**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 solution and its implications in the context of single-channel EEG usage on rats.

**Methodology & Findings:**

In order to classify unlabeled EEG segments from a cohort of 8 rodents into vigilance states, we first processed raw EEG readings to prepare it for a predictive algorithm. A smoothing function was applied, 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 individual 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. 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 two significant challenges in predicting sleep state purely based on EEG readings: dimensionality reduction and transformation of electrical levels into meaningful information. Processing a dataset of size 8640×5001, even with the simplest of models, demands computational power so intense that it would not be reasonable to train, test and alter the model for several cycles (until it met expectations) within the timeframe. Converting the first 5000 columns into a wave and decomposing it into its individual frequencies resolved both issues, however, reducing the swaths of data down to only several columns which were biologically and statistically relevant.

**Conclusion & Future Directions:**

By implementing the Butterworth filter and Random Forest successfully, we have demonstrated that this process can be automated to reduce the burden of manual classification. The F1 scores achieved show that our model is reliable in predicting sleep states.

In looking ahead, we believe there are current applications and future directions of this research . This study offers valuable insights for pharmaceutical companies, proving its utility in the development of targeted drugs by understanding the impact of medications on different sleep stages. Additionally, our findings have a practical implication for hospitals, offering potential for better and more accurate diagnoses of sleep disorders, which can lead to improved patient outcomes and a reduction of inpatient hospitalization.

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