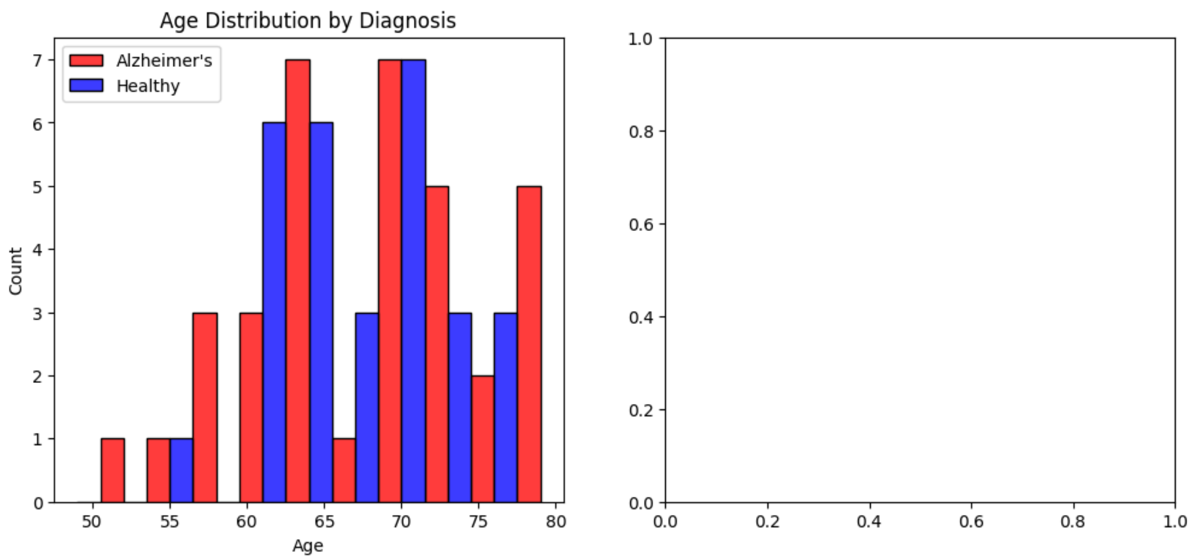
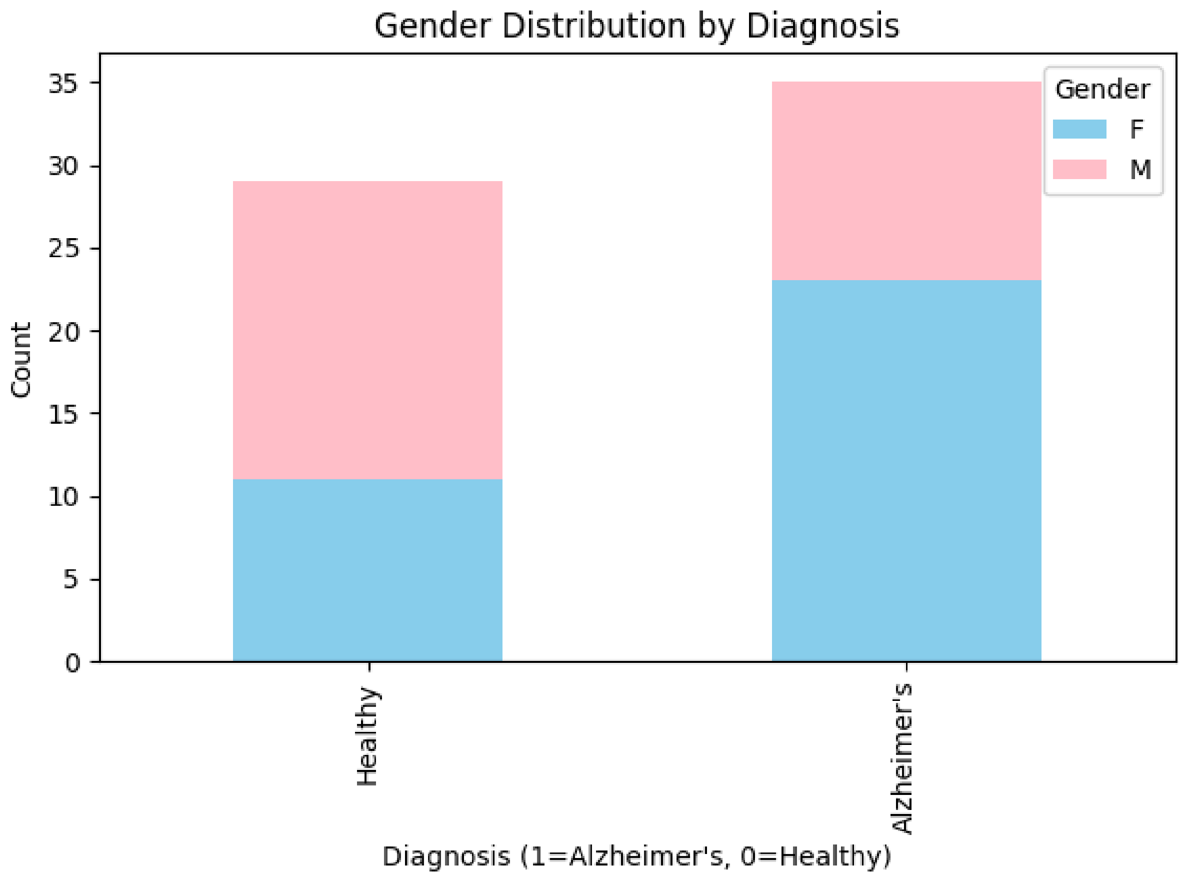
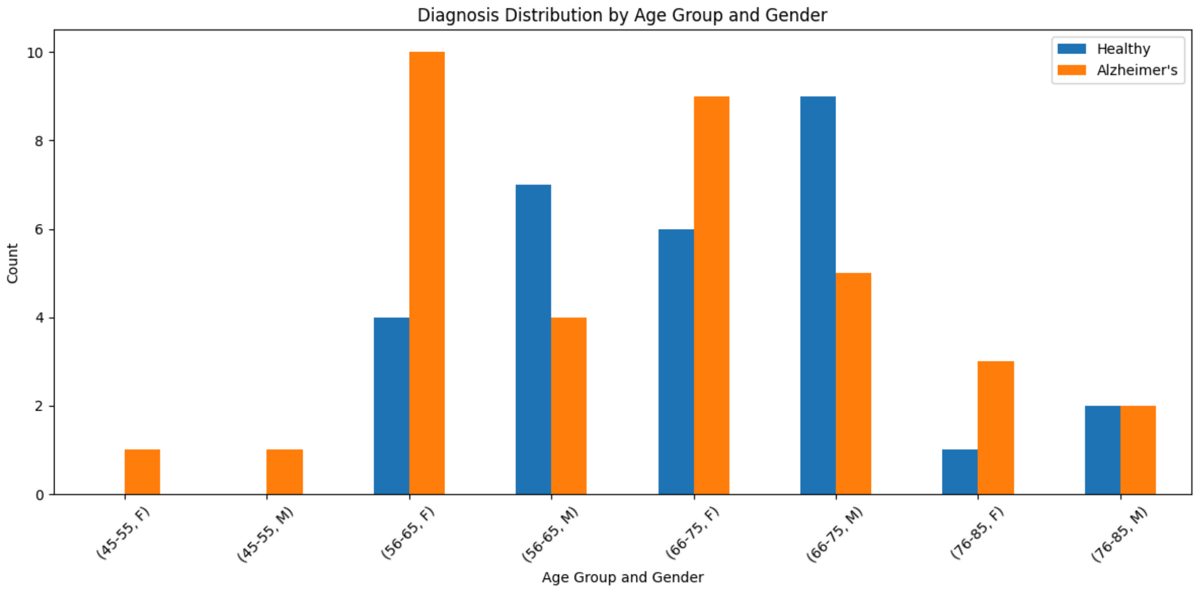
- Dataset Composition: The dataset consists of 64 samples and 15 features. The features include patient demographics (`Age`, `Gender`), an identifier (`patient_id`), the target variable (`label`), and several numerical features likely derived from biosignals (e.g., `log_band_power`, `std_dev`, `svd_entropy`).
- Label Distribution: The dataset includes 35 individuals diagnosed with Alzheimer's (label=1) and 29 control individuals (label=0). This indicates a slight imbalance, with Alzheimer's patients representing 54.69% of the dataset.
- Age Differences: Individuals diagnosed with Alzheimer's tend to be older, with an average age of 74.11 years, compared to the control group's average age of 69.48 years. The age range for Alzheimer's patients is 60 to 90 years, while for controls, it is 55 to 85 years.
- Gender Differences: There is a noticeable difference in gender distribution between the groups. The control group is predominantly male (62.1%), whereas the Alzheimer's group has a higher proportion of females (65.7%).



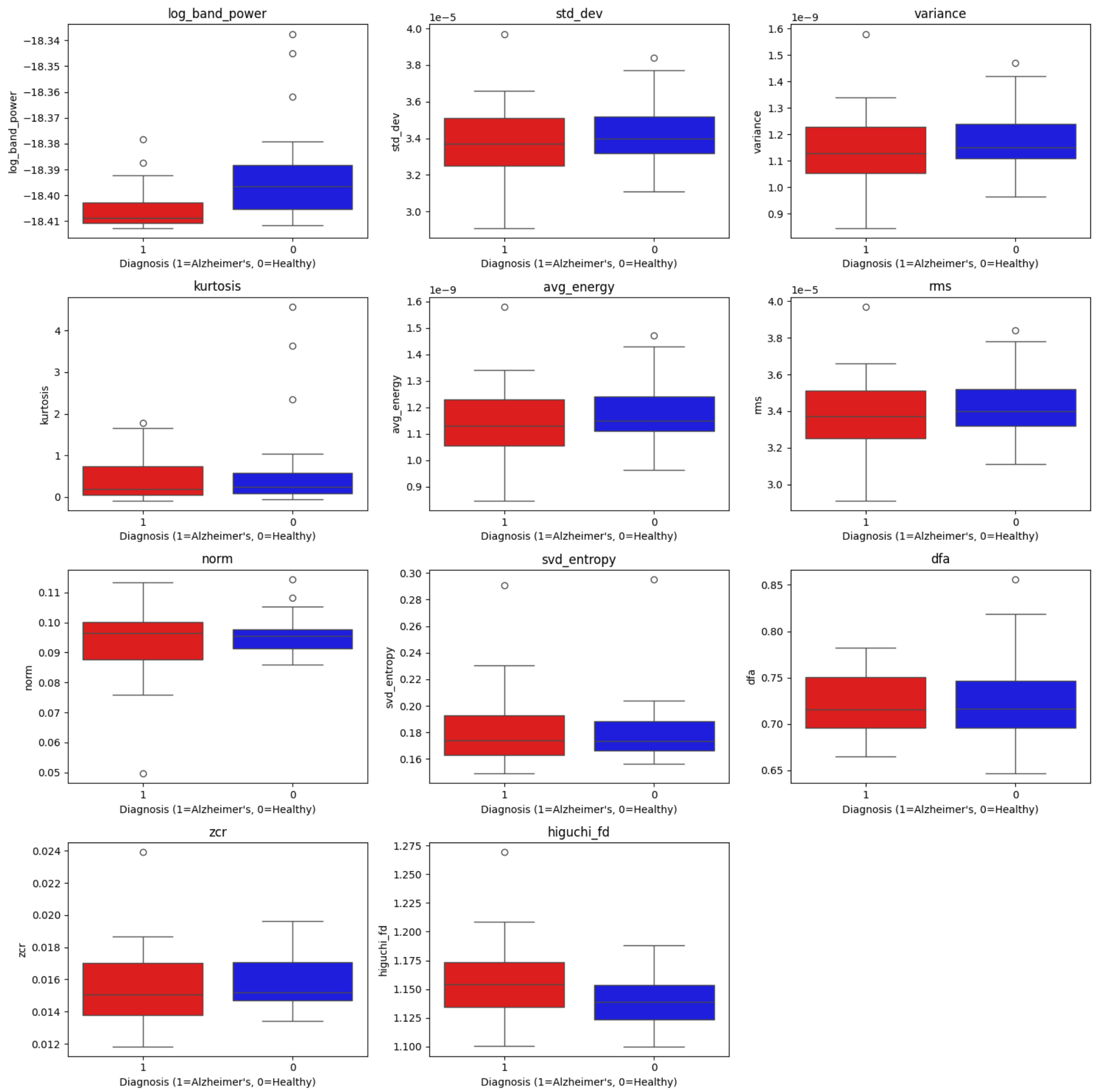
- Age Distribution: The histogram confirms that the Alzheimer's group generally includes older individuals compared to the healthy group[1]. The healthy group has peaks in the 60-65 and 65-70 age bins, while the Alzheimer's group shows peaks in the 60-65, 70-75, and 75-80 age bins[1].



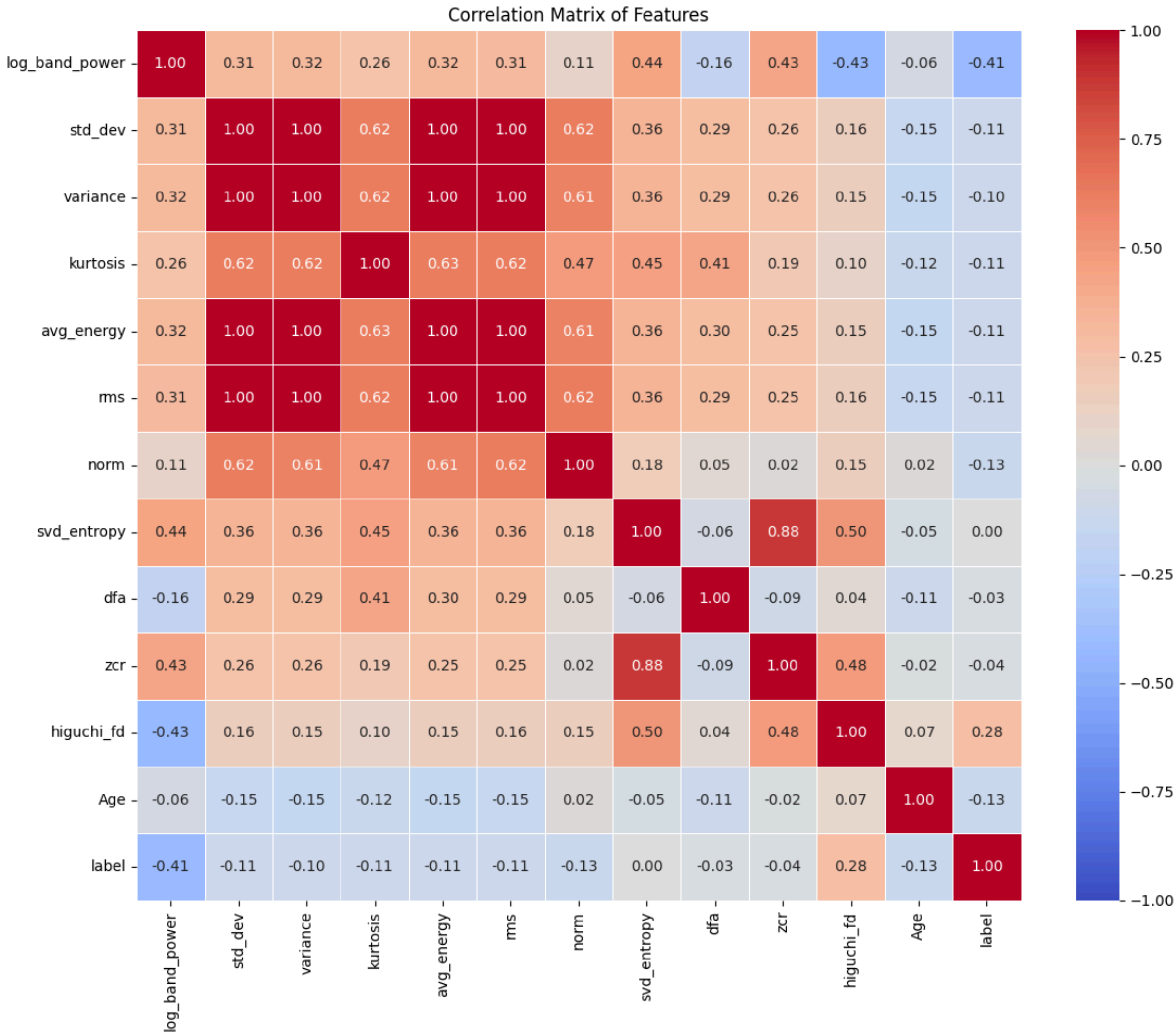
- Gender Distribution: This stacked bar chart clearly illustrates the gender disparity between the two diagnosis groups2. The healthy group has a higher proportion of males, while the Alzheimer's group has a significantly higher proportion of females2. This visually reinforces the previous cross-tabulation results.



- Diagnosis by Age Group and Gender : This chart provides a more granular view3:
    - Among females, Alzheimer's cases are notably high in the 56-65 and 66-75 age groups compared to healthy females in the same age ranges3.
    - Among males, healthy individuals outnumber Alzheimer's cases in the 56-65 and 66-75 age groups3.
    - In the oldest age groups (76-85), both genders show a presence of Alzheimer's cases, though the sample size appears smaller in these brackets3.

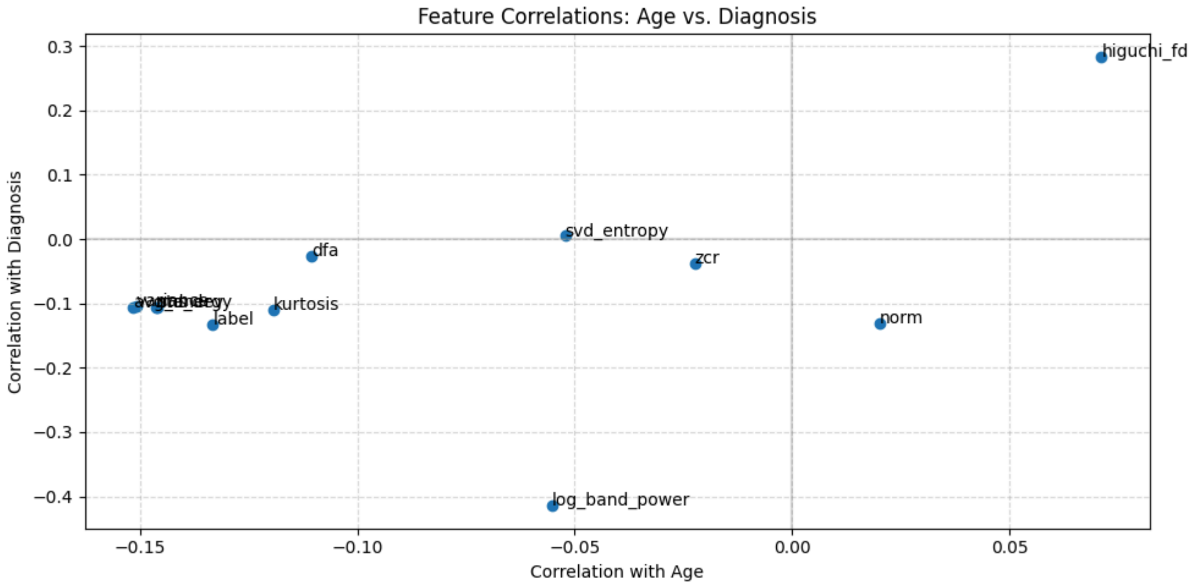


- Significant Differences: Only two features show a statistically significant difference between the groups ($p < 0.05$):
    - log_band_power: The mean `log_band_power` is significantly lower in the Alzheimer's group (-18.406) compared to the Healthy group (-18.393), with a p-value of 0.001854. The box plot for `log_band_power` visually confirms this, showing the distribution for the Alzheimer's group (red box) shifted slightly lower than the Healthy group (blue box)1.
    - higuchi_fd: The mean `higuchi_fd` (Higuchi Fractal Dimension) is significantly higher in the Alzheimer's group (1.157) compared to the Healthy group (1.140), with a p-value of 0.018912. The corresponding box plot also shows the distribution for the Alzheimer's group positioned slightly higher1.
- Non-Significant Differences: The majority of the features analyzed (9 out of 11) do *not* show a statistically significant difference between the Alzheimer's and Healthy groups ($p > 0.05$). These features include `std_dev`, `variance`, `kurtosis`, `avg_energy`, `rms`, `norm`, `svd_entropy`, `dfa`, and `zcr`.
- Visual Confirmation: The box plots for the non-significant features generally show substantial overlap between the distributions of the two groups (red and blue boxes)1. The median lines and interquartile ranges are quite similar, visually supporting the lack of statistically significant differences reported in the table.



- High Multicollinearity: Several features show perfect or near-perfect positive correlation (correlation coefficient of 1.00). Specifically, `std_dev`, `variance`, `avg_energy`, and `rms` are perfectly correlated with each other. This suggests they measure very similar aspects of the data and are likely redundant. Using all of them in a model could cause multicollinearity issues.
- Strong Feature Intercorrelations:
  - There is a very strong positive correlation between `svd_entropy` and `zcr` (0.88).
  - Moderate positive correlations exist between `kurtosis`, `norm`, and the group of (`std_dev`, `variance`, `avg_energy`, `rms`) (coefficients ranging from 0.47 to 0.63).
  - `higuchi_fd` shows moderate positive correlations with `svd_entropy` (0.50) and `zcr` (0.48).
- Correlations with Target Variable (`label`):

    - `log_band_power` exhibits the strongest correlation with the `label`, showing a moderate negative relationship (-0.41). This implies lower `log_band_power` values are associated with the Alzheimer's group (label=1).
    - `higuchi_fd` shows a moderate positive correlation with the `label` (0.28), suggesting higher values are somewhat associated with Alzheimer's.
    - Most other features, including `Age`, show weak correlations with the `label` (coefficients typically between -0.13 and 0.00).
- Notable Negative Correlations: `log_band_power` also shows moderate negative correlations with `svd_entropy` (-0.44), `zcr` (-0.43), and `higuchi_fd` (-0.43).
- Age: `Age` has very weak correlations with all other features and the `label`, indicating it doesn't strongly relate linearly to the other measured variables in this dataset.



- Key Diagnosis Correlates: `log_band_power` and `higuchi_fd` show the strongest correlations with the diagnosis label`1`.
    - `log_band_power` has a strong negative correlation (~ -0.41) with the diagnosis, meaning lower values are associated with Alzheimer's`1`. It also has a slight negative correlation with age`1`.
    - `higuchi_fd` has the strongest positive correlation (~ +0.28) with the diagnosis, indicating higher values are associated with Alzheimer's`1`. It shows a slight positive correlation with age`1`.

- Age Correlation: Most features exhibit only a weak correlation with `Age`, with most showing a slight negative correlation (points lie to the left of the vertical axis)1.
- Other Features: Features like `svd_entropy`, `zcr`, `dfa`, and `norm` cluster closer to the center, indicating weaker correlations with both age and diagnosis1. The highly intercorrelated features (`avg_energy`, `std_dev`, `variance`, `rms`) and `kurtosis` show weak negative correlations with both age and diagnosis1.
- Overall Pattern: The plot highlights `log_band_power` and `higuchi_fd` as potentially the most useful features for distinguishing the diagnosis groups, as they lie furthest from the horizontal axis (y=0). The relationship with age appears generally weak across most features1.

MRISHIKA

Insights from the ANOVA F-value Feature Selection

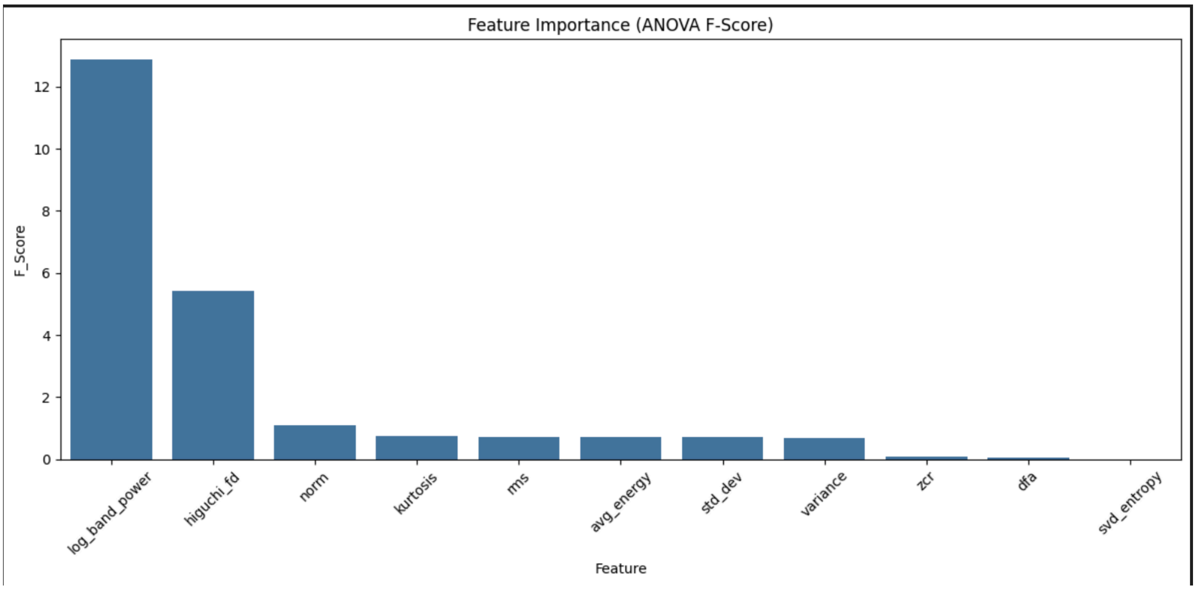
1. Feature Importance Ranking:
    - The feature `log_band_power` has the highest ANOVA F-score (~12.89) with a highly significant p-value (~0.00065).
        - Interpretation: This feature shows the strongest statistical difference between Alzheimer's and healthy groups, making it the most important feature for distinguishing the two.
    - The second most important feature is `higuchi_fd` (Higuchi fractal dimension) with a moderate F-score (~5.42) and a significant p-value (~0.023).
        - Interpretation: This suggests that fractal complexity of the signal differs significantly between the groups and is a relevant biomarker.
2. Other Features:
    - Features like `norm`, `kurtosis`, `rms`, `avg_energy`, `std_dev`, and `variance` have relatively low F-scores (around 0.6 to 1.1) and non-significant p-values (all > 0.29).

- Interpretation: These features do not show strong evidence of differing between Alzheimer's and healthy groups and thus are less useful for classification.

- Features such as `zcr`, `dfa`, and `svd_entropy` have very low F-scores close to zero and very high p-values (close to 1), indicating no meaningful difference between groups.

3. Overall Pattern:
   - The ANOVA F-test clearly identifies two features (`log_band_power` and `higuchi_fd`) as the most discriminative.
   - The majority of other features do not significantly contribute to distinguishing Alzheimer's from healthy controls based on this univariate test.
4. Visualization:
   - The bar plot visually confirms the dominance of `log_band_power` and `higuchi_fd` in terms of their F-scores.
   - The steep drop after these two features suggests a natural cutoff for feature selection.

Feature	F-Score	P-Value	Insight
log_band_power	12.89	0.00065	Strongest discriminator; key biomarker.
higuchi_fd	5.42	0.023	Significant fractal complexity difference.
norm	1.09	0.29	Not statistically significant.
kurtosis	0.75	0.39	Not statistically significant.
rms	0.71	0.40	Not statistically significant.
avg_energy	0.70	0.40	Not statistically significant.
std_dev	0.70	0.40	Not statistically significant.
variance	0.69	0.49	Not statistically significant.
zcr	0.08	0.77	No difference.
dfa	0.04	0.83	No difference.
svd_entropy	0.001	0.96	No difference.

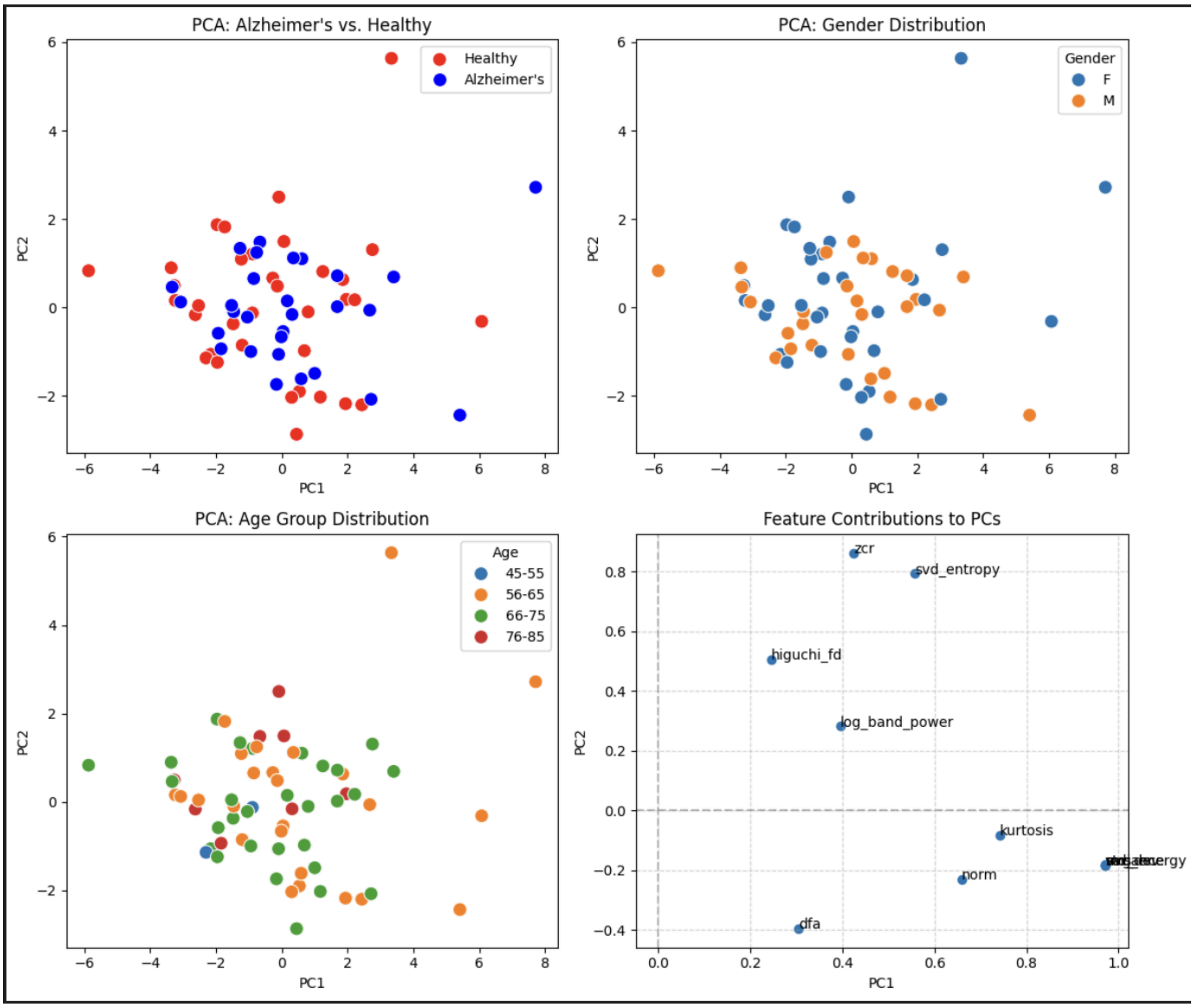


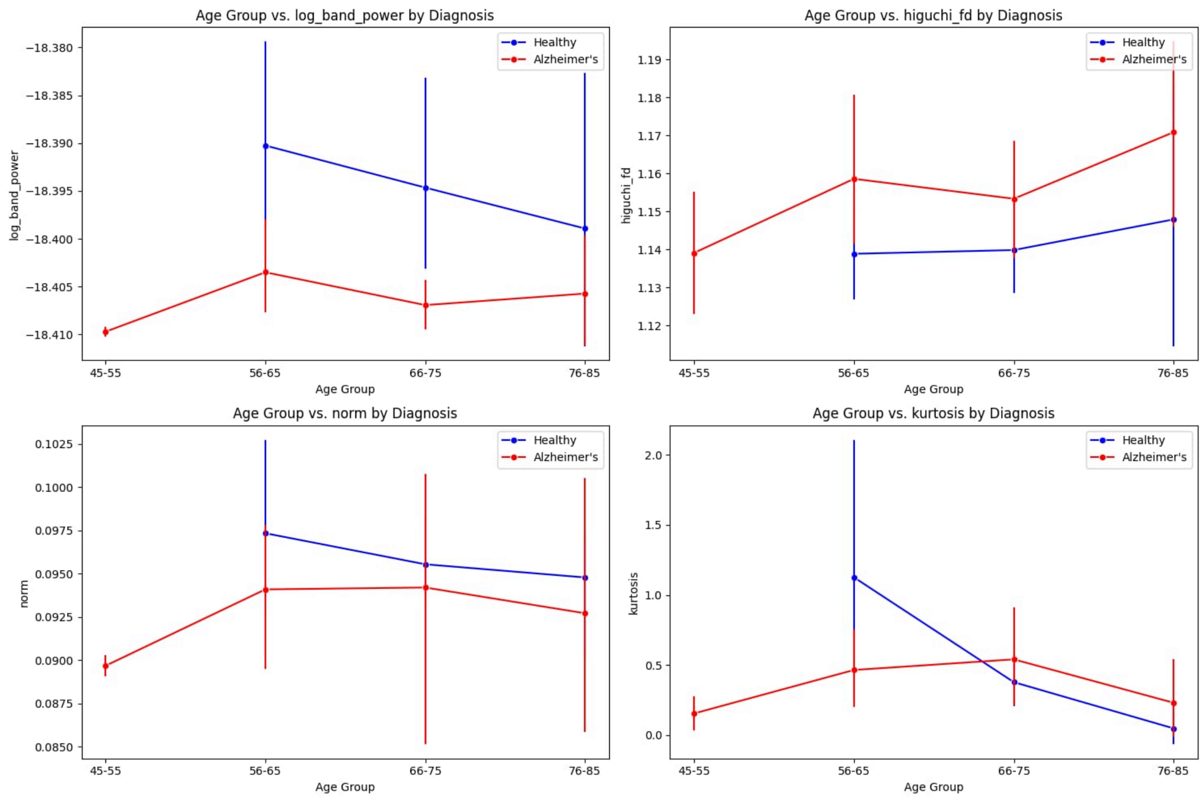
Principal Component Analysis (PCA) Insights

1. Variance Explained
    - The first two principal components (PC1 and PC2) together explain about 68% of the total variance in the dataset.
    - This means most of the important information from all features is captured in these two components.
2. Key Feature Contributions
    - PC1 is mainly influenced by features related to signal energy and spread:
      `avg_energy`, `std_dev`, `variance`, and `rms` all have very high positive contributions.
    - PC2 is mostly influenced by features related to signal complexity:
      `zcr`, `svd_entropy`, and `higuchi_fd` show the highest positive contributions.
3. Group Separation
    - The PCA scatter plot for Alzheimer's vs. Healthy shows some separation, but there is still overlap between the two groups.
    - This indicates that while the selected features help distinguish the groups, they do not provide perfect separation.
4. Demographic Effects

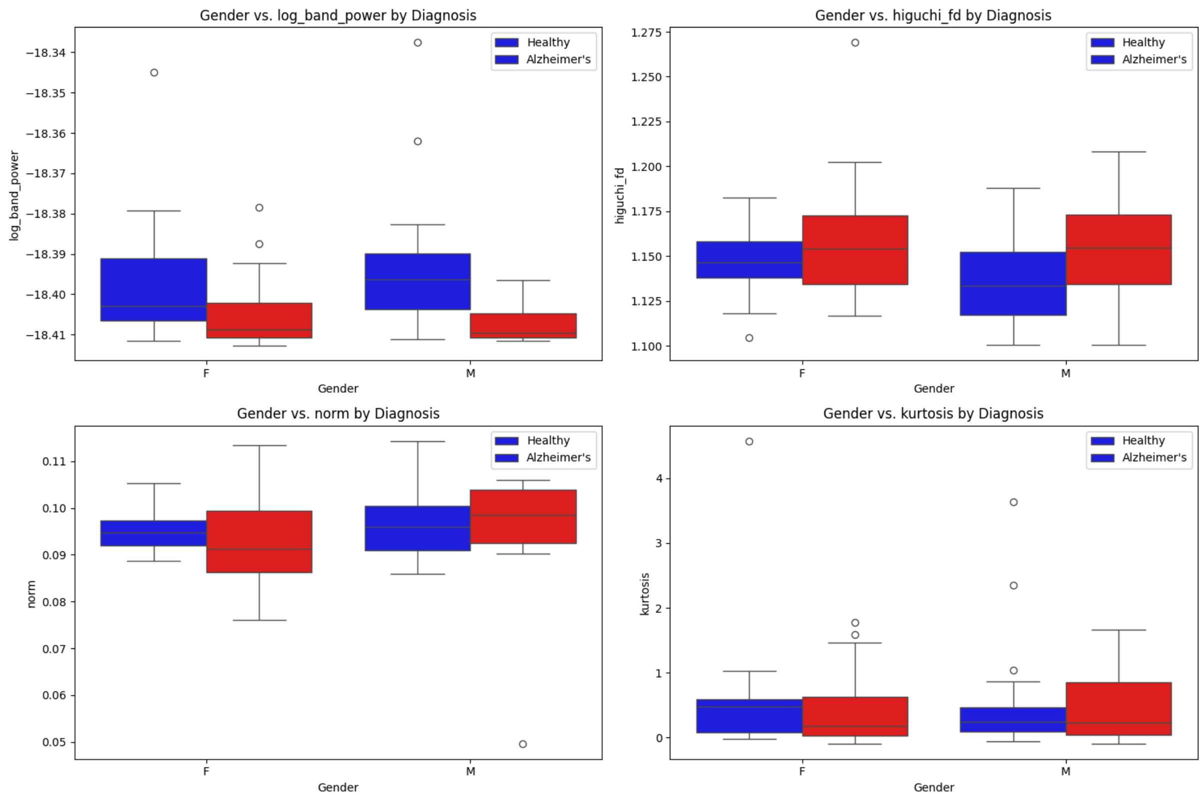
- The gender and age group scatter plots show that neither gender nor age group causes strong clustering in the PCA space.
- This suggests that the main patterns in the data are not dominated by gender or age.

5. Feature Interpretation
    - The “Feature Contributions to PCs” plot visually confirms which features most influence each principal component.
    - Signal energy features are most important for PC1, while complexity features are most important for PC2.





- Gender Differences in log_band_power: There is a statistically significant difference in `log_band_power` between healthy and Alzheimer's patients within the male group (p-value = 0.0041)`1`. However, this difference is not significant for females (p-value = 0.1259)`1`. This suggests that `log_band_power` may be a more useful biomarker for distinguishing Alzheimer's in males than in females.
- Age Group Differences in log_band_power: The `log_band_power` metric shows statistically significant differences between healthy individuals and Alzheimer's patients in the age groups 56-65 and 66-75 (p-value = 0.0360 and 0.0365, respectively)`1`. This indicates that `log_band_power` could be a relevant marker for Alzheimer's diagnosis in these age ranges. The age group 45-55 has insufficient samples for statistical testing`1`



Key Insights from EEG Metrics in Alzheimer's Disease

- **Log Band Power**:
  - Alzheimer's patients show consistently lower log_band_power values than healthy individuals across both genders
  - The separation between healthy and Alzheimer's patients appears more pronounced in males
  - Several outliers present in healthy groups suggest higher variability in normal brain function
- **Higuchi Fractal Dimension (FD)**:
  - Alzheimer's patients exhibit higher higuchi_fd values compared to healthy controls in both genders
  - The difference is consistent and visible in both gender groups
  - Alzheimer's patients show greater variability (wider boxes) in higuchi_fd values
- **Norm Metric**:

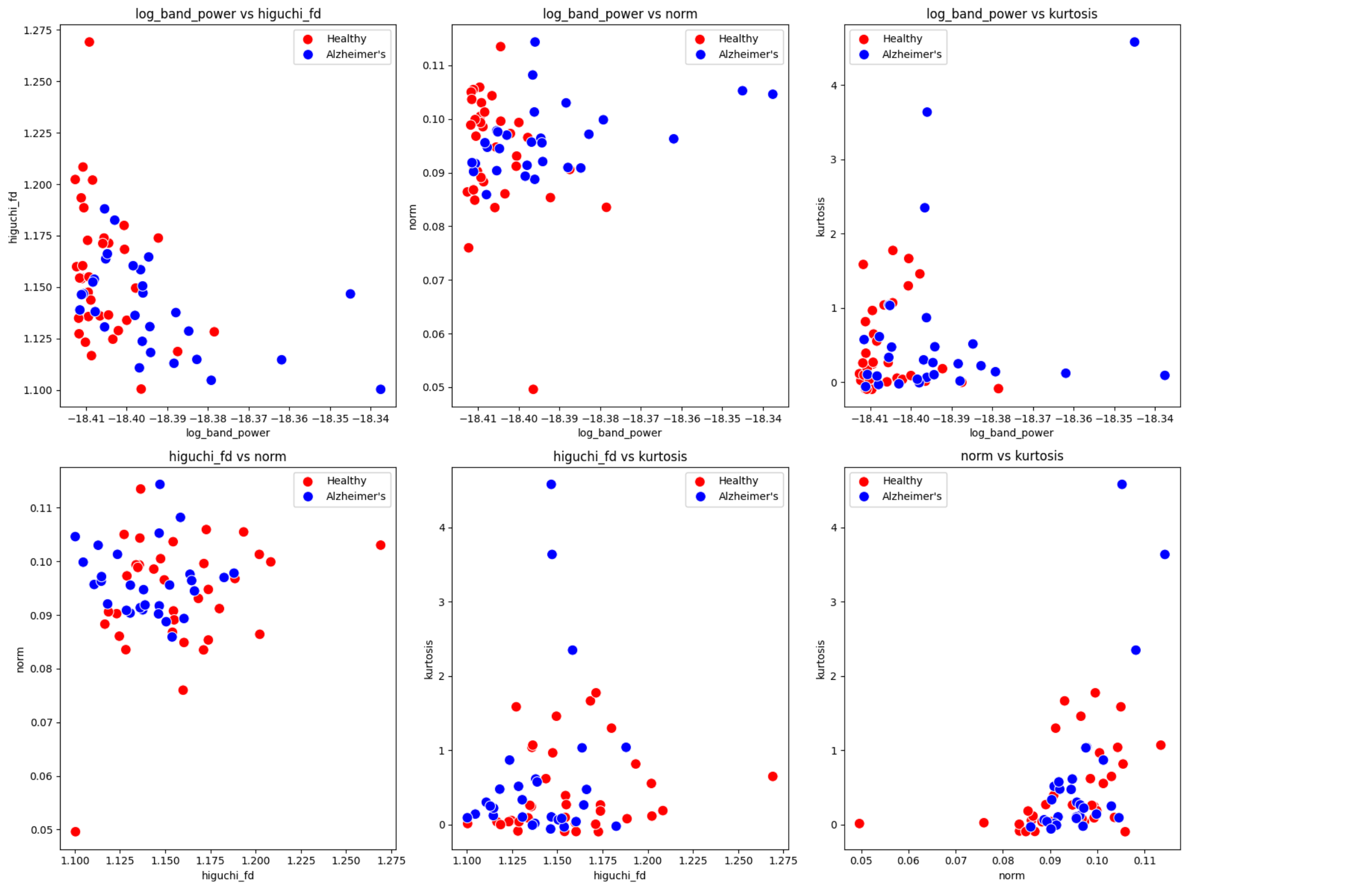
    - Male Alzheimer's patients display higher norm values and wider distribution than healthy males
    - Female Alzheimer's patients show greater variability in norm values compared to healthy females
    - There is less clear separation between groups in this metric compared to log_band_power
- **Kurtosis**:
    - Alzheimer's patients tend to have slightly higher kurtosis values across both genders
    - Multiple outliers present, especially in female groups
    - Less distinct separation between healthy and Alzheimer's groups compared to other metrics
- **Gender-Specific Patterns**:
    - Male patients show clearer distinction in log_band_power values between healthy and Alzheimer's groups
    - Females display more outliers across various metrics
    - Gender differences appear most pronounced in the log_band_power metric
- **Diagnostic Potential**:
    - Log_band_power appears most promising for distinguishing Alzheimer's patients, particularly in males
    - Higuchi_fd shows consistent patterns worth further investigation
    - Combined analysis of multiple metrics may provide more robust diagnostic capabilities than any single measure

Stratified Analysis By The Gender

Key Insights from Alzheimer's vs. Healthy EEG Analysis

- **Log Band Power (Strongest Biomarker):**
    - **Lower values in Alzheimer's patients across both genders1**
    - **Statistically significant in males (p=0.0041) but not females (p=0.1259)**
    - **Clear separation between healthy and Alzheimer's groups in males1**
    - **Also significant in age groups 56-65 (p=0.0360) and 66-75 (p=0.0365)**
- **Higuchi Fractal Dimension:**
    - **Consistently higher values in Alzheimer's patients for both genders1**
    - **Approaches significance in males (p=0.0980)**
    - **Greater variability (wider boxplots) in Alzheimer's patients1**
    - **Possible biomarker despite lacking statistical significance**
- **Norm Metric:**
    - **Alzheimer's patients show slightly higher values, especially in males1**
    - **Greater variability in female Alzheimer's patients1**
    - **Not statistically significant in any gender or age group**
    - **Less diagnostic value than log_band_power**
- **Kurtosis:**
    - **Slightly higher values in Alzheimer's patients1**
    - **Multiple outliers present, especially in females1**
    - **Not statistically significant across any gender or age group**
    - **Shows substantial overlap between healthy and Alzheimer's groups**
- **Gender-Specific Findings:**
    - **Males show stronger Alzheimer's-related differences in log_band_power1**
    - **Females display more outliers across multiple metrics1**

  - **Log_band_power appears most promising for male-specific diagnosis**
  - **Gender-stratified analysis recommended for future studies**
- **Clinical Implications:**
  - **Log_band_power is the most promising single biomarker**
  - **Males ages 56-75 show the most reliable EEG differences**
  - **Multi-metric approaches may improve diagnostic accuracy**
  - **Age and gender should be considered in diagnostic criteria**




```
# Multivariate logistic regression analysis
```

Key Insights from Logistic Regression Model

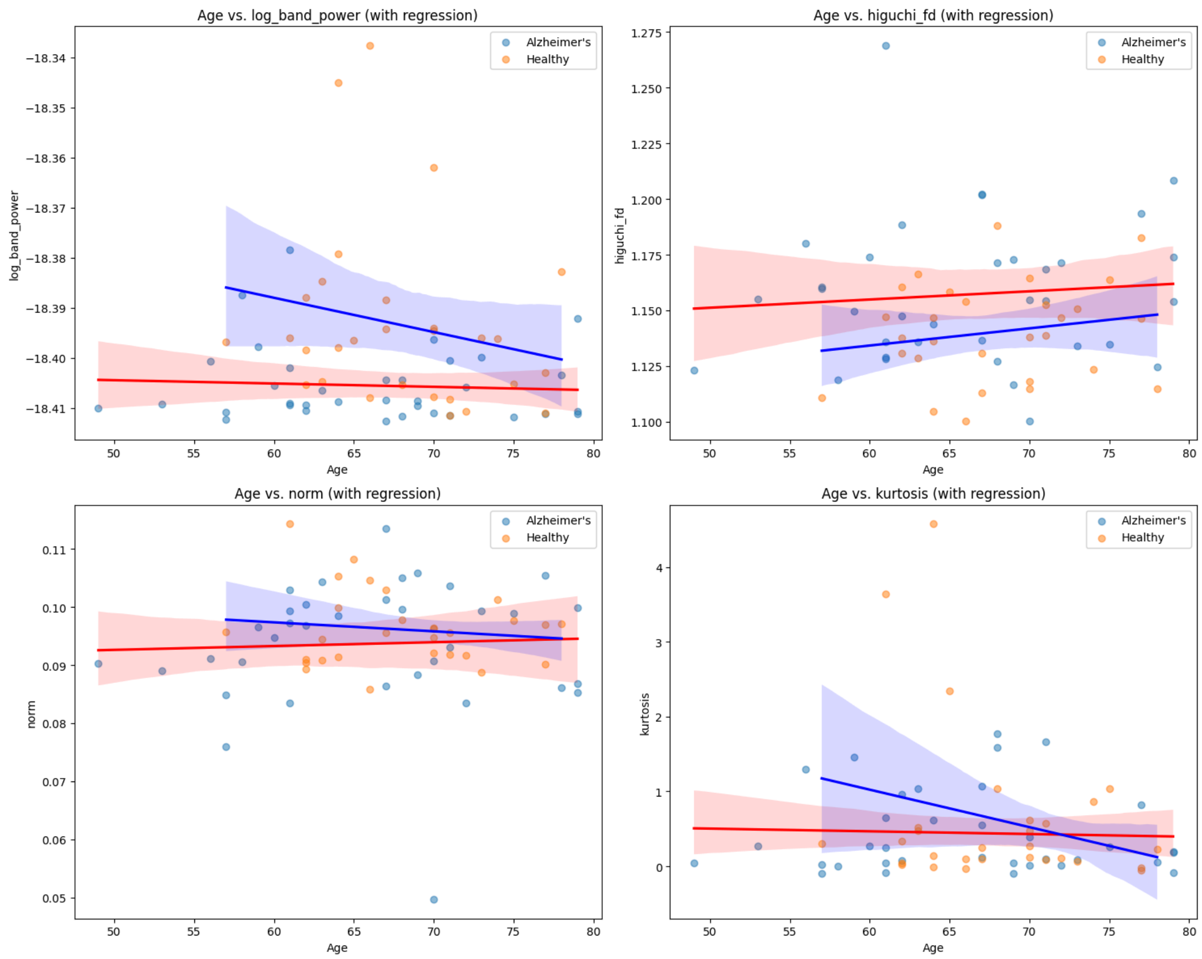
- **Model Performance:**

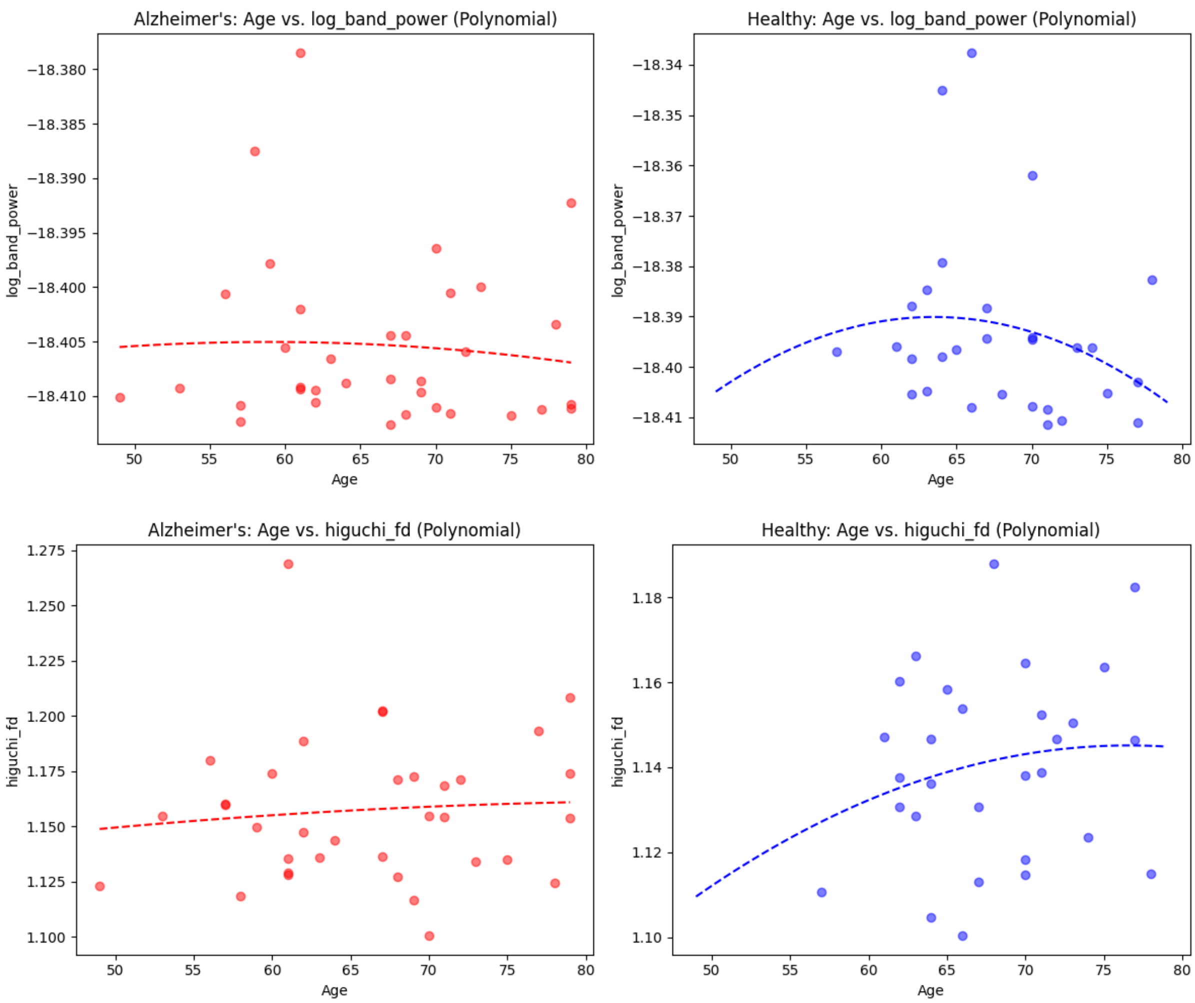
- **Moderate predictive power (Pseudo R-squared: 0.281)**
- **Failed to converge (Converged: 0.0000) after 35 iterations**
- **Marginally significant overall (LLR p-value: 0.073936)**
- **Small sample size (44 observations) limiting statistical power**

- **Significant Predictors:**
  - **Only x1 is statistically significant (p=0.0481)**
  - **x1 has negative association with Alzheimer's (coef: -3.6199)**
  - **For each unit increase in x1, odds of Alzheimer's decrease by 97.3% (odds ratio: 0.027)**
- **Potential Numerical Issues:**
  - **Several variables (x2, x3, x5, x6) show extremely large coefficients**
  - **Missing standard errors and p-values for these variables indicate perfect separation**
  - **Extremely large odds ratios (7.66×10^35 for x2, 4.44×10^55 for x3)**
  - **These suggest multicollinearity or complete separation problems**
- **Variable Relationships:**
  - **x8 shows positive association with Alzheimer's (coef: 3.1260, p=0.1560)**
  - **Most variables (x4, x7, x9, x10, x11) show negative but non-significant associations**
  - **x11 shows substantial negative effect (odds ratio: 0.221) but not significant (p=0.2121)**
- **Model Improvement Needed:**
  - **Consider feature selection to address multicollinearity**
  - **Larger sample size would improve reliability**
  - **Ridge or LASSO regularization could help stabilize estimates**
  - **Consider removing or combining highly correlated predictors**

- **Scatterplot Patterns:**
    - **Log_band_power differentiates groups with lower values in Alzheimer's patients**
    - **Some Alzheimer's cases show distinctly higher kurtosis values**
    - **Most effective visual separation appears in the log_band_power vs higuchi_fd plot**

Non Linear Relationships





Key Non-Linear Relationships in Alzheimer's EEG Analysis

- **Age vs. Log Band Power:**
  - **Healthy individuals show a pronounced inverted U-shaped relationship with peak values around ages 60-6545**
  - **Alzheimer's patients exhibit a relatively flat relationship with age, lacking the peak seen in healthy subjects4**
  - **This non-linear pattern suggests normal aging has different spectral power dynamics than pathological aging**
- **Age vs. Higuchi Fractal Dimension:**

- Healthy subjects show a distinct curvilinear relationship with increasing values that plateau around age 705
- Alzheimer's patients display a gentler positive slope with more variability across all ages5
- This difference highlights altered complexity in neural dynamics with disease progression

- Log Band Power vs. Higuchi FD:
  - Shows the clearest non-linear separation between groups among all paired metrics1
  - Lower complexity (lower higuchi_fd) combined with lower power appears characteristic of Alzheimer's cases17
  - This relationship aligns with research showing AD patients have "less chaotic" brain activity69

- Kurtosis Relationships:
  - Several Alzheimer's patients show extremely high kurtosis values (>3) not seen in healthy subjects12
  - Kurtosis decreases with age in Alzheimer's patients but remains relatively stable in healthy controls3
  - This non-linear pattern suggests potential peakedness abnormalities in signal distribution