

Popis datasetu:

Data pocházejí z projektu NSD – Nové syntetické drogy. Hlavním cílem projektu je vytvoření rozsáhlé databáze EEG záznamů dostupných drog testovaných na potkanech. Všechny testované drogy byly na základě farmakologických vlastností rozděleny do osmi skupin. V každém běhu byla pro každou skupinu vybrána jedna látka, která byla v cross-over designu aplikována osmnácti potkanům během deseti týdnů. Každý potkan dostal každou látku jednou, přičemž pořadí bylo randomizováno. K dispozici máme data od 54 potkanů (18 na každý běh, tedy tři runy). Každý záznam byl rozdělen na sedm segmentů, kde segment 0 představuje baseline před podáním látky a segmenty 1–6 odpovídají časovým úsekům (15 min) po jejím podání. Každý potkan byl vybaven osmnácti elektrodami, přičemž pro každého je k dispozici soubor o 1260 sloupcích, kde každý sloupec reprezentuje výkon EEG nebo jeho směrodatnou odchylku (SD) pro konkrétní frekvenční pásmo (delta, theta, alfa, beta, gamma), elektrody a segment. Například sloupce 1–180 patří segmentu 0 a obsahují údaje postupně pro každou elektrodu a pásmo (výkon a SD).

1. Formulace otázky

Například:

- „Jaký je efekt různých farmakologických skupin na výkon EEG v jednotlivých frekvenčních pásmech?“
- „Lze identifikovat specifické vzorce EEG aktivity, které korelují s konkrétními farmakologickými skupinami nebo látkami?“

2. Analýzy

- **Exploratorní analýza:** Prozkoumat základní statistiky (průměr, rozptyl) EEG výkonu v různých pásmech, mezi segmenty a mezi látkami.
- **Dimenzionální redukce:** Použít PCA/LDA (nebo kernel PCA/LDA) pro snížení počtu proměnných. Pomůže zjistit, zda existují latentní faktory (např. charakteristiky drog).
- **Shlukování:** Zkusit shlukování EEG výkonu mezi látkami (např. spektrální shlukování) pro identifikaci podobných farmakologických skupin.
- **ANOVA/MANOVA:** Porovnat, zda existují signifikantní rozdíly v EEG výkonu mezi látkami, farmakologickými skupinami a segmenty.
- **Diskriminační analýza:** Zkusit předpovědět farmakologickou skupinu látky na základě EEG výkonu.

- **Robustní statistika:** Identifikovat potenciální outliery (např. potkani s abnormálním EEG).

3. Konkrétní analýzy

1. Příprava dat:

- Normalizace EEG výkonu mezi potkany.
- Zahrnutí baseline (segment 0) jako kontrolní hodnoty.
- Rozdělení datasetu na tréninkovou a testovací část (pokud použiješ prediktivní modely).

2. Efekt drog na EEG:

- Využít smíšené modely (Mixed-Effects Models) s faktory „drogová skupina“, „segment“ a „elektroda“ (náhodný faktor: potkan).
- Zjistit, které frekvenční pásma jsou nejvíce ovlivněna.

3. Klasifikace:

- Použít LDA nebo logistickou regresi k odhadu farmakologické skupiny na základě EEG výkonu.

4. Redukce dimenze:

- Aplikovat PCA/LDA na všechna pásma a segmenty. Vizualizovat hlavní komponenty a zjistit, zda se látky nebo segmenty shlukují.

5. Shlukování:

- Identifikovat latentní vzorce EEG aktivity spojené s různými skupinami drog pomocí k-means nebo spektrálního shlukování.

4. Interpretace výsledků

- Vyhodnotit, zda jsou farmakologické skupiny odlišitelné na základě EEG.
- Porovnat vliv látek mezi segmenty (časový vývoj).
- Identifikovat limitace analýzy (např. velikost vzorku, kvalita záznamů).

5. Prezentace a report

- Vytvořit přehledný report s grafy (např. PCA bi-ploty, výsledky MANOVA, heatmapy výkonu).
- Prezentace by měla obsahovat interpretace efektů jednotlivých látek a vizuální přehled klíčových výsledků.

Závěr

Tento dataset nabízí dostatek dat pro různé statistické metody probírané v kurzu. Navíc umožňuje i výzkumový přínos, což dodá práci hodnotu.