

Tidur merupakan komponen vital dari kehidupan manusia yang berperan dalam menjaga sistem kekebalan tubuh, pengolahan kognitif, dan metabolisme tubuh. Proses tidur yang berkualitas membantu pengoptimalan fungsi kognitif, pemulihan energi, dan kestabilan emosi pada tubuh. Kualitas tidur yang buruk dapat menyebabkan gangguan tidur. Gangguan tidur dapat berdampak serius pada kesehatan fisik dan mental seseorang jika tidak ditangani dengan cepat dan tepat, sehingga analisis kualitas tidur menjadi hal yang krusial. Salah satu metode yang umum digunakan untuk menilai kualitas tidur adalah elektroensefalografi (EEG), yang merekam aktivitas listrik otak selama tidur. Sinyal EEG diklasifikasikan ke dalam tahapan tidur secara otomatis menggunakan *machine learning*. Namun, dalam proses klasifikasi terdapat permasalahan yang mempengaruhi performa model klasifikasi, yaitu keberadaan fitur yang tidak relevan dan berlebihan dapat meningkatkan kompleksitas model dan menurunkan akurasi klasifikasi. Oleh karena itu, penelitian ini bertujuan untuk mengintegrasikan metode seleksi fitur dalam proses klasifikasi untuk meningkatkan performa model klasifikasi.

Penelitian ini menggunakan dataset rekaman tidur dari Sleep-EDF Expanded yang berisi sinyal EEG dari kanal Fpz-Cz dan Pz-Oz dari kelompok *Sleep Cassette* dan *Sleep Telemetry*. Penelitian ini mencakup empat tahapan utama, yaitu *preprocessing*, ekstraksi fitur, seleksi fitur, dan klasifikasi. *Preprocessing* mencakup dua proses, yaitu filtering menggunakan filter bandpass butterworth untuk menghilangkan noise dan segmentasi data. Ekstraksi fitur dilakukan terhadap domain waktu dan frekuensi. Pada tahap seleksi fitur, dilakukan perbandingan kinerja lima metode seleksi fitur, yaitu MRMR, Information gain, Fisher Score, Chi-square, dan Relief-f. Pada tahap klasifikasi, digunakan algoritma SVM dan Random Forest.

Evaluasi kinerja metode dilakukan dengan membandingkan akurasi klasifikasi sebelum dan sesudah penerapan metode seleksi fitur. Data dalam penelitian ini dibagi menjadi data latih dan data uji dengan rasio 80:20. Kinerja diukur berdasarkan akurasi yang dihasilkan oleh model SVM dan Random Forest. Hasil dari penelitian ini menunjukkan bahwa Information Gain memiliki akurasi tertinggi untuk dataset SC dengan akurasi 88,06% untuk model SVM dan Relief-F untuk dataset ST dengan akurasi 88,08% untuk model Random Forest.

**Kata kunci:** EEG, SVM, Random Forest, seleksi fitur, Klasifikasi tahapan tidur

## ABSTRACT

*Sleep is a vital component of human life that plays a role in maintaining the immune system, cognitive processing, and metabolism. Quality sleep helps optimize cognitive function, energy recovery, and emotional stability in the body. Poor sleep quality can lead to sleep disorders, Sleep disorders can have a serious impact on a person's physical and mental health if not treated quickly and appropriately, so analyzing sleep quality is crucial. One commonly used method to assess sleep quality is electroencephalography (EEG), which records the brain's electrical activity during sleep. EEG signals are classified into sleep stages automatically using machine learning. However, in the classification process, some problems affect the performance of the classification model, namely the presence of irrelevant and redundant features that can increase the complexity of the model and reduce the classification accuracy. Therefore, this research aims to integrate feature selection methods in the classification process to improve the performance of the classification model.*

*This study uses the Sleep-EDF Expanded sleep recording dataset which contains EEG signals from the Fpz-Cz and Pz-Oz channels from the Sleep Cassette and Sleep Telemetry groups. This research includes four main stages, namely preprocessing, feature extraction, feature selection, and classification. Preprocessing includes two processes, namely filtering using a Butterworth bandpass filter to remove noise and data segmentation. Feature extraction is performed on the time and frequency domains. In the feature selection stage, a comparison of the performance of five feature selection methods, namely MRMR, Information gain, Fisher Score, Chi-square, and Relief-f, was conducted. In the classification stage, SVM and Random Forest algorithms are used.*

*The performance evaluation of the method is done by comparing the classification accuracy before and after the application of the feature selection method. The data in this study is divided into training data and test data with a ratio of 80:20. Performance is measured based on the accuracy generated by the SVM and Random Forest models. The results of this study show that Information Gain has the highest accuracy for SC dataset with 88.06% accuracy for SVM model and F-Relief for ST dataset with 88.08% accuracy for Random Forest model.*

**Keywords :** EEG, SVM, Random Forest, feature selection, sleep stage classification