Problem Overview:
Electroencephalography (EEG) provides noninvasive insight into the brain by attaching electrodes to the scalp which record electrical activity. Leveraged correctly, it can accurately denote the stage of sleep for a person or test animal. Processing the data produced by EEG to obtain these results, however, is time-intensive and rather menial, proving an ideal fit for automation. In this summary we outline our proposed method and its implications in the context of single-channel EEG usage on rats.

Methodology & Findings:
The first step of creating any predictive algorithm, in this case one which can transform raw EEG readings into vigilance states, is data processing. We applied a smoothing function as the primary form of accomplishing this: the Butterworth filter, chosen for its common use in EEG data. This filter selectively attenuates (lessens the amplitude) of the individual frequencies comprising the observed EEG wave; we constrained this attenuation to isolate waves in the 0.5 Hz to 200 Hz range, which represents known rat brain wave frequencies. We then Fourier transformed the filtered wave to separate it into its component frequencies. The five greatest amplitudes were recorded along with their corresponding frequencies. We subsequently recorded the average amplitude for each category of brain wave (alpha, beta, delta, etc.) in each epoch.

We used the Random Forest model to predict the sleep states of the rats based on these recorded features. To assess the performance of our model, we divided our data into an 80:20 train/validation split and calculated F1 scores for each sleep stage. The F1 scores are as follows: Phase REM (0.72), Slow-wave Sleep (0.94), Wakefulness (0.93). The F1 scores indicate a high level of accuracy across different sleep stages, with a strong performance in Slow-wave Sleep and Wakefulness phases. Based on our model, we can conclude that the sleep state of rats can be reliably predicted using the decomposed frequencies of their EEG wave.

Challenges & Solutions:
We faced one primary challenge in predicting sleep state based purely on EEG readings: complexity. Both time and space limitations quickly arose with certain models and techniques, imposed on us by the sheer scale of the data. This pushed us to utilize less complex models which could not address issues we knew were prevalent. Employing XGBoost, rather than Random Forest, for example, because of its ability to better account for imbalanced data was not a feasible pursuit due to training time. Instead, we implemented data padding, which resulted in no noticeable improvement but an increase in execution time, so it was subsequently snubbed.

Conclusion & Future Directions:

Based on the F1 Scores provided, we believe that it is evident our model is reliable in predicting sleep states. Utilized properly, this study has the potential to offer valuable insights for pharmaceutical companies in the development of targeted drugs by revealing the impact different medications have on sleep stages. Additionally, our findings would have practical implications for hospitals if extrapolated for human subjects, offering better and more accurate diagnoses of sleep disorders, directly improving health outcomes and reducing inpatient hospitalization.

Future directions for this research include expanding data collection to include additional physiological measures, exploring more advanced models like Convolutional Neural Networks (CNNs), investigating patterns related to sleep disorders, and establishing a baseline level of concordance between annotators as a reference to compare our model against. Ultimately, we hope this research will prove fruitful when translating our findings into human clinical trials for a practical application in healthcare.