A Multi-Modal AI Digital Twin for Personalized Cognitive Well-Being Prediction

Manjusha Madas[1], Rudrani Patre[2], Sakshi Rajole[3], Shreya Kadam[4]
[1,2,3,4]*Student, Department of Computer Applications (MCA), D. Y. Patil Institute of Master of Computer Applications and Management, Maharashtra, India.*

***Abstract*—Cognitive well-being is dynamic and multidimensional, involving stress, affect, workload, sleep-related behaviour and subjective experience. Existing computational approaches often rely on a single sensing modality or population-level models that do not adequately represent individual differences. This paper develops a multimodal artificial-intelligence (AI) digital-twin framework for personalized cognitive well-being monitoring. The framework is designed around five publicly documented secondary datasets: WESAD, StudentLife, SWELL-KW, DEAP and DREAMER. Together, these resources provide complementary physiological, behavioural, task-related and self-reported observations. Because the datasets were collected from different cohorts and under different protocols, they are treated as separate benchmark domains rather than merged into a fictitious common cohort. Dataset-specific encoders and task-appropriate labels feed a modality-aware fusion strategy with explicit availability masks. A global model is followed by subject-level calibration when sufficient individual history is available. The manuscript specifies preprocessing, subject-independent validation, baseline models, ablation experiments and reporting metrics. The dataset review establishes the feasibility and limitations of the proposed multi-domain design; however, no new model-performance values are claimed because the experiments have not yet been executed. The framework is intended as a research and monitoring aid, not a clinical diagnostic system.**

***Index Terms*—Human Digital Twin; Cognitive Well-Being; Multimodal Learning; Stress Recognition; Affective Computing; Personalization; Missing-Modality Learning; Wearable Sensing.**

I. INTRODUCTION

Cognitive well-being is influenced by physiological arousal, emotional state, sleep, daily activity, workload and subjective appraisal. These factors vary across individuals and within the same individual over time. Periodic questionnaires provide useful snapshots but may not capture short-term fluctuations, while sensor-based systems can provide more frequent observations at the cost of introducing noise, missing data and privacy concerns.

The digital-twin concept describes a virtual representation of a real-world entity that is updated using observations from that entity and can be used to monitor or estimate its state [1]. In a human-centred setting, a digital twin can represent selected physiological, behavioural and self-reported indicators. Such a representation must be carefully scoped: it is an estimate based on available measurements, not a complete representation of a person or a substitute for clinical assessment.

This study proposes a multimodal AI digital-twin framework that combines complementary datasets and learning methods while preserving the provenance of each dataset. It focuses on stress-, affect- and workload-related indicators that can reasonably be studied using public secondary data.

1.1 Research Gap and Problem Statement

Three limitations motivate this work: (i) many stress and affect models are trained on a single sensing source; (ii) population-level models may not reflect stable differences in individual baselines; and (iii) practical sensing streams are incomplete, yet missing modalities are not always handled explicitly. In addition, public datasets differ in participants, protocols, sensors and label definitions. Directly pooling their records can create invalid cross-subject or cross-task comparisons. The research problem is therefore to design a reproducible digital-twin pipeline that can learn from heterogeneous benchmark domains without implying that observations from different datasets belong to the same people.

1.2 Objectives

- Design a layered digital-twin architecture for cognitive well-being indicators.
- Identify and document public datasets covering physiological, behavioural, task-related and self-report information.
- Develop dataset-specific preprocessing and feature-extraction pipelines.
- Implement modality-aware fusion with availability masks and controlled modality dropout.
- Compare conventional machine-learning and temporal models using subject-independent evaluation.
- Investigate whether subject-level calibration improves performance without data leakage.
- Address privacy, dataset licensing, fairness and the non-diagnostic nature of the proposed system.

1.3 Contributions

- A modular architecture connecting physiological, behavioural, task and self-report encoders to a personalized digital-twin state.
- A dataset governance strategy that keeps distinct cohorts and label systems separate unless a valid harmonization protocol is demonstrated.
- A missing-modality-aware fusion design with explicit modality availability indicators.
- An evaluation plan covering subject-independent testing, ablation studies and personalization.

II. BACKGROUND AND RELATED WORK

2.1 Human Digital Twins

Digital twins originated in engineering as data-linked virtual counterparts used to monitor and simulate physical systems [1]. Human-centred applications extend this idea to selected physiological and behavioural variables. Their validity depends on data quality, temporal synchronization, model calibration, transparency and governance. For cognitive well-being, the term digital twin should therefore refer to a bounded computational profile rather than a clinical or psychological replica.

2.2 Multimodal Stress and Affect Recognition

WESAD provides wearable physiological and motion signals for baseline, stress and amusement conditions [2]. DEAP and DREAMER provide EEG and peripheral physiological recordings with participant-rated affect dimensions [5], [6]. These datasets support physiological and affective modelling, but their laboratory elicitation protocols should not be equated with long-term everyday mental-health monitoring.

2.3 Behavioural and Workload Sensing

StudentLife records smartphone-derived behavioural information and repeated self-report measures from a student cohort over an academic term [3]. SWELL-KW captures knowledge-work conditions using physiological, behavioural and task-related measurements [4]. They add context that is not represented by wearable physiology alone, although their settings and labels differ.

2.4 Temporal and Personalized Modelling

Long Short-Term Memory (LSTM) networks can represent sequential dependencies [7], while gradient-boosted decision trees provide strong baselines for engineered tabular features [8]. Attention mechanisms offer a way to weight information across time or modalities [9]. Personalization can be introduced through subject-specific normalization, calibration or fine-tuning, provided the test subject's future observations are never used during training.

2.5 Research Positioning

The proposed contribution is an evaluation-ready architecture and protocol. It does not assume that heterogeneous datasets can be concatenated into a single cohort. Instead, it uses dataset-specific tasks and encoders, reports within-dataset subject-independent performance, and treats cross-dataset transfer as a separate experiment requiring explicit label and feature compatibility.

III. MATERIALS AND METHODS

3.1 Study Design

This is a secondary-data methodological study. No new participant recruitment or primary data collection is described. The selected public datasets are used as separate benchmark domains. Before implementation, the research team must download each dataset from its official source, inspect its current licence and documentation, record the release/version, and retain a reproducibility log. Dataset access does not imply permission to redistribute the raw data.

3.2 Dataset Selection and Description

Five datasets were selected to cover complementary but non-identical aspects of cognitive well-being. The participant counts below refer to the published dataset descriptions and must be rechecked against the exact downloaded release before final submission.

Dataset	Participants / setting	Available information	Appropriate task in this study
WESAD [2]	15 participants; controlled wearable study	Chest/wrist ECG, EDA, EMG, respiration, temperature, BVP, acceleration; baseline, stress and amusement labels	Subject-independent stress/affect classification; physiological encoder
StudentLife [3]	48 college students; 10-week academic term	Smartphone sensing including activity, sleep, mobility/social proxies and repeated survey measures including PHQ-9/stress-related measures	Behavioural pattern modelling and association/prediction of available survey outcomes
SWELL-KW [4]	25 knowledge workers; simulated office tasks	Physiological responses, facial/postural/task and computer-interaction information; workload/stress conditions	Workload/stress classification using available physiological and task/behavioural features
DEAP [5]	32 participants; 40 music-video trials	32-channel EEG, peripheral physiological signals and self-rated valence/arousal/dominance	Affect classification/regression; EEG and peripheral-signal fusion
DREAMER [6]	23 participants; 18 audio-visual stimuli	EEG, ECG and self-rated valence/arousal/dominance	Affect classification/regression; EEG–ECG fusion

Important cohort rule: these datasets contain different participants. Their records must not be joined row-by-row or described as one integrated cohort. The primary evaluation is conducted separately within each dataset. A cross-dataset experiment may be reported only as transfer learning or external validation, with compatible features and labels explicitly defined. StudentLife survey labels, WESAD stress conditions, SWELL-KW workload conditions, and DEAP/DREAMER affect ratings are not interchangeable ground truth.

3.3 Data Sources and Access

WESAD is catalogued by the UCI Machine Learning Repository (DOI: 10.24432/C57K5T). StudentLife is described by the Dartmouth/Stanford research publication and its project page. SWELL-KW is associated with the original dataset paper. DEAP is distributed through the project website under its stated access terms. DREAMER is described in the IEEE Journal of Biomedical and Health Informatics publication. The team should cite the original dataset papers and comply with each repository's current access and usage conditions.

3.4 Preprocessing

- Quality control: inspect missingness, corrupted files, sensor saturation, duplicate records and documented participant exclusions.
- Segmentation: create fixed-duration windows appropriate to each sensor and task; record window length, overlap and sampling rate for reproducibility.
- Synchronization: synchronize modalities only within a dataset where timestamps and acquisition protocols support it. Do not synchronize across datasets.
- Normalization: estimate scaling parameters using training participants only. Subject-specific normalization at test time must use only the permitted calibration segment.
- Missingness: preserve a binary availability mask for each modality; imputation values and scaling parameters must be learned from training data only.
- Labels: retain original dataset labels for primary analyses. Any harmonized label mapping must be defined in advance and justified; avoid claiming that affect ratings are equivalent to clinical stress or depression.

3.5 Feature Extraction

Physiological features may include heart-rate and heart-rate-variability statistics where valid ECG/PPG is available, EDA tonic/phasic summaries, respiration statistics, temperature trends, acceleration summaries and EEG band-power features. Behavioural features may include daily activity aggregates, sleep-duration

summaries, mobility regularity and device-use summaries where those fields are present and permitted. Task-related features may summarize interaction patterns and workload conditions. Self-report features should use the original instrument scoring rules and should not be treated as diagnostic labels unless the dataset and instrument support that interpretation.

3.6 Multimodal Fusion and Missing-Modality Handling

Each available modality is processed by a dataset-appropriate encoder to produce an embedding. The fusion module receives the embeddings together with an availability vector. A simple concatenation model is retained as a baseline; an attention-weighted fusion model is evaluated as the proposed alternative. During training, modality dropout can simulate missing streams. Missing inputs are masked rather than interpreted as genuine zero-valued physiological measurements. Fusion is performed only among modalities that co-occur for the same participants and trials within a dataset.

3.7 Personalized Digital-Twin Update

A global model is trained using the training participants. For a held-out participant, the initial twin state is generated from the global model. If the protocol permits a calibration period, a small, temporally earlier portion of that participant's data may be used to estimate personal baselines or adapt a lightweight prediction head. The remaining later observations form the test period. Results must be reported separately for zero-shot global prediction and personalized calibration. No future test labels or windows may enter calibration.

3.8 Models and Evaluation Protocol

Component	Models / protocol
Classical baselines	Logistic regression, random forest and XGBoost on engineered features
Temporal models	LSTM on ordered windows; optional attention-based temporal/fusion model
Primary split	Subject-independent split; leave-one-subject-out (LOSO) where dataset size permits
Classification metrics	Accuracy, balanced accuracy, macro precision, macro recall, macro-F1 and ROC-AUC where class support permits
Regression metrics	MAE and RMSE for continuous affect or questionnaire outcomes, reported only when label scale supports regression
Uncertainty	Mean and standard deviation across held-out subjects/folds; confidence intervals where statistically appropriate
Ablations	Single modality; multimodal fusion; fusion without availability mask; fusion with mask; global versus calibrated twin

All preprocessing, feature selection, hyperparameter tuning and threshold selection must occur inside the training folds. Windows from one participant must never be split across training and test sets in a way that leaks subject identity. Report per-subject scores in addition to aggregate means where possible.

IV. PROPOSED SYSTEM ARCHITECTURE

The proposed system contains six logical layers:

- Data ingestion: imports approved physiological, behavioural, task and self-report records with dataset and participant identifiers.
- Preprocessing and feature layer: performs quality checks, windowing, normalization and modality-specific feature extraction.
- Modality encoder layer: maps each available modality into a latent representation using dataset-appropriate encoders.
- Fusion layer: combines embeddings with an availability mask and optional attention weights.
- Digital-twin core: maintains an individual baseline, recent history, uncertainty and calibrated prediction state.
- Prediction and feedback layer: displays task-specific estimates and trends with clear uncertainty and non-diagnostic wording.

Figure 1 should depict the six layers and show that each dataset enters through its own validated input pipeline. A dashed boundary should separate within-dataset fusion from optional cross-dataset transfer experiments.

V. RESULTS AND DISCUSSION

5.1 Dataset-Selection and Feasibility Results

The dataset review identifies complementary coverage across the selected benchmarks. WESAD provides

wearable physiological signals with explicit stress-related experimental conditions. StudentLife contributes longitudinal smartphone-derived behavioural observations and repeated self-report measures. SWELL-KW contributes a knowledge-work context with physiological and task-related information. DEAP and DREAMER provide EEG/peripheral physiology paired with affect self-ratings. This coverage supports the design of modality-specific encoders and benchmark-specific tasks.

Research dimension	Dataset evidence	Interpretation
Physiological stress response	WESAD; SWELL-KW	Supports stress/workload-related modelling in controlled or simulated settings
Behaviour over time	StudentLife	Supports longitudinal behavioural analysis in a student population
Physiology with affect ratings	DEAP; DREAMER	Supports valence/arousal/dominance prediction, not direct clinical diagnosis
Multiple modalities within one cohort	WESAD; SWELL-KW; DEAP; DREAMER	Enables within-dataset fusion where modalities are jointly recorded
Behaviour plus self-report	StudentLife	Enables analysis of behavioural indicators against available survey outcomes

5.2 Model-Performance Results

No model has been trained or evaluated as part of the material supplied for this revision. Consequently, accuracy, macro-F1, ROC-AUC, MAE and RMSE values cannot be reported as experimental findings at this stage. Reporting invented or literature-copied values as results of the proposed system would be scientifically misleading. The following table is the required results template and must be populated from the actual held-out-subject experiments.

Dataset / task	Model	Accuracy / balanced accuracy	Macro-F1 / MAE	ROC-AUC / RMSE
WESAD stress classification	Logistic regression / Random Forest / XGBoost / LSTM / proposed fusion	To be measured	To be measured	To be measured
StudentLife survey outcome	Baseline / temporal model / personalized twin	To be measured	To be measured	To be measured
SWELL-KW workload/stress	Baseline / multimodal fusion / personalized twin	To be measured	To be measured	To be measured
DEAP affect prediction	EEG-only / peripheral-only / fused model	To be measured	To be measured	To be measured
DREAMER affects prediction	EEG-only / ECG-only / fused model	To be measured	To be measured	To be measured

5.3 Ablation Results

Configuration	Required comparison	Result
Single-modality model	Each modality evaluated independently	To be measured
Multimodal fusion without availability mask	Standard fusion baseline	To be measured
Multimodal fusion with availability mask	Effect of explicit missing-modality handling	To be measured
Global model	No subject-level adaptation	To be measured
Personalized twin	Calibration using permitted earlier observations	To be measured

5.4 Discussion

The dataset audit indicates that a single universal label space would be inappropriate without further validation. Stress conditions in WESAD, workload manipulations in SWELL-KW, affect ratings in DEAP and DREAMER, and survey outcomes in StudentLife measure related but distinct constructs. The proposed architecture therefore emphasizes task-specific prediction heads and within-dataset evaluation. The principal empirical questions remain whether fusion improves over the strongest single modality, whether availability masks reduce degradation under missing inputs, and whether subject-level calibration improves held-out future predictions. These questions can be answered only after the experiments in Sections 3.8 and 5.2–5.3 have been run.

For final submission, include confusion matrices for classification tasks, per-subject performance plots, learning curves, and modality-ablation plots. Each figure must be generated from the final experimental run and accompanied by the split protocol, sample counts and uncertainty estimates.

VI. LIMITATIONS AND FUTURE WORK

- The datasets were collected from different populations, settings, devices and protocols; they do not form one common cohort.
- Some datasets are small at the participant level, making estimates sensitive to individual subjects and limiting deep-model generalization.
- Laboratory affect elicitation and simulated workload are not equivalent to naturalistic long-term cognitive well-being.
- Self-reported scores are subjective and may not be interchangeable across instruments or studies.
- Public dataset availability does not guarantee unrestricted redistribution; licences and access agreements must be checked.
- No clinical validation or prospective deployment is claimed.
- Future work should implement the pipeline, conduct subject-independent and external-transfer experiments, assess calibration and fairness, and validate the system prospectively under institutional ethics approval.

VII. ETHICAL AND PRIVACY CONSIDERATIONS

Physiological, behavioral and mental-health-related data are sensitive. The study should use only data obtained through authorized public repositories, minimize retained identifiers, store data securely and restrict access to the research team. Dataset-specific consent and licence terms must be respected. Any dashboard should communicate uncertainty and avoid diagnostic language. Predictions should not be used for employment, academic grading, insurance or other consequential decisions. The framework is intended for research and supportive self-monitoring, not diagnosis or treatment.

VIII. CONCLUSION

This paper presents a multimodal AI digital-twin framework for personalized cognitive well-being prediction and identifies five complementary public datasets for its evaluation. The design combines modality-specific encoders, missing-modality-aware fusion and carefully controlled subject-level calibration. A central methodological safeguard is to preserve dataset boundaries: the selected datasets are distinct benchmark domains, not a single shared population. The dataset review supports the feasibility of evaluating physiological, behavioural, task-related and self-report signals within their respective contexts. The effectiveness of the proposed models remains to be established through the specified experiments; no unperformed experimental results are claimed.

DECLARATIONS

- Funding: The authors should confirm the funding status. If no external funding supported this work, use: “This research received no external funding.”
- Conflict of interest: The authors should confirm that no financial or personal relationships could have influenced the work. If accurate, use: “The authors declare that they have no competing interests.”
- Ethics approval: This manuscript describes secondary analysis of publicly available datasets and no new human-participant recruitment. The authors must verify the institutional and target-journal requirements before stating that formal ethics approval was not required. Any dataset-specific data-use conditions remain applicable.
- Informed consent: No new informed consent was obtained by the authors because no primary participant data were collected. Consent procedures are those documented by the original dataset creators; this statement must be checked against each dataset's documentation and the journal's policy.
- Data availability: WESAD is available through the UCI Machine Learning Repository (https://archive.ics.uci.edu/dataset/465/wesad). StudentLife information and access are provided through the project publication/page (https://maplab.stanford.edu/publications/smartph

one/studentlife-assessing-mental-health-academic-performance-and-behavioral). SWELL-KW should be obtained from the original dataset source associated with Koldijk et al. DEAP access is provided through the official project page (https://www.eecs.qmul.ac.uk/mmv/datasets/deap/), subject to its terms. DREAMER is described in the original publication (DOI: 10.1109/JBHI.2017.2688239); use the authorized repository/source. Dataset access terms should be rechecked before submission. No new code repository is claimed; add a public code URL only after the code has been deposited.

- Author contributions: Proposed CRediT roles for confirmation: Manjusha Madas conceptualization, methodology, investigation, writing original draft; Rudrani Patre data curation, methodology, validation; Sakshi Rajole formal analysis, visualization, writing review and editing; Shreya Kadam literature review, validation, writing review and editing. All authors should review and approve the final allocation and manuscript.
- Acknowledgements: The authors may acknowledge their department, institution and research guide after confirming the preferred wording and obtaining permission where required.
- Corresponding author: Manjusha Madas.

REFERENCES

[1]. M. Grieves and J. Vickers, "Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems," in Transdisciplinary Perspectives on Complex Systems. Cham, Switzerland: Springer, 2017, pp. 85–113.

[2]. P. Schmidt, A. Reiss, R. Dürichen, C. Marberger, and K. Van Laerhoven, "Introducing WESAD, a multimodal dataset for wearable stress and affect detection," in Proc. 20th ACM Int. Conf. Multimodal Interaction (ICMI), 2018, pp. 400–408, doi: 10.1145/3242969.3242985.

[3]. R. Wang et al., "StudentLife: Assessing mental health, academic performance and behavioral trends of college students using smartphones," in Proc. ACM Int. Joint Conf. Pervasive and Ubiquitous Computing (UbiComp), 2014, pp. 3–14, doi: 10.1145/2632048.2632054.

[4]. S. Koldijk, M. A. Sappelli, S. Verberne, M. A. Neerincx, and W. Kraaij, "The SWELL knowledge work dataset for stress and user modeling research," in Proc. 16th ACM Int. Conf. Multimodal Interaction (ICMI), 2014, pp. 291–298.

[5]. S. Koelstra et al., "DEAP: A database for emotion analysis using physiological signals," IEEE Trans. Affect. Comput., vol. 3, no. 1, pp. 18–31, 2012, doi: 10.1109/T-AFFC.2011.15.

[6]. S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," IEEE J. Biomed. Health Inform., vol. 22, no. 1, pp. 98–107, 2018, doi: 10.1109/JBHI.2017.2688239.

[7]. S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Comput., vol. 9, no. 8, pp. 1735–1780, 1997.

[8]. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in Proc. ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining, 2016, pp. 785–794.

[9]. A. Vaswani et al., "Attention is all you need," in Advances in Neural Information Processing Systems, 2017.

[10]. C. Tang et al., "A roadmap for the development of human body digital twins," Nature Reviews Electrical Engineering, vol. 1, pp. 199–207, 2024, doi: 10.1038/s44287-024-00025-w.

[11]. "Concepts and applications of digital twins in healthcare and medicine," Patterns, vol. 5, no. 8, Art. no. 101028, 2024, doi: 10.1016/j.patter.2024.101028.

[12]. I. Moura, A. Teles, D. Viana, J. Marques, L. Coutinho, and F. Silva, "Digital phenotyping of mental health using multimodal sensing of multiple situations of interest: A systematic literature review," Journal of Biomedical Informatics, vol. 138, Art. no. 104278, 2023, doi: 10.1016/j.jbi.2022.104278.

[13]. T. Baltrušaitis, C. Ahuja, and L.-P. Morency, "Multimodal machine learning: A survey and taxonomy," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 41, no. 2, pp. 423–443, 2019, doi: 10.1109/TPAMI.2018.2798607.

[14]. R. Wu, H. Wang, H.-T. Chen, and G. Carneiro, "Deep multimodal learning with missing

modality: A survey," Transactions on Machine Learning Research, 2026.

[15]. M. Ma, J. Ren, L. Zhao, S. Tulyakov, C. Wu, and X. Peng, "SMIL: Multimodal learning with severely missing modality," in Proc. AAAI Conf. Artificial Intelligence, vol. 35, no. 3, pp. 2302–2310, 2021, doi: 10.1609/aaai. v35i3.16330.

[16]. A. O. Pataca et al., "Use of machine learning for predicting stress episodes based on wearable sensor data: A systematic review," Computers in Biology and Medicine, vol. 198, Art. no. 111166, 2025.

[17]. G. Udahemuka, K. Djouani, and A. M. Kurien, "Multimodal emotion recognition using visual, vocal and physiological signals: A review," Applied Sciences, vol. 14, no. 17, Art. no. 8071, 2024, doi: 10.3390/app14178071.

[18]. R. Büscher et al., "A systematic review on passive sensing for the prediction of suicidal thoughts and behaviors," npj Mental Health Research, vol. 3, Art. no. 42, 2024.

[19]. B. H. Tudor et al., "A scoping review of human digital twins in healthcare applications and usage patterns," npj Digital Medicine, vol. 8, Art. no. 587, 2025, doi: 10.1038/s41746-025-01910-w.