VILNIUS TECH
Vilniaus Gedimino technikos universitetas

Viroslava KAPUSTYNSKA

RESEARCH AND APPLICATION OF MACHINE LEARNING METHODS FOR MIGRAINE ATTACK PREDICTION

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (T 001)

Vilnius, 2026 2026-028-M

VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Viroslava KAPUSTYNSKA

RESEARCH AND APPLICATION OF MACHINE LEARNING METHODS FOR MIGRAINE ATTACK PREDICTION

DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (T 001)

Vilnius, 2026

The doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2021–2026.

Supervisor

Prof. Dr Šarūnas PAULIKAS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

Consultant

Prof. Dr Artūras SERACKIS (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

The Dissertation Defense Council of the Scientific Field of Electrical and Electronic Engineering of Vilnius Gediminas Technical University:

Chairman

Prof. Dr Vitalij NOVICKIJ (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

Members:

Dr Virginijus MARCINKEVIČIUS (Vilnius University, Informatics – N 009),
Dr Li-Ling Hope PAN (National Yang Ming Chiao Tung University, Taiwan, Medicine – M 001),
Assoc. Prof. Dr Aleksandr VASJANOV (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001),
Prof. Dr Nerija ŽURAUSKIENĖ (Vilnius Gediminas Technical University, Electrical and Electronic Engineering – T 001).

The dissertation will be defended at the public meeting of the Dissertation Defense Council of the Scientific Field of Electrical and Electronic Engineering in the *Aula Doctoralis* Meeting Hall of Vilnius Gediminas Technical University at **2 p.m. on 9 June 2026.**

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel.: +370 5 274 4956; fax +370 5 270 0112; e-mail: doktor@vilniustech.lt

A notification on the intended defense of the dissertation was sent on 8 May 2026. A copy of the doctoral dissertation is available for review at the Vilnius Gediminas Technical University repository (https://etalpykla.vilniustech.lt), the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania), and the Wroblewski Library of the Lithuanian Academy of Sciences (Žygimantų g. 1, LT-01102 Vilnius, Lithuania).

Vilnius Gediminas Technical University book No. 2026-028-M

https://doi.org/10.20334/2026-028-M

viroslava.kapustynska@vilniustech.lt

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Viroslava KAPUSTYNSKA

MAŠININIO MOKYMO METODŲ TYRIMAS IR TAIKYMAS MIGRENOS PRIEPUOLIUI PROGNOZUOTI

DAKTARO DISERTACIJA

TECHNOLOGIJOS MOKSLAI,
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (T 001)

Vilnius, 2026

Disertacija rengta 2021–2026 metais Vilniaus Gedimino technikos universitete.

Vadovas

prof. dr. Šarūnas PAULIKAS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Konsultantas

prof. dr. Artūras SERACKIS (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Vilniaus Gedimino technikos universiteto Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo taryba:

Pirmininkas

prof. dr. Vitalij NOVICKIJ (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Nariai:

dr. Virginijus MARCINKEVIČIUS (Vilniaus universitetas, Informatika – N 009),

dr. Li-Ling Hope PAN (Nacionalinis Jang Mingo Čiao Tungo universitetas, Taivanas, Medicina – M 001),

doc. dr. Aleksandr VASJANOV (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001),

prof. dr. Nerija ZURAUSKIENE (Vilniaus Gedimino technikos universitetas, Elektros ir elektronikos inžinerija – T 001).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties disertacijos gynimo tarybos posėdyje **2026 m. birželio 9 d. 14 val.** Vilniaus Gedimino technikos universiteto *Aula Doctoralis* posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (0 5) 274 4956; faksas (0 5) 270 0112; el. paštas doktor@vilniustech.lt

Pranešimai apie numatomą ginti disertaciją išsiųsti 2026 m. gegužės 8 d.
Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto talpykloje https://etalpykla.vilniustech.lt/, Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva) bei Lietuvos mokslų akademijos Vrublevskių bibliotekoje (Žygimantų g. 1, LT-01102 Vilnius, Lietuva).

Abstract

Migraine is a complex neurological disorder characterized by strong inter- and intra-individual variability, which makes early forecasting difficult using only clinical observations. Wearable biosensors combined with machine learning offer new opportunities to detect subtle physiological changes that may precede migraine attacks and to develop individualized prediction models.

This dissertation investigates migraine analysis and next-day prediction using physiological recordings collected under real-life monitoring conditions. Data were obtained with the Empatica Embrace Plus wearable device and include electrodermal activity, pulse rate, skin temperature, and movement-related signals. The analysis focuses on nocturnal recordings, since the night period provides a more stable physiological context with fewer external disturbances. Nights were standardized using sleep-based contextual selection and consistent night-level rules.

The experimental framework is organized in two stages. In the first stage, a window-level binary classification task is used as an exploratory methodological analysis to examine how design choices influence model performance. Night recordings are segmented into analysis frames ranging from 5 to 120 minutes, statistical features are extracted, and the influence of signal preprocessing and feature representation is evaluated across several classifier families, including Random Forest, XGBoost, histogram-based gradient boosting, support vector machines, and k-nearest neighbors.

In the second stage, the research evaluates next-day migraine prediction based on whole-night recordings. This stage refines the experimental methodology to obtain more reliable estimates of predictive performance under a stricter validation framework. The analysis focuses on the effect of temporal aggregation while comparing the same classifier families under consistent evaluation conditions.

The results demonstrate considerable variability across participants in achievable prediction performance and optimal modeling configurations. Shorter analysis frames generally preserve informative short-term physiological changes, whereas longer windows tend to smooth these variations. Signal preprocessing shows a window-dependent effect and does not consistently improve performance. Overall, the results highlight the importance of temporal resolution, rigorous validation, and individualized modeling for wearable-based migraine prediction systems.

Reziumė

Migrena yra sudėtingas neurologinis sutrikimas, pasižymintis didele tarpindividualine ir intraindividualine kintamumo variacija, todėl ankstyvas priepuolių prognozavimas remiantis vien klinikiniais stebėjimais yra sudėtingas. Nešiojamieji biosensoriai kartu su mašininio mokymosi metodais suteikia galimybę nustatyti subtilius fiziologinius pokyčius, galinčius pasireikšti prieš migrenos priepuolį, ir kurti individualizuotus prognozavimo metodus.

Disertacijoje tiriama migrenos analizė ir kitos dienos migrenos prognozavimas naudojant fiziologinius duomenis, surinktus realiomis gyvenimo sąlygomis. Duomenys buvo registruojami naudojant nešiojamąjį įrenginį *Empatica Embrace Plus* ir apima elektroderminės odos veiklos, pulso dažnio, odos temperatūros ir judesio signalus. Analizė orientuota į naktinius įrašus, nes nakties laikotarpis pasižymi stabilesnėmis fiziologinėmis sąlygomis ir mažesne išorinių veiksnių įtaka. Naktys buvo standartizuotos taikant miego pagrindu paremtą kontekstinį atrinkimą ir nuoseklias naktų parinkimo taisykles.

Eksperimentinė analizė organizuota dviem etapais. Pirmajame etape taikoma lango lygmens dvejetainė klasifikacijos užduotis, siekiant įvertinti, kaip metodiniai sprendimai veikia modelių veikimą. Naktiniai įrašai suskirstomi į analizės langus nuo penkių iki šimto dvidešimties minučių trukmės, apskaičiuojami statistiniai požymiai, o signalų išankstinio apdorojimo ir požymių reprezentacijos įtaka vertinama taikant kelias klasifikatorių šeimas, įskaitant *Random Forest, XGBoost,* histograminį gradientinį stiprinimą, atraminių vektorių mašinas ir artimiausių kaimynų metodą.

Antrajame etape vertinamas kitos dienos migrenos prognozavimas, remiantis visos nakties duomenimis. Šiame etape taikoma griežtesnė validavimo schema, siekiant gauti patikimesnius modelių veikimo įverčius, o analizėje daugiausia dėmesio skiriama laiko agregavimo poveikiui, lyginant tas pačias klasifikatorių šeimas nuoseklioje vertinimo aplinkoje.

Rezultatai rodo didelę dalyvių tarpusavio variaciją tiek prognozavimo tikslumo, tiek optimalių modelių konfigūracijų atžvilgiu. Trumpesni analizės langai dažniau išsaugo informatyvius trumpalaikius fiziologinius pokyčius, o ilgesni langai linkę šiuos svyravimus išlyginti. Signalų išankstinis apdorojimas pasižymi nuo lango trukmės priklausančiu poveikiu ir neužtikrina nuoseklaus rezultatų pagerėjimo. Gauti rezultatai pabrėžia laiko rezoliucijos, griežtos validacijos ir individualizuoto modeliavimo svarbą kuriant migrenos prognozavimo sistemas, paremtas nešiojamųjų įrenginių duomenimis.

Notations

Abbreviations

Acc – accelerometer (liet. *akselerometras*);
ACT – activity count (liet. *aktyvumo skaičius*);
AI – artificial intelligence (liet. *dirbtinis intelektas*);
ANS – autonomic nervous system (liet. *autonominė nervų sistema*);
ANOVA – analysis of variance (liet. *dispersinės analizės procedūra*);
bpm – beats per minute (liet. *dūžių per minutę*);
ECG – electrocardiogram (liet. *elektrokardiograma*);
EDA – electrodermal activity (liet. *elektrodermінė veikla*);
HR – heart rate (liet. *širdies ritmas*);
KNN – k-Nearest Neighbors (*k artimiausių kaimynų metodas*)
MET – metabolic equivalent of task (liet. *užduoties metabolinis ekvivalentas*);
ML – machine learning (liet. *mašininis mokymasis*);
NaN – not a number (liet. *ne skaičius*);
PPG – photoplethysmography (liet. *fotopletizmografija*);
PR – pulse rate (liet. *pulso dažnis*);
PRV – pulse rate variability (liet. *pulso dažnio kintamumas*);
RMSSD – root mean square of successive differences (liet. *nuoseklių skirtumų kvadratinių vidurkių šaknis*);

RR – respiratory rate (liet. *kvėpavimo dažnis*);
SC – step count (liet. *žingsnių skaičius*);
SDS – sleep detection stage (liet. *miego aptikimo etapas*);
SVM – Support Vector Machine (*atraminių vektorių metodas*)
SkinTEMP – skin temperature (liet. *odos temperatūra*);
SpO_2 – peripheral capillary oxygen saturation (liet. *periferių kapiliarų deguonies prisotinimas*);
TEMP – temperature (liet. *temperatūra*).

Contents

Introduction

Problem Formulation

Migraine is a complex neurological disorder that places a substantial burden on patients and healthcare systems. Its attacks are difficult to predict, while prodromal symptoms vary across individuals. Wearable devices offer continuous monitoring of physiological and behavioral signals that may capture pre-migraine changes, although real-world data are often noisy, incomplete, and influenced by sleep, activity, and circadian rhythms.

This dissertation addresses the problem of extracting reliable and interpretable pre-migraine information from wearable physiological data, for which the development of methods for robust feature extraction and interpretation is necessitated by the limitations of conventional signal processing and classification approaches when applied to noisy, non-stationary signals acquired in real-world conditions, making the problem highly relevant to Electrical and Electronic Engineering.

To address this problem, the dissertation develops a context-aware, multi-scale analytical framework that combines data auditing, context-sensitive labeling, and systematic evaluation of temporal aggregation, preprocessing, and feature representation in nocturnal data. Machine learning is employed to assess migraine classification and next-day migraine prediction, while feature importance analysis is utilized to identify informative signals, descriptors, and recurring patterns across participants and modeling settings.

Based on this formulation, the dissertation is guided by two hypotheses:

1. The detectability of pre-migraine physiological changes is strongly influenced by signal-processing and data-modeling choices, particularly temporal aggregation, preprocessing, feature extraction, and feature representation.
2. Despite inter-individual variability, recurrent physiological patterns and groups of informative features can be identified at the cohort level, reflecting shared structure in pre-migraine dynamics.

Relevance of the Dissertation

The relevance of this dissertation lies in the need to advance understanding of physiological changes preceding migraine attacks and address existing research gaps. Using physiological signals from prolonged real-world monitoring, it evaluates data quality, validation strategies, and the feasibility of migraine classification and next-day prediction, while assessing the effects of signal segmentation, filtering, and nocturnal feature importance.

Research Object

The research object of this dissertation is a data analysis framework for wearable physiological recordings, for identifying pre-migraine patterns and evaluating how analysis design choices influence their detection and the feasibility of next-day migraine prediction.

Aim of the Dissertation

The dissertation aims to develop and evaluate a reproducible framework for extracting, analyzing, and interpreting pre-migraine information from wearable-derived physiological recordings under naturalistic monitoring conditions.

Tasks of the Dissertation

To address the stated research problem and achieve the aim of the dissertation, the following tasks were formulated:

1. To develop a wearable-data processing pipeline for migraine analysis, including participant-level dataset construction, definition of nocturnal observations, and consistent labeling of pre-migraine and control periods.
2. To develop and evaluate binary classification models for distinguishing pre-migraine and control states using wearable-derived nocturnal physiological signals, with particular attention to individualized modeling.
3. To quantify the impact of key signal-processing and modeling choices, including temporal segmentation and preprocessing, on the detectability of pre-migraine patterns and classification performance.
4. To identify physiologically informative signals and feature groups associated with pre-migraine changes using statistical feature analysis and model-based relevance measures, and to evaluate the effect of feature reduction on model robustness and interpretability.
5. To validate the proposed approach on an expanded cohort under a unified training and validation protocol, and to assess the feasibility of next-day migraine prediction from nocturnal wearable data under temporally separated validation.

Research Methodology

The following methods were applied to meet the research objectives:

1. A structured review of migraine research was conducted based on data acquired from wearable devices, covering migraine phases, pre-migraine dynamics, measurable physiological signals, data-quality issues in real-world recordings, and common preprocessing, feature extraction, and machine-learning approaches.
2. Physiological data were collected under real-life conditions using the Empatica Embrace Plus device. The analysis focused on nocturnal recordings to reduce daytime variability. Signal completeness and missingness patterns were assessed for each participant and signal.
3. Preprocessing of the physiological signals ensured physiological plausibility by removing invalid samples and clipping implausible pulse-rate values. In addition, several filtering approaches were evaluated, including median, Savitzky–Golay, and Butterworth low-pass filtering.
4. Nocturnal recordings were segmented into fixed-length frames from 5 to 120 minutes. Statistical features were extracted, and differences between pre-migraine and control states were analyzed using descriptive statistics and one-way ANOVA.

5. Binary classification models were developed to distinguish pre-migraine from control states. Two validation strategies were applied: frame-level splitting for methodological comparison and night-level splitting for stricter evaluation. Hyperparameter optimization was performed using RandomizedSearchCV with GroupKFold cross-validation based on night identifiers. Performance was assessed using F1-score, recall, precision, ROC-AUC, and PR-AUC, with emphasis on the positive class.

Scientific Novelty of the Dissertation

In the dissertation, the following scientific contributions are presented:

1. A structured framework for next-day migraine prediction from nocturnal physiological data acquired from wearable devices, comprising a consistent night-to-next-day labeling scheme, exclusion of post-migraine nights, and explicit handling of ambiguous temporal intervals.
2. A structured protocol for characterizing datasets acquired from wearable devices, formalizing the assessment of signal completeness, missingness patterns, temporal coverage, and feature computability.
3. A multi-scale temporal aggregation scheme for migraine prediction based on physiological data acquired from wearable devices, using signal intervals from 5 to 120 minutes, and demonstrating that shorter windows preserve pre-migraine-related information more effectively, whereas longer windows reduce sensitivity to the minority class.
4. Cohort-level results demonstrating that, in nocturnal physiological recordings acquired from wearable devices, low-order statistical descriptors are more stable than higher-order features, and predictor composition varies with temporal scale.
5. Evidence from the extended cohort that conventional filtering of physiological biomarkers acquired from wearable devices does not improve migraine classification performance, indicating that the benefit of preprocessing must be empirically validated rather than assumed.

Practical Value of the Research Findings

1. The proposed data-auditing procedure can be used to assess the suitability of nocturnal physiological datasets acquired from wearable devices before model development and improve comparability across studies.

2. The analytical pipeline developed in this dissertation integrates sleep-based segmentation, transparent labeling, feature extraction, validation, and can be adapted for migraine studies and related episodic conditions using data acquired from wearable devices.
3. The compact and interpretable set of features identified in this dissertation can support the development of simpler and more transparent prediction models based on physiological data acquired from wearable devices without loss of performance.
4. The practical guidance on temporal segmentation supports the use of shorter nocturnal analysis windows for detecting pre-migraine states and can inform future monitoring pipelines based on wearable devices.
5. The two-stage modeling strategy combines cohort-level learning with subsequent individual adaptation for personalized next-day migraine prediction using wearable data.

Defended Statements

The following statements, based on the results of the present investigation, are to be defended:

1. Pre-migraine detection from nocturnal wearable data strongly depends on the temporal scale of analysis: increasing the window length from 5–10 to 30 min reduces positive-class recall by approximately 10–30% and F1-score by 5–15%, while extension to 60–120 min causes a further decline and increases inter-run variability.
2. Preprocessing did not improve positive-class recall or F1-score in any evaluated setting under strict validation, indicating that its benefit cannot be assumed a priori and must be critically evaluated.
3. Nocturnal physiological data contain a reproducible cohort-level structure of informative predictors, in which low-order statistical descriptors are more stable than higher-order features: at 5–10 min windows, recurrent EDA descriptors appear approximately 1.5 to 3 times more often than most higher-order or non-EDA features, whereas longer windows are associated with a more heterogeneous predictor structure.
4. Ensemble classifiers achieve the most robust and balanced performance in nocturnal wearable-based migraine modeling under strict validation, outperforming SVM and KNN in positive-class detection by approximately 20–60% and showing lower inter-run variability. SVM remains limited, whereas KNN does not detect the migraine class.

Approval of the Research Findings

The results of the research were published in four publications: two articles in journals indexed in the Clarivate Analytics Web of Science Core Collection with Impact Factors (Kapustynska et al., 2024, 2025). The author delivered two presentations at two international scientific conferences:

1. International Conference "Science in Motion: Classic and Modern Tools and Methods in Scientific Investigations", Vinnytsia–Vienna, 2024.
2. 15th Conference on Data Analysis Methods for Software Systems (DAMSS), Druskininkai, Lithuania, 2024.

Structure of the Dissertation

The dissertation consists of an introduction, three main chapters, general conclusions, and a summary in Lithuanian. The dissertation comprises 111 pages, including 35 figures, 4 formulas, and 17 tables. Additionally, it has 104 references.

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Prof. Dr. Šarūnas Paulikas, for his continuous support, encouragement, and trust. I am also deeply grateful to my scientific consultant, Prof. Dr. Artūras Serackis, for his guidance and thoughtful feedback, which greatly shaped this work. I sincerely thank the reviewers of this dissertation for their constructive comments and valuable suggestions, which helped strengthen both the study and my scientific reasoning.

I am grateful to Dr. Vytautas Abromavičius for our scientific collaboration and to neurologist Saulius Adruškevičius for providing and validating the clinical data used in this research. I also warmly thank my colleagues and the academic community at the faculty for their support, openness, and the constructive environment in which this work was carried out.

My deepest personal gratitude belongs to my husband, whose steadfast support, patience, and faith in me sustained me throughout this journey.

Finally, I thank my daughter, Petra. Living with Dravet syndrome, she has been my greatest source of inspiration. Her strength, courage, and quiet endurance have taught me resilience and the importance of moving forward even in the most difficult moments.

1

Literature Survey on Migraine Dynamics, Physiological Signals, and Prediction

This chapter reviews the current literature and describes the migraine phases, including the prodromal phase, the main focus of this dissertation, and the association of physiological signal changes with migraine. It then presents research on migraine forecasting, with particular attention given to the sensors and devices used, the recorded signals, extracted features, and the applied classifiers. A summary of existing research focuses on signal filtering techniques and selecting informative value ranges when processing physiological signals. Finally, the chapter concludes with a summary and the formulation of research hypotheses based on the conducted literature review. Parts of this literature review were previously presented in the author's publications (Kapustynska et al., 2024, 2025).

1.1. Migraine Phases and the Rationale for Predictive Approaches

The International Classification of Headache Disorders, the third edition (ICHD-3) of the International Headache Society (IHS), recognizes four primary headache disorders: migraine, tension-type headache, trigeminal autonomic cephalalgias, and other primary headache disorders (Imtiaz et al., 2025; Society (IHS), 2018). Migraine is increasingly conceptualized as a dynamic and continuous disorder with interictal periods that constitute functionally relevant phases of an ongoing pathological cycle rather than symptom-free intervals (Cortese et al., 2017; Hsiao et al., 2022; Puledda et al., 2023; van den Hoek et al., 2024; Vincent et al., 2022; Xie et al., n.d.). Clinically, migraine attacks are commonly described as progressing through four partially overlapping phases (Fig. 1.1): prodrome, aura, headache (ictal phase), and postdrome (Ferrari et al., 2015; Gao et al., 2024; Lipton et al., 2025; Raggi et al., 2024; Vincent et al., 2022).

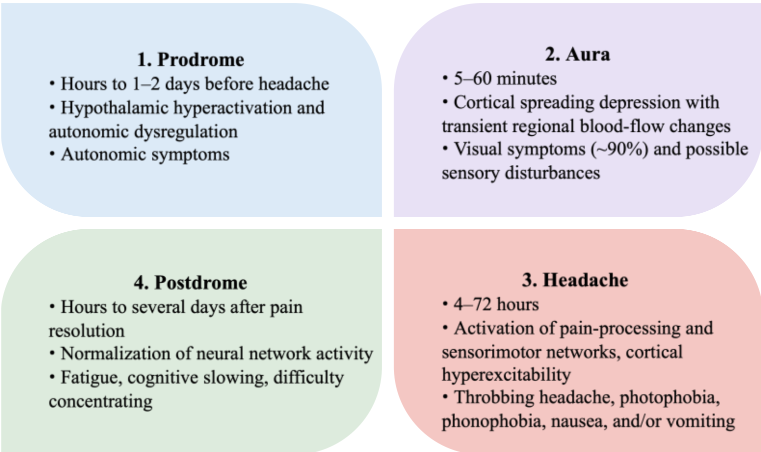


Fig. 1.1. Migraine phases (compiled by the author)

Migraine is classified as episodic when headaches occur on fewer than 15 days per month, and as chronic when headaches are present on at least 15 days per month for more than three months, with a minimum of eight migraine days per month (Butt et al., 2025; Dominguez et al., 2025; Society (IHS), 2018).

The prodromal phase is reported by approximately 70–80% of individuals with migraine and typically emerges 24–48 hours before headache onset (Gao et al., 2024; Laurell et al., 2016; Lipton et al., 2025; Mehnert et al., 2023; Raggi et al., 2024; van den Hoek et al., 2024). Prodromal manifestations encompass a heterogeneous set of symptoms, including changes in mood and activity, fatigue, yawning, neck stiffness, impaired concentration, sensory sensitivity, and autonomic symptoms, although their presence and combination vary markedly across individuals (Gao et al., 2024; Raggi et al., 2024; Society (IHS), 2018; Vincent et al., 2022).

Evidence for the temporal proximity of prodromal symptoms to headache onset was provided by the PRODROME trial, which demonstrated that more than 80% of identified prodromal events were followed by headache within a 1–6-hour window among participants who could reliably recognize their prodrome (Schwedt et al., 2023, 2025). Importantly, early intervention during this phase reduced headache severity and functional disability in this selected population, highlighting the prodrome as a clinically actionable window, albeit one that cannot be generalized to all migraineurs (Schwedt et al., 2025). At the same time, population-based studies and meta-analyses report substantially lower prevalence estimates, underscoring the non-specific nature of many prodromal symptoms, which often persist into the headache and postdromal phases, thereby blurring phase boundaries and limiting their reliability as standalone predictors (Eigenbrodt et al., 2022).

Approximately one-third of individuals with migraine experience aura, consisting of transient focal neurological disturbances, most commonly visual phenomena such as gradually expanding scotomas or zigzag patterns lasting less than one hour (Dodick, 2018; Dominguez et al., 2025; Hansen & Charles, 2019; Hougaard et al., 2025; Lucas, 2021; Rio & Cutrer, 2023; Viana et al., 2019). Aura is closely associated with cortical spreading depolarization and transient alterations in cerebral perfusion, providing a mechanistic explanation for its progressive and spatially evolving features (Hougaard et al., 2025; Karsan et al., 2023; Mehnert et al., 2023; van den Hoek et al., 2024). However, aura does not consistently precede headache and may occur during or after pain onset, and neuroimaging studies indicate that central alterations, including hypothalamic changes, precede both aura and headache, supporting the view that aura represents an epiphenomenon rather than a primary trigger of migraine attacks (Mehnert et al., 2023).

The headache phase constitutes the most clinically burdensome stage of the migraine attack and typically lasts between 4 and 72 hours. It is characterized by pulsating or throbbing pain, often unilateral, accompanied by nausea, photophobia, and phonophobia, which together drive functional impairment and care-seeking behavior (Kellier et al., 2023; Lipton et al., 2025). Neuroimaging evidence suggests that the ictal phase reflects downstream consequences of processes initiated earlier in the migraine cycle, as hypothalamic activation is reduced compared to the pre-ictal period (Mehnert et al., 2023; van den Hoek et al., 2024).

Following headache resolution, the postdromal phase may persist for hours or days and is characterized by fatigue, cognitive slowing, and a sense of incomplete recovery. Although less extensively studied, neurophysiological findings suggest a gradual normalization of brain network activity rather than an abrupt return to baseline, further supporting the concept of migraine as a continuous disorder extending beyond the period of active pain (Carvalho et al., 2022; Lipton et al., 2023, 2025; van den Hoek et al., 2024).

Taken together, the phase-based organization of migraine highlights a disorder driven by early, centrally mediated alterations that precede the onset of pain. While the headache and postdromal phases account for most clinical burden, they offer limited opportunities for anticipatory intervention. In contrast, pre-migraine changes, despite their heterogeneity and methodological challenges, represent the most promising temporal window for predictive modeling and proactive migraine management.

1.2. Physiological Signals and the Autonomic Nervous System

Studies have shown that alterations in biomedical signals regulated by the autonomic nervous system (ANS) are involved in a wide range of chronic conditions (Arnold & Young, 2025; Barakat et al., 2012; Bellocchi et al., 2022; D’Agnano et al., 2024; Jaychandran et al., 2016; Kannan et al., 2025; Sarkis et al., 2015; Singh & Roy, 2017; Stritzelberger et al., 2025), including migraine (Arakaki et al., 2023; Flamerz Arkawazi et al., 2021; Pagán et al., 2016; Vu et al., 2025), reflecting underlying dysregulation of sympathetic and parasympathetic balance. The ANS plays a central role in regulating cardiovascular, electrodermal, thermal, and respiratory functions, making related physiological signals particularly relevant for investigating migraine-related changes across different phases of the migraine cycle.

Among the physiological signals most frequently examined in migraine research, electrodermal activity (EDA), heart rate (HR), and heart rate variability (HRV) have strong physiological justification because of their direct association with autonomic regulation. EDA reflects sympathetic nervous system activation and is sensitive to changes in emotional arousal and stress-related states, while HR and HRV capture dynamic interactions between sympathetic and parasympathetic control of cardiac function (Pagán et al., 2016). Reductions in HRV and elevations in HR have been associated with autonomic imbalance, a phenomenon increasingly recognized in individuals with migraine.

Additional signals, such as skin temperature (SkinTEMP) and peripheral oxygen saturation (SpO_2), have been explored as indirect indicators of autonomic modulation, particularly through their relationship with peripheral blood flow and vascular regulation. Although these signals may be more context-dependent and sensitive to external factors, several studies suggest that their combined analysis with core autonomic markers can enhance the characterization of migraine-related physiological changes (Pagán et al., 2016).

Research on migraine prediction has identified multivariate approaches that integrate several autonomically regulated signals as more informative than single-signal analyses. In one of the earliest studies using physiological monitoring, Pagán et al. (2016) demonstrated that combinations of HR, EDA, SkinTEMP, and SpO_2 provided more meaningful information for migraine prediction than any single signal alone, although model performance varied substantially across participants and was associated with relatively high false-positive rates. These findings highlight both the potential and the limitations of autonomic signals when used in isolation.

Subsequent investigations have extended this line of research by employing wearable-based monitoring of autonomic signals such as HR, EDA, and SkinTEMP in real-world settings. For example, Gonzalez-Martinez et al. (2022) recorded continuous wrist-based physiological data from individuals with episodic migraine and reported that personalized models could identify migraine-related patterns preceding pain onset. While these studies primarily focused on predictive modeling, they also reinforced the relevance of autonomic signals as carriers of migraine-related physiological information.

Importantly, autonomic signal alterations have been reported across multiple temporal stages of migraine. Interictal recordings are often used to characterize baseline autonomic dysregulation, whereas pre-migraine changes may reflect centrally orchestrated modulation of autonomic tone preceding headache onset. During the ictal phase, systemic autonomic symptoms such as nausea and vomiting further indicate an acute imbalance within the ANS. Together, these observations support the view that migraine-related physiological changes are not confined to the headache phase but evolve dynamically across the migraine cycle.

1.2.1. Wearable Sensor Platforms for Physiological Data Acquisition

The increasing availability of personalized medical technologies, including smartphone applications and wearable devices, has enabled the large-scale collection of objective, prospective physiological data in naturalistic settings. This development has substantially reduced the cost and burden of long-term monitoring and has facilitated migraine research at both individual and population levels. Wearable devices enable continuous recording of physiological signals, including SkinTEMP, HR, EDA, physical activity, and sleep-related metrics, providing insight into physiological changes that may precede migraine onset.

Wearable sensors have become integral tools for physiological monitoring due to their non-invasive nature, portability, and continuous data acquisition. In migraine research, wrist-worn multi-sensor platforms are most commonly employed, as they enable simultaneous monitoring of autonomic and behavioral signals. Frequently reported sensor combinations include accelerometry, SkinTEMP, EDA, blood volume pulse, HR, and HRV, reflecting key dimensions of ANS regulation and daily activity patterns (Danelakis et al., 2025). In addition, some studies incorporate sleep-related measures and circadian indicators derived from passive nighttime monitoring, further extending the temporal coverage of physiological assessment (Harmsen et al., 2025).

Among the monitored physiological modalities, autonomic signals are most consistently represented. EDA serves as an indicator of sympathetic nervous system activation, while cardiovascular metrics such as HR and HRV are commonly derived from photoplethysmography-based blood volume pulse sensors (Arnold & Young, 2025; Stritzelberger et al., 2025). Peripheral SkinTEMP reflects autonomic regulation of vascular tone and has been used to capture stress- and state-related changes, although its interpretability depends on sensor placement and environmental context (Moser et al., 2024). Accelerometry is widely applied to quantify physical activity, rest–activity cycles, and sleep–wake patterns, supporting contextual interpretation of physiological signals (Faust et al., 2025).

Several wearable hardware platforms have been employed in migraine-related studies. Research-grade wrist-worn devices, such as the Empatica E4 and Empatica Embrace series, integrate multiple sensors, including electrodermal activity, photoplethysmography, accelerometry, and infrared thermopile-based skin temperature measurement, enabling synchronized acquisition of autonomic and behavioral data (Koskimäki et al., 2017; Stritzelberger et al., 2025). These devices have been used in both laboratory and naturalistic settings to capture physiological fluctuations across interictal, pre-ictal, and ictal periods. Complementary approaches have also included portable ECG systems, chest-strap heart rate monitors, and mobile EEG platforms, although such systems are typically less suitable for long-term, everyday use due to practical constraints (Arnold & Young, 2025; van den Hoek et al., 2024).

Importantly, wearable-based migraine studies emphasize feasibility and ecological validity. Prior investigations have demonstrated that combinations of autonomic signals, particularly EDA, HR, HRV, and SkinTEMP, are informative when collected continuously in natural environments, supporting their use in longitudinal migraine monitoring (Gonzalez-Martinez et al., 2022; Pagán et al., 2016). These findings highlight the relevance of multi-sensor wearable platforms for capturing pre-migraine physiological changes while minimizing participant burden.

In the present research, the Empatica Embrace Plus (Empatica, 2025) device was selected for its ability to continuously record key autonomic and activity-related signals on a single wrist-worn platform. Its sensor configuration aligns with physiological modalities most frequently reported in migraine prediction research. Meanwhile, its design supports long-term, non-invasive monitoring under real-world conditions, making it suitable for individualized and longitudinal analysis.

1.2.2. Data-Centric Analysis of Wearable Physiological Recordings

Recent studies using wearable devices are increasingly shifting the emphasis from purely comparing algorithmic accuracy to a more data-centric, structural analysis of physiological recordings (Abromavičius et al., 2023; Bota et al., 2024). This shift is driven by the fact that real-world signals such as EDA, pulse rate, HRV proxies, SkinTEMP, and Acc are highly context-dependent (sleep/wake, activity level, and wear compliance) and are frequently degraded by motion artifacts and loss of skin contact (Faust et al., 2025; Moser et al., 2024; Campanella et al., 2024). Consequently, before interpreting ANS dynamics or searching for pre-attack markers, it is essential to characterize data quality quantitatively, mechanisms of missingness, and biomarker validity (Stritzelberger et al., 2025; Faust et al., 2025).

A complementary line of work focuses on longitudinal, onset-aligned analysis, in which the primary objective is the temporal evolution of physiology in the hours to days preceding migraine onset (Mehnert et al., 2023; Schulte and May, 2016). Long-term monitoring studies highlight that identifying prodromal signatures requires not only sufficient observation time, but also circadian-matched control windows to separate potential pre-onset changes from time-of-day and behavioral confounding (Oosterlee et al., 2025; Mehnert et al., 2023). Notably, detailed single-subject or subject-level analyses can track physiological transitions up to 48 hours before pain-onset effects, which may be diluted in cohort aggregation (Mehnert et al., 2023; Faust et al., 2025).

A substantial body of literature also addresses the validity of digital biomarkers and modality-specific limitations. In particular, PPG-derived HRV metrics may differ systematically from ECG estimates (e.g., inflated RMSSD), a phenomenon commonly attributed to modality effects and vascular modulation (Stritzelberger et al., 2025). Therefore, appropriate interpretation of wearable HRV benefits from context-controlled analysis and an emphasis on within-subject dynamics (Stritzelberger et al., 2025).

Finally, recent studies emphasize that wearable-based analyses must explicitly address several methodological pitfalls, including non-wear intervals, informative missingness, motion artifacts, and Clever Hans effects in which models learn incidental device- or usage-related cues rather than underlying physiology, as discussed by (Dumkrieger et al., 2025; Faust et al., 2025). To mitigate these risks, recommended practices include explicit wear-state detection, often operationalized using skin-temperature thresholds in the range of 27 to 45 °C, contextual segmentation into sleep and activity states, and systematic robustness checks across aggregation and smoothing choices (Faust et al., 2025; Abromavičius et al., 2023). In addition, explainable AI is increasingly applied as a validation layer to verify that model decisions are driven by physiologically plausible signal structure, for example, EDA peak morphology, rise dynamics, and recovery behavior, rather than statistical artifacts (Moser et al., 2024; Kim et al., 2025; Ashina et al., 2024). Collectively, these directions motivate a data-centric methodology in which data quality, contextual validity, and biomarker interpretability are treated as primary determinants of reliable inference (Faust et al., 2025; Stritzelberger et al., 2025).

1.3. Filtering and Preprocessing of Physiological Signals

Filtering constitutes a fundamental preprocessing step in the analysis of physiological and biomedical signals, as unfiltered noise and artifacts may distort signal characteristics and compromise the reliability of subsequent clinical or analytical interpretations (Chaitanya & Sharma, 2022). In medical research, filtering is routinely applied to physiological recordings to suppress motion artifacts, sensor-related disturbances, and high-frequency noise, thereby enhancing the interpretability of signals derived from the ANS, cardiovascular activity, and neural processes (Moser et al., 2024; Ohal & Mantri, 2024; Vu et al., 2025).

Among commonly employed filtering techniques, median filtering is particularly effective for suppressing impulsive artifacts. As a nonlinear method, it replaces each data point with the median of its local neighborhood, enabling robust removal of abrupt outliers caused by motion or sensor instability while preserving sharp signal transitions (Feng et al., 2018; Ohal & Mantri, 2024). This property makes median filtering suitable for physiological signals, such as EDA and EEG, where isolated spikes may otherwise bias feature extraction.

The Savitzky–Golay filter provides an alternative smoothing approach when preservation of signal morphology is critical. By fitting local subsets of the signal with low-degree polynomials, it reduces noise while maintaining peak height, width, and waveform shape, which is essential for subsequent biomarker extraction in EEG, ECG, and optical biosignals (Agarwal et al., 2017; Chaitanya & Sharma, 2022; Liu & Li, 2021; Romo-Cardenas et al., 2018).

Low-pass Butterworth filters are widely used to attenuate high-frequency noise components, including electronic interference and muscle-related artifacts. Owing to their smooth amplitude–frequency response, these filters are well-suited to isolating slow physiological dynamics, such as tonic components of electrodermal activity and gradual autonomic fluctuations reflected in cardiovascular signals (Bota et al., 2024; Ohal & Mantri, 2024; Vu et al., 2025). In migraine research, such filtering has been applied to enhance the clarity of physiological trends associated with pre-ictal and ictal processes (Balam, 2025; van den Hoek et al., 2024).

In addition to these methods, a variety of other filtering approaches have been reported in biomedical signal processing. Spectral filters, including notch, high-pass, band-pass, and Chebyshev type II filters, are commonly applied to suppress power-line interference, baseline drift, and motion-related artifacts in ECG, EEG, and photoplethysmographic signals (Balam, 2025; Campanella et al., 2024; Chaitanya & Sharma, 2022; Vu et al., 2025). More advanced model-based and adaptive techniques, such as Kalman filtering, empirical mode decomposition, circulant singular spectrum analysis, and eigenvalue-based decompositions, have also been explored for separating noise from signal components in ECG and multimodal recordings (Chaitanya & Sharma, 2022; Liu & Li, 2021; Simon, 2025). Additionally, alternative denoising approaches, including binomial and non-local mean filtering, have been proposed to reduce stochastic noise while preserving morphological signal characteristics (Agarwal et al., 2017; Liu & Li, 2021).

In biomedical applications, the choice of filtering method is dictated by the dominant noise characteristics of the signal under investigation. Median filtering effectively mitigates motion-related artifacts, Savitzky–Golay filtering preserves diagnostically relevant waveform features, and Butterworth low-pass filtering facilitates the extraction of slow physiological trends (Agarwal et al., 2017; Chaitanya & Sharma, 2022; Ohal & Mantri, 2024). These complementary properties provide a methodological basis for experimentally evaluating the impact of filtering on derived physiological features and for comparing filtered and unfiltered biomarkers in predictive analyses.

1.4. Feature Extraction and Engineering for Physiological Signals

Effective feature extraction is a critical step in transforming physiological recordings into informative, stable representations suitable for predictive modeling in migraine and related neurological conditions. In migraine research, extracted features are designed to capture subtle autonomic, cortical, and behavioral changes across different phases of the migraine cycle while reducing noise, inter-individual variability, and sensor-related artifacts (Kapustynska et al., 2024; Menon et al., 2022; Siirtola et al., 2018).

Across migraine and broader neurological literature, feature extraction strategies span multiple signal modalities, including electroencephalography, cardiovascular signals, electrodermal activity, skin temperature, accelerometry, electronic diaries, and neuroimaging. A structured overview of commonly extracted features and their corresponding signal sources is provided in Table 1.1. The literature consistently emphasizes spectral EEG features such as band power and coherence to characterize cortical excitability (Balam, 2025; van den Hoek et al., 2024), heart rate variability indices to quantify autonomic regulation (Stritzelberger et al., 2025; Vu et al., 2025), and tonic–phasic decomposition of electrodermal activity to capture sympathetic arousal dynamics (Stritzelberger et al., 2025; Vu et al., 2025).

Table 1.1. Mapping of extracted features in migraine and sleep/stress-related studies

Feature category	Signal	Features reported
Time-domain statistical features	EDA, HR, SkinTEMP, Acc	Mean, median, minimum, maximum, standard deviation, RMS (Siirtola et al., 2018; Stritzelberger et al., 2025; Vu et al., 2025)
Higher-order distributional features	EDA, HR, SkinTEMP	Skewness, kurtosis (Liu & Li, 2021; Ohal & Mantri, 2024)
Peak-based and extreme-value features	EDA, HR, Acc	Peak amplitude, peak count, maximum deviation (Balam, 2025; Vu et al., 2025)
Variability-related features	HR / BVP	HRV metrics (RMSSD, SDNN), coefficient of variation (Gomes et al., 2019; Stritzelberger et al., 2025)
Energy-related features	Acc, EEG, EDA	Signal energy, spectral energy (Balam, 2025; van den Hoek et al., 2024)
Shape and ratio descriptors	Multimodal	Crest factor, shape factor, signal ratios (Baliarsingh et al., 2023; Ohal & Mantri, 2024)
Activity-related aggregates	Acc	Activity variance, movement intensity, MET proxies (Harmsen et al., 2025; Siirtola et al., 2018)
Temporal aggregation over windows	All wearable signals	Sliding windows (10–60 s), nocturnal aggregation, day/night summaries (Faust et al., 2025; Harmsen et al., 2025)

In wearable-based migraine studies, feature extraction predominantly relies on time-domain and frequency-domain characteristics derived from HR, HRV, SkinTEMP, EDA, and Acc (Gonzalez-Martinez et al., 2022; Siirtola et al., 2018). These features are typically computed within fixed or sliding temporal windows and aggregated over daily or nocturnal periods, reflecting the constraints of long-term, real-world monitoring and the temporal structure of migraine prodromes (Faust et al., 2025; Harmsen et al., 2025). Importantly, several studies highlight the added value of integrating subjective diary-derived features, such as pain intensity and stress exposure, with objective physiological metrics (Houle et al., 2017; Kellier et al., 2023).

Recent studies increasingly emphasize that the temporal scale of feature aggregation is an independent methodological factor that can materially affect the detectability of short-lived physiological dynamics. Long aggregation windows may dilute transient responses by averaging informative and non-informative segments within the same interval, whereas short windows can preserve temporally localized changes. In autonomic assessment, five-minute frames are widely treated as a standardized and sufficient unit for reliable heart-rate-variability estimation, as reflected in international guidance and explicitly discussed in wearable-validation work (Stritzelberger et al., 2025). Conversely, for electrodermal activity, extremely short windows have been justified when the objective is to capture a complete phasic response cycle, including latency, peak rise, and recovery (Moser et al., 2024). Evidence from multi-sensor stress-monitoring research further indicates that five-minute or shorter segments can improve sensitivity to acute physiological changes compared with coarse hourly aggregation, supporting the view that temporal resolution should be systematically evaluated (Abromavičius et al., 2023).

Although a wide range of candidate features has been proposed, existing evidence indicates that predictive performance depends not only on feature type but also on preprocessing choices, temporal aggregation strategies, and feature selection procedures (Kapustynska et al., 2024; Menon et al., 2022).

These observations motivate a systematic evaluation of feature extraction pipelines for migraine prediction, including comparisons across different settings and assessments of feature relevance within machine learning frameworks, as undertaken in the present research.

1.5. Machine Learning Approaches for Migraine Prediction

Machine learning (ML) approaches have been increasingly applied to migraine prediction to shift disease management from reactive treatment toward proactive intervention (Danelakis et al., 2025; Houle et al., 2017). Existing studies explore a wide range of classifiers and data modalities, including subjective headache diaries, physiological signals from wearable sensors, environmental factors, and, in some cases, neuroimaging data (Gonzalez-Martinez et al., 2022; Siirtola et al., 2018; Stubberud et al., 2023).

Classical ML algorithms remain the most commonly used methods in migraine prediction, largely due to their robustness on small datasets and their relative interpretability. Among these, Support Vector Machines (SVMs), Random Forests (RFs), and gradient boosting methods such as XGBoost have been widely employed.

SVM-based models have been applied to both diagnostic and predictive tasks using EEG-derived features, clinical variables, and headache diary data (Kim et al., 2025; Messina et al., 2023). When provided with informative handcrafted features, SVMs have demonstrated competitive performance in distinguishing migraine subtypes and predicting pain intensity, although their effectiveness depends strongly on feature selection and class balance (Balam, 2025; Pagán et al., 2016).

Random Forest (RF) models are frequently used as baseline or benchmark classifiers in migraine prediction studies. Their ability to model nonlinear feature interactions and to provide feature-importance estimates makes them particularly suitable for multimodal datasets that combine subjective reports and physiological measurements. For example, RF classifiers have been used to predict next-day migraine occurrence from diary entries and wearable sensor data, achieving moderate predictive performance, with reported AUC values typically ranging from 0.62 to 0.73 (Houle et al., 2017; Stubberud et al., 2023). These results highlight both the feasibility and the current limitations of ML-based forecasting in real-world migraine cohorts.

Gradient boosting approaches, including XGBoost, have been applied to more heterogeneous data sources, such as questionnaire data, environmental variables, and neuroimaging-derived features. In selected diagnostic settings, these models have achieved high classification accuracy for migraine subtypes, although such results are often obtained in controlled or cross-sectional datasets rather than in longitudinal prediction scenarios (Danelakis et al., 2025; Katsuki et al., 2023).

A key methodological characteristic of classical ML studies is the frequent use of personalized models, trained on data from individual patients. This approach reflects the pronounced inter-individual variability in migraine triggers and prodromal patterns. In contrast, generalized models trained across participants often show substantially reduced performance, sometimes approaching chance level, particularly when datasets are small or physiologically heterogeneous (Buse et al., 2020; Danelakis et al., 2025; Siirtola et al., 2018).

More recent studies have explored deep learning approaches, including convolutional neural networks (CNNs) and recurrent architectures such as long short-term memory (LSTM) networks. These models are primarily applied to raw or minimally processed physiological time series, including ECG, blood volume pulse, accelerometry, and EEG signals (Chiang et al., 2022; Faust et al., 2025; Gonzalez-Martinez et al., 2022). Deep learning models offer the advantage of automatic feature learning and the ability to capture temporal dependencies in sequential data. For instance, personalized LSTM models trained on wearable sensor signals have demonstrated high sensitivity for predicting migraine onset over short prediction horizons, such as two hours, in small cohorts of episodic migraine patients (Gonzalez-Martinez et al., 2022). However, these promising results are typically obtained in highly individualized settings and do not readily generalize to broader populations.

Generalized deep learning models are more commonly reported in studies with access to larger datasets, particularly in cardiovascular or neurological comorbidity research, where CNN- or LSTM-based architectures have achieved high AUC values (Chiang et al., 2022). In migraine-specific applications, however, the use of deep learning is constrained by limited sample sizes, class imbalance, and the lack of external validation (Danelakis et al., 2025; Faust et al., 2025).

Across both classical and deep learning approaches, several recurring methodological challenges are evident. Most migraine prediction studies rely on small cohorts (often fewer than 40 participants), which limits statistical power and increases the risk of overfitting (Faust et al., 2025; Stubberud et al., 2023). Class imbalance, with substantially more headache-free days than migraine days, further complicates model training and evaluation (Butt et al., 2025). In addition, deep learning models suffer from limited interpretability, which remains a critical concern in clinical decision-making contexts (Danelakis et al., 2025; Moser et al., 2024).

Taken together, the literature suggests that while deep learning approaches hold promise for capturing complex temporal patterns, classical ML models combined with carefully engineered features remain the dominant and most reliable choice for migraine prediction under current data constraints. The strong dependence of predictive performance on individual-specific patterns underscores the importance of personalized modeling strategies. It motivates experimental designs that explicitly compare filtered and unfiltered physiological features within subject-specific frameworks.

1.6. Conclusions of the First Chapter and Formulation of the Dissertation Tasks

Based on the conducted literature review, the following conclusions can be drawn.

1. Wearable multi-sensor physiological recordings provide access to autonomic and behavioral signals that may reflect pre-migraine changes. Cardiovascular measures, electrodermal activity, skin temperature, and activity-related variables are among the most frequently studied modalities. However, reported effect sizes and predictive performance vary substantially across studies and cohorts.
2. Strong inter-individual heterogeneity is consistently reported. Many studies demonstrate improved outcomes with participant-specific or personalized modeling approaches, suggesting that migraine-related physiological patterns differ considerably across individuals.
3. Comparability across studies remains limited due to heterogeneous problem formulations and labeling strategies. Differences in prediction horizon, migraine-onset definition, handling of post-attack intervals, and construction of control segments yield performance metrics that often address different clinical questions.
4. Publicly available wearable migraine datasets and fully reproducible analysis pipelines remain scarce. Most studies report aggregated results without providing raw wearable recordings or end-to-end processing code, limiting reproducibility and making transferability to independent datasets difficult to assess.
5. The influence of preprocessing and temporal segmentation is insufficiently explored in the literature. Signal filtering, smoothing, and the choice of analysis-frame duration can substantially alter feature distributions and model performance, yet these factors are rarely evaluated systematically and in a controlled manner.
6. Although personalized modeling approaches often outperform population-level models, the specific contributions of temporal segmentation, preprocessing, and feature design to predictive performance remain insufficiently disentangled.

Three hypotheses were formulated as a result of the performed literature survey:

1. The nocturnal regime, particularly when restricted to sleep intervals, provides a more stable physiological context than daytime recordings and is therefore expected to facilitate the detection of pre-migraine changes in wearable data.
2. The duration of the analysis window affects the detectability of pre-migraine patterns, as different temporal scales may either preserve short-

lived physiological deviations or attenuate them through temporal aggregation.

3. Signal preprocessing influences the stability and informativeness of extracted features; however, the direction and magnitude of this effect are not well established and require systematic evaluation across different temporal resolutions and modeling settings.

To test the proposed hypotheses, the following dissertation tasks were formulated.

1. To construct a reproducible wearable-data processing pipeline for migraine analysis, including participant-level dataset formation, definition of nocturnal observations, and consistent labeling of pre-migraine and control periods.
2. To develop and evaluate classification models for distinguishing pre-migraine and control states using wearable-derived physiological signals, and to quantify the effect of analysis-frame duration on classification performance.
3. To evaluate the influence of key methodological design choices, including signal-level preprocessing, and to identify physiologically informative signals and features associated with pre-migraine changes using statistical feature analysis and model-based relevance measures.
4. To assess the feasibility of night-level migraine risk estimation from nocturnal wearable data using aggregated frame-level information.

2

Dataset Characterization, Quality Control, and Temporal Segmentation

This chapter describes the structure and quality of the collected wearable dataset and provides an exploratory analysis of its main characteristics. The goal is to establish whether the data are sufficiently stable, complete, and physiologically coherent to support subsequent segmentation, feature extraction, and predictive modeling.

The chapter begins with an overview of the research design, participants, and device configuration. It then describes data export, structuring, labeling, and quality-control procedures. Next, signal distributions and circadian patterns are examined to assess inter-individual variability and the stability of nocturnal recordings. Finally, the effect of nighttime selection and temporal segmentation on data stability and feature behavior is evaluated. The research results presented in this chapter have already been partially demonstrated and discussed at the international conference "DAMSS" (Kapustynska et al., 2024) and published in (Kapustynska et al., 2024).

2.1. Research Overview and Analytical Workflow

This section presents the analytical workflow applied to the physiological data prior to further analysis. It summarizes the main stages of data handling, including data collection, cloud storage, data export and organization, and subsequent preprocessing, which precede segmentation, filtering, feature extraction, and machine learning experiments.

The complete data preparation workflow is illustrated in Figure 2.1.

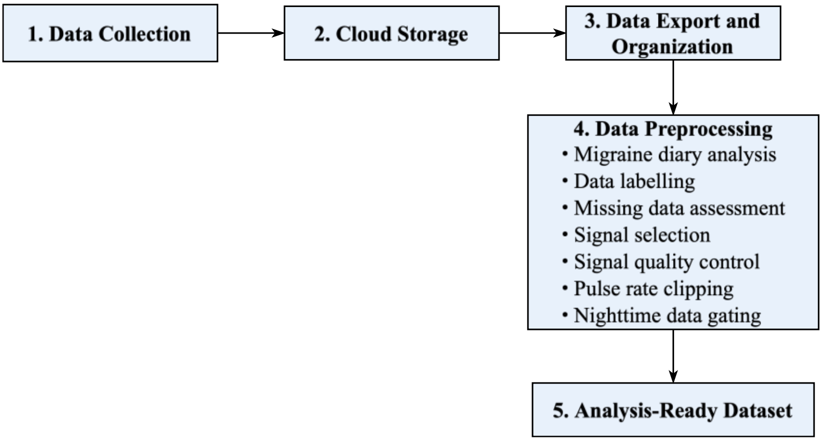


Fig. 2.1. Data preparation workflow

The data processing and analysis pipeline was implemented in Python using open-source scientific computing libraries.

2.2. Participants and Inclusion Criteria

Nineteen individuals with migraine were enrolled in this research according to specific inclusion criteria to assess the applicability of wearable biosensor technology to predict migraine attacks by monitoring changes in the ANS during the prodrome phase. All participants were instructed to wear the Empatica Embrace Plus device on the non-dominant wrist and to continue wearing it until at least three migraine episodes had been documented. The monitoring period per participant was heterogeneous, with an average duration of 31.4 ± 14.5 days. The demographic characteristics and the corresponding number of monitoring days per participant are provided in Table 2.1.

Table 2.1. Demographic information and monitoring days of study participants

Subject	Age	Sex	Migraine type	Migraine duration (years)	Monthly migraine headache days	Monitoring days
1	30	M	Without aura	5	7	7
2	29	M	With and without aura	8	6	21
3	33	F	Without aura	23	14	52
4	29	M	Without aura	23	5	33
5	22	F	Without aura	10	8	33
6	31	F	Without aura	10	12	68
7	39	F	Without aura	5	6	20
8	29	F	Without aura	20	5	40
9	62	M	Without aura	29	7	25
10	26	M	Without aura	9	6	23
11	62	M	Without aura	28	7	34
12	32	F	Without aura	5	6	49
13	35	M	Without aura	8	12	20
14	40	F	With and without aura	28	8	20
15	29	F	Without aura	4	4	22
16	49	F	Without aura	6	11	48
17	31	M	Without aura	12	5	27
18	27	F	With and without aura	11	10	30
19	27	F	Without aura	22	10	24

Note: F - Female, M – Male.

To minimize external influences on ANS variability and to ensure data reliability, participants were enrolled according to clearly defined inclusion and exclusion criteria. These criteria were designed to promote consistency in physiological responses and to control for potential confounders, including medication use and comorbid conditions. The recruitment process, medical eligibility assessment, and informed consent procedures were supervised by a board-certified neurologist to ensure compliance with clinical and ethical standards.

Participants were eligible for inclusion if they were 18 years of age or older, had a diagnosis of episodic migraine with or without aura according to the International Classification of Headache Disorders, 3rd edition (ICHD-3), reported at least four migraine attacks per month, and were able to understand and follow the research procedures in Lithuanian.

Exclusion criteria included pregnancy or breastfeeding, a diagnosis of chronic or hemiplegic migraine, other primary headache disorders (except infrequent episodic tension-type headache), the use of preventive migraine therapy, or the use of medications known to affect the autonomic nervous system, including antidepressants, anticholinergics, antihypertensives, antiepileptics, opioids, and muscle relaxants. Participants with other chronic pain conditions were also excluded.

2.3. Device and Physiological Signals

The Empatica Embrace Plus wearable device (Empatica Inc., Milan, Italy; Fig. 2.2) was selected for this research for its ability to continuously record raw physiological data over extended periods (up to 14 days) and to provide clinically validated digital biomarkers relevant to ANS activity.

Fig. 2.2. Empatica Embrace Plus wristband used for data collection

The Embrace Plus is certified as a Class IIa medical device under the European Medical Device Regulation (MDR 2017/745) and is approved for use in clinical investigations. This classification ensures that all physiological signals are acquired using medically validated sensors and signal-processing algorithms approved for medical use. Data were accessed via the Empatica Health Monitoring cloud Platform and subsequently used for data preparation.

In this research, digital biomarkers obtained from the Empatica Embrace Plus were used to capture a range of physiological parameters relevant to ANS dynamics and physical activity. The available raw and derived signals, along with their physiological interpretation, measurement units, and estimation principles, are described below.

1. Accelerometer magnitude standard deviation (Acc): Average standard deviation of the magnitude of the 3-axis acceleration signal, measured in gravitational units (g). Reflects physical movement and body orientation (Maczák et al., 2021).
2. Activity counts (ACT): Continuous estimation of movement intensity, derived from accelerometer data, expressed in arbitrary units (au).
3. Electrodermal activity (EDA): Continuous monitoring of the level of skin conductance, indicative of activation of the sympathetic nervous system, measured in microsiemens (μS) (Shaffer et al., 2016).
4. Metabolic equivalent of task (MET): Continuous estimation of energy expenditure during physical activity, expressed in metabolic equivalents (MET) (Lustrek et al., 2012).
5. Pulse rate variability (RMSSD): Intermittent monitoring of pulse rate variability, calculated as the root mean square of successive differences between consecutive systolic peaks, measured in milliseconds (ms) (Li et al., 2023; Zhang et al., 2024).
6. Pulse rate (PR): Continuous monitoring of heart rate, measured in beats per minute (bpm) (Sanches et al., 2023).
7. Respiratory rate (RR): Intermittent monitoring of breathing rate, measured in breaths per minute (breaths/min), typically derived from PPG and accelerometer signals (Kozumplík et al., 2021; Meredith et al., 2012).
8. Sleep detection stage (SDS): Classification of sleep periods and transitions (e.g., onset of sleep, awakening) based on activity and physiological signal patterns (Ahmadi et al., 2020; Radha et al., 2021).
9. Step counts: Continuous monitoring of the number of steps taken, measured in steps, derived from accelerometer data.
10. Skin temperature (SkinTEMP): Continuous monitoring of peripheral skin surface temperature, measured in degrees Celsius (°C).
11. Wearing detection: Estimation of device wear compliance, reported as the proportion of time the device is worn, expressed as a percentage (%).

2.4. Data Export, Organization, and Quality Control

This section describes the procedure for exporting physiological recordings from the Empatica Health Monitoring cloud Platform and organizing them into a structured format suitable for subsequent analysis. Recordings were exported as daily measurement streams and subsequently consolidated into unified data files for each participant.

Because the exported data were distributed across multiple recording sessions and device-specific folders, an automated organization procedure was applied to identify recordings, associate device identifiers with participants, and merge temporally aligned streams into continuous monitoring records.

The complete workflow is summarized in Algorithm 1.

Algorithm 1. Empatica data export, consolidation, and clinical cross-check procedure

Input: Empatica export directory *R* organized into date-wise folders; participant–device mapping *M* linking device identifiers to participant identifiers; diary labels *L*; clinician summary *C*
Output: Participant-level spreadsheet files $\{X_p\}$ and cross-check summary table *T*
1: Initialize an empty data structure *D* for participant-level recording sessions.
2: Automatically scan directory *R* to locate all exported recording folders.
3: **for detected recording folders do**
    Extract recording date and device identifier from the folder path.
    Verify participant identifier using mapping *M*.
    if the participant identifier is not found **then**
        Skip the folder and record an inconsistency.
    Load all CSV streams contained in the folder.
    Set *timestapm_iso* as the temporal index.
    Merge streams into a time-aligned session table.
    Remove auxiliary columns and duplicated streams.
    Append the session table to participant records *D[p]*.
    end for
4: **for all participants *p* in *D* do**
    Concatenate all sessions chronologically.
    Standardize temporal and annotation fields.
    Export consolidated records as a spreadsheet X_p.
    end for
5: Initialize cross-check summary table *T*.
6: **for all participants *p* do**
    Determine monitoring start and end dates.
    Compute the number of monitoring days.
    Count migraine attacks using diary labels.
    Add summary statistics to table *T*.
    end for
7: Compare table *T* with the clinician-provided summary *C*.
8: Report inconsistencies for verification.
9: return $\{X_p\}$, *T*.

Consolidating the recordings into a standardized structure enables systematic quality control, allowing the monitoring period to be tracked consistently, cohort- and participant-level completeness to be quantified, and the influence of subsequent preprocessing steps on data availability and stability to be evaluated.

After consolidation, 19 spreadsheets (.xlsx) were created, one per participant. Each file contains a standardized temporal index (*timestamp_datetime*), an annotation column (*Labels*), and a set of numeric streams corresponding to the device digital biomarkers and contextual quality signals, including Acc (*accelerometers_std_g*), ACT (*activity_counts*), EDA (*eda_scl_usiemens*), MET (*met*), PR (*pulse_rate_bpm*), PRV (*prv_rmssd_ms*), RR (*respiratory_rate_brpm*), SDS (*sleep_detection_stage*), step counts (*step_counts*), SkinTEMP (*temperature_celsius*), and wearing detection (*wearing_detection_percentage*).

The *Labels* column was obtained by digitizing participants' migraine diaries. A clinician verified the digitized diaries before further processing. A value of 0 indicates the absence of migraine, whereas a value of 1 indicates the presence of migraine.

2.4.1. Missingness and Signal Exclusion

Cohort-level missingness was then quantified for all biomarkers. The results indicate that RR and PRV contain a substantial proportion of missing values, with participant-level averages typically exceeding 70%. This level of incompleteness reduces their comparability and statistical reliability in cross-subject analyses.

Summary statistics of missingness for the signals across all 19 participants are reported in Table 2.2.

Table 2.2. Percentage of missing values across signals and all participants

Signal	Missing values, %
Accelerometers	12.37
Activity counts	12.37
Electrodermal activity	12.37
Metabolic equivalent of task	12.37
Pulse rate variability	**75.05**
Pulse rate	12.37
Respiratory rate	**72.47**
Sleep detection stage	12.37
Step counts	12.37
Skin temperature	12.37
Wearing detection	10.74

Note: Boldface highlights signals with excessive missing data excluded from further analysis.

Notably, data collection was performed using four Empatica Embrace Plus devices across the cohort. The high missingness observed in RR and PRV was present across recordings from multiple devices, suggesting that it is not specific to a single unit.

For each participant, the total number of observations was recorded. Because the RR and PRV streams exhibited excessive missingness, they were excluded from further analyses. A complete-case filter was then applied to the remaining streams by removing rows with missing values in any retained signal. The resulting participant-level completeness summary is reported in Table 2.3.

Table 2.3. Summary of total, missing, and complete observations per participant

Subject	Rows total	Rows with NaN	Rows complete	Removed, %
1	8640	1728	6912	20.00
2	30240	6561	23679	21.70
3	38880	5020	33860	12.91
4	47520	3751	43769	7.89
5	47520	3445	44075	7.25
6	97920	21454	76466	21.91
7	28800	2283	26517	7.93
8	57600	14421	43179	25.04
9	36000	2415	33585	6.71
10	33120	2529	30591	7.64
11	48960	5250	43710	10.72
12	70560	9808	60752	13.90
13	28800	4198	24602	14.58
14	28800	2568	26232	8.92
15	31680	2531	29149	7.99
16	69120	12442	56678	18.00
17	38880	6680	32200	17.18
18	43200	2850	40350	6.60
19	34560	3014	31546	8.72

Overall, missingness after excluding RR and PRV was moderate and broadly comparable across participants, suggesting sufficient consistency in recording coverage for cohort-level analyses. At the same time, the remaining proportion of complete observations varied by participant, indicating that downstream comparisons should account for between-subject heterogeneity in effective data availability.

2.4.2. Circadian Profiles and Inter-Individual Variability

The physiological and activity signals were examined to contextualize inter-individual variability and to screen for potential artifacts. For each signal retained for analysis, two complementary visual summaries were generated.

First, cohort-level circadian profiles were computed as hourly summaries. For each participant, signal values were aggregated by hour of day over the entire monitoring period, and hourly distribution statistics were estimated. Cohort-level profiles were then formed by taking, at each hour, the median across participants so that each participant contributed equally. An uncertainty band was defined using the interquartile range, corresponding to the 25th and 75th percentiles across participants. This representation highlights common time-of-day structure while limiting the influence of outliers and uneven recording density.

Second, participant-level heterogeneity was summarized using per-participant boxplots computed from all available observations for each participant. These boxplots report the median and interquartile range, with extreme values shown as outliers, thereby providing a compact view of between-subject differences in signal level and dispersion.

Across the cohort, pulse rate exhibited a clear circadian structure. The cohort median was lower during nighttime hours, increased in the morning, remained relatively stable throughout the day, and gradually declined toward late evening, as shown in Figure 2.3.

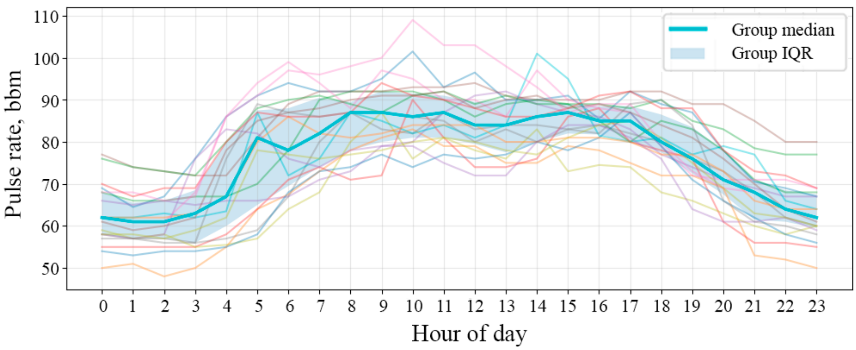


Fig. 2.3. Circadian pulse rate profile across the cohort

This circadian profile indicates that physiologically plausible diurnal dynamics were preserved in the recordings and suggests that the observed variability cannot be explained solely by noise or differences in data availability. Participant-level boxplots further revealed substantial inter-individual heterogeneity in both the central tendency and dispersion of PR, as shown in Figure 2.4, highlighting the need for normalization and explicit quality-control procedures in cross-subject comparisons.

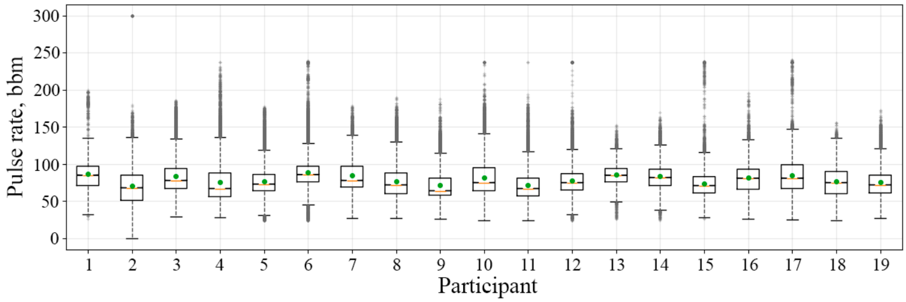


Fig. 2.4. Inter-individual distribution of pulse rate

Finally, the presence of occasional extreme outliers with implausibly high PR values, as shown in Figure 2.4, is consistent with PPG-related artifacts and or estimation errors during motion. Therefore, outlier handling using physiologically constrained rules, such as PR clipping, should be applied before feature extraction and model training to improve the stability and interpretability of downstream results.

Across the cohort, SkinTEMP exhibited a clear circadian pattern. The cohort-level median was higher during the late evening and night, decreased toward the morning, and then gradually increased again toward the evening, as shown in Figure 2.5.

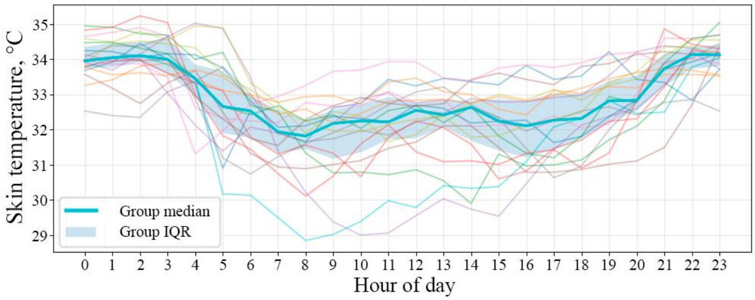


Fig. 2.5. Circadian skin temperature profile across the cohort

This pattern is consistent with physiologically plausible peripheral thermoregulation and indicates that the recordings preserve meaningful diurnal organization. The cohort interquartile range varied across the day, with broader between-subject dispersion typically observed during daytime compared to nocturnal hours, suggesting that inter-individual variability is not uniform and may be amplified by naturalistic contextual factors, including day-to-day differences in physical activity, ambient environmental conditions, and variability in device–skin contact stability. Participant-wise boxplots further demonstrated systematic differences in baseline SkinTEMP levels, with median shifts across individuals, as shown in Figure 2.6, underscoring the need for normalization and robust preprocessing in cross-subject analyses.

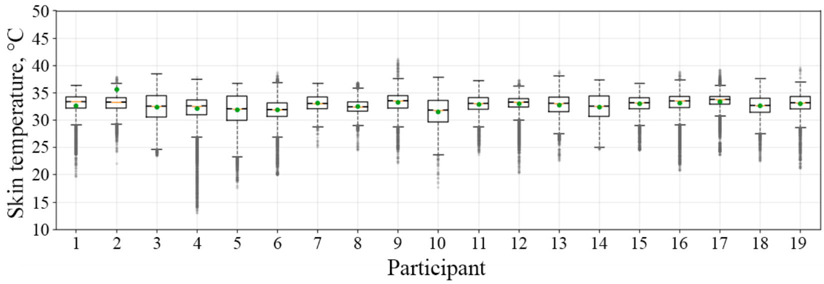


Fig. 2.6. Inter-individual distribution of skin temperature

Finally, occasional extreme values visible in the boxplots likely reflect measurement artifacts, such as transient loss of sensor contact or brief exposure to external temperature changes, and therefore, warrant explicit quality-control steps, including artifact handling and exclusion of unreliable intervals, before feature extraction and model training.

For the EDA visual analyses, a logarithmic transformation using log1p was applied before aggregation and plotting. This choice was motivated by the strongly right-skewed, heavy-tailed distribution of EDA values and by frequent high-amplitude excursions in the raw recordings. As illustrated by the participant-level boxplots in Figure 2.7, EDA exhibited pronounced upper tails and numerous extreme observations across participants, consistent with event-like sympathetic activations and, in part, potential measurement artifacts.

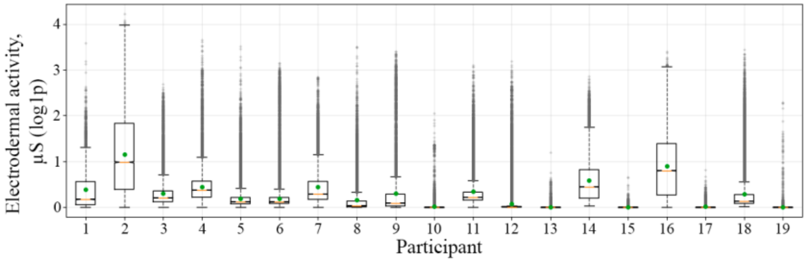


Fig. 2.7. Inter-individual distribution of electrodermal activity

Under these conditions, the log1p transform attenuated the disproportionate influence of rare high-amplitude peaks, reduced heteroscedasticity, and enabled more interpretable cohort-level summaries while preserving informative high-response episodes.

In contrast to pulse rate and skin temperature, which exhibited clear and physiologically plausible circadian organization at the cohort level in Figs. 2.4–2.7, EDA did not yield a smooth diurnal pattern. Instead, the cohort median remained comparatively low and only weakly dependent on time of day, whereas the interquartile range and participant-specific profiles revealed substantial between-subject heterogeneity and intermittent elevations, as shown in Figure 2.8.

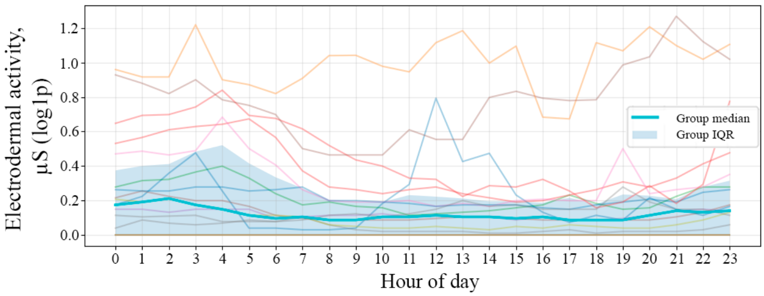


Fig. 2.8. Circadian electrodermal activity profile across the cohort

Therefore, EDA variability was governed less by a stable diurnal pattern and more by pronounced inter-individual differences and intermittently occurring excursions. This finding supports the use of robust preprocessing prior to downstream inference, including log transformation, participant-wise normalization, and explicit quality-control procedures, because analyses performed on the untransformed EDA scale may be unduly influenced by a limited number of high-amplitude episodes and participants with systematically elevated levels.

ACT displayed a well-defined circadian organization: values were near zero during nocturnal hours, increased after morning onset, remained elevated throughout the daytime period, and declined toward late evening, as shown in Figure 2.9. This profile indicates that the activity stream primarily reflects structured behavioral dynamics under unconstrained real-world monitoring conditions.

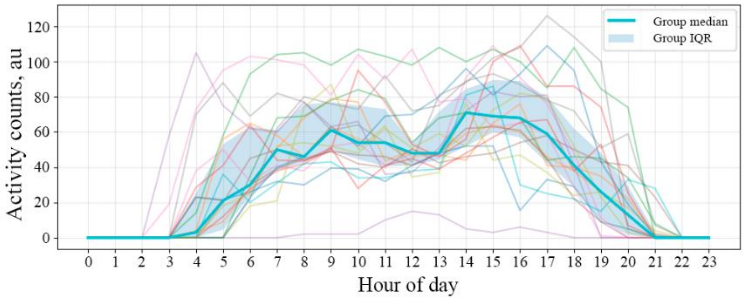


Fig. 2.9. Circadian activity counts profile across the cohort

Inter-participant variability in ACT was strongly time-dependent. The cohort IQR was minimal during nocturnal hours and maximal during daytime hours, which is consistent with activity being modulated by individual routines and heterogeneous daily demands, as well as by contextual influences such as mobility patterns and environment. This point is important for interpretation because daytime context is expected to amplify between-subject differences, thereby reducing comparability across participants and complicating cohort-level modeling when data are pooled without additional stratification or normalization.

Participant-level boxplots further confirmed substantial heterogeneity in both the central tendency and dispersion of ACT across subjects, as shown in Figure 2.10. Some participants exhibited wide distributions with pronounced upper tails, indicating high within-subject variability over the monitoring period, whereas others showed more compact ranges, consistent with more stable daily activity patterns or less variable exposure to activity-inducing contexts.

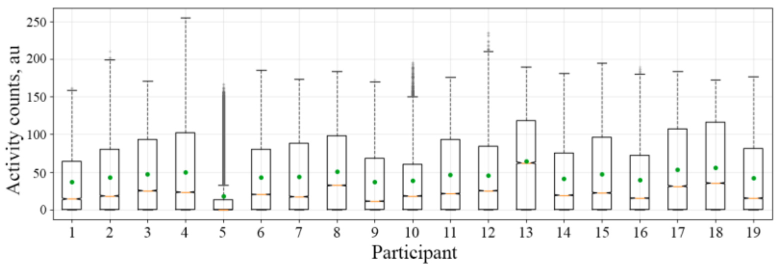


Fig. 2.10. Inter-individual distribution of activity counts

In addition, ACT distributions were typically right-skewed and contained occasional extreme peaks. This pattern supports the use of participant-wise normalization and robust artifact handling prior to feature extraction, particularly in cross-subject analyses where a small number of high-activity episodes can otherwise dominate summary statistics and downstream learning signals.

For the accelerometer-derived variability measure Acc, a physiologically expected circadian organization was observed, as shown in Figure 2.11. Values were close to zero during nocturnal hours, increased after morning onset, remained elevated with pronounced fluctuations throughout daytime, and decreased toward late evening.

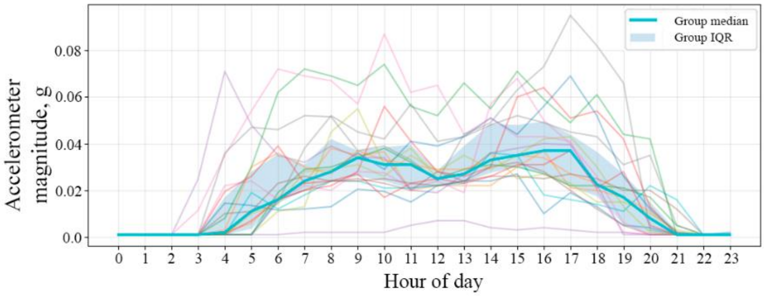


Fig. 2.11. Circadian accelerometer magnitude profile across the cohort

This pattern indicates that the Acc stream captures movement-related behavior and the sleep–wake regime. Inter-individual variability was not constant across the day. Nocturnal hours exhibited a more stable and constrained regime, whereas daytime variability increased, consistent with differences in daily routines, activity intensity, and naturalistic contextual factors, including occupational demands, commuting, and environmental constraints. Consequently, daytime intervals are expected to amplify cross-subject differences and reduce comparability unless participant-wise normalization and explicit quality control are applied.

Participant-level distributions of Acc, shown in Figure 2.12, further demonstrate substantial heterogeneity in both median level and dispersion. Some participants exhibited wide distributions with long upper tails, indicating greater within-subject variability over the monitoring period, whereas others showed more compact ranges, consistent with lower, more regular activity levels.

Rare high peaks and pronounced right tails were also evident in the participant-level distributions, as shown in Figure 2.12. Such behavior is typical for accelerometer-derived metrics and reflects episodic bursts of high-intensity movement. These properties motivate the use of robust scaling and explicit treatment of extreme values during preprocessing and feature extraction, particularly in cross-participant analyses where a small number of high-amplitude episodes can otherwise dominate cohort summaries and bias model fitting.

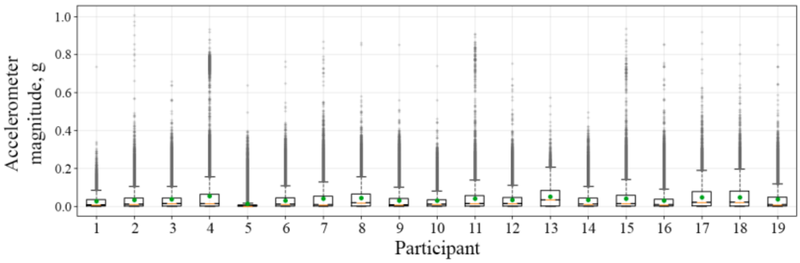


Fig. 2.12. Inter-individual distribution of accelerometer magnitude

At the cohort level, Acc remained consistently low during nocturnal hours, as shown in Figure 2.11, supporting prioritizing nocturnal segments for downstream modeling because the nighttime regime is more stable and less influenced by behavioral confounders, while daytime movement patterns introduce stronger contextual variability and inter-individual heterogeneity.

For the MET measure, an expected diurnal pattern was observed, consistent with its interpretation as a proxy for energy expenditure and activity intensity, as shown in Figure 2.13.

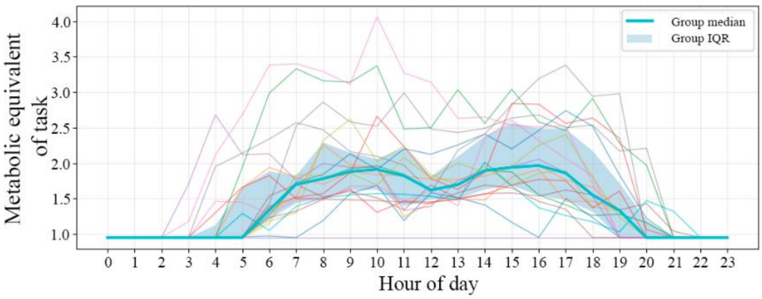


Fig. 2.13. Circadian metabolic equivalent of task profile across the cohort

During nocturnal hours, the cohort median remains close to the baseline level at approximately 1 MET, followed by a morning increase and a higher daytime plateau, with a gradual decline toward late evening, as shown in Figure 2.13. This profile indicates that the MET stream preserves physiologically plausible circadian organization and primarily reflects changes in behavioral regime.

Between-subject variability in MET is strongly time dependent. As shown in Figure 2.13, the interquartile range is minimal at night and increases substantially during daytime hours, as expected because daytime energy expenditure is shaped by individual routines, occupational demands, and broader contextual factors.

Consequently, daytime segments are likely to amplify cross-participant differences and, in pooled analyses, require normalization and robust preprocessing to ensure valid comparability.

Participant-level boxplots further confirm pronounced heterogeneity in MET distributions (Fig. 2.14). Participants differ both in their typical level and dispersion, and the distributions often exhibit long right tails with occasional high peaks. These properties motivate the use of participant-wise normalization and robust outlier-handling procedures before feature extraction and model training, particularly in cross-subject settings.

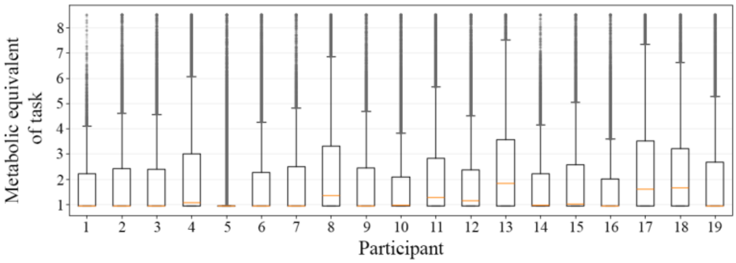


Fig. 2.14. Inter-individual distribution of metabolic equivalent of task

Such asymmetry is typical of activity-derived measures, which are dominated by low-to-moderate values interspersed with episodic bursts of higher intensity. This distributional shape supports robust scaling and the explicit handling of extreme observations during feature extraction, particularly in generalized modeling settings where between-subject differences can otherwise dominate the learning signal. Overall, the audited recordings preserved coherent circadian organization and sleep-related structure, while also exhibiting substantial inter-individual heterogeneity and occasional artifacts across multiple streams.

2.5. Data Labeling and Night Definition

This section focuses on the labeling of nocturnal segments to investigate whether differences exist between nights preceding migraine-free days and nights preceding migraine days. For this purpose, it was first necessary to define a consistent operational definition of nighttime sleep for subsequent labeling procedures.

2.5.1. Sleep Detection Stage and Sleep Gating

The SDS biomarker provided by the Empatica Embrace Plus device was used to identify sleep-related epochs. SDS represents a device-derived categorical state indicating whether the participant is classified as awake or in a sleep-related condition at a given time point. Table 2.4 summarizes the SDS states used in this research.

Table 2.4. Classification of the Sleep Detection Stage values

Value	Description
0	Wakefulness
101	Sleep – primary criterion for identifying valid sleep periods
102	Wake episodes occurring during sleep
300	Transient interruptions during sleep

In subsequent analyses, all non-zero SDS values were treated as sleep-related states. This rule enabled consistent identification of sleep-related intervals across participants.

To define nighttime sleep at the cohort level, the daily timing of non-zero SDS values was examined. For each participant, all timestamps were grouped by hour of day (0–23), and the share of samples with SDS not equal to zero was calculated for each hour. In Figure 2.15, this hourly profile is shown at the cohort level in two ways: (i) an average of the participant-specific hourly profiles, and (ii) a pooled estimate based on all samples combined, which gives more weight to participants with denser recordings. The results show that non-zero SDS values occur mainly at night and are rare during the day, supporting the use of SDS to identify nocturnal sleep intervals.

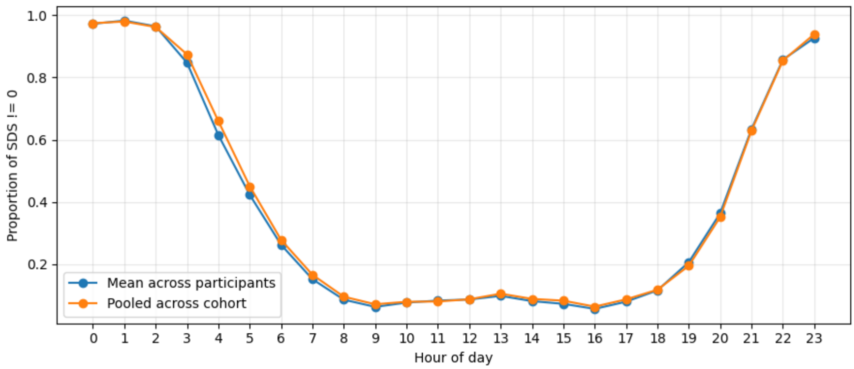


Fig. 2.15. Circadian profile of non-zero Sleep Detection Stage values

In addition, a participant-level heat map illustrating the hourly distribution of non-zero SDS values is shown in Figure 2.16. This visualization reveals inter-individual variability in sleep timing and transition periods while preserving a consistent circadian structure across the cohort. Most participants exhibited clearly defined nighttime sleep patterns, although several individuals demonstrated broader or shifted sleep intervals, indicating heterogeneity in sleep–wake behavior and recording conditions.

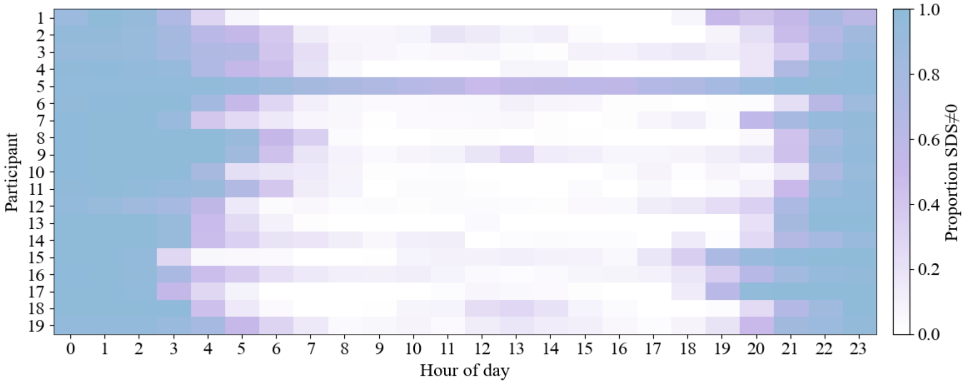


Fig. 2.16. Inter-individual variability in circadian SDS patterns

Based on these observations, nighttime sleep was operationally defined as intervals between 18:00 and 09:00 during which SDS values were non-zero. Nights following migraine days were excluded to minimize potential postdromal effects and to reduce deviations from baseline physiological conditions.

2.5.2. Class Counts Across Analysis Frames

After applying nighttime gating and labeling rules, the number of pre-migraine and baseline samples was quantified for each participant. Because this research further investigates the effect of dividing nocturnal sleep into analysis frames ranging from 5 to 120 minutes, Table 2.5 reports the number of pre-migraine and control samples obtained for each participant across analysis-frame durations.

Table 2.5. Class counts per participant and analysis frame duration (label 0/label 1)

Subject	5 min	10 min	30 min	60 min	90 min	120 min
1	72/269	37/137	13/49	7/27	5/18	4/15
2	1089/248	550/126	189/45	101/24	72/17	58/13

End of Table 2.5

Subject	5 min	10 min	30 min	60 min	90 min	120 min
3	1353/538	686/273	241/95	132/51	94/36	74/27
4	2671/293	1350/149	466/52	247/29	170/19	135/16
5	4202/470	2106/236	710/79	362/40	246/28	176/20
6	3746/692	1903/352	669/123	360/66	258/45	201/38
7	1141/420	578/212	199/73	103/38	74/28	55/20
8	2225/776	1124/392	394/138	208/72	146/52	110/40
9	1355/904	684/457	236/161	127/87	88/62	71/44
10	1121/501	568/253	197/89	103/48	78/35	58/28
11	2089/714	1057/363	368/125	193/66	139/47	110/37
12	2982/603	1510/304	527/105	278/54	200/40	158/30
13	628/517	319/261	110/92	58/49	42/35	33/28
14	960/589	485/300	167/106	88/55	61/40	48/28
15	1465/304	741/152	257/52	138/28	99/20	75/14
16	3486/321	1763/163	610/57	322/31	226/22	177/16
17	1171/904	591/458	202/160	107/85	77/61	58/45
18	1848/383	935/193	326/67	173/35	120/24	98/19
19	886/1119	444/566	154/196	83/105	57/77	44/57

2.6. Temporal Segmentation and Window-Level Analysis

Previous analyses showed that the number of nights associated with migraine or prodromal states remains limited in the available dataset. Consequently, the research operates under conditions of a relatively small number of positive observations and inherent class imbalance.

This analysis investigates how the choice of temporal analysis frame influences the statistical properties of physiological signals during nocturnal sleep and the feasibility of subsequent modeling.

2.6.1. Temporal Resolution and Signal Stability

The analysis was conducted on nocturnal recordings restricted to sleep intervals defined by non-zero SDS values. This restriction ensured that comparisons were performed within a physiologically consistent sleep regime.

For each night and each modality, recordings were partitioned into non-overlapping windows of 5, 10, 30, 60, 90, and 120 minutes.

The effect of window duration on signal stability is summarized in Figure 2.17, which presents cohort-level trends across six physiological modalities, such as PR, EDA, SkinTEMP, Acc, ACT, and MET. Stability was quantified using a robust coefficient of variation computed within each window and aggregated across participants.

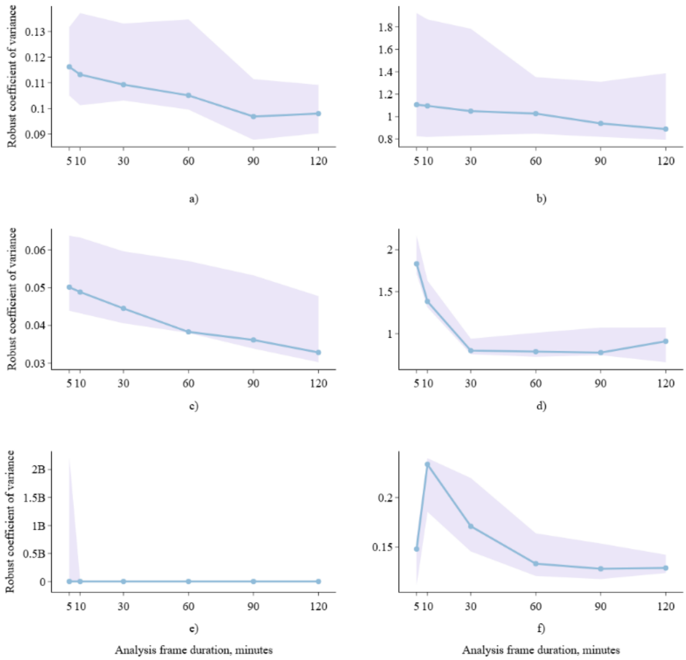


Fig. 2.17. Window-level stability across temporal resolutions during sleep: (a) pulse rate; (b) electrodermal activity; (c) skin temperature; (d) accelerometry magnitude; (e) activity counts; and (f) metabolic equivalent

As shown in Figure 2.17, short windows of 5–10 minutes preserve higher variability, reflecting short-term physiological fluctuations occurring during sleep. As window duration increases, variability decreases across most modalities, indicating stronger temporal averaging and smoother signal representations.

This transition becomes more pronounced beyond 30-minute windows, where signals exhibit more stable behavior and reduced sensitivity to transient changes. The effect is particularly strong in movement-related modalities, while physiological signals such as pulse rate and skin temperature demonstrate a more gradual stabilization.

These results indicate that window length represents a trade-off: shorter windows preserve temporal detail and provide more samples, whereas longer windows improve numerical stability but reduce data availability.

2.6.2. Data Availability and Signal-Level Differences

To evaluate the suitability of window-level representations for subsequent modeling, the analysis considered data availability and the presence of label-related differences.

Night recordings were segmented into non-overlapping windows of 5, 10, 30, 60, 90, and 120 minutes. For each window, a set of statistical descriptors was computed.

First, data availability was assessed. Data availability was defined as the proportion of participant–signal combinations for which statistical descriptors could be reliably computed at a given window length. A feature was considered valid only when at least 30 windows with finite values were available.

In general, high data availability was observed across most signals. However, as the window duration increased, availability gradually decreased due to fewer windows per night. This reduction reflects limited recording duration and implies that longer windows may favor participants with more complete data.

Next, label-related differences were examined at the signal level. For each signal and window length, the statistical descriptor providing the strongest separation between labels was identified. The strength of separation was quantified using the absolute value of Cliff's delta. Results are summarized in Figure 2.18.

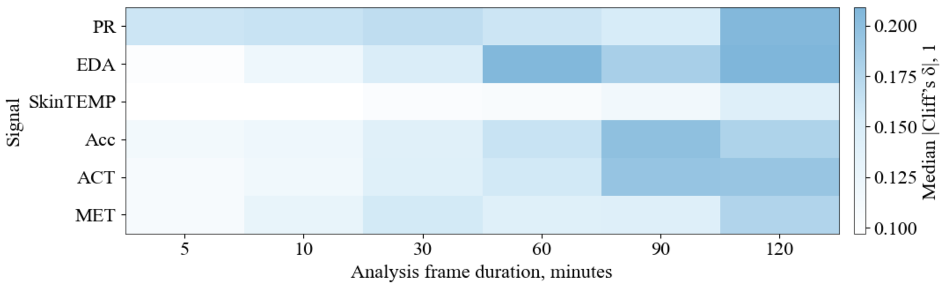


Fig. 2.18. Best label-separating feature by signal and window length

The results indicate that the degree of separation depends on both signal modality and window length. For several signals, larger differences appear at intermediate window durations, where sufficient data volume is retained while statistical estimates become less noisy.

No single signal demonstrates consistently strong separation across all window lengths, suggesting that informative patterns are signal-dependent and context-dependent.

Overall, the results show that the analysis frame duration affects both data availability and the detectability of label-related differences at the signal level. These findings provide a basis for selecting temporal resolution in subsequent modeling experiments.

2.7. Conclusions of the Second Chapter

1. Within the quality-control procedures, all Empatica Embrace Plus biomarkers were assessed for missingness. Most signals showed comparable and moderate incompleteness (~12.37%), whereas PRV and RR exhibited excessive missingness (>70%). After excluding PRV and RR, the remaining physiological and activity streams were sufficiently complete and internally consistent for cohort-level analysis. These retained signals also preserved physiologically plausible circadian organization, supporting their suitability for subsequent modeling.
2. The segmentation analysis showed that nighttime recordings are more stable than daytime data: within-segment variability is lower, and between-participant dispersion is reduced, suggesting that nighttime represents a more stationary physiological regime and therefore provides a more controlled setting for further analysis.
3. Further restricting nighttime data to sleep-related epochs (non-zero SDS) consistently reduced signal variability across most modalities, with the strongest stabilization observed in movement- and activity-related channels.
4. The analysis of temporal segmentation and signal-level differences indicates that the detectability of pre-migraine patterns depends on both the physiological modality and the chosen temporal resolution. No single signal or window length provides consistently strong differentiation, suggesting that informative patterns are context-dependent.

3

Experimental Framework and Evaluation

This chapter presents the experimental evaluation of the proposed framework for analyzing wearable physiological signals in migraine research. The experiments were designed to investigate the detectability of pre-migraine patterns in nocturnal data and to assess how methodological choices influence classification performance and the feasibility of risk estimation.

The evaluation was conducted in two stages: the first stage had an exploratory character and focused on initial analysis of classification performance under a less strict validation setting; and the second stage employed a stricter validation protocol based on independent nocturnal observations and was used to assess the robustness of the obtained results under more realistic conditions.

Across these stages, the experiments examined the effects of temporal segmentation, preprocessing, feature representation, and model choice. In addition to classification, the practical feasibility of estimating night-level migraine risk and predicting next-day outcomes from nocturnal wearable data was also evaluated (Kapustynska et al., 2024, 2025).

3.1. Objectives of the Experiment

The experimental research was designed to evaluate a unified methodological framework for migraine classification and prediction using wearable physiological signals. Due to the longitudinal nature of data collection and the cohort's gradual expansion, the evaluation was conducted in two phases.

Phase A included data from 10 participants and focused on exploratory individualized classification and statistical feature analysis. In this phase, nocturnal recordings were segmented into analysis frames of varying duration, and for each frame, a set of 78 statistical features was computed. Individualized binary classification was performed at the level of analysis frames derived from nocturnal recordings. Training and evaluation were conducted on frame-level samples, enabling comparisons across different analysis-frame durations and classifier configurations. In addition, one-way ANOVA was applied separately for each feature and each analysis-frame duration to identify class-related differences between pre-migraine and control nights. These analyses were previously reported in (Kapustynska et al., 2024, 2025).

Phase B expanded the cohort to 19 participants and introduced a stricter, more comprehensive evaluation framework. In this phase, classification analysis was conducted at the cohort level using frame-level splitting, followed by night-level splitting, treating each night as an independent observational unit. The feature space included 78 statistical features and 6 additional dynamic features. The effects of temporal segmentation, signal filtering, and feature representation on classification performance were systematically assessed on the expanded sample.

Phase B also included model-based feature analysis. Feature importance was evaluated across repeated runs, and the recurrence of top-ranked features was analyzed across classifiers and analysis-frame durations. In addition, a reduced feature subset was constructed from recurrent predictors and compared with the full feature set in order to assess the effect of feature-space reduction on model robustness and predictive performance.

In addition to cohort-level classification, Phase B included night-level estimation of migraine risk and individualized next-day prediction. In this setting, models were trained on analysis-frame-level feature vectors, while final night-level decisions were obtained by aggregating frame-level model outputs. Hyperparameter optimization was performed using RandomizedSearchCV with GroupKFold cross-validation, where grouping was defined by night identifiers to prevent information leakage between frames from the same night.

The objectives of the experimental research were defined as follows:

1. To develop and evaluate individualized and cohort-level classification models for distinguishing pre-migraine and control states using wearable

physiological signals, and to quantify the effect of analysis-frame duration on classification performance.

2. To evaluate the effect of signal filtering on classification performance.

3. To identify physiologically informative features and signals contributing to classification using feature-level statistical analysis and model-based relevance measures.

4. To assess the feasibility of night-level migraine risk estimation using aggregated frame-level information.

3.2. Experimental Design Overview

The experimental workflow for achieving the defined objectives is illustrated in Figure 3.1.

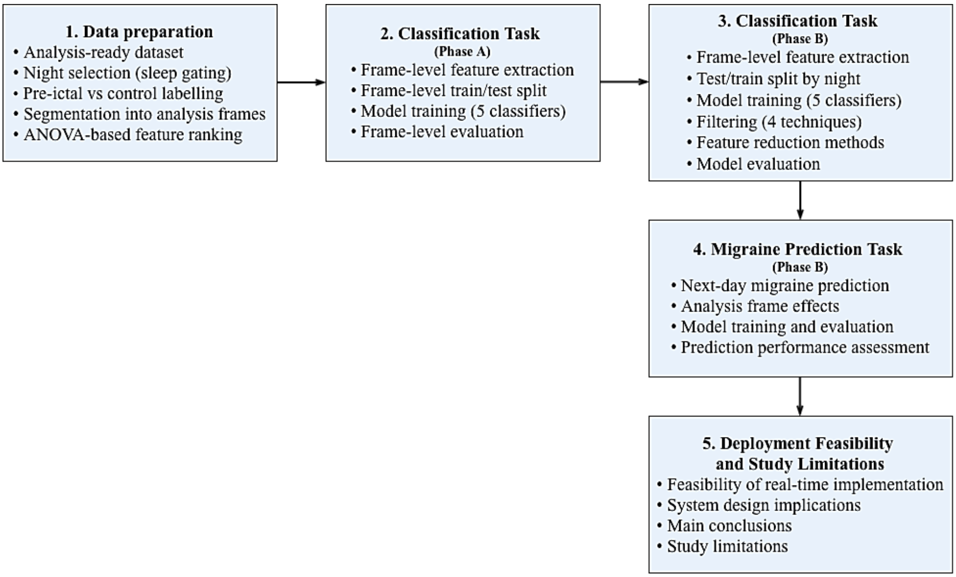


Fig. 3.1. Experimental workflow

The workflow is based on a dataset prepared using the preprocessing pipeline described in Chapter 2. All subsequent analyses operate on nocturnal recordings extracted using the sleep-gating procedure. Individual nights are considered the primary observational units; however, modeling is performed at multiple temporal resolutions defined by analysis-frame duration. Only valid gated nights are included in the experiments.

After night extraction, each night is segmented into analysis frames with durations ranging from 5 to 120 minutes. This segmentation enables investigation of how analysis-frame duration affects classification performance and subsequent night-level risk estimation. For each frame, feature vectors are computed based on physiological signals. Feature normalization is performed using parameters estimated solely from the training data and then applied to the test data, ensuring consistency in the evaluation procedure.

The workflow includes two main analytical components.

First, a binary classification task is performed to distinguish pre-migraine nights from control nights. In Phases A and B, classification is initially conducted at the analysis-frame level using frame-level samples. This setup enables a controlled comparison of methodological factors, such as analysis-frame duration and classifier choice, but does not enforce strict independence between the training and test data. In Phase B, the analysis is extended by introducing a stricter validation setting based on night-level splitting, in which entire nights are assigned exclusively to either the training or the test set. Under this protocol, classification is performed using night-based representations constructed from frame-level features, enabling evaluation under leakage-free conditions.

Five classifiers are evaluated under a unified procedure: XGBoost, Histogram Gradient Boosting, Random Forest, Support Vector Machine, and k-Nearest Neighbors. Classification performance is assessed using standard metrics, with primary emphasis on the F1-score. In addition, feature-level statistical analysis using one-way ANOVA (Fisher, 1992; Ridgman, 1990) is applied to identify features exhibiting class-related differences and to support interpretation of model behavior. In Phase B, additional experiments are conducted to evaluate the effects of signal filtering and feature-space reduction on classification performance.

Second, Phase B includes an additional experiment evaluating the feasibility of estimating night-level migraine risk and predicting next-day migraine. In this setting, models are trained on analysis-frame-level feature vectors, while final predictions are generated at the night level by aggregating frame-level model outputs. This setup allows assessment of how temporal aggregation and modeling choices influence predictive performance under night-based evaluation.

Finally, the experimental results are summarized from a system perspective to assess deployment feasibility and derive implications for a future real-time implementation, including considerations of temporal resolution, signal selection, and computational constraints.

3.3. Task Definition and Labeling Rules

This research defines binary classification and prediction tasks based on nocturnal wearable recordings. The same labeling rules are applied consistently across all experimental settings.

The overall objective is to use physiological data recorded the previous night to determine whether a migraine occurs the following day. Nighttime intervals are defined using the nocturnal gating procedure described in Section 2.5.1 and span the time window from 18:00 to 09:00.

For each night, the target label is assigned according to the calendar day following the end of that nocturnal interval. A night is labeled as positive if at least one migraine event is recorded at any time during the following calendar day. If no migraine events are recorded during that day, the night is labeled as a control.

Importantly, the labeling rule is based on the full next calendar day. Therefore, if a migraine event occurs early in the morning of the target day, for example, at 07:00, the corresponding preceding night is still assigned a positive label. Thus, each nocturnal recording is linked to the full next-day outcome.

3.4. Feature Representation

The feature extraction stage transforms wearable-derived physiological recordings into structured numerical representations suitable for supervised learning. Standard time-domain statistics and signal-shape characteristics were used to ensure interpretability and reproducibility.

All features were computed in Python using NumPy and pandas (Galli, 2021). For each analysis frame, the same feature set was computed independently for each signal channel. In the baseline representation, 13 features were extracted per signal, yielding 78 features across 6 signals. In selected experiments, a reduced subset was used to improve robustness.

The static features were grouped into three types. Central tendency features, such as mean and median, describe the typical signal level. Variability features: standard deviation, minimum, maximum, and root-mean-square describe dispersion. Shape-related features such as peak value, crest factor, clearance factor, impulse factor, shape factor, skewness, and kurtosis describe amplitude, asymmetry, and deviations from Gaussian behavior.

In Phase B, the feature space was extended with dynamic characteristics describing transitions between consecutive analysis frames within the same night. While the static feature set captures within-window properties of physiological signals, the dynamic representation was introduced to quantify short-term temporal changes across adjacent intervals.

For each analysis frame, a vector of mean values across all signal channels was computed. This vector was then compared with the corresponding vector from the preceding frame of the same night. The resulting dynamic features describe how the overall physiological profile evolves over time rather than its absolute level.

The dynamic representation includes several complementary descriptors. First, the magnitude of change between consecutive frames is captured using average, maximum, and Euclidean measures of absolute channel differences. Second, the structural similarity between frames is assessed using cosine similarity, which reflects whether the relative relationships between signals remain consistent. Third, normalized measures of relative change are computed to account for the scale of the preceding physiological state, enabling comparison of proportional rather than absolute variations.

For the first frame of each night, where no preceding frame is available, all dynamic features are set to zero.

These features were implemented using standard numerical operations such as means, norms, and dot products without relying on specialized libraries. Conceptually, they represent generic descriptors of vector changes between consecutive observations. Unlike static features, which describe signal properties within a single time window, dynamic features capture local non-stationarity and short-term transitions in physiological activity across the night.

Such transitions may reflect transient autonomic or behavioral fluctuations that are not captured by within-frame statistics alone.

3.5. Models and Training Configurations

The modeling framework was designed to provide a consistent basis for comparing classifier behavior across analysis-frame durations, feature representations, and preprocessing variants. The same training procedure was applied throughout the research, while only the factor under investigation was varied between experiments. Thus, the classifiers were primarily used as evaluation instruments to assess methodological choices rather than as independently optimized end-use systems.

All models were implemented in Python using the scikit-learn library (Pedregosa et al., 2012), while XGBoost was accessed through its official Python interface. Each analysis frame was represented by a fixed-length feature vector, while the corresponding label was defined at the night level.

To account for differences in numerical scale across physiological features, normalization was applied where required. Normalization parameters were estimated from the training data within each validation split and subsequently applied to the corresponding held-out data to prevent information leakage.

Five supervised classification algorithms were evaluated: Random Forest (De Filippis & Al Foysal, 2025; Sakalauskas & Kriksciuniene, 2025; Svensson & Borgström, 2026), Histogram Gradient Boosting, XGBoost (Badawi et al., 2025; Sharma & Verbeke, 2020; Tsai et al., 2025), Support Vector Machine with a radial basis function kernel (Fricke et al., 2021; Svensson & Borgström, 2026; Tsai et al., 2025), and k-Nearest Neighbors (Sarra et al., 2025; Tsai et al., 2025). These models were selected because they represent different learning principles commonly used in physiological time-series analysis and wearable-based health prediction studies.

Class imbalance was handled using class weighting, without synthetic oversampling or data duplication, to avoid introducing artificial structure into the data.

Model training was carried out using RandomizedSearchCV with GroupKFold cross-validation, grouping by night identifiers to ensure that all frames from the same night were assigned to the same fold, thereby preventing information leakage. Hyperparameter optimization was incorporated into the training pipeline and applied consistently across all experiments.

For models producing continuous decision scores, binary predictions were obtained using a data-driven threshold selected on the training data within each validation split. Threshold selection was guided by the F1-score for the positive class, subject to additional minimum precision and recall constraints, and the resulting threshold was then applied to the corresponding held-out data. This procedure was used to support stable minority-class detection under imbalanced conditions.

3.6. Signal Filtering Configurations

To assess whether model performance depends on preprocessing choices, several signal filtering configurations were evaluated in a sensitivity analysis. The aim of this step was not to modify the signals for their own sake, but to determine whether plausible denoising strategies influence feature extraction and downstream prediction results.

In all experiments, PR values were clipped to a physiologically plausible nocturnal range of 25–115 bpm. This clipping was applied as a fixed data-quality rule. The selected range follows the outlier analysis presented in The Second Chapter, which showed that extreme values originated primarily from sensor artifacts. The clipping procedure was kept identical across all experimental configurations to ensure comparability.

In addition to this mandatory quality-control step, baseline experiments were conducted using unfiltered signals. Filtering was introduced only in experiments designed to evaluate preprocessing effects. Whenever applied, filtering was performed before temporal segmentation and feature extraction so that smoothing effects were propagated consistently throughout the analysis pipeline. Three commonly used smoothing approaches were considered: median filtering, Savitzky–Golay smoothing, and Butterworth low-pass filtering. These methods were selected because they represent different practical ways of reducing noise in physiological time series while preserving meaningful signal structure.

Median filtering was included as a simple, robust technique for suppressing short spikes and isolated artifacts by replacing each sample with the median of its neighboring observations within a sliding window. This approach is widely used in biomedical signal processing to remove impulsive noise while preserving abrupt signal transitions (Justusson, 2006; Nodes & Gallagher, 1982).

Savitzky–Golay smoothing was used because it reduces noise while typically preserving the local shape of the waveform, including gradual trends and peaks. The method performs a local polynomial approximation within a moving window and evaluates the fitted polynomial at the central point. Two moderate configurations were evaluated, using window lengths of 11 and 9 samples with polynomial orders of 2 and 3, respectively (Gallagher, 2020; Savitzky & Golay, 1964).

Butterworth low-pass filtering was included as a standard frequency-based approach that attenuates higher-frequency fluctuations while preserving slower signal trends. In this research, a third-order low-pass Butterworth filter with a cutoff frequency of 0.003 Hz was applied using zero-phase forward–backward filtering in order to avoid phase distortion (Feng & Sun, 2024; Li Zhongshen, 2007).

All filtering procedures were implemented in Python using the SciPy signal-processing library. Median filtering was performed using *scipy.signal.medfilt*, Savitzky–Golay smoothing using *scipy.signal.savgol_filter*, and Butterworth filtering using *scipy.signal.butter* together with *scipy.signal.filtfilt*. The same implementation was applied across all participants and experimental conditions.

The parameter values were selected to represent typical preprocessing settings used in physiological signal analysis. This allowed evaluation of whether the modeling pipeline is robust to plausible preprocessing choices.

Figure 3.2 illustrates how the filters affect a representative nocturnal segment. It compares the raw EDA, SkinTEMP, and PR signals with the median, Savitzky–Golay, and Butterworth outputs for the same time interval.

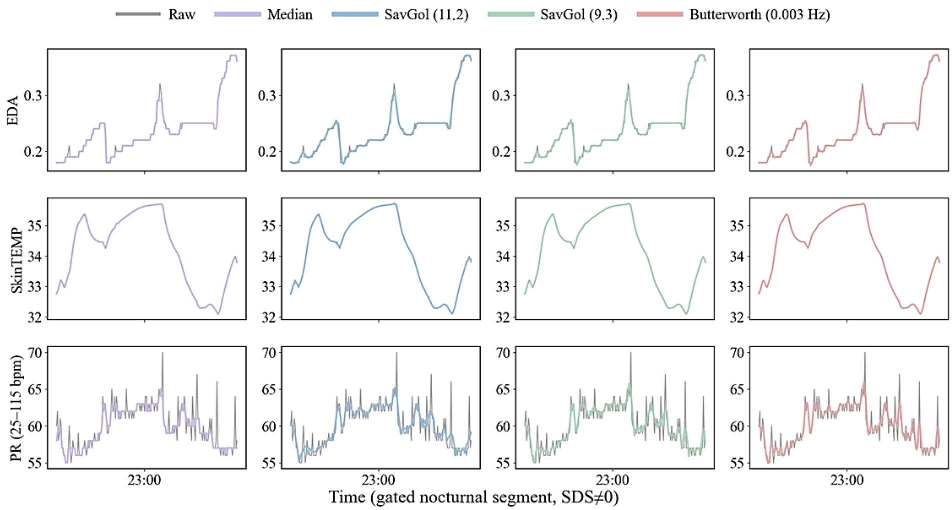


Fig. 3.2. Comparison of raw nocturnal signals and median, Savitzky–Golay, and Butterworth filtering

All subsequent experiments comparing filtering configurations used identical training, validation, and evaluation procedures to enable direct comparison of model behavior across preprocessing variants.

3.7. Validation Protocol and Metrics

Model performance was evaluated using a consistent validation protocol within each experimental setting to ensure comparability of results under fixed methodological configurations. Because migraine-positive nights occurred less frequently than control nights, the dataset was imbalanced, and the choice of evaluation metrics was adapted accordingly.

The primary evaluation metric was the F1-score, as it provides a balanced measure of precision and recall and is more informative than accuracy in imbalanced settings. Accuracy was reported for completeness but was not treated as a primary metric.

Macro-averaged precision, recall, and F1-score were used to characterize overall classification performance, while positive-class metrics were reported separately in the stricter validation setting to assess the ability to detect migraine-related nights. This distinction is important because acceptable overall performance may still correspond to poor detection of the minority class.

To support interpretation, results were evaluated relative to simple baselines. The majority-class baseline was used for classification metrics. For PR-AUC, performance was interpreted relative to the prevalence of the positive class, corresponding to the expected level for a non-informative ranking.

For prediction experiments, the evaluation focused on the positive class. In addition to threshold-based metrics, PR-AUC was used as a threshold-independent measure that reflects the trade-off between precision and recall and provides a more informative assessment under class imbalance.

Given the difficulty of the task and the heterogeneity of migraine-related patterns, positive-class metrics were interpreted relative to baseline performance rather than universal absolute thresholds. In this context, moderate values may still indicate meaningful predictive structure if they are consistent and exceed baseline.

The following standard classification metrics were reported according to the definitions commonly used in the machine learning literature (Foody, 2023; Vujovic, 2021).

Accuracy, defined as the proportion of correctly classified analysis frames,

$$Accuracy = \frac{T_p + T_n}{T_p + F_p + T_n + F_n}. \tag{3.1}$$

Precision, defined as the proportion of predicted positive analysis frames that are true positives,

$$Precision = \frac{T_p}{T_p + F_p}. \tag{3.2}$$

Recall (sensitivity), defined as the proportion of true positive analysis frames that were correctly detected,

$$Recall = \frac{T_p}{T_p + F_n}. \tag{3.3}$$

Binary F1-score, defined as the harmonic mean of precision and recall for the positive class,

$$F_1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall}. \tag{3.4}$$

For experiments reported with positive-class metrics, the binary definitions in Eqs. (3.2)–(3.4) were applied directly to Class 1.

All metrics were computed on held-out validation folds and averaged across folds. Training, normalization, and decision-threshold selection were performed using only the training data within each split. Validation data were not used during these steps.

3.8. Results

This section presents the results of the experimental evaluation of the proposed framework for analyzing nocturnal wearable physiological data in migraine research. The main objective of this stage was to determine whether stable pre-migraine patterns can be detected under different analytical settings and to identify which methodological factors most strongly affect classification performance. A separate part of the analysis assessed the feasibility of night-level migraine risk estimation from nocturnal wearable data.

3.8.1. Exploratory Classification Under Frame-Level Splitting

After extracting nocturnal sleep, each participant's recordings were segmented into analysis frames of 5, 10, 30, 60, 90, and 120 minutes. For each analysis-frame duration, informative features were computed, and several classifier families were trained. For each participant and each analysis-frame duration, the model with the highest F1-score was selected. The analysis was conducted separately for Phase A and Phase B to examine the consistency of the observed tendencies across different sample sizes.

In this exploratory setting, training and evaluation were performed using frame-level splitting, meaning that analysis frames originating from the same night could appear in both the training and the test sets. As a result, the findings reported in this section primarily reflect relative differences between methodological configurations, such as temporal segmentation and classifier choice, rather than providing an unbiased estimate of real-world predictive performance.

The distribution of classification performance metrics across participants is presented in Figure 3.3, including F1-score, Recall, Precision, and Accuracy for different analysis-frame durations in both phases.

Across Phase A and Phase B, a consistent tendency is observed: shorter analysis frames generally yield better classification performance. In particular, the highest median F1-scores are obtained for 5- and 10-minute windows, whereas longer frames of 60–120 minutes are associated with a noticeable decline in performance. A similar tendency is observed for Recall, suggesting reduced sensitivity to the positive class as temporal aggregation increases. Precision and Accuracy follow the same general pattern, although the effect is less pronounced.

This behavior suggests that informative physiological variations related to migraine may be temporally localized and therefore better captured at finer temporal resolution. Increasing the analysis-frame duration likely smooths short-term fluctuations and may reduce the amount of discriminative information available to the classifiers.

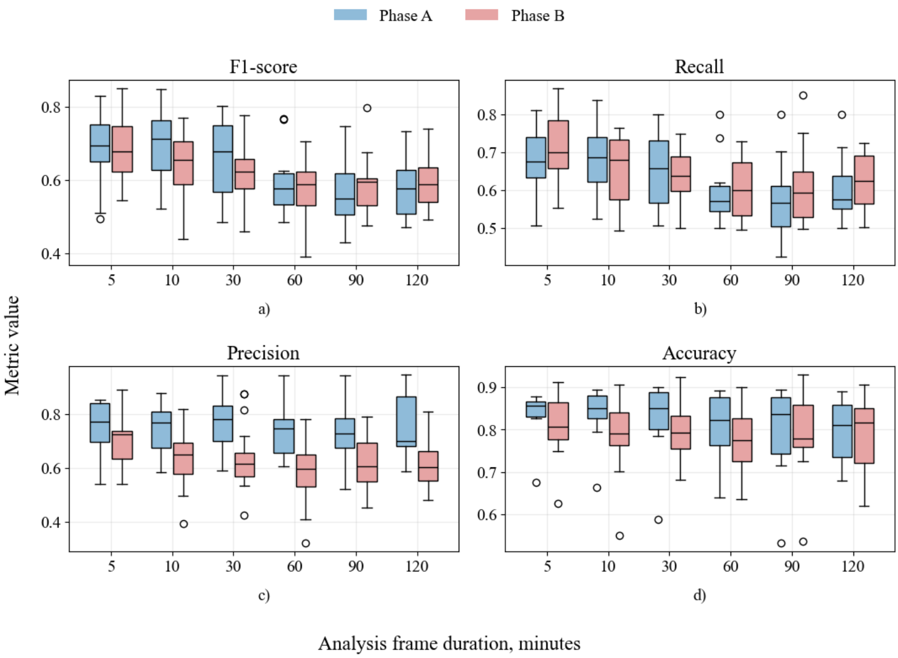


Fig. 3.3. Distribution of classification performance metrics across participants

However, this interpretation should be treated with caution. Due to the frame-level splitting strategy, segments from the same night may appear in both the training and test sets. Under such conditions, the apparent advantage of short analysis frames may partly reflect residual similarity between segments from the same recording rather than true generalization to unseen nights. In particular, improved performance for 5- and 10-minute windows may be influenced by partial recognition of within-night patterns when similar segments are shared between the training and test subsets.

Therefore, although the observed dependence on analysis-frame duration is consistent and physiologically plausible, it should be regarded as exploratory. The results reveal a notable tendency, but they do not yet allow a definitive conclusion regarding the optimal temporal resolution for prediction.

The distribution of classifiers selected as the best-performing models according to F1-score is shown in Figure 3.4.

Ensemble-based methods, particularly Random Forest and XGBoost, were most frequently selected as the best-performing models in both phases. This suggests that nonlinear relationships between physiological features may play an important role in the classification task. At the same time, no single classifier consistently outperformed the others across all analysis-frame durations, indicating that model performance depends on both algorithm choice and temporal segmentation.

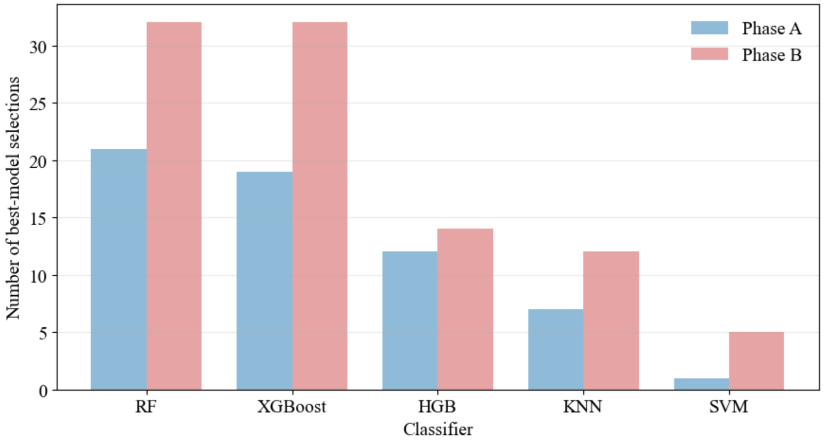


Fig. 3.4. Distribution of classifiers selected as the best models according to the F1-score in Phases A and B

Across participants, the best-performing participants varied considerably, indicating substantial inter-individual variability. In some cases, relatively high F1-scores were observed, whereas in others, performance remained moderate. Given the complexity of the task, including noisy wearable signals, class imbalance, and limited event counts, this variability suggests that informative patterns are present but are not uniformly strong or stable across individuals.

In summary, the exploratory analysis reveals a consistent tendency toward improved classification performance for shorter analysis frames. However, because the evaluation protocol allows overlap between the training and testing data at the night level, the possibility of information leakage cannot be ruled out. Consequently, the observed advantage of short analysis frames may reflect either genuinely localized physiological dynamics or partial recognition of segments from the same recordings.

To address this limitation, a stricter evaluation setting was subsequently introduced (Section 3.8.2), in which entire nights were treated as independent observational units and assigned exclusively to either the training or the test set. This approach eliminates overlap between subsets and enables assessment of whether the observed pattern persists under a more realistic, leakage-free validation framework.

3.8.2. Cohort Classification Under Night-Level Splitting

To verify whether the tendencies observed in the exploratory analysis persist under stricter conditions, a cohort-level evaluation was conducted using night-level splitting. In this setting, entire nights were treated as independent observational units, and all analysis frames belonging to a given night were assigned exclusively to either the training or the test set. This approach eliminates overlap between subsets at the night level and provides a more realistic assessment of model generalization to unseen data.

Under this validation protocol, classification performance was generally lower than in the frame-level setting, which is consistent with the increased difficulty of the task when temporal dependencies between training and testing samples are removed. At the same time, the overall structure of the results remained comparable, allowing meaningful analysis of methodological factors such as analysis-frame duration and classifier choice.

The obtained results are summarized in Figures 3.5 and 3.6. Figure 3.5 presents positive-class performance metrics, including F1-score, Recall, Precision, and PR-AUC, while Figure 3.6 shows the corresponding macro-averaged metrics. A random baseline is indicated in both figures to provide a reference level for comparison. Preliminary experiments also included the KNN classifier; however, because it did not demonstrate meaningful predictive ability for the positive class across the examined analysis-frame durations, it was not included in the detailed comparative analysis presented in this section.

As shown in Figure 3.5, which focuses on positive-class performance, shorter analysis frames of 5–10 minutes consistently yield higher F1 and recall scores than longer frames. This tendency is particularly pronounced for ensemble-based models. In particular, HistGradientBoosting and XGBoost achieve their highest positive-class F1-scores and Recall values at short temporal resolutions, followed by a gradual decline as the analysis-frame duration increases to 90–120 minutes.

A similar, although less pronounced, pattern can be observed in the macro-level results presented in Figure 3.6.

Macro F1 and ROC-AUC reach their highest values at shorter analysis frames and show a moderate decrease as temporal aggregation increases. At the same time, the comparison between Figures 3.5 and 3.6 highlights that the degradation is substantially stronger for positive-class metrics than for macro-averaged metrics.

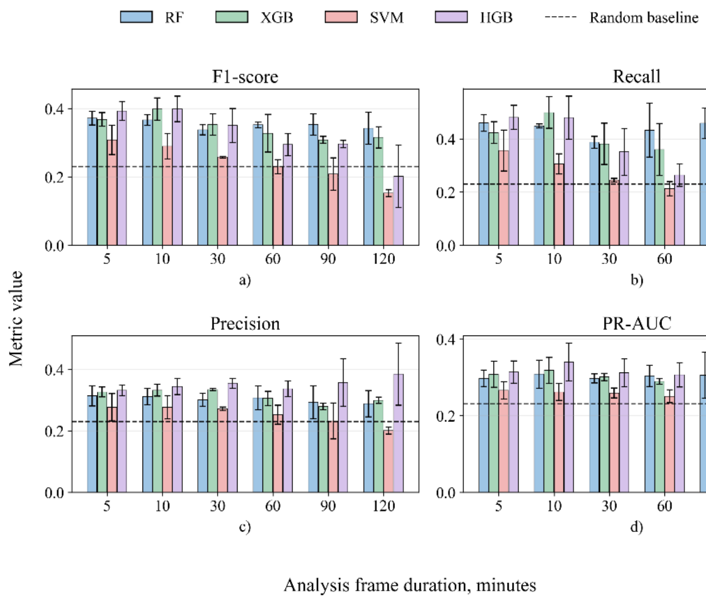


Fig. 3.5. Classification performance on the positive class across analysis-frame durations for different models: (a) F1-score; (b) Recall; (c) Precision; and (d) PR-AUC

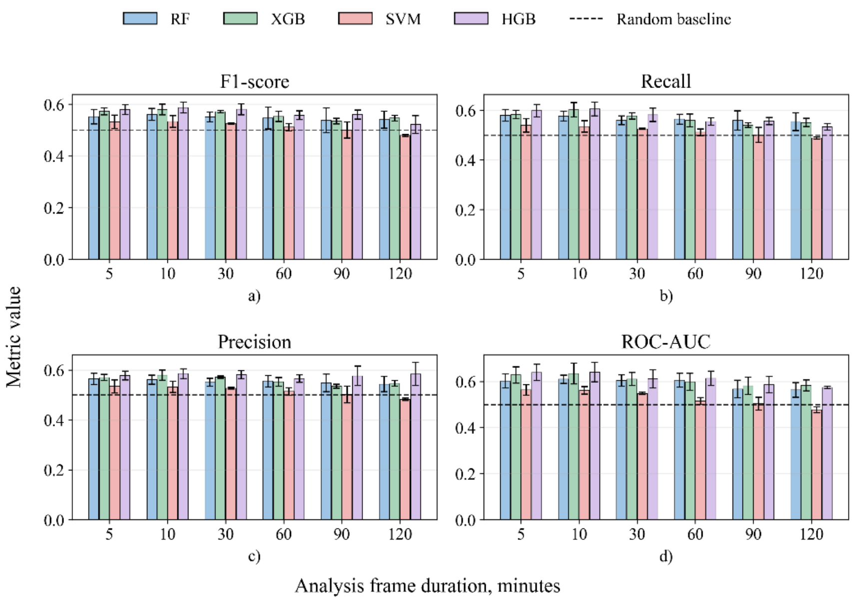


Fig. 3.6. Macro-averaged classification performance across analysis-frame durations for different models: (a) F1-score; (b) Recall; (c) Precision; and (d) ROC-AUC

To quantify this effect, the relative change in positive-class performance was examined across analysis-frame durations. A systematic degradation is observed for the best-performing models. For HistGradientBoosting, the positive-class F1-score decreases from approximately 0.40 at 10-minute frames to about 0.20 at 120-minute frames, corresponding to a reduction of nearly 50%. Over the same range, positive recall declines from approximately 0.48 to 0.16, which represents a decrease of about 65%.

For XGBoost, the decrease is less pronounced but remains substantial. The positive-class F1-score declines from approximately 0.40 at 10 minutes to around 0.32 at 120 minutes, corresponding to a reduction of about 20%, while positive recall decreases from approximately 0.50 to 0.34, indicating a reduction of roughly 30% between short and long analysis frames.

Interestingly, this degradation pattern is less pronounced for Random Forest. As shown in Figure 3.5, positive-class Recall for Random Forest does not exhibit a clear monotonic decrease with increasing analysis-frame duration and remains relatively stable across several temporal scales. A similar tendency is observed for positive-class F1-score, which shows more variability but no consistent downward trend. This behavior suggests that Random Forest is less sensitive to temporal resolution and to the loss of short-term physiological fluctuations. However, this apparent stability does not translate into superior predictive performance. Instead, it reflects a more conservative modeling behavior, in which the classifier captures more stable, coarse-grained patterns but is less effective at exploiting temporally localized signals compared to boosting-based methods.

The comparison with the random baseline further supports the validity of the obtained results. As shown in Figures 3.5 and 3.6, most models consistently outperform the baseline across all analysis-frame durations, indicating that the classification task is non-trivial and that meaningful predictive structure is present in the data. At the same time, the gap between the models and the baseline is considerably smaller for positive-class metrics than for macro-averaged metrics, highlighting the difficulty of detecting minority-class events.

In contrast to ensemble-based methods, SVM performs substantially worse. As illustrated in Figure 3.5, SVM yields consistently lower positive-class F1-scores and Recall values across all analysis-frame durations. Moreover, for longer frames (30–120 minutes), its performance approaches the random baseline, both in positive-class metrics and in macro-averaged metrics (Fig. 3.6). This indicates the limited ability of the SVM model to capture the underlying structure of the data, particularly under conditions of temporal aggregation and class imbalance.

A related effect can be observed for Accuracy, which remains relatively stable or even increases for longer analysis frames. However, this is accompanied by a reduction in positive-class Recall, indicating that models increasingly favor the majority class as temporal aggregation increases. In the presence of class imbalance, this limits the practical usefulness of longer analysis frames despite their acceptable global performance.

These findings suggest that shorter analysis frames preserve transient physiological variations that are important for discriminating migraine-related states. In contrast, longer frames introduce temporal smoothing, reducing the discriminative signal and leading to more conservative predictions. Importantly, as demonstrated in Figures 3.5 and 3.6, this tendency remains observable under the stricter night-level validation protocol, supporting the interpretation that it reflects a genuine property of the data rather than an artifact of the earlier evaluation setting.

Across classifiers, ensemble-based methods again demonstrate the most favorable results. HistGradientBoosting and XGBoost consistently achieve the best balance between overall discrimination and positive-class detection across most analysis-frame durations. Random Forest provides more stable but generally less optimal performance, while SVM shows consistently inferior results.

Across participants, variability remains substantial, reflecting inter-individual differences in physiological patterns and the inherent difficulty of the prediction problem. Nevertheless, the preservation of performance above the random baseline indicates that the models capture meaningful structure in the data.

In summary, the cohort-level analysis confirms that classification performance decreases under stricter, leakage-free validation, while preserving the key tendency observed in the exploratory analysis: shorter analysis frames of 5–10 minutes provide more informative representations for migraine prediction. The comparison between Figures 3.5 and 3.6 further demonstrates that this effect is primarily driven by degradation in positive-class detection. At the same time, the results highlight differences between classifier families, with boosting-based methods showing the highest sensitivity to temporally localized patterns and SVM approaching baseline performance under more challenging conditions.

3.8.3. Effect of Filtering on Classification

Signal filtering was investigated to assess its impact on migraine risk classification performance while keeping all other pipeline components unchanged. To isolate the effect of preprocessing, filtering was applied directly to the physiological signals prior to segmentation and feature extraction. All subsequent stages, including nocturnal window selection, feature computation, classifier configuration, and validation protocol, were held constant.

Four filtering strategies were evaluated: median filtering, Butterworth low-pass filtering, and Savitzky–Golay smoothing with two parameter configurations. The analysis was conducted under two validation settings: an exploratory frame-level protocol and a stricter cohort-level protocol with night-level splitting. In Phase B, three independent training and evaluation runs were performed to obtain a more robust estimate of performance. The SVM and KNN classifiers were excluded from further analysis due to their consistently low performance, as discussed in Section 3.8.2.

In the exploratory frame-level evaluation, filtering improved classification performance across several metrics, particularly for tree-based models. As shown in Tables 3.1–3.5, both Butterworth filtering and Savitzky–Golay smoothing with parameters (11,2) yielded higher F1-score, recall, and AUC values than the baseline configuration. For example, in the case of XGBoost, the F1-score increased from 0.649 without filtering to 0.678 with Butterworth filtering and to 0.695 with Savitzky–Golay smoothing. Comparable tendencies were observed for the other ensemble-based classifiers. Median filtering, by contrast, produced only limited deviations from the baseline and did not yield a stable improvement pattern.

Table 3.1. Overall metrics for each model without filtering

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.770	0.639	0.669	**0.649**	0.76
HGB	0.736	0.626	0.679	**0.635**	0.75
RF	0.838	0.766	0.579	0.595	0.78
SVM	0.817	0.608	0.511	0.480	0.54
KNN	0.805	0.610	0.548	0.551	0.62

Table 3.2. Overall metrics for each model with median filtering

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.769	0.634	0.660	**0.643**	0.75
HGB	0.740	0.623	0.669	**0.633**	0.75
RF	0.832	0.728	0.571	0.583	0.76
SVM	0.778	0.539	0.521	0.513	0.56
KNN	0.804	0.608	0.548	0.551	0.61

Table 3.3. Overall metrics for each model with Butterworth low-pass filtering

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.798	0.669	0.690	**0.678**	0.79
HGB	0.773	0.656	0.704	**0.670**	0.79
RF	0.851	0.806	0.616	**0.645**	0.82
SVM	0.755	0.560	0.552	0.555	0.59
KNN	0.822	0.680	0.595	**0.612**	0.69

Table 3.4. Overall metrics for each model with Savitzky–Golay filtering, window length 11, and polynomial order 2

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.806	0.683	0.712	**0.695**	0.81
HGB	0.781	0.666	0.719	**0.682**	0.81
RF	0.858	0.815	0.642	**0.677**	0.85
SVM	0.775	0.557	0.533	0.531	0.58
KNN	0.825	0.688	0.606	**0.625**	0.70

Table 3.5. Overall metrics for each model with Savitzky–Golay filtering, window length 9, and polynomial order 3

Model	Accuracy	Precision	Recall	F1-score	AUC
XGBoost	0.762	0.622	0.642	**0.629**	0.81
HGB	0.722	0.615	0.666	**0.622**	0.81
RF	0.836	0.766	0.570	0.581	0.84
SVM	0.817	0.619	0.512	0.482	0.58
KNN	0.809	0.627	0.555	0.561	0.70

However, these apparent gains were not preserved under the stricter validation scenario. In Phase B, performance was evaluated at the cohort level using night-level splitting, thereby providing a more realistic estimate of generalization to unseen nights. In addition to macro-level metrics, positive-class metrics were analyzed separately to assess the effect of filtering on the detection of pre-migraine nights, the primary classification target.

To summarize filtering effects in a compact, interpretable form, two heatmaps were constructed. Figure 3.7 shows the change in Macro F1-score relative to the baseline configuration and therefore reflects overall classification performance.

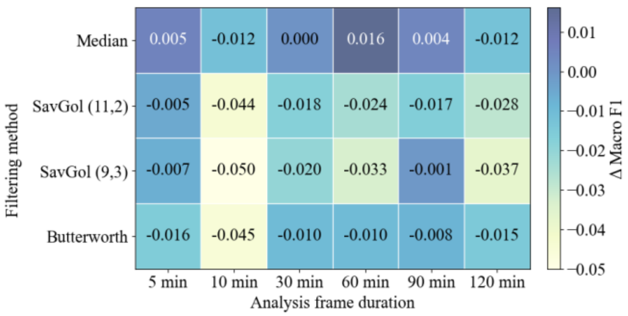


Fig. 3.7. Effect of signal filtering on overall classification performance across analysis-frame durations (ΔMacro F1 relative to baseline)

Figure 3.8 presents the corresponding change in positive-class recall and directly characterizes the model's ability to detect pre-migraine nights. The heatmap based on the Macro F1-score indicates that filtering exerts only a limited, non-uniform effect on overall classification quality.

Median filtering produces slight improvements for longer analysis windows, particularly at 60 and 90 minutes, whereas Butterworth filtering is associated with a small but consistently negative shift. Savitzky–Golay filtering yields the most pronounced reduction in the Macro F1-score across the majority of evaluated configurations. At the same time, the absolute magnitude of these changes remains relatively modest, indicating that filtering does not substantially alter global model performance.

A different pattern is observed for positive-class recall. As shown in Figure 3.8, filtering generally reduces the model's ability to detect pre-migraine nights.

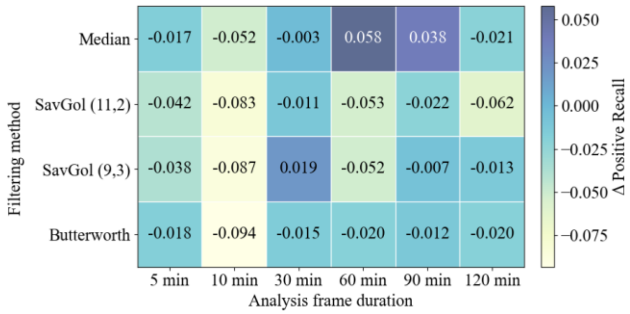


Fig. 3.8. Effect of signal filtering on pre-migraine detection performance across analysis-frame durations (ΔPositive Recall relative to baseline)

This negative effect is most pronounced for Savitzky–Golay filtering, which demonstrates the largest decline across most analysis windows. Butterworth filtering also leads to a stable deterioration in recall, although its impact is less severe. Median filtering exhibits mixed behavior, with moderate improvements observed for some longer windows; however, these effects are not consistent across all temporal scales.

The divergence between the Macro F1-score and positive-class recall is particularly important for interpreting the results. Aggregate metrics may suggest that certain filtering configurations are neutral or even mildly beneficial. Nevertheless, a closer inspection of the positive class reveals that these same configurations often reduce sensitivity to the clinically relevant outcome. This suggests that filtering attenuates subtle physiological signal characteristics that are informative for identifying pre-migraine states.

Overall, the results indicate that signal filtering does not provide a reliable benefit for migraine risk classification in this pipeline. Although some configurations yield small changes in macro-level performance, filtering generally has an unfavorable effect on the detection of pre-migraine nights. These findings suggest that preserving the original signal structure is more important than applying conventional smoothing techniques and highlight the importance of task-specific preprocessing in physiological signal analysis.

3.8.4. Feature Relevance and Feature Reduction

After analyzing the effects of analysis-frame duration and signal filtering, the next step was to investigate the structure of informative features that determine classification performance. The objective of this stage was to identify which physiological signals and statistical descriptors consistently contribute to class separability and to evaluate how this structure behaves under different analytical perspectives.

To address this, two complementary analyses were performed. In the exploratory stage, one-way ANOVA was used to assess the statistical separability of individual features between pre-migraine and control nights across physiological signals and analysis-frame durations, and the results were visualized as heatmaps of F-statistics and corresponding p-values. In Phase B, feature importance was evaluated within a predictive modeling framework under night-level splitting, using the recurrence of top-ranked features across models, analysis-frame durations, and repeated runs.

For EDA, the strongest and most consistent separability was observed for short analysis-frame durations, particularly 5 and 10 minutes, as shown in Figure 3.9.

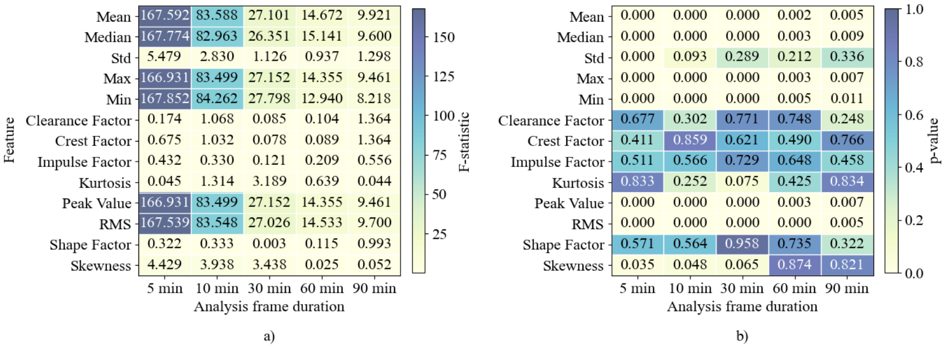


Fig. 3.9 ANOVA analysis of electrodermal activity features across various analysis frame durations: (a) F-statistics and (b) p-values

High F-statistics were obtained for low-order statistical descriptors such as Mean, Median, Min, Peak, and RMS. As the analysis-frame duration increased to

60–120 minutes, F-statistics decreased, and p-values increased, indicating a progressive reduction in discriminative power due to temporal averaging. This suggests that EDA-related pre-migraine changes are temporally localized and attenuate over longer windows.

For PR, separability was more moderate and less stable across analysis frames, as illustrated in Figure 3.10.

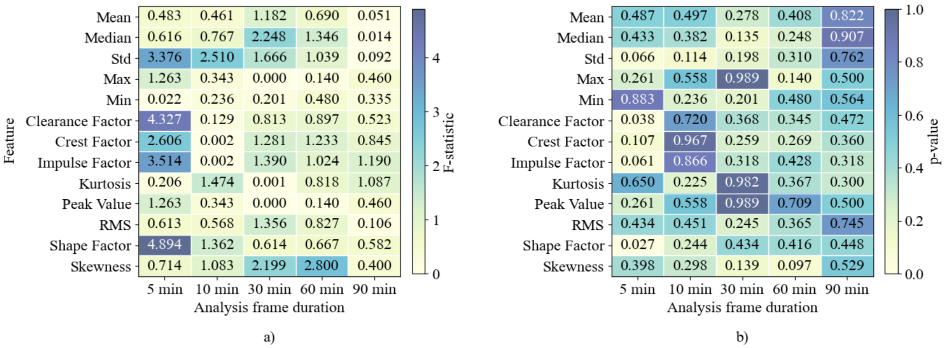


Fig. 3.10. ANOVA analysis of pulse rate features across various analysis frame durations: (a) F-statistics and (b) p-values

Although certain descriptors showed statistically significant differences, these effects were not consistent across all temporal resolutions, suggesting greater variability and a less dominant standalone role.

Skin temperature exhibited a more coherent pattern, similar to EDA, as shown in Figure 3.11.

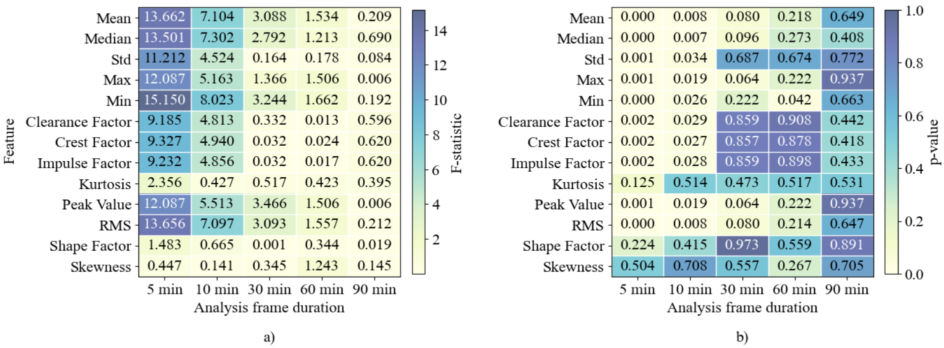


Fig. 3.11. ANOVA analysis of skin temperature features across various analysis frame durations: (a) F-statistics and (b) p-values

In short analysis frames, descriptors such as Mean, Median, Standard Deviation, Min, and RMS showed clear class differences. However, separability decreased with increasing window length, suggesting that thermoregulatory changes relevant to migraine prediction are also short-term phenomena.

MET demonstrated intermediate behavior (Fig. 3.12), with moderate separability observed for selected descriptors in short- and medium-length windows, but without a stable pattern across all temporal scales.

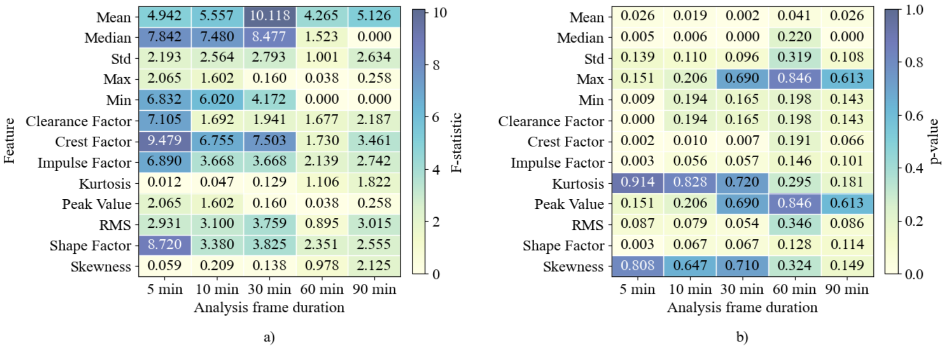


Fig. 3.12. ANOVA analysis of metabolic equivalent features across various analysis frame durations: (a) F-statistics and (b) p-values

A similar tendency was observed for ACT and accelerometer-derived features (Figs. 3.13 and 3.14), with stronger discriminative power in short analysis frames and decreasing power with stronger temporal aggregation.

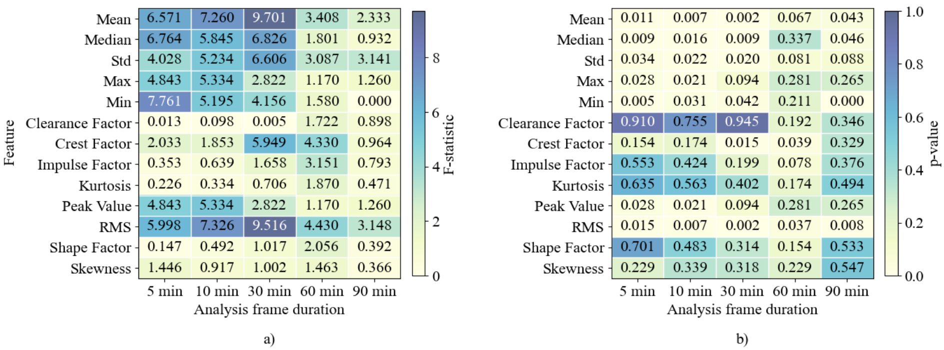


Fig. 3.13. ANOVA analysis of activity counts features across various analysis frame durations: (a) F-statistics and (b) p-values

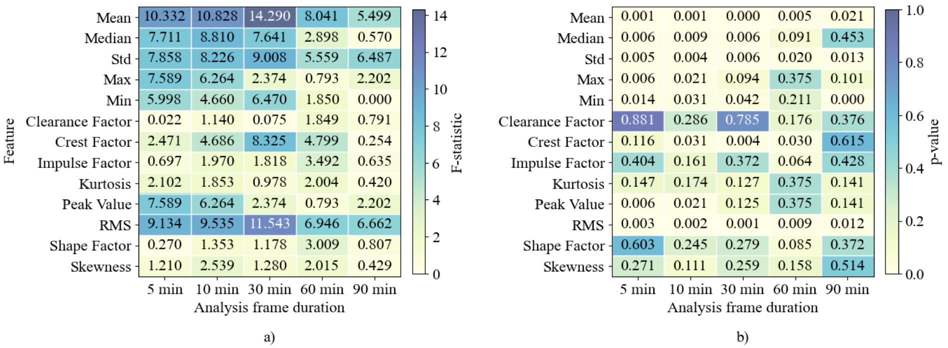


Fig. 3.14. ANOVA analysis of accelerometer features across various analysis frame durations: (a) F-statistics and (b) p-values

Across all signals, the ANOVA results reveal a consistent pattern: the highest class separability occurs in short analysis frames, and the most informative descriptors are low-order statistical measures related to signal level, central tendency, and amplitude.

In Phase B, feature importance was analyzed within the classification framework under night-level splitting. For each model and each analysis-frame duration, the top-ranked features for the positive class were identified, and their recurrence was aggregated across 12 model runs with three runs per six analysis frames and four classifiers. The results are summarized in Tables 3.6–3.11.

Table 3.6. Top-ranked features aggregated across model runs for the 5-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
eda_scl_usiemens__min	0.04770	12
eda_scl_usiemens__max	0.04005	12
eda_scl_usiemens__peak_value	0.05109	9
met__min	0.03936	9
activity_counts__max	0.05374	2
activity_counts__peak_value	0.05374	2
accelerometers_std_g__shape_factor	0.05964	3
temperature_celsius__shape_factor	0.05586	4
temperature_celsius__std	0.06153	2
pulse_rate_bpm__clearance_factor	0.10200	3

Table 3.7. Top-ranked features aggregated across model runs for the 10-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
eda_scl_usiemens__min	0.04029	12
eda_scl_usiemens__clearance_factor	0.03833	10
eda_scl_usiemens__peak_value	0.03763	9
activity_counts__min	0.03858	8
met__min	0.08659	9
temperature_celsius__shape_factor	0.05338	4
temperature_celsius__std	0.07024	3
accelerometers_std_g__shape_factor	0.04714	3

Table 3.8. Top-ranked features aggregated across model runs for the 30-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
eda_scl_usiemens__min	0.03555	12
eda_scl_usiemens__median	0.03563	10
eda_scl_usiemens__clearance_factor	0.04394	10
accelerometers_std_g__min	0.03423	8
activity_counts__median	0.03198	8
activity_counts__min	0.02892	6
accelerometers_std_g__median	0.03246	5

Table 3.9. Top-ranked features aggregated across model runs for the 60-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
accelerometers_std_g__min	0.03862	9
eda_scl_usiemens__clearance_factor	0.02951	9
dyn__mean_rel_delta_six	0.02960	7
met__median	0.02649	7
activity_counts__min	0.03017	5
eda_scl_usiemens__peak_value	0.02579	6

Table 3.10. Top-ranked features aggregated across model runs for the 90-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
met__median	0.03816	9
eda_scl_usiemens__median	0.02645	9
eda_scl_usiemens__min	0.02722	8
eda_scl_usiemens__mean	0.02660	8
activity_counts__min	0.02832	7
eda_scl_usiemens__clearance_factor	0.02621	7

Table 3.11. Top-ranked features aggregated across model runs for the 120-minute analysis frame

Feature	Mean importance	Recurrence (n/12)
activity_counts__median	0.03623	7
eda_scl_usiemens__clearance_factor	0.02882	7
accelerometers_std_g__median	0.04097	5
eda_scl_usiemens__peak_value	0.03298	5
eda_scl_usiemens__mean	0.02932	4
pulse_rate_bpm__kurtosis	0.02910	3

For short analysis frames of 5–10 minutes, which correspond to the most informative temporal scales identified in ANOVA, a highly consistent feature structure emerges. As shown in Tables 3.6 and 3.7, EDA-derived features dominate the top-ranked predictors. In particular, features such as *eda_scl_usiemens__min* appear in all model runs, while *eda_scl_usiemens__peak_value* and *eda_scl_usiemens__clearance_factor* also show high recurrence. These features correspond directly to the statistically significant descriptors identified in Figure 3.9, demonstrating strong agreement between Phase A and Phase B.

In addition to EDA, features derived from MET and movement-related signals, such as ACT and accelerometer data, also appear among the top-ranked predictors, particularly in the 10-minute window (Table 3.7). These features are typically associated with minimum values and central tendency, consistent with the moderate separability observed in Figures 3.12–3.14. At intermediate analysis frames of 30–60 minutes, the feature structure becomes more stable and shifts toward tonic descriptors. As shown in Tables 3.8 and 3.9, median- and minimum-based features become more prominent, particularly for EDA and accelerometer signals. EDA features remain highly recurrent, including *eda_scl_usiemens__median* and *eda_scl_usiemens__clearance_factor*, while movement-related features such as *accelerometers_std_g__min* and *activity_counts__min* increase in importance. This transition is consistent with the reduction in short-term variability observed in the ANOVA results.

At longer analysis frames of 90–120 minutes, the feature structure becomes more heterogeneous, as shown in Tables 3.10 and 3.11. Although EDA features remain present and recurrent, their dominance decreases relative to shorter windows. Movement-related and aggregated descriptors, such as *activity_counts__median* and *accelerometers_std_g__median*, become more prominent. This reflects the effect of temporal averaging, which attenuates localized physiological deviations and leads to a more diffuse representation of discriminative information.

A direct comparison between Phase A and Phase B reveals a systematic alignment. The features identified as statistically discriminative in ANOVA, primarily low-order EDA descriptors, are also the most recurrent and influential in the risk evaluation modeling framework. This agreement is particularly strong across all analysis-frame durations for EDA features and is most pronounced for short windows.

At the same time, movement-related signals exhibit partial agreement between the two analyses. While their statistical separability is moderate (Figs. 3.12–3.14), they consistently appear among the top-ranked features, especially at longer analysis-frame durations. This indicates that these signals provide complementary information that becomes more relevant under temporal aggregation.

In contrast, pulse rate and skin temperature features show weaker consistency between ANOVA and model-based importance. Although some descriptors demonstrate statistical significance, their recurrence across model runs is limited, suggesting lower stability in predictive settings.

Overall, the combined results of Phases A and B indicate that the most informative features for pre-migraine classification are low-order statistical descriptors, primarily derived from electrodermal activity and supported by movement-related signals. The strong agreement between statistical analysis and model-based feature importance confirms that the intrinsic feature separability observed at the cohort level is preserved in night-level classification based on nocturnal physiological data.

Furthermore, the observed evolution of feature structure across analysis-frame durations demonstrates that temporal aggregation plays a critical role in shaping feature relevance. Short analysis frames preserve localized physiological deviations and maximize discriminative power, whereas longer windows attenuate these effects and yield a more distributed feature representation.

These findings support both hypotheses of the dissertation. First, the detectability of pre-migraine physiological changes depends strongly on methodological choices, particularly temporal aggregation and feature representation. Second, despite pronounced inter-individual variability, recurrent cohort-level patterns can be identified, especially in electrodermal activity and low-order statistical descriptors, which consistently emerge as the most informative features across models and configurations.

To further validate the feature analysis results, an additional experiment was conducted using a reduced subset of 10 features selected from the recurrent predictors identified across Tables 3.6–3.11. The selection was based not on a single analysis-frame duration, but on the stability of feature recurrence across multiple temporal resolutions, classifier configurations, and repeated runs. As a result, the final subset included electrodermal activity descriptors such as eda_scl_usiemens__min, eda_scl_usiemens__median, eda_scl_usiemens__clearance_factor, and eda_scl_usiemens__peak_value, together with movement-related variables accelerometers_std_g__min, accelerometers_std_g__median, activity_counts__min, and activity_counts__median, as well as the complementary descriptors met__median and dyn__mean_rel_delta_six. This composition closely follows the structure revealed in the feature recurrence analysis, where EDA-derived features form the most stable and dominant component, while movement-related and aggregated descriptors provide additional support under temporal aggregation.

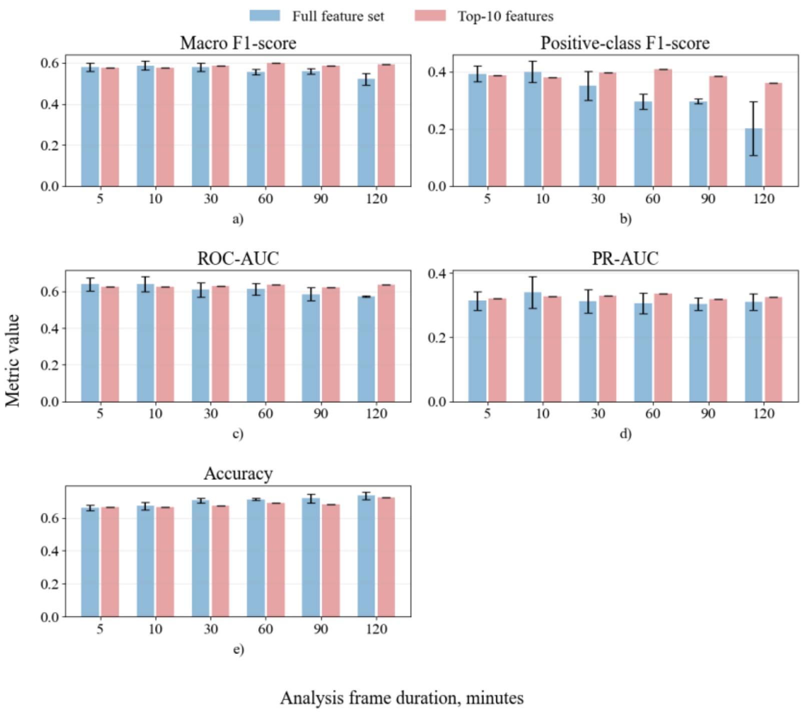


Fig. 3.15. Comparison of classification performance obtained with the full feature set and the reduced top-10 feature subset across different analysis-frame durations: (a) Macro F1-score, (b) Positive-class F1-score, (c) ROC-AUC, (d) PR-AUC, and (e) Accuracy

The effect of feature-space reduction on classification performance is illustrated in Figure 3.15, which compares the full feature set and the reduced top-10 subset across different analysis-frame durations.

The results demonstrate that reducing the number of features does not degrade model performance. On the contrary, for most configurations, the reduced feature subset yields comparable or higher values of Macro F1-score, ROC-AUC, PR-AUC, and Accuracy. The most pronounced improvement is observed for the positive-class F1-score, particularly at medium and longer analysis frames of 30–90 minutes, indicating better detection of the minority class.

The analysis further reveals that the benefit of feature reduction is not uniform across temporal resolutions. For short analysis frames of 5–10 minutes, the performance of the reduced subset remains close to that of the full feature set, suggesting that the strongest discriminative patterns are already captured by a small number of highly informative features. At intermediate durations of 30–60 minutes, the reduced subset often provides improved performance, reflecting a better balance between signal preservation and noise suppression. At longer analysis frames of 90–120 minutes, where temporal averaging attenuates localized physiological variations, the reduced feature set helps maintain model stability and prevents performance degradation caused by redundant descriptors.

These results indicate that the predictive structure of the problem is concentrated in a limited number of recurrent features, while the remaining variables in the full feature set contribute relatively little additional information. Instead, they may introduce variability and reduce model robustness. Consequently, the reduced feature subset can be interpreted as a compact and stable representation of pre-migraine physiological patterns.

Overall, the experiment confirms that the feature-relevance patterns identified in Phases A and B are not only statistically and structurally consistent but also sufficient to maintain classification performance. The strong agreement between feature recurrence and classification results supports the use of a compact feature representation as an effective strategy for building robust and interpretable predictive models.

3.8.5. Night-Based Migraine Risk Estimation

Night-level migraine risk estimation was evaluated by aggregating interval-based predictions into binary night labels across RF, XGB, and HGB models. For each combination of analysis frame duration and classifier, experiments were repeated across three independent runs, and results were averaged accordingly. The same evaluation protocol was applied to both cohort-level and individual participant models.

At the cohort level, night-level performance demonstrated moderate but consistent predictive capability across models and window lengths. As shown in Figure 3.16, the F1-score remained relatively stable, typically in the range of approximately 0.33–0.42, exceeding the prevalence-based baseline across all configurations.

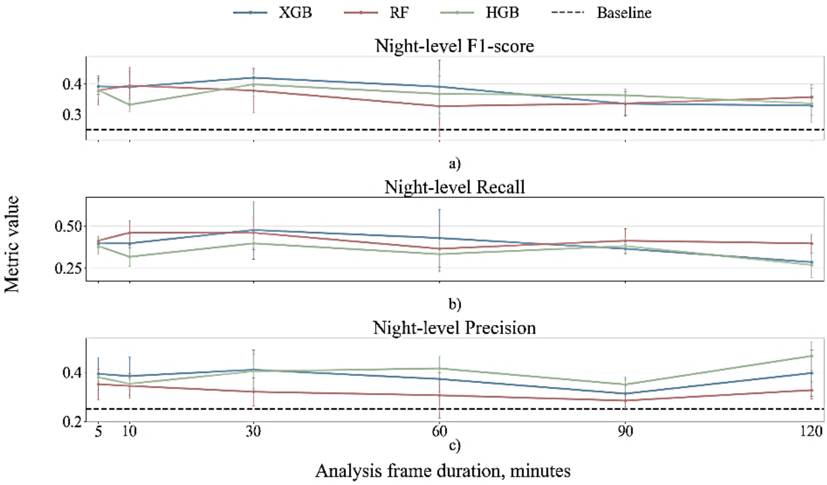


Fig. 3.16. Performance of night-level migraine risk estimation using ensemble models across different analysis-frame durations

The highest performance was generally observed at shorter to intermediate window lengths of 10–30 minutes, suggesting that these temporal resolutions provide a favorable balance between signal preservation and noise reduction.

Analysis of individual performance components revealed that recall tended to be higher at shorter window lengths, while precision showed a slight increase for longer windows. However, this gain in precision was often accompanied by a drop in recall, reflecting a trade-off between sensitivity and specificity in night-level classification. Overall, no single model consistently outperformed the others across all metrics, although ensemble voting provided stable aggregated results across runs.

At the individual participant level, the same multi-run evaluation protocol was applied; however, the resulting performance was substantially less stable. While a subset of participants achieved relatively strong night-level detection with F1-scores exceeding 0.5 at shorter window lengths, the majority of cases showed unstable behavior, frequently resulting in zero positive-class detection. This variability persisted despite averaging across runs and was primarily driven by limited per-subject sample sizes and pronounced class imbalance at the night level.

Importantly, comparison with the majority baseline indicates that improvements over trivial prediction strategies were not consistently achieved across participants, further highlighting the challenges of subject-specific modeling under limited data conditions.

Taken together, these findings suggest that the most reliable and reproducible signal is obtained at the cohort level, where aggregation across participants mitigates data sparsity and variability. At the same time, the presence of positive results for a subset of individuals indicates that a subject-specific signal does exist, albeit in a fragmented and data-limited form.

These results support a two-stage modeling strategy. First, a cohort-level model can capture shared predictive patterns across individuals. Second, this model can be progressively adapted to each participant as additional subject-specific data become available. Such an approach appears particularly well-suited for migraine risk estimation, where the condition is inherently individualized and personalized models are likely to provide greater clinical relevance than fully generic solutions.

3.8.6. System-Level Framework and Deployment Feasibility

The experimental results indicate that pre-migraine patterns can be detected from nocturnal wearable physiological signals, with the most reliable performance observed at the cohort level and at shorter analysis frames. These findings suggest that a practical system should combine a cohort-level modeling approach with subsequent personalization, rather than relying on a single universal configuration. In this context, the proposed framework can be interpreted as an adaptive signal-processing system that progressively refines migraine risk estimation based on individual data.

From a system perspective, the method can be implemented as a multichannel physiological signal-processing pipeline operating on six signals recorded by a wearable device: electrodermal activity, pulse rate, metabolic equivalent, skin temperature, accelerometer variability, and activity counts. These signals jointly represent both autonomic and behavioral aspects of the user’s physiological state during sleep, providing a multimodal basis for prediction. The wearable device continuously acquires signals, which are then transmitted to a local processing unit, such as a smartphone.

The processing pipeline begins with the extraction of the nocturnal interval, followed by segmentation of the signal into analysis frames. Based on the experimental findings, shorter temporal windows provide more informative representations and are therefore preferred in the operational configuration. For each analysis frame, statistical descriptors are computed for all available signals, forming a feature representation of short-term physiological dynamics.

Importantly, the feature extraction stage can be optimized based on the feature analysis results obtained in this research. The experiments demonstrated that the most informative predictors are low-order statistical descriptors, such as mean, median, minimum, and amplitude-related measures, particularly for electrodermal activity and, to a lesser extent, movement-related signals. This allows the system to initially restrict feature computation to a compact subset of low-order descriptors across all signals, reducing computational cost without sacrificing predictive performance.

In addition, the system can incorporate an adaptive feature selection mechanism. During periodic model updates, feature importance can be re-evaluated based on the recurrence of top-ranked features across validation runs. Features that consistently fail to appear among the most informative predictors can be removed from the feature set, while stable features are retained. This results in a progressively refined and participant-specific feature space, improving model robustness and reducing redundancy over time.

The extracted features are evaluated by a trained classification model that produces a probability estimate associated with migraine occurrence. Since the prediction target is defined at the night level, frame-level outputs are aggregated into a single nightly risk estimate. In this work, aggregation emphasizes intervals with stronger deviations by prioritizing higher-probability values, enabling the system to capture transient yet informative physiological patterns.

A key component of the system is the feedback loop that enables continuous adaptation. Each morning, the user receives a risk estimate based on the previous night. Each evening, the user provides a binary label indicating whether a migraine occurred during the day, thereby assigning a label to the preceding night. This process gradually accumulates labeled data, enabling the system to refine its predictive model over time.

Two deployment strategies can be considered. In a fully local implementation, all pipeline stages, including data storage, feature extraction, model inference, and periodic model updates, are performed on a smartphone or a similar edge device. Given the relatively small data volume and the low computational complexity of statistical feature extraction and classical machine learning models, this configuration is technically feasible without requiring cloud infrastructure. Alternatively, a cloud-assisted architecture may be used, in which computationally intensive tasks such as model reconfiguration or cohort-level retraining are performed remotely, while daily prediction remains local. A hybrid approach combining both strategies provides additional flexibility, allowing the system to operate independently while benefiting from optional centralized updates.

From a computational standpoint, the proposed framework is lightweight. The number of signals is limited, and the system operates on aggregated physiological data at one-minute resolution. Feature extraction relies on simple statistical operations, and therefore, both memory requirements and computational load remain modest, making real-time or near-real-time processing feasible on standard consumer devices. Importantly, the system does not rely on signal filtering or computationally expensive transformations, as experimental results demonstrated that filtering does not provide a consistent benefit and may reduce sensitivity to relevant patterns.

Overall, the proposed framework can be viewed as an edge-oriented biomedical signal-processing system that transforms nocturnal multichannel physiological data into a daily migraine risk estimate. The combination of cohort-level modeling, adaptive personalization, and dynamic feature-space optimization makes it suitable for practical deployment in wearable-based health-monitoring applications while maintaining a clear connection to principles of electrical and electronic engineering, including signal acquisition, segmentation, feature extraction, and embedded inference.

3.8.7. Discussion of the Results

Migraine machine-learning research can be broadly divided into two main directions that differ in data type, task formulation, and evaluation methodology. The first direction relies on structured clinical or diary-based datasets and typically addresses diagnostic classification using static features such as symptoms, demographics, or lifestyle factors. In such settings, performance is often evaluated using random train–test splits or standard cross-validation, and reported metrics frequently exceed 90% (Hasan et al., 2024; Imtiaz et al., 2025; Katsuki et al., 2023; Sarra et al., 2025). However, these results are not directly comparable to forecasting tasks, as they do not involve temporal dynamics or the prediction of future events.

The second direction, to which the present research belongs, focuses on wearable physiological signals and aims to predict migraine occurrence in a next-day forecasting setting. This formulation is inherently more challenging due to signal noise, inter-individual variability, class imbalance, and the need to extract predictive information from continuous time series. In this context, previously reported performance is typically moderate, with ROC-AUC values often in the 0.60–0.65 range (Stubberud et al., 2023).

The results obtained in this research are consistent with this range while providing a more detailed methodological analysis of the problem. At the cohort level, classification models achieved ROC-AUC values of approximately 0.62–0.70, with the highest values observed for boosting-based models at shorter analysis-frame durations. PR-AUC values were typically 0.32–0.37, indicating a moderate but consistent ability to detect the minority class under class-imbalance conditions.

Positive-class performance remained moderate, with F1-scores reaching approximately 0.40–0.44 for short analysis frames of 5–10 minutes. For example, HistGradientBoosting achieved a positive-class F1-score of approximately 0.43 at 5-minute windows, while XGBoost reached approximately 0.44 at 10-minute windows. As the analysis-frame duration increased, positive-class F1-scores decreased to approximately 0.32–0.38, primarily due to a reduction in recall. This behavior indicates that longer temporal aggregation reduces sensitivity to pre-migraine states.

Macro-averaged metrics were more stable across configurations, with Macro F1-scores typically in the range of approximately 0.58–0.64, while Accuracy ranged approximately from 0.70 to 0.80. The divergence between macro-level and positive-class metrics highlights the importance of evaluating minority-class performance separately, as global metrics alone may mask reduced sensitivity to clinically relevant events.

A central finding of the research is the strong dependence of predictive performance on temporal resolution. Across all classifiers, shorter analysis frames of 5–10 minutes consistently yielded better results, particularly in terms of positive-class recall and F1-score. This suggests that pre-migraine physiological changes are temporally localized and are better captured at finer temporal resolution. Increasing the analysis-frame duration introduces temporal smoothing, which reduces the discriminative signal. Importantly, this tendency remains observable under strict night-level splitting, indicating that it reflects an intrinsic property of the data rather than an artifact of the evaluation protocol.

Another important observation is the consistent advantage of ensemble-based models. Boosting algorithms, including HistGradientBoosting and XGBoost, achieved the best balance between precision and recall and the highest ROC-AUC values. This supports the interpretation that migraine-related physiological patterns are nonlinear and heterogeneous. In contrast, SVM consistently performed worse, indicating limited suitability for this type of data.

An additional contribution of the present research is the explicit treatment of nocturnal recordings as a distinct analytical context for pre-migraine detection. While many previous studies use daily or mixed-context wearable data, relatively few focus specifically on the night as a controlled physiological regime. The results show that, within this context, short analysis frames and low-order statistical descriptors provide the most informative representation for prediction. In particular, electrodermal activity features consistently emerged as the most stable predictors, supported by movement-related signals.

The applied data selection strategy, including sleep-based filtering and label construction, results in a cohort-level dataset of 455 nights, including 107 pre-migraine nights. This provides sufficient observations for training and evaluating models at the cohort level, particularly in a frame-based classification setting. At the same time, the modeling task involves predicting relatively rare events, which becomes more pronounced as we move from frame-level classification to night-level aggregation and, especially, to individual participant models.

The evaluation was performed at the level of analysis frames, while ensuring that all frames from a given night were assigned exclusively to either the training or the test set. This strategy preserves independence between training and test data at the night level while allowing each night to contribute multiple observations. As a result, the effective number of independent observations is determined by the number of nights, whereas the use of frame-level predictions improves the stability of performance estimates.

Feature analysis further demonstrated that the predictive structure of the problem is concentrated in a limited set of low-order statistical descriptors. Both ANOVA-based analysis and model-based feature importance consistently identified features related to signal level and amplitude, particularly for electrodermal activity, as the most informative. Reducing the feature space to a compact subset of recurrent predictors did not degrade performance and, in some cases, improved it, especially for positive-class detection. This indicates that a small number of stable features captures the essential information needed for prediction.

Overall, the results demonstrate that nocturnal wearable physiological signals contain meaningful information related to next-day migraine risk. While the achieved performance reflects the problem's inherent complexity, the research provides a systematic characterization of how temporal resolution, feature representation, and model choice influence predictive performance. The findings further indicate that cohort-level modeling is currently the most stable approach, whereas transitioning to night-level aggregation and individualized prediction requires additional data for reliable generalization.

3.8.7. Limitations of the Study

The results of this chapter should be interpreted in light of several methodological and practical limitations inherent to the research design and data characteristics.

First, the analysis was restricted to nocturnal physiological recordings extracted using a predefined sleep-gating procedure. This design choice was motivated by the need to reduce behavioral and environmental confounding and ensure comparable conditions across participants. However, the use of strict temporal boundaries and gating rules inevitably limits the number of valid observations. While this controlled setting improves signal consistency and interpretability, it also reduces the number of usable nights and may exclude potentially informative segments outside the defined nocturnal interval. Therefore, the reported findings are specific to the adopted night-selection protocol and may differ under alternative definitions of sleep or extended temporal contexts.

Second, although the cohort-level dataset includes sufficient nights for frame-level classification, the number of migraine-positive nights remains limited, particularly when aggregating nights and using individual participant models. This limitation is intrinsic to real-world longitudinal migraine data, where events are relatively sparse. As a result, while cohort-level patterns can be identified and evaluated with reasonable stability, the transition to night-level prediction and especially to personalized modeling leads to increased variability and reduced reliability of performance estimates.

Third, the modeling framework focused exclusively on wearable-derived physiological signals recorded during sleep. This approach ensures a controlled and consistent input space but does not account for additional factors known to influence migraine occurrence, such as daytime triggers, medication intake, stress levels, hormonal variations, or environmental conditions. Incorporating such multimodal information may improve predictive performance; however, it also introduces challenges related to data completeness, reliability, and synchronization between heterogeneous sources.

Fourth, migraine labels were based on participant-reported diary entries. Although this approach is standard in migraine research, it introduces potential uncertainty in the exact timing and characterization of migraine events. In particular, the definition of the pre-migraine interval depends on self-reported onset, which may vary in accuracy across participants and episodes.

Finally, preprocessing and feature-selection strategies were evaluated within a specific experimental framework that prioritizes statistical descriptors and avoids aggressive signal transformations. While this design was supported by the observed results, the interaction between preprocessing choices, temporal segmentation, and model performance remains complex and may require further investigation in larger datasets.

Despite these limitations, the research provides a structured and reproducible framework for analyzing nocturnal wearable physiological data in the context of migraine prediction. It highlights the importance of controlled data selection, temporal resolution, and feature representation, and it establishes a foundation for future work aimed at improving personalized prediction under real-world data constraints.

3.9. Conclusions of the Third Chapter

This section summarizes the main results of the experimental research investigating the influence of temporal segmentation, signal filtering, and feature representation on migraine classification and night-level prediction.

1. Across both exploratory and cohort-level settings, the highest performance was achieved with short analysis frames of 5–10 minutes. In cohort-level classification, positive-class F1-scores ranged from 0.43 to 0.44 for HistGradientBoosting and XGBoost at short windows, but increased to 90–120 minutes decreased them to approximately 0.32–0.38, primarily due to reduced recall. This indicates that pre-migraine physiological patterns are temporally localized and attenuate with stronger temporal aggregation.
2. When moving from frame-level splitting to night-level splitting, classification performance decreased, reflecting the increased difficulty of generalization to unseen nights. However, the relative advantage of short analysis frames and the superiority of ensemble-based models remained stable, indicating that these effects are not solely due to within-night data overlap.
3. HistGradientBoosting and XGBoost consistently achieved the best results across most analysis-frame durations, with ROC-AUC values in the range of approximately 0.68–0.70 and PR-AUC values around 0.34–0.37. Random Forest showed more stable, but generally lower, positive-class performance, while SVM approached baseline performance with longer analysis frames. KNN did not demonstrate meaningful predictive ability for the positive class.
4. Signal filtering does not provide a stable benefit under strict validation conditions. Although filtering improved performance in the exploratory frame-level setting, these gains were not preserved in the cohort-level evaluation. Under night-level splitting, filtering had a limited and inconsistent effect on Macro F1-score and generally reduced positive-class recall, particularly for Savitzky–Golay filtering. This indicates that smoothing may suppress informative physiological variability relevant to pre-migraine detection.
5. The predictive structure is concentrated in a compact set of low-order physiological features. Both ANOVA analysis and model-based feature importance consistently identified low-order statistical descriptors as the most informative features, particularly for electrodermal activity. Feature-space reduction experiments demonstrated that a compact subset of recurrent features preserves performance and, in some configurations, improves positive-class detection.
6. Night-level migraine risk estimation is feasible at the cohort level but remains unstable at the individual level. At the cohort level, night-level F1-scores ranged from approximately 0.33 to 0.42, exceeding the prevalence-based baseline across all configurations. The best performance was observed at intermediate analysis frames of 10–30 minutes. In contrast,

individual-level models showed high variability and frequently failed to detect positive cases, indicating insufficient data for stable personalized prediction.

7. Night-level migraine risk estimation is feasible at the cohort level but remains unstable at the individual level. At the cohort level, night-level F1-scores ranged from approximately 0.33 to 0.42, exceeding the prevalence-based baseline across all configurations. The best performance was observed at intermediate analysis frames of 10–30 minutes. In contrast, individual-level models showed high variability and frequently failed to detect positive cases, indicating insufficient data for stable personalized prediction.

General Conclusions

1. Migraine prediction using wearable data is highly dependent on methodological design and validation strategy. Comparing temporal segmentation windows ranging from 5 to 120 minutes, along with different preprocessing and validation settings, shows that these methodological choices significantly influence performance estimates. In particular, strict night-level validation reduces the optimistic bias observed in exploratory settings, whereas commonly used filtering techniques do not yield consistent improvements and may reduce sensitivity to the positive class.
2. Nocturnal wearable physiological signals contain detectable but moderately expressed pre-migraine patterns. Under cohort-level evaluation, models achieved ROC-AUC values of approximately 0.62–0.70 and PR-AUC values of approximately 0.32–0.37, indicating consistent but non-trivial predictive capability. However, night-level prediction remains unstable at the individual level, suggesting that the signal is present but not uniformly expressed across participants.
3. Temporal resolution is a critical factor in preserving discriminative information. The highest classification performance was consistently observed for short analysis frames of 5–10 minutes, with positive-class F1-scores reaching approximately 0.43–0.44. Increasing the window length to 90–

120 minutes reduced performance to approximately 0.32–0.38, demonstrating that temporal aggregation attenuates short-term physiological deviations.

4. Ensemble-based models are best suited for wearable-based migraine prediction. HistGradientBoosting and XGBoost provided the most reliable performance across configurations, while Random Forest showed stable but slightly lower results. In contrast, SVM demonstrated limited effectiveness.
5. Predictive information is concentrated in a compact set of low-order features dominated by electrodermal activity. Aggregated feature recurrence analysis shows that electrodermal activity features appear approximately 2–3 times more frequently than movement-related features, while temperature and pulse rate contribute less consistently. This indicates that autonomic nervous system activity is the primary source of predictive signal in the nocturnal context.
6. Cohort-level modeling provides stable and reproducible results, while individual-level prediction remains data-limited and variable. Therefore, practical systems should initially rely on cohort-derived patterns and progressively adapt to individual users as additional labeled data become available.

References

Abromavičius, V., Serackis, A., Katkevičius, A., Kazlauskas, M., & Sledevič, T. (2023). Prediction of exam scores using a multi-sensor approach for wearable exam stress dataset with uniform preprocessing. *Technology and Health Care*, *31*(6), 2499–2511. https://doi.org/10.3233/THC-235015

Agarwal, S., Rani, A., Singh, V., & Mittal, A. P. (2017). EEG signal enhancement using cascaded S-Golay filter. *Biomedical Signal Processing and Control*, *36*, 194–204. https://doi.org/https://doi.org/10.1016/j.bspc.2017.04.004

Ahmadi, M. N., Nathan, N., Sutherland, R., Wolfenden, L., & Trost, S. G. (2020). Non-wear or sleep? Evaluation of five non-wear detection algorithms for raw accelerometer data. *Journal of Sports Sciences*, *38*(4), 399–404. https://doi.org/10.1080/02640414.2019.1703301

Arakaki, X., Arechavala, R. J., Choy, E. H., Bautista, J., Bliss, B., Molloy, C., Wu, D.-A., Shimojo, S., Jiang, Y., Kleinman, M. T., & Kloner, R. A. (2023). The connection between heart rate variability (HRV), neurological health, and cognition: A literature review. *Frontiers in Neuroscience, Vol. 17-2023*. https://doi.org/10.3389/fnins.2023.1055445

Arnold, V. X., & Young, S. D. (2025). The Potential of Wearable Sensors for Detecting Cognitive Rumination: A Scoping Review. *Sensors*, *25*(3), 654. https://doi.org/10.3390/s25030654

Badawi, A., Elmoghazy, S., Choudhury, S., Elgazzar, S., Elgazzar, K., & Burhan, A. M. (2025). Multimodal Detection of Agitation in People With Dementia in Clinical Settings: Observational Pilot Study. *JMIR Aging*, *8*, e68156. https://doi.org/10.2196/68156

Balam, V. P. (2025). Automated EEG signal processing: A comprehensive investigation into preprocessing techniques and sub-band extraction for enhanced brain-computer interface applications. *Journal of Neuroscience Methods*, *424*, Article 110561. https://doi.org/10.1016/j.jneumeth.2025.110561

Baliarsingh, S., Panda, P. K., & Mohanty, M. N. (2023). Unsupervised technique for cardiac data compression with β wavelet transform for telemedicine. *Measurement: Sensors*, *26*, Article 100671. https://doi.org/10.1016/j.measen.2023.100671

Barakat, A., Vogelzangs, N., Licht, C. M. M., Geenen, R., Macfarlane, G. J., Geus, E. J. C. de, Smit, J. H., Penninx, B. W. J. H., & Dekker, J. (2012). Dysregulation of the autonomic nervous system and its association with the presence and intensity of chronic widespread pain. *Arthritis Care & Research*, *64*, *1209-1216*. https://doi.org/10.1002/acr.21669

Bellocchi, C., Carandina, A., Montinaro, B., Targetti, E., Furlan, L., Rodrigues, G. D., Tobaldini, E., & Montano, N. (2022). The Interplay between Autonomic Nervous System and Inflammation across Systemic Autoimmune Diseases. *International Journal of Molecular Sciences*, *23*(5), 2449. https://doi.org/10.3390/ijms23052449

Bota, P., Silva, R., Carreiras, C., Fred, A., & Silva, H. P. da. (2024). BioSPPy: A Python toolbox for physiological signal processing. *SoftwareX*, *26*, Article 101712. https://doi.org/10.1016/j.softx.2024.101712

Buse, D. C., McGinley, J. S., & Lipton, R. B. (2020). Predicting the Future of Migraine Attack Prediction. *Headache: The Journal of Head and Face Pain*, *60*(10), 2125–2128. https://doi.org/10.1111/head.14025

Butt, M. O., Mir, A., & Sujjada, A. (2025). A Novel Data-Driven Framework for Automated Migraines Classification Using Ensemble Learning. *Engineering Proceedings*, *107*(1), 25. https://doi.org/10.3390/engproc2025107025

Campanella, S., Altaleb, A., Belli, A., Pierleoni, P., & Palma, L. (2024). PPG and EDA dataset collected with Empatica E4 for stress assessment. *Data in Brief*, *53*, Article 110102. https://doi.org/10.1016/j.dib.2024.110102

Carvalho, I. V., Fernandes, C. S., Damas, D. P., Barros, F. M., Gomes, I. R., Gens, H. M., & Luzeiro, I. (2022). The migraine postdrome: Clinical characterization, influence of abortive treatment and impact in the quality of life. *Clinical Neurology and Neurosurgery*, *221*, Article 107408. https://doi.org/10.1016/j.clineuro.2022.107408

Chaitanya, M. K., & Sharma, L. D. (2022). Electrocardiogram signal filtering using circulant singular spectrum analysis and cascaded Savitzky-Golay filter. *Biomedical Signal Processing and Control*, *75*, Article 103583. https://doi.org/10.1016/j.bspc.2022.103583

Chiang, C.-C., Chhabra, N., Chao, C.-J., Wang, H., Zhang, N., Lim, E., Baez-Suarez, A., Attia, Z. I., Schwedt, T. J., Dodick, D. W., Cutrer, F. M., Friedman, P. A., & Noseworthy, P. A. (2022). Migraine with aura associates with a higher artificial intelligence: ECG atrial fibrillation prediction model output compared to migraine without aura in both women and men. *Headache: The Journal of Head and Face Pain*, *62*(8), 939–951. https://doi.org/10.1111/head.14339

Cortese, F., Coppola, G., Di Lenola, D., Serrao, M., Di Lorenzo, C., Parisi, V., & Pierelli, F. (2017). Excitability of the motor cortex in patients with migraine changes with the time elapsed from the last attack. *The Journal of Headache and Pain*, *18*(1), 2. https://doi.org/10.1186/s10194-016-0712-z

D'Agnano, D., Cernigliaro, F., Ferretti, A., Cascio, S. L., Correnti, E., Terrin, G., Santangelo, A., Bellone, G., Raieli, V., Sciruicchio, V., & others. (2024). The Role of the Autonomic Nervous System in Epilepsy and Migraine: A Narrative Review. *Journal of Integrative Neuroscience*, *23*(7), 128. https://doi.org/10.31083/j.jin2307128

Danelakis, A., Stubberud, A., Tronvik, E., & Matharu, M. (2025). The Emerging Clinical Relevance of Artificial Intelligence, Data Science, and Wearable Devices in Headache: A Narrative Review. *Life*, *15*(6). https://doi.org/10.3390/life15060909

De Filippis, R., & Al Foysal, A. (2025). The Impact of Machine Learning in Identifying Migraine Types: A Data-Driven Approach. *OALib*, *12*, 1–17. https://doi.org/10.4236/oalib.1113195

Dodick, D. W. (2018). A Phase-by-Phase Review of Migraine Pathophysiology. *Headache: The Journal of Head and Face Pain*, *58*(S1), 4–16. https://doi.org/10.1111/head.13300

Dominguez, M., Ashina, S., Yazdi, C., Simopoulos, T. T., Hasoon, J. J., Kaye, A. D., & Robinson, C. L. (2025). Introduction to migraine: Current concepts, definitions, and diagnosis. *Interventional Management of Migraines and Other Headache Disorders*, 1–9. https://doi.org/10.1016/B978-0-443-23557-3.00001-2

Eigenbrodt, A. K., Christensen, R. H., Ashina, H., Iljazi, A., Christensen, C. E., Steiner, T. J., Lipton, R. B., & Ashina, M. (2022). Premonitory symptoms in migraine: A systematic review and meta-analysis of observational studies reporting prevalence or relative frequency. *The Journal of Headache and Pain*, *23*(1), 140. https://doi.org/10.1186/s10194-022-01510-z

Empatica. (2025). *Empatica Embrace Plus*. https://www.empatica.com/en-int/embraceplus/

Faust, L., Cui, J., Knepper, C., Nasseri, M., Worrell, G., & Brinkmann, B. H. (2025). Detecting Diverse Seizure Types with Wrist-Worn Wearable Devices: A Comparison of Machine Learning Approaches. *Sensors*, *25*(17), 5562. https://doi.org/10.3390/s25175562

Feng, H., M. Golshan, H., & Mahoor, M. (2018). A Wavelet-based Approach to Emotion Classification using EDA Signals. *Expert Systems with Applications*, *112, 77-86*. https://doi.org/10.1016/j.eswa.2018.06.014

Feng, H., & Sun, L. (2024). Design and Simulation of Butterworth Low-pass Filter. *Academic Journal of Science and Technology*, *10*(3), 106–110. https://doi.org/10.54097/jpnnb366

Ferrari, M. D., Klever, R. R., Terwindt, G. M., Ayata, C., & van den Maagdenberg, A. M. (2015). Migraine pathophysiology: Lessons from mouse models and human genetics. *The Lancet Neurology*, *14*(1), 65–80. https://doi.org/10.1016/S1474-4422(14)70220-0

Fisher, R. A. (1992). Statistical Methods for Research Workers. In S. Kotz & N. L. Johnson (Eds), *Breakthroughs in Statistics: Methodology and Distribution* (pp. 66–70). Springer New York. https://doi.org/10.1007/978-1-4612-4380-9_6

Flamerz Arkawazi, B., Hatem, A., & Alkhazrajy, L. (2021). Autonomic symptoms in patients with migraine. *Revista Latinoamericana de Hipertension*, *15*, 211–215.

Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLOS One*, *18*(10), e0291908. https://doi.org/10.1371/journal.pone.0291908

Fricke, C., Alizadeh, J., Zakhary, N., Woost, T. B., Bogdan, M., & Classen, J. (2021). Evaluation of Three Machine Learning Algorithms for the Automatic Classification of EMG Patterns in Gait Disorders. *Frontiers in Neurology*, . https://doi.org/10.3389/fneur.2021.666458

Gallagher, N. B. (2020). Savitzky-Golay smoothing and differentiation filter. *Eigenvector Research Incorporated*, *2*, *4*.

Galli, S. (2021). Feature-engine: A Python package for feature engineering for machine learning. *Journal of Open Source Software*, *6*(65), 3642. https://doi.org/10.21105/joss.03642

Gao, L., Zhao, F., Tu, Y., & Liu, K. (2024). The prodrome of migraine: Mechanistic insights and emerging therapeutic strategies. *Frontiers in Neurology, 15, 1496401* . https://doi.org/10.3389/fneur.2024.1496401

Gomes, P., Margaritoff, P., & Silva, H. (2019). pyHRV: Development and evaluation of an open-source Python toolbox for heart rate variability (HRV). In *Proceedings of the International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)* (pp. 822–828). Veliko Gradište, Serbia.

Gonzalez-Martinez, A., Gálvez-Goicurría, J., Pagán, J., Quintas, S., Vieira, A., Ramiro, C. A., Sobrado, M., Ayala, J. L., Vivancos, J., & Gago-Veiga, A. B. (2022). P190 Prediction of a migraine crisis in real-time using a wearable device. *16th European Headache Congress 2022 Meeting Abstracts: Vienna, Austria. 7-10 December 2022. The Journal of Headache and Pain.*, *23*(Suppl 1). https://doi.org/10.1186/S10194-022-01527-4

Hansen, J. M., & Charles, A. (2019). Differences in treatment response between migraine with aura and migraine without aura: Lessons from clinical practice and RCTs. *The Journal of Headache and Pain*, *20*(1), 96. https://doi.org/10.1186/s10194-019-1046-4

Harmsen, R., Hansen, J. M., Matre, D., Arup, A. E. S. F., Garde, A. H., & Nabe-Nielsen, K. (2025). The acute effect of night work-related circadian misalignment on headache episodes: Results from the 1001 nights-cohort. *Headache: The Journal of Head and Face Pain*, *65*(9), 1554–1564. https://doi.org/https://doi.org/10.1111/head.15054

Hasan, F., Khatun, R., Salman, E. M., Mahmud, T., Raza, D. M., & Nahid, A. H. (2024). Deciphering the Symptom Spectrum: A Comprehensive Analysis of Migraine Patterns and Types. *American Journal of Data Mining and Knowledge Discovery*, *9*(1), 20–31. https://doi.org/10.11648/j.ajdmkd.20240901.12

Hougaard, A., Ayata, C., Brennan, K. C., van den Maagdenberg, A. M. J. M., & Ashina, M. (2025). The mysterious link between migraine aura and migraine headache. *PLOS Biology*, *23*(6), 1–19. https://doi.org/10.1371/journal.pbio.3003168

Houle, T. T., Turner, D. P., Golding, A. N., Porter, J. A. H., Martin, V. T., Penzien, D. B., & Tegeler, C. H. (2017). Forecasting Individual Headache Attacks Using Perceived

Stress: Development of a Multivariable Prediction Model for Persons With Episodic Migraine. *Headache: The Journal of Head and Face Pain*, *57*(7), 1041–1050. https://doi.org/https://doi.org/10.1111/head.13137

Hsiao, F.-J., Chen, W.-T., Pan, L.-L. H., Liu, H.-Y., Wang, Y.-F., Chen, S.-P., Lai, K.-L., Coppola, G., & Wang, S.-J. (2022). Dynamic brainstem and somatosensory cortical excitability during migraine cycles. *The Journal of Headache and Pain*, *23*(1), 21. https://doi.org/10.1186/s10194-022-01392-1

Imtiaz, I., Afzal, H., Rehman, A. U., & Insany, G. P. (2025). Evaluating the Role of Machine Learning in Migraine Detection and Classification. *Engineering Proceedings*, *107*(1*)*, 122. https://doi.org/10.3390/engproc2025107122

Jaychandran, R., Chaitanya, G., Satishchandra, P., Bharath, R. D., Thennarasu, K., & Sinha, S. (2016). Monitoring peri-ictal changes in heart rate variability, oxygen saturation and blood pressure in epilepsy monitoring unit. *Epilepsy Research*, *125*, 10–18. https://doi.org/10.1016/j.eplepsyres.2016.05.013

Justusson, B. (2006). Median filtering: Statistical properties. *Two-Dimensional Digital Signal Prcessing II: Transforms and Median Filters*, 161–196.

Kannan, A., West, D., Kumbhare, D., Ting, W.-T., Ali, M. Y., Kawsar, H. I., Singh, G., Shanthanna, H., Hapidou, E., & Howlader, M. M. R. (2025). An analytical review of biosensor-based chronic pain quantification in healthcare. *Healthcare Analytics*, *8*, Article 100419. https://doi.org/https://doi.org/10.1016/j.health.2025.100419

Kapustynska, V., Abromavičius, V., Serackis, A., Andruškevičius, S., Ryliškienė, K., & Paulikas, Š. (2025). Advancing a generalizable model for migraine prediction: Analysis of filtering techniques on physiological signals. *Technology and Health Care*, *33*(5), 2184–2193. https://doi.org/10.1177/09287329251332415

Kapustynska, V., Abromavičius, V., Serackis, A., Paulikas, Š., Ryliškienė, K., & Andruškevičius, S. (2024). Machine learning and wearable technology: Monitoring changes in biomedical signal patterns during pre-migraine nights. *Healthcare*, *12*, Article 1701.

Karsan, N., Silva, E., & Goadsby, P. J. (2023). Evaluating migraine with typical aura with neuroimaging. *Frontiers in Human Neuroscience*, *17, 1112790*. https://doi.org/10.3389/fnhum.2023.1112790

Katsuki, M., Matsumori, Y., Kawamura, S., Kashiwagi, K., Koh, A., Tachikawa, S., & Yamagishi, F. (2023). Developing an artificial intelligence–based diagnostic model of headaches from a dataset of clinic patients' records. *Headache: The Journal of Head and Face Pain*, *63*(8), 1097–1108. https://doi.org/https://doi.org/10.1111/head.14611

Kellier, D. J., de Prado, B. M., Haagen, D., Grabner, P., Raj, N., Lechtenberg, L., Velasquez, G., Hsu, J. Y., Farrar, J. T., Hershey, A. D., & Szperka, C. L. (2023). Validity of self-reported migraine in adolescents and children. *Headache: The Journal of Head and Face Pain*, *63*(5), 634–641. https://doi.org/10.1111/head.14498

Kim, J.-H., Kim, H.-S., Sohn, J.-H., Hwang, S.-M., Lee, J.-J., & Kwon, Y.-S. (2025). Functional Disability and Psychological Impact in Headache Patients: A Comparative Study Using Conventional Statistics and Machine Learning Analysis. *Medicina*, *61*(2), 188. https://doi.org/10.3390/medicina61020188

Koskimäki, H., Mönttinen, H., Siirtola, P., Huttunen, H.-L., Halonen, R., & Röning, J. (2017). Early detection of migraine attacks based on wearable sensors: Experiences of data collection using Empatica E4. *Proceedings of the 2017 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2017 ACM International Symposium on Wearable Computers,* 506–511. https://doi.org/10.1145/3123024.3124434

Kozumplík, J., Smital, L., Němcová, A., Ronzhina, M., Smisek, R., Maršánová, L., Kralik, M., & Vitek, M. (2021). *Respiratory Rate Estimation Using the Photoplethysmogram: Towards the Implementation in Wearables,* 1–4. https://doi.org/10.23919/CinC53138.2021.9662674

Laurell, K., Artto, V., Bendtsen, L., Hagen, K., Häggström, J., Linde, M., Söderström, L., Tronvik, E., Wessman, M., Zwart, J. A., & others. (2016). Premonitory symptoms in migraine: A cross-sectional study in 2714 persons. *Cephalalgia*, *36*(10), 951–959. https://doi.org/10.1177/0333102415620251

Li, K., Cardoso, C., Moctezuma-Ramirez, A., Elgalad, A., & Perin, E. (2023). Heart Rate Variability Measurement through a Smart Wearable Device: Another Breakthrough for Personal Health Monitoring? *International Journal of Environmental Research and Public Health*, *20*(24), 7146. https://doi.org/10.3390/ijerph20247146

Li Zhongshen. (2007). Design and Analysis of Improved Butterworth Low Pass Filter. *2007 8th International Conference on Electronic Measurement and Instruments*, 1-729-1-732. https://doi.org/10.1109/ICEMI.2007.4350554

Lipton, R. B., Lanteri-Minet, M., Leroux, E., Adams, A. M., Lama, J. C.-D., Reed, M. L., Fanning, K. M., & Buse, D. C. (2023). Pre- and post-headache phases of migraine: Multi-country results from the CaMEO - International Study. *The Journal of Headache and Pain*, *24*(1), 151. https://doi.org/10.1186/s10194-023-01683-1

Lipton, R. B., Stokes, J., Evans, C. J., Hribal, E., White, K., Keyloun, K., Parikh, K., Gandhi, P., & Dodick, D. W. (2025). Characterizing the patient experience during the prodrome phase of migraine: A qualitative study of symptoms and their timing. *Headache: The Journal of Head and Face Pain*, *65*(8), 1355–1368. https://doi.org/10.1111/head.15024

Liu, B., & Li, Y. (2021). ECG signal denoising based on similar segments cooperative filtering. *Biomedical Signal Processing and Control*, *68,* Article 102751. https://doi.org/10.1016/j.bspc.2021.102751

Lucas, C. (2021). Migraine with aura. *Revue Neurologique*, *177*(7), 779–784. https://doi.org/10.1016/j.neurol.2021.07.010

Lustrek, M., Cvetković, B., & Kozina, S. (2012). *Energy expenditure estimation with wearable accelerometers*, 5–8. https://doi.org/10.1109/ISCAS.2012.6271906

Maczák, B., Vadai, G., Dér, A., Szendi, I., & Gingl, Z. (2021). Detailed analysis and comparison of different activity metrics. *PLOS One, 16*(12), 1–28. https://doi.org/10.1371/journal.pone.0261718

Mehnert, J., Fischer-Schulte, L., & May, A. (2023). Aura phenomena do not initiate migraine attacks—Findings from neuroimaging. *Headache: The Journal of Head and Face Pain*, *63*(8), 1040–1044. https://doi.org/10.1111/head.14597

Menon, B., Pillai, A. S., Mathew, P. S., & Bartkowiak, A. M. (2022). Artificial intelligence–assisted headache classification: A review. *Augmenting Neurological Disorder Prediction and Rehabilitation Using Artificial Intelligence*, 145–162. https://doi.org/10.1016/B978-0-323-90037-9.00007-2

Meredith, D. J., Clifton, D., Charlton, P., Brooks, J., Pugh, C. W., & Tarassenko, L. (2012). Photoplethysmographic derivation of respiratory rate: A review of relevant physiology. *Journal of Medical Engineering & Technology*, *36*(1), 1–7. https://doi.org/10.3109/03091902.2011.638965

Messina, R., Sudre, C. H., Wei, D. Y., Filippi, M., Ourselin, S., & Goadsby, P. J. (2023). Biomarkers of Migraine and Cluster Headache: Differences and Similarities. *Annals of Neurology*, *93*(4), 729–742. https://doi.org/10.1002/ana.26583

Moser, M. K., Ehrhart, M., & Resch, B. (2024). An Explainable Deep Learning Approach for Stress Detection in Wearable Sensor Measurements. *Sensors*, *24*(16). https://doi.org/10.3390/s24165085

Nodes, T., & Gallagher, N. (1982). Median filters: Some modifications and their properties. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, *30*(5), 739–746. https://doi.org/10.1109/TASSP.1982.1163951

Ohal, H. S., & Mantri, S. (2024). Exploring EEG-Based biomarkers for improved early Alzheimer’s disease detection: A feature-based approach utilizing machine learning. *Measurement: Sensors*, *36*, Article 101403. https://doi.org/10.1016/j.measen.2024.101403

Pagán, J., Risco-Martín, J. L., Moya, J. M., & Ayala, J. L. (2016). Modeling methodology for the accurate and prompt prediction of symptomatic events in chronic diseases. *Journal of Biomedical Informatics*, *62*, 136–147. https://doi.org/10.1016/j.jbi.2016.05.008

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., & Louppe, G. (2012). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*.

Puledda, F., Viganò, A., Sebastianelli, G., Parisi, V., Hsiao, F.-J., Wang, S.-J., Chen, W.-T., Massimini, M., & Coppola, G. (2023). Electrophysiological findings in migraine may reflect abnormal synaptic plasticity mechanisms: A narrative review. *Cephalalgia*, *43*(8), 03331024231195780. https://doi.org/10.1177/03331024231195780

Radha, M., Fonseca, P., Moreau, A., Ross, M., Cerny, A., Anderer, P., Long, X., & Aarts, R. M. (2021). A deep transfer learning approach for wearable sleep stage classification with photoplethysmography. *Npj Digital Medicine*, *4*(1), 135. https://doi.org/10.1038/s41746-021-00510-8

Raggi, A., Leonardi, M., Arruda, M., Caponnetto, V., Castaldo, M., Coppola, G., Della Pietra, A., Fan, X., Garcia-Azorin, D., Gazerani, P., Grangeon, L., Grazzi, L., Hsiao, F.-J., Ihara, K., Labastida-Ramirez, A., Lange, K. S., Lisicki, M., Marcassoli, A., Montisano, D. A., … Martelletti, P. (2024). Hallmarks of primary headache: Part 1 – migraine. *The Journal of Headache and Pain*, *25*(1), 189. https://doi.org/10.1186/s10194-024-01889-x

Ridgman, W. (1990). Statistical Methods, 8th edn, by GW Snedecor & WG Cochran. Xx+ 503 pp. Ames: Iowa State University Press (1989). $44.95 (hard covers). ISBN 0 8138 1561 6. *The Journal of Agricultural Science*, *115*(1), 153–153. https://doi.org/10.1017/S0021859600074104

Rio, M. S. del, & Cutrer, F. M. (2023). Chapter 6—Pathophysiology of migraine aura. In J. W. Swanson & M. Matharu (Eds), *Migraine Biology, Diagnosis, and Co-Morbidities* (Vol. 198, pp. 71–83). Elsevier. https://doi.org/10.1016/B978-0-12-823356-6.00016-0

Romo-Cardenas, G., Avilés Rodríguez, G., Sanchez Lopez, J. de D., Cosío León, M., Luque, P., Gómez, C., Nieto, J., Vazquez, M., & Navarro, C. (2018). Nyquist-Shannon theorem application for Savitzky-Golay smoothing window size parameter determination in bio-optical signals. *Results in Physics, 11, 17-22*. https://doi.org/10.1016/j.rinp.2018.08.033

Sakalauskas, V., & Kriksciuniene, D. (2025). Binary Classification Models for Stroke Outcome Prediction. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *16*(2), 296–310. https://doi.org/10.70594/brain/16.2/22

Sanches, C. A., Silva, G. A., Librantz, A. F. H., Sampaio, L. M. M., & Belan, P. A. (2023). Wearable Devices to Diagnose and Monitor the Progression of COVID-19 Through Heart Rate Variability Measurement: Systematic Review and Meta-Analysis. *J Med Internet Res*, *25*, e47112. https://doi.org/10.2196/47112

Sarkis, R. A., Thome-Souza, S., Poh, M.-Z., Llewellyn, N., Klehm, J., Madsen, J. R., Picard, R., Pennell, P. B., Dworetzky, B. A., Loddenkemper, T., & Reinsberger, C. (2015). Autonomic changes following generalized tonic clonic seizures: An analysis of adult and pediatric patients with epilepsy. *Epilepsy Research*, *115*, 113–118. https://doi.org/10.1016/j.eplepsyres.2015.06.005

Sarra, R. R., Korial, A. E., Gorial, I. I., & Humaidi, A. J. (2025). Enhancing Migraine Classification Through Machine Learning: A Comparative Study of Ensemble Methods. *Technologies*, *13*(11), 500. https://doi.org/10.3390/technologies13110500

Savitzky, Abraham., & Golay, M. J. E. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Analytical Chemistry*, *36*(8), 1627–1639. https://doi.org/10.1021/ac60214a047

Schwedt, T. J., Lipton, R. B., Goadsby, P. J., Chiang, C.-C., Klein, B. C., Hussar, C., Liu, C., Yu, S. Y., Finnegan, M., & Trugman, J. M. (2025). Characterizing prodrome (premonitory phase) in migraine: Results from the PRODROME trial screening period. *Neurology: Clinical Practice*, *15*(1), e200359. https://doi.org/10.1212/CPJ.0000000000200359

Schwedt, T. J., Lipton, R. B., Goadsby, P. J., Chiang, C.-C., Klein, B., Liu, C., Yu, S. Y., Severt, L., Finnegan, M., & Trugman, J. M. (2023). Characterizing Prodrome (Premonitory Phase) in Migraine: Results From the PRODROME Trial Screening Period (S41.009). *Neurology*, *100*(17 Supplement 2): 1669. https://doi.org/10.1212/WNL.0000000000202046

Shaffer, F., Combatalade, D., Peper, E., & Meehan, Z. (2016). A Guide to Cleaner Electrodermal Activity Measurements. *Biofeedback*, *44*, 90–100. https://doi.org/10.5298/1081-5937-44.2.01

Sharma, A., & Verbeke, W. J. M. I. (2020). Improving Diagnosis of Depression With XGBOOST Machine Learning Model and a Large Biomarkers Dutch Dataset (n = 11,081). *Frontiers in Big Data*, *3*, 15. https://doi.org/10.3389/fdata.2020.00015

Siirtola, P., Koskimäki, H., Mönttinen, H., & Röning, J. (2018). Using Sleep Time Data from Wearable Sensors for Early Detection of Migraine Attacks. *Sensors (Basel, Switzerland)*, *18*(5), Article 1374. https://doi.org/10.3390/S18051374

Simon, L. (2025). From blood to body tissues: A dynamic framework for estimating volatile organic compound exposure using Kalman filtering and physiological models. *Computational Toxicology*, *34*, Article 100362. https://doi.org/10.1016/j.comtox.2025.100362

Singh, S., & Roy, A. (2017). Assessment of heart rate variability in the patients suffering with chronic pain of musculoskeletal origin. *National Journal of Physiology, Pharmacy and Pharmacology*, *7*, 1. https://doi.org/10.5455/njppp.2017.7.0204803032017

Society (IHS), H. C. C. of the I. H. (2018). The international classification of headache disorders, (beta version). *Cephalalgia*, *33*(9), 629–808. https://doi.org/10.1177/0333102413485658

Stritzelberger, J., Kirmse, M., Borutta, M. C., Gollwitzer, S., Reindl, C., Welte, T. M., Hamer, H. M., & Koehn, J. (2025). Validity of Empatica E4 Wristband for Detection of Autonomic Dysfunction Compared to Established Laboratory Testing. *Diagnostics*, *15*(20), 2604. https://doi.org/10.3390/diagnostics15202604

Stubberud, A., Ingvaldsen, S. H., Brenner, E., Winnberg, I., Olsen, A., Gravdahl, G. B., Matharu, M. S., Nachev, P., & Tronvik, E. (2023). Forecasting migraine with machine learning based on mobile phone diary and wearable data. *Cephalalgia*, *43*(5). https://doi.org/10.1177/03331024231169244

Svensson, M., & Borgström, O. (2026). *Predicting Migraine Onset Using Machine Learning and Daily Lifestyle Data*. Linnaeus University, Department of computer science and media technology (CM).

Tsai, Y.-H., Chiou, C.-L., Lee, J.-J., Yang, C.-M., Lin, kao-chang, Cheng, T.-J., & Chuang, C.-H. (2025). *Personalized prediction of the following-day migraine attacks: A machine learning approach based on digital headache diary records*. https://doi.org/10.21203/rs.3.rs-7226289/v1 van den Hoek, T. C., van de Ruit, M., Terwindt, G. M., & Tolner, E. A. (2024). EEG Changes in Migraine—Can EEG Help to Monitor Attack Susceptibility? *Brain Sciences*, *14*(5), 508. https://doi.org/10.3390/brainsci14050508

Viana, M., Tronvik, E. A., Do, T. P., Zecca, C., & Hougaard, A. (2019). Clinical features of visual migraine aura: A systematic review. *The Journal of Headache and Pain*, *20*(1), 64. https://doi.org/10.1186/s10194-019-1008-x

Vincent, M., Viktrup, L., Nicholson, R. A., Ossipov, M. H., & Vargas, B. B. (2022). The not so hidden impact of interictal burden in migraine: A narrative review. *Frontiers in Neurology, 13,* 1032103. https://doi.org/10.3389/fneur.2022.1032103

Vu, N., Huỳnh, V. T., Kim, S.-W., Shin, J.-E., Yang, H.-J., & Kim, S. H. (2025). Switch fusion for continuous emotion estimation from multiple physiological signals. *Biomedical Signal Processing and Control*, *107,* 107831. https://doi.org/10.1016/j.bspc.2025.107831

Vujovic, Ž. Đ. (2021). Classification Model Evaluation Metrics. *International Journal of Advanced Computer Science and Applications*, *12*(6), 599-606. https://doi.org/10.14569/IJACSA.2021.0120670

Xie, S., Huo, Y., Geng, X., Hu, L., Ao, R., Su, H., Li, D., Hu, M., Liu, Y., Chen, X., Liu, J., Li, Q., Lin, J., Yu, S., & Dong, Z. (n.d.). Exploring the alterations in microstate dynamics during the migraine cycle and detecting pre-ictal phases. *Headache: The Journal of Head and Face Pain*, *66, 2,* 457-469. https://doi.org/10.1111/head.15031

Zhang, Y., Xu, J., Xie, M., Zhu, D., Song, H., & Wang, W. (2024). Efficient and Direct Inference of Heart Rate Variability using Both Signal Processing and Machine Learning. *Proceedings of the 8th ACM/IEEE International Conference on Connected Health: Applications, Systems and Engineering Technologies, CHASE '23*, 102–113. https://doi.org/10.1145/3580252.3586971

List of Scientific Publications by the Author on the Topic of the Dissertation

Papers in the Reviewed Scientific Journals

Kapustynska, V., Abromavičius, V., Serackis, A., Andruškevičius, S., Ryliškienė, K., & Paulikas, Š. (2025). Advancing a generalizable model for migraine prediction: Analysis of filtering techniques on physiological signals. *Technology and Health Care, 33*(5), 2184–2193. https://doi.org/10.1177/09287329251332415

Kapustynska, V., Abromavičius, V., Serackis, A., Paulikas, Š., Ryliškienė, K., & Andruškevičius, S. (2024). Machine Learning and Wearable Technology: Monitoring Changes in Biomedical Signal Patterns during Pre-Migraine Nights. *Healthcare, 12*(17), Article 1701. https://doi.org/10.3390/healthcare12171701

Papers in Other Editions

Kapustynska, V., Abromavičius, V., Serackis, A., Paulikas, Š., Ryliškienė, K., & Andruškevičius, S. (2024). Detecting pre-migraine night patterns with wearable biosensors and machine learning. *Vilnius University Proceedings, 52* (*DAMSS: 15th conference on data analysis methods for software systems*, Druskininkai, Lithuania, November 28–30, 2024), 47–48. https://doi.org/10.15388/DAMSS.15.2024

Kapustynska, V., & Paulikas, Š. (2024). Analysis of electrodermal activity signal fluctuations prior to migraine onset. *Grail of Science*, *42*, 342–344. https://doi.org/10.36074/grail-of-science.02.08.2024.048

Summary in Lithuanian

Įvadas

Problemos formulavimas

Migrena yra sudėtingas neurologinis sutrikimas, sukeliantis didelę naštą pacientams ir sveikatos priežiūros sistemoms. Jos priepuolius sunku prognozuoti, o prodrominiai simptomai skiriasi tarp individų. Dėvimieji įrenginiai suteikia galimybę nuolat stebėti fiziologinius ir elgsenos signalus, kurie gali atspindėti priešmigreninius pokyčius, tačiau realaus pasaulio duomenys dažnai yra triukšmingi, neišsamūs ir stipriai veikiami miego, fizinio aktyvumo bei cirkadinių ritmų.

Šioje disertacijoje nagrinėjama patikimos ir interpretuojamos priešmigreninės informacijos išskyrimo iš dėvimųjų įrenginių fiziologinių duomenų problema, kuri lemia poreikį kurti metodus patikimam požymių išgavimui ir interpretavimui. Toks poreikis kyla dėl tradicinių signalų apdorojimo ir klasifikavimo metodų ribotumo, kai jie taikomi triukšmingiems, nestacionariems signalams, surinktiems realiomis sąlygomis, todėl ši problema yra itin aktuali elektros ir elektronikos inžinerijos srityje.

Šiai problemai spręsti disertacijoje kuriamas kontekstui jautrus, daugiapakopis analitinis pagrindas, apimantis duomenų auditą, kontekstui jautrų žymėjimą bei sistemingą laiko agregavimo, išankstinio apdorojimo ir požymių reprezentavimo vertinimą naktinių duomenų pagrindu. Migrenos klasifikacijai ir kitos dienos migrenos prognozei vertinti taikomi mašininio mokymosi metodai, o požymių svarbos analizė naudojama identifikuoti informatyvius signalus, jų charakteristikas ir pasikartojančius modelius tarp skirtingų dalyvių ir modeliavimo nustatymų.

Remiantis šia formuluote, disertacija grindžiama dviem hipotezėmis:

1. Priešmigreninių fiziologinių pokyčių aptinkamumas stipriai priklauso nuo signalų apdorojimo ir duomenų modeliavimo sprendimų, ypač nuo laiko agregavimo, išankstinio apdorojimo, požymių išgavimo ir jų reprezentavimo.
2. Nepaisant individualių skirtumų, kohortos lygmeniu galima identifikuoti pasikartojančius fiziologinius dėsningumus ir informatyvių požymių grupes, atspindinčias bendrą priešmigreninių procesų struktūrą.

Darbo aktualumas

Šios disertacijos aktualumą lemia poreikis gilinti supratimą apie fiziologinius pokyčius prieš migrenos priepuolius ir užpildyti esamas mokslinių tyrimų spragas. Naudojant fiziologinius signalus, surinktus ilgalaikio realaus gyvenimo stebėjimo metu, vertinama duomenų kokybė, validacijos strategijos ir migrenos klasifikavimo bei kitos dienos prognozavimo galimybės, taip pat analizuojamas signalų segmentavimo, filtravimo ir naktinių požymių svarbos poveikis.

Tyrimo objektas

Šios disertacijos tyrimo objektas yra dėvimųjų įrenginių fiziologinių duomenų analizės sistema, skirta priešmigreninių požymių nustatymui ir analizės sprendimų įtakos jų aptikimui bei kitos dienos migrenos prognozavimo galimybei vertinti.

Darbo tikslas

Šios disertacijos tikslas – sukurti ir įvertinti atkuriamą metodinę sistemą, skirtą priešmigreninei informacijai išgauti ir interpretuoti iš nešiojamųjų įrenginių registruojamų fiziologinių duomenų natūraliomis stebėjimo sąlygomis.

Darbo uždaviniai

1. Sukurti dėvimųjų įrenginių duomenų apdorojimo grandinę migrenos analizei, apimančią dalyvio lygmens duomenų rinkinio sudarymą, naktinių stebėjimų apibrėžimą ir nuoseklų priešmigreninių bei kontrolinių laikotarpių žymėjimą.
2. Sukurti ir įvertinti dvejetainius klasifikavimo modelius, skirtus priešmigreninėms ir kontrolinėms būsenoms atskirti, naudojant iš dėvimųjų įrenginių gautus naktinius fiziologinius signalus, ypatingą dėmesį skiriant individualizuotam modeliavimui.
3. Įvertinti pagrindinių signalų apdorojimo ir modeliavimo sprendimų, įskaitant laikinę segmentaciją ir išankstinį apdorojimą, poveikį priešmigreninių požymių aptikimui ir klasifikavimo rezultatams.
4. Nustatyti fiziologiškai informatyvius signalus ir požymių grupes, susijusias su priešmigreniniais pokyčiais, taikant statistinę požymių analizę ir modeliais pagrįstus reikšmingumo vertinimo metodus, taip pat įvertinti požymių mažinimo poveikį modelių patikimumui ir interpretuojamumui.
5. Patvirtinti siūlomą metodą išplėstame tiriamųjų rinkinyje taikant vieną mokymo ir validacijos protokolą bei įvertinti kitos dienos migrenos prognozavimo

galimybes naudojant naktinius dėvimųjų įrenginių duomenis, esant laiko atžvilgiu atskirtai validacijai.

Tyrimų metodika

Taikyti šie metodai, nagrinėjant darbo objektą:

- Atlikta struktūrinta migrenos tyrimų, paremtų dėvimų įrenginių duomenimis, apžvalga, apimanti migrenos fazes, priešmigreninę dinamiką, išmatuojamus fiziologinius signalus, realiomis sąlygomis surinktų duomenų kokybės problemas bei dažniausiai taikomus išankstinio apdorojimo, požymių išgavimo ir mašininio mokymosi metodus.
- Fiziologiniai duomenys buvo renkami realiomis sąlygomis naudojant „Empatica Embrace Plus“ įrenginį. Analizė buvo sutelkta į naktinius įrašus siekiant sumažinti dienos variaciją. Kiekvienam dalyviui ir signalui buvo įvertinta, ar yra visi duomenys ir trūkstamų duomenų struktūra.
- Fiziologinių signalų išankstinis apdorojimas užtikrino fiziologinį pagrįstumą, pašalinant netinkamus mėginius ir apribojant neįtikėtinas pulso dažnio reikšmes. Taip pat buvo įvertinti keli filtravimo metodai, įskaitant medianinį, Savitzky–Golay ir Butterworth žemadažnį filtravimą.
- Naktiniai įrašai buvo segmentuojami į fiksuoto ilgio intervalus nuo 5 iki 120 minučių. Buvo išgauti statistiniai požymiai, o skirtumai tarp priešmigreninių ir kontrolinių būsenų analizuoti naudojant aprašomąją statistiką ir vienfaktorinę dispersinę analizę (ANOVA).
- Sukurti dvejetainės klasifikacijos modeliai, skirti priešmigreninėms ir kontrolinėms būsenoms atskirti. Taikytos dvi validavimo strategijos: segmentų lygmens skaidymas metodiniam palyginimui ir naktų lygmens skaidymas griežtesniam vertinimui. Hiperparametrų optimizavimas atliktas naudojant „RandomizedSearchCV“ su „GroupKFold“ kryžmine validacija pagal naktų identifikatorius. Modelių veikimas vertintas naudojant F1 rodiklį, atkūrimą (*recall*), tikslumą (*precision*), ROC-AUC ir PR-AUC, ypatingą dėmesį skiriant teigiamai klasei.

Darbo mokslinis naujumas

- Struktūrinta kitos dienos migrenos prognozavimo sistema, paremta iš dėvimųjų įrenginių gautais naktiniais fiziologiniais duomenimis ir apimanti nuoseklią nakties–kitos dienos žymėjimo schemą, pomigreninių naktų eliminavimą ir aiškų dviprasmiškų laiko intervalų apdorojimą.
- Struktūrintas dėvimųjų įrenginių duomenų rinkinių charakterizavimo protokolas, formalizuojantis signalų pilnumo, trūkstamų duomenų struktūros, laikinio padengimo ir požymių skaičiavimo galimybių vertinimą.
- Sukurta daugiamastė laikinės agregacijos schema migrenai prognozuoti, paremta iš dėvimųjų įrenginių gautais fiziologiniais duomenimis, naudojant signalų intervalus nuo 5 iki 120 minučių, parodant, kad trumpesni intervalai geriau išsaugo su priešmigreniniais pokyčiais susijusią informaciją, o ilgesni intervalai mažina jautrumą mažumos klasei.

- Gauti tiriamųjų grupės lygmens rezultatai, rodantys, kad iš dėvimųjų įrenginių gautuose naktiniuose fiziologiniuose duomenyse žemesnės eilės statistiniai požymiai yra stabilesni nei aukštesnės eilės požymiai, o prediktorių sudėtis priklauso nuo laikinės skalės.
- Pateikti įrodymai iš išplėsto tiriamųjų rinkinio, kad įprastinis fiziologinių signalų filtravimas nepagerina migrenos klasifikavimo rezultatų, todėl išankstinio apdorojimo nauda turi būti empiriškai pagrįsta, o ne laikoma savaime suprantama.

Darbo rezultatų praktinė reikšmė

- Siūloma duomenų audito procedūra, kuri gali būti naudojama vertinant naktinių fiziologinių duomenų, gautų iš dėvimųjų įrenginių, tinkamumą prieš modelių kūrimą ir gerinant tyrimų palyginamumą.
- Disertacijoje sukurta analitinė grandinė, integruojanti miego pagrindu atliekamą segmentaciją, skaidrų žymėjimą, požymių išgavimą ir validaciją, kuri gali būti pritaikyta migrenos ir kitų epizodinių būklių tyrimuose, naudojant dėvimųjų įrenginių duomenis.
- Disertacijoje identifikuotas kompaktiškas ir interpretuojamas požymių rinkinys, leidžiantis kurti paprastesnius ir skaidresnius kitos dienos migrenos prognozavimo modelius neprarandant jų efektyvumo.
- Pateiktos praktinės rekomendacijos dėl laikinės segmentacijos, pagrindžiančios trumpesnių naktinių analizės langų taikymą priešmigreninėms būsenoms aptikti ir galinčios būti naudojamos būsimose dėvimųjų įrenginių stebėjimo sistemose.
- Sukurta dviejų etapų modeliavimo strategija, jungianti grupės lygmens mokymąsi su vėlesniu individualiu pritaikymu personalizuotam kitos dienos migrenos prognozavimui, naudojant dėvimųjų įrenginių duomenis.

Ginamieji teiginiai

1. Priešmigreninių būsenų aptikimas naudojant naktinius dėvimų įrenginių duomenis stipriai priklauso nuo analizės laiko mastelio: padidinus lango ilgį nuo 5–10 iki 30 minučių, teigiamos klasės atkūrimas sumažėja apie 10–30 %, o F1 rodiklis – apie 5–15 %, o dar labiau pailginus iki 60–120 minučių, šis sumažėjimas didėja ir didėja rezultatų variabilumas tarp modelių.
2. Išankstinis apdorojimas nė vienoje vertintoje situacijoje nepagerino teigiamos klasės atkūrimo ar F1 rodiklio esant griežtai validacijai, todėl jo nauda negali būti laikoma savaime suprantama ir turi būti kritiškai vertinama.
3. Naktiniai fiziologiniai duomenys pasižymi atkuriama kohortos lygmens informatyvių požymių struktūra, kurioje žemesnės eilės statistiniai deskriptoriai yra stabilesni nei aukštesnės eilės požymiai: 5–10 minučių languose EDA deskriptoriai pasireiškia maždaug 1,5–3 kartus dažniau nei dauguma kitų požymių, o ilgesni langai siejami su labiau heterogeniška požymių struktūra.
4. Ansambliniai klasifikatoriai užtikrina patikimiausią ir subalansuotą veikimą migrenos modeliavimo procese naudojant naktinius dėvimų įrenginių duomenis, griežtos validacijos sąlygomis apie 20–60 % pranokdami SVM ir KNN pagal

teigiamos klasės aptikimo kokybę ir pasižymėdami mažesniu rezultatų variabilumu. SVM išlieka riboto efektyvumo, o KNN neaptinka migrenos klasės.

Darbo rezultatų aprobavimas

Tyrimo rezultatai paskelbti keturiose publikacijose: dviejuose straipsniuose žurnaluose, įtrauktuose į „Clarivate Analytics Web of Science Core Collection“ (Kapustynska ir kt., 2024, 2025). Autorė taip pat pristatė rezultatus dviejose tarptautinėse mokslinėse konferencijose:

- Kapustynska, V., & Paulikas, Š. (2024). Elektroderminio aktyvumo signalo svyravimų analizė prieš migrenos priepuolio pradžią. Tarptautinė konferencija „Science in Motion“, Vinica–Viena.
- Kapustynska, V. ir kt. (2024). Priešmigreninių nakties modelių aptikimas naudojant dėvimuosius biosensorius ir mašininį mokymąsi. 15-oji konferencija DAMSS, Druskininkai, Lietuva.

Disertacijos struktūra

Darbas susideda iš įžangos, trijų teminių skyrių, bendrųjų išvadų ir santraukos lietuvių kalba, literatūros sąrašo, autoriaus publikacijų sąrašo ir trijų priedų. Be priedų, disertacija apima 111 puslapių, 35 iliustracijas, 4 formules ir 17 lentelių.

1. Literatūros apžvalga apie migrenos dinamiką, fiziologinius signalus ir prognozavimą

Šiuolaikiniai migrenos tyrimai vis dažniau apibūdina šį sutrikimą kaip dinaminį neurologinį procesą, kuriame klinikinis priepuolis yra tik viena platesnio patologinio ciklo dalis. Migrenos epizodas paprastai skirstomas į kelias fazes: prodrominę fazę, aurą, galvos skausmo (iktalinę) fazę ir postdrominę fazę. Prognozavimo požiūriu svarbiausia yra prodrominė fazė, kuri dažniausiai pasireiškia likus 24–48 valandoms iki skausmo pradžios ir gali būti lydima nuotaikos, aktyvumo bei autonominės nervų sistemos funkcijų pokyčių. Tačiau prodrominių požymių pasireiškimas ir jų kombinacijos tarp pacientų labai skiriasi, todėl universalūs prognozavimo sprendimai išlieka riboti.

Literatūra rodo, kad migrena yra susijusi su autonominės nervų sistemos disbalansu. Todėl fiziologiniai signalai, reguliuojami šios sistemos, laikomi potencialiais informacijos šaltiniais apie priešmigreninę būseną. Dažniausiai nagrinėjami signalai yra elektroderminė odos veikla, širdies susitraukimų dažnis ir širdies ritmo variabilumas, taip pat odos temperatūra bei fizinio aktyvumo rodikliai. Šie parametrai gali atspindėti simpatinės ir parasimpatinės reguliacijos pokyčius, kurie kartais pasireiškia dar iki priepuolio pradžios.

Nešiojamųjų sensorių technologijų plėtra sudarė galimybes nuolat registruoti tokius signalus realiomis gyvenimo sąlygomis. Daugiasensoriniai įrenginiai leidžia vienu metu stebėti autonominius ir elgsenos rodiklius, įskaitant miegą ir cirkadinius ritmus. Vis dėlto realaus gyvenimo duomenys pasižymi dideliu triukšmu ir kontekstiniu kintamumu: juos veikia judesio artefaktai, prietaiso nešiojimo pertrūkiai, nestabilus kontaktas su oda ir trūkstamos reikšmės, kurios dažnai nėra atsitiktinės. Todėl pastaraisiais metais vis labiau pabrėžiamas duomenų centrinis (*data-centric*) požiūris, kai prieš modeliavimą sistemingai vertinama duomenų kokybė, trūkstamumo struktūra ir biomarkerių patikimumas.

Svarbus fiziologinių signalų analizės etapas yra filtravimas ir išankstinis apdorojimas. Dažniausiai minimi metodai yra medianinis filtras, Savitzky–Golay glodinimas ir Butterworth žemų dažnių filtrai. Jie taikomi triukšmui ir artefaktams slopinti, tačiau filtravimo poveikis požymiams ir prognozavimo rezultatams migrenos kontekste nėra vienareikšmis ir dažnai priklauso nuo signalo tipo bei analizės lango trukmės.

Fiziologinės laiko eilutės paprastai transformuojamos į požymius, apskaičiuotus fiksuotuose ar slenkančiuose laiko languose. Plačiausiai naudojami paprasti statistiniai rodikliai (vidurkis, mediana, minimumas, maksimumas, standartinis nuokrypis, RMS). Vis daugiau dėmesio skiriama laiko agregavimo mastui: trumpi langai gali geriau išsaugoti trumpalaikius pokyčius, o ilgi langai suteikia stabilesnius įverčius, bet gali „išlyginti“ informatyvius svyravimus.

Migrenai prognozuoti taikomi įvairūs mašininio mokymosi metodai. Dažniausiai naudojami klasikiniai algoritmai (*Support Vector Machines*, *Random Forest*, gradientinio stiprinimo metodai), kurie gerai veikia su nedidelėmis imtimis ir leidžia analizuoti požymių svarbą. Giliojo mokymosi metodai taip pat taikomi, tačiau jų taikymą migrenos srityje dažnai riboja mažos imtys ir interpretavimo sudėtingumas. Daugelyje darbų pabrėžiama, kad individualizuoti modeliai dažnai pranoksta bendruosius, nes migrenos dinamika ir priešmigreniniai signalai yra stipriai individualūs.

Apibendrinant, literatūra patvirtina nešiojamaisiais įrenginiais registruojamų fiziologinių signalų potencialą aptikti priešmigreninius pokyčius, tačiau kartu pabrėžia metodologinių veiksnių svarbą: duomenų kokybę, kontekstinį segmentavimą (ypač miego ir cirkadinių ritmų įtaką), laiko agregavimo mastą, filtravimo pasirinkimą ir tinkamas validacijos strategijas. Šie aspektai pagrindžia poreikį sistemingai tirti, kaip analizės sprendimai keičia aptinkamumą ir prognozavimo rezultatų stabilumą.

2. Duomenų rinkinio apibūdinimas, kokybės kontrolė ir laiko segmentavimas

Antrajame skyriuje buvo atlikta dėvimo fiziologinio duomenų rinkinio struktūrizacija, kokybės auditas ir metodologinis parengimas prieš tolesnį modeliavimą. Analizės tikslas buvo įvertinti signalų išsamumą, stabilumą ir fiziologinį suderinamumą, taip pat pagrįsti naktinio režimo ir laikinės segmentacijos pasirinkimą kaip pagrindinius tolesnės analitinės schemos elementus. Po duomenų eksportavimo ir konsolidavimo buvo atliktas kiekybinis trūkstamų reikšmių įvertinimas visuose prieinamuose biomarkeriuose. Nustatyta, kad daugumai signalų būdinga panaši ir vidutinė trūkstamų duomenų dalis (apie 12 %), o pulso variabilumo (PRV) ir kvėpavimo dažnio (RR) rodikliai pasižymi itin dideliu nepilnumu, viršijančiu 70 %. Apibendrintos trūkstamų reikšmių proporcijos pateiktos S2.1 lentelėje.

S2.1 lentelė. Trūkstamų reikšmių procentinė dalis pagal signalus ir visus dalyvius

Signalas	Trūkstamos reikšmės, %
Akselerometrai	12,37
Aktyvumo skaitikliai	12,37

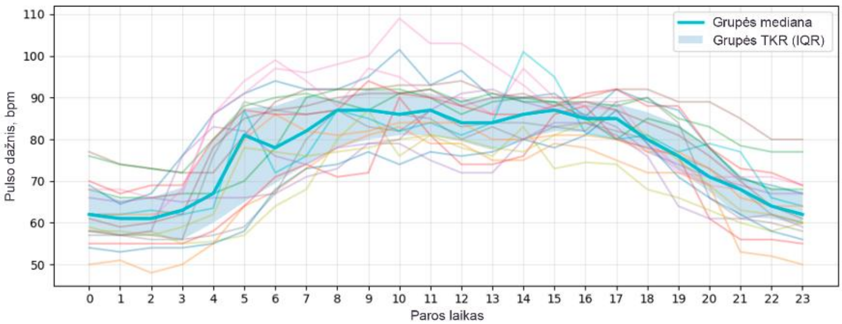
S2.1 lentelės pabaiga

Elektroderminė veikla	12,37
Metabolinis ekvivalentas	12,37
Pulso variabilumas	**75,05**
Pulsas	12,37
Kvėpavimo dažnis	**72,47**
Miego nustatymo stadija	12,37
Signalas	Trūkstamos reikšmės, %
Žingsnių skaičius	12,37
Odos temperatūra	12,37
Nešiojimo nustatymas	10,74

Pastaba: pusjuodžiu šriftu pažymėti signalai, turintys perteklinį trūkstamų duomenų kiekį ir todėl neįtraukti į tolesnę analizę.

Toks didelis trūkstamų duomenų kiekis reikšmingai riboja šių signalų statistinį patikimumą atliekant tarpindividinę analizę, todėl RR ir PRV buvo pašalinti iš tolesnio apdorojimo. Juos pašalinus, likusiems signalams buvo pritaikytas pilnų atvejų filtras, kuris leido išlaikyti palyginamą stebėjimų struktūrą tarp dalyvių. Nepaisant skirtingos pašalintų eilučių dalies tarp dalyvių, galutinis duomenų kiekis buvo pakankamas tolesnei kohortinei ir individualiai analizei.

Toliau buvo įvertintas signalų fiziologinis pagrįstumas ir tarpindividinė variacija. Buvo sudaryti kohortiniai cirkadiniai profiliai, remiantis medianinėmis reikšmėmis ir tarpkvartiliniu intervalu. Pulso dažnio pavyzdys (S2.1 pav.) rodo, kad įrašai išlaiko tikėtiną paros dinamiką: sumažėjimą nakties metu, rytinį pakilimą ir santykinę stabilizaciją dienos laikotarpiu.



S2.1 pav. Kohortos pulso dažnio cirkadinis profilis

Panaši cirkadinė struktūra stebima ir kituose fiziologiniuose bei elgesio signaluose, tai patvirtina registracijos korektiškumą ir rodo, kad artefaktai neturi dominuojančios įtakos duomenims. Kartu nustatytas ryškus tarpindividinis rodiklių lygių ir jų sklaidos heterogeniškumas, kuris pagrindžia individualizuoto modeliavimo poreikį tolesniuose analizės etapuose.

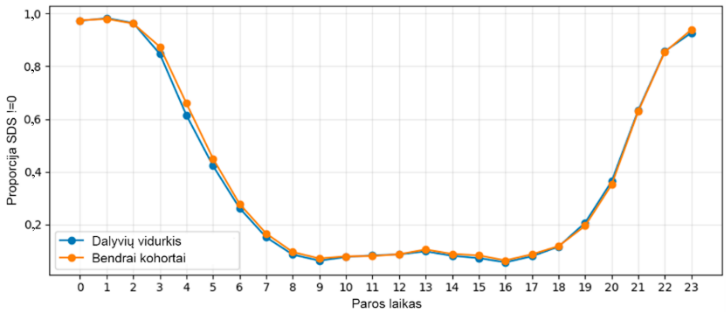
Atsižvelgiant į didesnę dienos duomenų variaciją, kurią lemia elgesio ir kontekstiniai veiksniai, tolesnė analizė buvo apribota naktiniu laikotarpiu. Objektyviam miego intervalų nustatymui buvo naudojamas miego aptikimo stadijos rodiklis (SDS). SDS reikšmių klasifikacija pateikiama S2.2 lentelėje.

S2.2 lentelė. Miego aptikimo stadijos (SDS) verčių klasifikacija

Reikšmė	Aprašymas
0	Būdravimas
101	Miegas — pagrindinis kriterijus, leidžiantis nustatyti tikrus miego laikotarpius
102	Miego metu pasikartojantys pabudimai
300	Laikinai sutrikę miego ciklai

Nulinės SDS reikšmės atitinka būdravimo būseną, o kitos reikšmės apibūdina skirtingas miego arba su miegu susijusias būsenas.

Nenulinių SDS reikšmių pasiskirstymas paros valandų atžvilgiu (S2.2 pav.) rodo jų aiškią koncentraciją nakties metu.



S2.2 pav. Nenulinių miego aptikimo stadijos (SDS) reikšmių cirkadinis profilis

Remiantis tuo, naktinis režimas buvo operaciškai apibrėžtas kaip intervalas nuo 18:00 iki 09:00, esant sąlygai SDS ≠ 0. Siekiant sumažinti galimą postdrominio laikotarpio įtaką ir išlaikyti fiziologiškai stabilesnes sąlygas, iš analizės taip pat buvo pašalintos naktys, einančios po migrenos epizodų. Papildoma analizė parodė, kad duomenų apribojimas tik miego intervalais sumažina signalų vidinę segmentų variaciją, palyginti su dienos duomenimis, ir patvirtina labiau stacionarų naktinio fiziologinio režimo pobūdį. Todėl naktinis miego būsenos pagrindu filtruotas kontekstas buvo pasirinktas kaip pagrindas požymiams formuoti ir modeliams kurti.

Nustačius naktinį režimą, buvo atlikta laikinė duomenų segmentacija. Naktiniai įrašai buvo suskirstyti į langus, kurie tarpusavyje nepersidengė ir kurių trukmė buvo 5, 10, 30, 60, 90 ir 120 minučių. Kiekvienam langui buvo apskaičiuotas fiksuotas statistinių deskriptorių rinkinys, paskui įvertintas jų padengimas, stabilumas ir klasių atskiriamumas. Nustatyta, kad didėjant lango trukmei mažėja prieinamų segmentų skaičius, tai matyti iš klasių pasiskirstymo tarp dalyvių ir analizės langų (S2.3 lentelė).

S2.3 lentelė. Klasės pavyzdžių skaičius kiekvienam dalyviui pagal analizės lango trukmę (žyma 0 / žyma 1)

Subjektas	5 min.	10 min.	30 min.	60 min.	90 min.	120 min.
1	72/269	37/137	13/49	7/27	5/18	4/15
2	1089/248	550/126	189/45	101/24	72/17	58/13
3	1353/538	686/273	241/95	132/51	94/36	74/27
4	2671/293	1350/149	466/52	247/29	170/19	135/16
5	4202/470	2106/236	710/79	362/40	246/28	176/20
6	3746/692	1903/352	669/123	360/66	258/45	201/38
7	1141/420	578/212	199/73	103/38	74/28	55/20
8	2225/776	1124/392	394/138	208/72	146/52	110/40
9	1355/904	684/457	236/161	127/87	88/62	71/44
10	1121/501	568/253	197/89	103/48	78/35	58/28
11	2089/714	1057/363	368/125	193/66	139/47	110/37
12	2982/603	1510/304	527/105	278/54	200/40	158/30
13	628/517	319/261	110/92	58/49	42/35	33/28
14	960/589	485/300	167/106	88/55	61/40	48/28
15	1465/304	741/152	257/52	138/28	99/20	75/14
16	3486/321	1763/163	610/57	322/31	226/22	177/16
17	1171/904	591/458	202/160	107/85	77/61	58/45
18	1848/383	935/193	326/67	173/35	120/24	98/19
19	886/1119	444/566	154/196	83/105	57/77	44/57

Kartu dviem iš 19 dalyvių segmentai nebuvo suformuoti nė vienam nagrinėtam analizės langui dėl nepakankamo validžių miego intervalų, atitinkančių įtraukimo kriterijus, skaičiaus.

Kartu požymių stabilumas didėja didėjant laiko lango trukmei, tačiau atskiriamumas tarp tarp priešmigreninių ir kontrolinių intervalų priklauso tiek nuo signalo tipo, tiek nuo pasirinktos laiko skalės. Praktikoje tai reiškia, kad kai kurie signalai ryškesnius skirtumus parodo trumpuose languose, kiti – vidutinės ar ilgesnės trukmės languose, todėl vieno universalaus optimalaus lango visoms modalybėms nustatyti nepavyko. Tai rodo, kad informatyvūs fiziologiniai modeliai yra nuo konteksto priklausomi.

Svarbu pabrėžti, kad antrojo skyriaus rezultatai yra pirminis įvertinimas prieš taikant mašininio mokymosi metodus. Šiame etape analizuojama duomenų kokybė, jų stabilumas, požymių elgsena ir statistiniai skirtumai tarp klasių. Toks analizės etapas leidžia nustatyti, ar duomenyse yra potencialiai informatyvus signalas, tačiau pats savaime dar nepadeda atsakyti į klausimą, kaip efektyviai šį signalą galės panaudoti klasifikavimo modeliai.

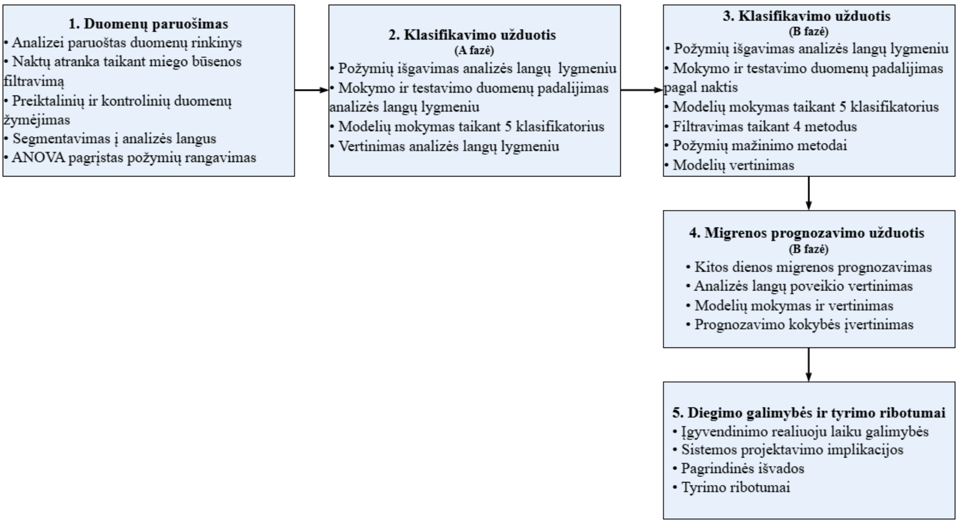
Apibendrinant galima teigti, kad antrojo skyriaus rezultatai rodo, jog po kokybės kontrolės procedūrų gauti nešiojamų įrenginių duomenys pasižymi pakankamu pilnumu ir fiziologiniu nuoseklumu tolesnei analizei; naktinis režimas, apribotas miego būsenos filtravimo procedūra, suteikia stabilesnę analitinę aplinką nei dienos duomenys; o laiko lango trukmė yra svarbus metodologinis veiksnys, darantis įtaką statistinėms požymių savybėms ir jų potencialiam atskiriamumui.

Toliau pasirinktas laiko langų intervalas (5–120 min.) naudojamas kaip valdomas parametras jau mašininio mokymosi uždaviniuose: pirmiausia dvejetainės klasifikacijos eksperimente (vertinant penkis klasifikatorius), o vėliau – griežtesnėje kitos dienos migrenos prognozavimo užduotyje.

3. Eksperimentinė metodika ir rezultatų vertinimas

Šiame skyriuje pateikiami eksperimentinio siūlomos metodinės sistemos vertinimo rezultatai, nagrinėjant, ar iš naktinių nešiojamųjų įrenginių fiziologinių duomenų galima aptikti stabilius priešmigreninius dėsningumus ir įvertinti kitos dienos migrenos riziką. Pagrindinis dėmesys skiriamas tam, kaip klasifikavimo rezultatus veikia analizės lango trukmė, signalų išankstinis apdorojimas ir požymių reprezentacija.

Bendras eksperimentinis darbo srautas pateiktas S3.1 pav.

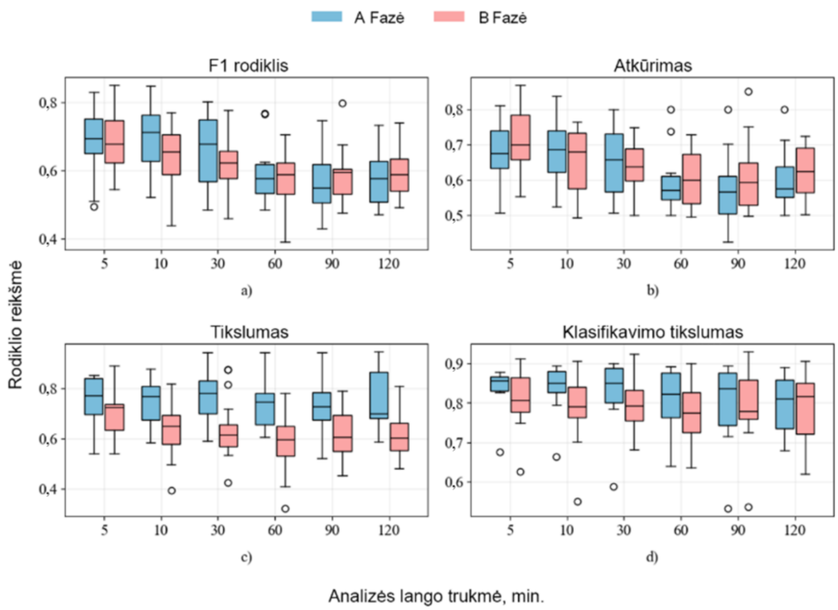


S3.1 pav. Eksperimentinės analizės darbo eiga

Šis analizės etapas remiasi 2 skyriuje aprašytu duomenų paruošimo etapu ir naudoja tik tuos naktinius įrašus, kurie buvo atrinkti taikant miego būsenos filtravimo procedūrą. Kiekviena naktis laikoma pagrindiniu stebėjimo vienetu, tačiau modeliavimas atliekamas skirtinguose laiko masteliuose, segmentuojant naktį į 5, 10, 30, 60, 90 ir 120 min. analizės langus. Toks sprendimas leido nuosekliai įvertinti, kaip laiko agregacija keičia klasių atskiriamumą ir prognozavimo kokybę.

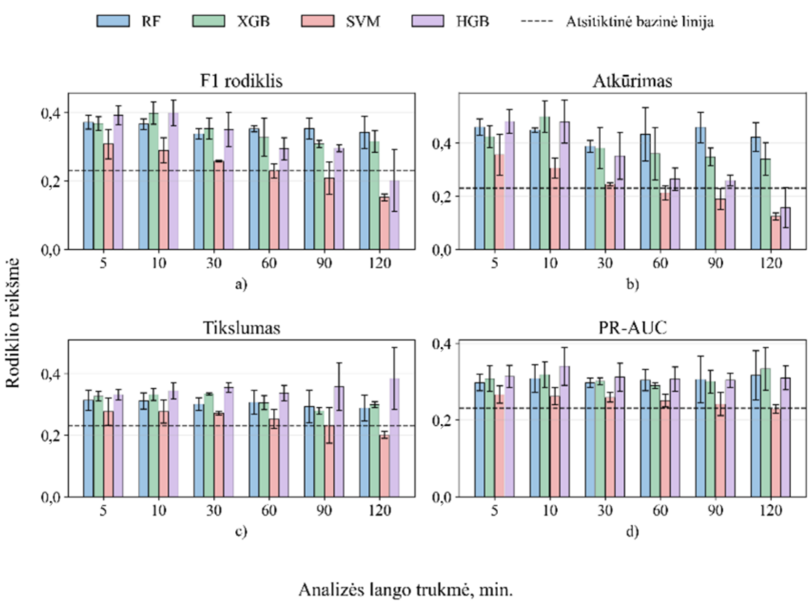
Eksploracinėje klasifikavimo schemoje, kai mokymo ir testavimo duomenys buvo skaidomi analizės langų lygmeniu, nustatyta aiški tendencija, kad trumpi analizės langai užtikrina geresnius rezultatus. Kaip parodyta S3.2 pav., didžiausios F1 ir jautrumo reikšmės dažniausiai buvo gaunamos naudojant 5 ir 10 min. langus, o didinant lango trukmę iki 60–120 min. klasifikavimo kokybė mažėjo.

Šis rezultatas rodo, kad priešmigreniniai fiziologiniai pokyčiai greičiausiai yra trumpalaikiai ir lokalūs laike, todėl jie geriau išsaugomi smulkesnėje laiko skiriamojoje geboje. Vis dėlto ši analizė laikoma eksploracine, nes toje pačioje naktyje esantys segmentai galėjo patekti ir į mokymo, ir į testavimo imtis.



S3.2 pav. Klasifikavimo metrikų pasiskirstymas tarp dalyvių esant skirtingoms analizės lango trukmėms

Todėl tolesniame etape buvo taikyta griežtesnė validacijos schema, kai visi vienai nakčiai priklausantys langai buvo priskiriami tik mokymo arba tik testavimo rinkiniui. Šios kohortinės klasifikacijos rezultatai pateikti S3.3 pav.

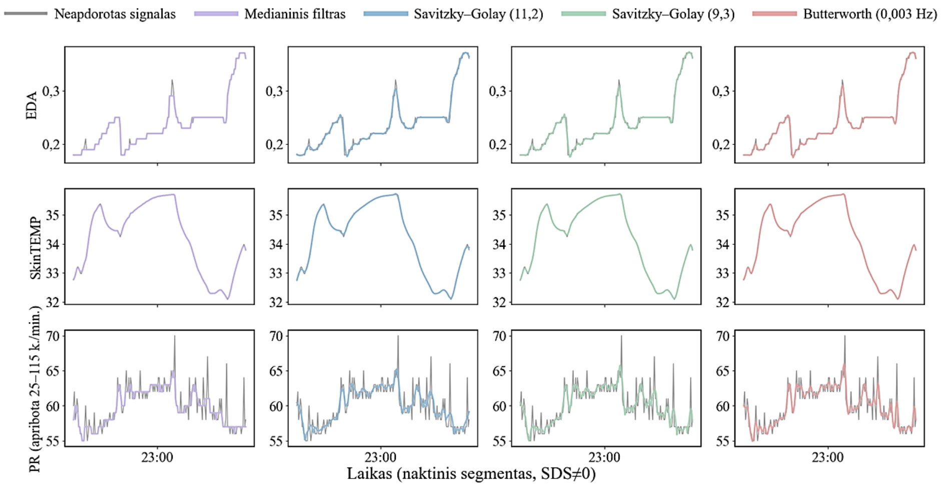


S3.3 pav. Klasifikavimo rezultatai kohortos lygmeniu naudojant nakties lygmens validaciją

Nors absoliutūs rodikliai sumažėjo, pagrindinė tendencija išliko tokia pati: trumpi 5–10 min. analizės langai ir toliau buvo palankiausi teigiamos klasės aptikimui. Didėjant analizės lango trukmei, ypač ryškiai mažėjo jautrumas ir teigiamos klasės F1 rodiklis, o tai rodo, kad ilgesni langai slopina trumpalaikius priešmigreninius nukrypimus. Makrolygmens rodikliai mažėjo ne taip stipriai, tai reiškia, kad laiko agregacija labiausiai kenkia būtent mažumos klasės aptikimui.

Modelių palyginimas parodė, kad ansambliniai metodai buvo tinkamiausi šiai užduočiai. *XGBoost* ir *Histogram Gradient Boosting* daugumoje konfigūracijų užtikrino geriausią pusiausvyrą tarp bendros diskriminacijos ir priešmigreninių naktų aptikimo. *Random Forest* taip pat rodė gana stabilų elgesį, nors dažniausiai nusileido *boosting* tipo modeliams. SVM veikimas buvo silpnesnis, ypač ilgesniuose analizės languose, o KNN neparodė pakankamo gebėjimo patikimai aptikti teigiamą klasę griežtesnėje validacijos schemoje.

Signalų filtravimo poveikis buvo vertinamas kaip papildoma jautrumo analizė. P3.4 pav. pateiktas pavyzdys rodo, kaip skirtingi filtravimo metodai keičia naktinių EDA, odos temperatūros ir pulso signalų formą.

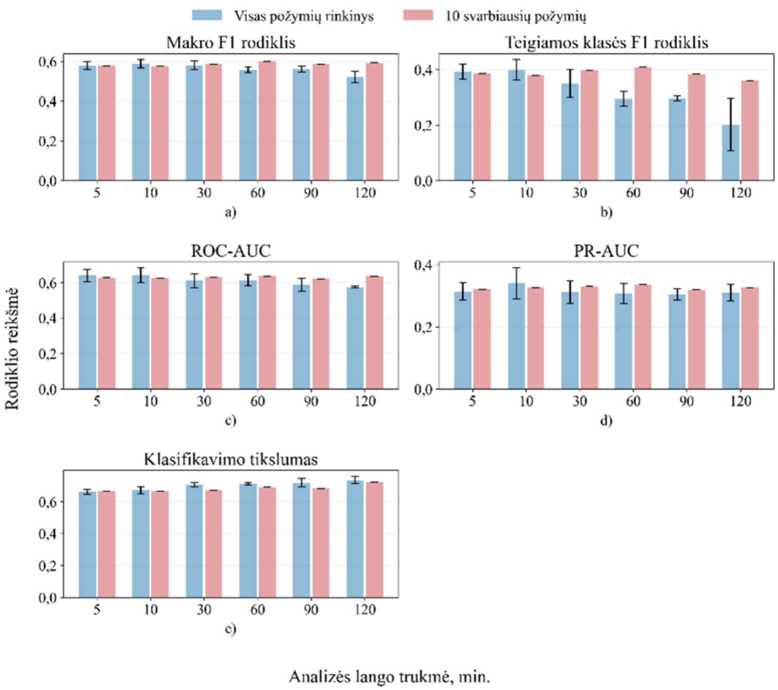


S3.4 pav. Žalių naktinių signalų palyginimas su medianiniu, Savitzky–Golay ir Butterworth filtavimu

Eksploracinėje schemoje kai kurie filtravimo metodai, ypač Butterworth ir Savitzky–Golay, tam tikrais atvejais pagerino bendrus klasifikavimo rodiklius. Tačiau griežtoje kohortinėje validacijoje šis efektas nebuvo stabilus. Nustatyta, kad filtravimas dažnai nemažai nepagerina bendros kokybės, o kai kuriais atvejais net sumažina teigiamos klasės jautrumą. Tai leidžia daryti išvadą, kad šioje užduotyje svarbu išsaugoti pirminę signalo struktūrą, nes glodinimas gali nuslopinti subtilius trumpalaikius fiziologinius pokyčius, svarbius priešmigreninių būsenų aptikimui.

Toliau buvo analizuojama, kokie požymiai yra informatyviausi. Nustatyta, kad didžiausią reikšmę turi žemos eilės statistiniai deskriptoriai, apibūdinantys signalo lygį, minimumą, medianą ar amplitudę. Ypač ryški buvo elektroderminės veiklos (EDA) svarba: trumpuose 5–10 min. languose būtent EDA požymiai dažniausiai dominavo tarp svarbiausių modelio kintamųjų. Didėjant analizės lango trukmei, šalia EDA didėjo ir judėjimo bei agreguotų signalų požymių svarba, tačiau EDA išliko nuosekliausiai pasikartojantis informatyvus signalas.

Papildomai buvo atliktas požymių mažinimo eksperimentas, kurio rezultatai pateikti S3.5 pav.



S3.5 pav. Klasifikavimo rezultatų palyginimas naudojant visą ir sumažintą požymių rinkinį skirtingoms analizės lango trukmėms

Čia lyginamas visas požymių rinkinys ir sumažintas, iš pasikartojančių svarbiausių požymių sudarytas pogrupis. Rezultatai parodė, kad sumažintas požymių rinkinys nedaro neigiamos įtakos, o kai kuriose konfigūracijose net pagerina modelių veikimą. Ryškiausias pagerėjimas buvo stebimas teigiamos klasės F1 rodiklyje, ypač vidutinės ir ilgesnės trukmės languose. Tai rodo, kad pagrindinė prognozinė informacija yra koncentruota nedideliame stabilių požymių rinkinyje, o pertekliniai požymiai gali didinti modelio variaciją ir mažinti jo robustiškumą.

Be klasifikavimo, buvo įvertintas ir nakties lygmens migrenos rizikos nustatymas, agreguojant langų lygmens prognozes į vieną nakties rezultatą. Kohortos lygmeniu gauti vidutiniai, bet nuoseklūs rezultatai, viršijantys bazinį lygį. Geriausios konfigūracijos dažniausiai buvo gaunamos naudojant 10–30 min. langus, tai rodo kompromisą tarp trumpalaikių signalų išsaugojimo ir triukšmo sumažinimo. Individualių dalyvių lygmeniu rezultatai buvo gerokai labiau kintantys. Daliai dalyvių buvo stebimi teigiami signalai, tačiau daugeliu atvejų prognozės išliko nestabilios dėl riboto duomenų kiekio ir ryškaus klasių disbalanso.

Apibendrinant galima teigti, kad eksperimentiniai rezultatai patvirtina dvi pagrindines išvadas. Pirma, priešmigreninių fiziologinių pokyčių aptinkamumas labai priklauso nuo metodologinių sprendimų, ypač nuo analizės lango trukmės ir požymių reprezentacijos. Antra, nepaisant ryškios tarpindividualios variacijos, kohortos lygmeniu galima nustatyti pasikartojančius ir interpretuojamus dėsningumus, ypač susijusius su elektrodermine veikla ir kompaktiškais žemos eilės statistiniais požymiais. Šie rezultatai rodo, kad naktiniai nešiojamųjų įrenginių duomenys turi realų potencialą migrenos rizikos vertinimui, tačiau stabiliausi rezultatai šiuo metu pasiekiami kohortos, o ne individualiame lygmenyje.

Bendrosios išvados

1. Migrenos prognozavimas naudojant nešiojamųjų įrenginių duomenis labai priklauso nuo metodologinio dizaino ir validavimo strategijos. Lyginant laikinės segmentacijos langus nuo 5 iki 120 min. kartu su skirtingais išankstinio apdorojimo ir validavimo nustatymais, nustatyta, kad šie metodologiniai sprendimai reikšmingai veikia modelių veikimo įverčius. Ypač griežtas validavimas nakties lygmeniu sumažina optimistinį šališkumą, pastebimą tiriamuosiuose nustatymuose, o dažniausiai naudojami filtravimo metodai nesuteikia nuoseklių pagerinimų ir gali sumažinti jautrumą teigiamai klasei.
2. Naktiniai dėvimųjų įrenginių fiziologiniai signalai turi aptinkamų, tačiau vidutiniškai išreikštų priešmigreninių požymių. Vertinant kohortos lygmeniu, modeliai pasiekė maždaug 0,62–0,70 ROC-AUC ir maždaug 0,32–0,37 PR-AUC reikšmes, o tai rodo nuoseklias, bet ne itin dideles prognozavimo galimybes. Tačiau prognozavimas nakties lygmeniu individualiu mastu išlieka nestabilus, todėl galima teigti, kad signalas egzistuoja, bet nėra vienodai išreikštas visų dalyvių duomenyse.
3. Laikinė raiška yra esminis veiksnys išsaugant diskriminacinę informaciją. Aukščiausi klasifikavimo rezultatai nuosekliai buvo pasiekiami naudojant trumpus, 5–10 min. analizės langus, kai teigiamos klasės F1 rodiklis siekė maždaug 0,43–0,44. Padidinus lango trukmę iki 90–120 min., rezultatai sumažėjo iki maždaug 0,32–0,38, o tai rodo, kad laikinis agregavimas susilpnina trumpalaikius fiziologinius nukrypimus.
4. Ansambliniai modeliai yra tinkamiausi migrenai prognozuoti pagal dėvimųjų įrenginių duomenis. *HistGradientBoosting* ir *XGBoost* pateikė patikimiausius rezultatus įvairiose konfigūracijose, o *Random Forest* rodė stabilius, bet šiek tiek silpnesnius rezultatus. SVM pademonstravo ribotą efektyvumą.
5. Prognozinė informacija sutelkta kompaktiškame žemos eilės požymių rinkinyje, kuriame dominuoja elektroderminio aktyvumo rodikliai. Agreguota požymių pasikartojimo analizė rodo, kad elektroderminio aktyvumo požymiai pasirodo maždaug 2–3 kartus dažniau nei su judesiu susiję požymiai, o temperatūra ir pulso dažnis prisideda mažiau nuosekliai. Tai rodo, kad autonominės nervų sistemos aktyvumas yra pagrindinis prognozinio signalo šaltinis naktiniame kontekste.
6. Modeliavimas kohortos lygmeniu pateikia stabilius ir atkuriamus rezultatus, o individualaus lygmens prognozavimas tebėra ribojamas duomenų kiekio ir pasižymi dideliu kintamumu. Todėl praktinės sistemos iš pradžių turėtų remtis kohortos pagrindu išvestais dėsningumais, o vėliau palaipsniui prisitaikyti prie individualių naudotojų, kai sukaupiama daugiau pažymėtų duomenų.

Viroslava KAPUSTYNSKA

RESEARCH AND APPLICATION OF MACHINE LEARNING
METHODS FOR MIGRAINE ATTACK PREDICTION

Doctoral Dissertation

Technological Sciences,
Electrical and Electronic Engineering (T 001)

MAŠININIO MOKYMO METODŲ TYRIMAS IR
TAIKYMAS MIGRENOS PRIEPUOLIUI PROGNOZUOTI

Daktaro disertacija

Technologijos mokslai,
Elektros ir Elektronikos Inžinerija (T 001)

Lietuvių kalbos redaktorė Dalia Markevičiūtė
Anglų kalbos redaktorė Jūratė Griškėnaitė

2026 05 08. 10,2 sp. l. Tiražas 20 egz.
Leidinio el. versija *https://doi.org/10.20334/2026-028-M*
Vilniaus Gedimino technikos universitetas
Saulėtekio al. 11, 10223 Vilnius
Spausdino UAB „Ciklonas“,
Žirmūnų g. 68, 09124 Vilnius