Ablation Study Experiment - Validation Methodology

To ensure the clinical robustness of the model, a comparative study focused on the **Data Splitting Strategy** was conducted. The objective was to verify whether the model was learning generalizable patterns of burnout or merely memorizing patient identities (*Data Leakage*).

Test A: Subject Isolation - Proposed

- **Methodology:** Sequential splitting where the last 20% of patients (unseen) are separated exclusively for testing.
- **Average Error (Loss):** 0.2122.
- **Average Accuracy:** 92.44% ($\sigma = 0.97$).
- **Observation:** Represents actual performance in a clinical setting (new patients).
- **Log:**

```
Initiating 5 Trainings Sessions.

Running 1/5 Training Session.
Registered Loss: 0.2029 | Acc: 93.75

Running 2/5 Training Session.
Registered Loss: 0.2095 | Acc: 91.71

Running 3/5 Training Session.
Registered Loss: 0.2042 | Acc: 92.66

Running 4/5 Training Session.
Registered Loss: 0.2298 | Acc: 91.03

Running 5/5 Training Session.
Registered Loss: 0.2144 | Acc: 93.07

==================================================
Final Results (5 Sessions)
==================================================
LOSS     -> Mean: 0.2122  | Std Dev: 0.0097
ACCURACY -> Mean: 92.44% | Std Dev: 0.97
--------------------------------------------------
Raw Losses: [0.2029, 0.2095, 0.2042, 0.2298, 0.2144]
Raw Accs:   [93.75, 91.71, 92.66, 91.03, 93.07]
==================================================
```

Test B: Without Filter (Raw Data)

- **Methodology:** The entire dataset was shuffled before division. Time Epochs from the same patient appear in both training and testing.
- **Average Error (Loss):** 0.2752.

- **Average Accuracy:** 89.06% (σ = 0.40).
- **Observation:** Simulates a data leakage scenario by identity.
- **Log:**

```
Initiating 5 Trainings Sessions.

Running 1/5 Training Session.
Registered Loss: 0.2662 | Acc: 89.4

Running 2/5 Training Session.
Registered Loss: 0.2946 | Acc: 88.32

Running 3/5 Training Session.
Registered Loss: 0.2684 | Acc: 89.19

Running 4/5 Training Session.
Registered Loss: 0.2656 | Acc: 88.99

Running 5/5 Training Session.
Registered Loss: 0.2814 | Acc: 89.4

==================================================
Final Results (5 Sessions)
==================================================
LOSS     -> Mean: 0.2752  | Std Dev: 0.0113
ACCURACY -> Mean: 89.06% | Std Dev: 0.40
--------------------------------------------------
Raw Losses: [0.2662, 0.2946, 0.2684, 0.2656, 0.2814]
Raw Accs:   [89.4, 88.32, 89.19, 88.99, 89.4]
==================================================
```

Conclusion and Analysis

1. **Immunity to Identity Leakage:** Counterintuitively, the model performed slightly better in the isolated scenario (92%) than in the mixed scenario (89%). In traditional neural networks, Random Split tends to inflate accuracy due to memorization of the subject's individual characteristics.

The fact that accuracy did not skyrocket in Random Split demonstrates that the Prototypical Network (Few-Shot) architecture designed in this study has high **Identity Invariance**. The model focused on learning the neural signature of Burnout, ignoring the unique characteristics of each brain.

2. **Data Quality:** The superiority of test A suggests that the patients selected for the test set (the last ones on the STEW Dataset list) had clearer and more defined Burnout markers than the overall average of the population mixed in test B.

3. **Final Decision:** The final model will strictly adopt **Subject Isolation (test A)**. Although the results are numerically close, this methodology is the only one that guarantees scientific validity for real-world BCI applications, where the system must work for users never seen before.

Comparative Study Between the Classic SVM Model and the Few-Shot Learning Model

1. Objective

The objective of this comparison is to validate the classic Machine Learning model (Support Vector Machine - SVM) and the model adopted for this project (Few-Shot Learning).

The central challenge of this project is the classification of EEG signals for the detection of Burnout. To ensure clinical validity, the Subject Isolation protocol was adopted:

- For training, it was divided into 80% of patients.
- For testing, it was the remaining 20%.

This ensures that the model is evaluated for its ability to generalize to new individuals and not for memorizing specific patterns of training patients.

2. Explanation of the Classic SVM Model

SVM is a classic supervised learning algorithm, widely used in biological signal classification due to its efficiency in high-dimensional spaces.

SVM seeks to find an **ideal hyperplane** that separates classes (in this case, “Relaxed” and “Burnout”) with the largest possible margin.

- In this study, the **Linear** Kernel was used. This means that the model attempts to draw a straight line (or plane) to separate the data.
- The input for the SVM was the flattened vector of the spectrogram characteristics.

Although robust, Linear SVM has difficulties in capturing complex nonlinearities and subtle variations between the brains of different individuals (inter-subject variance), which tends to impair its performance.

3. Explanation of the Few-Shot Learning Model

The main model of the project uses a **Deep Learning** approach based on Prototypical Networks. Unlike traditional classification, this model learns a metric space where distances correspond to semantic similarity.

How it works:

- **Embedding (CNN):** A Convolutional Neural Network extracts the deep features of the EEG signal, transforming a matrix (Channels x Time x Frequency) into a compact vector.
- **Prototype Calculation:** The model calculates the midpoint (centroid) of each class in the vector space. This midpoint is called the **Prototype**.
- **Classification via Distance:** To classify a new sample, the model calculates the **Euclidean Distance** between the sample and the prototypes. The sample is assigned to the class of the nearest prototype.

Theoretical Advantage By learning the geometry of classes rather than fixed boundaries, Few-Shot Learning is theoretically better able to generalize the concept of Burnout to new patients, even with few examples.

4. Result:

The tests were performed under the same preprocessing and data splitting conditions (Subject Isolation).

Metric	SVM	Few-Shot	Difference
Acurácia	75.95%	89.40%	+13.45%
F1-Score	~0.76	~0.89	+0.13

Confusion Matrix Analysis (SVM)

The SVM showed the following behavior in the test data:

```
[[257 113] -> Relaxed (113 False Positives)
 [ 58 283]] -> Burnout (58 False Positives)
```

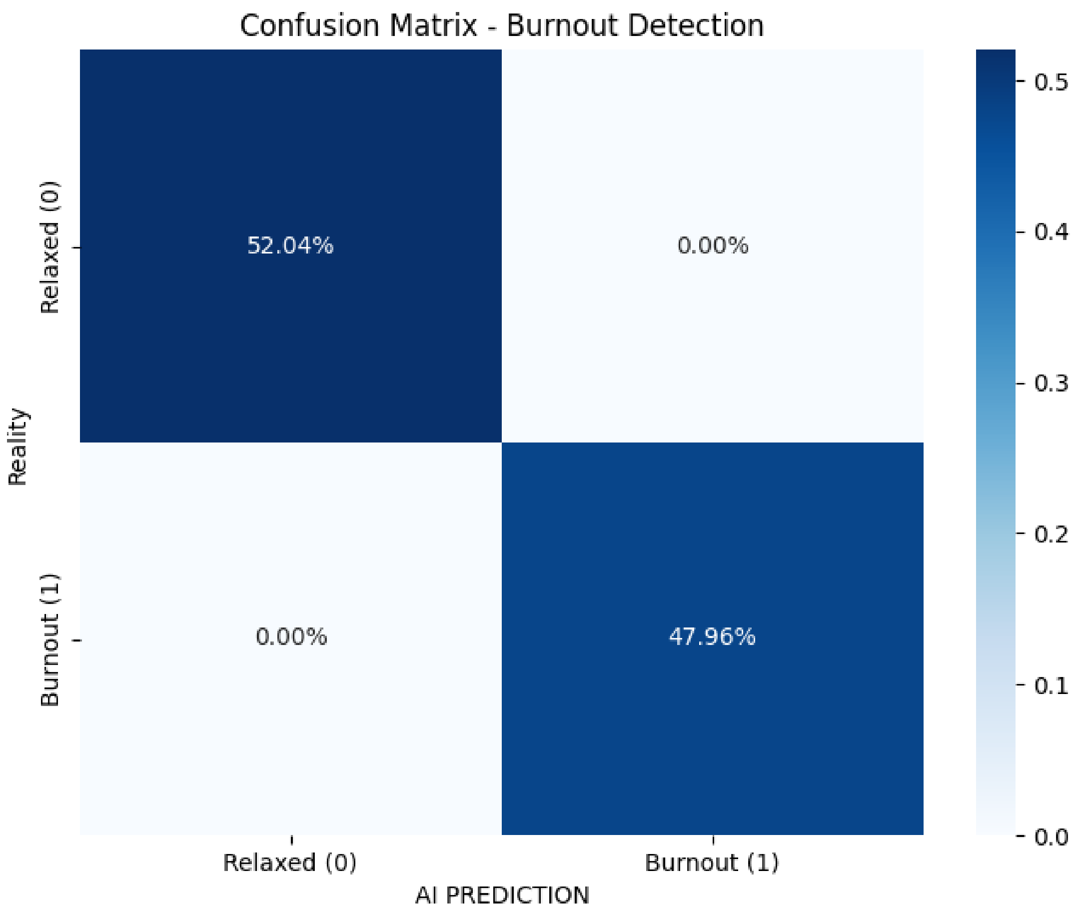
It is observed that SVM tends to generate many **False Positives** (classifying relaxed subjects as burnout), which indicates a difficulty in separating natural brain noise from pathological patterns in unknown patients.

Conclusion

The results demonstrate the statistical superiority of the **Few-Shot Learning** model for this task.

While SVM suffered a degradation in performance when faced with unseen subjects (falling to ~76%), the Prototypical Neural Network maintained high robustness (~89%). This confirms the hypothesis that deep feature extraction (Deep Learning) combined with distance-based learning (Few-Shot) is more suitable for dealing with the high variability of EEG signals among different humans.

Confusion Matrix



1- Interpretation of the Confusion Matrix Image:

The structure is a square divided into 4 quadrants:

1.1. Vertical/Left Axis:

The True Label. Represents the patient's actual state.

- 0 = Relaxed.
- 1 = Burnout.

1.2. Horizontal/Bottom Axis:

Predicted Label. Represents what the AI predicted.

- 0 = AI said it is Relaxed.
- 1 = AI said it is Burnout. Therefore:
- The upper left quadrant (0,0):
  - The patient was relaxed.
  - The AI said they were relaxed.
  - Conclusion: The AI correctly predicted the healthy state.
- The lower right quadrant (1,1):
  - The patient had Burnout.
  - The AI said the patient had Burnout.

    - Conclusion: The AI correctly predicted the Burnout state.
- The upper right quadrant (0,1):
    - The patient was relaxed.
    - The AI said the patient was experiencing burnout.
    - Conclusion: The AI was wrong in saying that the patient was experiencing burnout.
- The lower left quadrant (1.0):
    - The patient was experiencing burnout.
    - The AI said the patient was relaxed.
    - Conclusion: The AI was wrong in saying that the patient was relaxed.