Hybrid Autoencoder-LightGBM Approach for Predicting Sleep Health Metrics from Multimodal Lifelog Data

Minjeong Kim
EC21RNC
kmjng0712@gmail.com

Seongpil Han
Korea Testing Laboratory
seukwanghsp@gmail.com

***Abstract* - The analysis of lifelogs can yield valuable insights into an individual's daily life, particularly with regard to their sleep health and well-being. The accurate prediction of sleep quality metrics is necessitated by the effective integration of heterogeneous sensor data and robust feature extraction methodologies. To address this challenge, this study proposes a hybrid autoencoder-enhanced LightGBM framework for sleep health prediction from multimodal lifelog data. Our approach employs an autoencoder-based feature extraction methodology to capture latent representations from diverse sensor inputs. These extracted features are then combined with existing variables through feature fusion, followed by the use of optimized LightGBM classifiers for predicting sleep quality, fatigue, and stress levels. Experiments were conducted using the ETRI lifelog dataset comprising 12 distinct sensor modalities from smartphones and smartwatches to determine the optimal feature combinations and model configurations. Through comprehensive hyperparameter optimization using grid search, we attempted to improve prediction accuracy across both binary and multiclass classification tasks. Additionally, we identified the most influential sensor features for sleep health assessment through systematic feature importance analysis. As a result, our hybrid approach demonstrated superior performance compared to traditional single-model baselines, with improvements in Macro F1-scores for stress level prediction and sleep efficiency adherence. The detailed experimental setup and data analysis procedures of this study are thoroughly discussed in the modeling part of this methodology section. The complete implementation is available for reproducibility and further research.**

***Index Terms* - Lifelog, Sleep Health, Sleep Quality, Stress Prediction, Fatigue Assessment, Deep Learning, Autoencoder, LightGBM, Feature Engineering, Wearable Devices, Health Monitoring, Multimodal Sensor Fusion, Machine Learning, Digital Biomarkers**

I. Introduction

Sleep quality and stress levels represent fundamental indicators of human health and well-being, yet conventional assessment approaches frequently depend on subjective self-reporting or costly polysomnography performed in controlled clinical settings. The proliferation of smartphones and wearable technology has positioned multimodal lifelog data as a compelling alternative for continuous, objective health monitoring in naturalistic environments.

Contemporary developments in machine learning have showcased the capabilities of sensor-based methodologies for sleep and stress assessment. Nevertheless, most current studies concentrate on predicting individual metrics in isolation or employ restricted sensor modalities. For example, TzuAn Song's SLAMSS (Sequence-to-sequence LSTM for Automated Mobile Sleep Staging) model demonstrated 79% overall accuracy in three-class sleep staging using wrist-worn actigraphy and heart rate data PLOSPubMed [1]. Rohit Gupta's research on stress prediction using multimodal wearable sensors from wrist and chest devices achieved high classification performance, though evaluation was conducted under controlled laboratory conditions ResearchGateSemantic Scholar [2]. Hang Yuan developed a self-supervised deep neural network for sleep stage classification utilizing wrist-worn accelerometers, achieving fair to moderate agreement with polysomnography, with external validation showing a difference of 34.7 minutes in total sleep time prediction [3].

The current study addresses several critical limitations in standard research through the following hybrid deep learning architecture approach.

A. *Comprehensive Multi-target Prediction Framework*

Unlike previous studies that predict sleep quality or stress levels independently, this study's approach simultaneously predicts six distinct sleep health metrics from the 2025 ETRI lifelog challenge: three subjective survey-based measures (Q1: overall sleep quality, Q2: physical fatigue, Q3: stress level) and three objective sleep guideline adherence measures (S1: total sleep time, S2: sleep efficiency, S3: sleep onset latency).

The subjective metrics (Q1-Q3) were originally collected using 5-point Likert scales from pre- and post-sleep questionnaires but converted to binary classifications based on individual participant baselines. For instance, Q1 is assigned a value of 1 when an individual's self-reported sleep quality exceeds their personal average, and 0 when below average. Similarly, Q2 and Q3 indicate positive outcomes (1) when fatigue and stress levels are below individual averages. The objective metrics (S1-S3) reflect adherence to National Sleep Foundation guidelines [NSF Guidelines], with S1 uniquely employing a 3-class system (0: not recommended, 1: may be appropriate, 2: recommended) while S2 and S3 use binary classifications (0: inappropriate, 1: recommended).

B. *Novel Hybrid Deep Learning Architecture*

We propose a hybrid autoencoder-LightGBM framework that combines unsupervised feature learning with gradient boosting classification. While traditional approaches rely solely on hand-crafted features or end-to-end deep learning, our method leverages the complementary strengths of both paradigms by concatenating 20 original domain-specific features with 20 autoencoder-derived latent representations to create a comprehensive 40-dimensional feature space.

II. METHODLOGY

A. Datasets

The dataset was collected as part of the 2025 challenge using Android smartphones and smartwatches with sampling intervals of 1-10 minutes. Data timestamps follow Korea Standard Time (KST) in ISO 8601 format. The dataset exhibits inherent missing data and noise characteristics due to intermittent device usage and system interruptions. Given that these limitations are intrinsic to wearable sensor data collection, minimal preprocessing steps were applied before conducting the analysis. For privacy protection, GPS coordinates are provided as relative positions rather than absolute geographical locations. Data files are organized by participant ID and timestamp across multiple sensor modalities as shown in Table 1.

TABLE I. The detailed composition [4]

Items	Column	Data Type	Note
mACStatus	m_charging	integer	0: No, 1: Charging
mActivity	m_activity	integer	0: in_vehicle, 1: on_bicycle, 2: on_foot, 3: still, 4: unknown, 5: tilting, 7: walking, 8: running
mAmbience	m_ambience	object	List of ambient sound labels and their respective probabilities
mBle	m_ble	object	List of bluetooth device address, device_class, and rssi
mGps	m_gps	object	List of (altitude, latitude, longitude, speed)
mLight	m_light	float	Ambient light in lx unit
mScreenStatus	m_screen_use	integer	0: No, 1: Using screen
mUsageStats	m_usage_stats	object	List of app names and their respective usage times (in milliseconds unit)
mWifi	m_wifi	object	List of base station ID(bssid) and rssi
wHr	heart_rate	object	List of heart rate recordings
wLight	w_light	float	Ambient light in lx unit
wPedo	burned_calories	float	Number of calories
	distance	float	Distance in meters
	speed	float	Speed in km/h unit
	step	integer	Number of steps
	step_frequency	float	Step frequency in a minute

B. Data Preprocessing

The collected lifelog data was partitioned into training and testing sets based on subject-date pairs to ensure temporal consistency and prevent data leakage. Feature engineering was performed for each sensor modality to extract meaningful characteristics. Mobile device data yielded features including charging patterns, physical activity proportions, ambient sound/light measurements, Bluetooth/WiFi connectivity metrics, GPS mobility patterns, and screen/application usage statistics. Wearable device data was processed to generate heart rate statistics, light measurements, and pedometer metrics, segmented into time blocks (early morning, morning, afternoon, evening). All sensor-specific features were merged into a unified dataset using subject ID and date as keys through outer joins. Missing values were imputed with zeros to reflect sensor unavailability, and variable names were standardized for machine learning algorithm compatibility.

*The preprocessing methodology in this paper partially referenced the "Baseline" code shared by code7monkey on the DACON community forum (posted April 25, 2025) [5].

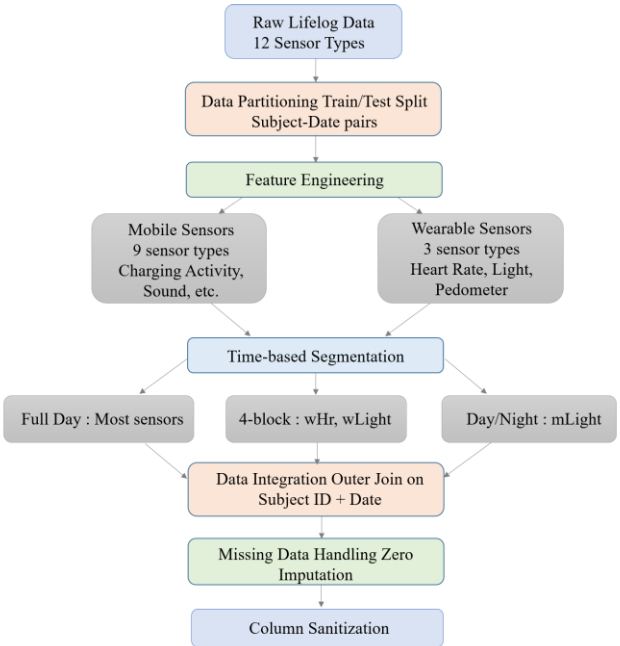

Fig 1 Preprocessing flow

C. feature Selection

To optimize model performance and reduce computational complexity, feature selection was conducted using feature importance analysis from preliminary machine learning models trained on the comprehensive feature set. A subset of 20 high-impact features was systematically identified based on their relative importance scores derived from gradient boosting algorithms. As tree-based algorithms, gradient boosting methods automatically learn non-linear relationships and interactions between features, enabling automatic discovery of non-linear patterns in data without requiring domain expertise while effectively identifying the most predictive features for sleep quality assessment.

The selection process prioritized features that demonstrated the highest predictive power for sleep quality classification (S1), encompassing six distinct categories: digital behavior patterns including screen usage metrics (screen_on_ratio, screen_on_duration_avg, screen_on_duration_max) and application usage patterns (message_time, others_time, Narration_monologue), connectivity indicators reflecting environmental stability through WiFi and Bluetooth signal characteristics (wifi_rssi_max, wifi_rssi_mean, wifidetected_cnt, rssi_mean), physical activity levels captured through movement state proportions (activity_3_ratio, activity_4_ratio), device charging behaviors serving as proxies for usage intensity and routine consistency (charging_ratio, max_charging_duration, avg_charging_duration), environmental light exposure patterns critical for circadian rhythm assessment (light_night_mean, light_max, light_std), and mobility characteristics indicating lifestyle regularity and spatial patterns (altitude_mean, speed_max_x). This feature importance-driven selection methodology ensures that the final feature set captures the most discriminative behavioral and physiological patterns while maintaining model interpretability and computational efficiency for practical sleep quality assessment applications.

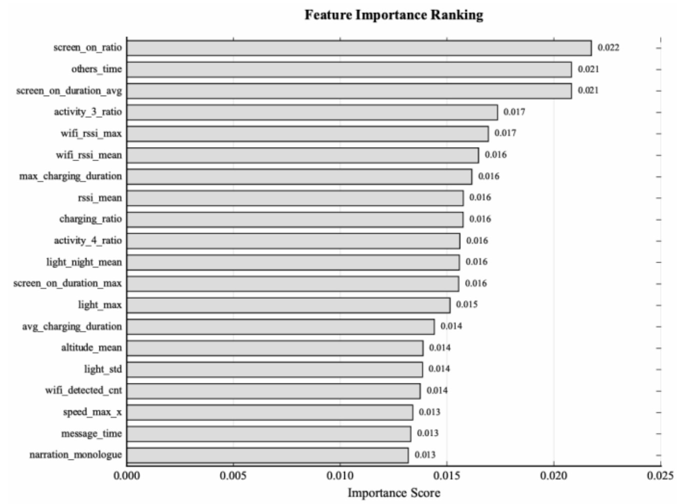

Fig 2 Feature importance

D. Autoencoder

Feature importance analysis was conducted to identify the top 20 most predictive features for sleep quality classification, encompassing digital behavior patterns (screen usage ratios and durations), connectivity metrics (WiFi and Bluetooth signal strengths), physical activity indicators (stillness and unknown activity ratios), device charging behaviors, environmental light exposure characteristics, and mobility patterns. To enhance these selected features, an autoencoder-based feature expansion approach was implemented utilizing a deep neural network architecture. The autoencoder consisted of an encoder pathway (20→128→64→20 dimensions with ReLU activation, batch normalization, and 0.3 dropout) and a decoder pathway (20→64→128→20 dimensions with sigmoid output activation), trained with an early stopping mechanism to prevent overfitting. The core methodology involved using the autoencoder to learn non-linear and compressed relationships among the 20 selected variables, generating 20 new transformed features that capture complex hidden patterns within the data. These autoencoder-derived features were then concatenated with the original 20 selected features to create an expanded 40-dimensional feature space. This feature fusion approach combines two complementary perspectives: the original variables provide direct and interpretable behavioral patterns, while the autoencoder-generated variables capture complex non-linear relationships and latent structures within the lifelog data. The autoencoder's hierarchical feature extraction capabilities enable automatic discovery of meaningful patterns without requiring domain expertise, effectively representing both explicit domain knowledge and hidden multi-dimensional relationships. This dual representation strategy, as demonstrated in multi-sensor human activity recognition research [6], enhances the model's ability to learn from enriched feature representations, potentially improving classification performance for sleep quality assessment.

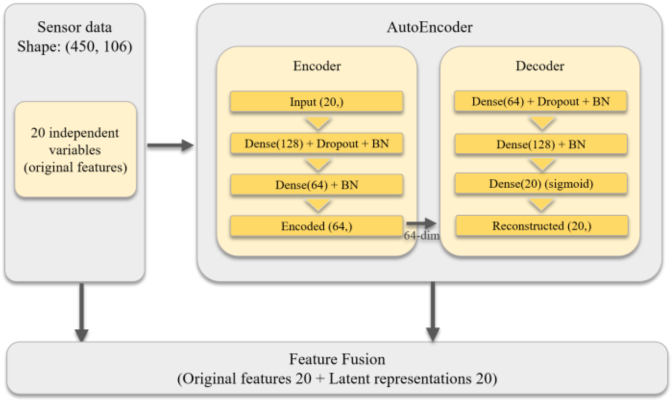

Fig 3 Autoencoder structure

E. LGBM Classifier

Following the autoencoder-based feature enhancement, LightGBM classifiers were employed for sleep health prediction tasks. LightGBM was selected due to its effectiveness in handling diverse feature types and robust processing of missing values. The classification framework consisted of five binary classification tasks (Q1, Q2, Q3, S2, S3) and one multiclass classification task (S1) with three categories. Hyperparameter optimization was performed using GridSearchCV with 3-fold stratified cross-validation. The parameter search included learning rates (0.01, 0.03), number of estimators (500, 1000), number of leaves (50, 100), maximum depth (-1, 5), and L1/L2 regularization parameters (0, 0.01, 0.1).

To evaluate the effectiveness of the proposed approach, comparative analysis was conducted against baseline methods. To evaluate the effectiveness of the proposed approach, comparative analysis was conducted against baseline methods. The performance results demonstrated that LSTM achieved a score of 0.4158, TabNet 0.4648, Random Forest 0.5744, and standalone LGBM 0.5822, while the proposed Autoencoder + LGBM framework achieved the highest public score of 0.6069. The hybrid approach achieved the highest public score, demonstrating superior performance over individual methods. The final model integrated the 40-dimensional enhanced feature space (20 original features + 20 autoencoder-generated features) with optimized LightGBM classifiers. Early stopping mechanisms and validation monitoring were implemented to prevent overfitting and ensure reproducible results. This approach effectively balanced model complexity and interpretability for practical sleep quality assessment applications.

TABLE II. Feature Selected Models

Model	**Public Score**
LSTM	0.4158
Tabnet	0.4648
Randomforest	0.5744
LGBM	0.5822
Autoencoder + LGBM	0.6069

III. Conclusion

This study presents a novel hybrid autoencoder-LightGBM framework for predicting sleep health metrics from multimodal lifelog data. The proposed approach successfully addresses the challenge of extracting meaningful patterns from heterogeneous sensor data by combining feature importance-driven selection with deep learning-based feature enhancement.

The methodology demonstrates several key contributions: First, systematic feature selection using gradient boosting algorithms identified 20 high-impact features across six behavioral categories, including digital behavior patterns, connectivity indicators, physical activity levels, charging behaviors, environmental light exposure, and mobility characteristics. Second, the autoencoder-based feature expansion technique effectively captured non-linear relationships and latent structures within the data, generating complementary representations that enhanced the original feature space from 20 to 40 dimensions. Third, the integration of interpretable domain features with learned hidden patterns provided a balanced approach between model performance and explainability. Experimental evaluation on the DACON 2025 Lifelog Challenge demonstrated the effectiveness of the proposed framework, achieving a public score of 0.6069, which outperformed baseline approaches including LSTM (0.4158), TabNet (0.4648), Random Forest (0.5744), and standalone LGBM (0.5822). The superior performance validates the effectiveness of combining unsupervised feature learning with optimized gradient boosting for sleep health prediction.

The findings suggest that lifestyle behavioral patterns captured through multimodal sensor data can serve as reliable indicators for sleep quality assessment. This work contributes to the growing field of digital health monitoring by providing a practical framework that leverages ubiquitous mobile and wearable devices for continuous sleep health evaluation, with potential applications in personalized healthcare and preventive medicine.

IV. Constraints and Limitations

The proposed hybrid autoencoder-LightGBM framework has several important limitations that should be acknowledged. The naturalistic data collection approach introduces challenges including substantial missing data, measurement noise, and irregular sampling patterns due to device charging or inconsistent wearing by participants. The evaluation conducted on the ETRI 2025 lifelog dataset with 10 participants may limit generalizability to broader populations, suggesting the need for validation across more diverse demographic groups and larger sample sizes. The framework's reliance on multiple sensor modalities presents practical considerations for deployment, as system performance may vary when certain sensors are unavailable or experiencing technical issues in real-world scenarios. Additionally, the current approach employs population-level modeling, indicating potential benefits from incorporating personalized adaptation mechanisms in future work. The comprehensive collection of behavioral and physiological data through this framework warrants careful consideration of privacy and ethical implications. Continuous monitoring of personal activities, location patterns, device usage, and health metrics requires thoughtful approaches to data security, informed consent, and responsible data management. Future research and implementation efforts should incorporate appropriate privacy-preserving mechanisms, transparent data governance frameworks, and user-controlled data sharing protocols to ensure ethical deployment while maintaining the system's analytical capabilities. These considerations represent important areas for continued development rather than insurmountable barriers to the technology's potential benefits.

ACKNOWLEDGMENT

The authors acknowledge the ETRI (Electronics and Telecommunications Research Institute) for providing the lifelog dataset used in this research. We also thank the DACON platform for organizing the 2025 lifelog challenge and facilitating access to this valuable resource for sleep health research.

REFERENCES

[1] Song, T. A., Chowdhury, S. R., Malekzadeh, M., Harrison, S., Hoge, T. B., Redline, S., ... & Dutta, J. (2023). AI-Driven sleep staging from actigraphy and heart rate. *PLOS One*, 18(5), e0285703

[2] Gupta, R., Bhongade, A., & Gandhi, T. K. (2023). Multimodal wearable sensors-based stress and affective states prediction model. *2023 9th International Conference on Advanced Computing and Communication Systems (ICACCS)*, 1, 30-35

[3] Yuan, H., Plekhanova, T., Walmsley, R., Reynolds, A. C., Maddison, K. J., Bucan, M., ... & Doherty, A. (2024). Self-supervised learning of accelerometer data provides new insights for sleep and its association with mortality. npj Digital Medicine, 7(1), 86

[4] 250430_ch2025_data_items_description.pdf

[5] https://dacon.io/competitions/official/236468/codeshare/12319?page=1&dtype=recent

[6] Arabzadeh, Saeed, Farshad Almasganj, and Mohammad Mahdi Ahmadi. "CNN Autoencoders for Hierarchical Feature Extraction and Fusion in Multi-sensor Human Activity Recognition." *arXiv preprint arXiv:2502.04489* (2025).