

Využití prediktivních modelů u pacientů s karcinomy slinných žláz v aplikaci Czech Salivary Gland Database

Západočeská univerzita v Plzni
Fakulta aplikovaných věd
Katedra informatiky a výpočetní techniky

Autor: Ing. Vojtěch Jelínek • Vedoucí: Ing. Petr Brůha • Klinický konzultant: MUDr. David Kalfeřt, Ph.D. (FN Motol)

MOTIVACE A PROBLÉM

Karcinomy slinných žláz jsou velmi vzácná a biologicky extrémně různorodá onemocnění, u kterých klinická praxe rozeznává až 21 různých podtypů. Kvůli nízkému výskytu a chybějícím rozsáhlým datům je pro lékaře nesmírně obtížné spolehlivě odhadnout vývoj a prognózu u konkrétního člověka. Zatímco u běžných typů rakoviny mají onkologové k dispozici standardizované výpočetní kalkulačky rizik, v této oblasti doposud podobný nástroj v České republice zcela chyběl.

Pro sběr dat sice vznikla starší verze programu Czech Salivary Gland Database, ta však sloužila pouze jako pasivní kartotéka. Její technické řešení navíc vykazovalo roztržitost a vysokou redundanci údajů, což přímo znemožňovalo efektivní extrakci dat pro strojové učení i pokročilé plánování preventivních pooperačních kontrol.

TECHNICKÉ ŘEŠENÍ

Původní statický registr byl transformován na dynamický systém podpory klinického rozhodování postavený na moderním desktopovém stacku **Electron, React a SQLite**.

Výpočetně náročná analýza přežití (*Survival Analysis*) byla úspěšně oddělena do samostatného predikčního jádra v jazyce Python s využitím architektonického vzoru **Sidecar**. Komunikace probíhá asynchronně přes standardní I/O toky pomocí strukturovaných JSON kontraktů. Systém integruje dva klíčové přístupy:

- **Random Survival Forest (RSF):** Neparametrický model strojového učení schopný zachytit komplexní nelineární vztahy. Při validaci na souboru 122 pacientů pomocí .632 bootstrap estimátoru dosáhl vysoké přesnosti s C-indexem **0,814**.
- **Cox Proportional Hazards Model:** Semi-parametrický statistický model sloužící jako srovnávací baseline.

DYNAMIKA A DOPAD V PRAXI

Aplikace byla úspěšně nasazena do **ostrého klinického provozu ve Fakultní nemocnici v Motole**, přičemž inženýrsky řeší klíčové výzvy reálného provozu:

- **Řešení studeného startu:** Při prvotním nasazení na novém pracovišti aplikace poskytuje vestavěné předtrénované modely z původní motolské kohorty, což umožňuje okamžitou predikci bez čekání na sběr vlastních dat.
- **Automatická evoluce:** Systém obsahuje spouštěče automatického přetrénování (retraining triggers). Jakmile lokální databáze naroste o 10 %, modely se samy adaptují na nová data.
- **Absolutní bezpečnost a GDPR:** Celá analýza probíhá 100% offline na lokální stanici lékaře, citlivá onkologická data tak nikdy nepouštějí zabezpečenou síť nemocnice.