ is used to estimate the probability of a microsleep event occurring after forecasting horizon τ . When $\tau = 0$, the task corresponds to onset detection; positive values of τ represent advance forecasting.	78
5.2	Full pipeline: causal multi-head self-attention is applied before the GRU to enrich temporal feature representations.	82

5.3	ROC and precision–recall curves for onset prediction	83
5.4	Illustrative frequency-domain changes surrounding microsleep episodes for an example subject. Panel A shows the alpha (8–12 Hz), beta (12–26 Hz), theta (4–8 Hz), and delta (0.5–4 Hz) band powers. Panel B shows the $\theta/(\alpha + \beta)$ ratio, with the red dashed line indicating the detection threshold of 0.22. Panel C shows the median frequency. Red shaded regions indicate microsleep intervals, while blue shaded regions indicate the one-second prediction windows preceding them. Purple regions represent temporal overlap between prediction windows and microsleep intervals. The blue arrows and “1.0 s Forecast” annotations indicate the forecast interval. Microsleep periods are characterised by an increased $\theta/(\alpha + \beta)$ ratio and a reduction in median frequency.	86
5.5	Raw EEG around microsleep episodes	87
5.6	Attention-weight heatmap across EEG features and five temporal input steps. Brighter colours indicate relatively greater model emphasis, whereas darker colours indicate lower relative emphasis. Several features exhibit alternating temporal patterns: spectral entropy, the delta/theta ratio, mean amplitude, kurtosis, and Hjorth activity receive greater emphasis at time steps 1 and 3, while line length, peak frequency, the theta/alpha ratio, and the alpha/beta ratio are emphasised mainly at time steps 0 and 2. Other features, including sample entropy, skewness, and frontal asymmetry, remain comparatively stable across the sequence.	89

6.1	Computation and interpretation of mutual information (MI) and transfer entropy (TE). The left panel shows that MI quantifies symmetric shared information between paired EEG and EOG observations using their joint and marginal distributions. The right panel shows that TE quantifies directional predictive information by comparing prediction of the target signal’s future from its own history with prediction using both the target and source histories.	100
6.2	Experimental stages and study design for microsleep prediction. The framework progresses from unimodal baselines to intra-modal mutual information augmentation, multimodal fusion strategies, transfer entropy integration, and statistical validation, while maintaining a consistent LightGBM classifier across all stages.	102
6.3	Baseline EOG versus EOG+MI	105
6.4	Stage 1 and Stage 2 results	106
6.5	Fusion architectures	107
6.6	Feature importance for Early Fusion MI Cross model	108
6.7	Transfer entropy augmentation	109
6.8	Statistical model comparison	110
7.1	Active learning pipeline for RQ4	118
7.2	Active-learning performance for the 1.0s prediction horizon as the percentage of labelled training data increases. The horizontal axis represents the labelling budget, and the vertical axis reports AUC–PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline obtained using 100% labelled data. The horizontal dotted lines denote 90% and 95% of the baseline performance. The red curve represents the proposed MD–AL disagreement strategy, while the remaining curves represent Random, Uncertainty, Diversity, and Committee sampling. .	121

7.3 Active-learning performance for the 2.0s prediction horizon as the percentage of labelled training data increases. The horizontal axis represents the labelling budget, and the vertical axis reports AUC-PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline obtained using 100% labelled data, while the horizontal dotted lines indicate 90% and 95% of its performance. The red curve represents the proposed MD-AL disagreement strategy, and the remaining curves represent Random, Uncertainty, Diversity, and Committee sampling. Performance improves most strongly at the lowest labelling budgets, after which the methods largely converge towards the fully supervised baseline. MD-AL remains competitive across the evaluated budgets and slightly exceeds the baseline at the larger budgets, although no strategy is uniformly superior across the complete budget range. . . . 122

7.4 Minimum labelling budget required by each acquisition strategy to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 1.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with 1% labelled data. At the 95% target, Random sampling requires 2%, whereas the remaining strategies, including the proposed MD-AL method, require 1%. Recovering 100% of the baseline is substantially more demanding: Random requires 40%, Committee sampling requires 80%, and Uncertainty, Diversity, and MD-AL require the full labelling budget. The red dashed line indicates the 100% labelling level. . . . 123

7.5 Minimum labelling budget required by each acquisition strategy to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 2.0 s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas the remaining strategies, including the proposed MD–AL method, require 1%. Recovering 100% of baseline performance requires larger and more strategy-dependent budgets: Committee sampling requires 20%, Random, Uncertainty, and MD–AL each require 60%, and Diversity requires the full 100% labelling budget. The red dashed line indicates the full-labelling level. 124

7.6 Final AUC–PR performance at the 100% labelling budget for the 1.0 s prediction horizon. Bars show the mean performance of Random, Uncertainty, Diversity, Committee, and the proposed MD–AL disagreement strategies, together with the fully supervised early-fusion reference model; error bars indicate performance variability. The orange dashed line marks the early-fusion baseline value of 0.7580. MD–AL achieves the highest numerical mean AUC–PR (0.7590), followed closely by Random (0.7588), Committee (0.7584), and Uncertainty (0.7582), while Diversity obtains 0.7502. The substantial overlap between the error bars and the very small differences between the mean values indicate broadly comparable performance; therefore, the marginal numerical advantage of MD–AL should not be interpreted as evidence of clear statistical superiority. 125

7.7 Final AUC–PR performance at the 100% labelling budget for the 2.0s prediction horizon. Bars show the mean performance of Random, Uncertainty, Diversity, Committee, and the proposed MD–AL disagreement strategies, together with the fully supervised early-fusion reference model; error bars indicate performance variability. The orange dashed line marks the early-fusion baseline value of 0.7580. Random sampling achieves the highest numerical mean AUC–PR (0.7619), followed closely by MD–AL (0.7617), Committee (0.7614), and Uncertainty (0.7610). Diversity achieves 0.7559, slightly below the reference baseline. The extensive overlap between the error bars and the small differences between the mean values indicate broadly comparable performance; consequently, the numerical differences should not be interpreted as evidence that any strategy is statistically superior. . 126

7.8 AUC–PR as a function of the percentage of labelled training data for the benchmark active-learning strategies at the 1.0s prediction horizon. The horizontal axis represents the labelling budget, and the vertical axis reports AUC–PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline, while the horizontal dotted lines indicate 90% and 95% of baseline performance. All strategies exceed the 95% reference level at very small labelling budgets and improve rapidly up to approximately 10–20% labelled data. Random sampling briefly exceeds the baseline at an intermediate budget, whereas Random, Uncertainty, and Committee sampling converge towards the baseline as the budget increases. Diversity sampling remains slightly below the other strategies at the largest budgets. 127

7.9 AUC–PR as a function of the percentage of labelled training data for the benchmark active-learning strategies at the 2.0s prediction horizon. The horizontal axis represents the labelling budget, and the vertical axis reports AUC–PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline, while the horizontal dotted lines indicate 90% and 95% of baseline performance. All strategies exceed the 95% reference level at very small labelling budgets and improve most strongly within the first 10–20% of labelled data. Committee sampling provides the highest numerical performance at several intermediate budgets and slightly exceeds the baseline from approximately 20% onwards. Random and Uncertainty sampling recover gradually and converge with Committee sampling at the largest budgets, while Diversity remains marginally below the other strategies. 128

7.10 Minimum labelling budget required by the benchmark acquisition strategies to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 1.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas Uncertainty, Diversity, and Committee sampling require 1%. Recovering 100% of the baseline is considerably more demanding: Random requires 40%, Committee requires 80%, and Uncertainty and Diversity require the full 100% labelling budget. The red dashed line indicates the full-labelling level. 129

7.11 Minimum labelling budget required by the benchmark acquisition strategies to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 2.0 s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas Uncertainty, Diversity, and Committee sampling require 1%. Recovering 100% of the baseline requires substantially larger and strategy-dependent budgets: Committee sampling requires only 20%, Random and Uncertainty require 60%, and Diversity requires the full 100% labelling budget. The red dashed line indicates the full-labelling level. 130

7.12 Ablation study of the proposed MD-AL acquisition strategy for the 1.0 s prediction horizon. Bars report mean AUC-PR and error bars indicate performance variability. V1 uses disagreement only, V2 combines disagreement with the boundary criterion, and V3 is the complete proposed method. V3 achieves the highest numerical mean at the 10%, 40%, and 100% budgets, whereas V2 is marginally higher at 20%. Performance improves gradually for all variants as the labelling budget increases. However, the differences between the variants are small and their error bars overlap substantially, indicating that the additional components provide modest numerical improvements rather than clear evidence of statistical superiority. 133

7.13 Learning curves for the three MD–AL ablation variants across increasing labelling budgets at the 1.0s prediction horizon. The horizontal axis represents the percentage of labelled training data, and the vertical axis reports mean AUC–PR; shaded regions indicate performance variability. V1 uses disagreement only, V2 combines disagreement with the boundary criterion, and V3 is the complete proposed method. All variants improve rapidly at the lowest budgets and then increase more gradually as additional labels are introduced. The three curves remain closely aligned across most budgets, with only small and inconsistent numerical differences between the variants. 134

List of Tables

4.1	Performance of feature families using LDA. These point estimates should be interpreted jointly across accuracy, sensitivity, and specificity; high accuracy alone is not sufficient for microsleep detection. .	63
4.2	Performance of feature families using Random Forest. Time-frequency features achieve high accuracy and specificity, while frequency-domain, nonlinear, and intensity-weighted features provide more balanced sensitivity–specificity behaviour.	63
4.3	Detection performance using time-domain features across classifiers. .	64
4.4	Detection performance using frequency-domain features across classifiers.	64
4.5	Detection performance using bispectrum features across classifiers. . .	65
4.6	Detection performance using nonlinear features across classifiers. . . .	65
4.7	Detection performance using time-frequency (wavelet) features across classifiers.	65
4.8	Detection performance using intensity-weighted features (IWMF, IWBW) across classifiers.	65
5.1	GRU baseline and proposed model across forecast horizons	84
5.2	CSP+SBL baseline versus proposed model	85
5.3	Results on the Sleepy Driver dataset	85
5.4	Ablation study for attention and focal loss	88
5.5	Attention-GRU versus EEG-TCNet	90

6.1	Unimodal EEG and EOG baselines	103
6.2	Effect of MI on unimodal models	104
6.3	Fusion model comparison	105
6.4	MI Cross versus MI+TE	109
7.1	Performance summary at reduced budgets	131
7.2	Wilcoxon test at 20% labelling	131
7.3	Friedman test at 1.0s	132
7.4	Friedman test at 2.0s	132
7.5	Ablation study of MD-AL	133

Abstract

Maram Awad Abdelrahman Mohamed

Enhancing Microsleep Prediction Performance Through Temporal Modeling, Multimodal Fusion, and Label-Efficient Learning

Microsleep episodes are brief involuntary transitions into sleep-like states that occur during wakefulness and are strongly associated with fatigue, reduced vigilance, and cognitive overload. Although microsleeps may last only a few seconds, they can severely impair attention, reaction time, perception, and decision-making, posing substantial risks in safety-critical environments such as transportation, industrial manufacturing, healthcare monitoring, aviation, and military operations. Accurate prediction of microsleep onset before behavioural failure remains a major challenge due to the transient, heterogeneous, and low signal-to-noise nature of the underlying physiological processes. Existing approaches are further constrained by limited subject generalisation, heavy dependence on handcrafted features, and insufficient understanding of multimodal physiological interactions underlying vigilance decline. This thesis addresses these challenges through the development of robust EEG–EOG-based machine learning and deep learning frameworks for multimodal microsleep prediction, with a focus on physiologically meaningful representation learning, temporal modelling, multimodal interaction analysis, and practical real-world applicability.

This thesis investigates microsleep prediction through a sequence of studies examining physiological feature representation, temporal forecasting, multimodal EEG–EOG integration, and label-efficient learning. The findings show that physiologically meaningful EEG features provide informative and discriminative representations for microsleep prediction, while temporal modelling substantially improves forecasting performance by capturing the progressive transition into mi-

crosleep states. The integration of EEG and EOG further improves predictive performance and provides insight into cross-modal physiological interactions preceding microsleep onset through information-theoretic measures including mutual information and transfer entropy. Finally, a multimodal disagreement-based active learning framework demonstrates that annotation requirements can be reduced considerably while maintaining competitive forecasting performance. Collectively, these results indicate that combining temporal modelling, multimodal physiological information, and label-efficient learning yields more robust and scalable microsleep prediction systems.

Overall, this thesis contributes a comprehensive framework for multimodal microsleep prediction that combines physiological feature analysis, temporal forecasting, multimodal interaction modelling, and label-efficient learning. The findings advance understanding of microsleep-related neurophysiology while providing practical insights for the development of future vigilance-monitoring systems. More broadly, the methodologies developed in this work may be applicable to fatigue monitoring, neuroergonomics, biomedical signal analysis, and related brain–computer interface applications.

Funding: This work was supported by Science Foundation Ireland under Grant No. SFI/12/RC/2289 P2.

Chapter 1

Introduction

1.1 Background and Motivation

Microsleeps are brief episodes of involuntary sleep-like activity that occur during wakefulness and are commonly associated with fatigue, reduced vigilance, and cognitive overload [1]. Although these events often last only a few seconds, they can severely impair attention, perception, reaction time, and motor control. Microsleeps are particularly dangerous in operational environments that require sustained attention, including driving, aviation, healthcare monitoring, industrial operations, and military systems [2], [3], [4], [5], [6]. Because microsleeps frequently occur without full awareness by the affected individual, their detection and prediction remain critical challenges in biomedical signal analysis and intelligent monitoring systems.

From an EEG perspective, microsleep onset is commonly associated with a shift in spectral activity from wake-related alpha and beta rhythms towards lower-frequency theta activity. This transition is illustrated in Figure 1.1. Microsleep should not be interpreted as a distinct frequency band; rather, it involves spectral reorganisation, including increased theta power, poorly defined or reduced alpha activity, and reduced beta activity relative to alert wakefulness.

Electroencephalography (EEG) has become one of the most widely used physiological modalities for studying vigilance decline and microsleep behaviour because it provides direct measurements of neural activity with high temporal resolution [7], [8].

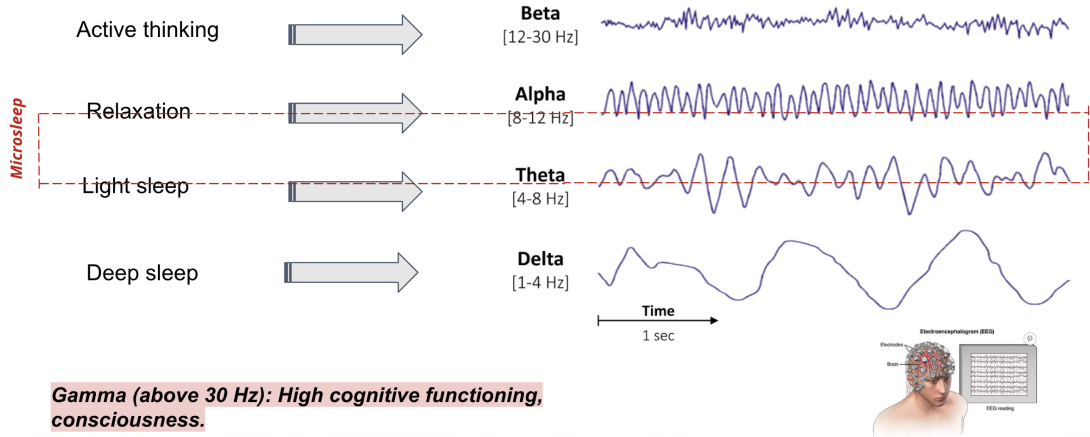


Figure 1.1: Illustrative EEG frequency bands and their broad associations with vigilance and sleep. Beta activity is commonly associated with active wakefulness, alpha activity with relaxed wakefulness, theta activity with drowsiness and light sleep, and delta activity with deeper sleep. Microsleep does not constitute a separate EEG frequency band; instead, it is commonly associated with a transition involving increased theta activity, reduced or poorly defined alpha activity, and reduced beta activity relative to alert wakefulness. The displayed waveforms are schematic and are intended to illustrate differences in frequency rather than represent recorded EEG segments.

EEG-based microsleep analysis has therefore attracted substantial research interest over the last decade. Early studies primarily focused on detecting microsleep states after onset using handcrafted spectral features and conventional machine learning classifiers. These approaches demonstrated that physiological changes associated with drowsiness and microsleep can be identified from neural signals, particularly through slowing of EEG activity and reductions in signal complexity [9].

However, despite progress in microsleep detection, several important limitations remain unresolved in the literature. First, many existing systems rely heavily on static classification frameworks that do not adequately model the progressive temporal transition into microsleep. Second, a large proportion of prior work focuses only on EEG, even though microsleep transitions involve coordinated neural and ocular processes that may be better captured through multimodal physiological analysis. Third, many systems require extensive manually annotated datasets, making model development expensive and difficult to scale. Finally, improvements in predictive

performance are often pursued primarily through increasingly complex classifiers rather than through systematic investigation of the underlying representational, temporal, multimodal, and annotation-related challenges.

These observations motivate the perspective adopted in this thesis, namely that microsleep prediction should not be treated as a single-stage classification problem. Instead, it should be viewed as a progressive modelling problem in which several interdependent components collectively determine predictive performance. Informative physiological feature representations must first be identified. Temporal structure must then be preserved to capture transitional dynamics. Complementary sensing modalities must be integrated in a physiologically meaningful manner. Finally, the resulting models must remain trainable under realistic annotation constraints.

This thesis addresses these challenges through a multi-stage framework for microsleep forecasting based on EEG and electrooculography (EOG) signals, which capture blink behaviour, eye movements, and eye-closure dynamics associated with vigilance decline [10]. The work is organised around four research questions that progressively examine feature representation, temporal modelling, multimodal integration, and label-efficient learning. Collectively, these studies aim to improve short-horizon microsleep forecasting, where prediction is performed several seconds before microsleep onset rather than at the moment a microsleep occurs, while maintaining interpretability, physiological relevance, and practical feasibility.

1.2 Problem Statement

Although substantial progress has been made in EEG-based vigilance monitoring, reliable short-horizon microsleep forecasting remains difficult. Existing approaches often suffer from several methodological limitations.

A major limitation concerns feature representation. Many previous systems rely on generic EEG descriptors without systematically evaluating whether those features capture physiologically meaningful indicators of vigilance decline [9], [11], [12].

Since microsleep transitions involve changes in neural oscillatory activity, reductions in complexity, and evolving cross-regional interactions, inadequate feature representation can limit downstream predictive performance.

A second limitation concerns temporal modelling. Traditional machine learning pipelines frequently treat EEG segments as independent observations, even though microsleep transitions occur progressively over time. Static classifiers therefore struggle to preserve temporally local information that may be critical for identifying imminent microsleep events [13].

A third limitation concerns modality usage and physiological interaction modelling. While EEG provides valuable information about neural activity, ocular behaviour measured through EOG also reflects fatigue-related processes associated with vigilance decline. However, many previous systems rely on either EEG or EOG independently, with limited research explicitly modelling the interaction between neural and ocular dynamics through combined EEG–EOG analysis for microsleep prediction. Furthermore, little attention has been given to understanding the temporal directionality of microsleep-related physiological transitions, specifically whether microsleep-related changes emerge first in cortical EEG activity, ocular EOG activity, or synchronously across both modalities. Investigating these relationships may provide important insight into the underlying physiology of microsleep onset and improve multimodal forecasting performance [10], [14], [15], [16], [17].

A fourth limitation concerns annotation efficiency. Microsleep datasets require labour-intensive expert labelling procedures, particularly for event forecasting tasks. Fully supervised training approaches therefore become increasingly impractical as datasets scale. Despite advances in predictive modelling, comparatively little attention has been given to reducing annotation requirements while preserving forecasting performance [18].

Together, these limitations indicate that microsleep prediction is constrained not only by classifier capacity, but also by feature quality, temporal structure preservation, multimodal dependency modelling, and efficient use of labelled data. Address-

ing these challenges requires an integrated modelling framework rather than isolated improvements to individual classifiers.

1.3 Research Aim

The overall aim of this thesis is to improve the prediction of microsleep events from physiological data, with particular emphasis on short-horizon forecasting under realistic operational and annotation constraints.

To achieve this aim, the thesis investigates microsleep prediction as a multi-stage modelling problem involving:

1. Identification of informative physiological feature representations.
2. Explicit modelling of temporal structure for forecasting.
3. Integration of complementary multimodal physiological information.
4. Reduction of annotation requirements through label-efficient learning.

The thesis therefore addresses microsleep prediction not only as a machine learning problem, but also as a physiologically grounded forecasting problem.

1.4 Research Questions

The thesis is structured around four interconnected research questions.

RQ1: Feature Representation *Which physiological EEG features provide the most informative representation for microsleep detection and form a robust foundation for subsequent forecasting models?*

This question investigates whether spectral, nonlinear, and complexity-based EEG features can reliably capture physiologically meaningful indicators of vigilance decline and microsleep transition.

RQ2: Temporal Modelling *How does explicit modelling of temporal structure influence short-horizon microsleep forecasting performance?*

This question evaluates whether sequence-sensitive architectures improve forecasting beyond static classification pipelines by preserving temporally local information relevant to microsleep onset.

RQ3: Multimodal Fusion *Can multimodal integration of EEG and EOG improve microsleep forecasting performance compared with unimodal approaches, and what physiological interaction patterns exist between neural and ocular activity during microsleep transitions?*

This question investigates whether complementary neural and ocular information improves forecasting performance and whether modelling cross-modal interaction provides additional benefits beyond simple feature fusion. In addition, the research examines the temporal directionality of microsleep-related physiological dynamics using information-theoretic measures such as Transfer Entropy (TE) to determine whether microsleep transitions originate predominantly from EEG activity, EOG activity, or emerge synchronously across both modalities.

RQ4: Label-Efficient Learning *Can active learning reduce annotation requirements while maintaining strong multimodal forecasting performance?*

This question examines whether intelligent sample-selection strategies can substantially reduce labelling effort without significant degradation in predictive performance.

1.5 Research Contributions

The contributions of this thesis are cumulative and progressively connected across the four research questions.

1.5.1 Contribution 1: Physiologically Meaningful Feature Representation

The thesis demonstrates that microsleep prediction depends strongly on physiologically relevant feature extraction. Spectral and nonlinear EEG features are shown to provide stable discriminative information associated with vigilance decline and reduced neural complexity.

1.5.2 Contribution 2: Sequence-Sensitive Temporal Forecasting

The thesis demonstrates that preserving temporally local information substantially improves forecasting performance. Temporal modelling approaches incorporating causal attention and recurrent processing improve sensitivity and discrimination across prediction horizons.

1.5.3 Contribution 3: Multimodal Physiological Fusion and Directionality Analysis

The thesis shows that EEG and EOG provide complementary physiological information for microsleep forecasting. The work further demonstrates that multimodal gain is strongest when cross-modal interaction is modelled explicitly through information-theoretic coupling measures rather than through simple concatenation alone. In addition, Transfer Entropy analysis is used to investigate the directional flow of information between EEG and EOG activity during microsleep transitions, providing insight into whether microsleep-related dynamics emerge predominantly from cortical activity, ocular behaviour, or synchronised multimodal processes.

1.5.4 Contribution 4: Label-Efficient Multimodal Learning

The thesis demonstrates that active learning can substantially reduce annotation requirements while retaining near-baseline forecasting performance. The proposed

multimodal disagreement-based acquisition strategy provides an interpretable and practically effective framework for label-efficient microsleep prediction.

1.5.5 Contribution 5: Integrated Perspective on Microsleep Forecasting

Beyond individual modelling improvements, the thesis contributes a unified perspective in which microsleep prediction is understood as a progressive modelling problem involving representation quality, temporal preservation, multimodal interaction, and annotation efficiency.

1.6 Significance of the Research

The significance of this research lies in both its scientific and practical contributions.

From a scientific perspective, the thesis advances understanding of how physiological representations, temporal dynamics, and multimodal dependencies contribute to microsleep forecasting. The work also contributes to explainable biomedical machine learning by emphasising interpretable physiological mechanisms alongside predictive accuracy.

From a practical perspective, reliable microsleep forecasting has direct relevance for fatigue monitoring systems in transportation, healthcare, defence, and industrial environments. Systems capable of forecasting microsleep transitions before full onset may support earlier intervention and improve operational safety.

Importantly, the thesis also addresses one of the major deployment barriers in physiological machine learning systems: annotation cost. By demonstrating that active learning can recover strong performance with substantially reduced labelled data, the work improves the feasibility of developing scalable real-world microsleep monitoring systems.

1.7 Relationship Between the Research Questions

Although each research question addresses a distinct modelling challenge, the thesis is intentionally cumulative in structure.

RQ1 establishes the representational foundation of the thesis by identifying informative physiological EEG features. These representations form the basis for all later modelling stages.

RQ2 builds upon this foundation by introducing temporal modelling approaches that preserve sequence-dependent information necessary for forecasting imminent microsleep events.

RQ3 extends the framework into multimodal learning by integrating complementary EEG and EOG information, modelling cross-modal synchrony during vigilance transitions, and investigating the directional flow of information between neural and ocular activity using information-theoretic interaction analysis.

RQ4 then addresses the practical challenge of annotation scalability by investigating whether strong multimodal forecasting systems can be trained efficiently under limited supervision.

Taken together, the research questions form a progression from representation, to temporal structure, to multimodal integration, and finally to efficient learning.

1.8 Thesis Structure

This thesis is organised into 9 chapters.

Chapter 1 – Introduction This chapter introduces the microsleep prediction problem, outlines the research motivation, presents the research questions and objectives, and summarises the overall contributions and structure of the thesis.

Chapter 2 – Foundations and Related Work This chapter reviews the literature related to physiological vigilance monitoring using EEG and EOG, microsleep detection, temporal modelling, multimodal physiological analysis, and

active learning in biomedical signal processing. The chapter also identifies the major research gaps addressed by the thesis.

Chapter 3 – Dataset and Evaluation Framework This chapter describes the datasets, preprocessing procedures, feature extraction methods, forecasting protocols, evaluation metrics, and statistical analysis procedures used throughout the thesis.

Chapter 4 – Feature Representation for Microsleep Prediction (RQ1) This chapter investigates physiologically meaningful EEG feature representations for microsleep prediction, with particular emphasis on spectral and nonlinear descriptors associated with vigilance decline.

Chapter 5 – Temporal Modelling for Microsleep Forecasting (RQ2) This chapter evaluates sequence-sensitive forecasting architectures and investigates how temporal modelling influences predictive performance across different forecast horizons.

Chapter 6 – Multimodal EEG–EOG Fusion for Forecasting (RQ3) This chapter investigates multimodal integration strategies for combining EEG and EOG information, including information-theoretic interaction measures, Transfer Entropy-based directionality analysis, and cross-modal fusion mechanisms for microsleep forecasting.

Chapter 7 – Active Learning for Label-Efficient Forecasting (RQ4) This chapter examines whether active learning can reduce annotation requirements while preserving multimodal forecasting performance. Benchmark active learning methods are compared with the proposed multimodal disagreement-based acquisition strategy.

Chapter 8 – Cross-Study Synthesis and Design Insights This chapter synthesises the findings across all research questions, discusses the major cross-cutting themes, outlines the methodological limitations of the thesis, and

presents implications for future research.

Chapter 9 – Conclusions and Future Work This chapter summarises the principal findings and contributions of the thesis, revisits the research questions, discusses key limitations, and outlines several directions for future research in microsleep forecasting, multimodal physiological modelling, and label-efficient learning.

1.9 Chapter Summary

This chapter introduced the problem of microsleep prediction and motivated the need for reliable short-horizon forecasting systems based on physiological monitoring. The chapter identified key limitations in existing approaches related to feature representation, temporal modelling, multimodal integration, and annotation efficiency.

To address these challenges, the thesis proposed a progressive modelling framework structured around four research questions investigating physiologically meaningful feature extraction, sequence-sensitive forecasting, multimodal EEG–EOG fusion, and label-efficient active learning.

The following chapter reviews the literature related to these research areas and identifies the theoretical and methodological gaps addressed throughout the thesis.

Chapter 2

Foundations and Related Work

2.1 Introduction

Microsleeps are transient episodes of involuntary sleep intrusion into wakefulness, typically lasting between 0.5 and 15 seconds, and are characterised by complete or partial loss of responsiveness, behavioural disengagement, and brief eye closure. These episodes are especially dangerous in environments requiring sustained attention such as driving, aviation, industrial monitoring, and healthcare operations because even a short lapse may lead to catastrophic consequences [2], [3], [4], [5], [6]. Microsleeps have therefore become a major research focus within neuroengineering, sleep science, and intelligent transportation systems [19], [20], [21].

Although drowsiness and fatigue have long been studied using behavioural and physiological measurements, microsleeps represent a more challenging phenomenon because they occur abruptly, are short-lived, and often emerge even in non-sleep-deprived individuals [9], [21], [22], [23]. Traditional behavioural indicators such as eyelid closure, steering deviation, or reaction time are often insufficient for reliable early detection. Consequently, electroencephalography (EEG) has emerged as one of the most informative modalities for microsleep analysis due to its direct measurement of cortical activity and high temporal resolution [7], [8]

Recent advances in machine learning and deep learning have accelerated research in automated microsleep detection and prediction. Early approaches relied heavily

on handcrafted spectral features and classical classifiers, whereas modern systems increasingly employ convolutional neural networks (CNNs), recurrent neural networks (RNNs), temporal convolutional networks (TCNs), attention mechanisms, and multimodal fusion frameworks [24], [25], [26], [27], [28], [29], [30]. Despite this progress, major challenges remain unresolved, including severe class imbalance, poor inter-subject generalisation, variability in microsleep definitions, lack of standardised evaluation protocols, and insufficient integration of neurophysiological knowledge into learning systems [31], [32], [33].

This chapter critically reviews the literature relevant to EEG-based microsleep detection and prediction. In addition to electroencephalography (EEG), the review also considers electrooculography (EOG), which captures eye-movement and ocular activity that are closely associated with vigilance decline and microsleep onset. The review covers physiological foundations of microsleeps, EEG and EOG characteristics, datasets and experimental paradigms, signal preprocessing, feature engineering, machine learning and deep learning methods, multimodal systems, active learning frameworks, and current methodological limitations. The chapter concludes by identifying research gaps that motivate the contributions of this thesis.

The remainder of this chapter is organised as follows. First, the physiological and neurocognitive foundations of microsleeps and vigilance degradation are reviewed. Next, EEG and EOG characteristics associated with microsleep episodes are discussed, followed by an overview of datasets and experimental paradigms used in microsleep research. The chapter then reviews traditional machine learning and modern deep learning approaches for microsleep detection and forecasting. Subsequently, multimodal systems integrating EEG with complementary modalities such as EOG are examined. The chapter further discusses active learning and label-efficient learning strategies for EEG analysis and driver monitoring applications. Finally, current challenges including subject variability, data imbalance, noisy real-world environments, and limited forecasting robustness are discussed, highlighting the research gaps addressed in this thesis.

2.2 Microsleeps and Vigilance Degradation

Microsleeps are considered brief intrusions of sleep into wakefulness and are commonly associated with degraded vigilance and impaired task responsiveness [22], [34]. Unlike ordinary drowsiness, microsleeps represent discrete behavioural failures during which individuals temporarily lose conscious control over task execution [9]. They are operationally characterised using combinations of eye closure, behavioural slowing, head nodding, and tracking failure [22].

The relationship between fatigue, vigilance decline, and accidents has been extensively documented in transportation research. Williamson et al. demonstrated that fatigue-related impairment may be comparable to alcohol intoxication in terms of reaction-time degradation and cognitive decline [3]. Similarly, Howard et al. showed that extended wakefulness significantly degrades simulated driving performance [35]. These findings support the importance of objective physiological monitoring systems capable of identifying microsleeps before behavioural failure becomes catastrophic.

Microsleeps are not restricted to severely sleep-deprived individuals. Peiris et al. demonstrated that healthy non-sleep-deprived subjects can still experience frequent microsleeps during monotonous continuous tracking tasks [9]. Similarly, Poudel et al. showed that microsleeps emerge even under rested conditions but become substantially more frequent following sleep restriction [36]. This observation suggests that microsleeps arise from dynamic instability between wake-promoting and sleep-promoting neural systems rather than from sleep deprivation alone. Recent machine learning studies have further demonstrated that microsleep episodes exhibit identifiable physiological signatures that can be automatically detected from EEG recordings using both feature-based and deep learning approaches [24], [37].

The concept of wake-state instability has therefore become central to microsleep research. According to this theory, sustained attention tasks produce fluctuating competition between wake-promoting neural systems (often referred to as arousal networks) and the homeostatic drive for sleep, resulting in transient lapses of consciousness [8]. Microsleeps may thus represent the behavioural manifestation of

unstable transitions between wakefulness and sleep [22], [34].

These findings collectively suggest that microsleeps are not isolated behavioural failures but rather manifestations of dynamic neurophysiological instability. Consequently, understanding the neural mechanisms underlying vigilance degradation is essential for developing reliable early-warning systems capable of predicting microsleeps before behavioural failure occurs.

2.3 EEG and EOG Signals in Microsleep Analysis

Electroencephalography (EEG) is one of the most widely used neurophysiological modalities for studying vigilance, fatigue, and microsleep phenomena due to its ability to directly measure cortical electrical activity with millisecond-level temporal resolution [7]. EEG signals are acquired non-invasively through electrodes placed on the scalp, where voltage fluctuations generated by synchronized postsynaptic activity of cortical pyramidal neurons are recorded. Because microsleeps emerge rapidly and involve transient alterations in neural activity, EEG is particularly well suited for detecting and forecasting these events.

A major advantage of EEG is its ability to capture dynamic temporal changes associated with vigilance degradation. Prior research has consistently demonstrated that declining alertness is accompanied by characteristic spectral changes, particularly increased low-frequency activity in the theta (4–8 Hz) and alpha (8–13 Hz) bands [9], [11], [38]. Theta-band increases are commonly associated with drowsiness and reduced attentional engagement, whereas alpha-band modulation has been linked to fluctuations in cortical arousal and visual attention [8]. More recent work has shown that microsleeps involve complex spatio-temporal neural dynamics extending beyond simple spectral slowing, including transient spindle-like activity, defined here as brief, waxing-and-waning bursts of rhythmic EEG activity resembling the sleep spindles typically observed during Stage-2 non-rapid eye movement sleep, as well as altered connectivity patterns and localised cortical activation changes [39], [40], [41], [42]. Furthermore, studies investigating vigilance instability and sleep de-

privation have reported altered neural oscillations, phase-amplitude coupling, and dynamic cortical state transitions associated with declining alertness [43].

EEG-based microsleep analysis has employed both single-channel and multichannel acquisition systems. Single-channel EEG systems offer simplicity, portability, and lower computational requirements, making them attractive for real-time monitoring applications such as driver vigilance systems [25]. However, their limited spatial coverage may fail to capture distributed neural processes associated with microsleep onset. Multichannel EEG systems, in contrast, provide richer spatial information and improved robustness by capturing interactions across multiple cortical regions [44]. Nevertheless, increased channel count introduces greater system complexity, setup time, and computational burden, which may limit practical deployment outside laboratory environments.

EEG signals are highly susceptible to contamination from physiological and environmental artifacts. Eye blinks, eye movements, muscle activity, electrode impedance fluctuations, and motion-related disturbances can substantially degrade signal quality [45]. This challenge becomes particularly important in real-world environments such as transportation, industrial monitoring, and healthcare settings, where subjects naturally move, speak, and interact with surrounding systems. Recent studies have highlighted the growing importance of robust artifact handling for wearable and mobile EEG systems operating under naturalistic conditions [14], [15], [16], [17]. Consequently, robust EEG-based microsleep forecasting requires methods capable of handling substantial signal variability and noise contamination.

Electrooculography (EOG) provides a complementary physiological modality for analysing vigilance and microsleep behaviour. EOG measures voltage differences generated by ocular movement and retinal-corneal potential changes using electrodes placed around the eyes [10]. Because eye behaviour changes significantly during fatigue and microsleep onset, EOG has become widely used in sleep research and fatigue monitoring applications.

Several ocular behaviours have been associated with declining vigilance, includ-

ing slow eye movements, blink duration changes, prolonged eye closures, and altered saccadic dynamics [22]. Sustained eyelid closure and reduced blink responsiveness are particularly strong indicators of microsleep onset. Video-based eye-aspect-ratio analysis has therefore gained popularity as a non-invasive behavioural monitoring approach [46]. However, camera-based systems are sensitive to illumination conditions, head pose variation, occlusion, and image quality degradation, limiting their reliability in uncontrolled environments [28].

Unlike video-based methods, EOG directly captures ocular electrophysiological activity and can therefore provide more robust measurements under challenging environmental conditions. EOG signals have been used extensively in sleep staging, fatigue monitoring, and ocular artifact analysis [10]. Deep learning approaches based on EOG have recently demonstrated strong performance in sleep-stage classification and vigilance analysis, highlighting the growing utility of EOG-driven representation learning frameworks [47], [48], [49]. Moreover, EOG recordings often contain informative pre-microsleep patterns associated with slow rolling eye movements and blink dynamics that may emerge before overt behavioural lapses occur.

Importantly, EEG and EOG capture distinct yet complementary aspects of the microsleep process. EEG reflects cortical vigilance dynamics and neural state transitions, whereas EOG captures behavioural and ocular manifestations of fatigue. Their joint analysis may therefore provide a more comprehensive representation of pre-microsleep physiology than either modality alone. Recent multimodal fatigue-monitoring systems integrating EEG with EOG, eye tracking, motion signals, or video-based behavioural analysis have demonstrated improved robustness and predictive performance compared with unimodal systems [29], [50], [51], [52], [53]. These findings suggest that multimodal integration can improve resilience under noisy and heterogeneous real-world conditions.

Despite this progress, most existing microsleep forecasting systems remain predominantly unimodal and rely mainly on EEG or behavioural measurements independently. Furthermore, relatively little attention has been given to explicitly

modelling the physiological interaction between neural and ocular dynamics during vigilance degradation. Consequently, the integration of EEG and EOG for physiologically grounded microsleep forecasting remains an important and underexplored research direction.

2.4 Neurophysiological Basis of Microsleeps

Understanding the neural basis of microsleeps is critical for designing physiologically grounded detection systems. Early EEG studies reported increases in low-frequency activity, particularly theta activity (4–8 Hz), commonly associated with drowsiness, and alpha activity (8–13 Hz), often linked to cortical arousal and attention, during microsleep episodes [9]. However, more recent work suggests that microsleeps involve considerably more complex neural dynamics [54], [55].

Jonmohamadi et al. demonstrated that microsleeps are frequently associated with alpha-band spindle activity originating from bilateral anterior temporal gyri and hippocampal regions, brain areas involved in memory processing and higher cognitive functions, resembling Stage-2 sleep spindles [39]. Their findings suggest that microsleeps may represent brief intrusions of genuine sleep mechanisms into wakefulness rather than simple reductions in alertness. Importantly, they also identified frontal-orbital theta activity associated with microsleep events, supporting the role of frontal attentional systems in vigilance failure. These observations are consistent with broader sleep physiology literature showing that sleep spindles and alpha rhythms arise from complex thalamo-cortical interactions [56], [57], [58].

Subsequent work by Zaky et al. expanded these findings by investigating source-localised EEG activity during microsleeps [40]. Their analysis revealed increases not only in delta, theta, and alpha bands but also in beta and gamma activity. The authors proposed that increased high-frequency activity may reflect unconscious cognitive processes attempting to restore wakefulness and consciousness during active task performance. This interpretation aligns with neuroimaging findings showing transient cortical hyperactivation during microsleeps and vigilance instability [59],

[60].

Functional neuroimaging studies further support the existence of compensatory cortical mechanisms. Poudel et al. observed transient activation in fronto-parietal, parahippocampal, and thalamo-cortical networks during microsleeps [36]. Importantly, they demonstrated that behavioural performance often improved immediately after microsleep recovery, suggesting that microsleeps may temporarily alleviate sleep pressure and restore vigilance. Similar observations were reported by Strauss et al., who showed that loss of responsiveness during sleep onset is associated with dynamic changes in large-scale neural connectivity [61]. These findings challenge simplistic interpretations of microsleeps as merely failures of wakefulness and instead position them as dynamic neurophysiological recovery phenomena.

Recent EEG studies have further demonstrated that microsleep-related vigilance decline is associated with distinct spectral reorganisation patterns involving theta and alpha activity [12], [41]. Dynamic analyses of neural connectivity and EEG microstates additionally suggest that vigilance transitions involve rapidly evolving spatio-temporal brain states rather than static spectral slowing alone [62], [63], [64]. Such findings reinforce the importance of modelling temporal dynamics and cross-regional neural interactions in computational microsleep prediction systems.

The distinction between behavioural microsleeps and ordinary drowsiness is therefore essential. Whereas drowsiness represents a tonic state of reduced alertness, microsleeps are phasic transitions involving transient disengagement from the environment and activation of sleep-related neural mechanisms [22]. This distinction has major implications for algorithm development because systems trained only on drowsiness-related features may fail to capture the discrete neural signatures associated with true microsleeps.

These findings indicate that microsleeps involve distributed and dynamic neural processes spanning cortical, thalamic, and hippocampal systems. Consequently, computational models designed for microsleep prediction must capture not only spectral slowing but also complex spatio-temporal interactions across multiple brain

regions and frequency bands.

2.5 EEG-Based Microsleep Detection and Forecasting

2.5.1 Traditional EEG-Based Detection Methods

Early microsleep detection systems primarily relied on handcrafted EEG features combined with traditional machine learning classifiers. Spectral slowing, particularly increased theta activity and reduced alpha power, has consistently been associated with vigilance decline and microsleep onset [9], [11]. Consequently, many early systems employed spectral power features extracted from canonical EEG frequency bands. Spectral EEG analysis during microsleep events further demonstrated characteristic low-frequency alterations associated with fatigue progression and reduced alertness [12].

Beyond conventional spectral analysis, researchers explored nonlinear and information-theoretic features to better capture the complexity of microsleep-related neural dynamics. Joint entropy and mutual information features demonstrated strong discriminative capability for distinguishing microsleep from responsive states [44]. Spatio-temporal approaches based on Common Spatial Patterns further improved classification performance by exploiting spatial covariance structures across EEG channels [13]. However, forecasting performance was shown to degrade substantially as prediction horizons increased, largely due to loss of fine-grained temporal information.

Ayyagari et al. investigated dimensionality reduction techniques for microsleep classification, comparing methods including principal component analysis, probabilistic PCA, and nonlinear embeddings [65]. Their findings demonstrated that reducing feature dimensionality could improve computational efficiency while preserving discriminative information. Similar investigations were conducted in subsequent

studies examining feature selection, channel optimisation, and statistical validation strategies for EEG-based microsleep analysis [66], [67], [68], [69]. Recent work has also explored handling class imbalance in microsleep datasets through oversampling and discriminative feature engineering strategies [32].

Several studies also explored blind source separation and independent component approaches for isolating microsleep-related EEG activity from artifacts and noise. Leong et al. proposed blind extraction methods capable of identifying microsleep-related signal components from mixed EEG recordings [70]. These approaches highlighted the importance of preprocessing and signal decomposition for improving microsleep detection reliability.

Traditional machine learning classifiers including support vector machines (SVM), k-nearest neighbours (KNN), linear discriminant analysis (LDA), and random forests have been widely employed for EEG-based fatigue and microsleep detection [66], [68], [69]. While these approaches often achieved promising within-subject performance, their ability to generalise across unseen individuals remained limited due to substantial inter-subject EEG variability.

Despite promising results, traditional handcrafted feature pipelines suffer from several limitations. Feature engineering often requires extensive domain expertise, handcrafted features may not generalise across subjects, and classical machine learning classifiers frequently struggle to capture complex nonlinear temporal dependencies present in EEG recordings. These limitations motivated the transition toward deep learning and end-to-end neural architectures.

2.5.2 Deep Learning for Microsleep Detection and Forecasting

Recent advances in deep learning have transformed EEG analysis by enabling end-to-end learning directly from raw or minimally processed signals. CNN-based architectures have demonstrated strong performance in extracting spatial and spectral representations from EEG recordings [24], [71]. Deep learning approaches reduce

reliance on handcrafted feature engineering and allow models to learn hierarchical signal representations automatically.

Single-channel CNN systems have shown that compact EEG acquisition setups can achieve practical microsleep and drowsiness detection performance [25], [72]. However, multichannel deep learning approaches generally achieve superior robustness by exploiting distributed neural information across multiple cortical regions. Temporal convolutional networks and recurrent neural networks have therefore become increasingly popular for modelling sequential EEG dynamics associated with vigilance decline.

Recurrent neural networks, particularly long short-term memory (LSTM) architectures and gated recurrent units (GRUs), have enabled modelling of temporal dependencies associated with behavioural lapses and vigilance degradation [73], [74], [75], [76]. These models are especially relevant for forecasting tasks because they preserve temporal context across sequential EEG windows.

Recent studies have also demonstrated the effectiveness of hybrid deep learning frameworks integrating convolutional feature extraction with temporal attention mechanisms [77], [78]. Such architectures improve representation learning by jointly modelling spatial, spectral, and temporal EEG characteristics associated with fatigue progression. Attention-based CNN frameworks have similarly shown improved performance in motor imagery EEG decoding and drowsiness recognition tasks [79], [80].

One of the few studies specifically addressing multi-horizon microsleep forecasting was conducted by Shoorangiz et al. [13], who examined prediction horizons ranging from 0.25 to 1 second prior to microsleep onset. Their findings demonstrated substantial degradation in sensitivity as the forecasting horizon increased, primarily due to reduced preservation of short-lived temporal cues preceding microsleep events. This highlighted the need for architectures capable of preserving temporal resolution while learning long-range sequential dependencies.

More recently, attention mechanisms and transformer-inspired architectures have

emerged as promising approaches for EEG decoding due to their ability to capture long-range temporal relationships and dynamically emphasise informative signal regions [81], [82]. Attention mechanisms enable models to selectively focus on temporally relevant EEG patterns while suppressing irrelevant or noisy information. Connectivity-aware graph neural networks have additionally been proposed for modelling dynamic interactions between EEG channels during drowsiness and vigilance decline [83].

Cross-subject and subject-independent learning has also become an important research direction due to the impracticality of subject-specific calibration in real-world deployments. Paulo et al. proposed a zero-calibration framework using spatio-temporal EEG image encoding for cross-subject drowsiness detection [84]. Similarly, several recent studies have explored subject-independent deep learning architectures capable of improving generalisation across unseen individuals [78], [85].

Multimodal deep learning approaches integrating EEG with additional physiological or behavioural modalities have further demonstrated improved robustness under real-world conditions. These include combinations of EEG with physiological measurements, eye tracking, and vehicular signals for fatigue and microsleep detection [86], [87].

Nevertheless, the application of advanced attention-based and multimodal architectures to short-horizon microsleep forecasting remains relatively limited compared with onset detection. Existing studies have primarily focused on general EEG classification tasks, drowsiness detection, or sleep-stage analysis rather than continuous microsleep forecasting across multiple horizons. Consequently, understanding how temporal attention mechanisms, multimodal physiological interaction modelling, and subject-independent learning influence microsleep forecasting robustness remains an important research direction.

2.6 Multimodal Microsleep Detection and Prediction

While EEG remains the dominant modality for microsleep analysis, complementary behavioural and physiological modalities have increasingly been explored to improve robustness under real-world conditions. Existing approaches include facial image analysis, steering-wheel pressure sensing, speech-based analysis, eye-blink monitoring, and video-based eye tracking systems [46], [88], [89]. These approaches aim to capture behavioural manifestations of vigilance decline that may complement neural activity measurements.

Eye-related measures are particularly relevant because ocular behaviour changes substantially during vigilance decline. Video-based eye-aspect-ratio analysis has demonstrated effectiveness for detecting prolonged eye closures associated with microsleep episodes [46]. More recent work has explored real-time blink tracking and event-based eye monitoring systems for fatigue detection [90]. Eye-state and yawning-based monitoring approaches have additionally shown promise for practical fatigue detection in real-world settings [89], [91], [92]. However, purely behavioural systems often fail under poor lighting conditions, occlusions, or subtle pre-onset vigilance changes.

Speech-based approaches have also been explored for fatigue and drowsiness monitoring. These systems typically analyse spectral, prosodic, acoustic, and emotional characteristics of speech to estimate fatigue-related cognitive decline. Related research in cognitive impairment and neurological monitoring has demonstrated that acoustic speech features may capture subtle neurophysiological changes associated with declining cognitive states [93], [94], [95]. Although promising, speech systems may be sensitive to environmental acoustics, recording quality, and inter-speaker variability.

Despite the growing interest in multimodal systems, most existing approaches remain effectively non-neural or unimodal from a neurophysiological perspective.

Hybrid physiological systems have recently begun integrating EEG with eye tracking, blink analysis, and additional biometric signals for fatigue monitoring [51], [96], [97]. Large-scale multimodal fatigue datasets incorporating behavioural, physiological, and video modalities have also recently emerged. However, to the best of current knowledge, very limited research has specifically investigated the fusion of EEG and EOG for short-horizon microsleep forecasting prior to onset.

EEG and EOG capture complementary aspects of the microsleep process. EEG reflects cortical vigilance dynamics and neural state transitions, whereas EOG captures ocular manifestations including slow eye movements, blink behaviour, and sustained eye closure [10]. Their joint analysis therefore offers the potential for more comprehensive modelling of pre-microsleep physiology. Recent multimodal fatigue-monitoring systems combining EEG with ocular and physiological measurements have demonstrated improved robustness compared to unimodal approaches [51], [86], [97].

Traditional multimodal fusion approaches often rely on direct feature concatenation, which treats modalities as largely independent sources of information. However, neural and ocular activity preceding microsleeps may exhibit complex nonlinear dependencies and directional interactions. Information-theoretic measures such as mutual information and transfer entropy therefore provide attractive frameworks for modelling EEG–EOG relationships [98].

Mutual information quantifies shared information between signals while capturing both linear and nonlinear dependencies without strong parametric assumptions [98]. Transfer entropy extends this framework by estimating directional information transfer and potential causal interactions between modalities [99]. Together, these measures may provide physiologically meaningful representations of neural–ocular coupling preceding microsleep onset.

Despite the promise of multimodal systems, robust fusion strategies remain underexplored in microsleep forecasting. Existing systems frequently focus on post-event detection rather than early prediction, and relatively little attention has been

given to physiologically interpretable multimodal interactions.

2.7 Active Learning and Label-Efficient EEG Analysis

2.7.1 Motivation for Active Learning in EEG

A major challenge in EEG-based machine learning is the high cost and difficulty of obtaining labelled data. EEG annotation often requires expert review, behavioural synchronisation, and substantial manual effort. In microsleep analysis, this challenge is further amplified by severe class imbalance and the rarity of microsleep events relative to normal wakefulness.

Active learning has therefore emerged as a promising strategy for reducing annotation burden while maintaining classification performance. Rather than randomly selecting samples for labelling, active learning methods iteratively identify the most informative unlabeled samples for annotation. This paradigm is particularly attractive in physiological signal analysis, where large quantities of unlabeled biosignals are often available but expert annotations remain expensive and time-consuming [18].

In EEG-based BCI systems, effective active learning strategies must balance several factors including informativeness, representativeness, and diversity. Samples near the classifier decision boundary may provide high informational value, while representative and diverse samples help avoid redundancy and outlier selection [100], [101], [102]. Batch-mode active learning approaches have therefore become increasingly important for practical annotation workflows involving groups of samples rather than isolated instances [103].

2.7.2 Active Learning in EEG and Brain–Computer Interfaces

Several studies have explored active learning within EEG-based classification and brain–computer interface systems. She et al. proposed a double-criteria active learning framework integrating uncertainty, representativeness, and diversity constraints to improve label efficiency in EEG classification [100]. Similarly, Wang et al. investigated neural decoding using active learning for intracortical brain–machine interfaces and demonstrated improvements in decoding performance under limited labeled data conditions [104].

Semi-supervised and progressive learning approaches have also attracted attention for EEG applications. Wei et al. proposed a semi-supervised progressive learning framework for BCI systems capable of leveraging both labeled and unlabeled EEG data to improve model robustness and calibration efficiency [105]. More recently, Long et al. explored active EEG annotation strategies jointly driven by pseudo-labels and manual annotations, demonstrating the potential of combining active learning with semi-supervised learning in practical EEG workflows [106].

2.7.3 Multimodal Learning and Active Learning

Recent advances in multimodal learning have highlighted the importance of jointly modelling complementary physiological modalities rather than relying solely on unimodal representations. Multimodal deep learning approaches have been explored across affective computing, biosignal analysis, fatigue monitoring, and BCI applications [107], [108].

A key challenge in multimodal systems concerns missing or incomplete modalities during acquisition. Ma et al. proposed SMIL, a multimodal learning framework capable of handling severely missing modalities while preserving multimodal representation quality [109]. Similarly, Liu et al. explored frequency-aware masked autoencoders for multimodal biosignal pretraining, demonstrating the effectiveness of self-supervised multimodal representation learning for physiological signals [110].

Recent work has also begun integrating active learning directly into multimodal systems. Jang et al. proposed cross-modal consistency-guided active learning for affective BCI systems, demonstrating that disagreement between modalities can provide informative acquisition signals for efficient annotation [111]. Similarly, Zhang and Zhu investigated multimodal active learning under limited paired-data settings, highlighting the importance of modality-aware acquisition strategies [112]. Rheude et al. further explored active modality acquisition strategies for multimodal learning systems [113].

2.7.4 Active Learning for Fatigue and Microsleep Monitoring

In fatigue and vigilance-monitoring applications, semi-supervised and active learning approaches have increasingly been explored as strategies for reducing manual annotation burden while exploiting large quantities of unlabeled physiological and behavioural data.

Li et al. proposed a fatigue-driving detection framework combining active sample selection with graph-based semi-supervised learning to improve fatigue classification performance under limited labeled data conditions [114]. Similarly, Rahmani et al. explored semi-supervised deep learning frameworks for driver drowsiness detection using convolutional neural networks [115]. Ahmed et al. and Hassan et al. further demonstrated the growing role of deep learning and transformer-based architectures in real-time drowsiness monitoring systems [116], [117].

Although these approaches demonstrate the potential of active and semi-supervised learning for fatigue monitoring, most existing studies focus primarily on drowsiness detection rather than microsleep forecasting across fixed future horizons.

2.8 Current Challenges in Microsleep Forecasting

Despite substantial progress in EEG-based microsleep analysis, several major challenges continue to limit the development of reliable real-world forecasting systems.

One of the most significant challenges is severe class imbalance. Microsleep events occur relatively infrequently compared with normal wakefulness, resulting in highly skewed datasets. Models trained on such datasets may become biased toward the majority class and fail to detect rare but critical microsleep events. Approaches such as focal loss [118], oversampling, synthetic data generation, and cost-sensitive learning have therefore been explored to mitigate imbalance effects [119], [120], [121], [122]. However, excessive oversampling may introduce redundancy and overfitting, while cost-sensitive strategies require careful optimisation for highly imbalanced physiological data.

Another major limitation is poor inter-subject generalisation. EEG signals exhibit substantial variability across individuals due to anatomical differences, electrode positioning, cognitive variability, and environmental conditions. This problem becomes even more challenging in real-world environments where noise and motion artifacts further increase signal variability. Recent studies have therefore explored transfer learning, unsupervised domain adaptation, adversarial learning, and domain-specific normalisation techniques to improve cross-subject robustness in EEG systems [123], [124], [125], [126], [127].

Forecasting microsleeps ahead of onset also remains substantially more difficult than instantaneous detection. As prediction horizons increase, discriminative temporal cues become weaker and more difficult to preserve. Existing studies consistently report declining forecasting performance at longer horizons [13], [128]. Developing architectures capable of maintaining temporal resolution while modelling long-range dependencies therefore remains an important research challenge. Recent advances in time-series forecasting and neural sequence modelling may offer useful directions for future microsleep forecasting architectures [129], [130].

Real-world deployment introduces additional challenges related to noise robust-

ness and ecological validity. Most existing microsleep datasets are collected under controlled laboratory conditions using constrained behavioural protocols. In practical environments such as transportation systems, industrial monitoring, healthcare environments, and safety-critical operations, EEG recordings become contaminated by motion artifacts, speech, muscle activity, and environmental interference [45]. Wearable and mobile EEG systems are particularly vulnerable to movement-related artifacts and electrode instability [14], [131], [132], [133]. Consequently, models trained exclusively on laboratory data may fail to generalise effectively under realistic operating conditions. Adaptive artifact suppression and multimodal physiological sensing strategies have therefore become increasingly important for robust deployment [29], [134].

Another underexplored issue concerns the volume of training data required for robust subject-independent forecasting. Deep learning models generally benefit from large datasets, yet collecting labelled EEG recordings remains expensive and time-consuming. Existing literature provides limited systematic investigation into how forecasting performance scales with training data volume, particularly under noisy conditions. This challenge is further amplified by the high variability of physiological fatigue responses across subjects and operational contexts [135], [136].

Finally, the trade-off between model complexity and practical deployment remains unresolved. While deep architectures and multimodal systems can achieve strong performance, they may require high computational resources, large electrode arrays, and extensive calibration procedures. Ensemble frameworks, transformer-based architectures, and hybrid deep learning systems may improve predictive performance but often introduce additional computational overhead and reduced interpretability [29], [137]. Real-world microsleep forecasting systems must therefore balance predictive performance with portability, interpretability, computational efficiency, and operational feasibility.

2.9 Research Gaps and Thesis Motivation

The literature reviewed in this chapter demonstrates substantial progress in EEG-based microsleep detection and prediction. Nevertheless, four gaps directly addressed in this thesis remain.

Gap 1: Systematic evaluation of physiologically meaningful EEG representations. Existing microsleep studies have examined spectral, statistical, and nonlinear EEG descriptors, but comparisons across feature families under a consistent subject-independent evaluation protocol remain limited. Consequently, it is unclear which EEG representations provide the most informative and robust indicators of microsleep-related activity. This gap is addressed by **RQ1**, which systematically evaluates spectral, nonlinear, complexity-based, and other hand-crafted EEG feature families for microsleep prediction.

Gap 2: Temporal modelling for multi-horizon microsleep forecasting. Existing research focuses predominantly on microsleep detection at or after event onset rather than forecasting across multiple pre-onset horizons. Forecasting performance also tends to deteriorate as the prediction horizon increases [13]. Although temporal attention, convolutional-transformer, and spatio-temporal models have shown potential in related EEG applications [81], [138], [139], their use for microsleep forecasting across multiple horizons remains limited. This gap is addressed by **RQ2**, which investigates attention-based temporal architectures designed to preserve information across sequential EEG windows.

Gap 3: Physiologically meaningful multimodal EEG–EOG forecasting. Despite growing interest in multimodal monitoring, the fusion of EEG and EOG for pre-onset microsleep forecasting remains underexplored. Existing multimodal systems have primarily addressed drowsiness, sleep staging, or general fatigue monitoring rather than microsleep forecasting [53], [86], [140]. Furthermore, relatively little work has modelled neural–ocular interactions using information-theoretic mea-

asures. This gap is addressed by **RQ3**, which evaluates EEG–EOG fusion strategies and interaction measures based on mutual information and transfer entropy.

Gap 4: Label-efficient multimodal microsleep forecasting. Current forecasting systems depend on substantial quantities of labelled physiological data, limiting their scalability. Although active learning has been examined in EEG classification and BCI calibration, its application to multimodal microsleep forecasting remains limited. In particular, modality-aware acquisition strategies that use disagreement between EEG and EOG have received little attention. Recent cross-modal active-learning and knowledge-distillation studies suggest that disagreement or consistency between modalities may support label-efficient learning [111], [141]. This gap is addressed by **RQ4**, which investigates disagreement-based active learning for label-efficient multimodal microsleep forecasting.

Subject-independent evaluation is used consistently across these four research questions to examine generalisation to unseen participants. This evaluation choice addresses cross-subject robustness within the available controlled dataset, but it does not constitute a complete investigation of real-world deployment.

Related gaps outside the scope of this thesis. Several broader challenges identified in the literature are not directly addressed in the present work. These include robustness to uncontrolled real-world artefacts, inter-session and cross-dataset generalisation, online adaptation, transfer learning with limited calibration data, and deployment using wearable or reduced-channel EEG systems [142], [143], [144], [145]. These remain important directions for future research but require datasets and experimental protocols beyond the scope of this thesis.

The four gaps addressed here therefore define a structured progression from physiological EEG representation (**RQ1**), through temporal forecasting (**RQ2**) and multimodal interaction modelling (**RQ3**), to label-efficient multimodal learning (**RQ4**). Collectively, the thesis investigates these questions under a consistent subject-independent evaluation framework.

2.10 Summary

This chapter reviewed the literature relevant to EEG-based microsleep detection and forecasting. The review began by examining the physiological and neurocognitive foundations of microsleeps and vigilance degradation, highlighting their significance in safety-critical environments. The chapter then discussed the neurophysiological basis of microsleeps, including spectral slowing, spindle-like activity, distributed cortical dynamics associated with transient vigilance failures, and evidence of brief post-microsleep neurophysiological recovery, referring to a temporary reduction in sleep pressure and a short-lived improvement in vigilance.

Subsequently, EEG and EOG signals were reviewed as complementary physiological modalities for microsleep analysis. Existing literature demonstrates that EEG provides direct access to cortical vigilance dynamics, whereas EOG captures ocular manifestations of fatigue and behavioural disengagement. Despite their complementary nature, their joint use in microsleep forecasting remains relatively unexplored.

The chapter further reviewed traditional machine learning approaches based on handcrafted spectral and information-theoretic features, followed by modern deep learning methods including CNNs, recurrent neural networks, and attention-based architectures. Although these approaches have substantially improved detection performance, forecasting microsleeps ahead of onset remains challenging, particularly at longer prediction horizons.

Multimodal systems integrating behavioural and physiological signals were also examined, together with active learning and label-efficient learning strategies for EEG analysis. While these approaches demonstrate considerable promise, important challenges remain regarding subject-independent generalisation, class imbalance, limited labelled data, and real-world robustness.

Finally, the chapter identified several research gaps motivating this thesis, including the need for improved multi-horizon forecasting architectures, physiologically meaningful multimodal fusion strategies, interpretable active learning frameworks, and robust subject-independent evaluation under ecologically valid condi-

tions. These gaps form the foundation for the methodologies and experiments presented in the subsequent chapters.

Chapter 3

Dataset and Evaluation Framework

3.1 Introduction

This chapter describes the dataset resources, annotation protocol, preprocessing pipeline, feature and representation construction, experimental protocol, statistical analysis, implementation details, and reproducibility measures used across this thesis. Its purpose is to provide a clear methodological foundation for the empirical work in Chapters 4–7, while also making explicit where individual research questions require task-specific adaptations.

The thesis investigates microsleep prediction from physiological signals, and this problem is especially sensitive to methodological choices. EEG and EOG signals vary substantially between subjects, microsleep events are short and difficult to separate from neighbouring drowsy states, and the resulting class distribution is imbalanced. These factors mean that small changes in preprocessing, windowing, labelling, or data splitting can strongly affect reported performance. For this reason, the thesis adopts a controlled framework based on subject-independent evaluation, consistent core preprocessing, and matched statistical comparisons.

At the same time, the framework should not be understood as a rigid claim that

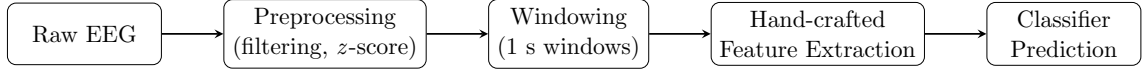


Figure 3.1: Simplified RQ1 pipeline for microsleep detection using hand-crafted EEG features. After preprocessing and one-second windowing, descriptive EEG features are extracted and used for classifier-based prediction.

every research question uses an identical pipeline. The core dataset, annotation logic, one-second base windowing, and subject-exclusive evaluation principles are shared across the thesis, but the representation and modelling stages differ according to the aim of each research question. RQ1 evaluates EEG feature families, RQ2 studies temporal forecasting, RQ3 investigates multimodal EEG–EOG fusion, and RQ4 examines label-efficient active learning. These differences are described explicitly in the relevant chapters rather than being presented under a single unified experimental methodology.

3.2 Datasets

3.2.1 Primary Dataset: Maintenance of Wakefulness Test (MWT)

The primary dataset used throughout this thesis is the Maintenance of Wakefulness Test (MWT) dataset. This dataset is a publicly available clinical resource collected under controlled laboratory conditions to capture transitions between wakefulness, drowsiness, and microsleep. Public datasets with expert-labelled microsleep events are rare, which makes the MWT dataset an important benchmark for reproducible microsleep research. Its cohort size of 76 participants also makes it suitable for subject-independent evaluation and statistical testing across folds [22], [146].

In the MWT protocol, each participant completes four trials across the day, with two trials before lunch and two trials after lunch. The trials are separated by at least two hours. During each trial, participants sit in a semi-darkened room and are instructed to remain awake for as long as possible. Each trial lasts up to 40 minutes, unless sleep-onset criteria are met earlier [22]. This setting is relevant to microsleep

research because it produces natural fluctuations in vigilance while maintaining a controlled acquisition environment.

The original recordings include EEG, EOG, that capture neural and ocular activity during the Maintenance of Wakefulness Test (MWT). The experiments in this thesis use the channels available in the released configuration used consistently in the associated studies: two occipital EEG derivations, O1–M2 and O2–M1, and two EOG derivations, E1–M1 and E2–M1. The signals are sampled at 200 Hz [146]. These channels are physiologically appropriate for microsleep modelling because occipital EEG captures changes in vigilance-related rhythms, while EOG captures ocular behaviour such as slow eye movements, eyelid closure, and other drowsiness-related changes.

3.2.2 Bern Scoring Criteria and Label Definition

A central strength of the MWT dataset is that the wake–sleep transition is annotated using the Bern continuous and high-resolution wake–sleep scoring criteria. This annotation framework is particularly important for this work because microsleep episodes are short events that may be missed or blurred by conventional sleep staging.

Traditional American Academy of Sleep Medicine (AASM) sleep scoring [22] uses fixed 30-second epochs for sleep-stage classification. This approach is appropriate for standard sleep staging, but it has limited temporal resolution for studying short events. Microsleep episodes may last only 1–15 seconds, and the transition from wakefulness to sleep may contain rapid fluctuations between alertness, drowsiness, and brief sleep-like patterns. A 30-second epoch can therefore obscure short microsleep-related events by averaging them with surrounding wakefulness.

The Bern criteria [22] were developed to address this limitation by providing continuous, high-resolution scoring of the wake–sleep transition zone. Rather than only assigning a coarse sleep-stage label to each 30-second epoch, the Bern approach identifies shorter microsleep-related events directly from the physiological

and behavioural evidence. This is particularly relevant in the MWT setting, where participants are attempting to stay awake and may repeatedly move between wakefulness, drowsiness, microsleep candidates, and microsleep episodes before reaching sustained sleep.

Under the Bern criteria, a microsleep episode (MSE) is defined as a short event, typically lasting 1–15 seconds, visible in at least one occipital EEG channel. The mandatory EEG characteristics include predominant, mostly irregular and polymorphic theta activity, poorly delimited or absent alpha activity compared with full wakefulness, and a simultaneous reduction in beta activity. The criteria also incorporate behavioural evidence, especially eyelid closure of at least 80% as estimated from face video. Slow eye movements may precede and persist during an MSE, and loss of muscle tone may also be observed [22].

The Bern framework also defines two intermediate categories that are important for modelling the transition zone. A microsleep episode candidate (MSEc) describes an event that resembles an MSE but does not satisfy all mandatory criteria. An episode of drowsiness (ED) describes a period with distinct boundaries but mixed and fluctuating EEG characteristics that cannot be clearly classified as wakefulness, MSE, or MSEc. These intermediate categories reflect the fact that microsleep is not always a sudden binary transition. Instead, it often emerges gradually through ambiguous physiological states.

The dataset representation used in this thesis contains four labels: wakefulness (0), microsleep episode (1), microsleep episode candidate (2), and episode of drowsiness (3). For the modelling experiments, wakefulness is treated as the negative class, while labels 1, 2, and 3 are merged into a single positive class representing a microsleep-transition state [146]. This label-merging decision is motivated by the early-warning objective of the thesis. In practical monitoring systems, the key requirement is to detect departure from stable wakefulness, including borderline or precursor states, rather than to distinguish every transitional subtype perfectly.

After segmentation into one-second windows and application of this binary label

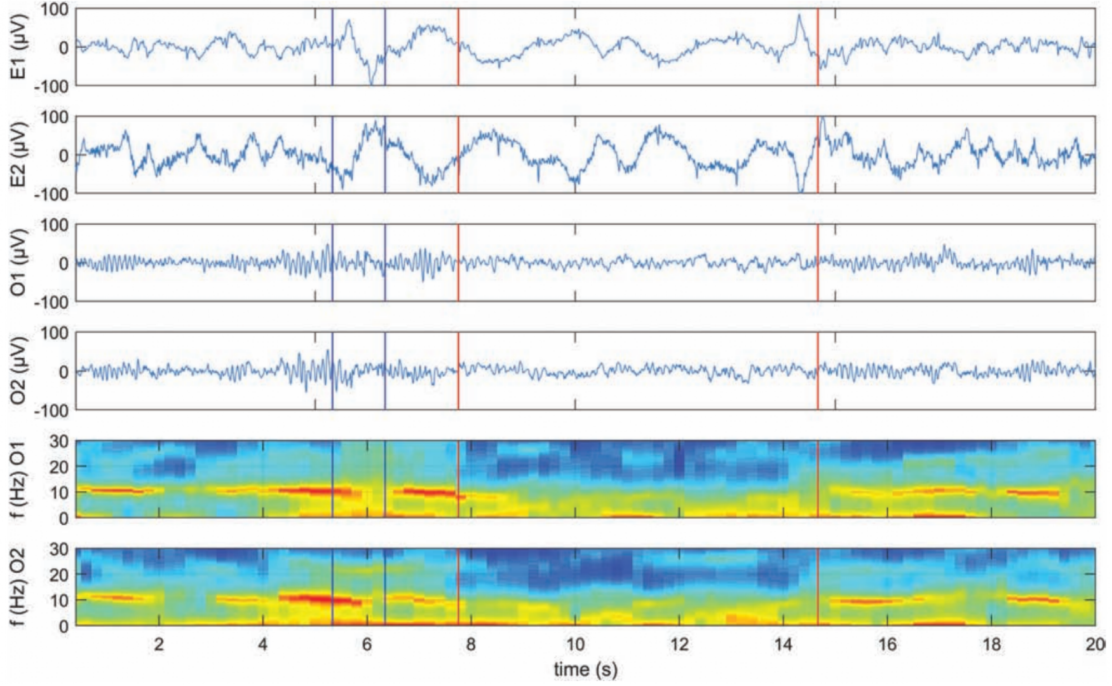


Figure 3.2: Example of microsleep-related events evaluated according to the Bern criteria. The blue vertical lines delimit a microsleep episode candidate (MSEc), and the red vertical lines delimit a microsleep episode (MSE). E1 and E2 show the EOG activity, O1 and O2 show the occipital EEG signals, and the lower panels present the corresponding O1 and O2 time–frequency profiles. During the MSEc interval, alpha activity becomes partially attenuated, but not all mandatory criteria for an MSE are satisfied. During the MSE interval, the occipital EEG demonstrates the principal electrophysiological characteristics of the Bern criteria: predominant irregular theta activity, poorly defined or absent alpha activity relative to wakefulness, and a simultaneous reduction in beta activity. The EOG channels provide complementary evidence of changes in eye-movement behaviour. The event boundaries and classifications were determined from the expert annotations supplied with the dataset.

formulation, the established MWT pipeline contains approximately 60,548 wakefulness windows and 37,456 microsleep-transition windows [146]. This creates a realistic imbalanced setting. The imbalance is important when interpreting results because accuracy alone may overstate model performance if the majority wakefulness class dominates the prediction. For this reason, later sections define sensitivity, specificity, F1-score, AUC-ROC, and AUC-PR as complementary evaluation metrics.

The MWT dataset is therefore well suited to the aims of this thesis. It provides clinically meaningful annotations, high temporal resolution, multimodal EEG-EOG recordings, and sufficient inter-subject variability to support subject-independent model evaluation.

3.2.3 Additional Datasets

To examine whether the conclusions regarding temporal modelling extend beyond the controlled MWT setting, RQ2 additionally reports experiments on the *Sleepy Driver EEG Brainwave Dataset* [128], [147]. This dataset provides a complementary evaluation scenario that differs substantially from MWT in terms of acquisition setup, signal characteristics, and data scale.

The Sleepy Driver dataset [147] was collected from four participants using the NeuroSky MindWave headset, a low-cost, single-channel consumer EEG device. Unlike the multi-channel clinical recordings used in MWT, this dataset relies on a single frontal electrode, with reference and ground electrodes placed on the earlobes. The recordings consist of approximately 5–8 minutes per subject in both awake and drowsy conditions, yielding labelled segments for binary classification (sleepy vs. not sleepy). The signals are sampled at approximately one-second intervals and represent power spectral density (PSD) measurements (in $\mu V^2/\text{Hz}$), rather than raw time-series EEG.

These characteristics introduce several important differences. The use of a single-channel device reduces spatial information and increases susceptibility to noise, while the shorter recording duration and smaller number of subjects limit statistical di-

versity. In addition, the dataset provides precomputed spectral features rather than continuous raw EEG, which constrains the range of feature extraction strategies that can be applied. As a result, this dataset represents a more constrained and lower-fidelity setting compared to MWT.

Despite these limitations, the dataset is valuable for assessing whether the relative behaviour of temporal models, particularly their ability to capture sequential structure, remains observable under reduced signal quality and simplified acquisition conditions. The goal is therefore not to directly compare absolute performance between datasets, but to evaluate whether the trends identified in MWT persist in a different and more resource-constrained context.

Aside from these cross-dataset experiments in RQ2, the thesis primarily relies on the MWT dataset to maintain consistency across RQ1, RQ3, and RQ4. This ensures that comparisons between feature representations, model architectures, and modality integration strategies are conducted under a controlled and uniform experimental framework. When multiple datasets are used, consistent methodological principles are maintained where possible, including subject-exclusive splitting, one-second segmentation, and the use of imbalance-aware evaluation metrics [128].

The results obtained from the Sleepy Driver dataset are interpreted cautiously. Rather than serving as definitive evidence of generalisation, they are treated as supportive observations that help validate whether the advantages of temporal modelling persist under different acquisition conditions. This distinction is important, as differences in hardware, signal representation, and dataset scale limit the extent to which conclusions can be transferred directly across datasets.

3.3 Signal Preprocessing Pipeline

All experimental pipelines begin with preprocessing of the raw physiological signals prior to feature extraction or model training. The purpose of preprocessing is to standardise signal characteristics, reduce avoidable noise, and support fair comparison across research questions. The main preprocessing choices are fixed at the thesis

level where possible, especially the use of one-second windows. However, some task-specific processing details differ between RQs because the modelling objectives are not identical.

3.3.1 Filtering

The MWT signals are sampled at 200 Hz [22]. Filtering is used to retain physiologically meaningful EEG and EOG activity while reducing low-frequency drift, high-frequency noise, and power-line interference. In the acquisition environment, standard clinical filtering practices are applied, including power-line notch filtering and impedance control [22].

In the thesis experiments, filtering follows the relevant modelling pipeline. For RQ3 and RQ4, which focus on multimodal EEG–EOG fusion and label-efficient learning, the filtering configuration follows the multimodal study protocol [146]. For RQ2, where temporal forecasting from EEG is central, the filtering configuration follows the temporal modelling pipeline to ensure consistent input characteristics for sequence-based models [128]. These differences are methodological rather than contradictory: the common objective is to preserve microsleep-relevant physiological activity while adapting filtering details to the specific representation used in each chapter.

3.3.2 Artifact Handling

Artifact handling is an important preprocessing step because EEG and EOG recordings are often contaminated by artifacts, including muscle activity, electrode movement, electrical interference, and other non-physiological signal disturbances [7]. Such artifacts can distort feature distributions and reduce model reliability if not addressed appropriately. To ensure reproducibility, artifact handling is implemented using deterministic and repeatable procedures rather than manual cleaning. All signals are standardised per subject using z-score normalisation, which reduces inter-individual amplitude differences and stabilises feature distributions [146]. This is

important because EEG and EOG amplitudes can vary across participants due to physiological variability, electrode placement, impedance, and recording conditions.

More complex artifact-correction methods, such as ICA-based ocular artifact removal, can be useful in some EEG applications. However, they also introduce additional parameter choices and may require manual inspection, which can reduce reproducibility and create fold-dependent decisions. For this reason, the thesis adopts conservative standardisation and robust modelling rather than aggressive manual artifact rejection. This choice supports a scalable and reproducible pipeline, while acknowledging that more specialised artifact correction could be investigated in future work [45].

3.3.3 Epoching Strategy

A consistent segmentation strategy is used to convert continuous recordings into fixed-duration windows suitable for detection and forecasting. For all MWT-based experiments, one-second windows form the base temporal unit. This choice is directly aligned with the Bern scoring criteria, which were designed to capture microsleep-related events at a much higher temporal resolution than conventional 30-second sleep staging.

The one-second window size is fixed across the thesis, but the use of overlap depends on the research question. In RQ1 and the feature-based multimodal pipelines, one-second windows are used as feature extraction units. In RQ2, overlapping windows are used to support temporal forecasting and sequence construction. In RQ4, the active learning pool is built from the same feature representation used in RQ3 so that query strategies can be compared under identical input conditions.

Window-level labels follow the formulation described in Section 3.2.2. Wakefulness is treated as the negative class and the merged microsleep-transition label is treated as the positive class. For forecasting studies, labels are shifted to construct future prediction targets. For example, RQ2 evaluates onset and short-horizon prediction, while RQ4 focuses on 1.0 s and 2.0 s ahead prediction in an active learning

setting. These forecasting targets are generated from the same one-second base segmentation, preserving a consistent temporal unit while allowing task-specific prediction horizons.

3.4 Feature Construction

Although each research question introduces a distinct modelling contribution, all representations originate from the same preprocessed physiological signals. This section summarises how task-specific inputs are constructed, while the detailed feature definitions and modelling choices are reserved for the corresponding RQ chapters.

3.4.1 Handcrafted Features (RQ1)

RQ1 focuses on the systematic evaluation of handcrafted EEG features for microsleep detection. The purpose is to determine which EEG-derived descriptors provide useful information under subject-independent evaluation. The feature families include time-domain descriptors, frequency-domain measures, time-frequency features, bispectrum descriptors, nonlinear dynamics features, and intensity-weighted spectral descriptors [148]. These features capture complementary properties of the EEG signal, including amplitude variability, spectral power distribution, transient time-frequency behaviour, higher-order frequency interactions, and complexity.

The full feature definitions, classifier comparisons, feature-family results, and statistical validation are presented in Chapter 4. In the present chapter, the important point is that RQ1 uses the shared MWT dataset and focuses on EEG-only handcrafted feature representations. This design choice was made to establish the contribution of EEG-derived features in isolation before introducing additional physiological information. Multimodal EEG–EOG integration is subsequently investigated in RQ3.

3.4.2 Time-Series Input for Temporal Models (RQ2)

RQ2 addresses a limitation of static window-based modelling: one-second windows treated independently may fail to capture the temporal evolution that precedes microsleep onset. To address this, RQ2 constructs sequence-based inputs from windowed EEG features and trains temporal models, including recurrent and attention-augmented architectures [128].

This design supports the forecasting objective of RQ2, where the model is required not only to detect microsleep at onset but also to predict events ahead of time. Prediction horizons corresponding to microsleep onset (0 sec) and short-term pre-onset forecasting intervals (1 sec and 2 sec before microsleep onset) are evaluated and an additional cross-dataset experiment is conducted. The details of the temporal architecture, focal loss, and horizon-specific evaluation are given in Chapter 5.

3.4.3 Multimodal Representation (RQ3)

RQ3 extends the thesis from EEG-only modelling to multimodal EEG–EOG fusion. Each one-second window is represented using EEG descriptors, EOG descriptors, and information-theoretic coupling measures. Mutual information is used to capture nonlinear dependencies within and across modalities, while transfer entropy is evaluated as a directional measure in the RQ3 study [146].

This representation supports controlled comparison between unimodal baselines, MI-enhanced unimodal models, early fusion, late fusion, and transfer-entropy augmentation. The motivation is that EEG and EOG capture complementary aspects of microsleep physiology: EEG reflects neural vigilance dynamics, while EOG reflects ocular behaviour associated with drowsiness and eye closure. The details of the multimodal feature set and fusion strategies are presented in Chapter 6.

3.4.4 Active Learning Sampling Pool (RQ4)

RQ4 builds directly on the multimodal representation established in RQ3. The goal is not to redesign the predictive architecture, but to evaluate whether com-

parable forecasting performance can be achieved with substantially fewer labelled samples. In this setting, features are computed once offline and reused across all active learning strategies. This ensures that differences in performance arise from the sample-selection policy rather than from representation drift.

The RQ4 active learning setting uses a pool-based framework under LOSO evaluation. Labels in the training pool are hidden, a small seed set is initially labelled, and query strategies progressively select additional samples for annotation. Standard strategies such as random sampling, uncertainty sampling, diversity sampling, and query-by-committee are compared with the proposed multimodal disagreement active learning strategy. The proposed strategy uses disagreement between EEG and EOG predictions, fusion-model boundary proximity, and a mild rare-event bias toward likely microsleep samples to guide sample selection. The detailed active learning design is presented in Chapter 7.

3.5 Experimental Protocol

3.5.1 Train–Validation–Test Split

Subject leakage is a major risk in microsleep modelling because adjacent windows from the same participant are highly correlated. If windows from one participant appear in both training and testing sets, the model may learn subject-specific characteristics rather than generalisable microsleep patterns. This can lead to over-optimistic performance estimates. For this reason, the thesis prioritises subject-exclusive evaluation throughout.

In RQ3 and RQ4, strict Leave-One-Subject-Out (LOSO) cross-validation is used. In each fold, one participant is held out entirely for testing, while the remaining participants form the training pool [146]. This design directly evaluates subject-independent generalisation and produces paired subject-wise results that can be used for matched statistical testing.

RQ1 and RQ2 also maintain subject exclusivity, although the exact split de-

sign is adapted to the task. RQ1 uses subject-exclusive evaluation for feature and classifier benchmarking. RQ2 uses subject-exclusive stratified splits and related subject-independent analyses for temporal modelling and cross-dataset evaluation [128], [146]. These differences are stated in the relevant chapters so that each result is interpreted in the correct experimental context.

3.5.2 Class Imbalance Handling

Microsleep-transition windows occur less frequently than wakefulness windows, producing a class imbalance that can bias models and distort accuracy-based interpretation. This is particularly important in safety-related applications, where missing microsleep events may be more consequential than misclassifying a small number of wakefulness windows.

Across the thesis, imbalance is addressed first through metric selection. Performance is evaluated using metrics that are more informative than accuracy in imbalanced settings, particularly Sensitivity, F1-score, and Area Under the Precision–Recall Curve (AUC–PR), which are defined in Section 3.5.3. In some experiments, model-level mechanisms are also used. For example, RQ2 uses focal loss to place greater emphasis on difficult minority-class samples [118], [128]. In RQ4, imbalance is handled carefully because the labelled subset changes as the active learning budget increases. Any imbalance control is applied consistently across query strategies so that performance differences reflect acquisition behaviour rather than unequal class treatment.

3.5.3 Evaluation Metrics

Performance is evaluated using a complementary set of metrics. Accuracy is reported because it provides an intuitive measure of overall correctness, but it is not treated as sufficient in isolation. In imbalanced microsleep prediction, high accuracy may occur even when the model under-detects microsleep-transition events. Therefore, accuracy is interpreted alongside class-specific and threshold-independent metrics.

Sensitivity, also referred to as recall or true positive rate, measures the proportion of microsleep-transition windows correctly identified. This metric is particularly important for early-warning systems because low sensitivity corresponds to missed events. Specificity measures the proportion of wakefulness windows correctly identified and is important for controlling false alarms.

Precision measures the proportion of predicted positive samples that are correctly classified as positive. Together, precision and recall (sensitivity) characterise the trade-off between false alarms and missed detections. The F1-score combines precision and recall through their harmonic mean,

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

and is particularly useful for imbalanced classification problems where both false positives and false negatives are important [149], [150].

The Area Under the Receiver Operating Characteristic Curve (AUC-ROC) measures discrimination across decision thresholds by considering the trade-off between sensitivity and false positive rate. However, AUC-ROC can remain overly optimistic in highly imbalanced settings. The Area Under the Precision-Recall Curve (AUC-PR) summarises the trade-off between precision and recall across decision thresholds. Unlike AUC-ROC, AUC-PR focuses directly on performance for the positive class and therefore provides a more informative assessment in highly imbalanced classification problems [151], [152].

In RQ4, additional active-learning-specific summaries such as the area under the learning curve and the annotation budget required to reach a fraction of baseline performance are used to quantify label efficiency.

3.6 Statistical Analysis

Formal statistical testing is used to support the interpretation of model comparisons. Under LOSO evaluation, each held-out subject produces one performance value per

model and metric. This creates paired subject-wise samples, allowing within-subject comparisons without assuming normality [146].

For comparisons involving three or more models, the Friedman test is used as a non-parametric omnibus test for detecting overall differences across multiple related samples without assuming normality [153], [154]. When the Friedman test indicates a significant difference, pairwise Wilcoxon signed-rank tests are applied to compare specific model pairs. The Wilcoxon signed-rank test is a non-parametric paired test that evaluates whether two related samples differ significantly in their median performance [154], [155].

Corrections for multiple comparisons, such as Bonferroni or Holm–Bonferroni correction, are used to control the family-wise error rate and reduce the probability of false-positive findings when multiple hypothesis tests are conducted [156].

Where appropriate, effect sizes such as Cohen’s d are reported to complement statistical significance with practical interpretation. Cohen’s d quantifies the magnitude of the observed difference, independent of sample size, allowing statistically significant results to be interpreted in terms of practical relevance [157].

This statistical framework is particularly important because the thesis does not rely only on point estimates. Subject-wise variability, paired comparisons, and corrected post-hoc tests help determine whether observed differences are likely to reflect systematic methodological effects rather than random variation across folds.

3.7 Implementation Details

All experiments are implemented in Python using standard scientific computing and machine learning tools. Classical pipelines and splitting procedures follow scikit-learn conventions. Temporal deep learning models in RQ2 are implemented using PyTorch [128]. Multimodal modelling in RQ3 and active learning experiments in RQ4 use LightGBM as the main classifier family [146]. Using a consistent classifier family in RQ3 and RQ4 helps ensure that improvements are attributable to the representation, fusion design, or query strategy rather than to changes in model

class.

Training procedures are controlled as far as possible through deterministic random seeds, fixed evaluation scripts, and consistent hyperparameter selection within each research question. Feature extraction is modular and, where appropriate, performed once offline before model comparison. This is especially important in RQ4, where all active learning strategies must operate on identical feature representations, labelled budgets, and subject splits to support fair comparison.

3.8 Reproducibility Statement

Reproducibility and controlled comparison are ensured through several principles. First, the MWT dataset and Bern-label formulation provide a consistent experimental foundation. Second, one-second windowing is used as the base temporal unit across the thesis. Third, subject-exclusive splitting prevents leakage between training and testing data. Fourth, statistical comparisons are conducted on paired subject-wise outcomes under matched folds. Finally, modular implementation ensures that dataset construction and feature extraction are reused across experiments wherever possible, especially in the multimodal and active learning chapters.

The thesis also explicitly documents task-specific differences between research questions. RQ1, RQ2, RQ3, and RQ4 address different methodological problems, and therefore use different representations and, in some cases, different forecasting horizons or split designs. These differences are not weaknesses; they are necessary for answering the specific research questions. The important requirement is that they are clearly stated, reproducible, and interpreted within the correct methodological context.

3.9 Chapter Summary

This chapter established the dataset resources, annotation decisions, preprocessing pipeline, representation pathways, evaluation protocol, statistical analysis, imple-

mentation details, and reproducibility measures used throughout the thesis. The MWT dataset serves as the primary benchmark because it combines controlled acquisition, EEG–EOG recordings, high-resolution Bern annotations, and a cohort size suitable for subject-independent evaluation [22], [146].

A key methodological point is the use of the Bern criteria rather than relying only on conventional 30-second AASM sleep staging. The Bern criteria provide a more suitable annotation framework for microsleep research because they capture short events and ambiguous wake–sleep transition states at higher temporal resolution. The thesis reformulates these annotations into a binary wakefulness versus microsleep-transition task to support early-warning modelling.

The next chapter builds on this framework to address RQ1, focusing on EEG feature evaluation for microsleep detection. It establishes the feature-level baseline that motivates the later temporal modelling, multimodal fusion, and label-efficient learning chapters.

Chapter 4

Feature Representation for Microsleep Prediction

Introduction

This chapter addresses the following research question:

RQ1: What EEG-derived features most effectively indicate microsleep episodes?

It establishes an EEG-only baseline for the thesis by systematically evaluating handcrafted EEG feature families and comparing classical and neural classifiers under a consistent experimental framework. The aim is not to claim that static feature-based modelling solves microsleep detection, but to identify which EEG descriptors provide useful evidence, where they fail, and how these limitations motivate the temporal modelling work in Chapter 5.

RQ1 focuses on microsleep detection and feature representation rather than forecasting. The objective is to identify informative EEG feature representations that can reliably distinguish microsleep and non-microsleep states and serve as the foundation for the forecasting models investigated in RQ2–RQ4. Although detection and forecasting are closely related tasks, they address different temporal objectives. In this thesis, detection refers to identifying microsleep events at or near onset, whereas

forecasting refers to predicting future microsleep occurrence before onset.

Microsleep episodes are brief and occur within the unstable transition between wakefulness and sleep. Their physiological signatures may appear as spectral slowing, changes in signal complexity, altered variability, or short-lived nonlinear patterns rather than as obvious changes in raw EEG amplitude. Under subject-independent evaluation, these signatures must remain informative despite differences between participants in baseline rhythms, signal amplitude, arousal state, and noise characteristics. This makes microsleep detection a demanding test case for EEG feature engineering.

Using the MWT dataset and the evaluation framework described in Chapter 3, this chapter evaluates six EEG feature families: time-domain, frequency-domain, bispectrum, nonlinear dynamics, time-frequency, and intensity-weighted spectral descriptors. These features are assessed using a common preprocessing and splitting strategy, and are paired with multiple classifiers to separate the effect of feature representation from the effect of model capacity. The experiments use the Maintenance of Wakefulness Test recordings [22], [158] and the binary wakefulness versus microsleep-transition formulation described in Chapter 3.

A central point of this chapter is that accuracy alone is not a reliable indicator of performance in this task. Several models obtain high accuracy by strongly favouring wakefulness, but this can coincide with poor sensitivity to microsleep-transition events. This is particularly clear for time-frequency features, which achieve high accuracy and specificity but can under-detect the positive class. Therefore, the interpretation throughout this chapter prioritises the trade-off between sensitivity and specificity rather than treating the highest accuracy as the best result.

4.1 Experimental Design and Reproducibility Details

The RQ1 experiments were designed to provide a controlled comparison of EEG feature representations. The analysis used the MWT dataset introduced in Chapter 3. The EEG channels used for feature extraction were the occipital derivations O1–M2 and O2–M1, sampled at 200 Hz. The task was formulated as binary classification, with wakefulness treated as the negative class and microsleep-related states treated as the positive class, following the merged-label formulation described previously.

Continuous EEG recordings were segmented into 1-second windows with 50% overlap. Windows containing mixed behavioural labels were excluded to ensure label consistency within each segment and to prevent assigning a single class label to physiologically heterogeneous data [148], [159]. This design choice improves the reliability of feature-label alignment and enables a controlled evaluation of feature representations under clearly defined conditions. It is important to note that this window-level formulation focuses on well-defined segments rather than transitional boundaries. While this simplifies the learning problem, it provides a necessary baseline for assessing the discriminative value of static EEG features in isolation. The challenge of modelling transitional dynamics is addressed in Chapter 5 through temporal modelling approaches.

The extracted EEG feature families include time-domain statistics, frequency-domain descriptors, bispectrum descriptors, nonlinear dynamics measures, time-frequency wavelet features, and intensity-weighted spectral descriptors. In the full RQ1 feature-evaluation pipeline, 164 features were extracted across these families [148]. Each feature family was evaluated separately so that its individual discriminative value could be assessed. This design avoids the interpretability problem that would arise if all features were concatenated from the beginning, where improved performance could not easily be attributed to a specific physiological representation.

To reduce subject-dependent bias, the data were split using a subject-exclusive

strategy. Data from the same participant were not allowed to appear in both training and testing partitions. Stratification was applied where appropriate to preserve class proportions. The RQ1 implementation used an 80% training and 20% testing split at the subject level, with all models evaluated on unseen subjects. This is less exhaustive than the LOSO protocol used in later chapters, but it preserves the essential requirement that the test set contains participants not seen during training.

For model evaluation, predictions were binarised using a threshold of 0.5 where applicable, and confusion matrices were used to derive sensitivity and specificity. Accuracy, sensitivity, and specificity are reported consistently across feature families and classifiers. Accuracy describes overall correctness, sensitivity measures the ability to detect microsleep-transition windows, and specificity measures the ability to correctly identify wakefulness. The interpretation in this chapter gives particular weight to sensitivity and specificity because high accuracy alone can be misleading under imbalance or conservative prediction behaviour.

4.2 Systematic Evaluation of EEG Feature Families

This section describes the six EEG feature families evaluated in RQ1. The features are motivated by different assumptions about how microsleep-related changes may appear in EEG. Some features capture simple amplitude statistics, others describe spectral redistribution, and others attempt to quantify complexity or non-linear structure.

4.2.1 Time-domain descriptors

Time-domain descriptors provide a compact statistical summary of the EEG amplitude distribution within each window. The evaluated statistics included the mean, standard deviation, maximum, minimum, and mean crossing rate [148]. These measures are computationally inexpensive and easy to interpret. The mean and extrema

capture shifts in signal level and transient excursions, while standard deviation reflects within-window variability. Mean crossing rate provides a coarse indication of oscillatory behaviour by counting how often the signal crosses its mean.

Despite these advantages, time-domain descriptors are limited for microsleep detection. Microsleep-related EEG changes are often expressed through frequency-specific slowing or altered signal complexity rather than through large amplitude changes alone. As a result, time-domain features provide a useful low-complexity baseline but are not expected to capture the full physiological structure of the wake-sleep transition [9].

4.2.2 Frequency-domain descriptors

Frequency-domain features were computed from power spectral density estimates and describe how EEG energy is distributed across frequency bands. The evaluated descriptors included absolute band power, relative band power, logarithmic power, peak frequency, and low/high frequency ratios [12], [148]. These features are physiologically motivated because reductions in vigilance are often associated with changes in theta, alpha, and beta activity [8], [9].

Relative band power is particularly useful in subject-independent settings because it partially reduces the influence of global amplitude differences between participants. However, stationary frequency-domain descriptors can still be limited when discriminative microsleep evidence occurs only during a short sub-interval of the window. In such cases, a single spectral summary may smooth over the transient pattern that is most relevant for detection.

4.2.3 Bispectrum descriptors

Bispectrum features were included to capture higher-order interactions between frequency components. Unlike standard power spectral features, which describe second-order structure, bispectrum features can reflect quadratic phase coupling and nonlinear frequency interactions. The evaluated descriptors included normalised

bispectrum entropy, logarithmic magnitude, diagonal log magnitude, and energy-weighted centres [148].

These features are motivated by the possibility that microsleep is associated not only with changes in power but also with changes in coupling structure. However, bispectrum estimation can be sensitive to window length and noise. In a subject-independent setting, such features may be unstable if coupling patterns vary strongly across individuals [142], [160].

4.2.4 Nonlinear dynamics descriptors

Nonlinear dynamics features were used to capture changes in EEG complexity and variability during reduced vigilance. This family included Poincaré plot indices, specifically SD1, SD2, and the SD1/SD2 ratio, as well as entropy measures such as Rényi and Tsallis entropy [148]. These descriptors are intended to represent the structure of the signal beyond simple amplitude or frequency summaries.

The Poincaré descriptors quantify short- and longer-term variability geometrically, while entropy measures quantify distributional irregularity. Such features are relevant because microsleep may involve changes in neural regularity and reduced signal complexity. In the feature importance analysis later in this chapter, Poincaré-derived measures appear among the strongest predictors, suggesting that signal complexity is an important EEG indicator in this dataset.

4.2.5 Time-frequency descriptors

Time-frequency descriptors were extracted using a discrete wavelet transform with a Daubechies-4 mother wavelet. Summary statistics, including mean, standard deviation, maximum, and minimum, were computed from the wavelet coefficients at each decomposition level [148]. These features are intended to capture transient and nonstationary changes that may be blurred by conventional spectral estimates.

The results later in this chapter show that time-frequency features can achieve very high accuracy and specificity. However, this is not sufficient evidence that they

are the best representation for microsleep detection. In several cases, high accuracy is accompanied by reduced sensitivity, meaning that the classifier is effective at identifying wakefulness but less reliable at detecting microsleep-transition windows. This illustrates why accuracy must be interpreted cautiously in this chapter.

4.2.6 Intensity-weighted spectral descriptors

Intensity-weighted descriptors summarise the spectral distribution through weighted moments rather than fixed band boundaries. The two descriptors evaluated were the *Intensity-Weighted Mean Frequency (IWMF)* and the *Intensity-Weighted Bandwidth (IWBW)* [148]. Let $P(f)$ denote spectral intensity at frequency f . The IWMF is defined as

$$\text{IWMF} = \frac{\sum_f f P(f)}{\sum_f P(f)}, \quad (4.1)$$

and the IWBW is defined as

$$\text{IWBW} = \sqrt{\frac{\sum_f (f - \text{IWMF})^2 P(f)}{\sum_f P(f)}}. \quad (4.2)$$

IWMF captures the spectral centre of mass, while IWBW captures dispersion around that centre. These features are useful because microsleep-related spectral slowing may shift the distribution of spectral energy without aligning perfectly with fixed frequency bands. In the consolidated results, intensity-weighted features provide a more balanced sensitivity–specificity profile than several other feature families.

4.3 Classifier Benchmarking

To isolate the effect of feature representation from model capacity, each feature family was evaluated using a consistent set of classifiers spanning probabilistic, margin-based, instance-based, ensemble, and neural approaches. All models were

implemented using standard libraries and applied without extensive hyperparameter tuning, so that differences in performance primarily reflect feature characteristics rather than model optimisation.

Classical machine learning models

The following classical classifiers were evaluated. They were selected to provide complementary assumptions and levels of model complexity.

- **Gaussian Naïve Bayes** [161]: Implemented with default parameters, this model assumes conditional independence between features and Gaussian class-conditional distributions. Its principal advantage is its low computational cost and ability to provide a simple probabilistic baseline, particularly for relatively small datasets. Its disadvantage is that the independence and Gaussian assumptions are unlikely to hold for correlated and nonlinearly distributed EEG features. In practice, the model often produced high sensitivity but very low specificity.
- **Linear Discriminant Analysis (LDA)** [162]: LDA models each class using a Gaussian distribution with a shared covariance matrix and constructs a linear decision boundary. Its principal advantage is that it is computationally efficient, interpretable, and able to account for correlations between features. Its disadvantage is that it assumes approximately Gaussian class distributions and equal covariance matrices, and its linear boundary may not capture complex nonlinear relationships in EEG data.
- **Support Vector Machine (SVM)** [163]: Implemented using a radial basis function (RBF) kernel with $C = 1.0$ and $\gamma = \text{scale}$. Probability estimation was enabled to support consistent thresholding. Its principal advantage is its ability to learn nonlinear decision boundaries and operate effectively in high-dimensional feature spaces. Its disadvantage is its sensitivity to feature scaling

and kernel hyperparameters, together with its limited interpretability and the additional computational cost of probability calibration.

- **k -Nearest Neighbours (KNN)** [164]: Implemented using $k = 5$ neighbours and Euclidean distance, with predictions determined by the majority class among the nearest observations. Its principal advantage is that it is a simple non-parametric method capable of representing local and nonlinear class structure. Its disadvantage is its sensitivity to feature scaling, irrelevant dimensions, and the choice of k . Its performance may also deteriorate in high-dimensional spaces because distances become less discriminative.
- **Random Forest** [165]: Implemented using 100 decision trees and a fixed random seed of 42. Default splitting criteria and tree-depth settings were retained. Its principal advantage is its ability to capture nonlinear relationships and feature interactions while reducing the variance of individual decision trees through bootstrap aggregation. It also provides estimates of feature importance. Its disadvantage is that the ensemble is less interpretable than an individual tree and may overfit noisy or high-dimensional features if model complexity is not appropriately controlled.
- **Gradient Boosted Trees (XGBoost)** [166]: Implemented using a logistic objective function for binary classification and log-loss as the evaluation metric. Default hyperparameters were retained. Its principal advantage is its strong ability to model complex nonlinear relationships by constructing trees sequentially to correct previous errors. Its disadvantage is its sensitivity to hyperparameter selection and its potential to overfit, particularly for small datasets. The resulting ensemble is also less directly interpretable than a linear model.

Together, these classifiers provide complementary perspectives, ranging from simple probabilistic and linear assumptions to local, kernel-based, and ensemble-based nonlinear learning.

Neural network models

Deep learning models have become increasingly popular in EEG analysis because they can automatically learn discriminative spatial and temporal representations from physiological signals without extensive manual feature engineering [167], [168]. Two neural architectures were included as baseline models for comparison with the classical machine learning approaches.

- **Convolutional Neural Network (CNN)** [169]: The CNN was implemented as a one-dimensional convolutional model operating on feature sequences. The architecture consisted of:
 - A Conv1D layer with 32 filters, kernel size 3, ReLU activation, and same padding
 - A MaxPooling1D layer with pool size 2
 - A second Conv1D layer with 64 filters and ReLU activation
 - A global average pooling layer
 - A fully connected output layer with sigmoid activation

The model was trained using the Adam optimiser and binary cross-entropy loss.

- **Long Short-Term Memory (LSTM)** [170]: The LSTM model was implemented as a stacked recurrent architecture consisting of:
 - An LSTM layer with 64 hidden units (returning sequences)
 - A second LSTM layer with 32 hidden units
 - A fully connected output layer with sigmoid activation

The model was trained using the Adam optimiser and binary cross-entropy loss.

For both neural models, outputs were converted to binary predictions using a threshold of 0.5.

Role of classifier diversity

The inclusion of multiple classifier families is important for interpreting the results. Simpler models such as Naïve Bayes highlight the effect of violated assumptions, while SVM and KNN provide alternative nonlinear and local decision strategies. Tree-based ensembles offer greater flexibility in modelling feature interactions, and neural models provide a baseline for learning hierarchical representations.

Evaluating all feature families across this diverse set of classifiers ensures that observed performance trends are not specific to a single modelling approach. Instead, consistent patterns across classifiers can be interpreted as evidence of the intrinsic discriminative value of the feature representation.

4.3.1 Feature-family performance under representative classifiers

Tables 4.1 and 4.2 report consolidated feature-family performance using LDA and Random Forest. These tables are useful because they show the behaviour of each representation under a simpler linear classifier and a nonlinear ensemble classifier.

The most important observation is that the highest accuracy does not necessarily correspond to the most useful microsleep detector. Time-frequency features achieve the highest accuracy, reaching 94% in both consolidated tables. However, this high accuracy is accompanied by low sensitivity, especially under LDA, where sensitivity is only 42%. This means that many microsleep-transition events are missed despite the strong overall accuracy. Therefore, the time-frequency result should be interpreted as a high-specificity result, not as evidence of universally superior microsleep detection.

In contrast, frequency-domain, nonlinear, and intensity-weighted features show more moderate but more balanced results. Frequency-domain features provide reasonable sensitivity and specificity under both LDA and Random Forest. Intensity-weighted features produce the highest sensitivity under LDA, while nonlinear and intensity-weighted features remain among the more balanced options across classi-

fiers. This suggests that the most useful static EEG indicators are not necessarily those that maximise accuracy, but those that maintain a workable trade-off between detecting microsleep and avoiding false alarms.

Table 4.1: Performance of feature families using LDA. These point estimates should be interpreted jointly across accuracy, sensitivity, and specificity; high accuracy alone is not sufficient for microsleep detection.

Feature Type	Accuracy	Sensitivity	Specificity
Time Domain	56%	69%	42%
Frequency Domain	67%	66%	68%
Bispectrum	57%	58%	56%
Nonlinear	59%	69%	50%
Time-Frequency	94%	42%	100%
IWMF, IWBW	63%	73%	54%

Table 4.2: Performance of feature families using Random Forest. Time-frequency features achieve high accuracy and specificity, while frequency-domain, nonlinear, and intensity-weighted features provide more balanced sensitivity–specificity behaviour.

Feature Type	Accuracy	Sensitivity	Specificity
Time Domain	56%	57%	55%
Frequency Domain	69%	68%	70%
Bispectrum	59%	61%	58%
Nonlinear	65%	60%	69%
Time-Frequency	94%	50%	99%
IWMF, IWBW	65%	64%	66%

4.3.2 Model performance across feature families

Tables 4.3–4.8 report the performance of the models for each feature family. These results highlight the main conclusion of this chapter: although static EEG features contain meaningful predictive information, models based on these features struggle to achieve consistently high sensitivity and specificity simultaneously.

Naïve Bayes often achieves very high sensitivity but extremely low specificity. For example, with bispectrum features it classifies nearly all positive samples correctly but fails completely on the negative class. This behaviour is not useful in practice because it would generate excessive false alarms. Conversely, several models using

time-frequency features show very high specificity but very low sensitivity, meaning they are conservative and miss many microsleep-transition events. These two failure modes show why both sensitivity and specificity must be reported.

Tree-based ensembles provide more stable behaviour across several feature families. Gradient Boost and Random Forest generally avoid the most extreme sensitivity-specificity imbalance seen in Naïve Bayes, SVM, CNN, and LSTM under some feature sets. Nevertheless, even these models do not solve the task completely under a static-window formulation. Their performance supports the existence of discriminative EEG information, but also reveals the limitations of treating each window independently.

Table 4.3: Detection performance using time-domain features across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	53%	96%	10%
SVM	56%	82%	30%
KNN	55%	56%	55%
Random Forest	57%	58%	57%
Gradient Boost	59%	62%	55%
CNN	56%	82%	30%
LSTM	55%	30%	80%

Table 4.4: Detection performance using frequency-domain features across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	53%	98%	9%
SVM	56%	94%	19%
KNN	64%	63%	64%
Random Forest	66%	64%	68%
Gradient Boost	69%	67%	71%
CNN	55%	11%	98%
LSTM	58%	82%	34%

4.4 Statistical Reporting

The results reported in Tables 4.1–4.8 are point estimates. These values are useful for identifying broad performance trends across feature families and classifiers, but they should be interpreted as summaries of the evaluated experimental configuration

Table 4.5: Detection performance using bispectrum features across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	50%	100%	0%
SVM	50%	100%	0%
KNN	54%	54%	54%
Random Forest	59%	57%	60%
Gradient Boost	59%	60%	58%
CNN	56%	74%	37%
LSTM	58%	62%	53%

Table 4.6: Detection performance using nonlinear features across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	56%	94%	17%
SVM	63%	71%	55%
KNN	60%	60%	61%
Random Forest	60%	59%	62%
Gradient Boost	64%	60%	68%
CNN	63%	58%	69%
LSTM	64%	64%	64%

Table 4.7: Detection performance using time-frequency (wavelet) features across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	89%	33%	96%
SVM	90%	8%	100%
KNN	91%	46%	97%
Random Forest	96%	79%	98%
Gradient Boosting	96%	75%	98%
CNN	89%	0%	100%
LSTM	89%	0%	100%

Table 4.8: Detection performance using intensity-weighted features (IWMF, IWBW) across classifiers.

Model	Accuracy	Sensitivity	Specificity
Naïve Bayes	58.16%	74.99%	41.34%
SVM	64.30%	65.85%	62.75%
KNN	60.17%	59.45%	60.90%
Random Forest	60.01%	58.70%	61.33%
Gradient Boost	63.75%	64.33%	63.17%
CNN	64.07%	69.14%	59.00%
LSTM	64.37%	67.16%	61.57%

rather than as complete estimates of uncertainty. Microsleep detection performance can vary across subjects because individuals differ in baseline EEG rhythms, fatigue

expression, signal quality, and the clarity of transition states.

This complements the point estimates with non-parametric statistical testing across classifiers. The Friedman test followed by Nemenyi post-hoc analysis was applied across accuracy, sensitivity, and specificity [148]. The Friedman test indicated significant differences among classifiers for all three metrics: $p = 0.0011$ for accuracy, $p = 0.0117$ for sensitivity, and $p = 0.0017$ for specificity. These results show that classifier choice has a statistically significant effect on performance under the evaluated conditions.

The Nemenyi post-hoc analysis, shown in Figure 4.1, provides pairwise comparisons between classifiers. The results support the conclusion that differences between simple baselines and stronger classifier families are not only numerical artefacts. However, the statistical tests must be interpreted alongside the sensitivity–specificity trade-offs. A classifier may be statistically different because it performs better on specificity or accuracy, but this does not necessarily make it preferable for a safety-related microsleep detection system if sensitivity is poor.

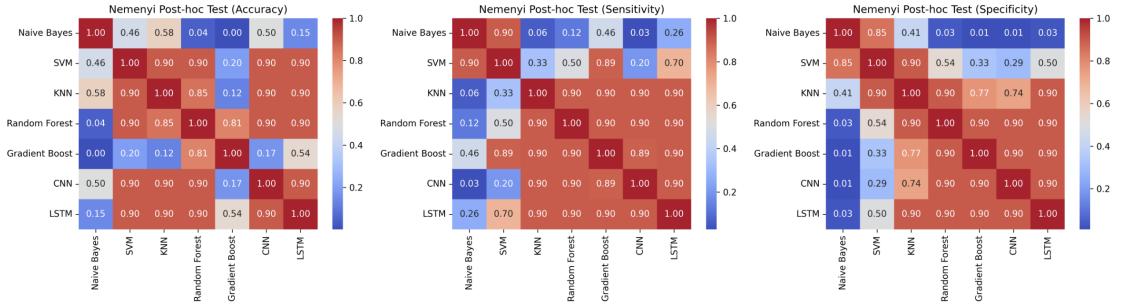


Figure 4.1: Nemenyi post-hoc test heatmaps showing pairwise comparisons for accuracy, sensitivity, and specificity across classifiers. Lower values indicate significant differences between classifier pairs.

4.5 Identification of Robust Predictive Indicators

Feature importance analysis was performed using Random Forest to identify which individual descriptors contributed most strongly to classification. Figure 4.2 shows the top-ranked features under the RQ1 configuration. The most prominent predictors are nonlinear dynamics descriptors, particularly Poincaré-derived features

such as SD1, SD2, and SD1/SD2. Entropy-related measures and intensity-weighted spectral descriptors, including IWMF and IWBW, also appear among the more informative features [148].

This feature-importance profile suggests that microsleep-related EEG patterns are not captured by one simple spectral marker alone. Instead, useful information appears to be distributed across complexity, variability, and spectral distribution measures. Time-domain statistics contribute moderately, while several simple band-power features appear less dominant in the Random Forest ranking.

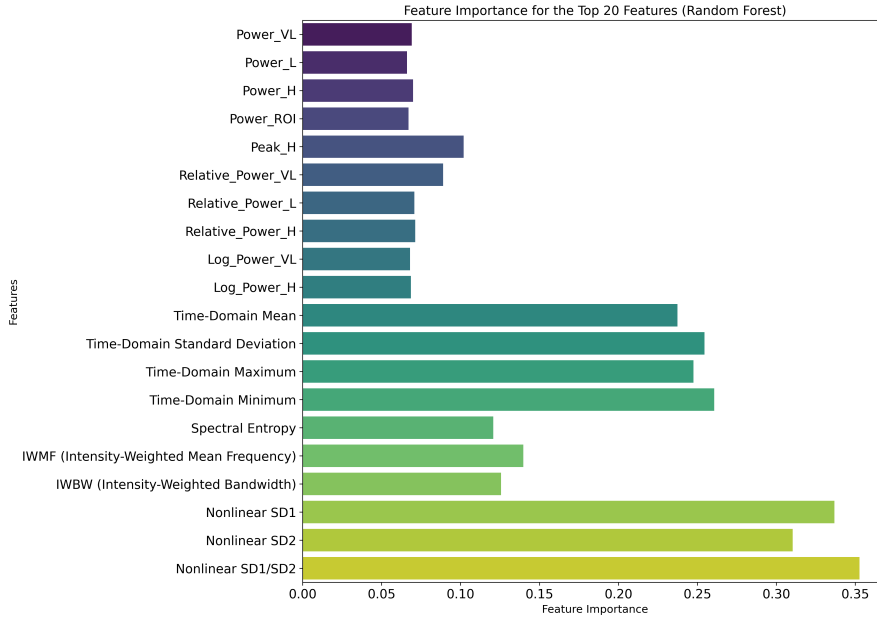


Figure 4.2: Top 20 features ranked by importance for the Random Forest classifier under the RQ1 configuration.

The confusion matrices in Figure 4.3 provide a complementary view of classifier behaviour using nonlinear features. They show that different classifiers fail in different ways. Some models favour the positive class and generate many false alarms, while others favour the negative class and miss microsleep-transition windows. This pattern supports the key conclusion of the chapter: static EEG features contain useful information, but they do not produce uniformly reliable microsleep detection when evaluated using sensitivity and specificity together.

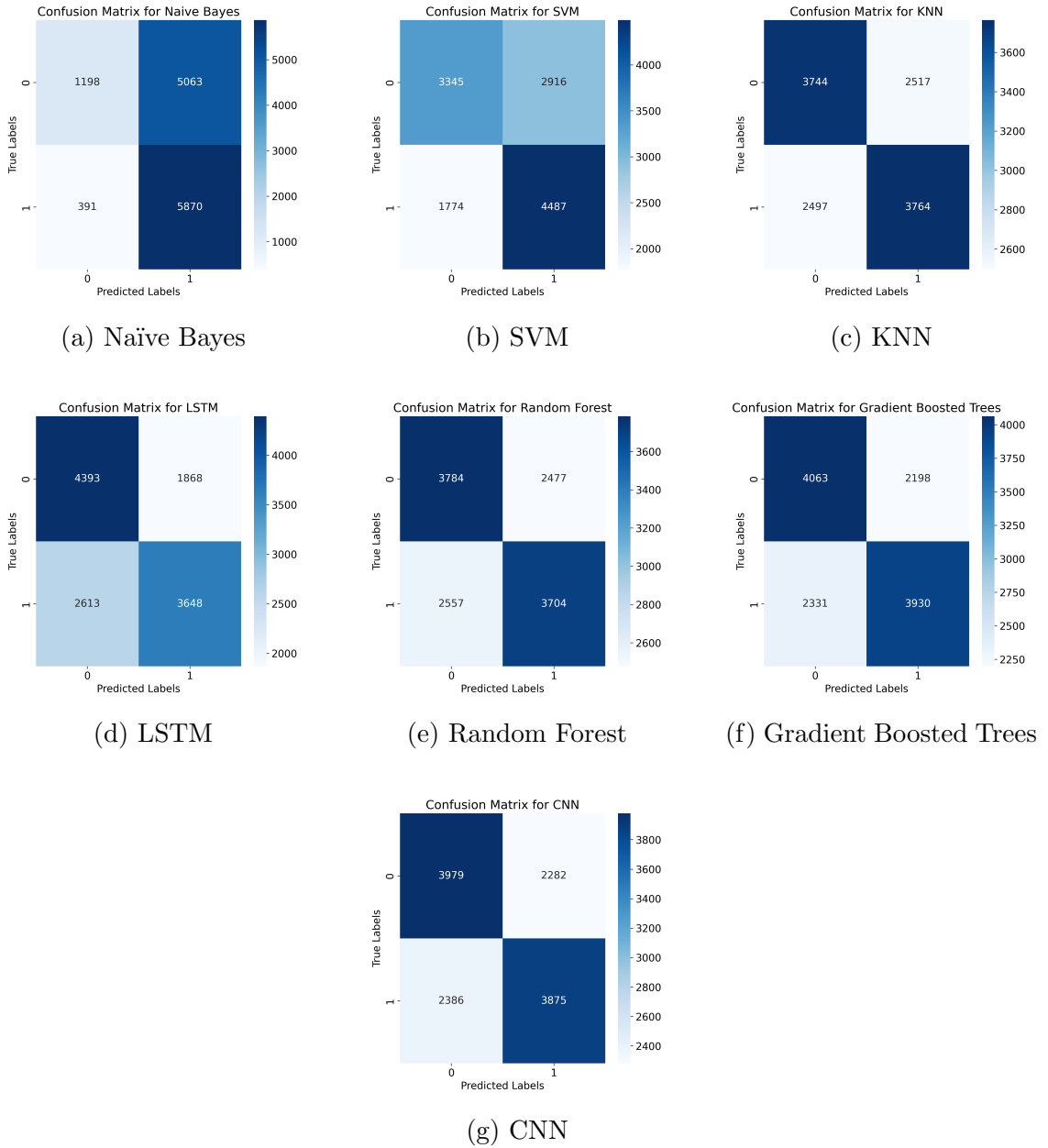


Figure 4.3: Confusion matrices for different classifiers using nonlinear features.

4.6 Interpretation of Static Feature Results

The main result of this chapter is not that one static EEG feature family solves microsleep detection. Rather, the results show that static feature approaches provide useful but incomplete evidence. Time-domain and bispectrum features are generally weak or unstable. Frequency-domain features perform better, but still show moderate sensitivity and specificity. Time-frequency features can produce very high accuracy and specificity, but this can come at the cost of sensitivity.

Nonlinear and intensity-weighted descriptors appear more balanced than simple time-domain or bispectrum features. The IWMF/IWBW results show moderate accuracy with comparatively stable sensitivity and specificity across several classifiers, particularly SVM, Gradient Boost, CNN, and LSTM. This supports the interpretation that spectral-distribution descriptors capture useful microsleep-related information, although their performance remains limited under a static-window formulation.

This pattern means that the EEG does contain microsleep-relevant information, but fixed-window descriptors are limited in how they represent the gradual onset and short duration of microsleep. Features computed independently for each segment cannot model temporal persistence, pre-event trends, or the sequence of physiological changes that occur as wakefulness deteriorates. A model may recognise stable wakefulness well but fail to detect early or ambiguous microsleep-transition windows.

This interpretation reframes the limitations of Chapter 4 as a strength of the thesis narrative. The static feature results establish the baseline and show that feature engineering alone is not enough. Chapter 5 therefore investigates whether temporal modelling can preserve information across neighbouring windows and improve early prediction.

4.7 Limitations of Static Feature Modelling

The first identified limitation is that all features in this chapter are computed independently for each window. This ignores temporal dependencies between adjacent segments. Microsleep episodes rarely appear as isolated, instantaneous events; they usually emerge through a gradual transition involving spectral slowing, reduced complexity, eye movement changes, and unstable wakefulness. Static windows cannot fully represent this evolution.

The second limitation is that the evaluation relies primarily on accuracy, sensitivity, and specificity. These metrics are informative, but they do not fully describe performance under imbalance. In later chapters, AUC-PR, F1-score, and horizon-

specific forecasting metrics provide a richer assessment of positive-class behaviour. For RQ1, the available metrics still reveal the key issue: high accuracy can coexist with poor microsleep sensitivity.

4.8 RQ1 Findings on EEG Indicators of Microsleep

The experiments in this chapter show that EEG-derived features contain useful information for identifying microsleep-transition states, but the usefulness of a feature family depends strongly on how performance is measured. If accuracy alone is considered, time-frequency features appear strongest. However, when sensitivity is examined, this conclusion becomes less clear because high accuracy is sometimes achieved by under-detecting microsleep events.

The more defensible finding is that nonlinear dynamics features and intensity-weighted spectral descriptors provide physiologically meaningful and relatively balanced evidence for microsleep detection. Nonlinear descriptors, including Poincaré-derived measures and entropy features, are among the most important predictors in the Random Forest analysis. Intensity-weighted spectral features also appear useful because they summarise spectral slowing and dispersion without relying entirely on fixed frequency bands. The IWMF/IWBW table shows that these features provide moderate but relatively balanced performance across several classifiers, supporting their role as complementary EEG indicators rather than as a complete standalone solution.

The chapter therefore answers RQ1 by identifying that EEG indicators of microsleep are not captured by a single universally optimal feature family. Instead, the most useful static indicators appear to involve a combination of complexity, variability, and spectral-distribution descriptors. Static features provide a necessary baseline, but their limitations motivate the temporal modelling work in Chapter 5.

Chapter Summary

This chapter established an EEG-only feature-based baseline for microsleep detection using the MWT dataset [22], [158]. Six feature families were evaluated across multiple classifiers under a controlled subject-exclusive protocol. The results show that static EEG features contain discriminative information, but also reveal important limitations.

The most important methodological lesson is that accuracy alone is not a reliable basis for interpreting microsleep detection performance. Time-frequency features achieve high accuracy and specificity, but may under-detect microsleep-transition windows. Nonlinear dynamics features and intensity-weighted spectral descriptors provide a more balanced and physiologically plausible interpretation. Tree-based classifiers are generally more stable than simple baselines, but no classifier-feature combination fully resolves the difficulty of static microsleep detection.

Overall, the chapter provides a necessary EEG-only baseline and clarifies why static feature modelling is insufficient on its own. These findings motivate Chapter 5, which moves from independent window-level descriptors to temporal modelling for early microsleep prediction.

Chapter 5

Temporal Modelling for Microsleep Forecasting

5.1 Introduction

Chapter 4 examined which EEG-derived features are most informative for microsleep detection. That analysis established that spectral and nonlinear descriptors contain discriminative information for separating microsleep from non-microsleep states, and it provided an evidence-based feature foundation for subsequent modelling. Building on that result, this chapter addresses the second research question by examining how temporal modelling can improve the prediction of microsleep events before their onset.

The practical motivation for forecasting, rather than only detecting, microsleep episodes is clear. Microsleeps are brief involuntary lapses in responsiveness that can occur during sustained attention tasks and may precede operational failures in settings such as driving, industrial monitoring, and clinical observation [1], [9], [171]. In such contexts, a prediction horizon of even one or two seconds may provide additional time for warning generation, system adaptation, or human intervention. However, forecasting is more difficult than onset detection because neural changes become less explicit as the time horizon moves further from the event [13]. In

practical terms, this means that models must rely on more subtle and distributed temporal patterns rather than clear event-related changes. As a result, both the representation of temporal dependencies and the preservation of fine-grained signal variations become central challenges in microsleep prediction.

Prior work has shown that EEG-based systems can achieve strong performance when detecting microsleeps at, or very close to, the onset of the event [24], [25]. By contrast, performance tends to degrade when models are required to forecast events ahead of time [13], [172]. A key reason is that conventional pipelines often aggregate EEG information in a way that reduces temporal precision, making it harder to preserve transient changes that may signal an approaching microsleep. This creates a methodological gap between effective event detection and reliable early prediction.

This chapter investigates whether temporal modelling with causal attention can mitigate that degradation. Specifically, it studies whether a multi-head causal self-attention mechanism, placed before a recurrent backbone, can preserve temporally local patterns that are relevant for forecasting. The chapter also evaluates the role of focal loss for handling class imbalance, since microsleep prediction remains a safety-critical and class-skewed problem.

A key hypothesis underlying this investigation is that the degradation observed at longer forecasting horizons is partly due to the loss of temporally localised predictive cues. Conventional recurrent models aggregate information sequentially, which can smooth or attenuate short-lived transitions that precede microsleep onset. A causal self-attention mechanism provides a way to selectively emphasise relevant time steps within the input sequence before this aggregation occurs. This is particularly important in forecasting settings, where weak pre-event signals must be preserved rather than averaged out. The proposed architecture is therefore designed to explicitly address this temporal resolution problem.

The work in this chapter therefore moves the thesis from *what to measure* toward *how to model change over time*. This is a necessary step toward the subsequent

multimodal work in Chapter 6, where temporal EEG modelling is extended through the integration of complementary modalities. Across these chapters, the overall objective remains the same: to improve prediction performance in a way that is methodologically justified and practically meaningful.

5.2 Research Question and Chapter Scope

This chapter addresses the following research question:

RQ2: How can temporal modelling improve microsleep prediction from EEG, particularly at extended forecast horizons, while maintaining sensitivity and limiting false alarms?

To address this question, the chapter has four aims:

1. to reproduce a published EEG microsleep forecasting baseline based on CSP and sparse Bayesian learning;
2. to establish a recurrent neural baseline using a GRU architecture;
3. to propose an attention-augmented temporal model for multi-horizon forecasting; and
4. to analyse the contribution of attention and focal loss through comparative and ablation experiments.

5.3 Related Work

Microsleep detection and prediction have been studied using a range of physiological and behavioural signals. EEG remains one of the most direct modalities for observing vigilance-related neural dynamics, and it has therefore been used extensively in the microsleep literature [9], [24]. Early work demonstrated that single-channel EEG systems can detect microsleep-related changes with useful accuracy, but such systems may not capture the broader spatial structure available in multichannel

recordings [25]. Multichannel EEG, although more demanding in terms of acquisition and processing, provides richer information about cortical activity and has therefore been used in several forecasting-oriented studies [13], [44].

Beyond EEG, alternative modalities have also been explored. Speech-based approaches have been used to identify sleepiness-related changes in vocal expression [173], while video-based systems have monitored eyelid closure and facial behaviour [46]. These studies indicate that fatigue and microsleep can be approached from multiple sensing perspectives, but direct comparisons with EEG-based forecasting remain limited. For the purposes of this chapter, EEG is used as the primary signal because it provides direct access to pre-event neural changes and because it connects naturally with the feature findings established in Chapter 4.

5.3.1 Feature Extraction for Microsleep Analysis

Spectral and information-theoretic EEG features have been repeatedly associated with changes in vigilance and microsleep onset. Pairwise joint entropy and mutual information between EEG channels achieved strong discrimination in prior work, illustrating that inter-channel relationships can be informative [44]. Common Spatial Patterns (CSP) have also been used to derive spatio-temporal components for forecasting, outperforming simpler log-power representations in some settings [13], [174]. Other studies have investigated blind source extraction, feature reduction, and dimensionality reduction pipelines to improve separability or reduce redundancy [65], [70].

The investigation conducted in RQ1 further showed that spectral and nonlinear features carry substantial predictive value for microsleep-related classification [148], [175]. These findings support their use as inputs to temporal forecasting models. However, informative features alone do not solve the forecasting problem if their temporal organisation is not preserved.

5.3.2 Forecasting Models and the Role of Temporal Dependence

A major limitation in the literature is the emphasis on event detection at the current time point rather than forecasting ahead of time. Deep learning methods, including convolutional networks, have shown strong performance for microsleep detection at onset [24]. Recurrent models such as LSTM-based and transformer-based approaches have also been explored for EEG sequence modelling, but forecasting horizons beyond immediate onset remain less extensively studied [139], [172], [176].

One of the few explicit multi-horizon microsleep forecasting studies is that of Shoorangiz et al. [13], who reported declining sensitivity as the forecast horizon increased from 0.25 s to 1.0 s. This result is important because it suggests that the main difficulty is not only class imbalance or model capacity, but also the preservation of temporally local information as the event becomes more distant. Recurrent models can capture sequence structure, but they may still smooth over short-lived transitions. Attention mechanisms offer a possible remedy by reweighting sequence elements according to their relevance. In EEG analysis, attention has often been used to model spatial channels or feature importance, but its use for causal microsleep forecasting remains limited.

This chapter therefore evaluates a causal multi-head self-attention mechanism placed before a GRU. The aim is not to replace recurrence entirely, but to enhance temporal feature resolution before recurrent integration.

5.4 Methodology

5.4.1 Datasets

Two datasets were used in this chapter to evaluate both forecasting performance and robustness across different recording conditions.

Maintenance of Wakefulness Test Dataset

The primary experiments in this chapter were conducted using the Maintenance of Wakefulness Test (MWT) dataset, which was introduced in Chapter 3. The MWT dataset contains simultaneous EEG and EOG recordings collected under controlled laboratory conditions and serves as the principal benchmark dataset throughout this thesis [22], [146]. The dataset description, acquisition protocol, channel configuration, and preprocessing procedures are therefore not repeated here.

For the present study, microsleep was treated as the positive class. Awake segments were treated as the negative class. After segmentation into 1-second windows, the resulting dataset comprised 60,548 wakefulness samples and 37,456 microsleep samples.

Sleepy Driver EEG Brainwave Dataset

To examine generalisation beyond the MWT recordings, additional experiments were carried out on the Sleepy Driver EEG Brainwave dataset [147]. This dataset was collected from four subjects using a NeuroSky MindWave headset, which provides a single-channel consumer-grade EEG signal. Each subject contributed short awake and drowsy recordings of approximately 5–8 minutes. The recordings were provided as one-second interval spectral measurements from a frontal electrode.

Although this dataset differs substantially from the MWT data in both scale and sensor configuration, it provides a useful robustness test. As in the main experiments, the task was formulated as a binary problem, with the sleepy condition treated as the positive state.

5.4.2 Problem Formulation

Let x_t denote the feature representation extracted from a short EEG segment at time step t . Given a sequence

$$\mathbf{X} = [x_1, x_2, \dots, x_T] \in \mathbb{R}^{T \times d}, \quad (5.1)$$

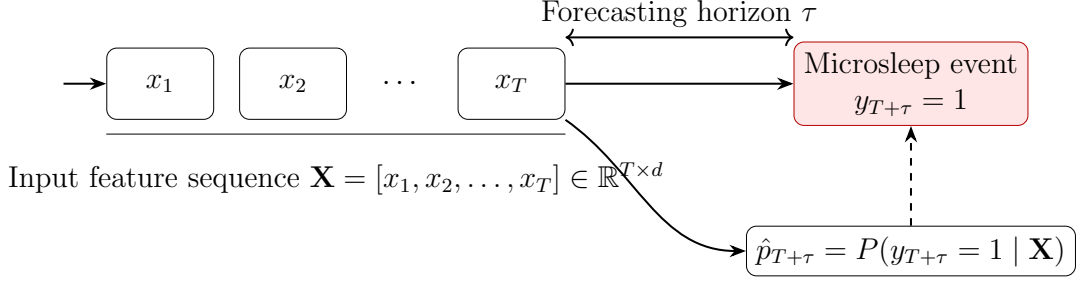


Figure 5.1: Illustration of the microsleep forecasting problem. A sequence of EEG feature vectors $\mathbf{X} = [x_1, \dots, x_T]$ is used to estimate the probability of a microsleep event occurring after forecasting horizon τ . When $\tau = 0$, the task corresponds to onset detection; positive values of τ represent advance forecasting.

the objective is to estimate the probability that a microsleep event will occur after a forecasting horizon τ . The experiments consider multiple horizons. For the replicated CSP+SBL baseline, horizons of 0.25 sec, 0.50 sec, 0.75 sec, and 1.00 sec are used following [13]. For the recurrent and attention-based models, horizons of 0 sec, 1 sec, and 2 sec are used to examine both onset detection and further-ahead forecasting.

5.4.3 Baseline Replication: CSP + SBL

The first baseline reproduces the forecasting framework reported by Shoorangiz et al. [13]. Multichannel EEG was band-pass filtered between 0.5 and 30 Hz and segmented into 1-second epochs. Labels were shifted according to the forecasting horizon. Common Spatial Patterns were then used to derive spatial filters, producing C components for each segment. The log-variance of the components formed the feature vector

$$\mathbf{f} \in \mathbb{R}^C,$$

which was classified using sparse Bayesian logistic regression. Hyperparameters were selected using cross-validation.

This baseline serves two roles in the chapter. First, it provides continuity with the published microsleep forecasting literature. Second, it offers a non-deep-learning comparator that is designed explicitly around short-horizon forecasting.

5.4.4 GRU Baseline

The second baseline is a recurrent neural model based on a two-layer Gated Recurrent Unit (GRU). Each EEG epoch is transformed into a sequence of features over time, producing an input tensor $\mathbf{X} \in \mathbb{R}^{T \times d}$. The GRU processes the sequence and outputs hidden states

$$\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_T.$$

The final hidden state $\mathbf{h}_T$ is then passed through a fully connected layer with sigmoid activation to estimate the microsleep probability.

The use of a GRU is motivated by the need to model sequential dependence with lower computational complexity than a comparable LSTM. This is consistent with the aim of developing a forecasting model that remains suitable for real-time deployment. In addition, GRUs have been shown to retain sufficient capacity for modelling temporal dependencies in physiological signals while reducing the risk of overfitting compared with more complex recurrent architectures. This makes them a suitable baseline for evaluating whether additional mechanisms, such as attention, provide measurable gains in forecasting performance.

5.4.5 Proposed Attention-Augmented GRU

The proposed model extends the GRU baseline by introducing a causal multi-head self-attention layer before the recurrent block. The goal is to reweight temporally local information before sequence integration, thereby reducing the loss of short-lived predictive cues at longer horizons.

Given the input sequence $\mathbf{X}$, attention is computed as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax} \left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{M} \right) \mathbf{V}, \quad (5.2)$$

where

$$\mathbf{Q} = \mathbf{X}\mathbf{W}_Q, \quad \mathbf{K} = \mathbf{X}\mathbf{W}_K, \quad \mathbf{V} = \mathbf{X}\mathbf{W}_V.$$

Here, d_k denotes the per-head dimensionality and $\mathbf{M}$ is a causal mask that prevents a time step from attending to future positions. The outputs from multiple heads are concatenated and linearly projected to form an attention-refined sequence $\tilde{\mathbf{X}}$, which is then passed to the GRU.

The causal mask is important because the task is forecasting. At prediction time, only current and past information should be available. The masking therefore ensures that the model does not exploit future information during training. By applying attention prior to recurrent processing, the model is encouraged to preserve and reweight temporally local information that may be predictive of an upcoming microsleep event. This design directly addresses the forecasting setting, where relevant cues are often weak and distributed across time rather than concentrated at the prediction point.

5.4.6 Feature Representation

For both the GRU baseline and the proposed model, each 1-second EEG epoch is converted into a feature sequence that combines nonlinear and frequency-domain descriptors. The extracted features include:

- five nonlinear complexity measures, including entropy-based descriptors; and
- fifteen frequency-domain features based on absolute and relative band power in the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–12 Hz), beta (12–26 Hz), and gamma (26–50 Hz) bands.

These feature groups were selected because they are supported by prior microsleep studies [44], [171] and by the feature evaluation carried out earlier in this thesis [148], [175]. Their combination allows the temporal model to operate on physiologically interpretable inputs rather than on raw signals alone.

5.4.7 Additional Comparator: EEG-TCNet Adaptation

An adapted version of EEG-TCNet [26] was also evaluated for forecasting. EEG-TCNet was originally proposed for EEG classification, but here it was modified to operate on temporally windowed EEG feature representations. The adapted architecture used convolutional layers to capture local feature interactions followed by a GRU stage for temporal modelling. The same forecasting protocol used for the other deep models was applied. This comparator was included to examine whether the observed gains were specific to the proposed architecture or reflected a broader benefit of temporal deep learning.

5.4.8 Loss Function

To reduce the effect of class imbalance, the proposed model uses focal loss [118]:

$$\mathcal{L}_{\text{FL}} = -\alpha_t(1 - p_t)^\gamma \log(p_t), \quad (5.3)$$

where p_t denotes the predicted probability assigned to the true class, $\alpha_t \in [0, 1]$ is a weighting factor used to balance the contribution of different classes, and $\gamma \geq 0$ is the focusing parameter that reduces the loss contribution from well-classified examples. When $\gamma = 0$, focal loss reduces to the standard cross-entropy loss. Consequently, focal loss places greater emphasis on difficult minority-class instances while reducing the influence of easy examples. In the present setting, this is intended to improve sensitivity to microsleep events while helping to control false alarms.

This is particularly relevant for microsleep forecasting, where missed positive events (false negatives) may have greater practical consequences than false alarms. By focusing learning on harder examples, focal loss supports improved sensitivity without requiring explicit resampling strategies.

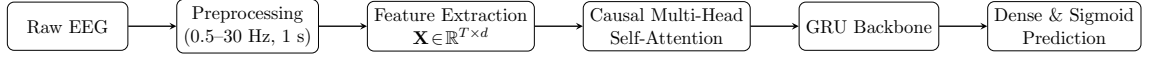


Figure 5.2: Full pipeline: causal multi-head self-attention is applied before the GRU to enrich temporal feature representations.

5.4.9 Evaluation Metrics

The models are evaluated using accuracy, sensitivity, specificity, geometric mean (GM), AUC–ROC, and AUC–PR. Sensitivity and specificity are especially important because the application is safety-related and because performance must be examined separately for positive and negative states. The geometric mean is included to summarise balanced performance across both classes, while AUC–PR is used in addition to AUC–ROC because of the class imbalance.

Figure ?? summarises the overall methodology used in this chapter, including the datasets, preprocessing stages, baseline models, the proposed attention-enhanced GRU framework, and the evaluation protocol across forecasting horizons.

5.4.10 Implementation Details

Experiments were implemented in Python using PyTorch for the deep models and scikit-learn for the CSP+SBL baseline. EEG preprocessing used a 0.5–30 Hz band-pass filter. Windows of one second were generated with 50% overlap for the deep models, and feature extraction was implemented using standard numerical libraries.

To reduce the risk of data leakage, subject-exclusive splits were applied, with 80% of the data used for training and 20% for testing while preserving class ratios. Deep models were trained with batch size 32 using the Adam optimiser with an initial learning rate of 10^{-3} . The GRU architecture used two recurrent layers with 64 and 128 units, dropout of 0.2, and early stopping based on validation AUC–ROC with patience of 10 epochs. The proposed model added a four-head causal attention layer with head dimension 64.

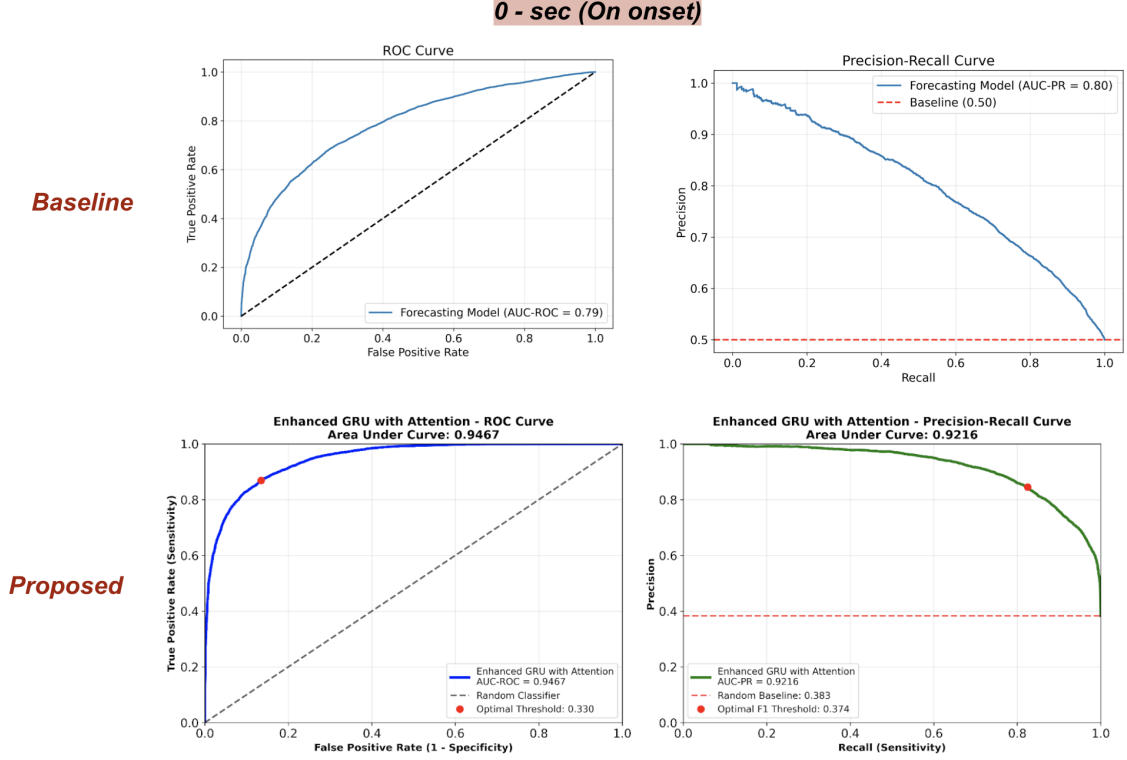


Figure 5.3: ROC and precision–recall curves for the baseline GRU (top row) and the proposed attention-augmented GRU (bottom row) at $\tau = 0$ s.

5.5 Results

This section evaluates whether the proposed temporal modelling approach improves microsleep forecasting performance across multiple horizons. The results are structured to first compare the proposed model with recurrent and classical baselines, followed by cross-dataset evaluation and ablation analysis to isolate the contributions of individual components.

5.5.1 GRU Baseline versus Attention-Augmented GRU

Table 5.1 compares the recurrent baseline with the proposed attention-augmented model across onset detection and two forecasting horizons. The proposed model improves all reported metrics at all horizons.

At $\tau = 0$ s, test accuracy increases from 0.71 to 0.85, sensitivity from 0.64 to 0.86, and AUC–ROC from 0.79 to 0.94. At $\tau = 1$ s, the performance gap becomes more pronounced: sensitivity increases from 0.59 to 0.84 and AUC–ROC from 0.76

Table 5.1: Performance of the GRU baseline and the proposed attention-augmented model at forecast horizons of 0 s, 1 s, and 2 s.

Metric	0 s (onset)		1 s		2 s	
	GRU	Proposed	GRU	Proposed	GRU	Proposed
Train Accuracy	0.72	0.87	0.70	0.84	0.68	0.82
Test Accuracy	0.71	0.85	0.69	0.83	0.67	0.81
Sensitivity	0.64	0.86	0.59	0.84	0.53	0.84
Specificity	0.79	0.97	0.79	0.95	0.81	0.93
Geometric Mean	0.71	0.87	0.68	0.85	0.65	0.84
AUC-ROC	0.79	0.94	0.76	0.93	0.73	0.92
AUC-PR	0.80	0.92	0.76	0.89	0.74	0.87

to 0.93. At $\tau = 2$ s, the proposed model maintains sensitivity at 0.84 compared with 0.53 for the GRU baseline, while also improving specificity and discrimination.

These results suggest that the causal attention layer helps preserve predictive cues that are otherwise attenuated as the forecasting horizon increases. Figure 5.3 illustrates this effect at onset through improved ROC and precision-recall behaviour.

5.5.2 Comparison with the CSP+SBL Forecasting Baseline

Table 5.2 compares the replicated CSP+SBL pipeline with the proposed model at shorter forecast horizons. The CSP+SBL baseline shows decreasing sensitivity as the horizon increases from 0.25 s to 1.00 s, falling from 0.51 to 0.45. By contrast, the proposed model maintains sensitivity between 0.60 and 0.65 across the same range.

The proposed model also improves AUC-PR at each horizon, indicating stronger positive-class retrieval under imbalance. The gains in AUC-ROC are more modest, but they remain consistent. This pattern is important: the main practical difficulty in microsleep forecasting is often not overall discrimination alone, but maintaining positive-class sensitivity as the warning horizon increases.

5.5.3 Evaluation on the Sleepy Driver Dataset

To examine whether similar modelling behaviour is observed beyond the MWT recordings, the GRU baseline and the proposed model were also evaluated on the

Table 5.2: Comparison of the CSP+SBL baseline and the proposed attention-augmented model across short forecast horizons.

Metric	0.25 s		0.50 s		0.75 s		1.00 s	
	CSP+SBL	Proposed	CSP+SBL	Proposed	CSP+SBL	Proposed	CSP+SBL	Proposed
Sensitivity	0.51	0.64	0.49	0.61	0.47	0.60	0.45	0.65
Specificity	0.72	0.77	0.70	0.75	0.70	0.73	0.69	0.71
Geometric Mean	0.66	0.71	0.65	0.69	0.64	0.70	0.61	0.68
AUC-ROC	0.77	0.80	0.76	0.77	0.75	0.77	0.75	0.77
AUC-PR	0.70	0.79	0.69	0.75	0.67	0.73	0.66	0.70

Table 5.3: Comparison of the GRU baseline and the proposed attention-augmented model on the Sleepy Driver EEG Brainwave dataset across forecast horizons.

Metric	0 s		1 s		2 s	
	GRU	Proposed	GRU	Proposed	GRU	Proposed
Sensitivity	0.77	0.79	0.76	0.78	0.78	0.79
Specificity	0.42	0.82	0.41	0.81	0.40	0.84
Geometric Mean	0.46	0.81	0.45	0.77	0.45	0.80
AUC-ROC	0.78	0.90	0.79	0.86	0.79	0.89
AUC-PR	0.66	0.86	0.65	0.82	0.64	0.85

Sleepy Driver EEG Brainwave dataset. The results are shown in Table 5.3.

The proposed model improves specificity, geometric mean, AUC-ROC, and AUC-PR at all reported horizons, while sensitivity changes only slightly. At $\tau = 0$ s, specificity improves from 0.42 to 0.82 and AUC-ROC from 0.78 to 0.90. Similar trends remain at $\tau = 1$ s and $\tau = 2$ s. These findings provide supportive evidence that the attention-based model may retain useful behaviour in a lower-cost EEG setting. However, given the limited size and single-channel nature of the dataset, they should be interpreted as indicative rather than conclusive evidence of generalisation.

The more modest sensitivity changes suggest that when only a single channel is available, attention has less opportunity to exploit richer temporal-spatial structure. The model therefore appears to retain value under constrained acquisition conditions, but its strongest benefit is observed in the multichannel clinical setting.

5.5.4 Illustrative EEG and Feature Trends

To complement the predictive results, Figures 5.4 and 5.5 illustrate representative temporal patterns observed around microsleep events in the MWT data. These figures are not used as evaluation measures, but they help contextualise why temporal

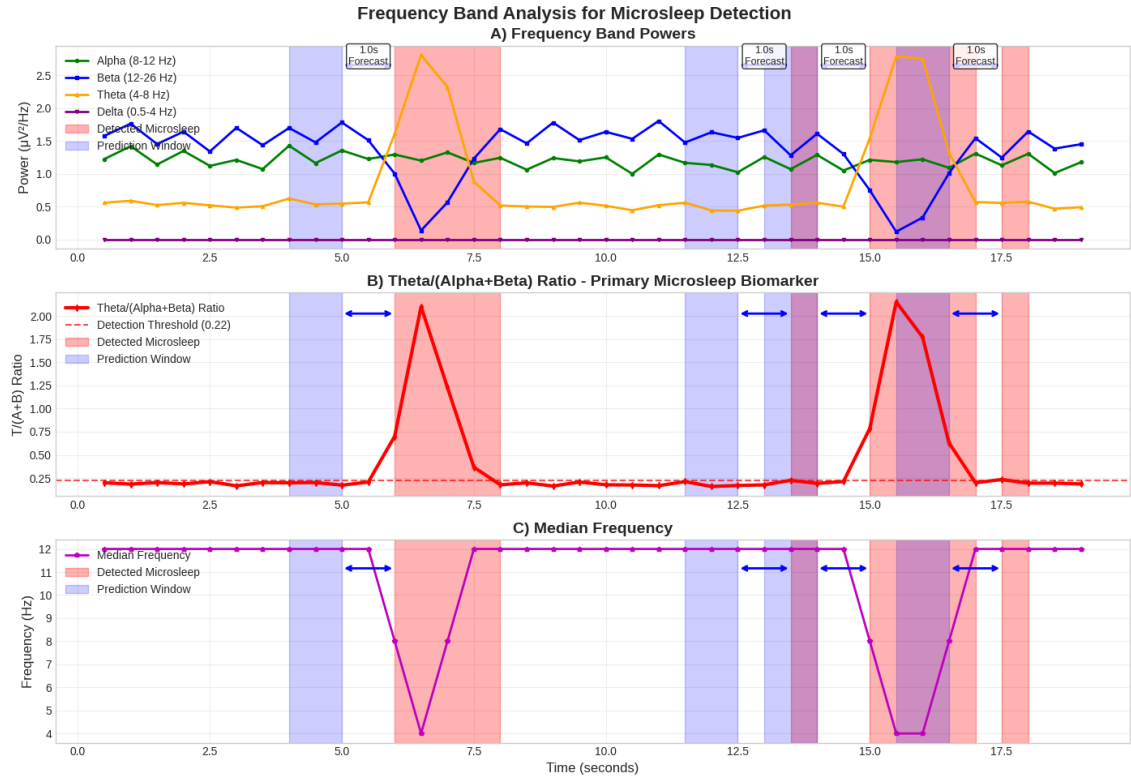


Figure 5.4: Illustrative frequency-domain changes surrounding microsleep episodes for an example subject. Panel A shows the alpha (8–12 Hz), beta (12–26 Hz), theta (4–8 Hz), and delta (0.5–4 Hz) band powers. Panel B shows the $\theta/(\alpha + \beta)$ ratio, with the red dashed line indicating the detection threshold of 0.22. Panel C shows the median frequency. Red shaded regions indicate microsleep intervals, while blue shaded regions indicate the one-second prediction windows preceding them. Purple regions represent temporal overlap between prediction windows and microsleep intervals. The blue arrows and “1.0 s Forecast” annotations indicate the forecast interval. Microsleep periods are characterised by an increased $\theta/(\alpha + \beta)$ ratio and a reduction in median frequency.

modelling is relevant.

Figure 5.4 shows the evolution of frequency-band powers and the theta/(alpha+beta) ratio for an example subject. During microsleep segments, theta activity increases while alpha and beta activity are reduced, causing the ratio to rise. Figure 5.5 shows the corresponding raw EEG trace, where the signal exhibits a greater dominance of slow oscillations and reduced rapid waveform variation during microsleep episodes. These observations are consistent with the feature analysis presented in Chapter 4 and support the use of temporally sensitive modelling.

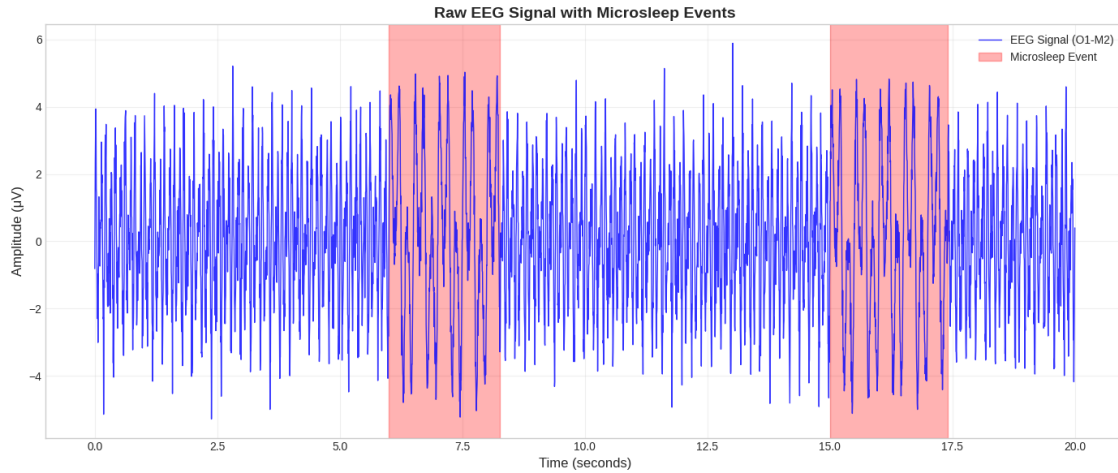


Figure 5.5: Raw EEG signal (O1–M2) from an example subject with microsleep episodes shaded in red.

5.5.5 Ablation Study

An ablation study was conducted to separate the effects of the attention mechanism and focal loss. Four configurations were compared: GRU with cross-entropy loss, GRU with focal loss, GRU with attention and cross-entropy, and GRU with attention and focal loss. The results are summarised in Table 5.4.

Relative to the plain GRU with cross-entropy, focal loss improves sensitivity from 0.54 to 0.64 and AUC–ROC from 0.70 to 0.79. Adding attention produces a larger change, increasing sensitivity to 0.83 and specificity to 0.95 even without focal loss. The full model, combining attention and focal loss, yields the strongest overall performance, including test accuracy of 0.85, sensitivity of 0.86, specificity of 0.97, and AUC–ROC of 0.94.

The ablation therefore indicates that focal loss and attention contribute in different but complementary ways. Focal loss primarily improves learning under imbalance, whereas attention improves temporal discrimination more broadly.

5.5.6 Attention Visualisation

To inspect which parts of the feature sequence are emphasised by the model, the output of the causal multi-head attention layer was visualised as a feature-by-time heatmap. Figure 5.6 shows that several features, including theta/alpha ratio,

Table 5.4: Ablation results showing the effect of attention and focal loss.

Metric	GRU-CE	GRU-Focal	GRU+Attn-CE	GRU+Attn-Focal
Train Accuracy	0.66	0.72	0.83	0.87
Test Accuracy	0.64	0.71	0.81	0.85
Sensitivity	0.54	0.64	0.83	0.86
Specificity	0.76	0.79	0.95	0.97
Geometric Mean	0.63	0.71	0.88	0.87
AUC-ROC	0.70	0.79	0.89	0.94
AUC-PR	0.71	0.80	0.91	0.92

delta/theta ratio, and spectral entropy, receive stronger attention weights at specific time steps.

This does not constitute a full explanation of the model, but it does suggest that the attention layer is not acting uniformly across the sequence. Instead, it appears to assign greater importance to physiologically plausible features during segments that are likely to precede a microsleep event. This is useful because it connects the temporal modelling result back to the feature-oriented findings of Chapter 4.

5.5.7 Latency Analysis

Inference latency was measured to assess whether the proposed model is consistent with real-time operation. On an NVIDIA RTX 3090 GPU, the model required a mean inference time of 17.7ms for a single 1-second EEG input window, with a minimum of 16.98ms and a maximum of 26.37ms over 100 runs. These timings are well below the input window duration, indicating that the method is computationally feasible for real-time forecasting under the tested conditions.

5.5.8 Comparison with EEG-TCNet

Table 5.5 compares the proposed attention-GRU with the adapted EEG-TCNet model. Both architectures perform well, but the proposed model achieves higher sensitivity, geometric mean, AUC-ROC, and AUC-PR at all three horizons. The difference is most relevant at the longer horizons of 1s and 2s, where maintaining sensitivity is central to the forecasting problem.

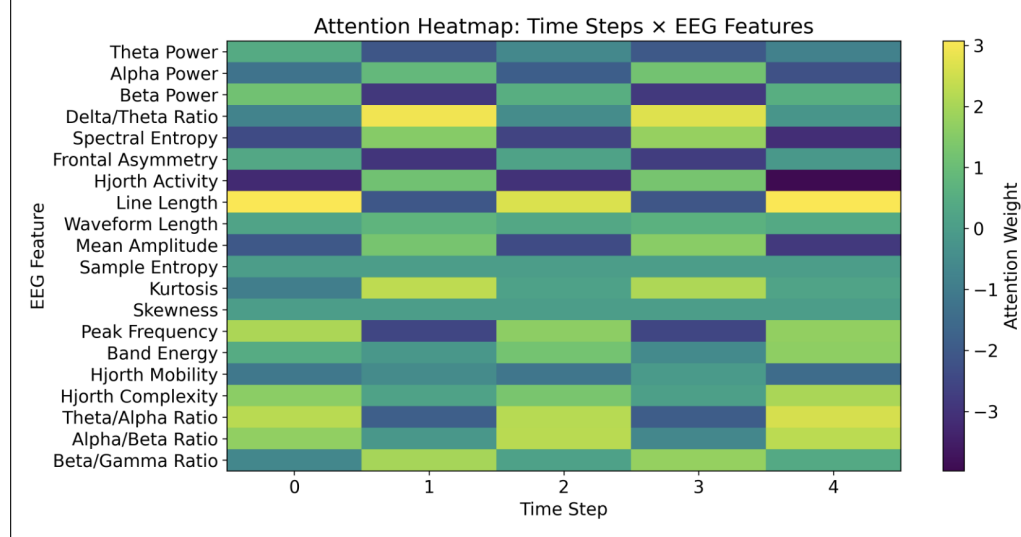


Figure 5.6: Attention-weight heatmap across EEG features and five temporal input steps. Brighter colours indicate relatively greater model emphasis, whereas darker colours indicate lower relative emphasis. Several features exhibit alternating temporal patterns: spectral entropy, the delta/theta ratio, mean amplitude, kurtosis, and Hjorth activity receive greater emphasis at time steps 1 and 3, while line length, peak frequency, the theta/alpha ratio, and the alpha/beta ratio are emphasised mainly at time steps 0 and 2. Other features, including sample entropy, skewness, and frontal asymmetry, remain comparatively stable across the sequence.

This comparison suggests that the gains are not attributable only to the use of deep learning, but specifically to the combination of causal attention and recurrent temporal integration used in the proposed architecture.

5.6 Discussion

5.6.1 Temporal Degradation and the Value of Causal Attention

The central result of this chapter is that forecasting performance can be improved by explicitly modelling temporal relevance before recurrent integration. Across the MWT experiments, the proposed model consistently outperformed both the GRU baseline and the replicated CSP+SBL pipeline. This was especially clear for sensitivity at extended horizons. The deterioration reported in earlier work [13] was reduced substantially in the proposed framework, suggesting that causal attention

Table 5.5: Comparison of forecasting performance between the proposed Attention-GRU and the adapted EEG-TCNet model.

Metric	Attention-GRU			EEG-TCNet		
	0 s	1 s	2 s	0 s	1 s	2 s
Train Accuracy	0.92	0.91	0.89	0.91	0.88	0.86
Test Accuracy	0.89	0.87	0.85	0.87	0.85	0.82
Sensitivity	0.87	0.84	0.81	0.83	0.78	0.75
Specificity	0.90	0.88	0.87	0.88	0.85	0.83
Geometric Mean	0.88	0.86	0.84	0.85	0.81	0.79
AUC-ROC	0.92	0.90	0.89	0.91	0.89	0.87
AUC-PR	0.90	0.87	0.85	0.88	0.85	0.82

helps preserve information that is otherwise weakened as the time-to-event increases.

Importantly, this suggests that the challenge in microsleep forecasting is not solely one of model capacity, but of how temporal information is represented and prioritised. The results therefore support the view that architectural choices which preserve temporal resolution are critical when moving from detection to prediction tasks.

This result is important in the context of the thesis as a whole. Chapter 4 showed that informative EEG features are available, but it did not address how those features should be organised over time. The present chapter shows that predictive performance depends not only on the quality of the features, but also on the temporal mechanism used to integrate them.

5.6.2 Interpretation of the Ablation Results

The ablation study indicates that attention and focal loss serve different functions. Focal loss improved minority-class detection, which is expected in a class-imbalanced setting. Attention, however, produced the larger overall change in performance, especially for sensitivity and specificity together. This suggests that the main challenge in RQ2 is not only class imbalance, but also temporal representation.

The full model therefore combines two design choices: one aimed at learning under skewed labels, and one aimed at preserving temporally local predictive structure. The results support retaining both elements in the final architecture.

5.6.3 Cross-Dataset Behaviour

The evaluation on the Sleepy Driver dataset provides a useful contrast with the MWT findings. Although the proposed model still improves most metrics, the behaviour differs from that seen on multichannel clinical EEG. The improvements on the Sleepy Driver data are driven primarily by gains in specificity and ranking-based metrics, while sensitivity changes only marginally. This is consistent with the smaller size and lower signal richness of the dataset.

The implication is not that attention is ineffective in low-cost EEG, but that its benefits depend on the quantity and structure of the available signal. In richer multichannel recordings, attention can exploit more informative temporal variations. In consumer-grade single-channel data, the forecasting problem remains more constrained. This boundary condition is important for interpreting the contribution realistically.

5.6.4 Practical Implications

From an application perspective, the ability to maintain useful sensitivity at 1–2 second horizons is relevant for warning systems, operator monitoring, and decision support. A warning that arrives before onset is more actionable than a decision made at the moment of behavioural lapse. At the same time, the specificity results matter because false alarms can reduce trust and usability.

The latency measurements further indicate that the proposed model can operate within real-time constraints under the tested hardware setting. This strengthens the practical case for temporal EEG forecasting, while also showing that improved performance need not require an excessively complex architecture.

5.7 Chapter Summary

This chapter addressed the second research question by investigating temporal modelling for microsleep forecasting from EEG. A published CSP+SBL baseline was

reproduced, a GRU-based recurrent baseline was established, and an attention-augmented GRU architecture with focal loss was proposed for multi-horizon prediction.

The results showed that the proposed model improved forecasting performance across onset and extended horizons on the MWT dataset, with particularly strong gains in sensitivity and discrimination. The ablation study demonstrated that attention and focal loss make complementary contributions, while the cross-dataset evaluation showed that the approach remains beneficial, though more constrained, in single-channel consumer EEG. The comparison with EEG-TCNet further supported the value of the proposed design.

Taken together, the findings of this chapter show that improving microsleep prediction performance requires more than informative features alone. Temporal organisation and causal sequence modelling are also central. Although temporal modelling substantially improves EEG-based forecasting performance, the results also suggest limits to what can be achieved using EEG alone, particularly under constrained recording conditions. This motivates the investigation of multimodal EEG–EOG integration in Chapter 6, where complementary physiological information is incorporated to determine whether prediction performance can be improved beyond what is achievable using EEG alone.

Chapter 6

Multimodal EEG–EOG Fusion for Forecasting

6.1 Introduction

Chapter 5 showed that temporal modelling improves microsleep forecasting from EEG by preserving fine-grained structure in the pre-onset period. In particular, the results demonstrated that forecasting performance depends not only on the quality of the extracted EEG features, but also on how those features are organized and weighted over time. That chapter therefore addressed the temporal dimension of the problem. The present chapter extends the investigation to the *multimodal* dimension by examining whether microsleep prediction can be improved further through the joint use of electroencephalography (EEG) and electrooculography (EOG).

Microsleep prediction remains challenging because the physiological changes preceding the event are subtle, variable across subjects, and distributed across multiple systems. The *Maintenance of Wakefulness Test* (MWT) provides a controlled environment for studying these transitions because it includes simultaneous EEG and EOG recordings under standardized vigilance conditions [158]. EEG reflects cortical changes associated with declining vigilance, whereas EOG captures ocular manifestations such as blink dynamics, slow eye movements, and eyelid closure [10],

[47], [177]. These complementary signals motivate the investigation of multimodal EEG–EOG fusion for microsleep forecasting.

Most prior microsleep studies have been effectively unimodal, with EEG being the dominant modality. Chapter 4 examined informative EEG features for detection, and Chapter 5 considered how temporal modelling can improve EEG-based forecasting. The next logical step is to test whether multimodal fusion improves prediction performance further. This chapter addresses that question through an interpretable fusion framework based on information-theoretic coupling measures [178]. Rather than treating EEG and EOG as independent feature sources, the chapter explicitly models intra-modal and cross-modal dependencies using mutual information (MI), and it further examines whether directional interactions measured by transfer entropy (TE) provide additional value [179].

This chapter addresses the following research question:

RQ3: Can multimodal fusion of EEG and EOG improve microsleep prediction performance under subject-independent evaluation, and what type of cross-modal interaction is most informative for this task?

The contribution of this chapter is threefold. First, it establishes a multimodal EEG–EOG microsleep prediction framework using MWT data. Second, it evaluates whether nonlinear coupling features based on mutual information improve performance beyond unimodal EEG or EOG models. Third, it investigates whether directional interaction features based on transfer entropy provide further discriminative value. In doing so, the chapter moves from temporal modelling within EEG to joint modelling across modalities.

6.2 Related Work

Microsleep detection and prediction have been studied using physiological, behavioural, and visual signals. EEG has been the most common modality because it directly reflects neural changes associated with vigilance decline. Early work showed that

spectral slowing, particularly increased theta activity and reduced alpha activity, is associated with drowsiness and microsleep [9], [11]. More recent EEG-based systems have used machine learning and deep learning to detect microsleep onset, although performance often declines as the prediction horizon increases [13], [180], [181]. Connectivity-oriented approaches have also been explored in EEG for microsleep characterization, indicating that relations between channels may contain useful information beyond local features [99].

Complementary modalities have also been explored. Image-based approaches have analysed facial behaviour and eye closure [51], [52], [182], while speech-based systems have attempted to detect fatigue-related changes in vocal expression [173]. Video-based systems using geometric descriptors such as the eye aspect ratio (EAR) have also been proposed for microsleep detection [46]. These approaches suggest that non-neural indicators of vigilance can be informative, particularly when direct neural monitoring is unavailable.

Despite this broader multimodal interest, prior microsleep work has rarely combined EEG and EOG for *prediction*. This is an important gap. EEG and EOG capture different aspects of the same transition: EEG reflects cortical slowing and altered neural coordination, while EOG reflects ocular behaviour that often accompanies reduced vigilance [10], [140]. Their combination is therefore physiologically plausible. However, direct feature concatenation may fail to capture the nonlinear dependencies between modalities.

Information-theoretic measures provide one way to address this limitation. Mutual information measures shared dependence between signals without assuming linearity, while transfer entropy estimates directional information flow. These tools have been used in related areas of physiological analysis, but their application to EEG–EOG microsleep prediction has remained limited. Recent multimodal work in sleep and physiological monitoring has also explored representation learning and knowledge distillation [183], and multimodal neural networks have been used in fatigue detection tasks [29]. However, many such approaches are less interpretable

than feature-level information-theoretic fusion.

The present chapter therefore examines whether explicit EEG–EOG coupling features improve microsleep prediction, and whether the relevant interaction is best characterised as symmetric synchrony or directed influence.

6.3 Methodology

6.3.1 Dataset

This study used the Maintenance of Wakefulness Test (MWT) dataset, which was introduced in Chapter 3 and used throughout the thesis [158]. The dataset comprises simultaneous EEG and EOG recordings collected under controlled laboratory conditions and provides the benchmark dataset for evaluating microsleep prediction. Full details of the acquisition protocol, channel configuration, signal preprocessing, and expert annotation procedure are in Chapter 3 .

For the present study, labels 1, 2, and 3 were grouped into the positive class, while wakefulness was retained as the negative class. This binary formulation was chosen because MSEc and ED represent early or partial manifestations of the same vigilance decline that culminates in microsleep onset. Recordings were segmented into 1-second non-overlapping epochs and relabelled by majority vote, yielding approximately 60,548 wakefulness samples and 37,456 microsleep-labelled samples.

To investigate short-horizon prediction rather than event detection at onset, a 2-second pre-onset window was extracted immediately before each microsleep-labelled epoch. This horizon was chosen to balance physiological discriminability with predictive relevance, following the rationale from the temporal forecasting literature [13]. All signals were z -scored per subject to reduce inter-individual amplitude variation.

A strict *leave-one-subject-out* (LOSO) cross-validation protocol was used throughout. Under this scheme, all samples from one participant formed the test set while the remaining participants were used for training. This provides a subject-

independent estimate of performance and directly evaluates cross-subject generalisation.

6.3.2 Feature Extraction

Feature extraction was performed separately for EEG, EOG, mutual information, and transfer entropy features.

EEG Features

From the two EEG channels, spectral power was computed in the delta (0.5–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) bands. In addition, Hjorth parameters (activity, mobility, complexity) and entropy-based descriptors (Shannon entropy and permutation entropy) were extracted. These choices are consistent with both the microsleep literature and the feature-analysis results reported earlier in this thesis [148], [175]. For each preprocessed O1 and O2 segment, the amplitude envelope was computed as the magnitude of the analytic signal obtained using the Hilbert transform. For a signal $x(t)$, the envelope was therefore defined as

$$A_x(t) = |x(t) + j\mathcal{H}\{x(t)\}|, \quad (6.1)$$

where $\mathcal{H}\{x(t)\}$ denotes the Hilbert transform. Intra-hemispheric mutual information was then computed between the O1 and O2 amplitude envelopes at time lags ranging from -500 to $+500$ ms to quantify nonlinear short-range neural coupling.

EOG Features

From the two EOG channels, blink rate, blink duration, blink amplitude, saccadic energy, ratios of slow to fast eye movements, and spectral power in the 0.5–35 Hz range were extracted. These features were intended to represent the ocular behaviour associated with vigilance reduction, including slow eye movement activity and prolonged closure tendencies. Eight intra-EOG MI features were also computed to capture nonlinear coordination between the vertical and horizontal channels.

Cross-Modal MI Features

To represent shared EEG–EOG behaviour, mutual information was computed for each EEG–EOG channel pair (O1/E1, O1/E2, O2/E1, O2/E2) using lags of ± 500 ms. These features quantify nonlinear neural–ocular synchrony and form the core of the multimodal coupling representation.

Transfer Entropy Features

To examine whether directional information transfer contributes additional predictive value, transfer entropy was estimated with a k -nearest-neighbour estimator at lags of 5 ms and 50 ms. Directional interactions included EEG→EOG, EOG→EEG, O1→O2, and E1→E2. These TE features were designed to test whether microsleep prediction benefits from directional interaction modelling beyond the symmetric dependency represented by MI.

Information-Theoretic Measures

To explicitly model nonlinear dependencies and interactions between signals, this study employs mutual information (MI) and transfer entropy (TE), which are well-established measures from information theory.

Mutual Information (MI) Mutual information quantifies the amount of shared information between two random variables without assuming linearity. For two variables X and Y , MI is defined as

$$I(X; Y) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} p(x, y) \log \frac{p(x, y)}{p(x)p(y)}, \quad (6.2)$$

where $p(x, y)$ is the joint probability distribution of X and Y , and $p(x)$ and $p(y)$ are their marginal distributions. In the present context, MI captures nonlinear synchrony between EEG–EEG, EOG–EOG, and EEG–EOG channel pairs.

Transfer Entropy (TE) Transfer entropy extends mutual information by quantifying directed information flow between time series. For two processes X and Y , the transfer entropy from X to Y is defined as

$$T_{X \rightarrow Y} = \sum p(y_{t+1}, y_t^{(k)}, x_t^{(l)}) \log \frac{p(y_{t+1} | y_t^{(k)}, x_t^{(l)})}{p(y_{t+1} | y_t^{(k)})}, \quad (6.3)$$

where $y_t^{(k)}$ and $x_t^{(l)}$ denote the past k and l observations of Y and X , respectively. TE therefore measures how much knowledge of the past of X reduces uncertainty about the future of Y , beyond what is already explained by the past of Y alone.

In this work, TE is used to evaluate whether directional neural–ocular interactions provide additional predictive information beyond the symmetric dependencies captured by MI.

Figure 6.1 illustrates the computation and interpretation of MI and TE. MI measures symmetric shared information, whereas TE determines whether the past of one signal provides additional information about the future of another signal beyond that contained in the target signal’s own past.

As shown in the left panel of Figure 6.1, MI is symmetric and therefore indicates the strength of statistical coupling without identifying which modality leads the interaction. In contrast, the right panel shows that TE conditions on the target signal’s own history and measures the additional reduction in uncertainty obtained from the source signal’s past. Consequently, EEG→EOG and EOG→EEG TE were computed separately.

6.3.3 Fusion Strategies

Multimodal fusion refers to the integration of information from two or more modalities to obtain a representation or decision that is more informative than that provided by either modality independently. Fusion approaches are commonly categorised according to the stage at which integration occurs. Early, or feature-level, fusion combines features from the modalities before classification. Intermediate fu-

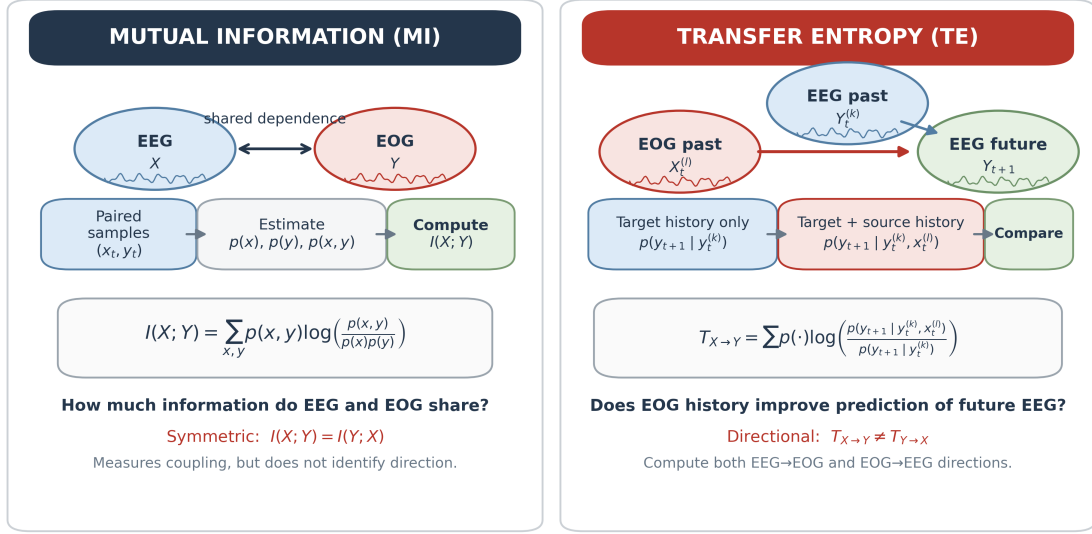


Figure 6.1: Computation and interpretation of mutual information (MI) and transfer entropy (TE). The left panel shows that MI quantifies symmetric shared information between paired EEG and EOG observations using their joint and marginal distributions. The right panel shows that TE quantifies directional predictive information by comparing prediction of the target signal’s future from its own history with prediction using both the target and source histories.

sion learns a joint representation within the model, while late, or decision-level, fusion combines the outputs of separately trained modality-specific models. Hybrid approaches may combine information at more than one of these stages. In this thesis, early and late fusion were selected to compare direct joint feature modelling with the combination of modality-specific predictions. Two fusion strategies were considered.

Early Fusion In early fusion, EEG, EOG, and MI-based cross-modal features were concatenated into a single representation prior to classification. This allows the classifier to model shared structure directly at the feature level.

Late Fusion In late fusion, separate unimodal classifiers were trained for EEG and EOG. Their probabilistic outputs were then combined using averaging or logistic stacking. This maintains modality-specific models and only combines decisions at the output level.

The comparison between early and late fusion was included to determine whether the predictive gain arises from explicit joint representation learning or simply from

aggregating unimodal decisions.

6.3.4 Classifier

All experiments used a Light Gradient Boosting Machine (LightGBM) classifier [184]. LightGBM is a gradient-boosted decision tree algorithm that constructs an ensemble of weak learners to model nonlinear relationships efficiently in tabular data. This model was selected for three reasons. First, it is suitable for tabular biomedical features with nonlinear interactions. Second, it performs well under class imbalance. Third, LightGBM supports post-hoc interpretation using SHapley Additive exPlanations (SHAP) [185]. SHAP assigns each feature a contribution value for an individual prediction using principles derived from cooperative game theory. Positive and negative SHAP values indicate whether a feature increases or decreases the model output relative to its baseline prediction, while aggregated absolute SHAP values provide an estimate of global feature importance. Hyperparameters were tuned using nested cross-validation within the training data.

6.3.5 Statistical Analysis

Performance was evaluated using accuracy, sensitivity, specificity, precision, F1-score, AUC-ROC, and AUC-PR. Model comparison used nonparametric statistics:

- the Friedman test for overall multi-model comparison;
- the Wilcoxon signed-rank test for pairwise model comparison with Bonferroni correction; and
- Cohen’s d for effect-size estimation.

This procedure was chosen because LOSO results are paired across subjects and may not satisfy normality assumptions.

The overall experimental design adopted throughout this work is illustrated in Fig. 6.2, highlighting the progressive integration of unimodal, multimodal, and information-theoretic modelling stages under a unified evaluation framework.

Experimental Stages and Study Design: Systematic Methodology for Microsleep Prediction

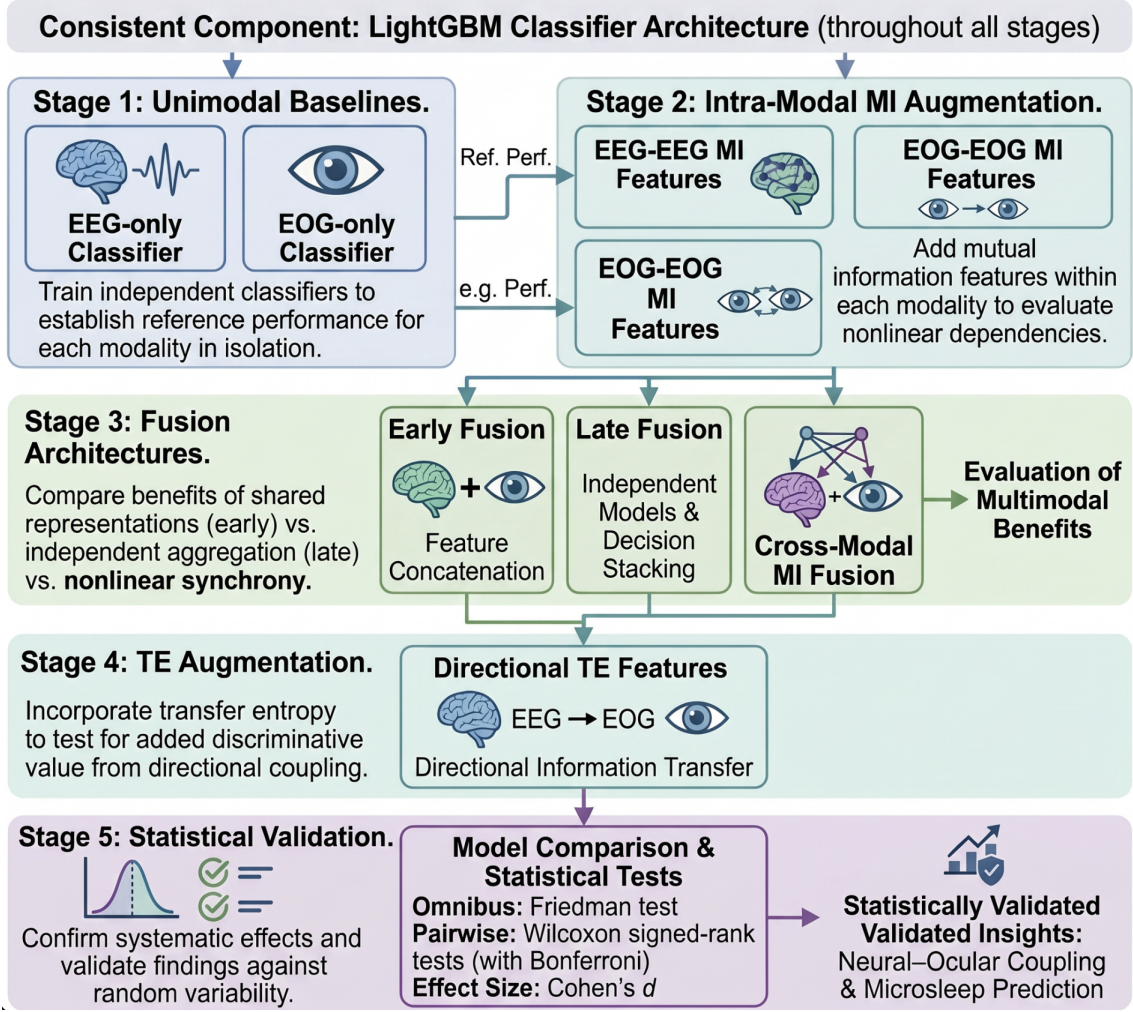


Figure 6.2: Experimental stages and study design for microsleep prediction. The framework progresses from unimodal baselines to intra-modal mutual information augmentation, multimodal fusion strategies, transfer entropy integration, and statistical validation, while maintaining a consistent LightGBM classifier across all stages.

6.3.6 Experimental Design

The study was organised into five stages:

1. **Stage 1: Unimodal baselines.** EEG-only and EOG-only models were trained independently.
2. **Stage 2: MI-enhanced unimodal models.** Intra-modal MI features were added within EEG and EOG.
3. **Stage 3: Fusion architectures.** Early fusion, late fusion, and MI-based

Table 6.1: Stage 1: Unimodal baselines under LOSO validation (mean $\pm$ standard deviation).

Metric	EEG	EOG
Accuracy	0.82 ± 0.15	0.68 ± 0.13
Sensitivity	0.14 ± 0.15	0.68 ± 0.24
Specificity	0.97 ± 0.06	0.63 ± 0.17

cross-modal fusion were compared.

4. **Stage 4: TE augmentation.** Transfer entropy features were added to test directional interaction effects.
5. **Stage 5: Statistical validation.** All configurations were compared using Friedman and Wilcoxon tests with effect sizes.

This staged design makes it possible to attribute any performance gains to a specific methodological component rather than to a simultaneous change in architecture and feature space.

6.4 Results

6.4.1 Stage 1: Unimodal Baselines

Table 6.1 summarises the LOSO results for the unimodal EEG and EOG classifiers. The EEG model achieved higher overall accuracy and very high specificity, but low sensitivity. This indicates that EEG alone behaved conservatively under subject-independent evaluation, identifying non-microsleep states reliably but missing many positive events.

The EOG model showed the opposite tendency. It achieved lower overall accuracy and specificity but substantially higher sensitivity, suggesting that ocular dynamics are more reactive to the microsleep transition. These complementary detection profiles provide a clear rationale for multimodal integration.

The interpretation of this is that EEG alone is stable and conservative, whereas EOG is more event-sensitive. This contrast suggests that each modality captures a

Table 6.2: Stage 2: Effect of intra-modal MI features on unimodal models.

Model	$\Delta F1$	ΔAUC_{pr}	$\Delta \text{Sensitivity}$	$p\text{-value}$
EEG + MI	+0.00	+0.02	-0.04	0.049*
EOG + MI	+0.31	+0.40	+0.63	<0.001***

different aspect of the microsleep process and provides motivation for multimodal fusion.

6.4.2 Stage 2: MI-Enhanced Unimodal Models

Mutual information features were then added within each modality to test whether nonlinear dependencies improved prediction. As shown in Table 6.2, MI augmentation improved both modalities, although the effect was more pronounced for EOG.

For EOG, MI substantially increased AUC-PR and sensitivity, indicating improved detection of the positive class under imbalance. For EEG, the gains were smaller and the effect on sensitivity was slightly negative, consistent with the already conservative behaviour of the EEG baseline. Figure 6.3 shows the subject-wise comparison between the baseline EOG model and the EOG+MI model, while Figure 6.4 summarises the improvement patterns across both modalities.

The interpretation of this is that MI features improve robustness by introducing nonlinear dependency structure within the modality, particularly in EOG where blink and eye-movement coordination appear strongly informative.

6.4.3 Stage 3: Fusion Architectures

Table 6.3 compares the principal fusion configurations. Both early and late fusion improved markedly over the unimodal baselines, but early fusion with cross-modal MI achieved the best overall balance. Specifically, *Early Fusion (MI Cross)* achieved an F1-score of 0.73 and an AUC-PR of 0.81, with sensitivity of 0.70 and specificity of 0.80. Late fusion produced similar but slightly lower values.

These results indicate that the main gain is not simply from combining two predictions after the fact, but from explicitly modelling shared neural-ocular structure

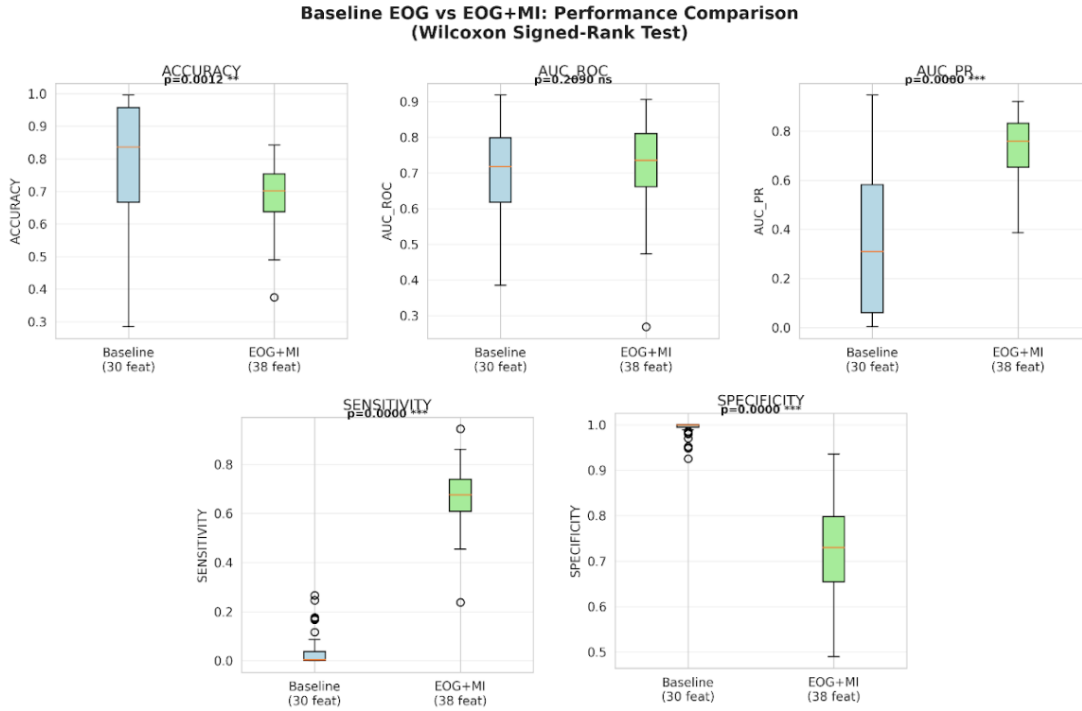


Figure 6.3: Comparison of baseline EOG and EOG+MI models under LOSO validation. Boxplots show subject-wise distributions for accuracy, AUC_{ROC} , AUC_{PR} , sensitivity, and specificity.

Table 6.3: Stage 3: Fusion models under LOSO evaluation.

Model	ACC	F1		SEN	SPEC
EEG	0.82	0.18	0.39	0.14	0.97
EOG	0.68	0.37	0.34	0.68	0.63
Early Fusion (MI Intra)	0.74	0.73	0.81	0.68	0.80
Early Fusion (MI Cross)	0.75	0.73	0.81	0.70	0.80
Late Fusion (avg)	0.74	0.73	0.80	0.70	0.79

at the feature level. Figure 6.5 shows the metric distributions across the evaluated fusion architectures.

Based on these results, Early Fusion (MI Cross) was selected as the best-performing configuration.

6.4.4 Feature Importance Analysis

To further investigate the interpretability of the selected multimodal model, feature importance analysis was performed using the gain-based importance scores produced by the best-performing LightGBM Early Fusion (MI Cross) model. Fig-

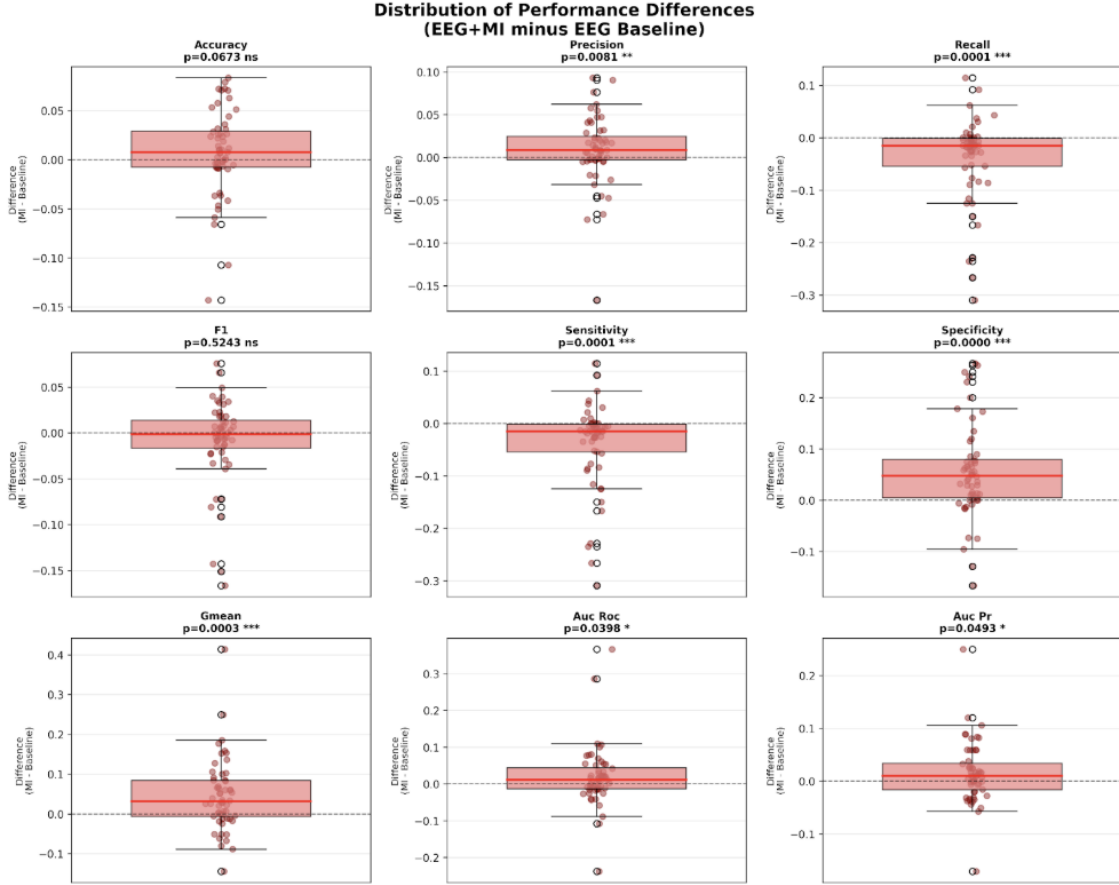


Figure 6.4: Stage 1–2 results: unimodal baselines and MI-enhanced models. MI improves both modalities, with the clearest effect observed for EOG sensitivity and AUC_{PR} .

ure 6.6 presents the highest-ranked features averaged across LOSO folds.

The results show that ocular dynamics contributed strongly to microsleep prediction. In particular, saccade rate was the most important feature, followed by EOG spectral activity and saccade-velocity measures. This indicates that ocular behaviour provides highly informative markers of the pre-microsleep transition. Several EEG nonlinear features from the occipital channels also appeared among the highest-ranked predictors, suggesting that cortical signal complexity remains important for distinguishing pre-microsleep states.

Importantly, cross-modal mutual-information features also appeared among the top-ranked predictors, including EEG–EOG MI coupling measures. This supports the interpretation that nonlinear neural–ocular synchrony contributes meaningful predictive information before microsleep onset. The feature importance analysis

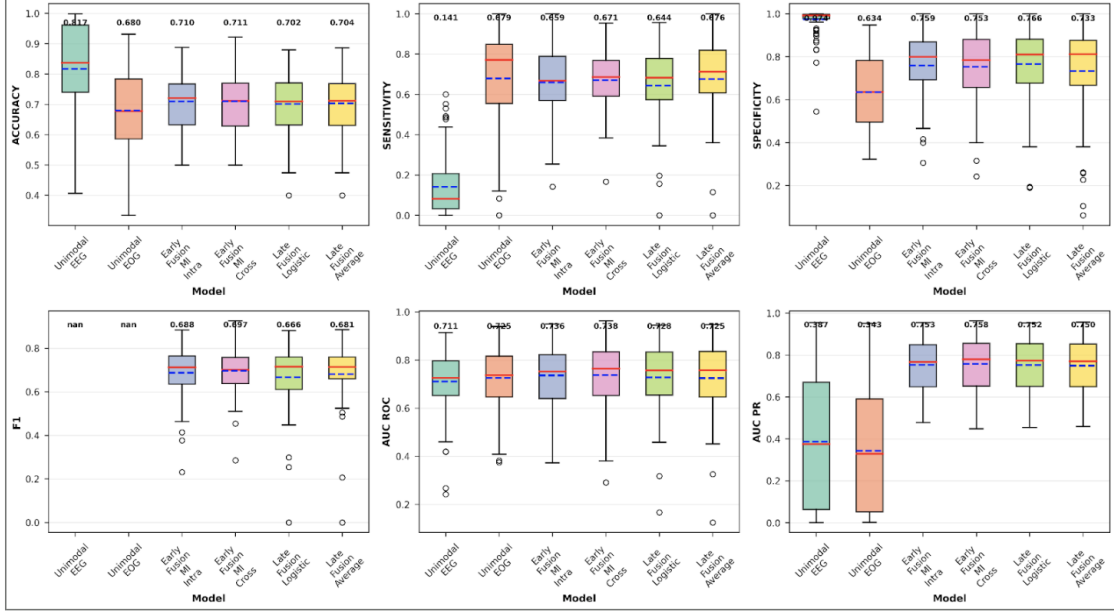


Figure 6.5: Stage 3 results: fusion architectures. Early fusion with cross-modal MI provides the strongest balance between sensitivity and specificity and the highest AUC_{PR} .

therefore provides additional support for the observed improvement of multimodal fusion over unimodal baselines and reinforces the interpretability of the proposed information-theoretic fusion framework.

6.4.5 Stage 4: Transfer Entropy Augmentation

Transfer entropy features were then added to test whether directional interactions improved prediction beyond symmetric MI-based coupling. A total of 40 TE features were added, increasing the feature count from 71 to 111. As shown in Table 6.4, none of the evaluated metrics improved significantly, and the average performance changed only marginally, with a slight decline overall.

This suggests that the relevant multimodal structure at the 2-second prediction horizon is not primarily directional. Instead, microsleep-related EEG–EOG interaction appears to be better represented as synchronous co-modulation. Figure 6.7 shows the subject-wise comparison between MI Cross and MI+TE.

Interpretation: At the studied horizon, multimodal microsleep prediction is better explained by bidirectional synchrony than by directional influence.

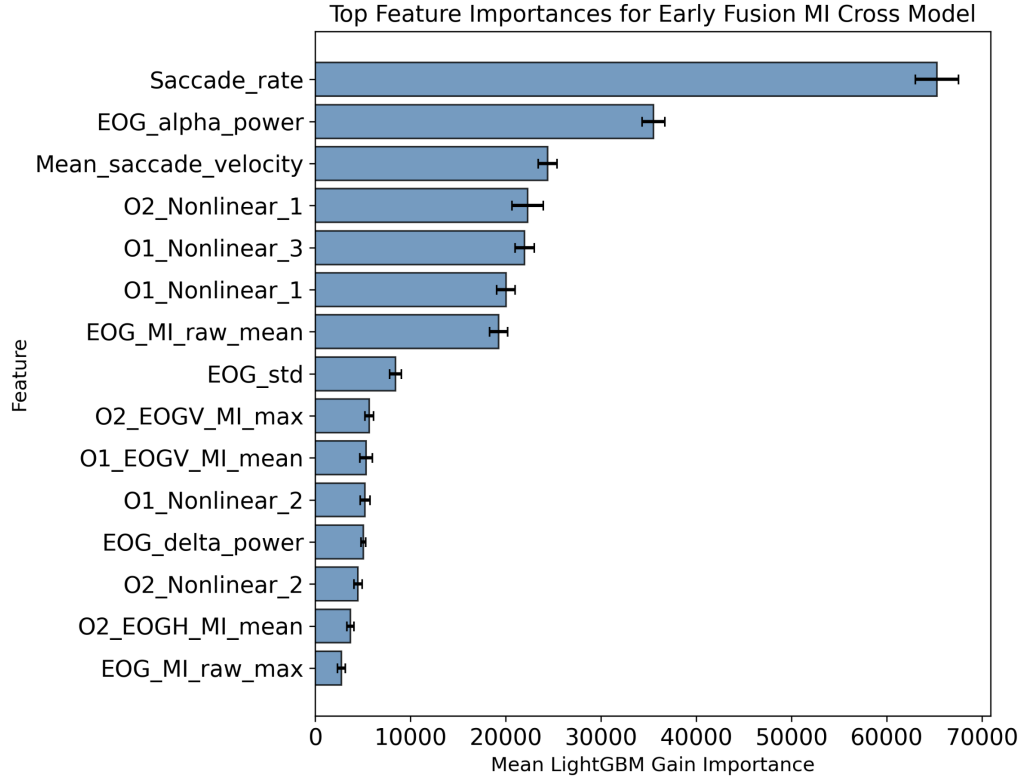


Figure 6.6: Top feature importance scores for the best-performing Early Fusion (MI Cross) LightGBM model. Importance values correspond to mean LightGBM gain importance across LOSO folds, with error bars indicating variability across folds. Ocular features, EEG nonlinear features, and EEG–EOG mutual-information features all contribute to the selected multimodal prediction model.

6.4.6 Stage 5: Statistical Validation

The Friedman test showed statistically significant differences across models for multiple metrics, including accuracy, sensitivity, specificity, and AUC–PR. Subsequent Wilcoxon tests showed that Early Fusion (MI Cross) significantly outperformed the unimodal baselines and the late-fusion alternatives, with corrected p -values below 0.005 in the main comparisons. Reported effect sizes were large, indicating that the differences were not only statistically detectable but also practically meaningful.

Figure 6.8 presents a multi-metric radar-style comparison of the main models and illustrates the balanced profile of the selected multimodal configuration.

Table 6.4: Stage 4: Comparison between the selected MI Cross model and the TE-augmented model.

Metric	MI Cross	MI+TE	<i>p</i> -value
	0.81	0.80	0.94 (ns)
F1	0.73	0.71	0.81 (ns)
Sensitivity	0.70	0.69	0.43 (ns)
Specificity	0.80	0.79	0.83 (ns)
Accuracy	0.75	0.74	0.90 (ns)

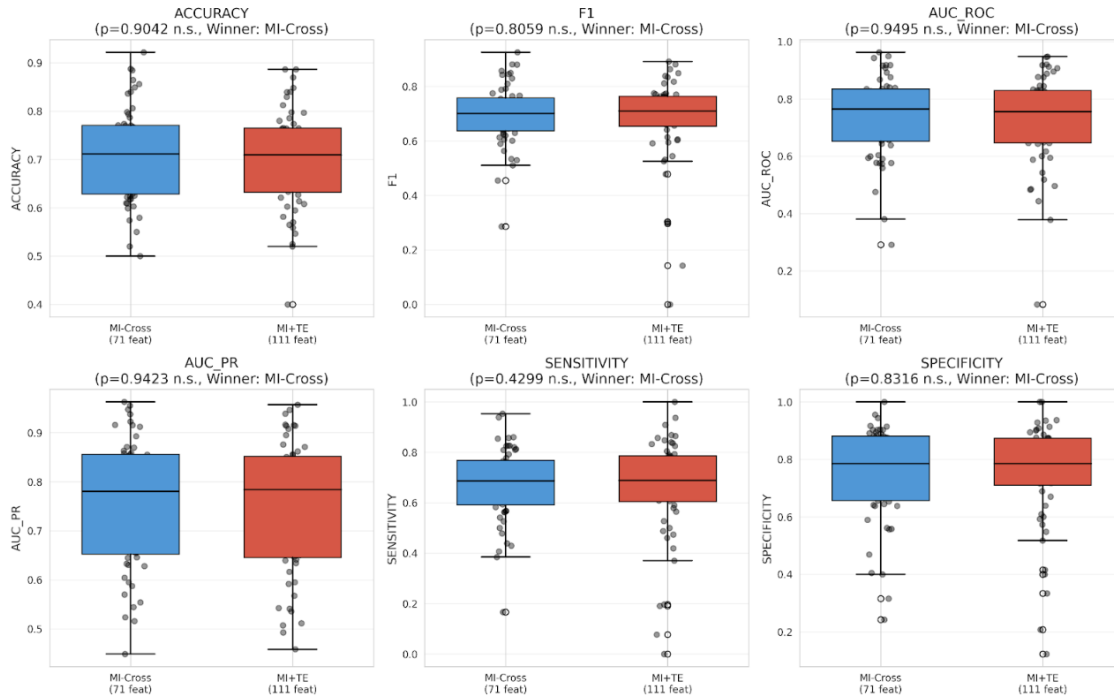


Figure 6.7: Stage 4 results: transfer entropy augmentation. Adding TE increased the feature count without improving predictive performance.

6.5 Discussion

6.5.1 Value of Multimodal Neural–Ocular Fusion

The main finding of this chapter is that multimodal EEG–EOG fusion improves microsleep prediction under subject-independent evaluation when cross-modal dependencies are modeled explicitly. The unimodal baselines showed complementary strengths: EEG alone was highly specific but insufficiently sensitive, whereas EOG alone was more sensitive but less specific. This complementarity is important because it indicates that each modality captures a different component of the pre-

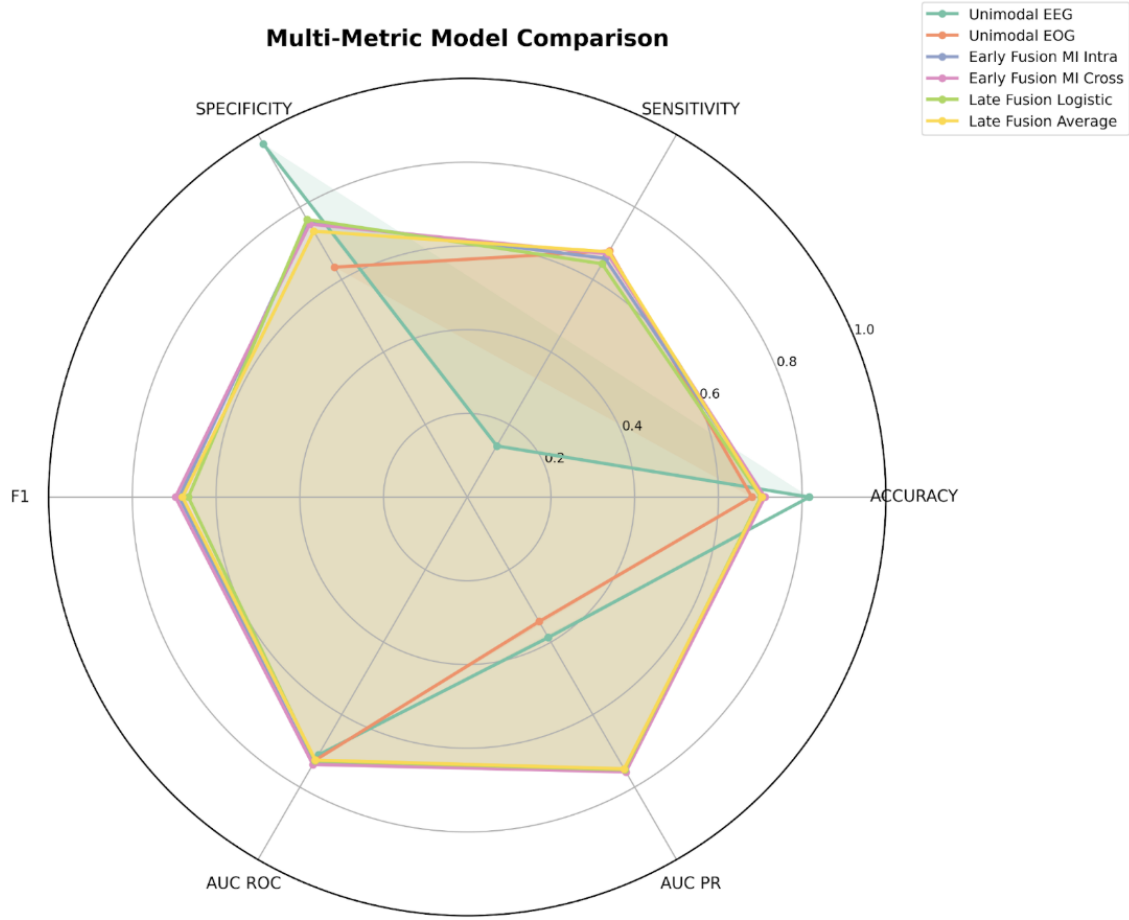


Figure 6.8: Stage 5 results: statistical comparisons across the principal models. Early Fusion (MI Cross) outperformed the unimodal and late-fusion baselines across the main evaluation metrics.

microsleep state.

The best results were achieved when those complementary signals were fused at the feature level using mutual information descriptors. This suggests that the predictive gain arises from the joint structure of neural and ocular activity rather than from the simple coexistence of two separate information sources. In other words, it is not only *what* EEG and EOG individually contain that matters, but also *how* they vary together before onset.

Chapter 5 addressed the challenge of preserving temporal resolution within EEG-based forecasting. The present chapter builds on that contribution by showing that, even with improved temporal modelling, additional predictive benefit can be ob-

tained from multimodal coupling. The two chapters therefore answer different parts of the broader problem. Chapter 5 improved the temporal representation of EEG features, while Chapter 6 improved the physiological representation by integrating neural and ocular information.

Taken together, these findings indicate that microsleep prediction performance depends on both dimensions: temporal structure and modality interaction. This provides a stronger foundation for later work on sample-efficient learning, where the question becomes how to make use of these richer representations without requiring exhaustive annotation.

6.5.2 Interpretation of MI and TE Findings

A central methodological question in this chapter was whether multimodal interaction is best represented by shared dependence or by directed influence. The results favoured the former. MI-based fusion consistently improved performance, whereas TE augmentation did not provide additional benefit.

This outcome suggests that the physiological transition into microsleep is characterised more by synchronous co-modulation than by a stable directional pathway from one modality to the other. In practical terms, cortical slowing and ocular behaviour appear to evolve together in the short pre-onset window, without a directional interaction pattern that is sufficiently consistent across subjects to improve prediction.

This interpretation also explains why the TE feature expansion may have introduced redundancy. Increasing feature dimensionality does not help if the added variables do not correspond to a stable discriminative pattern. Under LOSO validation, this is especially important because subject-independent generalisation penalises representations that overfit subject-specific interaction structures.

6.5.3 Methodological Implications

The study also provides methodological guidance for multimodal vigilance monitoring. First, feature-level early fusion is preferable to late decision fusion for this task, because it allows the model to use shared neural–ocular representations directly. Second, information-theoretic features can offer an interpretable alternative to less transparent multimodal deep architectures, particularly when the feature space is structured around known physiology. Third, subject-independent validation is essential. Many multimodal gains appear larger under subject-dependent settings, but the present results show that meaningful improvement can still be obtained under stricter LOSO evaluation.

6.6 Chapter Summary

This chapter addressed the third research question by investigating whether microsleep prediction can be improved through multimodal fusion of EEG and EOG. Using the MWT dataset under leave-one-subject-out evaluation, the chapter compared unimodal baselines, MI-enhanced unimodal models, feature-level and decision-level fusion architectures, and TE-augmented multimodal models.

The results showed that EEG and EOG provide complementary predictive information. EEG alone was highly specific but insensitive, whereas EOG alone was more sensitive but less specific. Mutual information features improved both modalities, with the clearest gains observed in EOG. The best overall performance was obtained with early fusion using cross-modal MI, which achieved the strongest balance across F1-score, sensitivity, specificity, and AUC–PR. Transfer entropy did not provide additional benefit, suggesting that the relevant pre-microsleep interaction is dominated by bidirectional synchrony rather than directed influence.

In relation to the previous chapter, these findings extend the thesis from temporal EEG modelling to multimodal physiological integration. The results demonstrate that richer multimodal representations can improve microsleep prediction perfor-

mance, but they also increase the importance of efficient use of labelled data. This practical consideration motivates the investigation of active learning strategies presented in Chapter 7.

Chapter 7

Active Learning for Label-Efficient Forecasting

7.1 Introduction

Having established a suitable feature representation, temporal structure, and multimodal integration strategy, the next practical question concerns *annotation efficiency*. The earlier chapters assumed that sufficiently large labelled datasets were available for model training and evaluation. In practice, however, the annotation of EEG and EOG recordings is labour-intensive, requires expert review, and becomes particularly costly when transitions between wakefulness, drowsiness, and microsleep are subtle. This creates a mismatch between the abundance of physiological recordings and the comparatively limited availability of verified labels.

This chapter addresses this by investigating active learning for multimodal microsleep forecasting. Active learning is a machine learning paradigm in which a model iteratively selects the most informative unlabeled samples for expert annotation. The goal is to maximise predictive performance while minimising labelling effort. Rather than treating all unlabeled samples as equally valuable, active learning seeks to select only those instances that are expected to provide the greatest improvement to the predictive model. In the present setting, this is especially relevant

because the forecasting task is imbalanced, physiologically variable across subjects, and evaluated under a strict leave-one-subject-out protocol.

A further motivation for this chapter arises directly from the multimodal findings of Chapter 6. If EEG and EOG capture complementary aspects of the transition into microsleep, then disagreement between the two modalities may itself provide information about which samples are ambiguous, informative, or close to the decision boundary. This suggests that multimodal disagreement can potentially be used not only for prediction, but also for label acquisition.

The central aim of this chapter is therefore to determine whether multimodal microsleep forecasting can be trained efficiently under limited supervision, and whether multimodal disagreement can be used as an interpretable acquisition mechanism.

7.2 Research Question and Chapter Scope

This chapter addresses the following research question:

RQ4: Can active learning reduce expert labelling requirements for multimodal microsleep forecasting while maintaining predictive performance, and can multimodal disagreement improve sample selection?

To address this question, the chapter has four objectives. First, it establishes a label-efficiency benchmark for multimodal microsleep forecasting using the best-performing model from Chapter 6 as a fully supervised reference. Second, it compares standard pool-based active learning strategies under subject-independent evaluation. Third, it introduces an interpretable multimodal disagreement-driven active learning method tailored to EEG–EOG fusion. Fourth, it analyses the contribution of each component of the proposed acquisition score through statistical comparison and ablation. In this way, the chapter extends the thesis from prediction performance alone toward training efficiency and practical deployment.

7.3 Methodology

7.3.1 Baseline Model

The active learning experiments in this chapter build directly on the strongest multimodal configuration identified in Chapter 6, namely the Early Fusion (MI Cross) LightGBM model. In that chapter, this model achieved the best subject-independent performance under leave-one-subject-out validation by combining EEG and EOG features with cross-modal mutual information descriptors. The role of the present chapter is not to redesign the predictive architecture, but to investigate how efficiently that architecture can be trained when only a limited subset of samples is labelled.

Using the same multimodal forecasting model throughout the experiments ensures that differences in performance can be attributed to the acquisition strategy rather than to changes in the representation or classifier.

7.3.2 Dataset and Forecasting Protocol

The experiments use the Maintenance of Wakefulness Test dataset introduced earlier in the thesis. As in Chapter 6, the signals are segmented into 1-second windows with a stride of 0.5 seconds, and the labels are shifted forward to construct two short-horizon forecasting tasks: 1.0-second-ahead prediction and 2.0-second-ahead prediction. Labels corresponding to microsleep episode, microsleep candidate, and drowsiness are grouped into the positive class, while wakefulness forms the negative class.

Evaluation follows a strict leave-one-subject-out protocol. In each fold, one subject is reserved entirely for testing and all remaining subjects form the training pool. This setting is intentionally demanding, as it tests whether an acquisition strategy remains effective under subject-independent generalisation rather than benefiting from subject-specific calibration.

7.3.3 Feature Representation

The same multimodal feature set developed in Chapter 6 is used here. Each sample is represented by 71 features. These include 21 EEG-derived features, 38 EOG-derived features, and 12 cross-modal mutual information descriptors. Transfer entropy features are not used, since the preceding chapter showed that they did not produce a statistically meaningful improvement in predictive performance. Fixing the feature representation across all experiments allows the present chapter to focus specifically on the label-selection problem.

7.3.4 Active Learning Framework

Within each LOSO fold, the training pool is initially treated as unlabeled except for a small balanced seed set, corresponding to approximately 1% of the available training samples. The model is trained on this seed set and then updated iteratively as additional samples are selected for annotation according to a chosen active learning strategy.

The labelling budgets considered in this chapter are

$$\mathcal{B} = \{1\%, 2\%, 5\%, 10\%, 20\%, 40\%, 60\%, 80\%, 100\%\}.$$

For each budget level $b \in \mathcal{B}$, the number of labeled samples is computed as

$$n_b = \max(20, \lfloor b N_{\text{train}} \rfloor), \quad (7.1)$$

where N_{train} denotes the size of the training pool in the corresponding LOSO fold. Performance is evaluated at every budget for both forecasting horizons.

The overall workflow is illustrated in Figure. 7.1. After multimodal feature extraction and LOSO splitting, labels are hidden in the training pool, candidate samples are queried according to the selected acquisition strategy, the early-fusion LightGBM model is retrained, and performance is evaluated on the held-out subject.

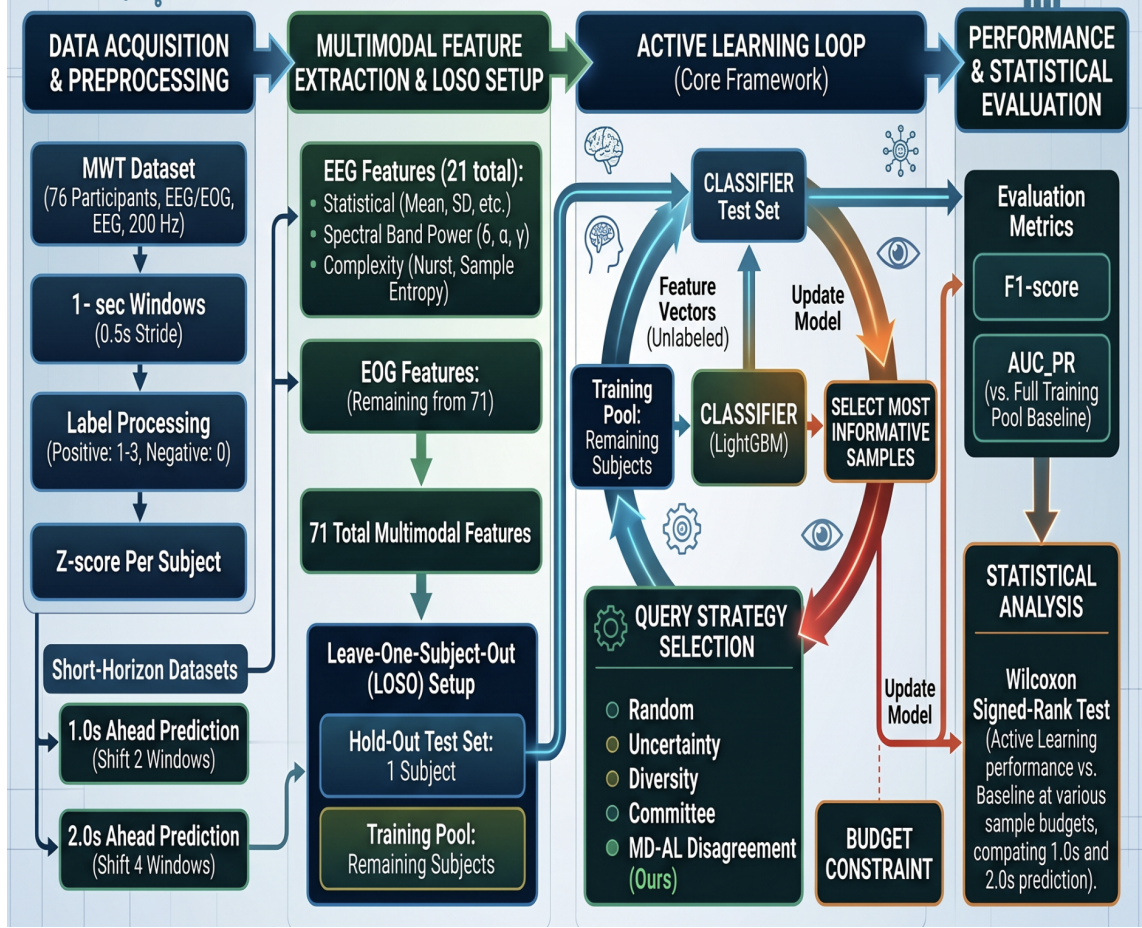


Figure 7.1: Overview of the active learning pipeline under leave-one-subject-out validation. Labels are hidden in the training pool and progressively revealed according to the acquisition strategy before retraining and evaluation on the held-out subject.

7.3.5 Standard Query Strategies

Four pool-based benchmark strategies are considered. Random sampling provides a passive baseline by selecting unlabeled samples uniformly. Uncertainty sampling selects instances with maximum predictive entropy, thereby prioritising points close to the decision boundary. Diversity sampling seeks to cover the feature space more broadly by selecting samples that are maximally dissimilar to previously chosen instances. Query-by-Committee estimates informativeness through disagreement among multiple LightGBM models with different hyperparameter settings.

These strategies are included because they represent common acquisition principles in active learning: randomness, uncertainty, representativeness, and model disagreement.

7.3.6 Proposed Multimodal Disagreement Active Learning

To exploit the multimodal structure of the forecasting problem, this chapter introduces *Multimodal Disagreement Active Learning* (MD-AL). The method is motivated by the observation that EEG and EOG may disagree most strongly on samples that correspond to transitional states, ambiguous evidence, or difficult decision regions. Such samples are plausible candidates for annotation because they may provide more information than samples on which all views already agree.

For each unlabeled sample x , three probability estimates are computed: $p_{EEG}(x)$ from an EEG-only model, $p_{EOG}(x)$ from an EOG-only model, and $p_{fusion}(x)$ from the multimodal early-fusion model. These probabilities correspond to the positive-class posterior probabilities predicted by the respective LightGBM models. The acquisition score is then defined as

$$S(x) = |p_{EEG}(x) - p_{EOG}(x)| \, 4p_{fusion}(x) (1 - p_{fusion}(x)) (0.5 + p_{fusion}(x)), \quad (7.2)$$

The first factor measures cross-modal disagreement. The second is a normalised boundary term that peaks when the fusion model is maximally uncertain. The third introduces a small rare-event bias so that likely microsleep samples are slightly favoured in the presence of class imbalance. The design is intentionally transparent: each term corresponds to a concept that can be interpreted physiologically or statistically.

Separate unimodal and multimodal models produce probability estimates which are combined into an interpretable query score. After ranking the unlabeled pool, a diversity filter based on feature-space distance is applied to reduce redundancy within each acquisition batch.

7.3.7 Ablation Design

To determine whether each component of the acquisition score is necessary, an ablation study is conducted using three variants: disagreement only, disagreement plus boundary-awareness, and the full MD-AL score. All other components of the pipeline remain unchanged. This makes it possible to analyse whether the gains of the proposed method arise from multimodal disagreement alone or from the interaction between disagreement, boundary uncertainty, and minority-event prioritisation.

7.3.8 Evaluation Metrics and Statistical Analysis

Performance is evaluated primarily using AUC-PR, since the forecasting task is imbalanced and the positive class is the operationally relevant target. AUC-ROC, F1-score, and recall at 90% specificity are also reported where relevant.

To assess statistical significance, Wilcoxon signed-rank tests are used to compare each active learning method against the fully supervised baseline under the one-sided alternative that active learning performance is lower. Friedman tests are then used to compare rankings across active learning strategies at matched budgets. This analysis is intended to distinguish between small numerical differences and differences that are statistically meaningful under the LOSO folds.

7.4 Results

7.4.1 Learning Curves Including the Proposed Method

The behaviour of the proposed MD-AL strategy is first examined in relation to the benchmark methods and the fully supervised baseline. Figure 7.2 shows the active learning curves for the 1.0-second forecasting horizon, and Fig. 7.3 shows the same comparison for the 2.0-second horizon.

At both horizons, performance rises sharply from the initial seed set and approaches the fully supervised baseline at relatively low labelling budgets. The most

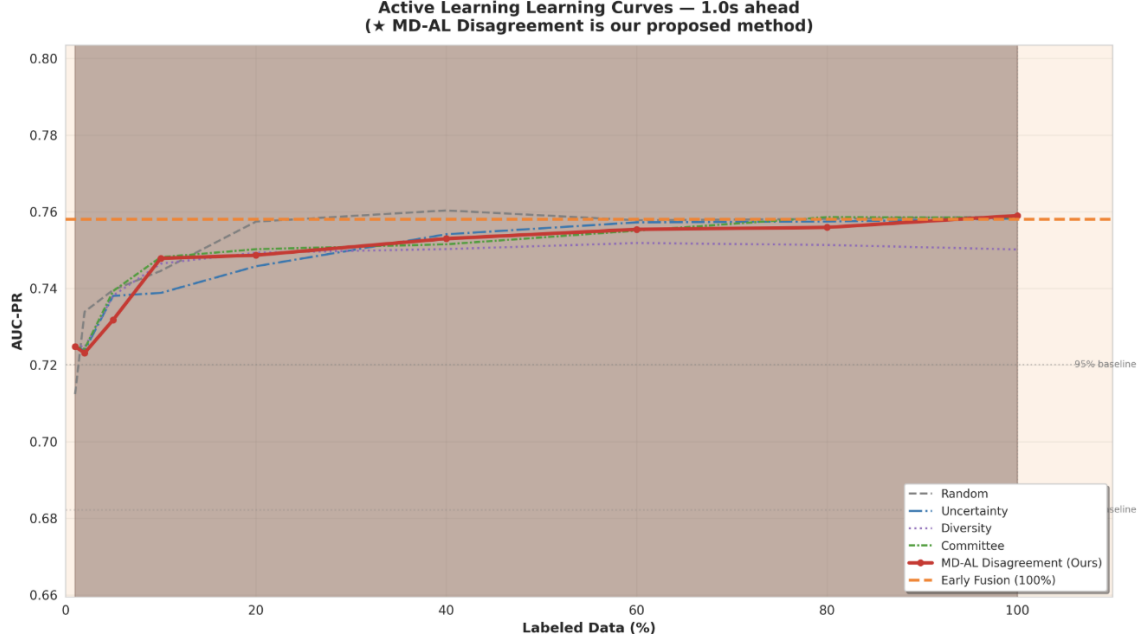


Figure 7.2: Active-learning performance for the 1.0s prediction horizon as the percentage of labelled training data increases. The horizontal axis represents the labelling budget, and the vertical axis reports AUC-PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline obtained using 100% labelled data. The horizontal dotted lines denote 90% and 95% of the baseline performance. The red curve represents the proposed MD-AL disagreement strategy, while the remaining curves represent Random, Uncertainty, Diversity, and Committee sampling.

noticeable improvement occurs between the smallest budgets, after which the learning curves flatten. This indicates that the majority of the decision structure can be recovered using only a small fraction of the available labels. The corresponding plots show that MD-AL remains close to the strongest benchmark methods across the budget range and converges near the fully supervised early-fusion baseline. For the 1.0-second task, MD-AL follows a smooth and stable trajectory and ends with one of the strongest final AUC-PR values. For the 2.0-second task, the same general pattern holds. This suggests that multimodal disagreement provides a viable acquisition signal not only at a single forecast horizon, but across the short-horizon forecasting setting examined in this thesis.

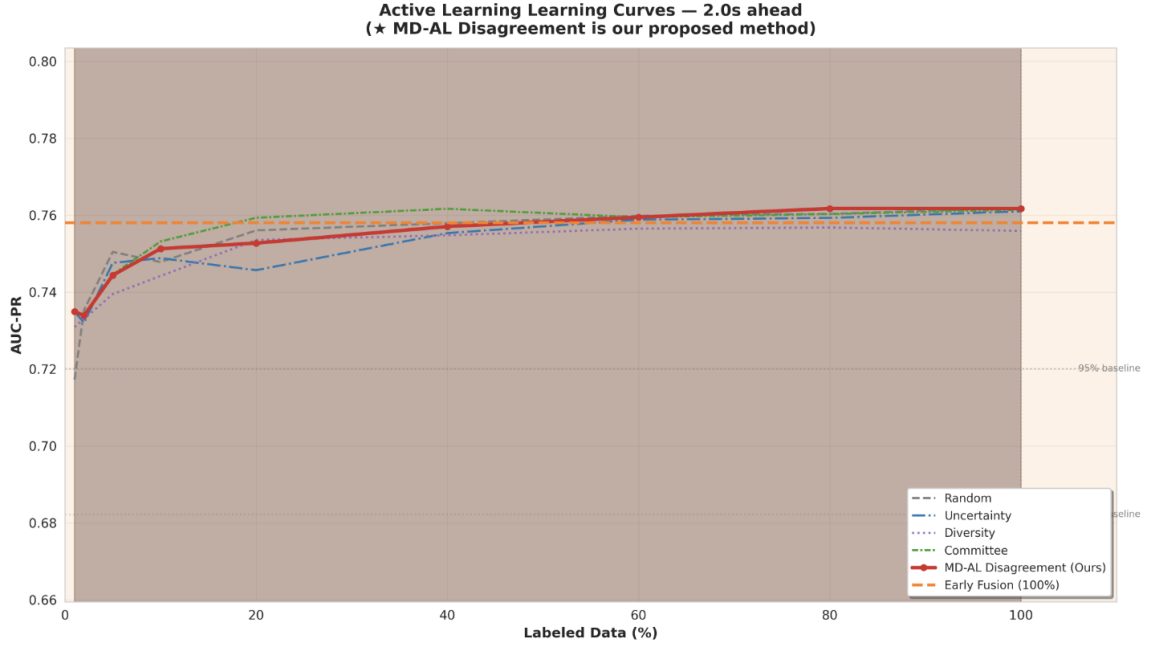


Figure 7.3: Active-learning performance for the 2.0s prediction horizon as the percentage of labelled training data increases. The horizontal axis represents the labelling budget, and the vertical axis reports AUC-PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline obtained using 100% labelled data, while the horizontal dotted lines indicate 90% and 95% of its performance. The red curve represents the proposed MD-AL disagreement strategy, and the remaining curves represent Random, Uncertainty, Diversity, and Committee sampling. Performance improves most strongly at the lowest labelling budgets, after which the methods largely converge towards the fully supervised baseline. MD-AL remains competitive across the evaluated budgets and slightly exceeds the baseline at the larger budgets, although no strategy is uniformly superior across the complete budget range.

7.4.2 Label-Efficiency Including the Proposed Method

The same behaviour is seen more explicitly in the efficiency plots of Figure. 7.4 and Fig. 7.5. These figures summarise the minimum labelling budget needed to reach 90%, 95%, and 100% of the fully supervised baseline.

Across both horizons, approximately 90% of baseline performance is reached with about 1% labeled data, and approximately 95% of baseline performance is reached with only 1–2% labeled data. The practical implication is that most of the forecasting capability can be retained with a very small annotation budget. At the stricter target of matching 100% of the fully supervised baseline, higher budgets are required, and the relative ordering of strategies becomes more apparent. In the

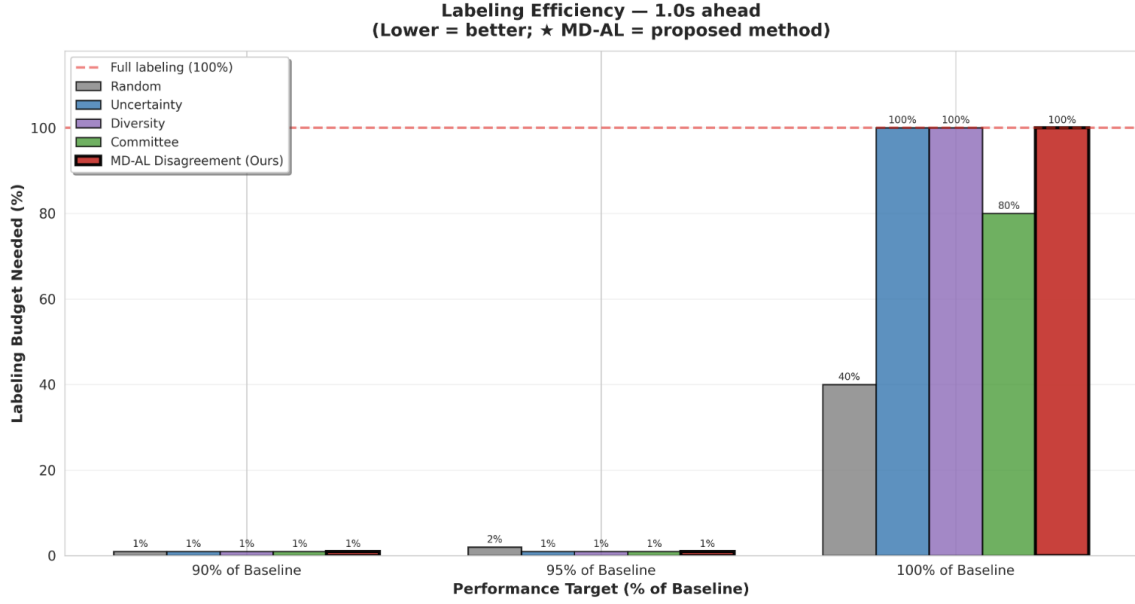


Figure 7.4: Minimum labelling budget required by each acquisition strategy to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 1.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with 1% labelled data. At the 95% target, Random sampling requires 2%, whereas the remaining strategies, including the proposed MD-AL method, require 1%. Recovering 100% of the baseline is substantially more demanding: Random requires 40%, Committee sampling requires 80%, and Uncertainty, Diversity, and MD-AL require the full labelling budget. The red dashed line indicates the 100% labelling level.

1.0-second task, MD-AL requires a larger budget than the Query-by-Committee and Random sampling strategies to match the baseline exactly, whereas in the 2.0-second task it reaches full baseline performance with a budget comparable to Random and Uncertainty but higher than Committee.

7.4.3 Final Performance Including the Proposed Method

At full labelling budget, the differences between methods become small. Figures 7.6 and 7.7 compare the final AUC-PR values of the benchmark strategies, the proposed MD-AL method, and the fully supervised early-fusion baseline.

For the 1.0-second horizon, MD-AL achieves one of the highest final AUC-PR values and is numerically very close to the fully supervised baseline. For the 2.0-second horizon, MD-AL again reaches a near-highest final value. These plots show

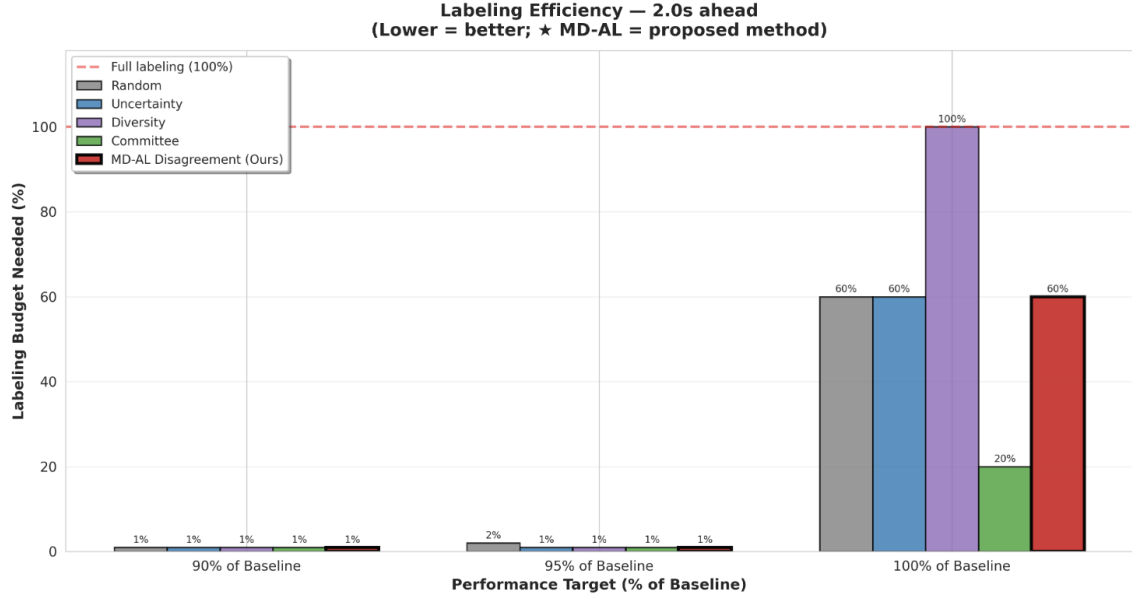


Figure 7.5: Minimum labelling budget required by each acquisition strategy to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 2.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas the remaining strategies, including the proposed MD-AL method, require 1%. Recovering 100% of baseline performance requires larger and more strategy-dependent budgets: Committee sampling requires 20%, Random, Uncertainty, and MD-AL each require 60%, and Diversity requires the full 100% labelling budget. The red dashed line indicates the full-labelling level.

that the proposed acquisition strategy is competitive with the strongest standard methods.

7.4.4 Benchmark Strategy Behaviour

To contextualise the proposed method, the benchmark strategies are examined separately. Figures 7.8 and 7.9 show the learning curves for Random, Uncertainty, Diversity, and Committee sampling relative to the fully supervised baseline.

At the 1.0-second horizon, all benchmark strategies approach the baseline by approximately 20% labelling. Random and Committee track the baseline closely beyond that point, whereas Uncertainty and Diversity converge more slowly. At the 2.0-second horizon, Committee sampling reaches baseline-level performance with lower budgets than the other standard methods, while Diversity remains less effi-

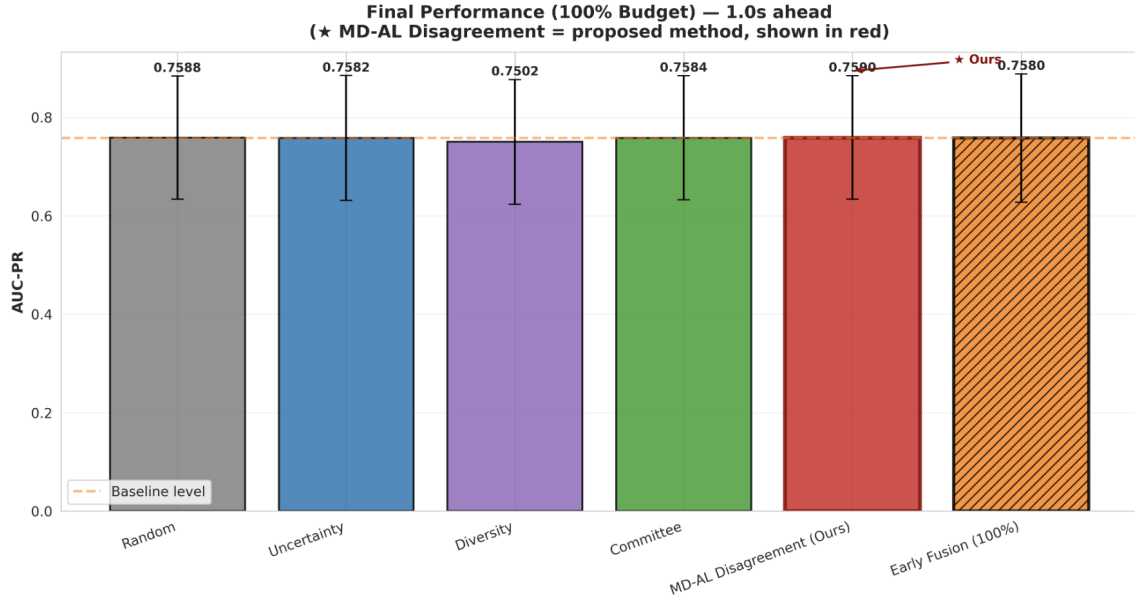


Figure 7.6: Final AUC-PR performance at the 100% labelling budget for the 1.0s prediction horizon. Bars show the mean performance of Random, Uncertainty, Diversity, Committee, and the proposed MD-AL disagreement strategies, together with the fully supervised early-fusion reference model; error bars indicate performance variability. The orange dashed line marks the early-fusion baseline value of 0.7580. MD-AL achieves the highest numerical mean AUC-PR (0.7590), followed closely by Random (0.7588), Committee (0.7584), and Uncertainty (0.7582), while Diversity obtains 0.7502. The substantial overlap between the error bars and the very small differences between the mean values indicate broadly comparable performance; therefore, the marginal numerical advantage of MD-AL should not be interpreted as evidence of clear statistical superiority.

cient.

These plots are useful because they show that the relative strength of a query strategy is most visible at moderate budgets. Once the budget becomes large enough, all methods converge to nearly the same endpoint, and the main difference lies in how quickly they get there.

The corresponding efficiency summaries are shown in Fig. 7.10 and Fig. 7.11. These figures indicate that very small budgets are sufficient to recover a large proportion of baseline performance, and that full baseline performance can sometimes be reached far below 100% labelling. In particular, Committee sampling reaches the full baseline at 20% labelling for the 2.0-second task, while Random reaches it at 40% labelling for the 1.0-second task. This confirms that large-scale exhaustive annotation is not necessary under the current multimodal representation.

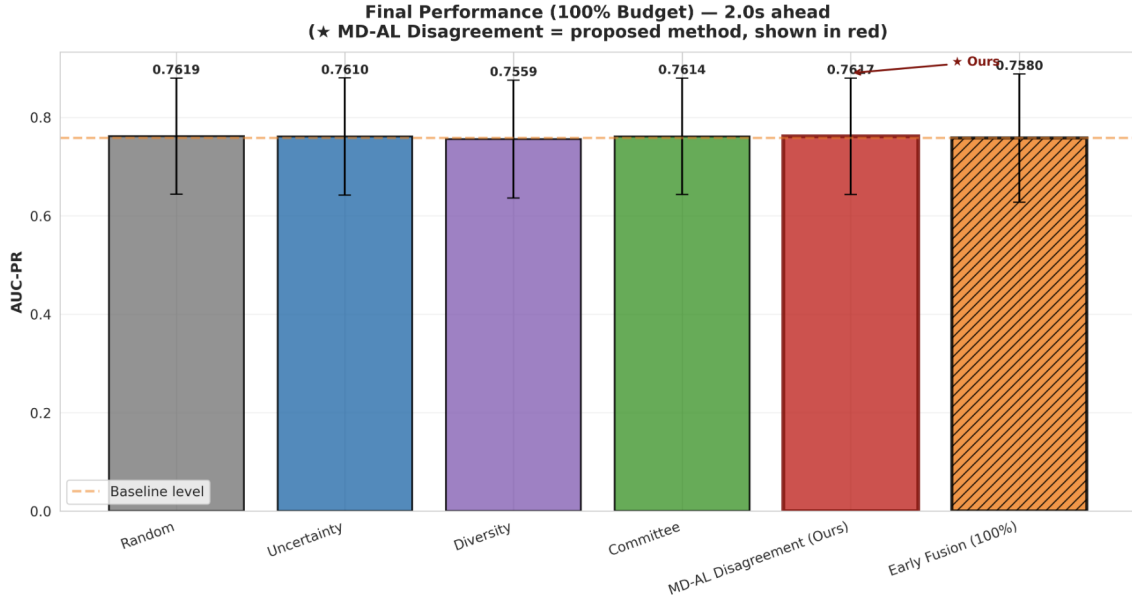


Figure 7.7: Final AUC–PR performance at the 100% labelling budget for the 2.0s prediction horizon. Bars show the mean performance of Random, Uncertainty, Diversity, Committee, and the proposed MD–AL disagreement strategies, together with the fully supervised early-fusion reference model; error bars indicate performance variability. The orange dashed line marks the early-fusion baseline value of 0.7580. Random sampling achieves the highest numerical mean AUC–PR (0.7619), followed closely by MD–AL (0.7617), Committee (0.7614), and Uncertainty (0.7610). Diversity achieves 0.7559, slightly below the reference baseline. The extensive overlap between the error bars and the small differences between the mean values indicate broadly comparable performance; consequently, the numerical differences should not be interpreted as evidence that any strategy is statistically superior.

The numerical comparison at reduced budgets is summarised in Table 7.1. At 20% labelling and 1.0-second prediction, Random reaches an AUC–PR of 0.7574, which is numerically almost identical to the baseline value of 0.7580. At the 2.0-second horizon, Committee achieves 0.7593 at 20% labelling, which is slightly above the baseline mean numerically. These values are best interpreted as evidence that some acquisition strategies can recover the baseline very closely with modest labelling budgets rather than as evidence that active learning systematically improves on full supervision.

7.4.5 Statistical Validation

To test whether the observed differences at reduced budgets are statistically meaningful, Wilcoxon signed-rank tests were performed at the 20% labelling level for the

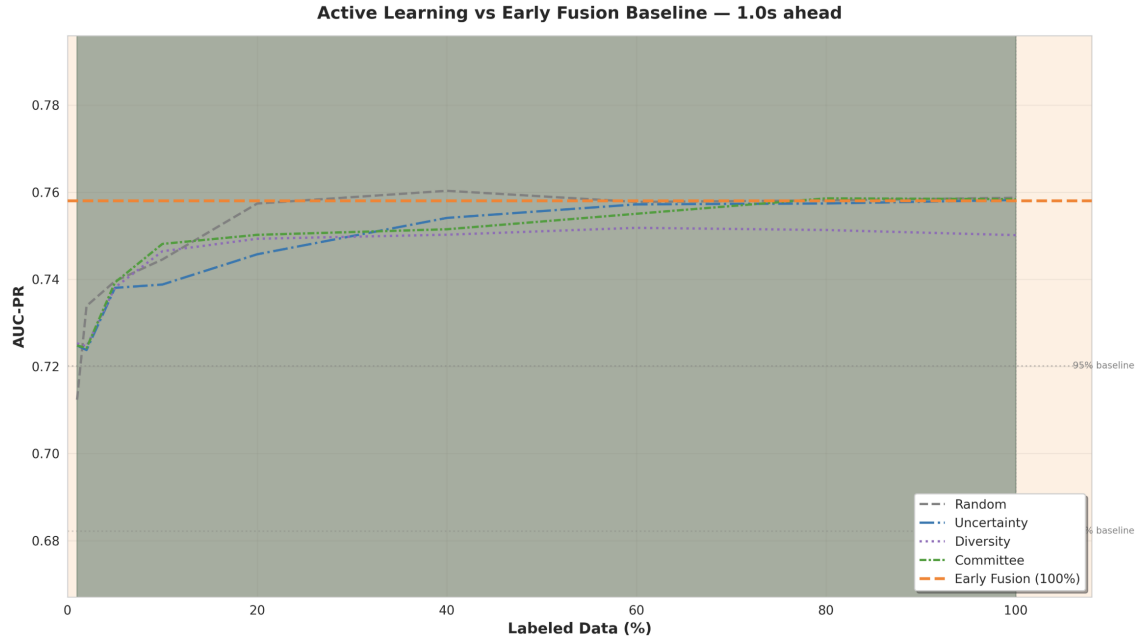


Figure 7.8: AUC-PR as a function of the percentage of labelled training data for the benchmark active-learning strategies at the 1.0s prediction horizon. The horizontal axis represents the labelling budget, and the vertical axis reports AUC-PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline, while the horizontal dotted lines indicate 90% and 95% of baseline performance. All strategies exceed the 95% reference level at very small labelling budgets and improve rapidly up to approximately 10–20% labelled data. Random sampling briefly exceeds the baseline at an intermediate budget, whereas Random, Uncertainty, and Committee sampling converge towards the baseline as the budget increases. Diversity sampling remains slightly below the other strategies at the largest budgets.

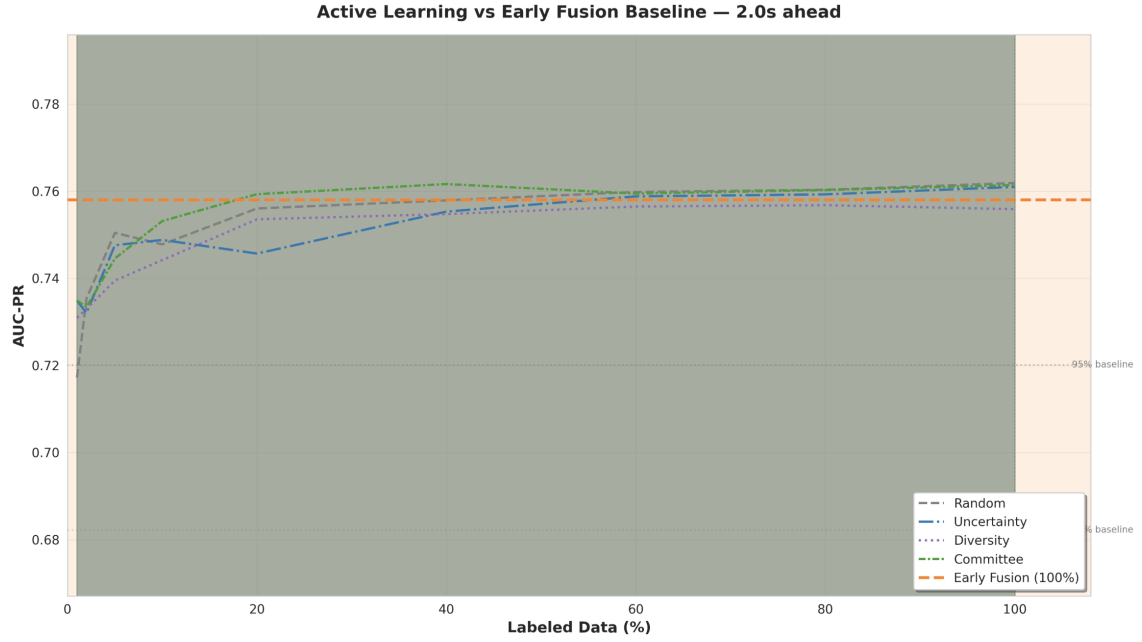


Figure 7.9: AUC-PR as a function of the percentage of labelled training data for the benchmark active-learning strategies at the 2.0s prediction horizon. The horizontal axis represents the labelling budget, and the vertical axis reports AUC-PR, for which higher values indicate better performance. The orange dashed line denotes the fully supervised early-fusion baseline, while the horizontal dotted lines indicate 90% and 95% of baseline performance. All strategies exceed the 95% reference level at very small labelling budgets and improve most strongly within the first 10–20% of labelled data. Committee sampling provides the highest numerical performance at several intermediate budgets and slightly exceeds the baseline from approximately 20% onwards. Random and Uncertainty sampling recover gradually and converge with Committee sampling at the largest budgets, while Diversity remains marginally below the other strategies.

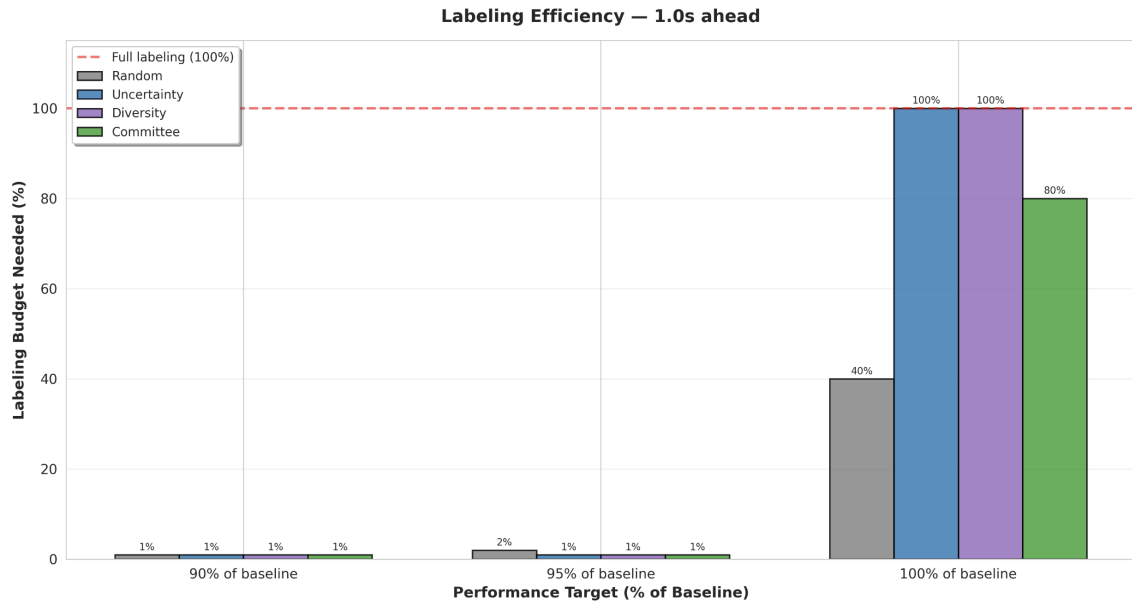


Figure 7.10: Minimum labelling budget required by the benchmark acquisition strategies to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 1.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas Uncertainty, Diversity, and Committee sampling require 1%. Recovering 100% of the baseline is considerably more demanding: Random requires 40%, Committee requires 80%, and Uncertainty and Diversity require the full 100% labelling budget. The red dashed line indicates the full-labelling level.

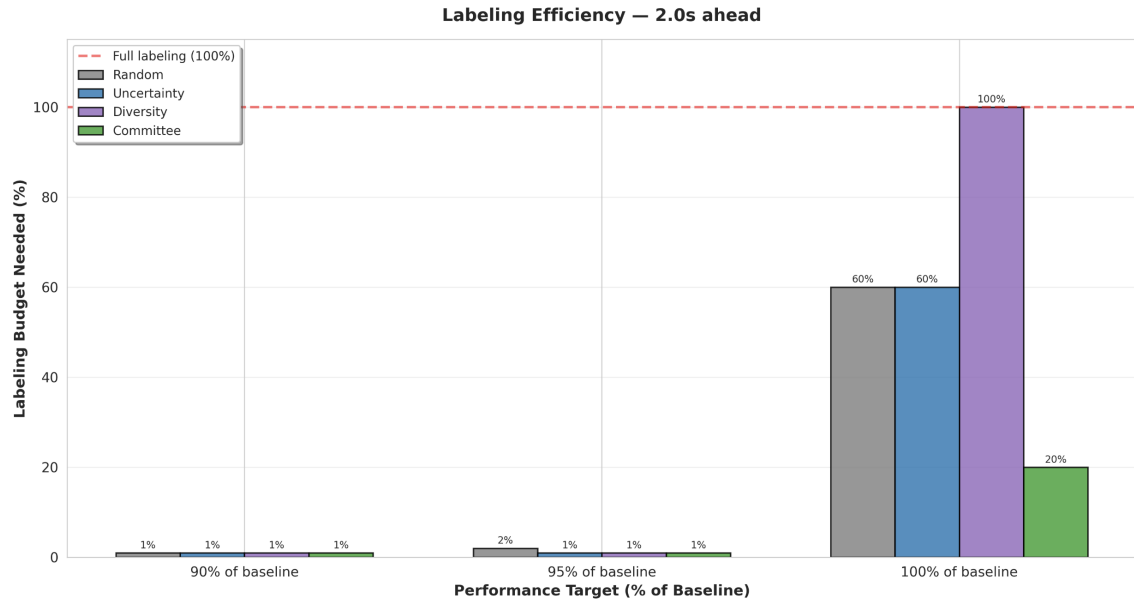


Figure 7.11: Minimum labelling budget required by the benchmark acquisition strategies to reach 90%, 95%, and 100% of the fully supervised early-fusion baseline for the 2.0s prediction horizon. The horizontal axis shows the target proportion of baseline performance, while the vertical axis shows the percentage of labelled data required; therefore, lower bars indicate greater label efficiency. All strategies reach 90% of baseline performance with only 1% labelled data. At the 95% target, Random sampling requires 2%, whereas Uncertainty, Diversity, and Committee sampling require 1%. Recovering 100% of the baseline requires substantially larger and strategy-dependent budgets: Committee sampling requires only 20%, Random and Uncertainty require 60%, and Diversity requires the full 100% labelling budget. The red dashed line indicates the full-labelling level.

Table 7.1: Performance summary at 10% and 20% labelling budgets for both forecasting horizons.

Horizon	Method	Budget	AUC-PR	AUC-ROC	F1	Recall@90%Spec
1.0 s	Early Fusion	100%	0.7580	0.7377	0.6971	0.4226
	Random	10%	0.7446	0.7245	0.6572	0.3988
	Random	20%	0.7574	0.7374	0.6723	0.4228
	Uncertainty	10%	0.7388	0.7149	0.6692	0.3801
	Uncertainty	20%	0.7458	0.7233	0.6752	0.4015
	Diversity	10%	0.7465	0.7247	0.6725	0.3960
	Diversity	20%	0.7493	0.7277	0.6729	0.4024
	Committee	10%	0.7481	0.7283	0.6638	0.3939
	Committee	20%	0.7502	0.7291	0.6723	0.4088
2.0 s	Early Fusion	100%	0.7580	0.7377	0.6971	0.4226
	Random	10%	0.7478	0.7192	0.6615	0.3899
	Random	20%	0.7560	0.7232	0.6636	0.4072
	Uncertainty	10%	0.7488	0.7203	0.6642	0.3773
	Uncertainty	20%	0.7457	0.7116	0.6683	0.3874
	Diversity	10%	0.7442	0.7119	0.6662	0.3871
	Diversity	20%	0.7536	0.7210	0.6719	0.3949
	Committee	10%	0.7532	0.7280	0.6650	0.4026
	Committee	20%	0.7593	0.7351	0.6699	0.3972

Table 7.2: Wilcoxon signed-rank test at 20% labelling for the 1.0s horizon. The alternative hypothesis is that active learning performance is lower than the fully supervised baseline.

Method	AUC-PR	Δ	p	Worse?
Random	0.7574	-0.0006	0.4287	No
Committee	0.7502	-0.0078	0.1638	No
Diversity	0.7493	-0.0087	0.0303	Yes
Uncertainty	0.7458	-0.0123	0.0092	Yes
Baseline	0.7580	—	—	—

1.0-second horizon. The results are shown in Table 7.2. Random and Committee are not significantly worse than the fully supervised baseline, whereas Diversity and Uncertainty are significantly worse under the one-sided alternative.

A Friedman test was also used to compare the ranking of the active learning strategies at matched budgets, and significant differences were found for both horizons. The corresponding results are given in Tables 7.3 and 7.4. These tests confirm that query strategy matters, even if the mean performance values are numerically close.

Table 7.3: Friedman test across active learning strategies at 20% labelling for the 1.0 s horizon.

Statistic	Value
χ^2 (df = 3)	18.0857
p -value	4.2230×10^{-4}
Significant?	Yes

Table 7.4: Friedman test across active learning strategies at 20% labelling for the 2.0 s horizon.

Statistic	Value
χ^2 (df = 3)	13.2643
p -value	4.0986×10^{-3}
Significant?	Yes

These statistical results support the main practical claim of the chapter: substantial reductions in labelling effort are achievable without significant loss in forecasting performance, particularly when Random or Committee sampling is used.

7.4.6 Ablation Study of the MD-AL Query Score

The ablation analysis examines whether the three components of MD-AL each contribute to the final acquisition behaviour. Table 7.5 reports the AUC-PR values for the disagreement-only, disagreement-plus-boundary, and full MD-AL variants at the 1.0-second horizon. Figures 7.12 and 7.13 visualise the same pattern.

The disagreement-only formulation already performs reasonably well, confirming that cross-modal conflict is informative. Adding the boundary-awareness term improves performance consistently, indicating that disagreement becomes more useful when constrained to uncertain decision regions. The full MD-AL formulation then yields the strongest or near-strongest result at all reported budgets. Although the numerical improvements are modest, they are stable, which suggests that the rare-event term contributes to acquisition balance rather than to isolated gains.

The ablation therefore shows that the method is not driven by disagreement alone. Its strength comes from the interaction between multimodal conflict, boundary uncertainty, and class-aware prioritisation.

Table 7.5: Ablation study of the MD-AL query score (AUC-PR, 1.0s horizon).

Method	10%	20%	40%	100%
Disagreement Only	0.7431	0.7473	0.7509	0.7584
Disagreement + Boundary	0.7452	0.7495	0.7502	0.7587
Full MD-AL (Proposed)	0.7478	0.7487	0.7530	0.7590

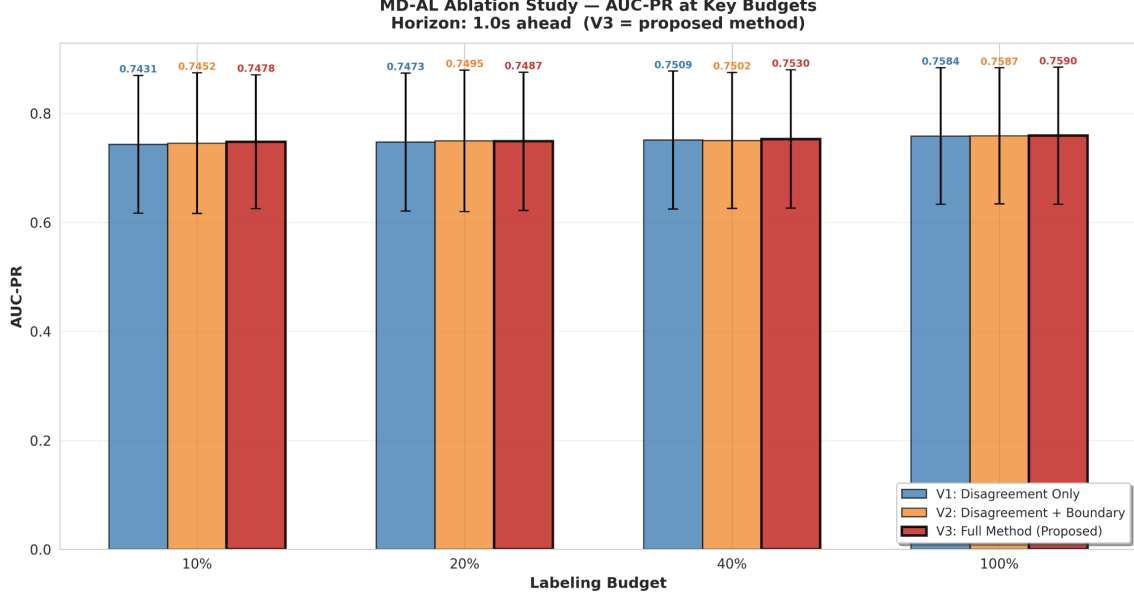


Figure 7.12: Ablation study of the proposed MD-AL acquisition strategy for the 1.0s prediction horizon. Bars report mean AUC-PR and error bars indicate performance variability. V1 uses disagreement only, V2 combines disagreement with the boundary criterion, and V3 is the complete proposed method. V3 achieves the highest numerical mean at the 10%, 40%, and 100% budgets, whereas V2 is marginally higher at 20%. Performance improves gradually for all variants as the labelling budget increases. However, the differences between the variants are small and their error bars overlap substantially, indicating that the additional components provide modest numerical improvements rather than clear evidence of statistical superiority.

7.5 Discussion

The results demonstrate that active learning provides an effective way to reduce annotation requirements in multimodal microsleep forecasting. Across both prediction horizons, the learning curves saturate early, and near-baseline performance is reached with only a small proportion of labelled data. In practical terms, this means that the predictive structure learned by the multimodal early-fusion model can be recovered without exhaustive annotation.

One of the most notable findings is the strong performance of Random sampling.

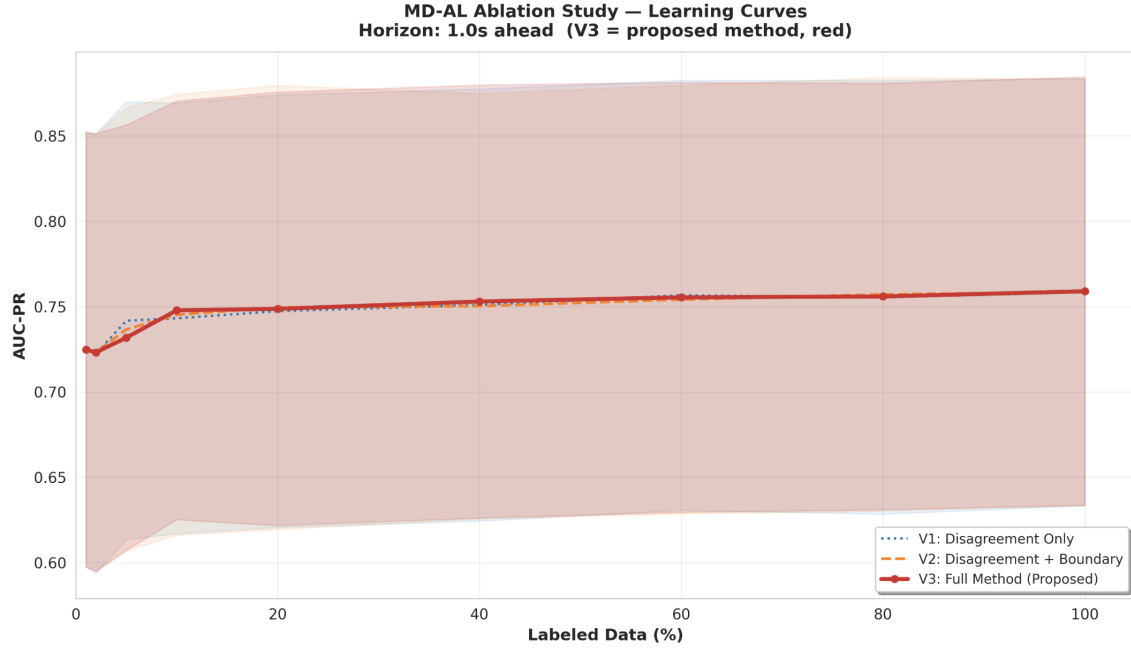


Figure 7.13: Learning curves for the three MD-AL ablation variants across increasing labelling budgets at the 1.0s prediction horizon. The horizontal axis represents the percentage of labelled training data, and the vertical axis reports mean AUC-PR; shaded regions indicate performance variability. V1 uses disagreement only, V2 combines disagreement with the boundary criterion, and V3 is the complete proposed method. All variants improve rapidly at the lowest budgets and then increase more gradually as additional labels are introduced. The three curves remain closely aligned across most budgets, with only small and inconsistent numerical differences between the variants.

This suggests that the multimodal feature representation developed in Chapter 6 produces a well-structured feature space in which informative samples are not confined to a narrow or highly irregular region. As a result, even simple unbiased selection can recover much of the fully supervised decision boundary. This does not make active learning unimportant; rather, it indicates that the value of the representation itself is high.

At the same time, the proposed MD-AL method contributes something different. Its advantage is not that it consistently dominates every benchmark strategy at every budget, but that it remains competitive while offering an interpretable and domain-aware acquisition rule. In biomedical applications, this matters. A query rule that can be explained in terms of cross-modal disagreement, uncertainty, and likely microsleep relevance is more transparent than a purely generic entropy

heuristic. In that sense, MD-AL contributes not only to label efficiency, but also to explainability.

The behaviour across the 1.0-second and 2.0-second forecasting horizons is also informative. The relative stability of the active learning results suggests that extending the short prediction window does not dramatically increase sample complexity under the current feature set. This aligns with the earlier thesis findings that the multimodal representation captures physiologically meaningful pre-microsleep patterns that remain informative across short horizons.

In relation to the preceding chapter, the present results extend the role of multimodality. Chapter 6 showed that EEG-EOG fusion improves predictive performance under full supervision. The current chapter shows that the same multimodal structure also supports efficient learning under limited supervision. Multimodality therefore contributes not only to performance, but also to sample efficiency and practical feasibility.

Some limitations should nevertheless be acknowledged. First, the experiments are conducted on a controlled MWT dataset. Although LOSO validation is strict with respect to subject independence, additional work is needed to determine whether the same acquisition behaviour transfers to more naturalistic or cross-dataset settings. Second, the current study uses a fixed handcrafted feature representation and a tree-based classifier. Active learning behaviour may differ in end-to-end neural models or in systems where the representation itself evolves jointly with acquisition. Third, the proposed MD-AL strategy is a task-specific and interpretable acquisition method designed for paired EEG-EOG microsleep forecasting, rather than a generic active-learning strategy intended to be universally dominant across modalities and application domains. Its value lies in combining strong performance with a clear rationale for why samples are queried.

7.6 Chapter Summary

This chapter addressed the fourth research question by investigating whether multimodal microsleep forecasting can be trained efficiently under limited supervision. Using the early-fusion multimodal model established in Chapter 6 as a fully supervised reference, several active learning strategies were evaluated across progressive labelling budgets at 1.0-second and 2.0-second forecasting horizons.

The results showed that near-baseline performance can be achieved with very small labelled subsets, and that substantial reductions in annotation effort are possible without statistically significant loss in forecasting performance. Random and Committee sampling proved especially effective in this regard. The chapter also introduced MD-AL, an interpretable acquisition strategy that combines multimodal disagreement, boundary-awareness, and rare-event bias. While its performance was broadly comparable to the strongest standard strategies rather than uniformly superior, it provided a physiologically meaningful rationale for sample selection.

Taken together with the preceding chapters, the thesis has now progressed from identifying informative features, to modelling temporal structure, to integrating modalities, and finally to reducing the labelling burden required to train the resulting system. The following chapter synthesises these findings and presents the overall conclusions of the thesis.

Chapter 8

Cross-Study Synthesis and Design Insights

8.1 Introduction

This thesis investigated microsleep prediction through four connected research questions, each addressing a different layer of the overall problem. Chapter 4 examined which EEG-derived features are most informative for microsleep detection. Chapter 5 then studied how temporal modelling can improve forecasting across short prediction horizons. Chapter 6 extended the problem to multimodal learning by integrating EEG and EOG within an information-theoretic fusion framework. Finally, Chapter 7 addressed annotation efficiency through active learning, with the aim of reducing labelling effort while maintaining predictive performance.

Taken individually, each chapter answered a specific methodological question. Taken together, they form a progression from representation, to temporal structure, to multimodal integration, and finally to training efficiency. The purpose of this chapter is to synthesise those findings across the four research questions, identify the main themes that emerge from the thesis as a whole, and discuss their implications for microsleep prediction as both a scientific and applied problem.

The central argument developed across the thesis is that reliable microsleep pre-

diction does not depend on a single modelling decision. Rather, it depends on the interaction of several factors: the quality of the underlying physiological features, the preservation of temporal information, the integration of complementary sensing modalities, and the efficient use of labelled data. This chapter therefore moves beyond chapter-specific results and considers how the findings jointly contribute to the broader objective of improving microsleep forecasting in a way that is interpretable, robust, and practically feasible.

8.2 Restatement of the Thesis Aim

The overall aim of this thesis was to improve the prediction of microsleep events from physiological data, with particular emphasis on short-horizon forecasting under realistic constraints. This required addressing four related challenges. The first was determining which signal features are informative for identifying microsleep-relevant states. The second was understanding how predictive performance changes when temporal structure is modelled explicitly rather than treated as static classification. The third was evaluating whether multimodal fusion can capture aspects of the microsleep transition that are not available from EEG alone. The fourth was determining whether strong multimodal forecasting models can be trained efficiently when labels are limited and expensive to obtain.

The thesis therefore addresses microsleep prediction not only as a machine learning problem, but as a multi-stage modelling problem in which feature representation, temporal dependence, multimodal coupling, and annotation cost all affect final performance.

8.3 Synthesis of Findings Across the Research Questions

8.3.1 RQ1: Feature Representation as the Foundation of Prediction

The first research question established that microsleep detection depends strongly on the quality and physiological relevance of the extracted EEG features. The results in Chapter 4 showed that spectral and nonlinear features provide the most consistent discriminative value, particularly features associated with slowing of vigilance-related activity and reduced signal complexity. This finding was important for two reasons.

First, it demonstrated that the microsleep problem is not adequately represented by arbitrary generic EEG descriptors. Features must reflect the known physiological correlates of vigilance decline. Second, the RQ1 findings provided the representational basis for the later chapters. The temporal models examined in RQ2, the multimodal fusion strategies examined in RQ3, and the active learning study in RQ4 all depend on the availability of a stable and informative feature space.

A key implication of RQ1 is that predictive performance in later chapters should not be interpreted only as an effect of model architecture. The success of later methods is partly enabled by the fact that the underlying feature representation already captures physiologically meaningful variation.

8.3.2 RQ2: Temporal Modelling Improves Forecasting Beyond Static Detection

The second research question addressed the limitations of static or near-static prediction pipelines by explicitly modelling temporal structure. Chapter 5 showed that forecasting performance degrades as the prediction horizon extends away from the onset of the event, which is consistent with earlier literature. However, it also showed

that this degradation can be reduced through temporal modelling, particularly when causal attention is used before recurrent processing.

The main contribution of RQ2 was therefore not simply that one temporal model outperformed another, but that preserving temporally local information is central to short-horizon microsleep forecasting. The results suggested that a substantial part of the forecasting difficulty lies in the loss of fine-grained temporal cues when the model aggregates information too coarsely. The causal attention mechanism improved sensitivity and discrimination across horizons, supporting the interpretation that forecasting depends on recognising transitional changes rather than only endpoint states.

This chapter also provided an important bridge in the thesis. RQ1 established what to represent, while RQ2 addressed how those representations should be organised over time. The temporal findings therefore shifted the thesis from feature quality alone toward sequence-sensitive modelling.

8.3.3 RQ3: Multimodal Fusion Adds Complementary Physiological Information

The third research question examined whether EEG-only forecasting could be improved by integrating EOG in a principled multimodal framework. Chapter 6 showed that EEG and EOG contribute complementary information. EEG alone was highly specific but comparatively insensitive, whereas EOG alone was more sensitive but less specific. When the modalities were fused, especially through cross-modal mutual information features, overall predictive performance improved and the balance between sensitivity and specificity became stronger.

The most important outcome of RQ3 was that multimodal gain was not obtained merely by concatenating two sources of data. Rather, the strongest results arose when the interaction between modalities was modelled explicitly using mutual information. This suggests that the relevant signal is not only the presence of multiple modalities, but the nonlinear synchrony between them during the transition into

microsleep.

Equally important was the negative result concerning transfer entropy. The lack of additional improvement from directional interaction features implies that, at the time scales considered here, the microsleep transition is better characterised as bidirectional synchrony than as a stable directional process. This is a useful scientific result because it narrows the type of multimodal dependency that appears to matter most for short-horizon prediction.

8.3.4 RQ4: Label Efficiency Can Be Improved Without Sacrificing Performance

The fourth research question shifted attention from predictive modelling to training efficiency. Chapter 7 demonstrated that active learning can substantially reduce annotation requirements while retaining near-baseline predictive performance. Across both forecast horizons, approximately 90% to 95% of fully supervised performance was recoverable with a very small fraction of labeled data, and in several conditions around 80% of the annotation burden could be removed without statistically significant loss in the main evaluation metric.

This result is practically important because it addresses one of the main barriers to deployment and scaling. Even when a strong multimodal predictor exists, the system remains difficult to develop if it requires exhaustive expert labelling. The active learning experiments showed that this requirement can be relaxed considerably.

A second outcome of RQ4 was methodological. The proposed multimodal disagreement active learning strategy did not dominate every benchmark at every budget, but it remained competitive while offering a clear and interpretable acquisition rule. This reinforces one of the broader themes of the thesis: in biomedical signal analysis, methods should not be judged only by whether they produce the single highest metric value, but also by whether their behaviour is transparent and physiologically meaningful.

8.4 Cross-Cutting Themes

8.4.1 Microsleep Prediction is a Progressive Modelling Problem

One of the clearest findings of the thesis is that microsleep prediction should not be treated as a single-stage classification problem. Performance depends on a progression of modelling decisions. Informative physiological features are required first. Those features must then be preserved across time rather than collapsed prematurely. Complementary modalities can then be integrated to improve robustness and capture more of the underlying transition. Finally, the resulting model must be trainable under realistic annotation constraints.

This progression is reflected in the structure of the thesis itself. Each research question contributes a necessary part of the overall system. None of the later improvements would be as meaningful without the earlier stages. In this sense, the thesis is cumulative rather than a collection of unrelated experiments.

8.4.2 Interpretability Matters Alongside Accuracy

A recurring theme across the thesis is the importance of interpretability. In RQ1, interpretability appears through the use of physiologically meaningful EEG features. In RQ2, it appears through the examination of temporally relevant cues and the attention-based emphasis on specific features over time. In RQ3, it appears through information-theoretic coupling measures that provide an interpretable account of neural-ocular synchrony. In RQ4, it appears through the design of an acquisition score whose components correspond to modality disagreement, uncertainty, and minority-event bias.

This matters because microsleep prediction sits at the boundary between machine learning and physiological monitoring. In such settings, predictive performance alone is not sufficient. Domain experts must also be able to understand what aspect of the signal contributes to the prediction or to the sample-selection process. The

thesis therefore contributes to the problem not only by improving metrics, but also by maintaining a modelling perspective that remains explainable.

8.4.3 Multimodality Improves Both Prediction and Data Efficiency

Another cross-cutting theme is that multimodality provides benefits beyond raw predictive accuracy. Chapter 6 showed that EEG and EOG fusion improves subject-independent prediction. Chapter 7 then showed that the same multimodal structure can also support efficient learning under limited supervision. This is important because it suggests that multimodal design has two separate benefits: it enriches the representation of the physiological state, and it improves the efficiency with which informative samples can be identified. This dual role of multimodality is one of the central contributions of the thesis as a whole.

8.4.4 The Main Constraint is Not Only Model Capacity

The results across the thesis suggest that microsleep prediction is constrained by more than the expressiveness of the classifier. In RQ1, the limitation was representational quality. In RQ2, it was preservation of temporal resolution. In RQ3, it was the ability to model cross-modal dependence rather than only independent feature sets. In RQ4, it was the availability of labels. These observations imply that simply increasing model complexity is unlikely to solve the entire problem. Improvement instead comes from addressing the right bottleneck at each stage.

8.5 Overall Contributions of the Thesis

The contributions of the thesis can be summarised at four levels.

At the representational level, the thesis provided a systematic analysis of EEG features for microsleep detection and established that spectral and nonlinear descriptors form a strong basis for downstream modelling.

At the temporal level, it showed that forecasting degradation across prediction horizons can be reduced through causal attention and recurrent modelling, thereby improving short-horizon microsleep prediction beyond static detection pipelines.

At the multimodal level, it established that EEG–EOG fusion improves subject-independent prediction when cross-modal dependencies are modelled explicitly through mutual information, and that the relevant interaction is better characterised by synchrony than by directional coupling.

At the training-efficiency level, it showed that strong multimodal forecasting performance can be retained with substantially fewer labels, and introduced an interpretable multimodal disagreement strategy for active learning.

Taken together, these contributions form a coherent framework for microsleep prediction that moves from signal representation to practical training considerations.

8.6 Concluding Remarks

This thesis set out to improve microsleep prediction by addressing four linked questions concerning feature representation, temporal modelling, multimodal integration, and annotation efficiency. Across these questions, the work showed that strong predictive performance arises not from any one component in isolation, but from the coordinated treatment of physiological representation, temporal structure, modality interaction, and data efficiency.

The main conclusion is that microsleep forecasting can be made more accurate, more informative, and more practical when these dimensions are addressed together. EEG features with physiological relevance provide the representational basis. Temporal modelling preserves the transitional structure needed for forecasting. EEG–EOG fusion captures complementary aspects of the pre-microsleep state. Active learning then reduces the cost of building and maintaining the resulting system.

In this sense, the thesis contributes both methodologically and practically. It provides a structured account of how microsleep prediction can be improved step by step, and it offers evidence that such improvement can be achieved without

losing interpretability or requiring exhaustive annotation. These findings support the broader view that microsleep prediction should be approached as an integrated modelling problem rather than as a single isolated classification task.

The findings synthesised in this chapter provide an integrated perspective on microsleep forecasting across the four research questions. The principal limitations of the thesis and directions for future research are discussed in Chapter 9.

Chapter 9

Conclusions and Future Work

This thesis investigated the problem of microsleep prediction using multimodal physiological signals, with the aim of improving prediction performance, physiological interpretability, and practical applicability under realistic conditions. Using EEG and EOG recordings, the research explored microsleep prediction as a progressive modelling problem involving physiological feature representation, temporal forecasting, multimodal interaction analysis, and label-efficient learning. Across four research questions, the thesis examined how complementary physiological information can be leveraged to improve forecasting performance while maintaining subject-independent generalisation and reducing annotation requirements.

9.1 Thesis-Wide Contributions

The principal contributions of this thesis are summarised as follows:

1. Physiologically informative EEG feature representations were systematically investigated for microsleep prediction. Spectral, nonlinear, and complexity-based features were shown to capture meaningful indicators associated with vigilance decline and microsleep transition.
2. Temporal modelling approaches for short-horizon microsleep forecasting were investigated using sequence-sensitive learning frameworks. The findings demon-

strated that preserving temporal structure substantially improves forecasting performance compared with static classification approaches.

3. A multimodal EEG–EOG framework for microsleep prediction was developed to investigate the complementary relationship between neural and ocular physiological activity. The results demonstrated that multimodal integration consistently outperformed unimodal approaches.
4. Transfer entropy analysis was investigated as a mechanism for examining the temporal directionality of microsleep-related physiological interactions. This analysis provided insight into whether microsleep-associated changes emerge first in cortical EEG activity, ocular EOG activity, or synchronously across both modalities.
5. Subject-independent microsleep prediction was systematically evaluated under Leave-One-Subject-Out (LOSO) validation using a broad range of machine learning and deep learning architectures.
6. A multimodal disagreement-based active learning framework was proposed for label-efficient microsleep prediction. The proposed approach utilised disagreement between EEG- and EOG-based predictions to identify informative samples for annotation, substantially reducing labelling requirements while preserving competitive prediction performance.

9.2 Cross-RQ Synthesis

Across the four research questions, several consistent themes emerged. First, physiological representations derived from EEG and EOG signals contain complementary information relevant to microsleep prediction. While EEG captures cortical dynamics associated with vigilance regulation, EOG reflects fatigue-related ocular behaviour. Combining both modalities consistently improved predictive performance compared with unimodal approaches.

Second, temporal context was shown to play a critical role in forecasting microsleep events. Models capable of preserving temporal structure achieved stronger performance than approaches based on isolated physiological segments, highlighting the progressive nature of vigilance decline.

Third, subject-independent evaluation demonstrated that robust microsleep prediction remains challenging due to substantial inter-subject variability. Nevertheless, multimodal approaches consistently improved generalisation performance, suggesting that complementary physiological information can partially mitigate subject-specific differences.

Finally, the active learning investigation demonstrated that effective microsleep prediction can be achieved with substantially reduced annotation effort. This finding is particularly important for real-world deployment, where expert physiological annotation remains expensive and time-consuming.

Taken together, the findings indicate that accurate and practical microsleep prediction systems benefit from the integration of multimodal physiological information, temporal modelling, subject-independent evaluation, and label-efficient learning strategies.

9.3 Future Work

Several future research directions emerge from this work.

9.3.1 Larger and More Diverse Multimodal Datasets

Future work should investigate larger and more diverse multimodal datasets incorporating broader demographic variation, different operational environments, and wider physiological variability. Expanding dataset diversity will improve understanding of inter-subject variability and strengthen the robustness of subject-independent microsleep prediction systems across applications such as transportation, industrial manufacturing, healthcare monitoring, and other vigilance-sensitive domains.

9.3.2 Dynamic Physiological Interaction Modelling

Although this thesis investigated multimodal interaction using transfer entropy analysis, future work should explore more advanced dynamic interaction representations capable of capturing time-varying physiological dependencies preceding microsleep onset. In particular, Cross-Temporal Coupling Maps (CTCMs) [186] provide a two-dimensional representation of the coupling between signals at different time points or temporal lags. Each map indicates how strongly activity in one modality at a given time is associated with activity in another modality at an earlier or later time. Applied to EEG and EOG, CTCMs could reveal when neural–ocular coupling emerges before microsleep onset and whether ocular changes systematically precede or follow cortical changes. This would provide a more detailed representation of evolving multimodal interactions than a single summary coupling value. Such approaches may provide deeper insight into evolving multimodal physiological interactions associated with vigilance decline.

9.3.3 Advanced Multimodal Learning

Future work may further investigate advanced multimodal learning frameworks through the integration of additional physiological modalities beyond EEG and EOG, including electrocardiography (ECG), respiration signals, and electromyography (EMG). Incorporating complementary physiological measurements may provide a more comprehensive representation of fatigue-related processes and improve robustness under noisy and heterogeneous conditions. In parallel, advanced multimodal learning approaches such as cross-attention fusion mechanisms, graph-based physiological modelling, and self-supervised multimodal representation learning may further enhance the modelling of complex inter-modal relationships while preserving physiological interpretability.

9.3.4 Real-Time and Embedded Deployment

Practical deployment requires models that balance predictive performance with computational efficiency. Future work should therefore investigate lightweight architectures, model compression strategies, and edge-deployment optimisation suitable for real-time physiological monitoring systems operating in embedded and mobile environments.

9.4 Closing Statement

This thesis investigated microsleep prediction through the integration of physiological representation learning, temporal modelling, multimodal EEG–EOG interaction analysis, and label-efficient learning. The findings demonstrate that combining neural and ocular physiological information provides meaningful advantages for forecasting microsleep-related vigilance decline and that preserving temporal structure is important for modelling short-horizon physiological transitions.

Beyond prediction performance alone, this work contributes to improved understanding of the interaction between cortical and ocular dynamics preceding microsleep onset. The thesis also demonstrates the potential of multimodal active learning strategies for reducing annotation requirements in physiological monitoring systems.

Overall, the methodologies and findings presented in this thesis contribute toward the development of more reliable, interpretable, and practically deployable microsleep prediction systems suitable for a broad range of vigilance-monitoring applications.

Bibliography

- [1] T. B. Young, “Epidemiology of daytime sleepiness: Definitions, symptomatology, and prevalence,” *Journal of Clinical Psychiatry*, vol. 65, no. Suppl 16, pp. 12–16, 2004.
- [2] J. M. Lyznicki, T. C. Doege, R. M. Davis, and M. A. Williams, “Sleepiness, driving, and motor vehicle crashes,” *JAMA*, vol. 279, no. 23, pp. 1908–1913, 1998. DOI: 10.1001/jama.279.23.1908.
- [3] A. M. Williamson, A.-M. Feyer, R. P. Mattick, R. Friswell, and S. Finlay-Brown, “Developing measures of fatigue using an alcohol comparison to validate the effects of fatigue on performance,” *Accident Analysis & Prevention*, vol. 33, no. 3, pp. 313–326, 2001. DOI: 10.1016/S0001-4575(00)00046-2.
- [4] S. K. Lal and A. Craig, “A critical review of the psychophysiology of driver fatigue,” *Biological Psychology*, vol. 55, no. 3, pp. 173–194, 2001. DOI: 10.1016/S0301-0511(00)00085-5.
- [5] G. Hidalgo-Gadea, A. Kreuder, J. Krajewski, and C. Vorstius, “Towards better microsleep predictions in fatigued drivers: Exploring benefits of personality traits and iq,” *Ergonomics*, vol. 64, no. 6, pp. 778–792, 2021. DOI: 10.1080/00140139.2021.1882707.
- [6] J. Skorucak, A. Hertig-Godeschalk, P. Achermann, J. Mathis, and D. Schreier, “Automatically detected microsleep episodes in the fitness-to-drive assessment,” *Frontiers in Neuroscience*, vol. 14, 2020. DOI: 10.3389/fnins.2020.00008.

- [7] S. Sanei and J. A. Chambers, *EEG Signal Processing and Machine Learning*, 2nd ed. Wiley, 2021, As cited in manuscript list.
- [8] B. S. Oken, M. C. Salinsky, and S.-M. Elsas, “Vigilance, alertness, or sustained attention: Physiological basis and measurement,” *Clinical Neurophysiology*, vol. 117, no. 9, pp. 1885–1901, 2006. DOI: 10.1016/j.clinph.2006.01.017.
- [9] M. R. Peiris, R. D. Jones, P. R. Davidson, and P. J. Bones, “Detecting behavioral microsleeps from eeg power spectra,” in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2006, pp. 5723–5726. DOI: 10.1109/IEMBS.2006.260411.
- [10] M. Holub, M. Šrutová, and L. Lhotská, “Microsleep classifier using eeg channel recording: A feasibility study,” in *Information Technologies in Biomedicine*, ser. Advances in Intelligent Systems and Computing, vol. 471, Cham: Springer, 2015, pp. 107–118. DOI: 10.1007/978-3-319-19387-8_10.
- [11] G. R. Poudel, R. D. Jones, C. R. H. Innes, R. Watts, and T. L. Signal, “Fmri correlates of behavioral microsleeps during a continuous visuomotor task,” in *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, IEEE, 2009, pp. 1325–1328. DOI: 10.1109/IEMBS.2009.5333774.
- [12] C. Wang, B. Guragain, A. Verma, L. Archer, S. Majumder, and J. Song, “Spectral analysis of eeg during microsleep events annotated via driver monitoring system to characterize drowsiness,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 2, pp. 1346–1356, 2020. DOI: 10.1109/TAES.2019.2933960.
- [13] R. Shoorangiz, A. B. Buriro, S. J. Weddell, and R. D. Jones, “Detection and prediction of microsleeps from eeg using spatio-temporal patterns,” in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 522–525. DOI: 10.1109/EMBC.2019.8857962.

- [14] P. Arpaia et al., “A systematic review of techniques for artifact detection and artifact category identification in electroencephalography from wearable devices,” *Sensors*, vol. 25, no. 18, p. 5770, 2025. DOI: 10.3390/s25185770.
- [15] X. Chen, X. Xu, A. Liu, et al., “Removal of muscle artifacts from the eeg: A review and recommendations,” *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5353–5368, 2019. DOI: 10.1109/JSEN.2019.2906572.
- [16] S. Mahmud, M. Hossain, M. Chowdhury, and M. Reaz, “Mlmrs-net: Electroencephalography (eeg) motion artifacts removal using a multi-layer multi-resolution spatially pooled 1d signal reconstruction network,” *Neural Computing and Applications*, vol. 35, pp. 8371–8388, 2022. DOI: 10.1007/s00521-022-08111-6.
- [17] L. Xing and A. Casson, “Deep autoencoder for real-time single-channel eeg cleaning and its smartphone implementation using tensorflow lite with hardware/software acceleration,” *IEEE Transactions on Biomedical Engineering*, vol. 71, no. 11, pp. 3111–3122, 2024. DOI: 10.1109/TBME.2024.3408331.
- [18] B. Gu, Z. Zhai, C. Deng, and H. Huang, “Efficient active learning by querying discriminative and representative samples and fully exploiting unlabeled data,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 9, pp. 4111–4122, 2020. DOI: 10.1109/TNNLS.2020.3016928.
- [19] P. Alvaro et al., “Driver education: Enhancing knowledge of sleep, fatigue and risky behaviour to improve decision making in young drivers,” *Accident Analysis & Prevention*, vol. 112, pp. 77–83, 2018. DOI: 10.1016/j.aap.2017.12.017.
- [20] C. Anderson et al., “Feeling sleepy? stop driving—awareness of fall-asleep crashes,” *Sleep*, vol. 46, no. 9, 2023. DOI: 10.1093/sleep/zsad136.
- [21] R. Minhas, S. Arbatli, Y. Çelik, Y. Peker, and B. Semiz, “A novel approach to quantify microsleep in drivers with obstructive sleep apnea by concurrent analysis of eeg patterns and driving attributes,” *IEEE Journal of Biomedical*

- and Health Informatics*, vol. 28, no. 3, pp. 1341–1352, 2024. DOI: 10.1109/JBHI.2024.3352081.
- [22] A. Hertig-Godeschalk, J. Skorucak, A. Malafeev, P. Achermann, J. Mathis, and D. R. Schreier, “Microsleep episodes in the borderland between wakefulness and sleep,” *Sleep*, vol. 43, no. 1, 2020. DOI: 10.1093/sleep/zsz163.
- [23] S. Aprilia and D. Fitriana, “Drowsiness detection in low-light conditions using yolo and clahe augmentation for real-time applications,” in *2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, 2024, pp. 742–746. DOI: 10.1109/ISRITI64779.2024.10963648.
- [24] A. Malafeev, A. Hertig-Godeschalk, D. R. Schreier, J. Skorucak, J. Mathis, and P. Achermann, “Automatic detection of microsleep episodes with deep learning,” *Frontiers in Neuroscience*, vol. 15, 2021. DOI: 10.3389/fnins.2021.564098.
- [25] A. Balaji, U. Tripathi, V. Chamola, A. Benslimane, and M. Guizani, “Toward safer vehicular transit: Implementing deep learning on single channel eeg systems for microsleep detection,” *IEEE Transactions on Intelligent Transportation Systems*, 2021, Volume/pages listed as PP in manuscript. DOI: 10.1109/TITS.2021.3125126.
- [26] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, “Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces,” in *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2020, pp. 2958–2965. DOI: 10.1109/SMC42975.2020.9283028.
- [27] L. Alzubaidi et al., “Review of deep learning: Concepts, cnn architectures, challenges, applications, future directions,” *Journal of Big Data*, vol. 8, no. 53, 2021. DOI: 10.1186/s40537-021-00444-8.

- [28] S. Dubey, V. Samadhia, S. Chauhan, N. Deshmukh, and D. Sati, “Micro-sleep accident prevention by smart vehicle using ai and image processing,” in *2023 3rd Asian Conference on Innovation in Technology (ASIANCON)*, 2023, pp. 1–6. DOI: 10.1109/ASIANCON58793.2023.10270583.
- [29] S. Cao, P. Feng, W. Kang, Z. Chen, and B. Wang, “Optimized driver fatigue detection method using multimodal neural networks,” *Scientific Reports*, vol. 15, 2025. DOI: 10.1038/s41598-025-86709-1.
- [30] J. Shah, A. Chougule, V. Chamola, and A. Hussain, “Novel welch-transform based enhanced spectro-temporal analysis for cognitive microsleep detection using a single electrode eeg,” *Neurocomputing*, vol. 549, p. 126 387, 2023. DOI: 10.1016/j.neucom.2023.126387.
- [31] M. Hasan, M. Islam, N. Sulaiman, M. Hossain, and J. Mendes, “Real-time eeg signal analysis for microsleep detection: Hyper-opt-ann as a key solution,” *IEEE Access*, vol. 13, pp. 66 354–66 372, 2025. DOI: 10.1109/ACCESS.2025.3559619.
- [32] M. Hasan, S. Khandaker, N. Sulaiman, M. Hossain, and A. Islam, “Addressing imbalanced eeg data for improved microsleep detection: An adasyn, fft and lda-based approach,” *Diyala Journal of Engineering Sciences*, 2024. DOI: 10.24237/djes.2024.17304.
- [33] C. M. Michel and D. Brunet, “Eeg source imaging: A practical review of the analysis steps,” *Frontiers in Neurology*, vol. 10, 2019. DOI: 10.3389/fneur.2019.00325.
- [34] D. Schreier, “Microsleep episodes and wakefulness: The daytime job of sleep-wake medicine,” *Sleep Medicine Reviews*, vol. 82, p. 102 120, 2025. DOI: 10.1016/j.smrv.2025.102120.
- [35] M. E. Howard, M. L. Jackson, G. A. Kennedy, P. Swann, M. Barnes, and R. J. Pierce, “The interactive effects of extended wakefulness and low-dose

- alcohol on simulated driving and vigilance,” *Sleep*, vol. 30, no. 10, pp. 1334–1340, 2007. DOI: 10.1093/sleep/30.10.1334.
- [36] G. Poudel, C. Innes, and R. Jones, “Temporal evolution of neural activity and connectivity during microsleeps when rested and following sleep restriction,” *NeuroImage*, vol. 174, pp. 263–273, 2018. DOI: 10.1016/j.neuroimage.2018.03.031.
- [37] J. Skorucak, A. Hertig-Godeschalk, D. Schreier, A. Malafeev, J. Mathis, and P. Achermann, “Automatic detection of microsleep episodes with feature-based machine learning,” *Sleep*, vol. 43, no. 2, 2020. DOI: 10.1093/sleep/zsz225.
- [38] M. H. Zaky, R. Shoorangiz, G. R. Poudel, L. Yang, and R. D. Jones, “Investigating the neural signature of microsleeps using eeg,” in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2021, pp. 6293–6296. DOI: 10.1109/EMBC46164.2021.9630401.
- [39] Y. Jonmohamadi, G. Poudel, C. Innes, and R. Jones, “Microsleeps are associated with stage-2 sleep spindles from hippocampal-temporal network,” *International Journal of Neural Systems*, vol. 26, no. 4, p. 1650015, 2016. DOI: 10.1142/S0129065716500155.
- [40] M. Zaky, R. Shoorangiz, G. Poudel, L. Yang, C. Innes, and R. Jones, “Increased cerebral activity during microsleeps reflects an unconscious drive to re-establish consciousness,” *International Journal of Psychophysiology*, vol. 190, pp. 1–9, 2023. DOI: 10.1016/j.ijpsycho.2023.05.349.
- [41] E. Beldzik, Z. Yang, S. Williams, and L. Lewis, “Distinct spectral pattern of cognitive, drowsiness, and fatigue-related theta/alpha eeg activity during wakefulness,” *Sleep*, 2024. DOI: 10.1093/sleep/zsae067.0182.
- [42] Z. Liu, Y. Zhou, and C. Hao, “Alteration in neural oscillatory activity and phase-amplitude coupling after sleep deprivation: Evidence for impairment

- and compensation effects,” *Journal of Sleep Research*, vol. 34, 2024. DOI: 10.1111/jsr.14264.
- [43] S. Snipes, E. Krugliakova, E. Meier, and R. Huber, “The theta paradox: 4–8 hz eeg oscillations reflect both sleep pressure and cognitive control,” *Journal of Neuroscience*, vol. 42, no. 45, pp. 8569–8586, 2022. DOI: 10.1523/JNEUROSCI.1063-22.2022.
- [44] A. Baseer, S. J. Weddell, and R. D. Jones, “Prediction of microsleeps using pairwise joint entropy and mutual information between eeg channels,” in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2017, pp. 4495–4498. DOI: 10.1109/EMBC.2017.8037855.
- [45] R. Ranjan, B. C. Sahana, and A. K. Bhandari, “Ocular artifact elimination from electroencephalography signals: A systematic review,” *Biocybernetics and Biomedical Engineering*, vol. 41, no. 3, pp. 960–996, 2021.
- [46] M. A. Ngasri, I. S. B. Isa, S. N. B. Sulaiman, and Z. H. C. Soh, “Automated stand-alone video-based microsleep detection system by using ear technique,” *2019 9th IEEE International Conference on Control System, Computing and Engineering (ICCSCE)*, pp. 78–83, 2019. [Online]. Available: <https://api.semanticscholar.org/CorpusID:215815659>.
- [47] J. Fan, C. Sun, M. Long, C. Chen, and W. Chen, “Eognet: A novel deep learning model for sleep stage classification based on single-channel eog signal,” *Frontiers in Neuroscience*, vol. 15, 2021. DOI: 10.3389/fnins.2021.573194.
- [48] H. Van Gorp, M. Van Gilst, S. Overeem, et al., “Single-channel eog sleep staging on a heterogeneous cohort of subjects with sleep disorders,” *Physiological Measurement*, vol. 45, 2024. DOI: 10.1088/1361-6579/ad4251.
- [49] S. Maiti, S. Sharma, and R. Bapi, “Enhancing healthcare with eog: A novel approach to sleep stage classification,” in *ICASSP 2024 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2024, pp. 2305–2309. DOI: 10.1109/ICASSP48485.2024.10446703.

- [50] N. Karuppusamy and B.-H. Kang, “Multimodal system to detect driver fatigue using eeg, gyroscope, and image processing,” *IEEE Access*, vol. 8, pp. 129 645–129 667, 2020. DOI: 10.1109/ACCESS.2020.3009226.
- [51] Z. Lian, T. Xu, Z. Yuan, J. Li, N. Thakor, and H. Wang, “Driving fatigue detection based on hybrid electroencephalography and eye tracking,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 11, pp. 6568–6580, 2024. DOI: 10.1109/JBHI.2024.3446952.
- [52] S. Shanthini and S. Subhashini, “Multimodal fusion for robust driver drowsiness detection: Integrating eeg and video signals with enhanced accuracy,” *Proceedings of ACCAI*, pp. 1–4, 2024. DOI: 10.1109/ACCAI61061.2024.10601871.
- [53] D. Lu, H. Huang, Z. Yi, and J. Pan, “Confidence-aware multimodal network for driver drowsiness detection: Leveraging eeg-eog signals enhanced by knowledge distillation,” *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–13, 2025. DOI: 10.1109/TIM.2025.3591857.
- [54] Y. Nir and L. De Lecea, “Sleep and vigilance states: Embracing spatiotemporal dynamics,” *Neuron*, vol. 111, no. 12, pp. 1826–1845, 2023. DOI: 10.1016/j.neuron.2023.04.012.
- [55] G. Poudel, C. Innes, P. Bones, R. Watts, and R. Jones, “Losing the struggle to stay awake: Divergent thalamic and cortical activity during microsleeps,” *Human Brain Mapping*, vol. 35, no. 1, pp. 257–269, 2014. DOI: 10.1002/hbm.22178.
- [56] L. Fernandez and A. Lüthi, “Sleep spindles: Mechanisms and functions,” *Physiological Reviews*, vol. 100, no. 2, pp. 805–868, 2019. DOI: 10.1152/physrev.00042.2018.
- [57] I. Manshanden, J. C. De Munck, N. R. Simon, and F. L. Da Silva, “Source localization of meg sleep spindles and the relation to sources of alpha band

- rhythms,” *Clinical Neurophysiology*, vol. 113, no. 12, pp. 1937–1947, 2002. DOI: 10.1016/S1388-2457(02)00304-8.
- [58] D. Mylonas et al., “A novel approach to estimating the cortical sources of sleep spindles using simultaneous eeg/meg,” *Frontiers in Neurology*, vol. 13, 2022. DOI: 10.3389/fneur.2022.871166.
- [59] L. Yao, Y. Wang, Y. Gao, H. Gao, and X. Guo, “The role of the fronto-parietal network in modulating sustained attention under sleep deprivation: An fmri study,” *Frontiers in Psychiatry*, vol. 14, 2023. DOI: 10.3389/fpsy.2023.1289300.
- [60] Y. Chen and L. Pan, “Altered effective connectivity of thalamus with vigilance impairments after sleep deprivation,” *Journal of Sleep Research*, vol. 31, 2022. DOI: 10.1111/jsr.13693.
- [61] M. Strauss, J. Sitt, L. Naccache, and F. Raimondo, “Predicting the loss of responsiveness when falling asleep in humans,” *NeuroImage*, vol. 251, p. 119003, 2022. DOI: 10.1016/j.neuroimage.2022.119003.
- [62] M. Wiemers, H. Laufs, and F. Von Wegner, “Frequency analysis of eeg microstate sequences in wakefulness and nrem sleep,” *Brain Topography*, vol. 37, pp. 312–328, 2023. DOI: 10.1007/s10548-023-00971-y.
- [63] S. Zhou et al., “Dynamic functional connectivity states characterize nrem sleep and wakefulness,” *Human Brain Mapping*, vol. 40, no. 18, pp. 5256–5268, 2019. DOI: 10.1002/hbm.24770.
- [64] E. Bullmore and O. Sporns, “Complex brain networks: Graph theoretical analysis of structural and functional systems,” *Nature Reviews Neuroscience*, vol. 10, no. 3, pp. 186–198, 2009. DOI: 10.1038/nrn2575.
- [65] S. Ayyagari, R. Jones, and S. Weddell, “Detection of microsleep states from the eeg: A comparison of feature reduction methods,” *Medical & Biological Engineering & Computing*, vol. 59, pp. 1643–1657, 2021. DOI: 10.1007/s11517-021-02386-y.

- [66] B. Guragain, A. Rad, C. Wang, et al., “Eeg-based classification of microsleep by means of feature selection: An application in aviation,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2019, pp. 4060–4063. DOI: 10.1109/EMBC.2019.8856429.
- [67] Y. Zheng, J. Cammon, S. Zhang, et al., “A new feature selection approach for driving fatigue eeg detection with a modified machine learning algorithm,” *Computers in Biology and Medicine*, vol. 147, p. 105718, 2022. DOI: 10.1016/j.combiomed.2022.105718.
- [68] M. Hasan, A. Islam, M. Asad, et al., “Optimizing eeg-based microsleep detection: A comparative analysis of time and frequency domain features utilizing svm,” in *2024 IEEE International Conference on Signal Processing, Information, Communication and Systems (SPICSCON)*, 2024, pp. 1–5. DOI: 10.1109/SPICSCON64195.2024.10940738.
- [69] M. Hasan, M. Hossain, N. Sulaiman, et al., “Automatic microsleep detection based on knn classifier utilizing selected and effective eeg channels,” *Jurnal Teknologi*, vol. 86, 2024. DOI: 10.11113/jurnalteknologi.v86.22154.
- [70] W. Y. Leong, D. P. Mandic, M. Golz, and D. Sommer, “Blind extraction of microsleep events,” in *15th International Conference on Digital Signal Processing*, 2007, pp. 207–210. DOI: 10.1109/ICDSP.2007.4288555.
- [71] S. Chaabene, B. Bouaziz, A. Boudaya, et al., “Convolutional neural network for drowsiness detection using eeg signals,” *Sensors*, vol. 21, 2021. DOI: 10.3390/s21051734.
- [72] P. Gong, P. Wang, Y. Zhou, et al., “Tfac-net: A temporal-frequential attentional convolutional network for driver drowsiness recognition with single-channel eeg,” *IEEE Transactions on Intelligent Transportation Systems*, vol. 25, no. 7, pp. 7004–7016, 2024. DOI: 10.1109/TITS.2023.3347075.

- [73] P. Davidson, R. Jones, and M. Peiris, “Detecting behavioral microsleeps using eeg and lstm recurrent neural networks,” in *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 2005, pp. 5754–5757. DOI: 10.1109/IEMBS.2005.1615795.
- [74] R. Hefron, B. Borghetti, J. Christensen, and C. Schubert-Kabban, “Deep long short-term memory structures model temporal dependencies improving cognitive workload estimation,” *Pattern Recognition Letters*, vol. 94, pp. 96–104, 2017. DOI: 10.1016/j.patrec.2017.05.020.
- [75] J. Chung, C. Gülçehre, K. Cho, and Y. Bengio, “Empirical evaluation of gated recurrent neural networks on sequence modeling,” *arXiv preprint arXiv:1412.3555*, 2014.
- [76] B. Ghojogh and A. Ghodsi, “Recurrent neural networks and long short-term memory networks: Tutorial and survey,” *arXiv preprint arXiv:2304.11461*, 2023. DOI: 10.48550/arXiv.2304.11461.
- [77] D. Gao, K. Wang, M. Wang, et al., “Sft-net: A network for detecting fatigue from eeg signals by combining 4d feature flow and attention mechanism,” *IEEE Journal of Biomedical and Health Informatics*, vol. 28, no. 8, pp. 4444–4455, 2023. DOI: 10.1109/JBHI.2023.3285268.
- [78] Y. Rama, M. Reddy, M. Srinivas, et al., “An effective hybrid deep learning model for single-channel eeg-based subject-independent drowsiness recognition,” *Brain Topography*, vol. 37, pp. 1–18, 2023. DOI: 10.1007/s10548-023-01016-0.
- [79] X. Xu, W. Chen, Z. Pei, et al., “A temporal dependency learning cnn with attention mechanism for mi-eeg decoding,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3188–3200, 2023. DOI: 10.1109/TNSRE.2023.3299355.
- [80] E. Eldele, Z. Chen, C. Liu, et al., “An attention-based deep learning approach for sleep stage classification with single-channel eeg,” *IEEE Transactions on*

- Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 809–818, 2021. DOI: 10.1109/TNSRE.2021.3076234.
- [81] B. Abibullaev, A. Keutayeva, and A. Zollanvari, “Deep learning in eeg-based bcis: A comprehensive review of transformer models, advantages, challenges, and applications,” *IEEE Access*, vol. 11, pp. 127 271–127 301, 2023. DOI: 10.1109/ACCESS.2023.3329678.
 - [82] M. Pfeffer, S. Ling, and J. Wong, “Exploring the frontier: Transformer-based models in eeg signal analysis for brain–computer interfaces,” *Computers in Biology and Medicine*, vol. 178, p. 108 705, 2024. DOI: 10.1016/j.combiomed.2024.108705.
 - [83] Z. Zhuang, Y. Wang, Y. Chang, et al., “A connectivity-aware graph neural network for real-time drowsiness classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 83–93, 2023. DOI: 10.1109/TNSRE.2023.3336897.
 - [84] J. Paulo, G. Pires, and U. Nunes, “Cross-subject zero calibration driver’s drowsiness detection: Exploring spatiotemporal image encoding of eeg signals for convolutional neural network classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 905–915, 2021. DOI: 10.1109/TNSRE.2021.3079505.
 - [85] H. Obaidan, M. Hussain, and R. Almajed, “Eeg_dmnet: A deep multi-scale convolutional neural network for electroencephalography-based driver drowsiness detection,” *Electronics*, vol. 13, 2024. DOI: 10.3390/electronics13112084.
 - [86] N. Said, R. Suleiman, and S. Suhaimie, “Multimodal deep learning for microsleep detection using eeg and physiological signals,” in *2025 IEEE 6th International Conference in Robotics and Manufacturing Automation (ROMA)*, 2025, pp. 304–309. DOI: 10.1109/ROMA66616.2025.11155317.
 - [87] A. Chougule, J. Shah, V. Chamola, and S. Kanhere, “Enabling safe its: Eeg-based microsleep detection in vanets,” *IEEE Transactions on Intelli-*

- gent Transportation Systems*, vol. 24, no. 12, pp. 15 773–15 783, 2023. DOI: 10.1109/TITS.2022.3230259.
- [88] F. Safarov, A. Farkhod, A. Abdusalomov, R. Nasimov, and Y. I. Cho, “Real-time deep learning-based drowsiness detection: Leveraging computer-vision and eye-blink analyses for enhanced road safety,” *Sensors*, vol. 23, no. 14, p. 6459, 2023. DOI: 10.3390/s23146459.
 - [89] F. Makhmudov, D. Turimov, M. Xamidov, F. Nazarov, and Y. I. Cho, “Real-time fatigue detection algorithms using machine learning for yawning and eye state,” *Sensors*, vol. 24, no. 23, p. 7810, 2024. DOI: 10.3390/s24237810.
 - [90] C. Ryan et al., “Real-time face and eye tracking and blink detection using event cameras,” *Neural Networks*, vol. 141, pp. 87–97, 2021. DOI: 10.1016/j.neunet.2021.03.019.
 - [91] N. Mondal, “Real-time drowsiness detection system using eye-blink sensing and microcontroller-based alert mechanism,” *International Journal of Scientific Research in Engineering and Management*, 2025. DOI: 10.55041/IJSREM48132.
 - [92] M. Naidu, “Drowsiness detection and real-time system for monitoring driver vigilance using raspberry pi,” *International Journal for Research in Applied Science and Engineering Technology*, 2025. DOI: 10.22214/IJRASET.2025.67319.
 - [93] H. Ding et al., “Detection of mild cognitive impairment from non-semantic, acoustic voice features: The framingham heart study,” *JMIR Aging*, vol. 7, e55126, 2023. DOI: 10.2196/55126.
 - [94] S. Luz, F. Haider, S. Fuente, D. Fromm, and B. MacWhinney, “Detecting cognitive decline using speech only: The addresso challenge,” *arXiv preprint arXiv:2104.09356*, 2021. DOI: 10.1101/2021.03.24.21254263.

- [95] R. Wang, C. Kuang, C. Guo, Y. Chen, C. Li, Y. Matsumura, et al., “Automatic detection of putative mild cognitive impairment from speech acoustic features in mandarin-speaking elders,” *Journal of Alzheimer’s Disease*, vol. 95, no. 3, pp. 901–914, 2023. DOI: 10.3233/JAD-230373.
- [96] M. Hasan, C. Watling, and G. Larue, “Physiological signal-based drowsiness detection using machine learning: Singular and hybrid signal approaches,” *Journal of Safety Research*, vol. 80, pp. 215–225, 2021. DOI: 10.1016/j.jsr.2021.12.001.
- [97] M. Shahbakhti, M. Beiramvand, E. Nasiri, S. Far, W. Chen, J. Solé-Casals, et al., “Fusion of eeg and eye blink analysis for detection of driver fatigue,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 2037–2046, 2023. DOI: 10.1109/TNSRE.2023.3267114.
- [98] W. Y. Leong, D. P. Mandic, M. Golz, and D. Sommer, “Eeg-based driver fatigue estimation using synchronisation measures,” in *Proceedings of the 15th International Conference on Digital Signal Processing (DSP)*, Cardiff, UK, 2007, pp. 207–210. DOI: 10.1109/ICDSP.2007.4288555.
- [99] U. Venkatasubramanian, “Effective connectivity based characterization and detection/prediction of microsleeps using eeg,” Ph.D. dissertation, University of Otago, Dunedin, New Zealand, 2022.
- [100] Q. She, K. Chen, Z. Luo, T. Nguyen, T. Potter, and Y. Zhang, “Double-criteria active learning for multiclass brain-computer interfaces,” *Computational Intelligence and Neuroscience*, vol. 2020, p. 3287589, 2020. DOI: 10.1155/2020/3287589.
- [101] R. Luo and X. Wang, “Batch active learning with two-stage sampling,” *IEEE Access*, vol. 8, pp. 46518–46528, 2020. DOI: 10.1109/ACCESS.2020.2979315.
- [102] Z. He, Z. Xia, Y. Du, and X. Zhang, “Cluster optimized batch mode active learning sample selection method,” *Infrared Physics & Technology*, 2025. DOI: 10.1016/j.infrared.2025.105746.

- [103] L. Zhang et al., “A novel machine-learning-based batch selection method in sparse near-field scanning,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 11, pp. 5019–5028, 2022. DOI: 10.1109/TMTT.2022.3205909.
- [104] W. Wang, J. Huang, D. Wu, and P. Zhang, “Neural decoding based on active learning for intracortical brain-machine interfaces,” in *2020 IEEE International Conference on Mechatronics and Automation (ICMA)*, 2020, pp. 1425–1430. DOI: 10.1109/ICMA49215.2020.9233606.
- [105] Y. Wei et al., “A semi-supervised progressive learning algorithm for brain-computer interface,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2067–2076, 2022. DOI: 10.1109/TNSRE.2022.3192448.
- [106] Y. Long, J. Lin, C. Chen, and Z. Huang, “Active learning model for eeg annotation driven jointly by pseudo-labels and manual annotations,” in *2025 18th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, 2025, pp. 1–6. DOI: 10.1109/CISP-BMEI68103.2025.11259213.
- [107] R. Pillalamarri and S. Udhayakumar, “A review on eeg-based multimodal learning for emotion recognition,” *Artificial Intelligence Review*, vol. 58, 2025. DOI: 10.1007/s10462-025-11126-9.
- [108] L. Mou et al., “Isotropic self-supervised learning for driver drowsiness detection with attention-based multimodal fusion,” *IEEE Transactions on Multimedia*, vol. 25, pp. 529–542, 2023. DOI: 10.1109/TMM.2021.3128738.
- [109] M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, and X. Peng, “Smil: Multimodal learning with severely missing modality,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, no. 3, pp. 2302–2310, 2021. DOI: 10.1609/aaai.v35i3.16330.

- [110] R. Liu et al., “Frequency-aware masked autoencoders for multimodal pre-training on biosignals,” *arXiv preprint arXiv:2309.05927*, 2023. DOI: 10.48550/arxiv.2309.05927.
- [111] H.-J. Jang, H.-B. Shin, and K. Yin, “Cross-modal consistency-guided active learning for affective bci systems,” *2026 14th International Conference on Brain-Computer Interface (BCI)*, pp. 1–4, 2025. DOI: 10.48550/arxiv.2511.15138. [Online]. Available: <https://api.semanticscholar.org/CorpusID:283103808>.
- [112] J. Zhang and Y. Zhu, *Towards multimodal active learning: Efficient learning with limited paired data*, 2025. DOI: 10.48550/arxiv.2510.03247. [Online]. Available: <https://openreview.net/forum?id=ajPeZCRvZk>.
- [113] T. Rheude, R. Eils, and B. Wild, “Cohort-based active modality acquisition,” *arXiv preprint arXiv:2505.16791*, 2025. DOI: 10.48550/arxiv.2505.16791.
- [114] X. Li, L. Hong, J. Wang, and X. Liu, “Fatigue driving detection model based on multi-feature fusion and semi-supervised active learning,” *IET Intelligent Transport Systems*, vol. 13, no. 12, pp. 1671–1679, 2019. DOI: 10.1049/iet-its.2018.5590.
- [115] C. Rahmani, A. Benlamoudi, S. Bekhouche, Y. Bounab, D. Samai, and F. Dornaika, “Advancing road safety: A cnn-based semi-supervised learning approach for drivers’ drowsiness detection,” in *2024 1st International Conference on Electrical, Computer, Telecommunication and Energy Technologies (ECTE-Tech)*, 2024, pp. 1–6. DOI: 10.1109/ECTE-TECH62477.2024.10850976.
- [116] M. Ahmed et al., “A deep-learning approach to driver drowsiness detection,” *Safety*, vol. 9, no. 3, p. 65, 2023. DOI: 10.3390/safety9030065.
- [117] O. Hassan, A. Ibrahim, A. Gomaa, M. Makhoulouf, and B. Hafiz, “Real-time driver drowsiness detection using transformer architectures: A novel deep

- learning approach,” *Scientific Reports*, vol. 15, pp. 1–15, 2025. DOI: 10.1038/s41598-025-02111-x.
- [118] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, “Focal loss for dense object detection,” in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 2980–2988.
 - [119] R. Abdulsadig and E. Rodríguez-Villegas, “A comparative study in class imbalance mitigation when working with physiological signals,” *Frontiers in Digital Health*, vol. 6, 2024. DOI: 10.3389/fdgth.2024.1377165.
 - [120] M. Altalhan, A. Algarni, and T. Alouane, “Imbalanced data problem in machine learning: A review,” *IEEE Access*, vol. 13, pp. 13 686–13 699, 2025. DOI: 10.1109/ACCESS.2025.3531662.
 - [121] I. Araf, A. Idri, and I. Chairi, “Cost-sensitive learning for imbalanced medical data: A review,” *Artificial Intelligence Review*, vol. 57, 2024. DOI: 10.1007/s10462-023-10652-8.
 - [122] A. Tarawneh, A. Hassanat, G. Altarawneh, and A. Almuhaimeed, “Stop oversampling for class imbalance learning: A review,” *IEEE Access*, vol. 10, pp. 47 643–47 660, 2022. DOI: 10.1109/ACCESS.2022.3169512.
 - [123] R. Kobler, J. Hirayama, Q. Zhao, and M. Kawanabe, “Spd domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg,” *arXiv preprint arXiv:2206.01323*, 2022. DOI: 10.48550/arXiv.2206.01323.
 - [124] Z. Liang, Z. Zheng, W. Chen, Z. Pei, J. Wang, and J. Chen, “A novel deep transfer learning framework integrating general and domain-specific features for eeg-based brain-computer interface,” *Biomedical Signal Processing and Control*, vol. 95, p. 106 311, 2024. DOI: 10.1016/j.bspc.2024.106311.
 - [125] Y. Shi, Y. Wang, and H. Meng, “Single-source unsupervised domain adaptation for cross-subject mi-eeg classification based on discriminative informa-

- tion,” *Applied Intelligence*, vol. 54, pp. 10 438–10 454, 2024. DOI: 10.1007/s10489-024-05662-0.
- [126] Y. Liu, L. Qin, X. Chen, L. Jeannés, J. Coatrieux, and H. Shu, “Advancing cross-subject domain generalization in brain–computer interfaces with multiadversarial strategies,” *IEEE Transactions on Instrumentation and Measurement*, vol. 74, pp. 1–12, 2025. DOI: 10.1109/TIM.2025.3566804.
 - [127] J. Zhang, J. Fang, S. Liu, D. Liu, H. Wu, and J. Long, “Toward cross-brain-computer interface: A prototype-supervised adversarial transfer learning approach with multiple sources,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, 2024. DOI: 10.1109/TIM.2024.3451593.
 - [128] M. A. Mohamed, P. Redmond, M. Maniparambil, T. Ward, and N. E. O’Connor, *Attention-based temporal resolution enhancement for microsleep forecasting across multiple horizons*, AICS 2025 Companion Proceedings, Dublin City University, Accepted for publication (final metadata pending)., 2025.
 - [129] C. Challu, K. Olivares, B. Oreshkin, F. Garza, M. Mergenthaler-Canseco, and A. Dubrawski, “N-hits: Neural hierarchical interpolation for time series forecasting,” *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 37, no. 6, pp. 6989–6997, 2022. DOI: 10.1609/aaai.v37i6.25854.
 - [130] M. Nasser et al., “Forecasting epileptic seizures with wearable devices: A hybrid short- and long-horizon pseudo-prospective approach,” *Epilepsia*, vol. 66, pp. 3293–3308, 2025. DOI: 10.1111/epi.18466.
 - [131] D. Seok, S. Lee, M. Kim, J. Cho, and C. Kim, “Motion artifact removal techniques for wearable eeg and ppg sensor systems,” *Frontiers in Electronics*, vol. 2, 2021. DOI: 10.3389/felec.2021.685513.
 - [132] A. Studnicki, R. Downey, and D. Ferris, “Characterizing and removing artifacts using dual-layer eeg during table tennis,” *Sensors*, vol. 22, no. 15, p. 5867, 2022. DOI: 10.3390/s22155867.

- [133] S. Yang and Y. Lin, “Movement artifact suppression in wearable low-density and dry eeg recordings using active electrodes and artifact subspace reconstruction,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 3844–3853, 2023. DOI: 10.1109/TNSRE.2023.3319355.
- [134] X. An, Y. Liu, Y. Zhao, S. Lu, G. Stylios, and Q. Liu, “Adaptive motion artifact reduction in wearable eeg measurements using impedance pneumography signal,” *Sensors*, vol. 22, no. 15, p. 5493, 2022. DOI: 10.3390/s22155493.
- [135] A. Hamann and N. Carstengerdes, “Assessing the development of mental fatigue during simulated flights with concurrent eeg-fnirs measurement,” *Scientific Reports*, vol. 13, 2023. DOI: 10.1038/s41598-023-31264-w.
- [136] I. Mehmood et al., “Deep learning-based construction equipment operators’ mental fatigue classification using wearable eeg sensor data,” *Advanced Engineering Informatics*, vol. 56, p. 101978, 2023. DOI: 10.1016/j.aei.2023.101978.
- [137] M. Liu et al., “Combining meta and ensemble learning to classify eeg for seizure detection,” *Scientific Reports*, vol. 15, 2025. DOI: 10.1038/s41598-025-88270-3.
- [138] J. Fan, K. Zhang, H. Yipan, Y. Zhu, and B. Chen, “Parallel spatio-temporal attention-based tcn for multivariate time series prediction,” *Neural Computing and Applications*, vol. 35, pp. 13109–13118, 2021. DOI: 10.1007/s00521-021-05958-z.
- [139] Y. Ding et al., “Eeg-deformer: A dense convolutional transformer for brain-computer interfaces,” *IEEE Journal of Biomedical and Health Informatics*, vol. 29, pp. 1909–1918, 2024. DOI: 10.1109/JBHI.2024.3504604.
- [140] W. Yu and J. Yang, “Casleepnet: A cross attention-based multimodal fusion approach for sleep staging with eeg and eog,” in *ICASSP 2025 – 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2025, pp. 1–5. DOI: 10.1109/ICASSP49660.2025.10889570.

- [141] H. Jang, H. Shin, and S. Lee, “Uncertainty-aware cross-modal knowledge distillation with prototype learning for multimodal brain-computer interfaces,” in *2025 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2025, pp. 6682–6687. DOI: 10.1109/SMC58881.2025.11343095.
- [142] R. Maswanganyi, C. Tu, P. Owolawi, and S. Du, “Statistical evaluation of factors influencing inter-session and inter-subject variability in eeg-based brain computer interface,” *IEEE Access*, vol. 10, pp. 96 821–96 839, 2022. DOI: 10.1109/ACCESS.2022.3205734.
- [143] L. Hu, W. Gao, Z. Lu, C. Shan, W. Zhang, and Y. Li, “Subject-independent wearable p300 brain–computer interface based on convolutional neural network and metric learning,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 32, pp. 3543–3553, 2024. DOI: 10.1109/TNSRE.2024.3457502.
- [144] A. Keutayeva, N. Fakhrutdinov, and B. Abibullaev, “Compact convolutional transformer for subject-independent motor imagery eeg-based bcis,” *Scientific Reports*, vol. 14, 2024. DOI: 10.1038/s41598-024-73755-4.
- [145] K. Zhang, N. Robinson, S. Lee, and C. Guan, “Adaptive transfer learning for eeg motor imagery classification with deep convolutional neural network,” *Neural Networks*, vol. 136, pp. 1–10, 2020. DOI: 10.1016/j.neunet.2020.12.013.
- [146] M. A. Mohamed, P. Redmond, T. Ward, and N. E. O’Connor, *Mutual-information-guided multimodal fusion of eeg and eog for robust microsleep prediction in maintenance of wakefulness test (mwt) data*, RQ3 paper (submission/manuscript version used in thesis), Submitted to IEEE Access, 2026.
- [147] M. Nada, *Sleepy driver eeg brainwave data*, Kaggle dataset, Kaggle DOI provided in manuscript., 2023. DOI: 10.34740/KAGGLE/DSV/6391469.
- [148] M. A. Mohamed, P. Redmond, M. Maniparambil, T. Ward, and N. E. O’Connor, *Comprehensive evaluation of eeg features for microsleep detection: Statisti-*

- cal validation and methodological improvements*, AICS 2025 Companion Proceedings, Dublin City University, Accepted for publication (final metadata pending)., 2025.
- [149] D. Powers, “Evaluation: From precision, recall and f-measure to roc, informedness, markedness and correlation,” *Journal of Machine Learning Technologies*, 2011.
 - [150] M. Sokolova and G. Lapalme, “A systematic analysis of performance measures for classification tasks,” *Information Processing and Management*, 2009.
 - [151] T. Saito and M. Rehmsmeier, “The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets,” *PLoS One*, 2015.
 - [152] J. Davis and M. Goadrich, “The relationship between precision-recall and roc curves,” *ICML*, 2006.
 - [153] M. Friedman, “A comparison of alternative tests of significance for the problem of m rankings,” *The Annals of Mathematical Statistics*, 1940.
 - [154] J. Demšar, “Statistical comparisons of classifiers over multiple data sets,” *Journal of Machine Learning Research*, 2006.
 - [155] F. Wilcoxon, “Individual comparisons by ranking methods,” *Biometrics Bulletin*, 1945.
 - [156] S. Holm, “A simple sequentially rejective multiple test procedure,” *Scandinavian Journal of Statistics*, 1979.
 - [157] J. Cohen, *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.
 - [158] A. Hertig-Godeschalk et al., *Maintenance of wakefulness test (mwt) recordings*, Zenodo dataset (Version 1), 2019. DOI: 10.5281/zenodo.3251716.

- [159] M. P. Pauli, C. Pohl, and M. Golz, “Optimal eeg segmentation for microsleep detection based on machine learning: An empirical optimization investigation,” *Current Directions in Biomedical Engineering*, vol. 8, no. 2, pp. 749–752, 2022. DOI: 10.1515/cdbme-2022-1191.
- [160] M. Golz, A. Schenka, F. Haselbeck, and M. P. Pauli, “Inter-individual variability of eeg features during microsleep events,” *Current Directions in Biomedical Engineering*, vol. 5, no. 1, pp. 13–16, 2019. DOI: 10.1515/cdbme-2019-0004.
- [161] G. H. John and P. Langley, “Estimating continuous distributions in bayesian classifiers,” *Proceedings of UAI*, 1995.
- [162] R. A. Fisher, “The use of multiple measurements in taxonomic problems,” *Annals of Eugenics*, vol. 7, no. 2, pp. 179–188, 1936. DOI: 10.1111/j.1469-1809.1936.tb02137.x.
- [163] C. Cortes and V. Vapnik, “Support-vector networks,” *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [164] T. Cover and P. Hart, “Nearest neighbor pattern classification,” *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [165] L. Breiman, “Random forests,” *Machine Learning*, vol. 45, pp. 5–32, 2001.
- [166] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *KDD*, 2016.
- [167] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning,” *Nature*, vol. 521, pp. 436–444, 2015.
- [168] R. T. Schirrmeister et al., “Deep learning with convolutional neural networks for eeg decoding and visualization,” *Human Brain Mapping*, 2017.
- [169] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” *Proceedings of the IEEE*, 1998.

- [170] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Computation*, 1997.
- [171] Z. Zhang, C. Wang, Y. Li, and B. Chen, “A vehicle active safety model: Vehicle speed control based on driver vigilance detection using wearable eeg and sparse representation,” *Sensors*, vol. 16, no. 2, p. 242, 2016. DOI: 10.3390/s16020242.
- [172] R. Shoorangiz, S. J. Weddell, and R. D. Jones, “Prediction of microsleeps from eeg: Preliminary results,” in *Proceedings of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2016, pp. 4650–4653. DOI: 10.1109/EMBC.2016.7591764.
- [173] J. Krajewski, A. Batliner, and R. Wieland, “Multiple classifier applied on predicting microsleep from speech,” in *Proceedings of the International Conference on Pattern Recognition (ICPR)*, 2009, pp. 1–4. DOI: 10.1109/ICPR.2008.4761639.
- [174] J. Xie et al., “A transformer-based approach combining deep learning network and spatial-temporal information for raw eeg classification,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 30, pp. 2126–2136, 2022. DOI: 10.1109/TNSRE.2022.3194600.
- [175] A. Mohamed, N. E. O’Connor, and P. Redmond, *In-depth analysis of the most indicative features for detecting microsleeps from eeg signals*, SSRN preprint, Also available at SSRN (abstract 5169927)., 2025. DOI: 10.2139/ssrn.5169927.
- [176] Y. Song, Q. Zheng, B. Liu, and X. Gao, “Eeg conformer: Convolutional transformer for eeg decoding and visualization,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 710–719, 2023. DOI: 10.1109/TNSRE.2022.3230250.
- [177] M. Golz, S. Wollner, D. Sommer, and S. Schnieder, “Eeg feature relevance determination for microsleep detection,” *Current Directions in Biomedical*

- Engineering*, vol. 3, no. 2, pp. 815–818, 2017. DOI: 10.1515/cdbme-2017-0172.
- [178] J. Toppi, L. Astolfi, G. R. Poudel, C. Innes, F. Babiloni, and R. D. Jones, “Time-varying effective connectivity of the cortical neuroelectric activity associated with behavioural microsleeps,” *NeuroImage*, vol. 124, pp. 421–432, 2016. DOI: 10.1016/j.neuroimage.2015.08.059.
 - [179] G. Chiarion, L. Sparacino, Y. Antonacci, L. Faes, and L. Mesin, “Connectivity analysis in eeg data: A tutorial review of the state of the art and emerging trends,” *Bioengineering*, vol. 10, no. 3, p. 372, 2023. DOI: 10.3390/bioengineering10030372.
 - [180] V. Krishnamoorthy, R. Shoorangiz, S. Weddell, L. Beckert, and R. D. Jones, “Deep learning with convolutional neural network for detecting microsleep states from eeg: A comparison between the oversampling technique and cost-based learning,” in *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, IEEE, 2019, pp. 4760–4763. DOI: 110.1109/EMBC.2019.8857588.
 - [181] A. B. Buriro, R. Shoorangiz, S. J. Weddell, and R. D. Jones, “Predicting microsleep states using eeg inter-channel relationships,” *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 26, no. 12, pp. 2260–2271, 2018. DOI: 10.1109/TNSRE.2018.2878587.
 - [182] G. R. Poudel and R. D. Jones, “Multimodal neuroimaging with simultaneous fmri and eeg,” in *Handbook of Neuroengineering*, P. A. Hui et al., Eds., Singapore: Springer, 2022. DOI: 10.1007/978-981-15-2848-4_81-1.
 - [183] R. Thapa et al., “A multimodal sleep foundation model for disease prediction,” *Nature Medicine*, vol. 32, no. 2, pp. 752–762, 2026. DOI: 10.1038/s41591-025-04133-4.
 - [184] G. Ke et al., “Lightgbm: A highly efficient gradient boosting decision tree,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

- [185] S. M. Lundberg and S.-I. Lee, “A unified approach to interpreting model predictions,” in *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [186] A. Khorasani and T. Demazure, *Cross-temporal attention fusion (ctaf) for multimodal physiological signals in self-supervised learning*, 2026. arXiv: 2602.02784 [cs.LG]. [Online]. Available: <https://arxiv.org/abs/2602.02784>.