scientific reports

OPEN

Handcrafted MRI radiomics of enlarged perivascular spaces and machine learning predict cognitive impairment and sleep disturbance in young adults

Li Li[1,6], Jiaojiao Wu[2,6], Bin Li[3], Rui Hua[2], Feng Shi[2], Lizhou Chen[4✉] & Yeke Wu[5✉]

Our study seeks to develop a predictive model using MRI-derived enlarged perivascular spaces (EPVSs) measurements and machine learning to assess cognitive impairment, subjective sleep quality, and excessive daytime sleepiness in young adults with long-time mobile phone use (LTMPU). We enrolled 82 participants and employed a pretrained deep learning model (VB-Net) to automatically segment EPVSs lesions across 17 brain subregions, extracting four handcrafted radiomic features – predefined based on morphological properties – per subregion (EPVSs count, volume, mean length, and mean curvature). The cohort was randomly divided into training (80%) and testing (20%) sets. Through minimum redundancy maximum relevance (mRMR) feature selection, six key biomarkers from 68 initial EPVSs metrics were identified combined with sex and age covariates. Final models were constructed using a Gaussian process (GP) classifier for cognitive impairment and decision tree (DT) algorithms for sleep quality and excessive sleepiness assessment. In testing, the GP model achieved an AUC of 0.818 (95% confidence interval [CI] 0.610-1) for cognitive impairment prediction. The DT models showed AUCs of 0.826 (95% CI: 0.616-1) for sleep quality and 0.875 (95% CI: 0.718-1) for daytime sleepiness. This automated radiomics pipeline demonstrates EPVSs morphological features as potential biomarkers for evaluating mobile phone exposure-related neurocognitive dysfunction. This automated radiomics pipeline suggests that EPVSs morphological features might be beneficial for evaluating mobile phone exposure-related neurocognitive dysfunction.

Keywords Enlarged perivascular spaces, Magnetic resonance imaging, Machine learning, Cognitive impairment, Sleep disturbance

Mobile phones have become ubiquitous in contemporary society, offering unparalleled convenience while simultaneously sparking health-related inquiries to people's lives. Approximately 80% of internet users engaged various social media platforms through their mobile phones with young adults aged 18–29, the predominant demographic, spending an average of 181 min daily to social media interactions[1]. Long-time mobile phone use (LTMPU), defined as engaging with a mobile device ≥ 4 h/day is consistently linked to sleep disturbances and mental distress[2]. A recent systematic review published in 2023 has uncovered a significant association between diminished sleep quality and mobile phone usage patterns[3]. Recent evidence also underscores a robust correlation between mobile phone dependency and excessive daytime sleepiness[4]. while emerging evidence suggests that the overuse of mobile phone is related with cognitive impairments[5]. Notably, sleep disturbances are a well-recognized risk factor for dementia, which is a condition with no curative treatments currently available[6], thus preventive measures assume particular importance. Given the high prevalence and potential severity of

[1]Department of Radiology, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610072, Sichuan, China. [2]Department of Research and Development, United Imaging Intelligence, Shanghai 200232, China. [3]Department of Geriatrics, Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610072, Sichuan, China. [4]Department of Radiology, West China Hospital of Sichuan University, No. 37, Guoxue Alley, Chengdu 610041, Sichuan, China. [5]Department of Stomatology, Hospital of Chengdu University of Traditional Chinese Medicine, No. 39, Shierqiao Rd, Chengdu 610072, Sichuan, China. [6]Li Li and Jiaojiao Wu contributed equally to this work. ✉email: 357850956@qq.com; wuyeke@cdutcm.edu.cn

1

these associated risks, there is a compelling public health rationale for deeply studying this specific, high-risk population. This study therefore focuses specifically on young adults with LTMPU as a critical first step to characterize this ubiquitous at-risk group and identify potential biomarkers for early intervention.

Dementia, which is characterized as impaired memory, language, problem-solving abilities, and an overall decline in cognitive function, stands as a prevalent cause of disability and mortality among the elderly[7]. As China progressively transitions into an aging society, with the elderly population reaching 14% of the total populace and projected to escalate to 22% by 2033[8], the imperative for early detection and intervention in dementia thus becomes increasingly critical. Given the absence of a curative treatment, the emphasis is on the early identification of cognitive decline, underpinned by the recognition that sleep is integral to cognitive performance[9]. Suboptimal sleep quality is increasingly linked to cognitive deficits[10], while excessive daytime sleepiness is recognized to exert adverse effects on cognitive functions[11]. Consequently, precise evaluation of cognitive impairment, sleep quality, and the severity of daytime sleepiness symptoms is essential for devising preventive strategies aimed at pinpointing modifiable risk factors, thereby potentially decelerating or mitigating the advancement of dementia.

In clinical practice, the Montreal Cognitive Assessment (MoCA)[12] is commonly utilized to assess the severity of cognitive impairment. The Epworth Sleepiness Scale (ESS) is employed to gauge the severity of excessive daytime sleepiness symptoms[13], while the subjective sleep quality can be assessed by either Pittsburg Sleep Quality Index (PSQI) or the Insomnia severity index (ISI)[14]. However, a significant drawback of these assessment tools is their reliance on subjective reports, which may lack precision, accuracy, and reliability in certain contexts[15,16]. This shortfall has intensified the quest for more objective and precise diagnostic techniques to evaluate the severity of cognitive impairment, subjective sleep quality, and daytime sleepiness symptoms.

Growing evidence indicates that perivascular spaces (PVSs) dysfunction is associated with the pathogenesis of sleep disturbances and impairments in cognitive function. PVSs, which are fluid-filled cavities encircling penetrating cerebral arterioles and venules, are hypothesized to facilitate a drainage network crucial for the clearance of metabolic byproducts and cerebrospinal fluid from the brain[17,18], particularly during sleep[19]. Some indirect evidence suggests that increased visibility of PVSs is linked to obstructive sleep apnoea and with reduced sleep efficiency, indicating PVSs dysfunction during disrupted human sleep[20,21]. A review in 2023 further highlights sleep's potential key role in PVSs(the glymphatic system) function. It also indicates that neuroinflammatory conditions, neurodegenerative diseases, and cognitive dysfunction may be associated with underlying glymphatic system dysfunction, suggesting sleep disorders could be a target for intervention[22]. The clearance of cellular byproducts, such as amyloid-βand tau, which are related to neurodegenerative processes and cognitive impairment, depends in part on intact glymphatic function, with PVSs serving as a key component[23,24].

Machine learning-based radiomics, encompassing the development of predictive models and the discovery of meaningful patterns within datasets through computational techniques, offers an objective approach to data analysis[25]. MRI presents significant potential for elucidating the nexus between sleep disorders and the risk of dementia in vivo[26]. Enlarged perivascular spaces (EPVSs), detectable through MRI (MRI-visible EPVSs) as indicators of EPVSs dysfunction[19]. EPVSs are related to cognitive function, vascular risk factors, vascular and neurodegenerative brain lesions, sleep patterns and cerebral haemodynamics[19]. A substantial body of evidence implicates a negative correlation between EPVSs and cognitive functions, as well as sleep processes. A important population-based study has established that the presence of EPVSs in the basal ganglia and white matter correlates with a notably elevated risk of developing dementia[27]. Although research on the association between EPVSs and excessive daytime sleepiness is nascent, indirect evidence points to EPVSs dysfunction during sleep disruption[19], which is intricately associated with sleep quality. Previous studies have indicated morphological changes, characterized by the enlargement of basal ganglia-EPVSs, in individuals experiencing persistent poor sleep quality following coronavirus disease[28]. The quantification of EPVSs is increasingly being addressed through automated methodologies[18], which may enhance the precision and objectivity of assessment in this research domain.

This study aims to develop a predictive model utilizing MRI-quantified EPVSs metrics and machine learning to evaluate the severity of cognitive impairment, self-reported subjective sleep quality, and the intensity of excessive daytime sleepiness symptoms in young adults with LTMPU. Through this innovative methodology, we hope to elucidate the potential correlation between EPVSs and cognitive, sleep quality, and excessive daytime sleepiness in individuals addicted to mobile phone use. This research may pave the way for more precise, objective assessments and could potentially inform preventive strategies and interventions in this demographic.

Results
Participants characteristics
We recruited 82 participants who underwent MRI examinations from the Affiliated Hospital of Chengdu University of Traditional Chinese Medicine between October 2021 and May 2022 (Fig. S1). The demographics and clinical scales of each participant were collected and presented in Table 1. The median age of all participants is 38.0 years (Fig. 1a), with 29.3% (24/82) being male. The distribution of cognitive impairment, poor sleep quality, insomnia, and sleepiness among all participants is visualized in Fig. 1b, with occurrences of 55, 54, 36, and 40, respectively (Fig. S2). It demonstrates that one participant suffers from multiple disorders at the same time, revealing an intrinsic correlation among them. The demographics of participants in each disorder are summarized in Table 1. Notably, the median age of the cognitive normalization group (MoCA ≥ 26) is 33.0 years, whereas the median age of the cognitive impairment group (MoCA < 26) is 40.0 years, which is a significant difference between the two ages. There are no significant differences in sex distribution between the cognitive normalization and cognitive impairment groups, between the good sleep (PSQI ≤ 5) and poor sleep (PSQI > 5) groups, between the non-insomnia (ISI ≤ 7) and insomnia (ISI > 7) groups, and between the non-sleepiness (ESS ≤ 6) and sleepiness (ESS > 6) groups. The range, mean, and standard deviation of scores for all clinical scales are summarized in Table S1.

Variables	Label = 0	Label = 1	p
MoCA	Cognitive normalization (MoCA ≥ 26)	Cognitive impairment (MoCA < 26)	
No. of participants	27	55	–
MoCA score	Range: 26–30 Mean ± SD: 27.2 ± 1.3	Range: 13–25 Mean ± SD: 22.1 ± 2.9	< 0.001
Age (median, years)	33.0 (26.0, 39.0)	40.0 (36.0, 44.0)	< 0.001
Sex (male, %)	7 (25.9%)	17 (30.9%)	0.642
Education (median, years)	16.0 (16.0, 19.5)	15.0 (12.0, 16.0)	< 0.001
PSQI	Good sleep (PSQI ≤ 5)	Poor sleep (PSQI > 5)	
No. of participants	28	54	–
PSQI score	Range: 1–5 Mean ± SD: 3.7 ± 1.3	Range: 6–21 Mean ± SD: 12.3 ± 4.1	< 0.001
Age (median, years)	36.0 (29.0, 43.5)	39.0 (33.0, 43.0)	0.358
Sex (male, %)	6 (21.4%)	18 (33.3%)	0.261
ISI	Non-insomnia (ISI ≤ 7)	Insomnia (ISI > 7)	
No. of participants	46	36	–
ISI score	Range: 0–7 Mean ± SD: 2.3 ± 2.2	Range: 8–21 Mean ± SD: 14.0 ± 3.8	< 0.001
Age (median, years)	36.0 (30.5, 42.0)	40.0 (34.0, 44.0)	0.120
Sex (male, %)	15 (32.6%)	9 (25.0%)	0.452
ESS	Non-sleepiness (ESS ≤ 6)	Sleepiness (ESS > 6)	
No. of participants	42	40	–
ESS score	Range: 0–6 Mean ± SD: 4.1 ± 1.6	Range: 7–24 Mean ± SD: 10.4 ± 3.6	< 0.001
Age (median, years)	39.0 (34.0, 43.0)	36.0 (31.0, 43.5)	0.215
Sex (male, %)	13 (30.9%)	11 (27.5%)	0.343

Table 1. Demographics of participants. Continuous variables were compared using Mann-Whitney U tests; sex distribution was compared using the chi-square test. A two-tailed p-value < 0.05 was considered statistically significant.
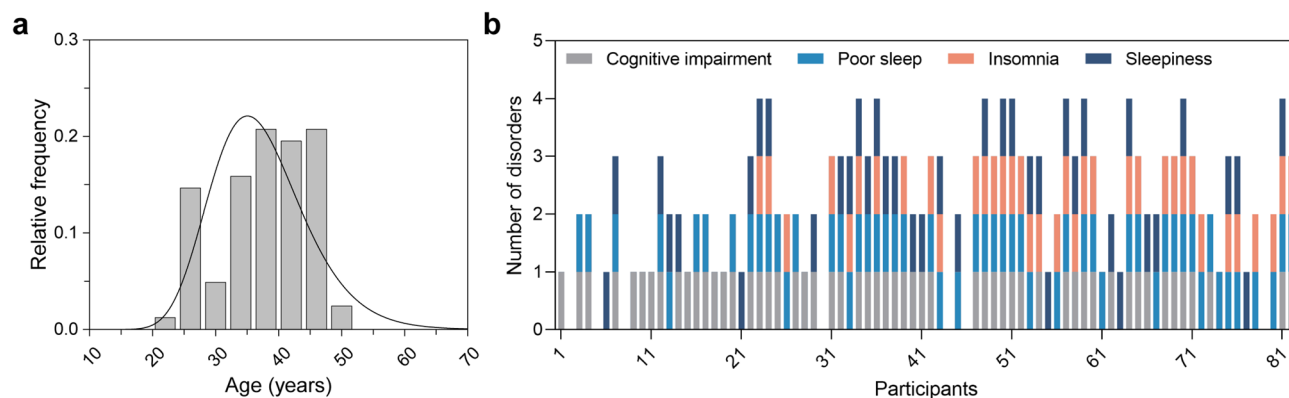

Fig. 1. Characteristics of the participants. (**a**) Age distribution of all participants. (**b**) Distribution of four disorders in all participants.

Gaussian process model in predicting cognitive impairment

To explore the ability of EPVSs features to predict cognitive impairment severity, subjective sleep quality, and excessive daytime sleepiness symptoms severity in young adults with LTMPU, machine learning-based radiomics analyses are conducted. A total of 70 EPVSs features combined with easily accessible participant demographics (i.e., sex, age) are used as inputs to select the most valuable features to construct the machine learning model.

The cognitive function can be classified into two categories based on MoCA scores, with MoCA ≥ 26 being the cognitive normalization group and MoCA < 26 being the cognitive impairment group. To identify participants with cognitive impairment, the mRMR method is used to select the six most valuable features (Fig. 2a), whose correlation matrix is shown in Fig. 2b. It is clear that three pairs of features are significantly correlated. The

distributions of all six features in the training (Fig. S3a–f) and testing datasets (Fig. S3g–l) are visualized, with statistical comparisons between cognitive normalization and impairment groups. Key features show varying effect sizes (Table S2): age exhibits a large effect (d = 0.932, 95% CI 0.442–1.421), while the average length of EPVSs in left centrum semiovale shows a small effect (d = −0.442, 95% CI −0.914–0.031) in distinguishing the two groups. Subsequently, a classification model is constructed using the Gaussian process (GP) algorithm. Since early detection of cognitive impairment relies on high sensitivity to reduce false negatives, we set the classification threshold using Youden index from the training dataset to balance sensitivity and specificity. The receiver operating characteristic (ROC) curves are plotted in Fig. 2c. Specifically, the area under the ROC curve (AUC) values of the GP model are 0.949 with a 95% confidence interval (CI) of 0.900–0.998.900.998 and 0.818 (95% CI 0.610–1.000.610.000) in the training and testing datasets, respectively. In the testing dataset, the model achieves a sensitivity of 0.727, specificity of 0.667, and accuracy of 0.706 (Table 2), indicating it can identify 72.7% of true cognitive impairment cases – critical for early intervention. The calibration curves show that the positive incidence predicted by the GP model deviates somewhat from the actual incidence of cognitive impairment in the training and testing datasets, indicating that the accuracy of the model prediction needs to be further improved (Fig. 2d). Nonetheless, the model is still able to achieve a net clinical benefit within a threshold range of 0.3 to 0.9 (Fig. 2e). The confusion matrix of the training dataset (Fig. 2f) and its corresponding performance metrics (Fig. 2g) demonstrate that all metrics, including sensitivity (0.864) and specificity (0.905), are above 0.80, reflecting robust performance in the training phase. In the training dataset, 6 out of 44 true cognitive impairment cases were misclassified as normal (false negatives), and 2 out of 21 true normal cases were misclassified as impaired (false positives). For the testing dataset, the confusion matrix (Fig. 2h) and metrics (Fig. 2i) confirm the model's ability to maintain a reasonable balance between sensitivity and specificity. In the testing dataset, 3 out of 11 true cognitive impairment cases are false negatives, and 2 out of 6 true normal cases are false positives. There are no significant differences between false negatives and correctly classified cognitive impairment cases in terms of age or key EPVSs features ($p > 0.05$ for all comparisons), and the small number of misclassified cases limits our ability to detect subtle systematic patterns. To interpret model predictions, we conducted SHAP analyses: (1) A summary plot (training dataset) shows average length of EPVSs in the left basal ganglia, average curvature of EPVSs in the left centrum semiovale, age, and volume of EPVSs in the left frontal lobe as the top features (Fig. 2j); (2) A SHAP heatmap (testing dataset) demonstrates feature contributions across all testing samples (Fig. 2k); (3) A force plot (testing dataset, representative sample) illustrates how individual features push the prediction toward "cognitive impairment" (Fig. 2l).

Decision tree model in predicting subjective sleep quality (PSQI)

Radiomics analysis is also used to categorize participants with poor subjective sleep quality (PSQI > 5) and good subjective sleep quality (PSQI ≤ 5). Six features are chosen through the mRMR method (Fig. 3a), and the correlation matrix, displayed in Fig. 3b, reveals significant correlations between five pairs of features. The distribution of the six features used to construct the machine learning model in the training and testing datasets is shown in Fig. S4. Statistical comparisons reveal small or negligible effect sizes for all PSQI-related features (Table S2), with the largest effect observed for the average curvature of EPVSs in the left centrum semiovale (d = −0.399, 95% CI −0.867–0.068). A decision tree (DT) algorithm is employed to build a classification model, with ROC curves depicted in Fig. 3c. The DT model exhibits AUC values of 0.865 (95% CI 0.770–0.959) and 0.826 (95% CI 0.616–1.000) in the training and testing datasets, respectively. The calibration curve demonstrates strong agreement between the predicted positive incidence rates by the DT model and the actual incidence rates of poor subjective sleep quality in the training dataset, with slight deviation observed in the testing dataset (Fig. 3d). Overall, the model yields high prediction accuracy rates of 0.846 and 0.824 in the two datasets, as shown in Table 2. Furthermore, decision curves illustrated in Fig. 3e indicate that the model delivers substantial clinical net benefits across a range of thresholds (0.2–1.0.2.0) in both training and testing datasets, showcasing its potential for enhancing patient care and decision-making support effectively. Detailed quantitative metrics from confusion matrices visualized in Fig. 3f,g for training data and Fig. 3h,i for testing data are presented in Table 2. In the training dataset, 4 out of 43 poor sleep quality cases are misclassified as false negatives, and 6 out of 22 good sleep quality cases are misclassified as false positives. In the testing dataset, 1 out of 11 poor sleep quality cases are false negatives, and 2 out of 6 good sleep quality cases are false positives. Overall, except for specificity, all metrics surpass a threshold of above 0.8 in both datasets, indicating a relatively high prediction performance by the model despite some false positives being present. SHAP analyses further clarify feature contributions for this model (Fig. 3j–l): the average curvature of EPVSs in the left centrum semiovale is the most influential feature, and a force plot for a representative testing sample visualizes how EPVSs metrics drive the prediction of "poor sleep".

Gaussian process model in predicting subjective sleep quality (ISI)

Similar feature selection and modeling procedures are conducted to distinguish between the non-insomnia group (ISI ≤ 7) and the insomnia group (ISI > 7). Using the mRMR method, a total of six features are selected (Fig. 4a), with the corresponding correlation matrix presented in Fig. 4b, indicating no significant correlation among the selected features, underscoring their unique and independent information contribution. The distribution of the six features used to construct the machine learning model in the training and testing datasets is shown in Fig. S5. Subsequently, a classification model is built using the GP algorithm, using the Youden index of the training dataset as the classification threshold. The AUC values for the GP model are calculated as 0.947 (95% CI 0.888–1.000.888.000) in the training dataset and 0.757 (95% CI 0.492–1.000.492.000) in the testing dataset, as shown in Fig. 4c. Notably, the wide CI in the testing dataset likely reflects the small sample size ($n = 17$) and underscores the need for cautious interpretation of this model's generalizability. Although calibration curves show narrower prediction intervals and acceptable accuracy (Fig. 4d), decision curves reveal
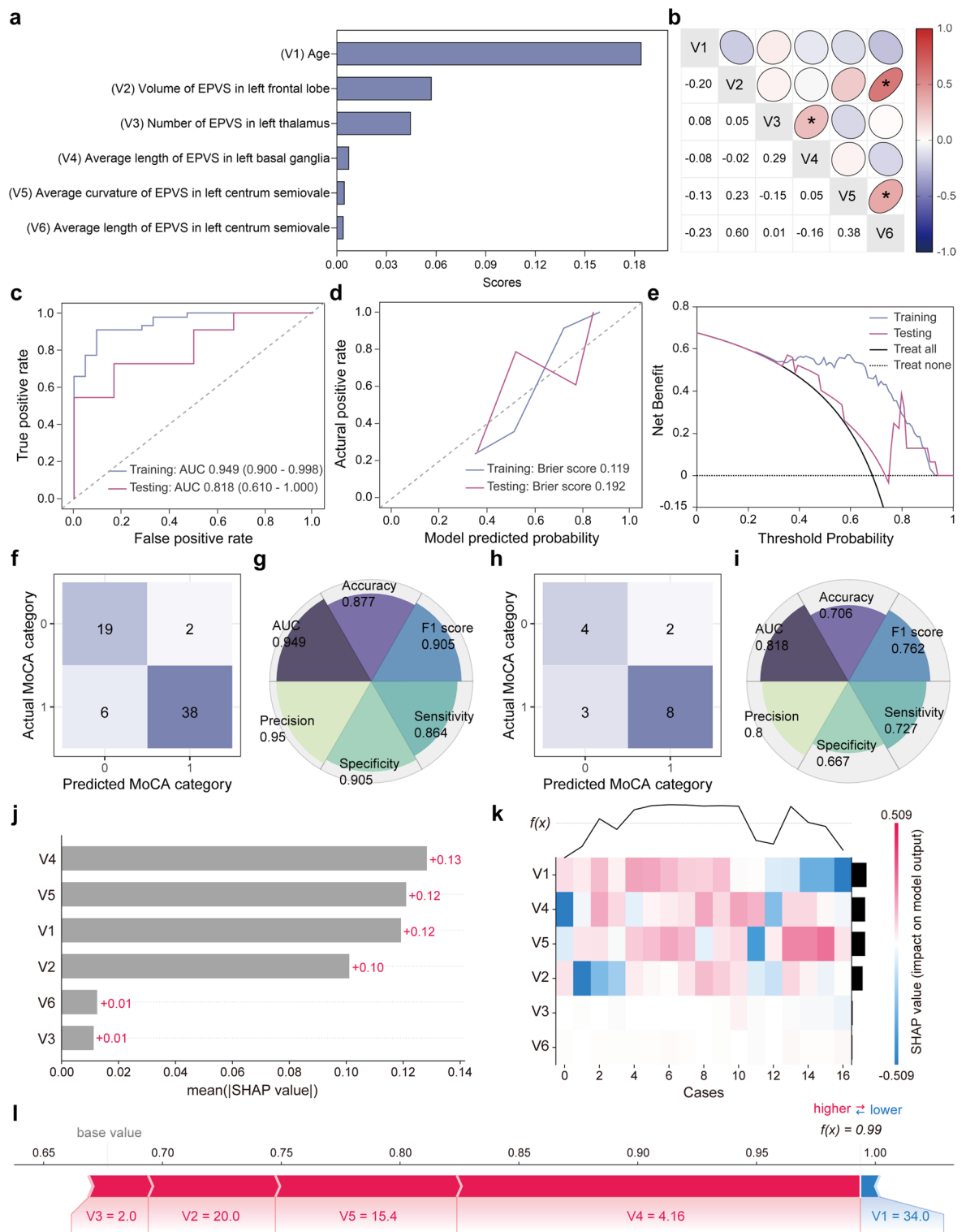
Fig. 2. Model construction and evaluation in identifying cognitive impairment. (**a**) Six features selected by the mRMR method. (**b**) Correlation heatmap of selected features. Pearson or Spearman correlation analyses were performed and * indicates $p < 0.05$. Numbers labeled in the plots represent correlation coefficients. (**c**) ROC curves evaluating the trade-off between sensitivity and specificity of the GP model, with a higher AUC indicating a better discrimination ability of the model across different threshold settings. (**d**) Calibration curves evaluating the consistency of predicted probability and the actual cognitive impairment rate. (**e**) Decision curves showing the clinical net benefit. (**f**) Confusion matrix of the training dataset. The "0" and "1" represent cognitive normalization and impairment, respectively. (**g**) Metrics of the training dataset. (**h**) Confusion matrix and (**i**) metrics of the testing dataset. (**j**) SHAP summary plot (training dataset) showing feature importance. (**k**) SHAP heatmap (testing dataset) visualizing feature contributions across all testing samples. (**l**) SHAP force plot (testing dataset, representative sample) illustrating individual prediction drivers.

	AUC (95% CI)	Accuracy	F1 score	Sensitivity	Specificity	Precision
GP for predicting cognitive impairment (MoCA < 26 vs. MoCA ≥ 26)						
Training	0.949 (0.900–0.998)	0.877	0.905	0.864	0.905	0.950
Testing	0.818 (0.610–1.000)	0.706	0.762	0.727	0.667	0.800
DT for predicting subjective sleep quality (PSQI > 5 vs. PSQI ≤ 5)						
Training	0.865 (0.770–0.959)	0.846	0.886	0.907	0.727	0.867
Testing	0.826 (0.616–1.000)	0.824	0.870	0.909	0.667	0.833
GP for predicting subjective sleep quality (ISI > 7 vs. ISI ≤ 7)						
Training	0.947 (0.888–1.000)	0.862	0.830	0.759	0.944	0.917
Testing	0.757 (0.492–1.000)	0.765	0.714	0.713	0.800	0.714
DT for predicting excessive daytime sleepiness symptoms (ESS > 6 vs. ESS ≤ 6)						
Training	0.923 (0.867–0.978)	0.877	0.857	0.750	1.000	1.000
Testing	0.875 (0.718–1.000)	0.824	0.800	0.750	0.889	0.857

Table 2. Performance of four machine learning models in predicting cognitive and sleep disorders.

a clinical net benefit in both datasets, particularly within a threshold range of 0.3–0.6, defining a practical clinical utility window where the model adds value beyond default strategies (e.g., treating all or no patients) (Fig. 4e). This range thus serves as a preliminary threshold for potential clinical application, pending validation in larger cohorts. Performance metrics—including accuracy, F1-score, sensitivity, specificity, and precision—are derived from the training dataset's confusion matrix (Fig. 4f) and displayed in Fig. 4g, and further validated by the testing dataset's confusion matrix (Fig. 4h) and corresponding metrics (Fig. 4i). In the training dataset, 7 out of 29 insomnia cases are misclassified as false negatives, and 2 out of 36 non-insomnia cases are misclassified as false positives. In the testing dataset, 2 out of 7 insomnia cases are false negatives, and 2 out of 10 non-insomnia cases are false positives. In testing, the model achieves a sensitivity of 0.713, specificity of 0.800, and accuracy of 0.765 (Table 2), reflecting its ability to correctly identify 71.3% of true insomnia cases, which is important for early detection, while maintaining a reasonable level of specificity to avoid excessive false positives. SHAP analyses reveal key drivers of this model (Fig. 4j–l): the average length of EPVSs in the left basal ganglia, number of EPVSs, and average curvature of EPVSs in the right frontal lobe are considered as primary features, and a force plot for a representative sample demonstrates how EPVSs metrics influence the insomnia prediction.

Decision tree model in predicting excessive daytime sleepiness symptoms

Feature selection and modeling procedures are applied to differentiate between the non-sleepiness group (ESS ≤ 6) and the sleepiness group (ESS > 6). Six features are selected using the mRMR method, as illustrated in Fig. 5a, with the correlation matrix presented in Fig. 5b showing no significant correlations among these selected features. This absence of significant correlation highlights the unique and independent information contributed by each feature. The distributions of the six features used to construct the machine learning model in the training (Fig. S6a –f) and testing datasets (Fig. S6g –l~) are shown, with statistical comparisons between the non-sleepiness and sleepiness groups. Notably, a significant difference is observed in the average length of EPVSs lesions in the left centrum semiovale (Fig. S6h). The largest effect among ESS-related features is observed for the average length of EPVSs in the left centrum semiovale (d = 0.406, 95% CI −0.038–0.850; Table S2). Following this, a classification model is constructed using the DT algorithm. The AUC values for the DT model are determined as 0.923 (95% CI 0.867–0.978) in the training dataset and 0.875 (95% CI 0.718–1.000.718.000) in the testing dataset, as depicted in Fig. 5c. The calibration curves demonstrate a strong alignment between the model's predicted likelihood of sleepiness and the actual prevalence of sleepiness in both the training and testing datasets, as shown in Fig. 5d. This alignment underscores the model's ability to accurately estimate an individual's probability of belonging to the sleepiness group across different datasets, indicating its reliability in evaluating sleep disorders. Furthermore, decision curves showcase that the model provides clinical utility and benefit across a wide range of thresholds (0.1–0.8), suggesting its potential positive impact on clinical decision-making processes (Fig. 5e). Quantitative metrics such as accuracy, F1-score, sensitivity, specificity, and precision are computed based on confusion matrix of the training dataset (Fig. 5f), with its corresponding performance metrics shown in Fig. 5g. Similarly, the confusion matrix for the testing dataset (Fig. 5h) and its metrics (Fig. 5i) confirm the model's robustness. In the testing dataset, 2 out of 8 sleepiness cases are false negatives, and 1 out of 9 non-sleepiness cases are false positives. Overall performance evaluation reveals that except for sensitivity, all metrics exceed a threshold of 0.8 in both datasets, indicating a strong performance across various assessment criteria. SHAP analyses explain feature impacts for this model (Fig. 5j,l): a training dataset summary plot identifies average curvature of EPVSs in the left frontal lobe and right centrum semiovale as the top features, and a force plot for a representative sample illustrates how individual features drive the sleepiness prediction.

Discussion

To our knowledge, this paper is the first study that presents a novel approach to classify cognitive impairment severity, subjective sleep quality, and excessive daytime sleepiness symptoms severity in young adults with LTMPU by integrating MRI-based quantification of EPVSs and machine learning algorithms. Our model has exhibited accuracy in these classifications, presenting a promising path for non-invasive and objective
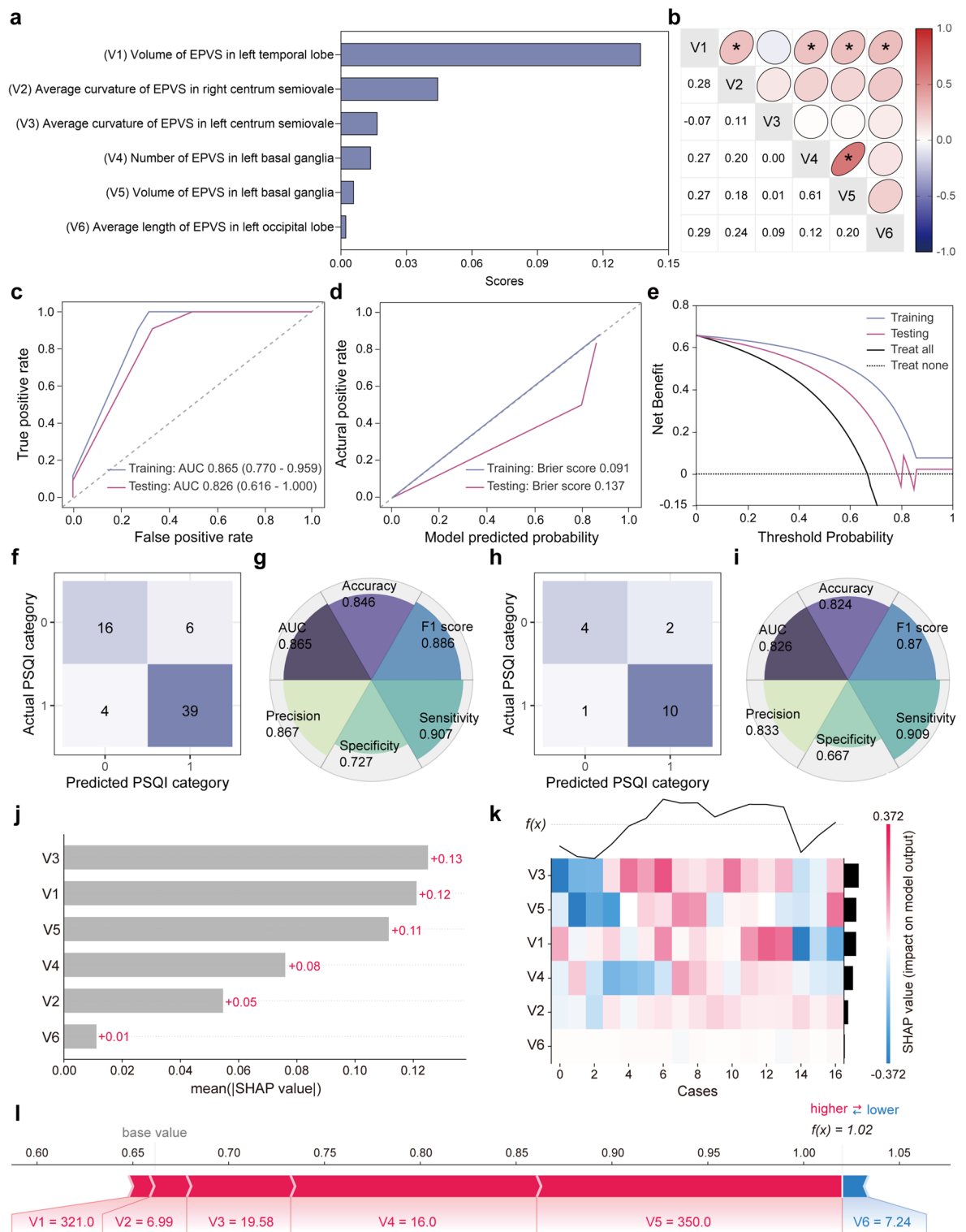
Fig. 3. Model construction and evaluation in classifying poor subjective sleep quality and good subjective sleep quality (PSQI). (**a**) Six features selected by the mRMR method. (**b**) Correlation heatmap of selected features. Pearson or Spearman correlation analyses were performed and * indicates $p < 0.05$. Numbers labeled in the plots represent correlation coefficients. (**c**) ROC curves evaluating the trade-off between sensitivity and specificity of the DT model. (**d**) Calibration curves evaluating the consistency of predicted probability and the actual poor subjective sleep quality rate. (**e**) Decision curves showing the clinical net benefit. (**f**) Confusion matrix of the training dataset. The "0" represents good sleep and "1" represents poor sleep. (**g**) Six metrics of the training dataset. (**h**) Confusion matrix of the testing dataset. (**i**) Six metrics of the testing dataset. (**j**) SHAP summary plot (training dataset) showing feature importance. (**k**) SHAP heatmap (testing dataset) visualizing feature contributions across all testing samples. (**l**) SHAP force plot (testing dataset, representative sample) illustrating how individual features influence the prediction outcome.
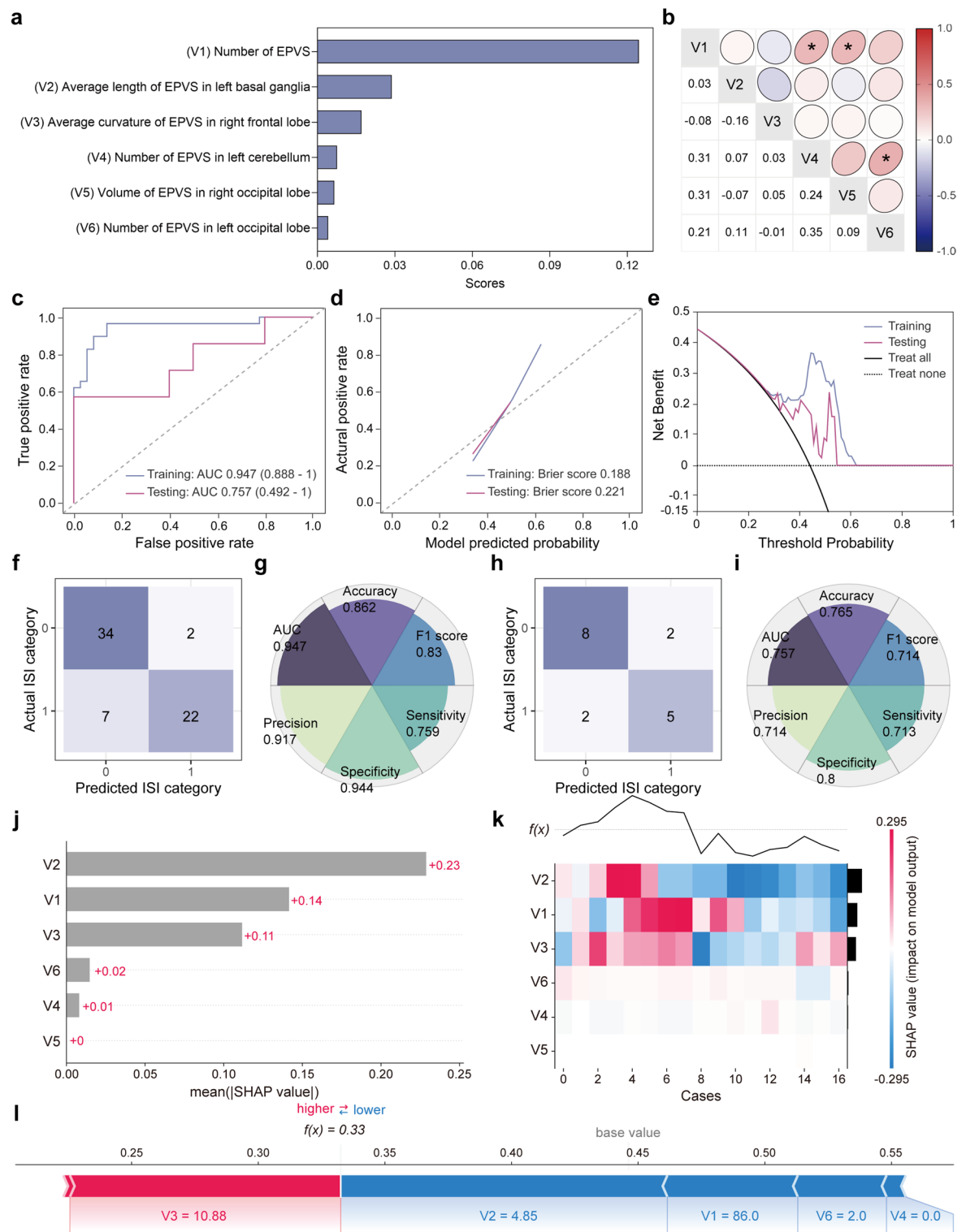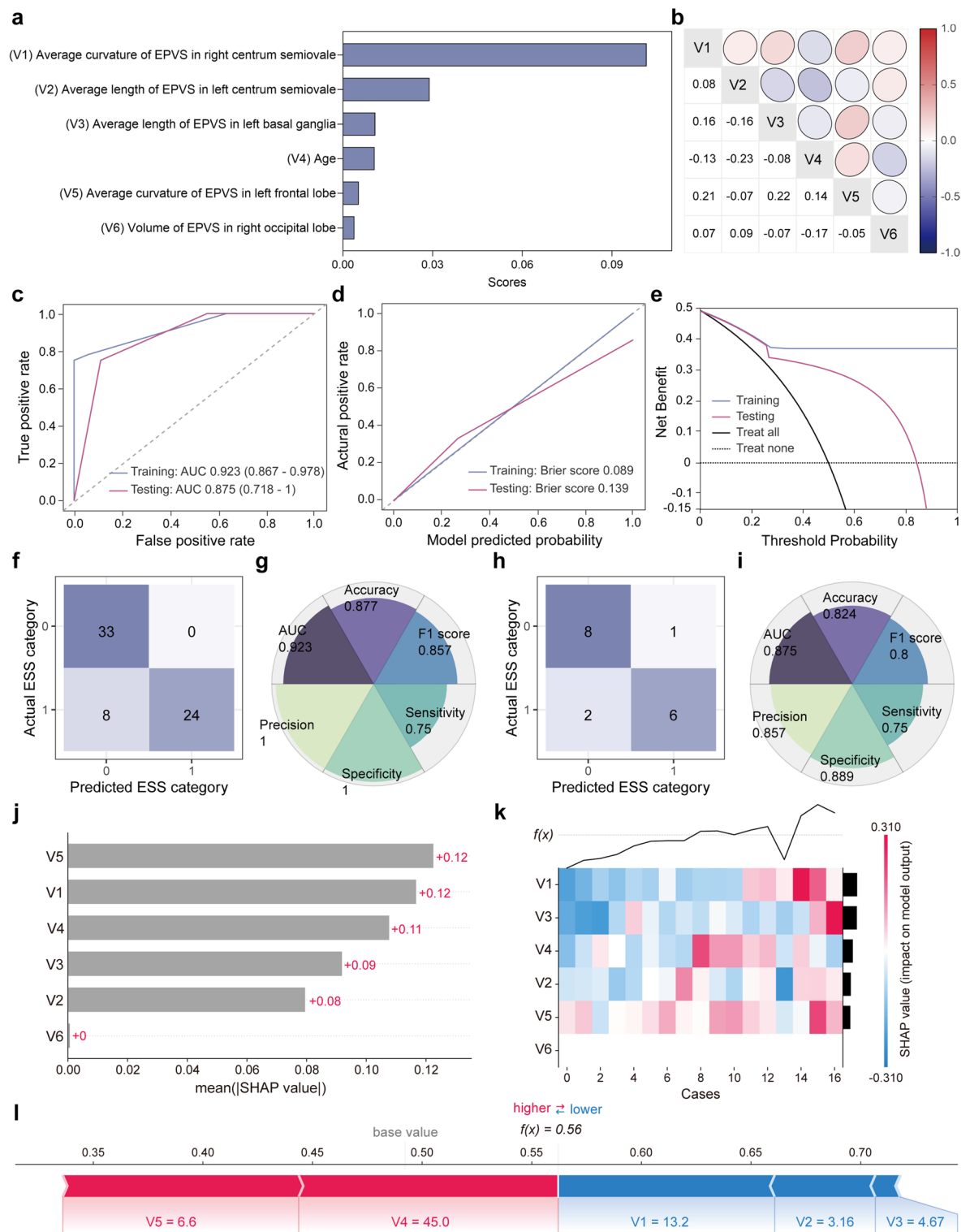
Fig. 4. Model construction and evaluation in identifying subjective sleep quality (ISI). (**a**) Six features selected by the mRMR method. (**b**) Correlation heatmap of selected features. Pearson or Spearman correlation analyses were performed and * indicates $p < 0.05$. Numbers labeled in the plots represent correlation coefficients. (**c**) ROC curves evaluating the trade-off between sensitivity and specificity of the GP model. (**d**) Calibration curves evaluating the consistency of predicted probability and the actual insomnia rate. (**e**) Decision curves showing the clinical net benefit. (**f**) Confusion matrix of the training dataset. The "0" represents non-insomnia and "1" represents insomnia. (**g**) Six metrics of the training dataset. (**h**) Confusion matrix of the testing dataset. (**i**) Six metrics of the testing dataset. (**j**) SHAP summary plot (training dataset) showing feature importance. (**k**) SHAP heatmap (testing dataset) visualizing feature contributions across all testing samples. (**l**) SHAP force plot (testing dataset, representative sample) illustrating how individual features influence the prediction outcome.

Fig. 5. Model construction and evaluation in identifying sleepiness. (**a**) Six features selected by the mRMR method. (**b**) Correlation heatmap of selected features. Pearson or Spearman correlation analyses were performed and * indicates $p < 0.05$. Numbers labeled in the plots represent correlation coefficients. (**c**) ROC curves evaluating the trade-off between sensitivity and specificity of the DT model. (**d**) Calibration curves evaluating the consistency of predicted probability and the actual sleepiness rate. (**e**) Decision curves showing the clinical net benefit. (**f**) Confusion matrix of the training dataset. The "0" represents non-sleepiness and "1" represents sleepiness. (**g**) Six metrics of the training dataset. (**h**) Confusion matrix of the testing dataset. (**i**) Six metrics of the testing dataset. (**j**) SHAP summary plot (training dataset) showing feature importance. (**k**) SHAP heatmap (testing dataset) visualizing feature contributions across all testing samples. (**l**) SHAP force plot (testing dataset, representative sample) illustrating individual prediction drivers.

assessment methodologies. The integration of MRI data with advanced computational models represents a significant advancement in the field, as no prior studies have been known to harness machine learning to such an end with respect to sleep and cognitive function. Furthermore, our research show that EPVSs are associated with subjective symptoms of excessive daytime sleepiness, an area that has received scant attention in previous studies, as far as we know.

We further identified the most predictive EPVSs features across tasks and their key anatomical correlates, including the centrum semiovale (ESS: average EPVSs curvature), frontal lobe (MoCA: EPVSs volume; ISI: average EPVSs curvature), thalamus (MoCA: EPVSs count), basal ganglia (ISI: average EPVSs length), and temporal lobe (PSQI: EPVSs volume), which are as shown in Figs. 2, 3, 4, 5. Previous studies have reported an association between sleep efficiency and EPVSs burden, predominantly in the centrum semiovale[29]; as a large brain region, its subregions may also be involved in sleep regulation[30]. Lesions in the frontal lobe, such as brain tumors, have been shown to alter REM sleep rhythmogenesis[31]. Following subcortical basal ganglia stroke, patients with post-stroke cognitive impairment exhibit reduced structural-functional coupling in the frontal lobe, which correlates with multidimensional cognitive deficits[32]. A review published in 2024 highlighted the thalamus' involvement in nearly all aspects of cognitive functioning and behavior[33], while volume loss and altered functional connectivity of thalamus in Parkinson's disease patients with mild cognitive impairment patients correlates with global cognitive performance[34]. Poor sleep efficiency is independently associated with EPVSs in the basal ganglia, suggesting that sleep may influence structural changes in these fluid-filled cavities[35]. Additionally, increased severity of obstructive sleep apnea, a common sleep-disordered breathing condition, is linked to larger volumes of medial temporal structure (hippocampus and entorhinal cortex) in women[36]. However, data on the implications of key EPVSs features (e.g., average curvature) remain limited, warranting further investigation.

In current study, EPVSs in centrum semiovale, frontal lobe, thalamus, basal ganglia, temporal lobe are found to be associated with sleep disturbances and cognitive function. However, due to the heterogeneity of EPVSs and their diverse risk factors, EPVSs in these anatomical structures may also be linked to other neurological diseases. Multiple risk factors are associated with EPVSs. These include patient age, EPVSs location, and scan indication, as well as clinical factors such as hypertension history, blood pressure, and other vascular risk factors[19]. Additionally, features of small vessel disease, cognitive impairment, and systemic inflammation are also linked to EPVSs[19]. EPVSs burden increases with age in the basal ganglia, centrum semiovale and hippocampus[18,19]. Specifically, hypertension, systemic inflammatory markers, lacunar stroke, and vascular dementia show stronger associations with EPVSs in the basal ganglia[19]. In contrast, high visibility of EPVSs in the centrum semiovale is linked with probable cerebral amyloid angiopathy, recurrent intracerebral haemorrhage and cognitive impairment[18,19]. Thus, the rationale for interpreting EPVSs would be enhanced by discussions on the heterogeneity in EPVSs interpretation across studies and potential confounders such as hypertension or other vascular comorbidities.

Our EPVSs quantification addressed reliability through a validated segmentation model (VB-Net: recall 0.953, precision 0.923) and radiologist review – critical for accurate morphological measurements. To capture regional variations, we focused on region-specific metrics (e.g., the average length of EPVSs in left centrum semiovale), aligning with evidence that EPVSs biology differs by brain region[37]. In our current research, we referenced a previous study for the precise measurement of EPVSs quantification[38], thereby aiming to advance the understanding of EPVSs in the context of subjective sleep quality, cognitive function, and their potential implications in neurodegenerative processes. Fig. S7 further confirms that EPVSs change in the target subregions (i.e., centrum semiovale, basal ganglia, frontal lobes) across all groups, with visible differences that align with the quantitative features used in our models.

Compared to visual rating scales, a growing body of research showed that these computational quantification of EPVSs offers increased sensitivity and precision. To date, the quantification of EPVSs on MRI in research to study has mainly relied on visual rating scales. However, such qualitative scores of visual rating scales are relatively insensitive, limited by floor and ceiling effect, and manual counting of perivascular spaces within individual scan slices is overly time-consuming, particularly in large studies[19]. These computational methods exhibit enhanced sensitivity in detecting associations with white matter hyperintensities(WMH) and retinal vessel diameters[18]. Specifically, computational measures reflecting individual EPVSs features, such as size, length and width show stronger association with WMH, stroke and hypertension than using visual rating scales[39]. Advancements in automated segmentation algorithms, such as the VB-Net architecture, have enabled high-precision volumetric and morphological analyses of EPVSs[38].

In recent years, many studies have employed machine learning techniques to develop predictive models for classifying the severity of cognitive impairment[40]. This study represents the first attempt to utilize MRI quantified EPVSs volumes and machine learning to accurately classify subjective sleep quality and the severity of excessive daytime sleepiness symptoms in young adults with LTMPU, which may hold significant potential for clinical applications. Our study provides preliminary evidence suggesting a relationship between EPVSs, a biomarker indicative of glymphatic dysfunction, and the severity of cognitive impairment, sleep quality, and the severity of excessive daytime sleepiness symptoms in young adults with LTMPU. This is in line with the burgeoning body of research that points to a bidirectional relationship between sleep disturbances and the risk of dementia[26]. The absence of curative treatments for dementia underscores the critical need for preventive interventions[26]. By integrating MRI-quantified EPVSs volumes with machine learning algorithms, our model may offer insights into the early stages of Alzheimer's disease, potentially identifying syndromal conversion in cognitively unimpaired subjects—a domain where data are exceedingly scarce. This approach harnesses the power of neuroimaging to detect preclinical neurodegenerative changes, facilitating both the early diagnosis of Alzheimer's and the monitoring of sleep health. Excessive daytime sleepiness is a public health issue, which is often undervalued, infrequently diagnosed, and inadequately addressed[13]. The observed relationship between EPVSs and excessive daytime sleepiness symptoms in our study suggests that EPVSs could serve as a promising

biomarker for this condition. Further exploration of this association could deepen our understanding of the neurobiology underlying excessive daytime sleepiness, ultimately aiding in the development of improved diagnostic and therapeutic strategies for affected patients.

This study, while pioneering in its approach, has several limitations. Firstly, the modest sample size employed may restrict the generalizability of our findings, with small testing datasets contributing to wide CIs for some models. This small sample also introduces overfitting risk, which may inflate AUC estimates. Though we mitigated this *via* stratified sampling, 5-fold cross-validation (Table S3), selection of 6 stable features (Fig. S8), and use of simple models, the risk cannot be fully eliminated. This warrants validation on a larger scale to enhance the model's robustness and applicability. Second, the observational design precludes causal inferences between daytime sleepiness and EPVSs. The analysis did not examine EPVSs location due to sample size constraints; future work should include more detailed spatial analysis. Third, the demographic homogeneity (Han Chinese, young, right-handed) limits the generalizability of our findings, necessitating future studies with more diverse populations. Fourth, the cognitive impairment model exhibited relatively low sensitivity, which may affect clinical interpretability and applicability. Finally, the single-center design and the lack of a control group or internal stratification by usage time also limit the broader applicability of our conclusion. External validation and more comparative analyses are required in future research.

Conclusions

Our study introduces an innovative analytical framework by integrating MRI-quantified EPVSs metrics with machine learning algorithms, offering a new methodological paradigm for classifying the severity of cognitive impairment, subjective sleep quality and excessive daytime sleepiness symptoms in young adults with LTMPU. The insights gained from this preliminary investigation set the stage for more extensive inquiries into the complex interplay between EPVSs, cognitive function, and sleep quality in the context of LTMPU. Further research with expanded cohorts and multi-centric approaches are imperative for substantiating the reliability and generalizability of our model. Such endeavors will be important in validating the predictive capabilities of our model across various populations and healthcare settings, thereby enhancing its potential impact on clinical diagnostics and therapeutic interventions.

Materials and methods

This study adheres to the CLEAR guidelines for radiomics research, with the completed checklist provided in Supplementary Information.

Participants

This retrospective cross-sectional study was conducted from October 2021 to May 2022 at a medical college in Wen jiang District, Chengdu, China. This study was approved by the Institutional Review Board (IRB, No. EC-20230525-1014) and Ethics Committee of Hospital of Chengdu University of Traditional Chinese Medicine (No.2021KL-093). All relevant institutional IRB and ethics committees granted ethical approval. The studies were conducted in accordance with the local legislation and institutional requirements. All participants provided written informed consent, with de-identified data stored on password-protected servers accessible only to the research team.

From 165 initially recruited students and young teachers (18–50 years), 146 completed baseline assessments through classroom-administered questionnaires (88.5% response rate). The inclusion criteria in this study were as follows: (a) with LTMPU. The duration of mobile phone use per day was obtained by the following question: How long do you usually spend on using a mobile phone per day? The response categories for this question were: less than 2 h, 2 to 4 h, 4 to 6 h, and more than 6 h. LTMPU was defined as using a mobile phone ≥ 4 h per day in consideration of the recent findings[2]. (b) ethnic Han. (c) free of any psychoactive medication at least 2 weeks before and during the study. (d) right-handedness assessed with the Edinburgh Handedness Inventory[41]. Exclusion criteria in this study were as follows: (a) with coronavirus disease 2019 (COVID-19) infections; (b) any significant neuropsychiatric disease formally diagnosed by a psychiatrist or neurologist, including but not limited to major neurocognitive disorder (dementia), insomnia disorder, obstructive sleep apnea, major depressive disorder, bipolar disorder, and schizophrenia, or brain structural abnormality; (c) with MRI contraindications.

At baseline, 91 out of 146 participants (62.3%) reported using a mobile phone ≥ 4 h per day (LTMPU). Each participant with LTMPU completed informed written consent before undergoing magnetic resonance (MR) imaging (within two weeks of completing the scale). Nine participants were excluded because of MRI motion artifacts. Finally, 82 participants with LTMPU were included. Figure 6 illustrates the technical pipeline encompassing EPVSs processing and machine learning modeling. It shows representative EPVSs segmentation results in key brain subregions, confirming the anatomical localization and quality of segmented lesions.

Sample size justification

A power analysis was performed to justify the sample size[42]. The rational for selecting key EPVSs metrics in sample size calculations is detailed in Method S1. We hypothesized that "average_length_of_EPVSs_in_Left_ centrum_semiovale" could be an effective factor in identifying cognitive impairment. The statistical power for the given parameters was calculated using G*Power software (https://www.psychologie.hhu.de/arbeitsgruppe n/allgemeine-psychologie-und-arbeitspsychologie/gpower). As shown in Fig. S9, the sample sizes of group 1 (cognitive normalization) and group 2 (cognitive impairment) were 27 and 55, respectively. And the means of "average_length_of_EPVSs_in_Left_centrum_semiovale" for the two groups were 3.08 (μ_0) and 2.24
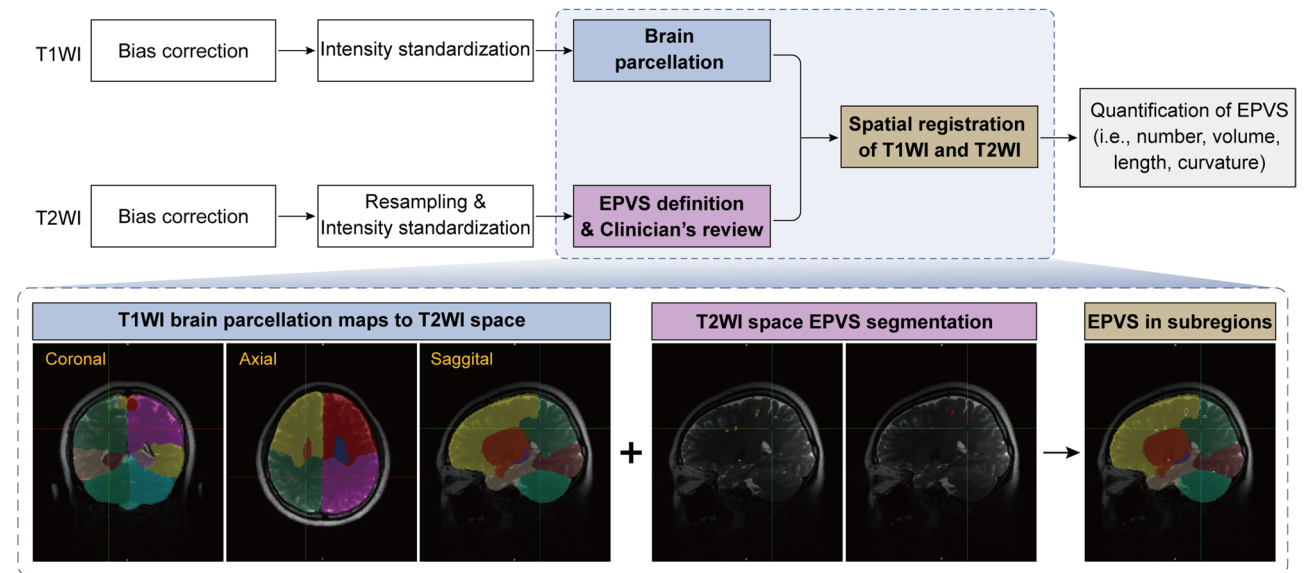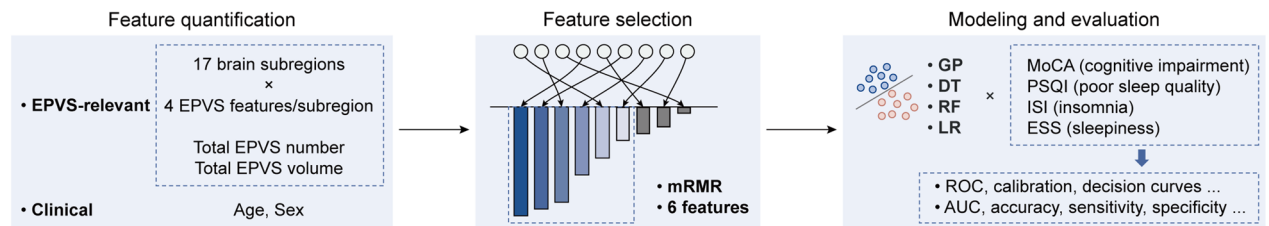
Fig. 6. Flowchart for technical pipeline. (**a**) EPVSs processing, with representative EPVSs segmentation in key brain subregions. (**b**) Machine learning modeling for four classification tasks.

(μ_A), respectively, and the effect size ($d = \frac{\mu_A - \mu_0}{\sigma}$) was 0.913. At a setting of $\alpha = 0.05$, the statistical power ($1 - \beta$) could reach 0.956.

Similar operations were conducted for the other three scales, including PSQI (using "average_curvature_of_EPVSs_in_Left_centrum_semiovale" as a distinguishing factor, ISI (using "average_curvature_of_EPVSs_in_Right_frontal_lobe" as a distinguishing factor), and ESS (using "volume_of_EPVSs_in_Right_occipital_lobe" as a distinguishing factor). The calculated total sample sizes ranged from 74 to 82 given the predefined α, means, σ, and power thresholds (Table S4). We also explored the effect of the sample size on power (Fig. S10). As the sample size increases, so is power because the overlap between the two distributions is decreased. At $n = 82$ with $d = 0.913$ (MoCA), power reached 0.96 ($\alpha = 0.05$). Corresponding values were 0.91 for PSQI ($d = 0.789$), 0.86 for ISI ($d = 0.700$), and 0.89 for ESS ($d = 0.725$), demonstrating balanced specificity (α control) and sensitivity (1 - β) across tasks.

Clinical assessments

To evaluate cognitive and sleep status, all participants were asked to complete the MoCA, the ESS, the PSQI, and the ISI. The severity of cognitive impairment was assessed by the MoCA. The total score of MoCA is in the range of 0 to 30. when the score falls below 26, cognitive impairment is present. The lower the MoCA score is, the worse the cognitive function[43]. The severity of excessive daytime sleepiness symptoms was assessed by the ESS. The total score of ESS is in the range of 0 to 24. An ESS score of more than 6, 11, and 16 was defined as sleepiness, excessive sleepiness, and risky sleepiness, respectively[44]. The severity of subjective sleep quality was assessed by the PSQI. The total score of PSQI is in the range of 0 to 21. A score > 5 suggests poor sleep quality[45]. The severity of subjective sleep quality was assessed by the ISI. The total of ISI is in the range of 0 to 28. An ISI score ≤ 7 indicates absence of insomnia; 8–14 indicates sub-threshold insomnia; 15–21 indicates moderate insomnia; 22–28 indicates severe insomnia[45].

Sleep quality is a complex, multifaceted construct that poses challenges for objective quantification due to inter-individual variability and its inherently subjective nature. The PSQI and ISI are two commonly used instruments of subjective self-report sleep quality. The PSQI, a widely recognized questionnaire for gauging subjective sleep quality, has demonstrated robust reliability and validity, particularly in known-group comparisons. However, concerns regarding its factor model, the large recall period, and the scoring system challenge the value of the global PSQI score for distinguishing poor and good sleepers. The ISI, on the other hand, quantifies perceived

insomnia severity by focusing on the level of disturbance to the sleep pattern, consequences of insomnia, and the degree of concern and distress related to the sleep problem. The ISI has exhibited significant correlations with various sleep questionnaires including PSQI (albeit with low correlation coefficients with ESS), as well as with psychological, health, and psychopathological assessments. Future studies are needed to clarify the factor structure of ISI. In our study, PSQI and ISI are utilized to evaluated the severity of subjective sleep quality.

MR imaging

All patients were examined using a single 3.0 T whole-body scanner (Discovery MR750, GE Healthcare, Milwaukee, WI) equipped with a 32-channel phased array head coil. No contrast agents were administered. T2-weighted images (T2WI) acquisition parameters were: repetition time (TR) = 5613 ms, echo time (TE) = 116 ms, slice thickness = 5.0 mm, voxel size = 0.8 mm×0.8 mm×0.8 mm, slice spacing = 1.5 mm, FOV = 26 mm. 3D T1-weighted imaging (T1WI) was acquired using spoiled gradient echo sequence with voxel size = 1.0 mm×1.0 mm×1.0 mm, TR = 2.9 ms, TE = 3.0 ms, inversion time = 450 ms, flip angle = 8°, slice thickness = 1 mm, matrix = 250 × 250, FOV = 22 cm × 22 cm.

Data preprocessing and EPVSs quantification

This private dataset (n = 82) originates from a retrospective cohort investigating neuroimaging biomarkers in long-term mobile phone users, with partial overlap (imaging raw data) with our prior anxiety-depression study (DOI: https://doi.org/10.3389/fpsyt.2025.1532256).

The image preprocessing procedure was implemented through the uAI research portal (version 20240730, United Imaging Intelligence, https://urp.united-imaging.com/#/)[46] – a commercially available standardized software platform –consisting of several steps, as outlined below. First, N4 bias field corrections were applied to both T1WI and T2WI to remove magnetic field inhomogeneity. Next, grayscale values were standardized by normalizing intensities to the range of [−1, 1] through clipping at 0.1%−99.9%, after resampling T2WI to 1 mm³ isotropic voxels *via* cubic B-spline interpolation. Utilizing a deep learning model VB-Net[47], the skull was removed from T1WI and the whole brain was segmented into 109 regions of interest (ROIs) based on the DK atlas[48].These regions were then consolidated into 17 brain subregions detailed in Table S5, including bilateral frontal lobes, parietal lobes, occipital lobes, temporal lobes, basal ganglia, cerebellum, thalamus, centrum semiovale, and brainstem. Subsequently, EPVSs lesions were automatically segmented from the T2WI image using a built-in VB-Net model[38], which demonstrated high accuracy for EPVSs segmentation with recall and precision of 0.953 and 0.923, respectively (recall = 0.953, precision = 0.923). The AI-generated masks were first reviewed and modified by a radiologist with 5 years of experience, then double-checked by another radiologist with 10 years of experience. Both radiologists were blinded to all clinical data to avoid bias. In case of inconsistent opinions on mask modifications, they consulted to reach a consensus. This two-step validation (automated segmentation + expert review) ensures reliability of subsequent morphological measurements. Furthermore, T1WI and T2WI images were co-registered using a registration algorithm[49], transforming the segmentation mask from the T1WI space to the T2WI space. Finally, a total of 70 quantitative metrics of EPVSs lesions were extracted from original images without spatial filtering. These metrics were defined as "handcrafted radiomics" – features manually designed based on known morphological and anatomical properties of EPVSs, as opposed to features automatically learned by deep learning models. They encompassed the total number and total volume of EPVSs lesions in the whole brain, as well as the number, volume, average length, and average curvature of EPVSs lesions for each of the 17 brain subregions (Table S5). Details regarding the rationale for omitting normalization of subregional EPVSs metrics are provided in Method S2. Each feature was defined and extracted as follows: (1) Number: Number of discrete EPVSs lesions (independent clusters separated by normal tissue); (2) Volume: Total voxel volume of EPVSs lesions calculated as total voxel count × voxel volume, normalized to 1 mm³ isotropic voxels; (3) Average length: Average longest axis of lesions, measured by skeletonizing each lesion and calculating the maximum distance between endpoints; (4) Average curvature: Average curvature of lesion boundaries, defined as the EPVSs length divided by the shortest distance between the begin and end points of the EPVSs. No intensity discretization was applied due to prior min-max normalization. Non-imaging predictors included age (median 38.0 years) and sex (29.3% male). Cognitive-sleep outcomes followed Chinese population thresholds: MoCA < 26 (cognitive impairment), PSQI > 5 (poor sleep), ISI > 7 (insomnia symptoms), ESS > 6 (excessive sleepiness).

Predictive modeling analysis

Quantitative imaging analysis was used to investigate the ability of EPVSs characteristics to predict cognitive impairment, sleep quality, insomnia, and sleepiness symptoms in young adults with LTMPU. The pipeline consisted of two core components: (1) EPVSs segmentation and feature quantification *via* the commercial uAI research portal (version 20240730)[46]; (2) statistical analysis and predictive modeling *via* custom open-source Python code. The modeling workflow was implemented as follows:

Data grouping

All 82 participants had complete imaging and clinical data without missing values necessitating imputation or exclusion. To minimize sampling bias in small datasets, the cohort was split into a training dataset (80%) and a testing dataset (20%) using stratified sampling, which preserved the class distribution of each outcome between datasets. The training dataset was used for feature selection and model construction, and the testing dataset was used to evaluate the robustness and generalizability of the model. No class balancing or oversampling techniques were applied as the original class distributions were preserved for clinical interpretability. The selected models are relatively robust to mild imbalances, and we optimized the classification threshold *via* the Youden index to balance sensitivity and specificity across classes.

Feature selection
The 70 EPVSs quantitative features in conjunction with 2 available clinical features (i.e., sex and age) served as the input to identify the most valuable biomarkers for clinical outcomes. Notably, feature standardization was first conducted using z-score normalization to eliminate the effect of magnitudes between different features. Then, the minimum redundancy maximum relevance (mRMR) method was employed to select the most relevant feature combinations (6 features). The mRMR method is based on two key assumptions: features should be highly relevant to the outcome of interest, and there should be low redundancy between selected features. To evaluate the stability of the selected features, we adopted five-fold cross-validation. The training dataset was divided into five subsets, and mRMR feature selection was performed on each subset separately. For all prediction tasks, the original top 6 features each appeared in at least 3 folds (Fig. S8), with an average occurrence rate exceeding 70%, confirming the stability of the selected features across different data partitions.

Model construction
Based on the selected features, two machine learning algorithms (i.e., Gaussian process [GP], decision tree [DT]) were used to construct the classification models with fixed random seed (20). The choice of GP and DT as final models was guided by our cohort characteristics: (1) Their robustness to small sample sizes ($n = 82$) reduces overfitting risks, unlike ensemble methods (e.g., XGBOOST) or deep learning, which require larger datasets; (2) DT offers interpretable decision rules, and GP provides probabilistic outputs with uncertainty quantification – both critical for clinical interpretation; (3) They achieve stable performance with our 6-feature set without excessive computational demands. Hyperparameter optimization was performed *via* grid search on the training dataset. For each classification task, we retained the model with the highest discriminative performance based on the area under the curve (AUC) on the training dataset, where the GP model was used for the MoCA and ISI classification, and the DT model for the PSQI and ESS classification. The hyperparameters of each model are detailed in Table S6.

Model evaluation
The performance of models was evaluated in the internal testing dataset, which could reflect the robustness and generalizability of models. To address potential variability from small testing datasets, we supplemented this with 5-fold cross-validation on the training dataset, which showed consistent performance across folds (Table S3). The receiver operating characteristic (ROC) curve was first plotted, where the AUC with 95% confidence intervals (CI) could be calculated quantitatively *via* 1,000 bootstrap resamples. Five metrics were calculated to evaluate the consistency between the actual label and predictive label, including accuracy, sensitivity, specificity, precision, and F1-score. These metrics were defined as follows (Eqs. 1–5):

$$Accuracy = \frac{TP + TN}{TP + PF + TN + FN} , \tag{1}$$

$$Sensitivity = Recall = \frac{TP}{TP + FN} , \tag{2}$$

$$Specificity = \frac{TN}{TN + FP} , \tag{3}$$

$$Precision = \frac{TP}{TP + FP} , \tag{4}$$

$$F1score = \frac{2 * Precision * Recall}{Precision + Recall} , \tag{5}$$

where TP represented true positive, TN represented true negative, FP represented false positive, and FN represented false negative. Feature importance was quantified through LASSO regression coefficients, revealing six key predictors in each classification task. Calibration curves were also used to compare the predictive output and the actual outcome. Finally, the decision curves were utilized to show the clinical net benefit for predicting outcomes.

Model interpretability
The SHapley Additive exPlanations (SHAP) analyses were performed for all models[50]. SHAP values quantify the contribution of each feature to individual predictions, where a positive SHAP value indicates that the feature promotes the prediction (increases the probability of the target outcome), and a negative SHAP value indicates that the feature suppresses the prediction (decreases the probability of the target outcome). Visualizations included summary plots (training dataset) for global feature importance, heatmaps (testing dataset) for cross-sample consistency, and force plots (testing dataset) for individual sample explanations.

Statistical analysis
The Shapiro-Wilk tests were used to check the normal distribution of continuous variables. For continuous variables that were approximately normally distributed, they were represented as mean ± standard deviation. For continuous variables with asymmetrical distributions, they were represented as median (25th, 75th percentiles). Categorical variables were represented as counts (percentages), and compared using chi-square tests. The correlation analysis utilized Pearson's method when both variables satisfied normal distribution assumptions; otherwise, Spearman's method was applied. To evaluate the classification performance of machine learning

models, six quantitative metrics (i.e., AUC, accuracy, sensitivity, specificity, precision, and F1-score) were calculated. All statistical analyses were implemented using SPSS (version 26.0, https://www.ibm.com/spss) and R (version 4.2.2, https://www.R-project.org). All figures were plotted using GraphPad Prism 9 (https://www.graphpad.com/), Origin 2021 (https://www.originlab.com/), and Adobe Illustrator CC 2019 (https://www.adobe.com/products/illustrator.html).

Data availability
All data generated or analysed during this study are included in this published article and its supplementary information files. The data presented in this study are available on request from the corresponding author. Analysis utilized uAI research portal (v20240730) with open-source Python/scikit-learn/PyTorch implementations. De-identified data available upon request.

Code availability
The analytical workflow consists of two reproducible components: (1) EPVSs segmentation and feature quantification: Implemented *via* the uAI research portal (version 20240730, United Imaging Intelligence) – a commercially available standardized software platform. Any user with access to this software can obtain the same segmentation results and feature matrices (in tabular format) directly, without the need for custom coding or script manipulation. (2) Statistical analysis and predictive modeling: Custom code for this component is available on GitHub (https://github.com/simonsf/EPVS-Radiomics-ML). The code is built on open-source libraries: Python (v3.9.5), scikit-learn (v1.6.1), and PyTorch (v1.12.1). It includes complete implementations of z-score normalization, mRMR feature selection, GP/DT model training, model evaluation, and result visualization.

References
1. Joshi, S. C., Woltering, S. & Woodward, J. Cell phone social media use and psychological Well-Being in young adults: implications for Internet-Related disorders. *Int J. Environ. Res. Public. Health* **20** (2023).
2. Liu, S. et al. The associations of long-time mobile phone use with sleep disturbances and mental distress in technical college students: a prospective cohort study. *Sleep* **42** (2019).
3. Brautsch, L. A. et al. Digital media use and sleep in late adolescence and young adulthood: A systematic review. *Sleep. Med. Rev.* **68**, 101742 (2023).
4. Nowak, M. et al. Correlations between problematic mobile phone use and depressiveness and daytime Sleepiness, as well as perceived social support in adolescents. *Int J. Environ. Res. Public. Health* **19** (2022).
5. Wacks, Y. & Weinstein, A. M. Excessive smartphone use is associated with health problems in adolescents and young adults. *Front. Psychiatry*. **12**, 669042 (2021).
6. Clarke, A. J. et al. Risk factors for the neurodegenerative dementias in the Western Pacific region. *Lancet Reg. Health West. Pac.* **50**, 101051 (2024).
7. Sadeghmousavi, S., Eskian, M., Rahmani, F. & Rezaei, N. The effect of insomnia on development of alzheimer's disease. *J. Neuroinflammation*. **17**, 289 (2020).
8. Zhao, M., Li, J., Xiang, L., Zhang, Z. H. & Peng, S. L. A diagnosis model of dementia via machine learning. *Front. Aging Neurosci.* **14**, 984894 (2022).
9. Mason, G. M., Lokhandwala, S., Riggins, T. & Spencer, R. M. C. Sleep and human cognitive development. *Sleep. Med. Rev.* **57**, 10147 (2021).
10. Li, M., Wang, N. & Dupre, M. E. Association between the self-reported duration and quality of sleep and cognitive function among middle-aged and older adults in China. *J. Affect. Disord.* **304**, 20–27 (2022).
11. Slonkova, J. et al. Hypocretin-1/orexin-A, sleep and excessive daytime sleepiness in patients with nonconvulsive status epilepticus: A cross-sectional cohort study. *Sleep. Med.* **119**, 192–200 (2024).
12. Cai, Y. et al. Comparing machine learning-derived MRI-based and blood-based neurodegeneration biomarkers in predicting syndromal conversion in early AD. *Alzheimers Dement.* **19**, 4987–4998 (2023).
13. Pérez-Carbonell, L., Mignot, E., Leschziner, G. & Dauvilliers, Y. Understanding and approaching excessive daytime sleepiness. *Lancet* **400**, 1033–1046 (2022).
14. McCarter, S. J. et al. Physiological markers of sleep quality: A scoping review. *Sleep. Med. Rev.* **64**, 101657 (2022).
15. Liu, X. et al. Gait can reveal sleep quality with machine learning models. *PLoS One*. **14**, e0223012 (2019).
16. Jeon, J. et al. Accuracy of machine learning using the Montreal cognitive assessment for the diagnosis of cognitive impairment in parkinson's disease. *J. Mov. Disord.* **15**, 132–139 (2022).
17. Ding, J. et al. Large perivascular spaces visible on magnetic resonance Imaging, cerebral small vessel disease Progression, and risk of dementia: the Age, Gene/Environment Susceptibility-Reykjavik study. *JAMA Neurol.* **74**, 1105–1112 (2017).
18. Duering, M. et al. Neuroimaging standards for research into small vessel disease-advances since 2013. *Lancet Neurol.* **22**, 602–618 (2023).
19. Wardlaw, J. M. et al. Perivascular spaces in the brain: anatomy, physiology and pathology. *Nat. Rev. Neurol.* **16**, 137–153 (2020).
20. Song, T. J. et al. Moderate-to-severe obstructive sleep apnea is associated with cerebral small vessel disease. *Sleep. Med.* **30**, 36–42 (2017).
21. Berezuk, C. et al. Virchow-Robin spaces: correlations with Polysomnography-Derived sleep parameters. *Sleep* **38**, 853–858 (2015).
22. Dredla, B. K., Brutto, D., Castillo, P. R. & O. H. & Sleep and perivascular spaces. *Curr. Neurol. Neurosci. Rep.* **23**, 607–615 (2023).
23. Piantino, J. A., Iliff, J. J. & Lim, M. M. The bidirectional link between sleep disturbances and traumatic brain injury symptoms: A role for glymphatic dysfunction? *Biol. Psychiatry*. **91**, 478–487 (2022).
24. Wang, X. X. et al. MRI-visible enlarged perivascular spaces: imaging marker to predict cognitive impairment in older chronic insomnia patients. *Eur. Radiol.* **32**, 5446–5457 (2022).
25. Greener, J. G., Kandathil, S. M., Moffat, L. & Jones, D. T. A guide to machine learning for biologists. *Nat. Rev. Mol. Cell. Biol.* **23**, 40–55 (2022).
26. Ward, S. A. & Pase, M. P. Advances in pathophysiology and neuroimaging: implications for sleep and dementia. *Respirology* **25**, 580–592 (2020).
27. Zhu, Y. C. et al. High degree of dilated Virchow-Robin spaces on MRI is associated with increased risk of dementia. *J. Alzheimers Dis.* **22**, 663–672 (2010).

28. Del Brutto, O. H. et al. Long coronavirus disease-related persistent poor sleep quality and progression of enlarged perivascular spaces. A longitudinal study. *Sleep* **45** (2022).
29. Aribisala, B. S. et al. Sleep quality, perivascular spaces and brain health markers in ageing - A longitudinal study in the Lothian birth cohort 1936. *Sleep. Med.* **106**, 123–131 (2023).
30. Shih, N. C. et al. Effects of sleep on brain perivascular space in a cognitively healthy population. *Sleep. Med.* **111**, 170–179 (2023).
31. Ortega-Leonard, L. V. Del Río-Portilla, Y. EEG spectral power during REM sleep in patients with frontal brain tumor. *BMC Neurol.* **23**, 195 (2023).
32. Liu, C. et al. Brain structural-functional coupling mechanism in mild subcortical stroke and its relationship with cognition. *Brain Res.* **1845**, 149167 (2024).
33. Biesbroek, J. M., Verhagen, M. G., van der Stigchel, S. & Biessels, G. J. When the central integrator disintegrates: A review of the role of the thalamus in cognition and dementia. *Alzheimers Dement.* **20**, 2209–2222 (2024).
34. Li, M. G. et al. Structural and functional thalamic changes in parkinson's disease with mild cognitive impairment. *J. Magn. Reson. Imaging.* **52**, 1207–1215 (2020).
35. Del Brutto, O. H., Mera, R. M., Brutto, D., Castillo, P. R. & V. J. & Enlarged basal ganglia perivascular spaces and sleep parameters. A population-based study. *Clin. Neurol. Neurosurg.* **182**, 53–57 (2019).
36. Martineau-Dussault, M. et al. Medial Temporal lobe and obstructive sleep apnea: effect of sex, age, cognitive status and free-water. *Neuroimage Clin.* **36**, 103235 (2022).
37. Pham, W. et al. A critical guide to the automated quantification of perivascular spaces in magnetic resonance imaging. *Front. Neurosci.* **16**, 1021311 (2022).
38. Zhang, Z. et al. Quantitative analysis of multimodal MRI markers and clinical risk factors for cerebral small vessel disease based on deep learning. *Int. J. Gen. Med.* **17**, 739–750 (2024).
39. Ballerini, L. et al. Computational quantification of brain perivascular space morphologies: associations with vascular risk factors and white matter hyperintensities. A study in the Lothian birth cohort 1936. *Neuroimage Clin.* **25**, 102120 (2020).
40. Levy, B. et al. Machine learning enhances the efficiency of cognitive screenings for primary care. *J. Geriatr. Psychiatry Neurol.* **32**, 137–144 (2019).
41. Oldfield, R. C. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* **9**, 97–113 (1971).
42. Krzywinski, M. & Altman, N. Points of significance: importance of being uncertain. *Nat. Methods.* **10**, 809–810 (2013).
43. Chen, X. Y., Chen, Z. Y., Dong, Z., Liu, M. Q. & Yu, S. Y. Regional volume changes of the brain in migraine chronification. *Neural Regen Res.* **15**, 1701–1708 (2020).
44. Hou, A. et al. Widespread aberrant functional connectivity throughout the whole brain in obstructive sleep apnea. *Front. Neurosci.* **16**, 920765 (2022).
45. Morin, C. M., Belleville, G., Bélanger, L. & Ivers, H. The insomnia severity index: psychometric indicators to detect insomnia cases and evaluate treatment response. *Sleep* **34**, 601–608 (2011).
46. Wu, J. et al. uRP: an integrated research platform for one-stop analysis of medical images. *Front. Radiol.* **3**, 1153784 (2023).
47. Shi, F. et al. Deep learning empowered volume delineation of whole-body organs-at-risk for accelerated radiotherapy. *Nat. Commun.* **13**, 6566 (2022).
48. Desikan, R. S. et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage* **31**, 968–980 (2006).
49. Avants, B. B. et al. The insight toolkit image registration framework. *Front. Neuroinform.* **8**, 44 (2014).
50. Michalski, A., Duraj, K. & Kupcewicz, B. Leukocyte deep learning classification assessment using Shapley additive explanations algorithm. *Int. J. Lab. Hematol.* **45**, 297–302 (2023).

Author contributions

Li Li: Writing—review & editing, Writing—original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. Jiaojiao Wu: Writing—review & editing, Writing—original draft, Visualization, Software, Methodology, Formal analysis, Data curation, Conceptualization. Bin Li: Writing—review & editing, Writing—original draft, Visualization, Formal analysis, Data curation. Rui Hua: Writing—review & editing, Writing—original draft, Visualization, Formal analysis, Data curation. Feng Shi: Writing—review & editing, Writing—original draft, Visualization, Formal analysis, Data curation. Lizhou Chen: Writing—review & editing, Supervision, Project administration, Methodology, Formal analysis. Yeke Wu: Writing—review & editing, Supervision, Project administration, Methodology, Formal analysis.

Funding

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-026-35845-3.

Correspondence and requests for materials should be addressed to L.C. or Y.W.

Reprints and permissions information is available at www.nature.com/reprints.