International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

Using Multi-Physiological Signals and Linear Discriminant Analysis for Automatic Sleep Stages Classification

[1]*Husam Sadig, [2]Mohammed Gafer Hadra
[1]Doctor of Philosophy, Professor (Assistant) at Dhofar University.
[2]Center for Preparatory Studies, Sultan Qaboos University.

Abstract

Background: Sleep specialists often identify sleep stages manually by visually inspecting human electro-biological signals. This manual process is tedious, requires high experience, and is error-prone. Automation procedures may solely use Electroencephalogram (EEG) signals, which may not be able to capture sleep-related data in other physiological pathways.
Objective: This study aims to analyze whether using multimodal physiological measures, EEG, Electrooculography (EOG), and Electromyography (EMG), in conjunction with entropy-based features, can enhance the six-class automatic classification of sleep stages.
Method: A subset of 15 polysomnographic records of the Sleep-EDF Expanded database was used to obtain data. Having eliminated epochs where there were no values of the entropy, there were 2918 epochs to analyze. Ten features were extracted, including Range Entropy (RangeEn) and Sample Entropy (SampEn) of EEG (Fpz-Cz, Pz-Oz), EOG, and EMG. A Linear Discriminant Analysis (LDA) classifier was used and trained with the help of Leave-One-Out Cross-Validation (LOOCV). Sensitivity, specificity, and accuracy were used to measure performance.
Results: The multimodal LDA classifier reached an accuracy of 84%, with sensitivity at 64% and specificity at 90% after 2918 epochs. Best performance was in Wake stages (82% sensitivity, 99% specificity) and moderate in REM (61% sensitivity, 88% specificity). Challenges persisted in classifying deep sleep stages S3 and S4 with sensitivities of 55% and 69%, respectively. Significant predictors included entropy parameters, particularly beta-band RangeEn. Excluding EOG and EMG entropy features lowered overall accuracy to 79% and reduced sensitivities for Wake and REM stages, highlighting the importance of multimodal entropy information.
Conclusion: These results indicate the usefulness of multimodal PSG signals in automated sleep scoring. Future research should evaluate the generalization to larger cohorts and attempt to evaluate how robust it is to inter-subject variability.

Keywords: Sleep Quality; Sleep Disorder; Automatic Sleep Assessment; Sleep Scoring; Physiological Signals; Sleep Stages

Introduction

Sleep disorder is a widespread disease with a significant effect on the quality of human life (Alnawwar et al., 2023). Studying and diagnosing sleep disorders such as narcolepsy, excessive snoring, sleep apnea, and insomnia requires precise knowledge of the patient's sleep stages. Sleep scoring is used to identify the different stages throughout the hours spent in sleep (Fiorillo et al., 2021). The clinical sleep staging is conventionally conducted by trained technicians who visually examine recordings of the overnight Polysomnography (PSG) and put 30-second epochs into the characteristic stages (Cheng et al., 2024). This manual process, though the gold standard, is time consuming, expensive, and prone to significant inter-rater variance, of particular concern when dealing with transitional or ambiguous epochs (Leino et al., 2022). These constraints have stimulated the creation of automated sleep staging systems that have the potential to be more efficient, objective, and promote the field of sleep research and clinical decision-making on large scale. Existing literature evaluated automated sleep staging, with a large portion of it being based on the Electroencephalogram (EEG), which records electrical activity of the brain and is conventionally discussed as the main signal in classifying sleep (Gaiduk et al., 2023). Many computation techniques have been offered based on the time, frequency, and time-frequency domain features, nonlinear, and entropy based measures. Machine learning models like Support Vector

Husam Sadig
Doctor of Philosophy, Professor (Assistant) at Dhofar University.
Email: sabir_siddiqui@du.edu.om

Received: 4-Dec-2025 Revised: 15-Dec-2025 Accepted: 23-Dec-2025

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

Machines (SVM), Random Forests, Extreme Learning Machines (ELM), and Linear Discriminant Analysis (LDA) have been shown to be promising with regard to classifying EEG epochs into different sleep phases (Hidalgo Rogel et al., 2024). In a study, deep learning models, such as convolutional and recurrent neural networks, have been used on raw EEG or time-frequency representations, with additional accuracy gains being made (Walther et al., 2023). Approximate entropy, sample entropy, fuzzy entropy, permutation entropy, and their multiscale variations have been especially useful in quantifying the intrinsic complexity and irregularity of EEG dynamics at a variety of levels of sleep depth and awareness (Cofré & Destexhe, 2025).

Despite these measures, most automated scoring systems make use of EEG. This is a significant weakness since the rules of sleep staging, be it the classical Rechtschaffen and Kales (RandK) criteria or the new American Academy of Sleep Medicine (AASM) rules, explicitly use other physiological indicators (Van der Lande et al., 2022). Examples include eye movements, which are essential to differentiate between REM, N1 (or S1) stage. Similarly, the EMG activity offers invaluable data on the status of muscle tone as the status distinguishing between the REM (muscle atonia) and the state of wake and NREM (Siddiqui et al., 2025). EEG-only systems will not achieve the ability of multimodal physiological signatures of human sleep unless these complementary modalities are included (Sun et al., 2025). This may result in decreased precision of some of the stages, especially those with fine EEGs or those characterized by the extra-cerebral physiological features.

The use of entropy measures has been extensively studied in EEG data, but it has not been applied to Electrooculography (EOG) and Electromyography (EMG) data (Wu, 2024). However, these indicators also have characteristic patterns of complexities that correlate with changes in the depth of sleep, rapid eye movement, and muscle tone. Although most of the previous literature assesses the classification of automatic sleep by the simplified or binary classification problem (e.g., wake vs sleep, REM vs NREM), there are fewer studies that use the complete six-stage model established by the R&K manual: Wake, S1, S2, S3, S4, and REM (Loh et al., 2022). Fine-grained staging is more clinically applicable and more conducive to the analysis of sleep architecture, although it also has significant difficulties, specifically; the difficulty to distinguish between S3 and S4 deep sleep stages, which are physiologically identical, and may be confounded in contemporary scoring systems (Ganglberger et al., 2024). Consequently, there is an apparent necessity for the research which should study multimodal entropy characteristics with regard to the whole six-class staging and analyze their performance with the help of strict validation techniques. Therefore, this study aimed to examine the role of multimodal physiological aspects based on EEG signals, EO signals, and EMG signals in automatic classification of sleep stages.

Material and Method

Dataset

The data used in this study were collected in the Sleep-EDF (expanded) database, which is found on PhysioNet (Detti, 2020). This database contains polysomnography (PSG) recordings, which were recorded in 1987-2002, and it contains recordings of the original Sleep-EDF database. Two PSGs of about 20 hours each were obtained on each subject, nighttime sleep across two days at the homes of the subjects. The PSG recordings contain EEG channels Fpz-Cz and Pz-Oz, horizontal EOG, submental chin EMG, and event marker signals. Oro-nasal respiration and rectal body temperature are also found in many recordings. All signals are sampled at 100 Hz. In the current analysis, EEG Fpz-Cz, EEG Pz-Oz, EOG and EMG of fifteen PSG recordings were utilized. The hypnogram files that are related give annotations of sleep stages to each of the subjects. These phases are Wake, S1, S2, S3, S4 and REM and each phase is associated with 30 seconds. All data were in European Data Format (EDF).

Pre-processing of Data

The data files of 22 to 24 hours of each signal channel were complete recordings of the signal channel. The state of every 30-second segment, which is called an epoch, was described in the corresponding hypnogram file. These states were Wake, S1, S2, S3, S4, and REM according to the Rechtschaffen & Kales (R&K) manual. These states were scattered all over the signal, and in this experiment, epochs of the same state in all the chosen signals were clustered together. The main product of this pre-processing phase is a set of columns which denote the same alertness states. The epochs of the same state were piled into a matrix, which we referred to as $D\epsilon R^{mn}$, for each signal. Where m is the number of samples per epoch, n is the number of epochs of the same state per subject. In this study, m=3000 (30 seconds at 100 Hz). The recording was

done automatically to reject artifacts and noise generated by other sources with threshold-based algorithms. The time of each alertness state differs among subjects. In order to have a balanced dataset, we picked equal epochs of each subject. In particular, we had 15 polysomnography (PSG) recordings, and 35 epochs of each sleep stage: Wake, S1, S2, S3, S4, and REM.

Sample Entropy and Range Entropy

Sample Entropy

Sample Entropy is a method proposed by Richman and Moorman and is used to estimate the complexity of a time series using the principle of reconstructed phase space (Richman & Moorman, 2000). It involves counting the matched state space vector in m-dimensional phase space. Here, 'm' is also known as the embedding dimension. It represents the length of the state vector in state space. The method can be defined as "the negative natural logarithm of the conditional probabilities that two sequences that are similar for 'm' points remain similar at the next point". Mathematically, SampEn is defined as shown in Equation 1.

$$SampEn=\lim_{n\to\infty}-\ln(A_i/B_i) \quad (i)$$

Where A_i is the sum of conditional probabilities of the (m+1) long segments, and B_i is the sum of conditional probabilities of segments of length m. SampEn is employed in this study to measure the complexity of EEG, EOG, and EMG signals in the various sleep stages and this gives a measure of the dynamical irregularity of each epoch.

Range Entropy

Range Entropy B is a new modification of SampEn to overcome some of its limitations. The main difference is that RangeEn B calculates the normalized pseudo difference between two state vectors instead of using Chebyshev distance. The pseudo difference between two vectors X_i^m and X_j^mcan be seen in Equation 2:

$$d\_range\ (X^{i}\hat{}m, X\_j\hat{}m) = \max_k |x\_(i+k)-x\_(j+k)| \Big/ \left(\max_k |x\_(i+k)-x\_(j+k)| - \min_k |x\_(i+k)-x\_(j+k)|\right) \quad (ii)$$

For k = 0, 1, …, m − 1 where m is the embedding dimension.

Feature Extraction

High-dimensional time series such as EEG, EOG, and EMG signals have a lot of redundancy and noise. Since the vast majority of physiologically significant data is represented in their time-varying frequency content and nonlinear dynamics, the first step before classification is the derivation of compact and discriminative features. The dimensionality reduction of feature extraction, improved computational efficiency, and better performance of classifiers are achieved by describing each epoch by a few informative descriptors.

Methods of feature extraction may be broadly divided into linear and nonlinear methods. Time-domain features, spectral (frequency-domain) features, transformed features and time-frequency representations are linear methods. Complex temporal properties are captured in nonlinear methods, which contain entropy measures, fractal dimensions, and correlation dimensions. This study concentrated on nonlinear complexity measures because they have been shown to capture the variation in neural activation, muscle tone, and ocular dynamics between sleep stages.

Two entropy based features were obtained including Sample Entropy (SampEn) and Range Entropy (RangeEn). EEG beta band (1330Hz) is closely related to wakefulness, cognitive activities, and alertness changes. Both EEG channels (FpzCz and PzOz) were band-pass filtered into the beta range with a zero-phase digital filter in order to capture this physiologically important activity. RangeEn of the signal segments in the beta-band was calculated, and SampEn and RangeEn of the entire-band EEG epochs was also calculated. This enabled to record global EEG complexity as well as high-frequency alertness-related dynamics.

SampEn and RangeEn were calculated based on Equations (1) and (2), respectively to each 30-second epoch of EOG and EMG signals. These modalities provide additional physiological data: EOG provides ocular activity that is important in the differentiation of REM and light sleep, whereas EMG records muscle tone changes that are typical of REM and stage changes. The addition of entropy characteristics of these channels thus adds complementary information to that which can be obtained by EEG.

The results were a ten-feature data set as presented in Table 1. Six of these features were extracted from the EEG channels, and the other four features were extracted from the EOG and EMG channels. Namely, the extracted features are: RangeEn of EEG (Pz-Oz), SampEn of EEG (Pz-Oz), RangeEn of Beta sub-band in EEG (Pz-Oz), RangeEn of EEG (Fpz-Cz), SampEn of EEG (Fpz-Cz), RangeEn of Beta sub-band in EEG (Fpz-Cz), RangeEn of

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

EOG, SampEn of EOG, RangeEn of EMG, and SampEn of EMG. This multimodal entropy representation was created to represent cortical (EEG) and extracerebral (EOG, EMG) physiological dynamics of interest to sleep staging.

Table 1. Extracted Entropy Features from EEG, EOG, and EMG Signals

Feature Name	Signal Source	Short Definition	Unit / Expected Range
RangeEn (Pz–Oz)	EEG (Pz–Oz)	Range Entropy computed on full-band EEG to quantify amplitude-normalized variability.	Dimensionless; typically 0–2
SampEn (Pz–Oz)	EEG (Pz–Oz)	Sample Entropy measures irregularity of EEG signal patterns	Dimensionless; typically 0–3
RangeEn-Beta (Pz–Oz)	EEG (Pz–Oz, 13–30 Hz)	Range Entropy computed after beta-band filtering to capture alertness-related activity.	Dimensionless; 0–2
RangeEn (Fpz–Cz)	EEG (Fpz–Cz)	Full-band Range Entropy reflecting frontal cortical complexity	Dimensionless; 0–2
SampEn (Fpz–Cz)	EEG (Fpz–Cz)	Sample Entropy quantifying irregularity in frontal EEG	Dimensionless; 0–3
RangeEn-Beta (Fpz–Cz)	EEG (Fpz–Cz, 13–30 Hz)	Beta-band Range Entropy for alertness-related frontal activity	Dimensionless; 0–2
RangeEn (EOG)	EOG	Range Entropy capturing variability in ocular movement patterns	Dimensionless; 0–1.5
SampEn (EOG)	EOG	Sample Entropy reflecting irregularity of eye-movement dynamics	Dimensionless; 0–2
RangeEn (EMG)	EMG	Range Entropy captures muscle tone variability and transitions	Dimensionless; 0–1.5
SampEn (EMG)	EMG	Sample Entropy representing irregularity in muscle activation patterns	Dimensionless; 0–2

Entropy features of all subjects were combined by sleep stage (Wake, S1, S2, S3, S4, REM) to enhance generalizability and minimize inter-subject bias, as has been done in [26,27]. The first dataset comprised 3150 epochs (15 PSG recordings, 35 epochs, 6 stages). In the process of feature extraction, there were epochs with undefined SampEn or RangeEn values because of the lack of variability in the signal or degeneracy in template matching. These led to the loss of 232 cases that were eliminated through listwise deletion. The last dataset to be analysed had 2918 valid epochs. Figure 1 gives a summary of the preprocessing and entropy feature extraction process, such as beta-band filtering and calculation of SampEn and RangeEn of all modalities.

Classification and validation methods

Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a machine learning method that is extensively applied in multi-class classification and is supervised. LDA classifies cases by using a linear combination of continuous predictor variables. The groups are represented by the categories of a categorical variable. Within the framework of sleep staging, the groups are the six stages of sleep (Wake, S1, S2, S3, S4, REM), and the predictors are the entropy features extracted. LDA approximates a discriminant function of each group on the assumption that the data are the result of multivariate normal distributions with equal covariance matrices. The discriminant function is given in Equation 3:

$$D_i(x)=f(w_0+\Sigma^k_{j=1}\ w_j x_j) \quad \text{(iii)}$$

Where, $D_i(x)$ is the linear discriminant score for the ith

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

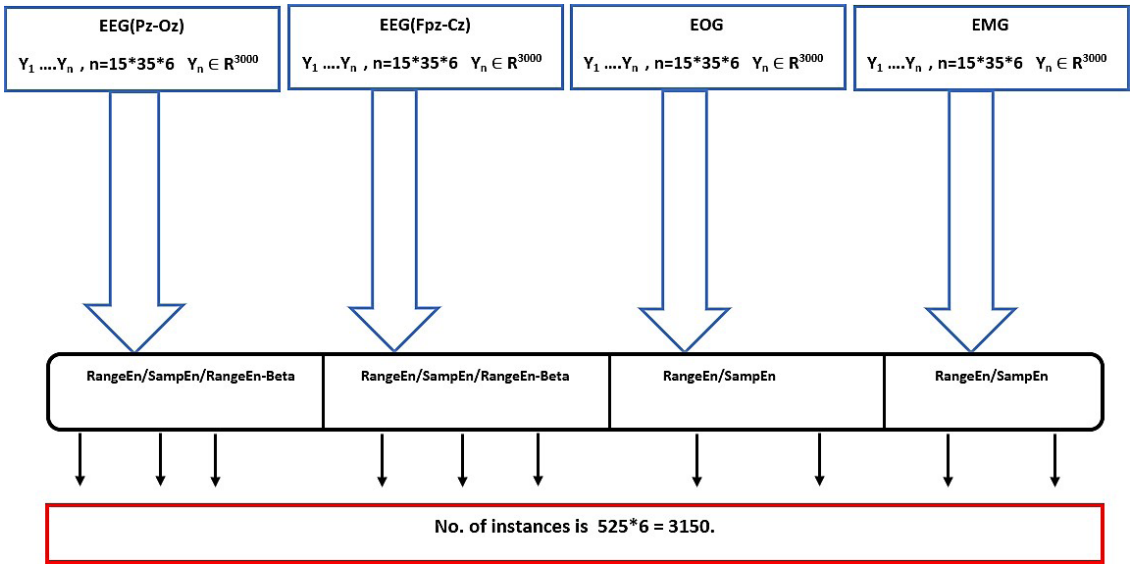

Figure 1. Feature Extraction Method

case in the sample (i = 1, 2, …, n); w is the weight vector (or the coefficient vector); w_0 is the threshold weight; the coefficients w_j are the components of the weight vector $w_;$ and w_j are the predictor variables (or classification variables). The observation is put in class with the highest score in the discriminant function. The discriminant coefficients are maximized to achieve the maximum separation between the groups, hence, enhancing the classification performance.

The predictors in this research are the ten entropy-based features derived from EEG (Pz 2 Oz, Fpz 2 Cz), EOG, and EMG channels: RangeEn and SampEn of each channel, and RangeEn of the beta-band-filtered EEG segments. Though Logistic Regression (LR) can be classified, standard LR is inherently a binary classification and cannot effectively solve multi-class problems without some extension (e.g., one-vs-rest approaches). LDA, in its turn, is intrinsically multi-class and is more interpretable and computationally efficient in this case.

Cross-Validation

Cross-validation is a statistical resampling method that is applied to approximate the predictive model's generalization to unknown data. It is especially significant in cases where the dataset at hand is small, since it gives an objective estimate of the performance of the classifier. In k-fold cross-validation, the data is divided into k equally sized folds. Each fold is then taken as a test set, and the rest of the k-1 folds are taken as training. The average of the performance is taken over all folds.

We used Leave-One-Out Cross-Validation (LOOCV) in this study, which is a severe form of k-fold cross-validation, when k is equal to the number of data instances. In LOOCV, an epoch is separated out once as the test instance, and the rest of the N-1 epochs are used to train the classifier. This is continued until all the epochs have been used as the test case once.

LOOCV has two major strengths:

- Optimal use of data to train, which is important when the sample sizes are small.
- Dependable performance estimation, since every observation is tested separately.

Sensitivity, Specificity, and Accuracy

In order to measure the performance of the classifier, we calculated three common measures based on the confusion matrix, namely, sensitivity, specificity, and accuracy. A confusion matrix is a summary of the association between the real and the predicted class labels, as shown in Table 2.

The classifier is used to predict the sleep stage class for the sample cases. For a given sleep stage class, the prediction will either be 'positive' (if the classifier predicted the class) or 'negative' (if the classifier predicted another class). In many instances, the classifier makes 'true' predictions. However, in each prediction, the classifier is subject to an error (i.e. false prediction). Corollary, the interpretation of the confusion matrix elements is as follows:

True Positive (TP): Classifier predicted a positive sleep stage, and it is true (correct classification). True Negative

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

Table 2. Confusion Matrix Structure

	Predicted Positive	**Predicted Negative**
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

(TN): Classifier predicted a negative sleep stage, and it is true (correct classification). False Positive (FP): The Classifier predicted a positive sleep stage, and it is false (wrong classification). False Negative (FN): The Classifier predicted a negative sleep stage, and it is false (wrong classification).

The mathematical expressions of the performance indicators are as follows: Sensitivity (True Positive Rate) = TP / (TP + FN)

Specificity (True Negative Rate) = TN/ (TN+FP) Accuracy = (TP+TN) / (TP +TN +FP +FN)

Results

LDA Model Estimates

The LDA model presented in Table 4 is estimated with 2918 valid epochs. Both unstandardized and standardized coefficients are presented. Normalized coefficients show how important each of the entropy features is in differentiating among the six sleep stages. Univariate ANOVA tests indicate that all 10 predictors significantly differentiated the sleep stages (all $p < 0.01$). The discriminatory ability of the LDA model is high, with the eigenvalue of 2.841 and a canonical correlation of 0.86, which shows that the discriminant function accounts for 65.7 percent of the total between-group variance.

Table 3. Estimated LDA Coefficients and Univariate ANOVA

Predictor	**Unstandardized Coefficient**	**Standardized Coefficient**	**F-value (ANOVA)**
RangeEn Pz–Oz	–5.253	–0.803	10.02***
RangeEn Fpz–Cz	0.373	0.057	643.65***
RangeEn EOG	1.537	0.253	30.17***
RangeEn EMG	42.635	0.068	17.20***
RangeEn Beta Pz–Oz	3.303	0.305	587.76***
RangeEn Beta Fpz–Cz	2.961	0.302	708.64***
SampEn Pz–Oz	–0.106	–0.040	179.97***
SampEn Fpz–Cz	0.211	0.100	62.30***
SampEn EOG	–0.186	–0.073	138.58***
SampEn EMG	25.144	0.113	37.43***
Constant	–7.980	—	—

*($df_1 = 5$, $df_2 = 2912$ for all F-tests; *** $p < 0.01$)*

These findings prove that complexity-based characteristics based on EEG, EOG, and EMG give valuable discriminatory data on sleep stages. The ANOVA values, particularly of beta-band RangeEn, indicate that nonlinear variability in the beta range is very informative in the difference between wake and sleep. The strong canonical correlation (0.86) supports the fact that the LDA model represents a meaningful linear separation among the six R&K stages.

Classification Using All Cases

Table 4 shows the results of the classification based on all valid epochs. Proper percentages of classification were between:

- Wake: 82.7%
- S4: 70.0%
- REM: 61.5%
- S1: 60.2%
- S2: 58.4%

- S3: 57.4%

Misclassifications were most prominent between S3 and S4, where:

- 29.6% of S4 → S3
- 32.4% of S3 → S4

Table 4. Classification results using all cases

Actual → Predicted	S4	S3	S2	S1	Wake	REM	Total
S4	366 (70%)	155 (29.6%)	0 (0%)	2 (0.4%)	0 (0%)	0 (0%)	523
S3	170 (32.4%)	301 (57.4%)	43 (8.2%)	4 (0.8%)	1 (0.2%)	5 (1%)	524
S2	20 (3.8%)	102 (19.5%)	305 (58.4%)	38 (7.3%)	2 (0.4%)	55 (10.5%)	522
S1	7 (1.4%)	18 (3.5%)	28 (5.4%)	312 (60.2%)	16 (3.1%)	137 (26.4%)	518
Wake	2 (0.7%)	1 (0.3%)	0 (0%)	43 (14.1%)	253 (82.7%)	7 (2.3%)	306
REM	4 (0.8%)	10 (1.9%)	56 (10.7%)	130 (24.8%)	2 (0.4%)	323 (61.5%)	525

The entries are the counts and percentages (in brackets) of cases in the data that are correctly and incorrectly classified.

The high accuracy of Wake is due to its unique physiological characteristics (high beta activity, higher EMG tone), which entropy can capture. The confusion of S3 and S4 is symmetric, which is explained by the fact that these two stages are physiologically similar, which is in line with the modern unification of the AASM into a single N3 stage.

LOOCV Classification Results

The results of the Leave-One-Out Cross-Validation (LOOCV) are presented in Table 5. There was a small (0.3–1.2%) but consistent performance reduction between stages, indicating a constant generalization.

LOOCV proves that the LDA model is generalizable across epochs in the dataset. The accuracy decrease is small, indicating that there is little overfitting. Nevertheless, the LOOCV does not test generalization

Table 5. LOOCV Classification Results

Actual → Predicted	S4	S3	S2	S1	Wake	REM	Total
S4	360 (68.8%)	161 (30.8%)	0 (0%)	2 (0.4%)	0 (0%)	0 (0%)	523
S3	179 (34.2%)	290 (55.3%)	45 (8.6%)	4 (0.8%)	1 (0.2%)	5 (1%)	524
S2	20 (3.8%)	104 (19.9%)	301 (57.7%)	38 (7.3%)	2 (0.4%)	57 (10.9%)	522
S1	7 (1.4%)	19 (3.7%)	28 (5.4%)	310 (59.8%)	16 (3.1%)	138 (26.6%)	518
Wake	4 (1.3%)	1 (0.3%)	0 (0%)	42 (13.7%)	252 (82.4%)	7 (2.3%)	306
REM	4 (0.8%)	10 (1.9%)	56 (10.7%)	133 (25.3%)	2 (0.4%)	320 (61%)	525

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

between subjects, and external validation on independent PSG records is an essential step to be undertaken in future studies.

Similarly, exclusion of 232 cases may have resulted in selection bias in case missingness was conditional on the behavior of entropy. The stability between re-substitution and LOOCV, however, indicates that the effect is minimal.

Classifier Performance Measures

Table 6 is a summary of sensitivity, specificity, and accuracy by sleep stage. Wake was the most performing in all the measures.

Table 6. Performance Metrics Using All Modalities

State	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
S4	360	1473	214	163	69%	87%	83%
S3	290	1543	295	234	55%	84%	78%
S2	301	1532	129	221	58%	92%	84%
S1	310	1523	219	208	60%	87%	81%
Wake	252	1581	21	54	82%	99%	96%
REM	320	1513	207	205	61%	88%	82%
Overall	—	—	—	—	**64%**	**90%**	**84%**

The values of specificity are high, indicating that the classifier does not often mislead a particular stage with the rest of the stages. Patterns of sensitivity indicate physiological variations: Wake is the most recognizable, and deep sleep (S3/S4) is the most difficult, which is in line with clinical scoring difficulty.

Effect of Removing EOG and EMG Modalities

When the LDA model was retrained with EEG features only, the general accuracy dropped to 79%, which is a 5% downgrade compared to 84%. Drops were also noticed at stage levels.

Table 7. Performance measures without EOG and EMG

State	TP	TN	FP	FN	Sensitivity	Specificity	Accuracy
S4	307	1401	274	218	58%	83%	77%
S3	230	1498	355	279	45%	80%	73%
S2	250	1502	180	251	49%	89%	80%
S1	275	1490	254	241	53%	85%	78%
Wake	201	1541	71	94	68%	95%	91%
REM	280	1570	247	248	53%	86%	78%
Overall	—	—	—	—	**54%**	**86%**	**79%**

The performance reduction proves that EOG and EMG provide distinctive and complementary data to EEG-based entropy features. The steep decrease in Wake sensitivity (82% to 68%) and REM sensitivity (61% to 53%) reveals the significance of muscle tone and eye-movement irregularity, neither of which is included in EEG features only. In this way, there is undoubtedly a performance advantage to the multimodal approach.

Discussion

The current experiment investigated the ability of entropy-based features of multimodal physiological signals, which are EEG, EOG, and EMG, to enhance six-class automatic classification of sleep stages using an LDA model. In 2918 epochs of the Sleep-EDF data, all ten entropy features were found to have significant discriminatory ability, and the multimodal classifier had a total accuracy of 84%, sensitivity of 64% and specificity of 90%. These findings shed some light on some of the most important discoveries regarding the physiological significance of entropy measures, the usefulness of multimodal information, and the difficulty of distinguishing particular sleep stages.

Interpretation of Multimodal Entropy Features

Entropy measures indicate the extent of irregularity or complexity of physiological signals. Their high level of statistical significance (all ANOVAs $p < 0.01$ univariate) indicates that sleep stages are not only different in their amplitude and frequency, as it is traditionally defined, but also in their nonlinear dynamics. The significance of the RangeEn and SampEn measured in both full-band and beta-band EEG channels is in agreement with the established neurophysiological properties of sleep (Figorilli et al., 2021). An example is the increased complexity and beta activity that is evident in wakefulness and reflected by the increased entropy in the beta-subband features. Deep NREM sleep stages (S3 and S4), on the other hand, have slower and more regular oscillatory patterns, which lead to lower entropy values.

The high scores of EOG and EMG entropy features also help to prove that sleep phases indicate alterations in various physiological systems. REM sleep, with its rapid eye movements and muscle atonia, had entropy patterns that were in line with the distinctive ocular and muscular dynamics of the sleep. EMG-based entropy features also represented well the wake and light sleep, which exhibit more EMG variability. The results are consistent with the known clinical scoring criteria that focus on the significance of EOG and EMG in addition to EEG (Sharma, Yadav, et al., 2022).

Performance and Physiological Interpretation of Classification

The LDA classifier did best with Wake (82% sensitivity, 99% specificity), then S4, and REM. Such a trend is in line with the clear physiological markings of these phases (Ekram, 2024). Wake has high-frequency EEG activity, increased EMG tone, and typical eye-movement patterns- aspects that entropy is effective at capturing. The REM sleep also exhibited moderate classification (61% sensitivity), which is a characteristic of the rapid eye movement and the inhibited muscle activity (Figorilli et al., 2021).

S3 and S4 were the most difficult steps of the classifier. These high levels of cross-classification (around 30% two-way confusion) are a reflection of the physiological similarity between these deep-sleep phases (Morokuma et al., 2023). S3 and S4 also have only a small difference in the percentage of delta activity in clinical scoring, and their EEG dynamics are significantly similar. In fact, the current AASM guidelines have combined these steps into one N3 category, as it is generally accepted that the difference is subtle and sometimes challenging even for human scorers (Malhotra, 2024). Thus, the misclassification of S3 and S4 is natural, and it cannot be considered a weakness of the entropy features or LDA classifier, but it is a property of the scoring system.

Comparison to Existing Automated Sleep Scoring Approaches

Previous studies on automated sleep staging have shown that EEG-based classifiers tend to have accuracies of between 75 and 90% based on feature sets, type of classifier, and number of stages (Yazdi et al., 2024). The performance of studies that are based solely on linear or spectral features tends to report lower performance in differentiating transitional or deep-sleep stages (Soleimani et al., 2023). Other nonlinear-based works (including approximate entropy, permutation entropy, and multiscale entropy) have demonstrated better discrimination, especially of Wake and REM, but still have difficulty in differentiating mid-stage NREM (Zandbagleh et al., 2025). Although not studied as extensively as the EEG-only models, multimodal methods usually claim better classification results because of the complementary character of the EOG and EMG signals (Muhammad et al., 2025). A study involving the combination of EOG and EMG has demonstrated a higher ability to detect REM and Wake stages as indicators of the physiological significance of eye movement and muscle tone patterns during sleep (Sharma, Darji, et al., 2022). These findings have been validated in the current study through entropy features, where both REM and Wake are more sensitive when multimodal signals are considered.

LDA is simpler to use than the current deep-learning architectures, but it provides interpretability and transparency. Studies have demonstrated that the use of simpler linear models can be competitive when applied with appropriate features and proper preprocessing (James et al., 2023; Van Der Donckt et al., 2023). The findings of this study demonstrated that although upon using a classical classifier, the multimodal entropy features were as accurate as or more than most of the more traditional and nonlinear EEG-only systems.

Added Value of EOG and EMG Modalities

One of the main contributions of the research is the ability to show that the use of entropy features based

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

on EOG and EMG significantly enhances the classification performance. The elimination of these modalities decreased overall accuracy by 84% to 79% and sensitivity in Wake (82% to 68%) and REM (61% to 53%) dropped significantly. This conforms to the previous multimodal PSG research that EOG is critical in the determination of REM and light sleep, and EMG is critical in the determination of Wake and REM because of the variation in muscle tone (Samaee et al., 2025).

In contrast to most of the current literature that uses EOG or EMG amplitude or spectral measures, this study assesses their nonlinear complexity, which shows that irregularity and variability patterns in both ocular and muscular activity do have stage-specific information (Kose et al., 2021). This builds on the literature by emphasizing entropy as a useful multimodal feature extraction methodology, not just to cortical dynamics, but also to extracerebral physiology.

Generalization and Model Stability

The results of LOOCV were close to the full-sample assessment, which shows that the LDA model is well generalized in the dataset. This stability is important as most previous automated sleep staging studies use small hold-out test sets or scanty cross-validation, which leads to overly optimistic performance estimates (Gaiduk et al., 2023; Zhai, 2023). LOOCV is a more rigorous test of model reliability since every epoch is used as a test case. But, LOOCV measures generalization on an epoch level as opposed to a subject level. Earlier research has pointed out that subject-wise cross-validation tends to be less accurate because of the inter-individual differences in sleep architecture, EEG morphology, and signal noise (Cisotto et al., 2024). Therefore, the model should be tested on invisible subjects in future work to find out whether the multimodal entropy approach is inter-population strong.

A potential source of bias is the removal of 232 epochs because of the undefined entropy values. Although the change in LOOCV is minimal, indicating that the effect is not that large, previous studies on entropy-based features indicate that missingness is frequently caused by low-variance or artifact-contaminated segments. The entropy algorithms (e.g., changing tolerance parameters) could be improved in the future to minimize missing data and enhance inference.

Methodological and Practical Implications

Although LDA is not as complex as nonlinear classifiers or deep learning systems, it has the benefits of interpretability, computational efficiency, and stability, highlighting its importance in becoming more and more crucial to clinical applications (Shani & Moradi, 2023). Many studies have indicated that classical models can be competitive with informative features (Ahmed et al., 2023) (Cisotto et al., 2024). This is confirmed in the current study as it employed a linear model, the multimodal entropy features provided accuracy rates that were higher than a variety of EEG-only or linear-feature models.

The findings of the study have clinical implications, as they emphasize the need to use a variety of physiological channels in the process of automated sleep staging. This is consistent with classical PSG scoring criteria, which underscore the fact that sleep is a multimodal physiological condition and not a phenomenon of the brain (Figorilli et al., 2021).

Limitations and Future Research

One of the major strengths of this study is that multimodal entropy features based on EEG, EOG, and EMG were used, which gave complementary physiological measures and better classification rates than using EEG alone. One limitation lies in the fact that a dataset with a small number of subjects was used, which limits the external generalizability and might not reflect inter-individual variability. These results should be confirmed in larger independent cohorts in future studies, subject-wise cross-validation should be examined, entropy extraction should be optimized to minimize missing values, and it should be determined whether nonlinear or deep-learning models can further improve the performance of multimodal classification.

Conclusion

This study showed that EEG, EOG, and EMG can be used together in a multimodal framework to add more theoretical power to sleep-stage classification systems based on nonlinear measures of complexity. The combination of several entropy characteristics confirms their possible ability to differentiate the characteristic dynamics of various sleep conditions, as well as emphasizes the natural challenge of distinguishing deep-sleep phases because of their physiological similarity. The study also confirms the stability of the method and emphasizes the informational complementation of the inclusion of ocular and muscular activity. The results indicate that multimodal entropy-based approaches should be implemented in the future automated sleep-scoring systems to improve

interpretability and methodological soundness.

Declaration Statement

The authors have nothing to declare.

Funding

No funds or grants were received for the completion of this study.

Data Availability Statement

Data will be made available upon request from the corresponding author.

References

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., Mofijur, M., Shawkat Ali, A., & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521-13617.

Alnawwar, M. A., Alraddadi, M. I., Algethmi, R. A., Salem, G. A., Salem, M. A., & Alharbi, A. A. (2023). The effect of physical activity on sleep quality and sleep disorder: a systematic review. Cureus, 15(8).

Cheng, H., Yang, Y., Shi, J., Li, Z., Feng, Y., & Wang, X. (2024). Comparison of automated deep neural network against manual sleep stage scoring in clinical data. Computers in Biology and Medicine, 179, 108855.

Cisotto, G., Zancanaro, A., Zoppis, I. F., & Manzoni, S. L. (2024). hvEEGNet: a novel deep learning model for high-fidelity EEG reconstruction. Frontiers in Neuroinformatics, 18, 1459970.

Cofré, R., & Destexhe, A. (2025). Entropy and Complexity Tools Across Scales in Neuroscience: A Review. Entropy, 27(2), 115.

Detti, P. (2020). Siena scalp EEG database. physionet, 10, 493.

Ekram, S. S. (2024). N2 Stage Impact on REM Sleep Behavior Disorders Detection Politecnico di Torino].

Figorilli, M., Lanza, G., Congiu, P., Lecca, R., Casaglia, E., Mogavero, M. P., Puligheddu, M., & Ferri, R. (2021). Neurophysiological aspects of REM sleep behavior disorder (RBD): a narrative review. Brain Sciences, 11(12), 1588.

Fiorillo, L., Favaro, P., & Faraci, F. D. (2021). Deepsleepnet-lite: A simplified automatic sleep stage scoring model with uncertainty estimates. IEEE transactions on neural systems and rehabilitation engineering, 29, 2076-2085.

Gaiduk, M., Serrano Alarcón, Á., Seepold, R., & Martínez Madrid, N. (2023). Current status and prospects of automatic sleep stages scoring. Biomedical engineering letters, 13(3), 247-272.

Ganglberger, W., Nasiri, S., Sun, H., Kim, S., Shin, C., Westover, M. B., & Thomas, R. J. (2024). Refining sleep staging accuracy: transfer learning coupled with scorability models. Sleep, 47(11), zsae202.

Hidalgo Rogel, J. M., Martínez Beltrán, E. T., Quiles Pérez, M., López Bernal, S., Martínez Pérez, G., & Huertas Celdrán, A. (2024). Studying drowsiness detection performance while driving through scalable machine learning models using electroencephalography. Cognitive Computation, 16(3), 1253-1267.

James, G., Witten, D., Hastie, T., Tibshirani, R., & Taylor, J. (2023). Linear model selection and regularization. In An Introduction to Statistical Learning: with Applications in Python (pp. 229-288). Springer.

Kose, M. R., Ahirwal, M. K., & Kumar, A. (2021). A new approach for emotions recognition through EOG and EMG signals. Signal, Image and Video Processing, 15(8), 1863-1871.

Leino, A., Korkalainen, H., Kalevo, L., Nikkonen, S., Kainulainen, S., Ryan, A., Duce, B., Sipilä, K., Ahlberg, J., & Sahlman, J. (2022). Deep learning enables accurate automatic sleep staging based on ambulatory forehead EEG. IEEE Access, 10, 26554-26566.

Loh, H. W., Ooi, C. P., Dhok, S. G., Sharma, M., Bhurane, A. A., & Rajendra, A. U. (2022). Automated detection of cyclic alternating pattern and classification of sleep stages using deep neural network. Applied Intelligence, 52(3), 2903-2917.

International Journal of Medical Discoveries
Volume 2, Issue 1
ISSN: 3067-7912

Malhotra, R. K. (2024). AASM Scoring Manual 3: a step forward for advancing sleep care for patients with obstructive sleep apnea. Journal of Clinical Sleep Medicine, 20(5), 835-836.

Morokuma, S., Hayashi, T., Kanegae, M., Mizukami, Y., Asano, S., Kimura, I., Tateizumi, Y., Ueno, H., Ikeda, S., & Niizeki, K. (2023). Deep learning-based sleep stage classification with cardiorespiratory and body movement activities in individuals with suspected sleep disorders. Scientific reports, 13(1), 17730.

Muhammad, G., Almuntasheri, S., Alenezi, F., Alhadi, N., & Leung, V. C. (2025). EEG-based Multimodal Emotion Recognition: Recent Progress, Challenges, and Future Directions. ACM Transactions on Multimedia Computing, Communications and Applications.

Richman, J. S., & Moorman, J. R. (2000). Physiological time-series analysis using approximate entropy and sample entropy. American journal of physiology-heart and circulatory physiology, 278(6), H2039-H2049.

Samaee, M., Yazdi, M., & Massicotte, D. (2025). Multi-modal signal integration for enhanced sleep stage classification: Leveraging EOG and 2-channel EEG data with advanced feature extraction. Artificial Intelligence in Medicine, 103152.

Shani, N. S. J., & Moradi, M. H. (2023). Biomedical Signal Processing for Automated Detection of Sleep Arousals. Advances in Non-Invasive Biomedical Signal Sensing and Processing with Machine Learning, 263.

Sharma, M., Darji, J., Thakrar, M., & Acharya, U. R. (2022). Automated identification of sleep disorders using wavelet-based features extracted from electrooculogram and electromyogram signals. Computers in Biology and Medicine, 143, 105224.

Sharma, M., Yadav, A., Tiwari, J., Karabatak, M., Yildirim, O., & Acharya, U. R. (2022). An automated wavelet-based sleep scoring model using EEG, EMG, and EOG signals with more than 8000 subjects. International journal of environmental research and public health, 19(12), 7176.

Siddiqui, M. I. H., Sakib, A. H., Akter, S., Debnath, J., & Mahmud, M. R. (2025). Comparative analysis of traditional machine learning Vs deep learning for sleep stage classification. International Journal of Science and Research Archive [Internet], 1778-1789.

Soleimani, R., Barahona, J., Chen, Y., Bozkurt, A., Daniele, M., Pozdin, V., & Lobaton, E. (2023). Advances in modeling and interpretability of deep neural sleep staging: A systematic review. Physiologia, 4(1), 1-42.

Sun, H., Parekh, A., & Thomas, R. J. (2025). Artificial intelligence can drive sleep medicine. Sleep Medicine Clinics, 20(1), 81-91.

Van Der Donckt, J., Van Der Donckt, J., Deprost, E., Vandenbussche, N., Rademaker, M., Vandewiele, G., & Van Hoecke, S. (2023). Do not sleep on traditional machine learning: Simple and interpretable techniques are competitive to deep learning for sleep scoring. Biomedical Signal Processing and Control, 81, 104429.

Van der Lande, G. J., Blume, C., & Annen, J. (2022). Sleep and circadian disturbance in disorders of consciousness: current methods and the way towards clinical implementation. Seminars in neurology,

Walther, D., Viehweg, J., Haueisen, J., & Mäder, P. (2023). A systematic comparison of deep learning methods for EEG time series analysis. Frontiers in Neuroinformatics, 17, 1067095.

Wu, H. (2024). Multiscale entropy with electrocardiograph, electromyography, electroencephalography, and photoplethysmography signals in healthcare: A twelve-year systematic review. Biomedical Signal Processing and Control, 93, 106124.

Yazdi, M., Samaee, M., & Massicotte, D. (2024). A review on automated sleep study. Annals of Biomedical Engineering, 52(6), 1463-1491.

Zandbagleh, A., Sanei, S., Penalba-Sánchez, L., Rodrigues, P. M., Crook-Rumsey, M., & Azami, H. (2025). Intra-and inter-regional complexity in multi-channel awake EEG through multivariate multiscale dispersion entropy for assessing sleep quality and aging. Biosensors, 15(4), 240.

Zhai, B. (2023). Towards automated sleep stage assessment using ubiquitous computing technologies Newcastle University].