MULTIDISCIPLINARY RESEARCH INSTITUTE INDIA

Open Access International Journal of Science and Engineering

ISSN: 2456-3293

Volume 9 Issue 1s (2026) | Pages: 20-25

A Responsible and Risk-Aware Artifical Intelligence Framework for Multimodal Sleep Cycle Monitoring Using Wearable and Contactless Physiological Sensing

Soham Chouk[1], Aditi Gupta[2], Eva Singh[3], Rajesh B. Ghongade[4], Ay L. Prabhakar[5]

[1,2,3,4,5]Department of Electronics and Telecommunication Engineering, Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, India

[1]sachouk22-etc@bvucoep.edu.in, [2]agupta22-etc@bvucoep.edu.in, [3]esingh22-etc@bvucoep.edu.in, [4]rbghongade@bvucoep.edu.in, [5]ayprabhakar@bvucoep.edu.in

Peer Review Information

Type: *Article*
Received: *29 March 2026*
Revised: *26 April 2026*
Accepted: *30 May 2026*
Published: *19 June 2026*

Abstract

The use of artificial intelligence has made persistent and tailored monitoring of physiological functions possible, though with the problem of reliability and safety in deployment . The aspect of sleep monitoring is the most sensitive since it is closely linked with cardiovascular, metabolic, and cognitive well-being. Most current automated systems are based on a small number of sensing modalities and non-transparent learning models which limit robustness and interpretability in real-word settings.

This paper introduces a multimodal artificial intelligence sleep monitoring system that combines heterogenous physiological and behavioral indicators of sleep cycles in a late-fusion learning system. The framework maintains modality- specific representations and enables adaptive temporal integration to enhance resilience in unconstrained conditions. In order to raise its reliability, the system integrates uncertainty-aware inference and intrinsic explainability, enabling predictive confidence estimation and features contributions supplied by individual modalities.

Instead of creating an autonomous diagnostic system, the offered method acts under the version of the decision-support tool because it reveals interpretable outputs and highlights the predictions of low level of confidence to be reviewed by experts. Experimental findings show better performance compared to unimodal baselines especially in physiologically ambiguous sleep states, while maintaining transparency and reliability.

This framework adds a deployable and interpretable solution of intelligent sleep monitoring solution to healthcare.

Keywords: Missing Person Detection; YOLOv8; CCTV Surveillance; Clothing Attribute Recognition; Computer Vision; Artificial Intelligence.

How to Cite This Article

Chouk, S., Gupta, A., Singh, E., Ghongade, R. B., & Prabhakar, A. L. (2026). A Responsible and Risk-Aware Artificial Intelligence Framework for Multimodal Sleep Cycle Monitoring Using Wearable and Contactless Physiological Sensing. *Open Access International Journal of Science and Engineering*, 9(1s), 20-25.

Introduction

The growing use of artificial intelligence in the healthcare sector has helped to establish the ongoing and customized tracking of physiological processes, which has opened up new prospects for prophylactic and evidence-based health evaluation. Sleep monitoring has been one of these applications, particularly because sleep quality has a close relationship with cardiovascular health, metabolic control, cognitive functioning, and psychological state [1], [3]. Although conventional techniques of sleep assessment were considered to be clinically valid, they tend to be obtrusive and ineffective in terms of long-term perspectives of monitoring, which prompts the creation of wearable and ambient sensor-related solutions [1].

Technically, developments in AI in the recent past have enabled sleep analysis based on physiological signals acquired in practical settings. It was established in a number of studies that machine learning and deep learning methods have the potential to be applied in sleep-stage classification and sleep-wake sensors [5], [6]. But the various systems that are presently in place make use of single-modality detection, or limited mixtures of cues, are susceptible to noise, signal loss, and inter-individual disparity [7]. What is more, this rise in the use of sophisticated deep learning models has created issues regarding interpretability, transparency, and trust, especially in health-sensitive aspects of decision-support systems [11].

Simultaneously, there has been the increase of interest in the responsible use of artificial intelligence in healthcare. Issues of model opacity, algorithmic bias, uncertainty and regulatory appropriateness have suggested the necessity of AI systems that are both true to themselves but also transparent, risk-conscious, and responsible [8], [10], [11]. Current sleep monitoring devices that offer web-based and mobile applications are seldom based on the tenets of responsible AI, therefore leaving a discrepancy between technical operation and credible real-world application [7].

To specifically fill the specified gap, the current paper suggests a responsible and risk- conscious artificial intelligence model for multimodal sleep cycle monitoring that can combine heterogeneous physiological and behavioral cues. The proposed solution with the integration of multimodal information, explainable and uncertainty-sensitive inference will improve resilience, interpretability, and safety in unconstrained monitoring scenarios. An ethical and governance strategies, such as avoiding privacy-sensitive data handling and fairness evaluation will be established throughout the system lifecycle in connection with the new standards of trustworthy artificial intelligence [12]-[15]. With this combination, the work aims at furthering credible and responsible sleep tracking technologies that can facilitate effective expert oversight.

Literature Review

Sleep monitoring using artificial intelligence has developed from classic actigraphy and statistical classifiers in favor of deep learning models that can learn complicated temporal relationships in physiological signals [1], [6]. Although such methods have led to better accuracy in the classification of sleep-stages, most have been limited to the use of single-sensing modalities or highly combined inputs, which restricts their performance in real-world situations [7].

Multimodal learning has turned out to be a promising trend because it incorporates the use of heterogeneous physiological cues to enhance resilience and generalization [4]. Nevertheless, the current reality with multimodal systems is the use of early-fusion approaches or opaque integration approaches such whereby the contribution of modalities is not observed and interpretation is limited [9].

The attentive-based fusion and explainable learning methods have been presented in recent work to enhance the clarity of health-related artificial intelligence [8], [11]. However, they have been little used in previous sleep monitoring and incorporate explicit modeling of predictive uncertainty or deployment risk [10].

Moreover, other aspects of ethics and governance, including protection of privacy, fair- ness, and accountability, are not frequently addressed in system design as they are treated as minor issues [12], [13]. The identified gap encourages the creation of sleep monitoring frameworks that will serve all three goals performance, transparency, and reliability, as incorporated in this work.

Problem Statement and objectives

Existing artificial intelligence-based sleep monitoring systems have three major shortcomings: a lack of robustness due to the use of single or weakly combined sensing modalities, the lack of interpretability due to the use of opaque learning models, combined with the inadequate treatment of predictive uncertainty in health-critical settings [11], [14]. These issues limit clinical trust and safe deployment in real-world situation.

This study tries to overcome these limitations with the following objectives:

- To develop a multimodal framework for reliable sleep monitoring based on heterogeneous physiological and behavioral heterogeneous signals [6], [4].

- To create a late-fusion learning architecture to maintain modality-specific interpret- ability while also enabling adaptive integration [9].
- To make uncertainty-aware inference to mitigate overconfident misclassification [10].
- To support clear and trustworthy decision-making with the help of explainability and human review [8], [11].

Materials And Methods

This section introduces the proposed sleep monitoring framework, which covers the system architecture, the use of a multimodal learning strategy, and the research hypotheses for the study.

Overview of System Architecture

The proposed system takes a modular approach to sleep cycle monitoring based on diverse physiological and contextual inputs through a common analysis pipeline [7]. At the data acquisition layer, the system combines both body-worn and ambient sensing sources to acquire multiple dimensions of sleep-related behavior such as thermoregulatory activity, autonomic response, respiratory patterns, and physical movement, as well as postural changes[1], [3], [4]. Of all these techniques, the use of complementary sensing sources reduces the dependency on one or only a few data sources and increases the fault tolerance [7].

All the acquired signals are processed through a preprocessing stage which performs noise suppression, artifact removal, normalization and temporal synchronization [6]. This step deals with designing around the issue of interference from motion and environmental effects. Only the information that is relevant to the task is kept for further analysis [12]. A feature extraction layer is modality-specific representations while physiological meaning must still be preserved [8]. Each data stream is processed separately in order to preserve the semantic clarity and make it possible to trace the links between the input signals and the model output, without performing premature fusion of the heterogeneous information [9].

Learning in a Conceptual Framework using the Multimodal Learning Framework.

So in order to model the dynamic characteristics of sleep physiology the framework uses a late-fusion learning strategy in which the modality-specific representations are learned separately before their integration takes place [7], [9]. By doing so, information can be maintained throughout every stream of data so that the temporal and statistical properties of any given data set are retained [6].

Each modality is processed by a dedicated encoder responsible for transforming inputs into latent temporal representations which are relevant for the task of sleep-stages classification [5], [6]. These encoders contain both short-term changes and long-range temporal dependencies related to transitions in sleep [6], [7].

The modality specific latent representations are then combined using an attention-based temporal fusion mechanism [8], [9]. This allows adaptive weighting of the contributions of different modalities over time so that the contributions of the most informative signals can be focused on during the different phases of sleep [7].

To improve the reliability, the framework introduces uncertainty-aware inference algorithms that both estimate predictive confidence in addition to classification answers [10]. These estimations help to identify ambiguous or low confidence epochs, reducing the likelihood of overconfident prediction [11], [14].

The learning framework is highly flexible and serves as a decision support system vs. an autonomous diagnostic tool [11], [14]. Explainable outputs and confidence measures are explicitly said to be accessible for human interpretation and oversight [15].

Multimodal Learning Framework

To reflect the dynamic nature of sleep physiology the modelling adopts a modular late- fusion learning process [7], [9], where modality specific representations are acquired separately before combined [7].

Let each sensing modality generate a temporally segmented input sequence [5], [6].

$$X_m = \{x_{m,t}\}_{t=1}^{T}$$

T= Epochs of sleep. A special encoder network processes each of the modality $fm(\cdot)$ [7], [9], producing a latent representation.

$$Z_m = f_m(X_m)$$

This design allows the learning of modality-specific temporal and physiological properties to be done in isolation and semantic interpretability is maintained.

The temporal fusion mechanism based on the integration of the modality-specific latent representations is an attention-based mechanism [8], [9], that calculates context-dependent attention weights α(m,t) at every time step of the modalities. The fused representation *Z*t is defined as:

$$Zt = \Sigma \sum_{m=1}^{M} \alpha m,t \cdot Zm,t \quad (1)$$

The modulation of modality relevance by sleep stage by this formulation enables the model to dynamically regulate modality relevance [6], [7], in response to changes in the sleep physiology that is non-stationary. A classification head is used to make predictions of sleep stages. Simultaneously, an uncertainty estimator is a predictive uncertainty estimator [10], [14], that calculates confidence scores by probabilistic inference algorithms like ensemble variance or distributional entropy [10]. Predictions that have a high level of uncertainty that is above a set of uncertainty levels are labeled as low-confidence, and sent to human inspection [11], [14], [15].

Explainability is obtained with the analysis of attention weights [8], [9], [11]. and feature-level attributions that allow understanding of the contribution of each modality to every prediction. The model thereby incorporates forecasting, elucidation, and risk identification in a single inference framework [8].

Research Hypotheses

The proposed research study is guided by the following research hypotheses:

- H1: Multimodal fusion of heterogeneous physiological signal contents produce bet- ter sleep stage classification performance than unimodal ones under real conditions [4], [6], [7].
- H2: Late-fusion architectures that use modality-specific encoders and attention- based fusion make them more robust to noise and signal degradation than early-fusion models [7], [9].
- H3: Incorporation of the uncertainty estimation decreases overconfident misclassification and produces reliable physiologically ambiguous sleep stage judgments [10], [11], [14].
- H4: Explicit modeling of modality contributions improve interpretability without hurting predictive performance [8], [9], [11].

These hypotheses are experimentally tested using comparative experiments and predictive confidence, attention distributions, and classification accuracy analysis [6], [8], [10].

Explainability and Risk-Aware Modeling

Transparency and reliability are crucial for domains such as sleep monitoring where artificial intelligence is used in healthcare applications [11]. To address the drawbacks of the opaque predictive models, the proposed framework adds both explainability and risk awareness directly into the learning pipeline instead of having them as a post hoc project addition.

Explainability is given via attention-based attribution and feature-level analysis that makes it possible to find modality-specific contributions to sleep stage prediction through time [8], [9]. By revealing how various physiological inputs manifest in model outputs at varying phases of sleep makes it easy to physiologically consistent interpret learned representations. This aids validation of the model and clinical plausibility [9], [11], by providing the ability for domain specialists to evaluate whether behaviour of the system conforms to known physiology of sleep.

The attention mechanism can further be used to analyse temporal changes in signal relevance, with potential to provide information on how the importance of modalities changes between sleep stages. Such intrinsic interpretability has special importance in health care applications, where understanding the basis of automated decisions is of key importance in terms of trust and adoption. In parallel, the framework includes uncertainty-aware inference, a way for coping with the variability and noise in actual physiological data. Estimation of predictive confidence is done alongside sleep stage outputs, allowing for the identification of ambiguous or low-confidence epochs, especially in the case of outcomes for transitional sleep stages or cases of degraded signal [10].

By being able to quantify uncertainty, the system will reduce overconfident misclassification and enable safer decision-making [14]. Predictions that cross preset limits on uncertainty are marked for expert consideration thus maintaining the importance of using the framework as an aid to decision making versus a standalone diagnostic tool [11], [14], [15]. Through the joint integration of explainability and uncertainty modeling, the proposed approach promotes a transparent, reliable and trusted method, without reducing the predictive performance, which strengthens its suitability for healthcare deployment.

Ethics and Governance

The implementation of artificial intelligence in healthcare requires the inclusion of ethical and governance issues in the system design [12], [13]. The ethical principles of the suggested structure are enforced by means of technical and procedural protection that guarantees responsible use of data and transparent model behavior [15].

Preservation of privacy is met by privacy-by-design measures that are carried out around data acquisition, preprocessing as well as storage of data [12]. Instead, only task-relevant physiological information is stored and only access to raw data is restricted to curb the misuse. Safe data processing protocols minimize the release of confidential data and assist in meeting the requirements of data protection in healthcare [14].

Fairness has become a performance metric by assessing the consistency of model behavior across subgroups of demographics that allows systematic differences to be dis- covered and corrective refinements to be made in the course of development [13], [15].

Audit-ready logging, explainable outputs, and clearly defined uncertainty estimation provide accountability and traceability [11], [14]. Imprecise predictions or predictions of low confidence are identified and left to be reviewed by an expert keeping the decision of critical communications in the hands of the human being. With these steps, the principles of ethics are converted into operationally appropriate safeguards applicable to controlled healthcare implementation.

Experimental Evaluation and Results

Experimental Setup

To assess the proposed framework, observational data of sleep have been used, which was gathered in the absence of constraints that represent the real world conditions [5], [6]. These data are synchronized physiological and behavioral signals that have fixed length epochs that are annotated with clinically defined sleep-stage.

Unimodal baseline models are built on single physiological streams whereas multi- modal models use the proposed late-fusion structure [6], [7]. All models are trained with similar network capacity and with the same optimization settings.

Accuracy, macro-averaged F1-score, and confusion matrices are used to evaluate the performance with special attention to transitional and rapid-eye-movement sleep phases[5], [7]. The measure of robustness is evaluated by additive noise and partial signal dropout controlled perturbations [7].

Results and Analysis

Multimodal integration is always superior to unimodal baselines in all measures meaning that it has better generalization and tolerance to the variations in the signals [6], [7]. The highest recovery of performance in physiologically ambiguous stages where unimodal models exhibits rise in rates of misclassification.

Attention-based fusion presents sleep stage-specific and physiologically stable relevance of modalities, which can be interpreted by the human [9]. Uncertainty-aware inference also improves the level of reliability since it is able to detect the uncertain or noisy cases and decrease overconfident misclassification [10], [14].

Interestingly, explainability, and uncertainty estimation do not reduce along with predictive performance, which proves that transparency and reliability could be attained and consistency in classification can remain high [9], [10].

Discussion

These results imply that predictive performance is complementary and not contradictory in relation to transparency and reliability [9], [15]. The results confirm that sleep is a naturally multidimensional phenomenon and cannot be robustly characterised using isolated sensing modalities [1], [6]. The enhanced performance of the multimodal framework is shown as an example of using heterogeneous physiological and behavioral signals for reliable sleep-stage classification [5], [7].

Interpretability is enhanced by attention-based fusion that produces physiologically plausible patterns of modality contributions across the phases of sleep [8], [9]. This correspondence between learned representations and domain knowledge helps to sup- port the clinical plausibility of the model [3], [9].

Uncertainty-aware inference provides safety by allowing for selective human review of predictions that are ambiguous, especially in unconstrained environments where the quality of signals varies [10], [11]. These findings indicate that the principles of transparency and reliability are not mutually exclusive and thus do not come at the expense of predictive performance [9], [15].

Conclusion

This paper proposes a multimodal artificial intelligence framework for sleep cycle monitoring to improve robustness, interpretability, and deployment reliability [7], [11]. By incorporating heterogeneous physiological and behavioral measures through late fusion architecture, the proposed technique will be able to achieve better performance than unimodal baselines (especially in ambiguous sleep stages) [5], [6].

Experimental results prove consistent improvement in classification accuracy and re- duction in misclassification along with attention-based fusion and uncertainty-aware inference modelling to improve interpretability and predictability [8], [9], [10]. The framework supports human-centered decision-making, and the outputs are explainable and have confidence estimates, instead of operating as an autonomous decision diagnostic tool [11], [14].

Overall, the proposed approach presents a technically sound and practical solution for intelligent sleep monitoring which has significant potential to be further extended in the future to scalable and personalized healthcare applications [12], [15].

References

1. M. Ancoli-Israel et al., "The role of actigraphy in the study of sleep and circadian rhythms," Sleep, vol. 38, no. 5, pp. 665–680, 2015.
2. K. Kräuchi, "How is the circadian rhythm of core body temperature regulated?" Sleep Medicine Reviews, vol. 15, no. 5, pp. 343–355, 2011.
3. H. Hong, Y. Lin, and C. Gu, "Noncontact sleep monitoring using impulse radio radar," IEEE Microwave Magazine, vol. 20, no. 8, pp. 55–68, 2019.
4. C. Sano and R. W. Picard, "Stress recognition using wearable sensors and mobile phones," IEEE Trans. Affective Computing, vol. 4, no. 3, pp. 301–314, 2013.
5. M. Radha et al., "Sleep stage classification from heart rate variability using deep learn- ing," npj Digital Medicine, vol. 4, no. 1, 2021.
6. Y. Sun et al., "Sleep staging from physiological signals using deep neural networks," Physiological Measurement, vol. 40, no. 12, 2019.
7. R. Zhai et al., "Challenges and opportunities of deep learning for sleep analysis," npj Digital Medicine, vol. 7, 2024.

A. Holzinger, "Explainable AI and multi-modal causability in medicine," i-com, vol. 19, no. 3, pp. 171–179, 2020.

8. W. Samek et al., Explainable AI: Interpreting, Explaining and Visualizing Deep Learn- ing, Springer, 2019.
9. A. Kendall and Y. Gal, "What uncertainties do we need in Bayesian deep learning," NeurIPS, pp. 5574–5584, 2017.
10. C. Rudin, "Stop explaining black box machine learning models," Nature Machine Intel- ligence, vol. 1, pp. 206–215, 2019.
11. European Commission, "Ethics guidelines for trustworthy AI," 2019.
12. OECD, "OECD principles on artificial intelligence," 2019.
13. U.S. FDA, "Good machine learning practice for medical device development," 2021.
14. T. Grote and P. Berens, "On the ethics of algorithmic decision-making in healthcare," Journal of Medical Ethics, vol. 46, no. 3, pp. 205–211, 2020.