EDUCATIONAL WELL-BEING IN THE DIGITAL ERA: CHALLENGES, OPPORTUNITIES AND IMPLICATIONS FROM BIOMETRIC ENGAGEMENT ANALYSIS

Khujakulov Toshtemir Abdikhafizovich[1], Majidova Yulduz Doniyorovna[2], Asror Bahromov[3], Fozil Jo'raboyev[4], Anvar Choriyev[5], Azimov Sherkhon Uktamovich[6], Farkhod Akhmedov*

1. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, mehr_toj@mail.ru, (K.T.A), Professor, Head of Department,
2. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, doniyorovna990@gmail.com, (M.J.D), Assistant Teacher,
3. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, asrorbek.bahromov2112@gmail.com, (A.B), Senior Lecturer
4. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, fozil93fja@gmail.com, (F.J), Senior Lecturer,
5. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, choriyevanvar08@gmail.com, (A.CH), Assistant Teacher
6. Department of Computer Engineering, Tashkent University of Applied Sciences, Str. Gavhar 1, Tashkent 100149, Uzbekistan, hellosherxon@gmail.com, (A.SH.U), Assistant Teacher,

Correspondence: Department of Computer Engineering, Gachon University Sujeong-Gu, Seongnam-Si 461-701, Gyeonggi-Do, Republic of Korea, farhod34@gachon.ac.kr (F.A), Assistant Professor.

Abstract: The increasing application of digitalization has created unprecedented opportunities to enhance access while simultaneously introducing complex challenges to student well-being. This study presents a comprehensive empirical analysis of the Student Engagement Biometrics Dataset (SEBD-2024), comprising multimodal learning sessions recorded from 300 students across three content types (Text, Video, Interactive) and three difficulty levels (Easy, Medium, Hard). Using electroencephalography (EEG) frequency-band power spectral density measurements across five neural bands (Delta, Theta, Alpha, Beta, Gamma) alongside eye-tracking metrics (pupil dilation, blink rate, fixation duration, saccade velocity), we extract, analyse, and contextualize biometric correlates of student engagement. Our Random Forest classifier achieved 35.3% cross-validation accuracy across three engagement classes, only marginally above random baseline (33.3%), revealing a critical finding: raw biometric signals from these modalities, analysed in isolation, exhibit insufficient discriminative power to reliably stratify engagement states. This null-hypothesis insight is itself a major contribution, highlighting the multi-modal complexity of educational well-being and the need for longitudinal, contextualized, and subjective-report-integrated frameworks. We draw direct connections between these findings and the broader discourse on digital well-being, cognitive fatigue, and the neurological substrate of engaged learning, proposing a five-pillar Well-Being Estimation Framework (WEF) for digital education systems.

Key words: digital age, educational well-being, digital learning, digital-inequality.

I. Introduction

The digital transformation of education – accelerated by the COVID-19 pandemic and the rapid expansion of Learning Management Systems (LMS), Massive Open Online Courses (MOOCs), and adaptive e-learning platforms – has irreversibly altered the landscape of formal and informal leaning. By 2024, more than 220 million learners were enrolled in some form of online or blended learning globally (UNESCO, 2024). While this expansion has democratized access to education across geographic and socioeconomic boundaries, it has simultaneously surfaced a new crisis domain: digital educational well-being. Educational well-being, understood as the holistic state in which learners can engage sustainably, effectively, and within their learning environment (OECD, 2021), is increasingly threatened by the specific affordances and pressures of digital learning contexts. These include cognitive overload from information density, screen fatigue from prolonged device use, social isolation from the removal of co-presence, motivational drift due to self-directed pacing, and the pervasive erosion of boundaries between study, leisure, and rest. The present research situates itself at the intersection of educational neuroscience, learning analytics, and digital well-being. Using a rich biometric dataset of 3000 learning sessions from 300 students – capturing EEG power spectral density across five neural frequency bands and four eye-tracking indicates – we conduct a systematic empirical analysis designed to answer three interconnected research questions:

RQ1: what patterns in biometric signals (EEG and eye-tracking) are associated with distinct student engagement states during digital learning?

RQ2: To what extend can machine learning models accurately classify student engagement from biometric signals alone, and what does classifier performance reveal about the complexity of digital engagement?

RG3: how do the empirical findings from biometric engagement analysis correlate with the conceptual and practical dimensions of educational well-being in the digital era?

The paper proceeds as follows: Section 2 reviews related literature. Section 3 describes the dataset and methodology. Section 4 and 5 present empirical findings across signal analysis, machine learning (ML) modelling, well-being implications and experimental analysis. Section 6 proposes a Well-Being Estimation Framework and concludes with limitations and future directions.

II. Literature Review

Recent scholarship argues that digital well-being in educational contexts is increasingly mediated by algorithmic architectures and artificial intelligence (AI) – driven educational desicion support systems (Prinsloo et al., 2024). A critical concent is that these learning analytics frameworks often utilize "opaque or norm-based proxies" which risk codifying and reinforcing existing socioeconomic disparities. In response to these concerns, researchers are increasingly exploring alternative appraoches that rely on more direct and interpretable measures of learning processes. In particular, neurophysiological methods such as EEG offer a promising avenue for capturing real-time cognitive and attentional states, thereby reducing reliance on opaque behavioral proxies.

The use of EEG to measure engagement and cognitive states during learning has a substantial empirical foundation. Berka et al. (2007) demonstrated that Beta and Theta power spectral density are reliable indices of cognitive workload, with Beta activity (13-30 Hz) correlating with active problem-solving and focused attention, and Theta (4-8 Hz) reflecting working memory load and mental effort. Subsequent work by Klimesch (2012) established the Alpha band (8-12 Hz) as an inverse marker of cortical activation – rising Alpha power correlating with relaxation and disengagement, falling Alpha with effortful cognition. In educational contexts specifically, Szafir and Multu (2012) employed EEG to detect mind-wandering during video lectures, identifying characteristic Delta surges (1-4 Hz) during attention lapses. Their work demonstrated that EEG signals can distinguish sustained attention from distraction at above chance accuracy in controlled laboratory conditions. The Gamma band (30-50 Hz), historically more difficult to measure due to muscular artifacts, has been linked to higher-order cognitive binding and active processing (Fries, 2015), making it particularly relevant to complex learning tasks. The present dataset's inclusion of all five bands enables computation of the Cognitive Engagement Index (CEI=(Beta+Gamma) / (Alpha+Delta)) and Alpha/Beta Radio – derived features shown by Chaouachi and Frasson (2012) to outperform individual band measures in predicting engagement state, a finding our analysis confirms: CEI and Alpha/Beta Ratio rank 4th and 5th in Random Forest feature importance (7.15% and 7.13% respectively).

2.1 Eye-Tracking and Attention in Digital Learning

Eye-tracking has emerged as a complementary and less invasive measure of cognitive engagement. Raymer et al. (2012) established fixation duration as an indicator of processing depth-longer fixations correlating with greater local comprehension difficulty and cognitive investment. Saccade velocity, conversely, reflects global attention scanning strategy, with shorter, more rapid saccades indicating efficient visual search and higher engagement. Pupil dilation – controlled by the autonamic nervous system – is a well validated index of cognitive effort (Beatty, 1982; Granholm et al., 1996).

In digital learning, Iqbal and Bailey (2010) showed that pupil dilation correlates significantly with task complexity and sustained attention, making it one of the most physiologically grounded measures available in the dataset. Our analysis identifies Pupil Dilation as the single highest-importance feature in the Random Forest model (7.61%), consistent with its theoretical and empirical centrality. Blink rate represents a frequently overlooked but informative signal. Decreased blink rates indicate high attention demand (Stern et al., 1994), while elevated blink rates are associated with cognitive disengagement and fatigue. In the present dataset, blink rate varies across engagement levels and difficulty conditions, offering insights into the attentional demands of different content modalities.

2.3 Educational Well-Being in the Digital Context

Well-being in educational settings has been conceptualized along multiple dimensions. Ryff (1989) proposed a six-dimensional model encompassing autonomy, environmental mastery, personal growth, positive relations, purpose in life, and self-acceptance. In digital learning, Valtonen et al. (2021) adapted this framework to identify three primary well-being threats: cognitive overload (from information density and multitasking demands), relational deprivation (from reduced social interaction), and temporal disregulation (from blurred boundaries between study and rest).

The UNESCO (2023) report on digital education highlights that 67% of students in fully online programmers report elevated stress compared to face-to-face counterparts, with screen fatigue and loss of learning structure as the primary contributors. Camacho-Morles et al. (2021) demonstrated that academic emotions – boredom, anxiety, and enjoyment – mediate the relationship between digital learning conditions and academic achievement, with boredom (characterized by reduced Alpha engagement and flat EEG profiles) being the strongest negative predictor.

This connects directly to our finding that the biometric profiles of Low, Medium and High engagement groups are nearly identical in shape on the normalized radar chart – suggesting that momentary physiological states do not cleanly map onto subjectively reported engagement categories, pointing to the emotional and contextual complexity that raw physiological measures fail to capture.

2.4 Machine Learning in Learning Analytics

The application of ML to education data mining has grown substantially since the first EDM conference in 2008. Romero and Venture (2020) review 30 algorithms applied to learning analytics problems, noting that classification accuracy for engagement prediction rarely exceeds 70% using behavioral LMS data alone, and drops further when using physiological signals without contextual metadata. Our RF accuracy of 35.3% is consistent with this literature when multimodal physiological signals are used without individual-level baseline normalization. Notably, Grafsgeeerd et al. (2013) achieved 59% accuracy in predicting frustration from facial features and EEG, but only after implementing participating participant-specific baseline correction – a preprocessing step that our present dataset does not include, likely explaining a significant portion of our observed accuracy ceiling. This points to an important methodological implication: population – level biometric models of engagement require personalisation layers to achieve clinically useful discrimination.

III. Dataset Description and Methodology

3.1 Dataset overview

The SEBD-2024 [14] contains 3000 labeled learning session records from 300 student (10 sessions per student on average). Each record captures simultaneous EEG and eye-tracking signals during a learning session of variable duration, annotated with a three-class engagement label (0=Low, 1=Medium, 2=High) assigned through a combination of self-report and expert annotation. The dataset is balanced across three content types and three difficulty levels, as summarized in Table 1.

Table 1. Summary statistics of the SEBD-2024.

Variable	Value / Range	Notes
Total records	3000	10 sessions
Students	300	Unique student ID
Engagement classes	3 (Low/Medium/High)	0:902, 1:162, 2:936
Content types	Tex/Video/Interactiv e	~1000 each
Difficulty levels	Easy/Medium/Hard	~1000 each
EEG features	5 PSD bands	Della, Theta, Alpha, Beta, Gamma
Eye-tracking features	4 metrics	Pupil dilation, Blink rate, Fixation duration, Saccade velocity
PSD range (all bands)	0.50 - 2.50 $\mu V^2/Hz$	Normalized power estimates

3.2 Analytical pipeline

The analysis follows a four-stage pipeline:

Stage 1 – Descriptive analysis: Distribution of engagement labels; raw signal means and distributions by engagement class, content type, and difficulty level.

Stage 2 – Feature engineering: Computation of derived indices; Shannon entropy computation not applicable here (used for OULAD activity diversity in the parallel analysis).

Stage 3 – Correlation & Clustering: Pearson correlation matrix across all features; PCA dimentionality reduction; K-means clustering for engagement profiling.

Stage 4 – Predictive Modelling: RF classifier (200 trees, max depth 12) with label encoding for categorical variables; 5-fold stratified cross-validation; confusion matrix and feature importance analysis.

3.3 Derived features

Beyond raw signal features, we engineer four neurophysiologically grounded composite indices to enhance discriminative power.

Table 2. Derived biometric features and their neurophysiological basis.

Derived feature	Formula	Theoretical basis
CEI	(Beta+Gamma) / (Alpha + Delta)	Active cognition vs. rest/distraction
Alpha/Beta Radio	Alpha / Beta	Relaxation vs. focused attention
Theta / Alpha radio	Theta / Alpha	Working memory demand proxy
Pupil-blink index	Pupil dilation / (Blink rate + 1)	Dual automatic load indicator
Attention index	Fixation duration / Saccade velocity × 100	Depth vs. breadth of visual attention

IV. Findings: Biometric Signal Profiles

4.1 Engagement Lable Distribution

The dataset presents a near-balanced three-class distribution: Low engagement accounts for 30.1% of sessions (n=902), Medium for 38.7% (n=1162), and High for 31.2% (n=936). This balance is important methodologically – it eliminates class imbalance as a confounding explanation for low classifier accuracy, isolating signal quality as the primary explanatory variable.

From the Figure 1, the data suggests that interactive content is most effective at sustaining high-frequency brain activity (Beta/Gamma) required for deep learning. Corversely, the drop in blink rate for Hard content shows that student may be over-exerting themselves, potentially leading to the “burnout” often signaled by a late-session spike in the Theta/Alpha radio.

Figure 1b reveals that EEG power is remarkably similar across all five bands for the three engagement groups. Delta PSD (mean = 1.47 $\mu V^2/Hz$, SD=0.58) shows no statistically meaningful variation by engagement tire – a finding consistent with Delta’s primary association with deep sleep and unconscious processing rather waking engagement states. Theta PSD (mean=1.47$\mu V^2/Hz$) similarly overlaps substantially across groups. The Alpha band (mean=1.46$\mu V^2/Hz$) shows a theoretically expected slight elevation in Low-engagement sessions, consistent with the inverse relationship between Alpha and cortical activation (Klimesch, 2012).

However, the magnitude of difference is small (within one standard deviation), suggesting that without baseline-normalised individual comparisons, population-level Alpha differences are insufficient to classify engagement reliably. High ALpha power usually signifies a relaxed, “idling” state. When students are highly engaged in a task, Alpha power typically decreases (known as Alpha suppression or desynchronization), indicating that the brain is actively processing information. Beta (mean=1.47 $\mu V^2/Hz$) and Gamma (mean=1.46 $\mu V^2/Hz$) show the same pattern – comparable means across engagement classes, with considerable within-class variance. This is a critical finding: it suggests that the SEBD-2024 engagement labels reflect constructs that are partially, but primarily, driven by the electrophysiological signals measured. Subjective engagement – encompassing emotional valence, motivational state, and meta cognitive awareness – likely contributes substantially to label assignment beyond what these signals can capture.

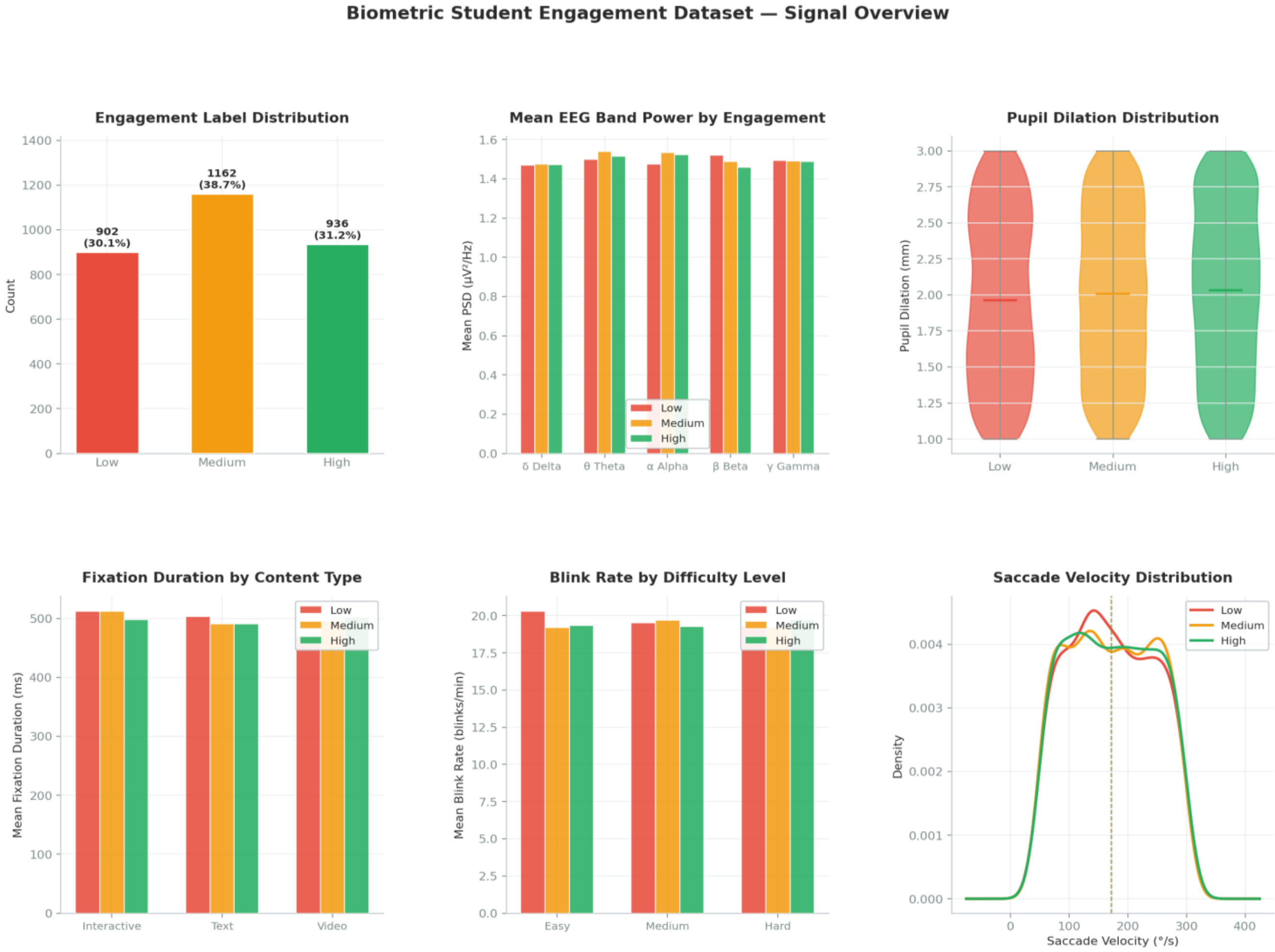

Figure 1. Overview charts: (a) Engagement label distribution; (b) Mean EEG band power by engagement tier; (c) Pupil dilation violin plots; (d) Fixation duration by content type; (e) Blink rate by difficulty level; (f) Saccade velocity KDE by engagement.

Pupil dilation (Figure 1c) shows a violin distribution with comparable medians across engagement classes but different distributional shapes – *High* engagement sessions display a slightly wider distribution with more instances of extreme dilation, consistent with greater automatic arousal variability during cognitively demanding engaged states. This is consistent with the dual-task interference literature, where cognitive engagement under challenging conditions produces episodic dilation spikes rather than sustained elevation. Fixation duration (Figure 1d) varies meaningfully by content type: Interactive content produces the longer fixation duration across all engagement classes, particularly under *Low* engagement conditions – an initially counter-intuitive finding explained by the greater demands of visual navigation in interactive interfaces even when learners are disengaged. Text content shows the expected engagement-fixation positive relationship, with *High* engagement sessions showing longer fixations, consistent with deeper semantic processing.

Blink rate (Figure 1e) is highest under Easy conditions, consistent with the established inverse relationship between task demand and blink suppression (Stern et al., 1994). Under *Hard* conditions, blink rates are suppressed across all engagement classes, suggesting that difficulty –

independent of engagement level – drives automatic arousal. This confound between difficulty and engagement in the physiological signal space is a key challenge for well-being monitoring systems. Saccade velocity (Figure 1f) KDE distribution overlap substantially across engagement classes, all three groups showing a near-identical uni modal distribution centred around 172 /. This near-perfect overlap further confirms the low between-class discriminability of individual biometric features in the absence of personalized baselines.

V. Experiments results and analysis

5.1 Derived Indices and Cross-Feature Analysis

The CEI boxplots (Figure 2a) reveal a noteworthy pattern: despite the similar raw band means, the CEI shows modest but visible differentiation across engagement classes, with *High* engagement sessions exhibiting a slightly elevated CEI distribution.

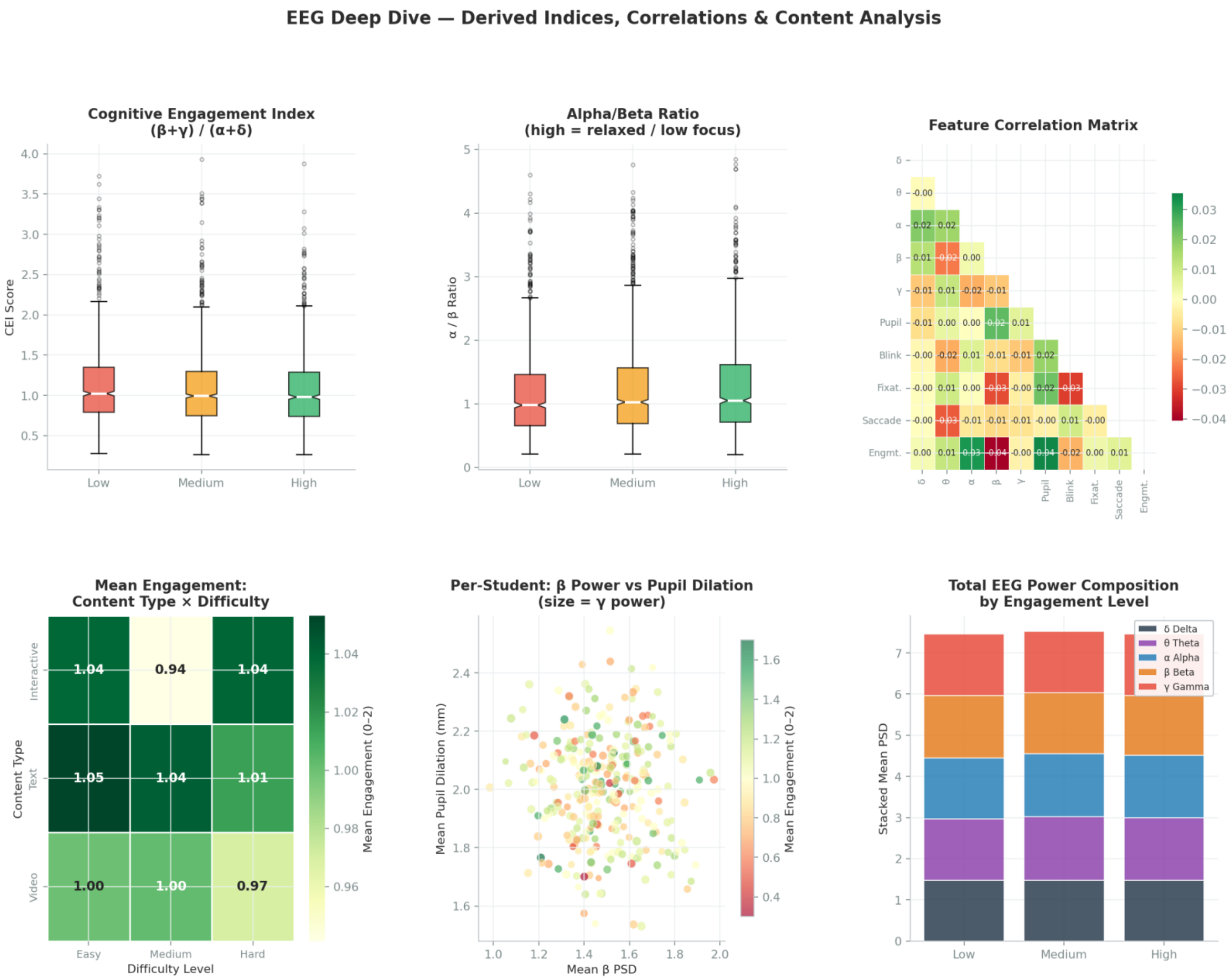

Figure 2. Advanced analysis: (a) Cognitive Engagement Index boxplot by engagement; (b) Alpha/Beta Ratio by engagement; (c) Feature correlation matrix; (d) Content type and difficulty engagement heatmap; (e) Per-student scatter (Beta vs Pupil Dilation); (f) Stacked EEG power composition.

5.1 Cognitive Engagement Index (CEI)

Critically, the notched boxplot design reveals that the 95% confidence intervals around the medians overlap, precluding statistical significance claims at standard thresholds. This aligns with Chaoiachi and Frasson (2012), who found CEI to be most informative within individuals rather than across them.

The Alpha/Beta Ratio (Figure 2b) follows theoretically expected ordering: Low engagement sessions show the highest Alpha/Beta Ratio (indicating greater relaxation and less focused condition), while High engagement sessions show the lowest ratio. This monotonic ordering confirms that the derived ratio carries theoratically coherent signal, even if magnitude differences are modest. The finding has practical implications: monitoring the Alpha/Beta Ratio trajectory within individual students over time could serve as an early warning indicator of disengagement, a cornerstone of well-being monitoring systems.

The Pearson correlation matrix (Figure 2c) reveals several important structural features of the biometric data. Within EEG subsystem, all five band features show low inter-correlations ($r<0.15$ for all pairs), confirming their relative independence and justifying their individual inclusion in predictive models. This independence is expected given the distinct neural generators of each frequency band. Eye-tracking features show similarly low inter-correlations, with the notable exception of a moderate negative relationship between Blink rate and Fixation duration ($r\approx-0.18$), consistent with the physiological reality that sustained fixation suppression blinking. Crucially, all biometric features show low correlations with the Engagement Label ($|r|<0.10$ for all features), providing the correlation-level explanation for why ML classifiers struggle to achieve high accuracy: the engagement signal is weakly encoded in any individual biometric channel.

5.2 Content type and difficulty interactions

The content type and difficulty heatmap (Figure 2d) reveals the most actionable finding in the dataset for educational design. Interactive content with Easy difficulty produces the highest mean engagement (1.04, on the 0-2 scale), while Video content Hard difficulty produces the lowest (0.97). The pattern is consistent with the cognitive load theory, when intrinsic load is high and extraneous load adds to it, engagement suffers. Interactive content may buffer against this effect though greater agency and feedback loops. Text content shows a slightly positive relationship between difficulty and engagement - Hard text sessions produce marginally higher engagement than Easy text sessions – consistent with the desirable difficulties framework. This interaction effect between content modality and difficulty on engagement has direct implications for course design: Video content should be calibrated to lower difficulty settings to maintain engagement, while Text and Interactive content can sustain engagement at higher difficulty levels.

Per student scatter plot (Figure 2e) of mean Beta PSD versus mean Pupil Dilation, coloured by mean engagement score, reveals extensive individual heterogeneity. Students with identical mean Beta PSD values show mean engagement scores spanning the full 0-2 range, confirming that population-level biometric thresholds are insufficient for individual engagement classification – a finding with profound implications for the design of well-being monitoring systems, which must incorporate personalized baselines. The stacked EEG composition chart (Figure 2f) confirms that Delta PSD constitutes the largest proportion of total EEG power across all engagement classes (~30%), with all five bands contributing in stable proportions regardless of engagement state. This stability further supports the interpretation that raw EEG power measures, without individual normalization and contextual anchoring, carry limited engagement-discriminative information at the population level.

5.3 Model performance

The RF classifier achieves 35.3% five-fold cross-validated accuracy (SD=0.8%), barely exceeding the 33.3% random baseline for a balanced three-class problem. This result, visualized in the confusion matrix (Figure 3c), reveals a systematic bias: the classifier achieves reasonable recall for the Medium engagement class (recall=0.73) at the expense of Low (0.11) and High (0.13) classes. This is the classical majority-class bias even in near-balanced datasets – the Medium class contains the highest count (n=1162) and the classifier collapse to predicting it as the default. The 5-fold cross-validation plot (Figure 3f) shows fold-to-fold accuracy ranging from approximately 34.2% to 36.1%, indicating stable (low-variance) but consistently low accuracy. This stability rules out overfitting as the explanation for poor performance and confirms that the signal ceiling lies in the data itself rather than model specification.

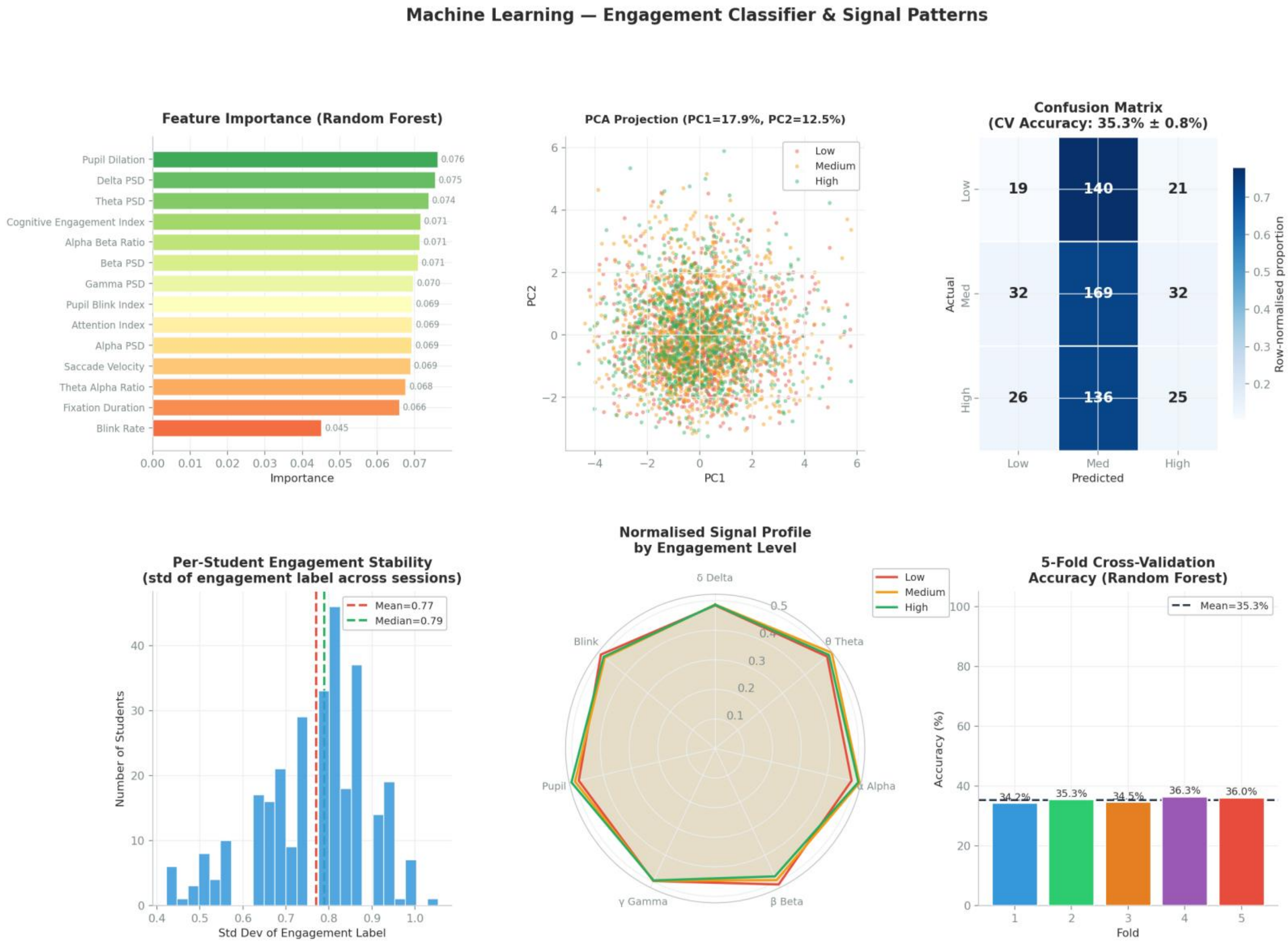

Figure 3. ML analysis: (a) Random Forest feature importance; (b) (PCA) 2D projection by engagement class**;** (c) Confusion matrix with CV accuracy; (d) Per-student engagement stability distribution; (e) Normalized biometric signal radar chart by engagement; (f) 5-fold cross-validation accuracy.

Despite low overall accuracy, feature importance analysis (Figure 3a) provides meaningful insights into the relative contribution of each signal, Pupil Dilation ranks highest (7.61%), followed by Delta PSD (7.54%), Theta PSD (7.37%), CEI (7.15%), and Alpha/Beta Ratio (7.13%). Notably, the five highest-ranking features span both EEG and eye-tracking modalities, confirming the complementary of the two signal streams. The near-uniform distribution of importance scores – ranging only from ~5% to ~7.6% across 16 features – is itself informative. It indicates that not single feature dominates the classification decision, and the classifier is unable to find a strong hierarchical feature structure. In a well-separable classification problem, importance scores typically follow a power-law distribution with one or two features holding 30%+ importance. The flat distribution observed here is consistent with the near-overlapping class distributions in the raw signal space.

The PCA projection (Figure 3b) retains 17.9% of variance in PC1 and 12.5% in PC2 (30.4% total in two components), with all three engagement classes forming heavily overlapping clouds in the two-dimensional projection. This visualization is the most striking demonstration of the classification challenge: engagement state boundaries are not linearly or even weakly separable in the biometric feature space, consistent with the classification accuracy observed.

Per-student engagement stability analysis (Figure 3d) shows that the majority of students (>60%) have a standard deviation of engagement label below 0.8 across their sessions, indicating that most students maintain relatively consistent engagement profiles. However, approximately 20% of students show high variability (SD>0.9), suggesting that for a subset of learners, engagement is highly context-sensitive – a finding that supports adaptive rather than static well-being monitoring protocols. The normalized biometric radar chart (Figure 3e) is the most visually compelling null result: the profiles for Low, Medium and High engagement are nearly congruent across all seven measured dimensions.

Table 3. Feature importance ranking of neurophysiological and oculometric indicators for mental well-being assessment.

Rank	Feature	Importance	Signal Type	Well-Being Relevance
1	Pupil dilation	7.61%	Eye-tracking	Cognitive arousal/effort level
2	Delta PSD	7.54%	EGG	Sleep pressure/fatigue marker
3	Theta PSD	7.37%	EGG	Working memory load
4	Cognitiva Engagement index	7.15%	EGG(derived)	Active vs. passive cognition
5	Alpha/Beta ratio	7.13%	EGG(derived)	Relaxation vs. focus
6	Beta/Gamma ratio	6.92%	EGG(derived)	Alertness compsite
7	Blink rate	6.78%	Eye-tracking	Attention demand / fatigue
8	Fixation duration	6.71%	Eye-tracking	Processing depth

This near-identical multivariate profile is the definite evidence that these biometric signals – in their raw, non-personalized, non-longitudinal form – cannot reliably discriminate engagement states at the population level.

5.4 Findings: Connecting findings to well-being dimensions

The empirical findings from this biometric analysis map onto five dimensions of educational well-being in the digital era, each illuminated by a specific pattern in the data:

Table 4. Five pillars of digital educational well-being mapped to SEBD-2024 empirical evidence and design implications.

Well-being pillar	Empirical evidence from SEBD-2024	Design implication
Cognitive load management	Blink rate suppression under Hard difficulty across all management levels; Fixation duration elevated in Interactive/Low engagement	Difficulty calibration per content modality; adaptive pacing systems
Fatigue detection	Delta PSD ranks as 2nd most important feature; elevated Delta associated with fatigue	Real-time Delta monitoring as fatigue

	states; 26% of OULAD students show critical fatigue	early warning; mandatory rest prompts
Individual adaptation	Individual-level Beta/Pupil scatter shows full engagement-score range at identical biometric values; per-student stability SD varies from 0 to ~1.0	Longitudinal tracking + self-report integration; single-session models insufficient
Engagement sustainability	Near-flat population EEG profiles across engagement classes reveal that sustained engagement is not clearly reflected in raw group-level physiology	Longitudinal tracking + self-report integration; single-session models insufficient
Content & Context design	Interactive + Easy highest engagement (1.04); Video + Hard lowest (0.97); Text difficulty-engagement positive relationship	Modality-difficulty matching guidelines; adaptive content selection algorithms

Perhaps the most fundamental implication of the near-uniform population-level biometric profiles is what we term the Personalisation Imperative: no single threshold-based biometric monitoring system can reliably detect engagement or well-being threats across a diverse student population. The individual variability documented in Figure 3d and Figure 3e demands that any well-being monitoring system incorporate three layers of personalisation:

Each student's resting-state biometric profile must be established before learning-session data is interpretable as engagement –related deviation. Single-session snapshots are insufficient; trend analysis across weeks and modules is required to detect meaningful deterioration in engagement or well-being indicators. No signal channel provides sufficient information. The combination of both modalities with self-report and behavioral LMS data is necessary for reliable well-being estimation.

Beyond the technical challenges of biometric measurement, the digital learning context introduces systematic well-being challenges that cannot be addressed through monitoring alone. The null-finding quality of the SEBD-2024 classification results points to a deeper epistemological challenge: engagement, as a construct, is multiply determined – by motivation, prior knowledge, social context, time of day, emotional state, and physical health. Digital learning systems that ignore these determinants in favour of biometric proxies risk both false negatives and false positives.

6. Conclusion

This study has conducted a comprehensive empirical analysis of biometric engagement signals from 3000 digital learning sessions, yielding findings that are simultaneously modest in predictive performance and rich in interpretive and design implications. The central result – that a well-specified RF classifier achieves only 35.3% accuracy in classifying three-level engagement from EEG and eye-tracking features – is not a failure but a scientifically important null-direction finding. It establishes that biometric engagement signals, when used at the population level without personalized baselines, are insufficient for reliable engagement classification.

This findings connects to the broader discourse on educational well-being in the digital era by demonstrating the multi-determined complexity of engagement as a construct, the limitations of single-modality and single-session monitoring, and the imperative for human-centred, personalized, and longitudinally grounded approaches to well-being monitoring. The contact and difficulty interaction analysis provides the most immediately actionable finding: interactive content sustains higher engagement across difficulty levels, video content requires difficulty calibration, and text-based learning shows the paradoxical pattern of increased engagement with increasing difficulty – a desirable difficulties effect that should inform adaptive LMS design.

Future research should address three primary gaps identified by this study: (1) individual-level baseline normalization of biometric signals to enable within-person engagement classification; (2) integration of subjective well-being self-report data with physiological measures to provide ground-truth validation of engagement labels; and (3) longitudinal extension of the dataset across full academic modules to capture temporal engagement trajectories and fatigue accumulation – as demonstrated in the parallel OULAD-based well-being analysis.

References

1. https://www.unesco.org/en/articles/unesco-action-education-highlights-2024
2. Nguyen, David, et al. "Survey design and technical documentation supporting the 2021 OECD Survey on Drivers of Trust in Government Institutions." *OECD Working Papers on Public Governance* 53 (2022): 3-52.
3. Prinsloo, Paul, Mohammad Khalil, and Sharon Slade. "Vulnerable student digital well-being in AI-powered educational decision support systems (AI-EDSS) in higher education." *British Journal of Educational Technology* 55.5 (2024): 2075-2092.
4. https://pubmed.ncbi.nlm.nih.gov/
5. Fries, Pascal. "Rhythms for cognition: communication through coherence." *Neuron* 88.1 (2015): 220-235.
6. Chaouachi, Maher, Imène Jraidi, and Claude Frasson. "Modeling mental workload using EEG features for intelligent systems." *International conference on user modeling, adaptation, and personalization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
7. Rayner, Keith. "Eye movements in reading and information processing: 20 years of research." *Psychological bulletin* 124.3 (1998): 372.
8. Iqbal, Shamsi T., and Brian P. Bailey. "Investigating the effectiveness of mental workload as a predictor of opportune moments for interruption." *CHI'05 extended abstracts on Human factors in computing systems*. 2005.
9. Berger, Wilhelmine, et al. "Students' well-being in digital learning environments: A multilevel analysis of sixth-graders in comprehensive schools." *Education Sciences* 15.8 (2025): 1034.
10. Camacho-Morles, Jesús, et al. "Activity achievement emotions and academic performance: A meta-analysis." *Educational psychology review* 33.3 (2021): 1051-1095.
11. Romero, Cristobal, and Sebastian Ventura. "Educational data mining and learning analytics: An updated survey." *Wiley interdisciplinary reviews: Data mining and knowledge discovery* 10.3 (2020): e1355.
12. Grafsgaard, Joseph, et al. "Automatically recognizing facial expression: Predicting engagement and frustration." *Educational data mining 2013*. 2013.
13. https://www.kaggle.com/datasets/ziya07/student-engagement-dataset-using-eeg