Žilinská Univerzita

Fakulta riadenia a informatiky

(Biomedicínska informatika)

Obrázok, na ktorom je čierny, temnota

Automaticky generovaný popis

**Dokumentácia k semestrálnej práci**

*Vytvorenie prototypu pre vyhodnocovanie predikcie ochorenia hereditárnej hemochromatózy*

**Vypracovali:** Martin Krištof, Andrej Farský

**Predmet:** Softvérové spracovanie biomedicínskych údajov

**Školský rok:** 2024/2025

Obsah

[1. Pochopenie údajov a problému 3](#_Toc198487253)

[1.1 Štruktúra údajov 3](#_Toc198487254)

[2. Spracovanie údajov 4](#_Toc198487255)

[2.1 Úvodná úprava a konverzia 4](#_Toc198487256)

[2.2 Načítanie údajov 4](#_Toc198487257)

[2.3 Čistenie údajov 4](#_Toc198487258)

[2.4 Normalizácia údajov 4](#_Toc198487259)

[2.5 Validácia kódu diagnóz 5](#_Toc198487260)

[2.6 Zhrnutie spracovania údajov 5](#_Toc198487261)

[2.6.1 Podrobnosti o vyradených údajoch 5](#_Toc198487262)

[3. Použitá technika 6](#_Toc198487263)

[3.1 Streamlit pre UI a interaktivitu 6](#_Toc198487264)

[3.2 Formátovanie dát a tabuľky 6](#_Toc198487265)

[3.3 Práca s dátovými typmi 6](#_Toc198487266)

[4. Návrh a popis prototypu riešenia 7](#_Toc198487267)

[4.1 Navigácia 7](#_Toc198487268)

[4.2 Stránka „Zoznam pacientov“ 8](#_Toc198487269)

[4.3 Stránka „HWE testy“ 10](#_Toc198487270)

[4.4 Stránka „Diagnózy“ 11](#_Toc198487271)

[4.5 Stránka „Vizualizácia genotypov vo vzťahu“ 12](#_Toc198487272)

[4.6 Stránka „Analýza diagnóz v čase“ 14](#_Toc198487273)

[5. Zistenia a vyhodnotenie výsledkov 15](#_Toc198487274)

[5.1 Zistenia ohľadom ochorenia HH pre jednotlivých pacientov 15](#_Toc198487275)

[5.2 Analýza zastúpenia genotypov v Hardy-Weinbergovej rovnováhe 16](#_Toc198487276)

[5.3 Analýza vzťahu medzi genetickým rizikom a diagnózami 17](#_Toc198487277)

[5.4 Zistenia z vizuálnej analýzy rozdelenia genotypov 18](#_Toc198487278)

[5.5 Zistenia z analýzy diagnóz v čase 19](#_Toc198487279)

[5.6 Záver 20](#_Toc198487280)

1. Pochopenie údajov a problému

Cieľom práce bolo vytvoriť interaktívny prototyp na filtrovanie, analýzu a vizualizáciu zdravotných údajov pacientov, súvisiacich s hereditárnou hemochromatózou (**HH**). Táto geneticky podmienená porucha metabolizmu železa spôsobuje jeho nadmerné vstrebávanie z potravy, čo vedie k jeho hromadeniu v tkanivách a orgánoch, ako sú pečeň, pankreas či srdce. V prípade neliečenia môže dôjsť k vážnym zdravotným komplikáciám, vrátane cirhózy pečene, cukrovky alebo srdcových ochorení.

Diagnostika kladie dôraz najmä o genetické testovanie prítomnosti mutácií v **HFE géne**, ktoré súvisia so zvýšeným vstrebávaním železa. Najčastejšie sledované mutácie sú:

* **C282Y**, **H63D**, **S65C**

Na základe prítomnosti týchto mutácií možno pacientov rozdeliť do kategórií:

* **Bez rizika** – pacient nevykazuje žiadnu zo sledovaných mutácií,
* **Prenášač** – prítomná je jedna mutácia, čo však nemusí viesť k ochoreniu,
* **S predispozíciou** – kombinácia mutácií naznačuje vyššie riziko rozvoja HH.

## 1.1 Štruktúra údajov

Poskytnuté dáta pochádzali z laboratórnych vyšetrení a obsahovali viaceré premenné relevantné pre klinickú aj výskumnú analýzu:

* identifikátor pacienta (ID) – jednoznačné označenie osoby,
* vek a pohlavie – demografické údaje,
* kódy diagnóz (MKCH-10) – diagnostické označenia podľa medzinárodnej klasifikácie,
* dátumy a časy prijatia vzorky a validácie výsledku – časové informácie o priebehu testovania,
* genetické mutácie – výsledky analýzy HFE génu pre mutácie **C282Y**, **H63D** a **S65C**.

Pre potreby spracovania bolo nevyhnutné:

* určiť typ každého stĺpca (číselné, textové, časové, kategorizované údaje),
* analyzovať frekvencie výskytu hodnôt (napr. časté diagnózy, mutačné kombinácie),
* odhaliť a riešiť chýbajúce alebo nesprávne údaje,
* identifikovať intervaly veku a dátumov, v ktorých sa údaje pohybujú.

Vytvorený prototyp má slúžiť odborným lekárom a výskumníkom, ktorí potrebujú intuitívne vyhľadávať a vizualizovať dáta podľa zvolených kritérií – ako sú prítomné mutácie, diagnózy, časové obdobie, vek pacienta alebo pohlavie.

# 2. Spracovanie údajov

Predtým, ako bolo možné pracovať s údajmi v samotnej aplikácii, bolo potrebné ich upraviť a pripraviť na spracovanie. Poskytnutý súbor v Exceli obsahoval viaceré nedostatky a nepresnosti, ktoré znemožňovali jeho priame načítanie do systému. Z tohto dôvodu prešli údaje viacerými krokmi úpravy, čistenia, validácie a transformácie, ktoré sú popísané v nasledujúcich podkapitolách.

## 2.1 Úvodná úprava a konverzia

Pôvodný súbor bol vo formáte .xlsx (Excel súbor), avšak obsahoval viaceré problémy, ako napríklad:

* preklepy v názvoch stĺpcov,
* prázdne stĺpce medzi údajmi,
* nejednotné pomenovanie alebo formátovanie hlavičiek.

Tieto problémy boli vyriešené manuálnou úpravou súboru priamo v Exceli, pričom bol následne exportovaný ako CSV súbor s oddeľovačom ;.

## 2.2 Načítanie údajov

CSV súbor bol načítaný do programu pomocou knižnice pandas. Počas načítania došlo k:

* odstráneniu medzier z názvov stĺpcov,
* transformácii názvov stĺpcov na malé písmená a nahradeniu medzier podčiarkovníkmi,
* explicitnému nastaveniu stĺpca „id“ ako reťazec, kvôli zachovaniu unikátneho označenia pacienta

Obrázok, na ktorom je text, snímka obrazovky, písmo

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

## 2.3 Čistenie údajov

V rámci čistenia boli odstránené všetky riadky, ktoré obsahovali aspoň jednu chýbajúcu hodnotu. Voliteľne bolo, zobraziť podrobnosti o tom, ktoré stĺpce v jednotlivých riadkoch chýbali.



## 2.4 Normalizácia údajov

V tomto kroku boli údaje transformované do jednotného formátu:

* dátumy a časy boli skonvertované do štandardného typu **datetime** / **time**,
* vek bol prevedený na desatinné číslo (**float**) s presnosťou na dve desatinné miesta,
* pohlavie bolo prevedené na veľké písmená (**str**) a zbavené nadbytočných medzier,
* hodnoty mutácií boli transformované na malé písmená (**str**) bez medzier,
* kódy diagnóz boli očistené a prevedené na textový formát (**str**).



## 2.5 Validácia kódu diagnóz

Kódy diagnóz boli validované podľa formátu MKCH-10, ktorý predpokladá nasledujúcu štruktúru:

* jedno veľké písmeno,
* dve číslice,
* voliteľne desatinná časť vo formáte .XX (maximálne dve číslice).

Všetky riadky, ktorých kód diagnózy nezodpovedal tejto štruktúre, boli z datasetu vyradené.



## 2.6 Zhrnutie spracovania údajov

Vstupný dataset obsahoval pôvodne **1668 riadkov**. Po vykonaní jednotlivých krokov spracovania – čistení, normalizácii a validácii – zostalo v konečnom datasete **1646 riadkov**, čo znamená, že **bolo odstránených 22 riadkov** (približne 1,32 % z pôvodného počtu).

### 2.6.1 Podrobnosti o vyradených údajoch

* **Chýbajúce hodnoty:**  
  Z databázy bolo odstránených 7 riadkov, ktoré obsahovali prázdne hodnoty v kritických stĺpcoch ako id, validovany\_vysledok alebo diagnoza\_mkch-10.
* **Nevalidné diagnózy (MKCH-10):**  
  Ďalších 15 riadkov bolo odstránených z dôvodu, že obsahovali nevalidný formát kódu diagnózy – konkrétne všetky mali kód K30.

Toto spracovanie údajov zabezpečilo, že do programu vstupoval len konzistentný, formátovo správny a bezchybný dataset, čo je nevyhnutné pre správne analytické spracovanie.

# 3. Použitá technika

Pri vývoji aplikácie sme zvolili kniznicu **Streamlit**, ktorá umožňuje rýchly vývoj interaktívnych dátových aplikácií s dôrazom na jednoduchosť a užívateľské rozhranie. Použité techniky možno rozdeliť do niekoľkých častí:

3.1 Streamlit pre UI a interaktivitu

Streamlit poskytol komponenty ako st.selectbox, st.multiselect, st.slider, st.date\_input, st.expander a st.button, ktoré sme využili na vytvorenie flexibilného systému filtrov a interakcií v aplikácii.

* **st.selectbox** umožňuje používateľovi vybrať jednu možnosť z ponuky, napríklad konkrétnu mutáciu alebo skupinu pacientov, čo zjednodušuje navigáciu a výber dát na analýzu.
* **st.multiselect** umožňuje vybrať viacero možností naraz, čo je užitočné pri porovnávaní viacerých mutácií alebo filtrov podľa rôznych kategórií.
* **st.slider** slúži na intuitívne nastavenie rozsahu numerických hodnôt, napríklad filtra veku pacientov.
* **st.date\_input** poskytuje možnosť výberu dátumov alebo časových intervalov pre analýzu časovo závislých údajov.
* **st.expander** umožňuje zoskupiť rozšírené nastavenia alebo doplnkové informácie, ktoré sa dajú podľa potreby skryť alebo zobraziť, čo pomáha udržiavať UI prehľadné.
* **st.button** slúži na spustenie špecifických akcií, ako napríklad spustenie analýzy alebo obnovenie výsledkov podľa nastavených filtrov.

Táto kombinácia komponentov priniesla rýchly a jednoduchý spôsob, ako vytvoriť interaktívne používateľské rozhranie bez nutnosti znalosti komplexných front-end technológií. Výhodou bolo hlavne rýchlosť a možnosť jednoduchej rozšíriteľnosti aplikácie.

3.2 Formátovanie dát a tabuľky

Na zobrazenie dát v aplikácii bola použitá funkcia **st.dataframe** spolu s vlastným formátovaním cez style.format na zabezpečenie čitateľnosti a prehľadnosti údajov. Pre zvýraznenie genetického\_statusu (bez rizika, prenášač a predispozícia) sa využilo prispôsobené štýlovanie, ktoré umožnilo farebne odlíšiť hodnoty v tabuľke.

3.3 Práca s dátovými typmi

Dátumy boli spracované ako datetime objekty a časy ako time, čo umožnilo presné porovnávanie vo filtroch. Textové stĺpce boli očistené a kategorizované na jednoduchšiu manipuláciu.

Celý vývoj aplikácie prebiehal v prostredí Python s využitím knižníc **pandas**, **numpy** a **streamlit**. Pandas zabezpečilo manipuláciu s dátami, numpy poskytlo nástroje na numerické operácie a streamlit sa postaral o zobrazenie a interaktivitu.

# 4. Návrh a popis prototypu riešenia

Navrhnuté riešenie má podobu **webovej aplikácie**, ktorá umožňuje používateľovi dynamicky filtrovať a prezerať zdravotné údaje pacientov na základe zvolených kritérií. Náhľad úvodnej stránky webovej aplikácie:

Obrázok, na ktorom je text, snímka obrazovky, písmo, číslo

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

## 4.1 Navigácia

Navigácia v aplikácii je riešená pomocou komponentu st.sidebar, ktorý umožňuje používateľovi prepínať medzi jednotlivými sekciami aplikácie v bočnom paneli.

Hlavne menu aplikácie



Používateľ má na výber tieto hlavné pohľady:

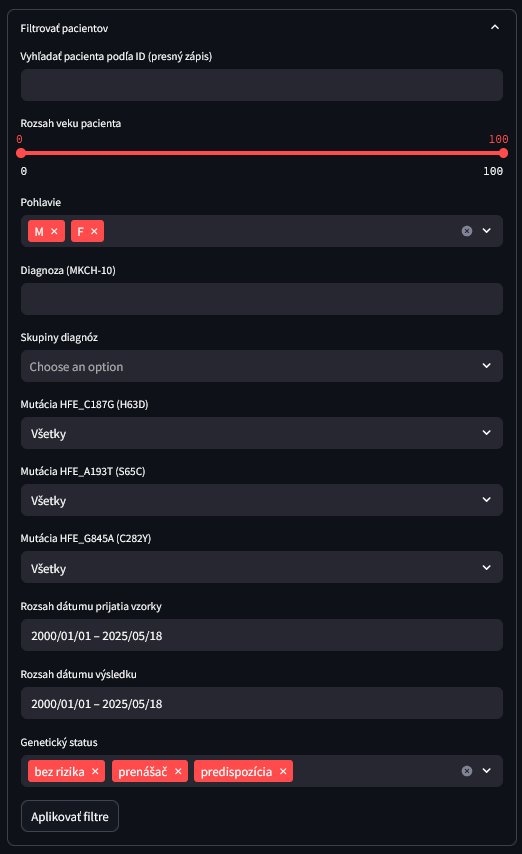
* **Domov** – úvodná stránka s prehľadom,
* **Zoznam pacientov** – detailný výpis a filtrovanie pacientskych dát,
* **HWE testy** – analýza Hardy-Weinbergovej rovnováhy,
* **Diagnózy** – zobrazenie a analýza diagnóz v dátach,
* **Vizualizácia genotypov vo vzťahu** – grafické zobrazenie vzťahov mutácií,
* **Analýza diagnóz v čase** – časová analýza vývoja diagnóz.

Tento prístup zabezpečuje jednoduchú a intuitívnu orientáciu používateľa medzi rôznymi časťami aplikácie, pričom všetky hlavné funkcie sú dostupné na jednom mieste a prepínajú sa okamžite podľa výberu.

## 4.2 Stránka „Zoznam pacientov“

Táto podstránka slúži na prehľad a filtrovanie pacientskych údajov, čo je kľúčová funkcionalita prototypu. Stránka „Zoznam pacientov“ je rozdelená do troch hlavných častí:

**1. Filtre pacientov**



Interaktívna sekcia s rozbaľovacími menu a vstupnými poľami, kde používateľ môže:

* Filtrovať pacientov podľa ID (pri zadaní konkrétneho ID sa deaktivujú ostatné filtre, čo zaručuje rýchle a presné vyhľadávanie).
* Nastaviť rozsah veku pacienta pomocou st.slider.
* Vybrať pohlavie pacienta (st.multiselect).
* Filtrovať podľa konkrétnej diagnózy (zadáva sa kód MKCH-10).
* Skupiny diagnóz (st.multiselect) – implementované pomocou slovníka DG\_SKUPINY, čo umožňuje filtrovať na základe skupín diagnóz.
* Vybrať jednotlivé mutácie HFE génu (st.selectbox), vrátane H63D, S65C a C282Y.
* Filtrovať podľa rozsahu dátumu prijatia vzorky a validácie výsledku (st.date\_input).
* Filtrovať podľa genetického statusu pacienta (st.multiselect), kde možnosti sú napríklad „bez rizika“, „prenášač“, „predispozícia“.
* Tlačidlo Aplikovať filtre, ktoré vykoná filtrovanie na základe zadaných hodnôt.

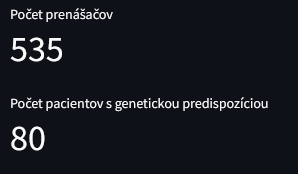
**2. Tabuľka s pacientmi**



Dynamicky aktualizovaná tabuľka, ktorá zobrazuje výsledky filtrovania.

* Dátumy sú formátované na štandardný formát dd.mm.rrrr a časy na hh:mm.
* Vek pacienta je zaokrúhlený na 2 desatinné miesta, aby sa zlepšila čitateľnosť.
* Použitý je štýlovací mechanizmus, ktorý zafarbuje genetický status podľa jeho hodnoty:
  + **Zelená:** „bez rizika“
  + **Oranžová:** „prenášač“
  + **Červená:** „predispozícia“

**3. Sumarizácia genetických statusov**



Metriky, ktoré poskytujú:

* Počet pacientov s genetickou predispozíciou.
* Počet prenášačov.

## 4.3 Stránka „HWE testy“

Táto podstránka je zameraná na analýzu genetických údajov prostredníctvom Hardy-Weinbergovho testu (HWE) a analýzy distribúcie genotypov. Používateľ môže získať prehľad o rozložení genotypov pre tri kľúčové mutácie a zároveň vykonať test HWE na zistenie, či sú genetické údaje v rovnováhe.

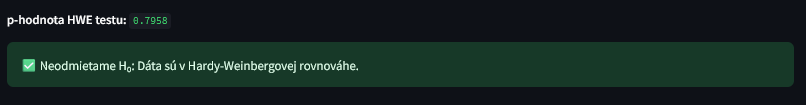
Pre každú mutáciu je generovaný samostatný blok



ktorý obsahuje:

* Podnadpis s názvom mutácie.
* Prehľad o rozložení genotypov (normal, heterozygot, mutant).
* Zobrazenie počtov a percentuálneho zastúpenia genotypov.

Po zobrazení rozdelenia genotypov sa pre každú mutáciu vykonáva Hardy-Weinbergov test. Výsledok HWE testu je zobrazený ako p-hodnota.



Ak je p-hodnota nižšia ako 0.05, aplikácia zobrazuje upozornenie, že H₀ je odmietnutá a dáta nie sú v HWE rovnováhe. Ak je p-hodnota vyššia alebo rovná 0.05, zobrazuje sa správa, že H₀ nie je odmietnutá a dáta sú v HWE rovnováhe. V prípade nedostatku údajov pre niektorú z mutácií sa zobrazí varovanie: "Nebolo možné vypočítať HWE test – žiadne alebo neúplné dáta."

## 4.4 Stránka „Diagnózy“

Táto podstránka poskytuje prehľad o tom, ako genetické riziko súvisí s vybranými skupinami diagnóz. Umožňuje používateľovi analyzovať distribúciu diagnóz vo vybranej oblasti a porovnať ju s genetickým rizikom na základe sledovaných mutácií.

**Výber skupiny diagnóz**



Komponent st.selectbox() umožňuje výber skupiny diagnóz. Po výbere sa zobrazí textová informácia o vybranej kapitole a príslušnom intervale MKCH-10 kódov.

**Tabuľka s analýzou diagnóz**



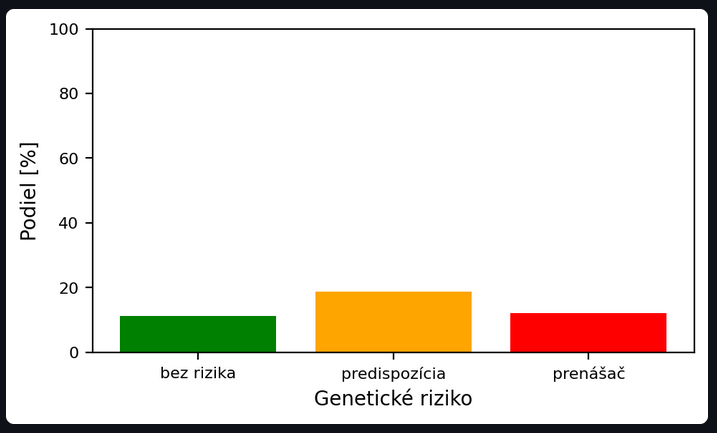
Po určení genetického rizika sa dáta agregujú na základe genetického rizika a výskytu vybraných diagnóz. Kategórie genetického rizika:

* **bez rizika:** Všetky genotypy sú normal.
* **prenášač:** Presne 1 heterozygot a ostatné genotypy sú normal.
* **predispozícia:** Viac ako 1 heterozygot alebo aspoň 1 mutantný genotyp.

Výsledná tabuľka obsahuje nasledujúce stĺpce:

* Počet s diagnózou – Počet pacientov s diagnózou v danej rizikovej skupine.
* Počet v skupine – Celkový počet pacientov v rizikovej skupine.
* Podiel [%] – Percentuálny podiel pacientov s diagnózou v danej rizikovej skupine.

**Grafické zobrazenie vzťahu**



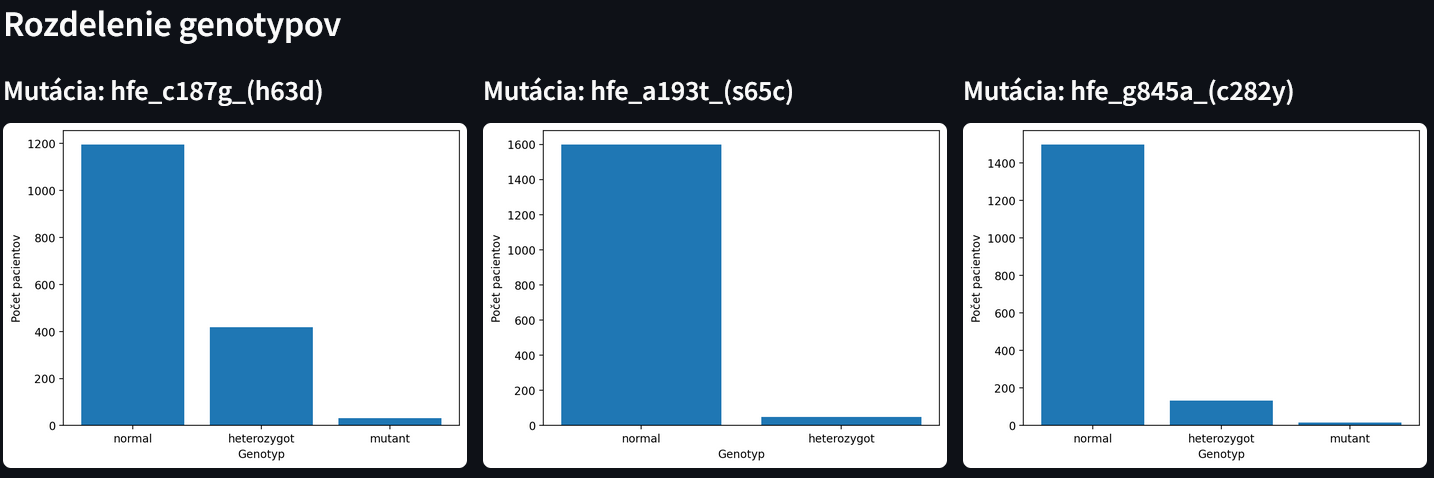
Osa X zobrazuje kategórie genetického rizika (bez rizika, prenášač, predispozícia). Osa Y zobrazuje percentuálny podiel pacientov s diagnózou (podiel). Použité farby pre kategórie:

* Zelená – bez rizika
* Oranžová – prenášač
* Červená – predispozícia

## 4.5 Stránka „Vizualizácia genotypov vo vzťahu“

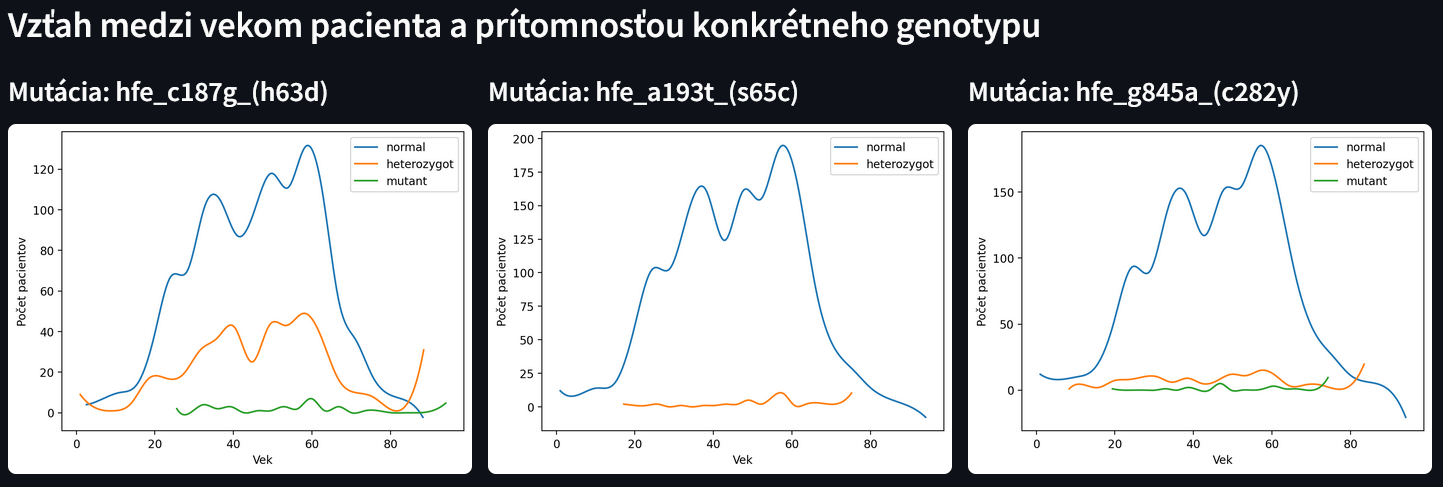
Cieľom tejto podstránky je vizuálne prezentovať vzťahy medzi genetickými údajmi (mutácie HFE génu) a vybranými faktormi, ako sú vek, pohlavie a diagnostické skupiny. Vizualizácia umožňuje lepšie porozumieť rozloženiu genotypov a identifikovať potenciálne asociácie s diagnostickými kategóriami. Prezentované grafy poskytujú nasledujúce prehľady:

**Rozdelenie genotypov**



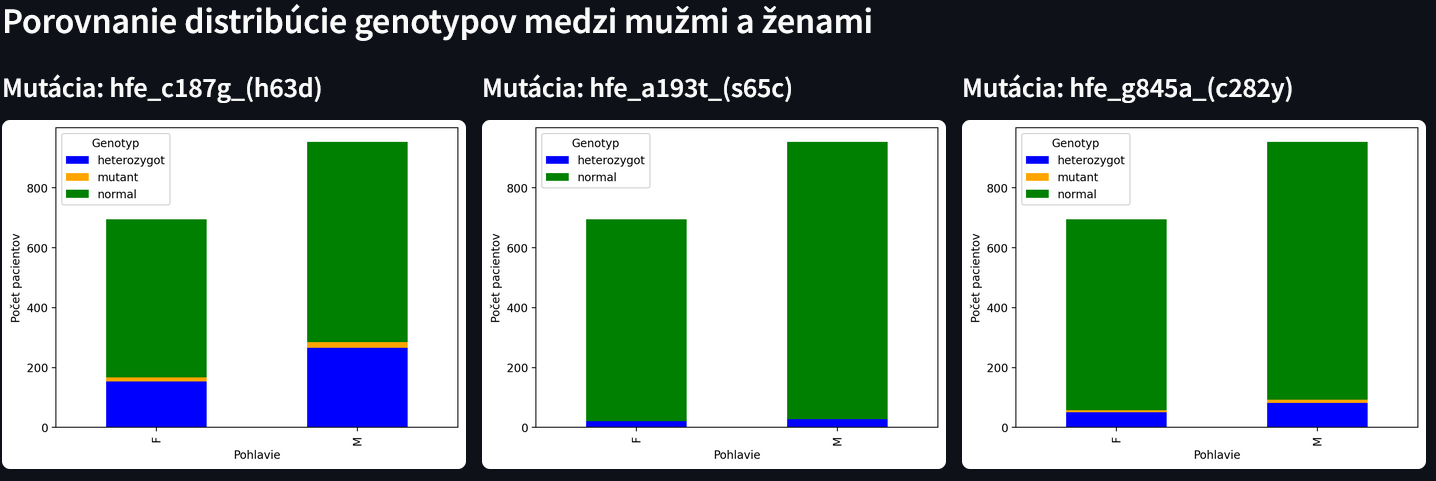
Pre každú mutáciu (hfe\_c187g\_(h63d), hfe\_a193t\_(s65c), hfe\_g845a\_(c282y)) sa vytvorí jednoduchý stĺpcový graf so zastúpením genotypov (normal, heterozygot, mutant). Vykreslenie je oddelené pre každú mutáciu, každý má svoj graf.

**Vek vs. Genotyp**



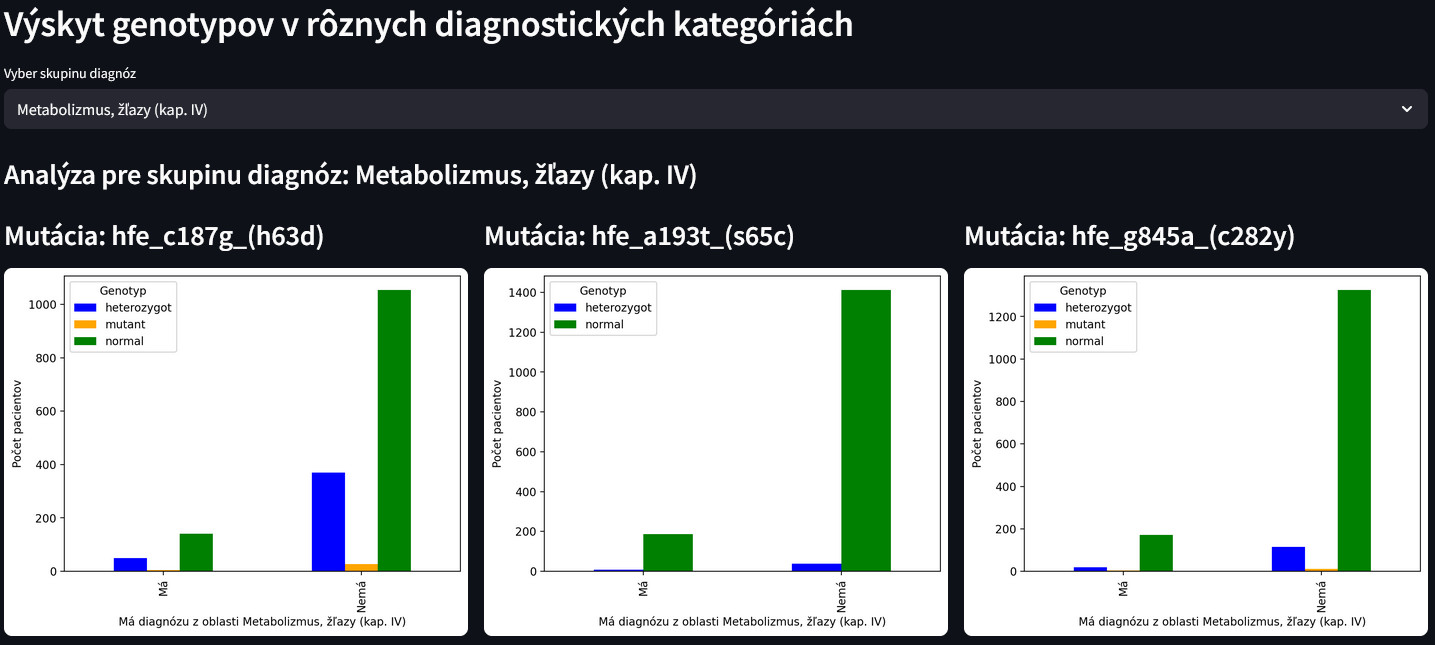
Vizualizuje distribúciu genotypov v rôznych vekových kategóriách, čím umožňuje sledovať, či sú určité genotypy častejšie zastúpené v špecifických vekových skupinách. Krivky sú vykreslené pre jednotlivé genotypy (normal, heterozygot, mutant).

**Pohlavie vs. Genotyp**



Porovnáva rozloženie genotypov medzi mužmi a ženami, čo môže indikovať potenciálne rozdiely v genetickom riziku na základe pohlavia.

**Výskyt genotypov v rôznych diagnostických kategóriách**



Skúma zastúpenie genotypov v rôznych klinických oblastiach (napr. nádorové ochorenia, metabolické poruchy) na základe klasifikácie MKCH-10.

## 4.6 Stránka „Analýza diagnóz v čase“

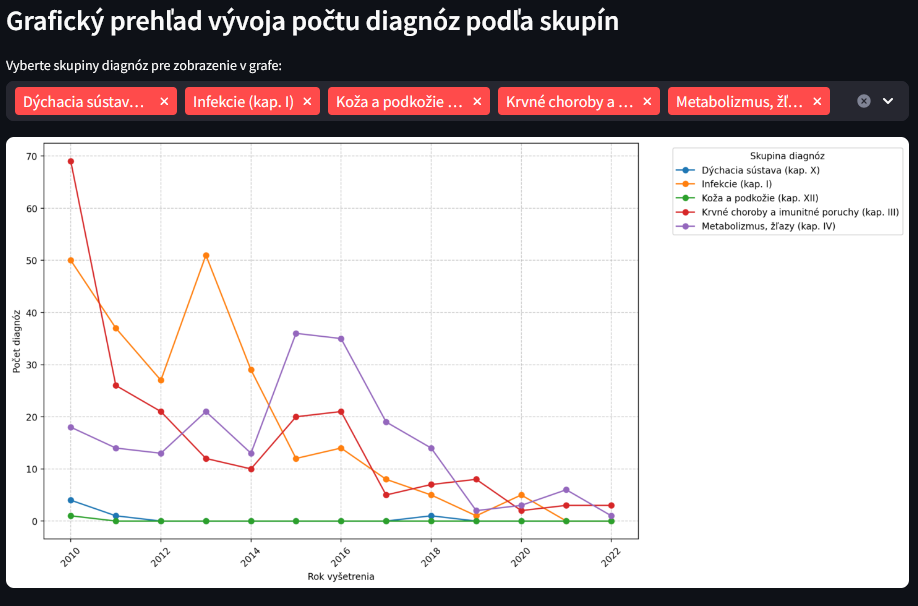
Táto podstránka sa zameriava na skúmanie vývoja počtu diagnóz v rôznych kategóriách MKCH-10 v priebehu rokov. Analýza umožňuje identifikovať trendy v jednotlivých diagnostických skupinách a potenciálne zmeny vo frekvencii výskytu konkrétnych diagnóz.

**Tabuľkový prehľad počtu diagnóz podľa rokov a skupín**



Pre každý rok sa zobrazuje počet diagnóz v jednotlivých skupinách (napr. Infekcie, Nádory, Krvné choroby a pod.). Stĺpce reprezentujú skupiny diagnóz, riadky roky vyšetrenia. V prípade absencie dát pre niektorý rok alebo skupinu sú hodnoty v tabuľke automaticky nastavené na nulu.

**Grafický prehľad vývoja počtu diagnóz**



Používateľ si môže vybrať, ktoré skupiny diagnóz chce zobraziť v grafe. Pre každý vybraný rok sú zobrazené samostatné línie, kde každá farba reprezentuje inú skupinu diagnóz. Graf je doplnený legendou a mriežkou pre jednoduchšiu orientáciu. Os X zobrazuje roky vyšetrenia a os Y počet diagnostikovaných prípadov.

# 5. Zistenia a vyhodnotenie výsledkov

Prototyp umožnil interaktívnu prácu so zdravotnými údajmi a výrazne zjednodušil proces ich analýzy a filtrovania. Používateľ tak získava možnosť:

* prehľadne filtrovať podľa množstva klinických a genetických parametrov
* rýchlo identifikovať relevantných pacientov
* pozorovať súvislosti medzi genetickými mutáciami a diagnózami

## 5.1 Zistenia ohľadom ochorenia HH pre jednotlivých pacientov

Kľúčovou analytickou funkciou bolo stanovenie tzv. **genetického statusu pacienta** – teda kategorizácia na základe prítomnosti mutácií v géne **HFE**, ktoré súvisia s hereditárnou hemochromatózou (HH).

Genetický status bol určovaný podľa nasledovných pravidiel:

* **Predispozícia**: pacient má aspoň jednu mutáciu (homozygot alebo zložený heterozygot), alebo je dvojnásobný heterozygot,
* **Prenášač**: pacient je heterozygot len pre jednu mutáciu,
* **Bez rizika**: v genetickom profile nie je detekovaná žiadna relevantná mutácia.

Na základe tejto klasifikácie boli získané nasledovné štatistiky zo spracovaného datasetu:

|  |  |  |
| --- | --- | --- |
| Genetický statusu | Počet pacientov | Percentuálny podiel |
| Bez rizika | 1031 | 62,6 % |
| Prenášač | 535 | 32,5 % |
| Predispozícia | 80 | 4,9 % |
| Spolu | **1646** | **100 %** |

Z týchto údajov vyplýva, že:

* Takmer tretina populácie (32,5 %) sú **prenášači** – majú jednu z mutácií, ktoré môžu byť rizikové pri ďalšej genetickej kombinácii (napr. v prípade, že druhý rodič má tiež mutáciu),
* 4,9 % pacientov má **genetickú predispozíciu** na rozvoj HH a mali by byť ďalej klinicky sledovaní,
* Väčšina pacientov (vyše 60 %) nevykazuje genetické známky súvisiace s HH.

## 5.2 Analýza zastúpenia genotypov v Hardy-Weinbergovej rovnováhe

V praxi slúži **HWE test na overenie**, či sa pozorované genetické údaje významne odlišujú od toho, čo by sme očakávali v „rovnovážnej“ populácii. Teda očakávame, že keď vieme koľko ľudí má mutáciu, vieme predpovedať, koľko by malo byť heterozygotov a homozygotov.

Výsledky HWE testov na hladine významnosti 5% pre jednotlivé mutácie:

|  |  |  |  |
| --- | --- | --- | --- |
| Mutácia | HWE p-hodnota | Rovnováha | Interpretácia |
| H63D | 0.7958 | Áno | Výskyt zodpovedá populácii |
| S65C | 0.8351 | Áno | Zriedkavá mutácia, rovnovážna |
| C282Y | 0.0000 | Nie | Pravdepodobne selektívny výber alebo klinický význam |

Môžeme si všimnúť, že pri mutácii **C282Y**, rozloženie genotypov nezodpovedá tomu, čo by sme očakávali v bežnej populácii bez vonkajších vplyvov (ako je výber, migrácia alebo chorobnosť). Toto môže naznačovať, že:

* daná mutácia je spojená s ochorením (HH), a teda sa častejšie vyskytuje u testovaných pacientov, než v bežnej populácii,
* alebo že na ňu pôsobia selekčné tlaky – napr. tí, ktorí majú túto mutáciu, častejšie končia na genetickom testovaní.

## 5.3 Analýza vzťahu medzi genetickým rizikom a diagnózami

Jednou z úloh bolo zistiť, či existuje súvis medzi genetickým rizikom pre HH a konkrétnymi typmi diagnóz, ktoré pacienti v dátach mali. Najčastejšou oblasťou diagnóz u pacientov s predispozíciou na HH bola **kapitola 6 – Choroby tráviaceho systému** (K00–K93), kde až **50**% z nich malo diagnózu z tejto kategórie. Rovnako vysoké podiely sa však objavili aj u prenášačov a pacientov bez rizika, čo naznačuje, že diagnózy z tejto oblasti sú v populácii veľmi bežné. Z tohto dôvodu je ťažké jednoznačne prisúdiť túto skupinu ochorení výlučne genetickému riziku HH.

Obrázok, na ktorom je text, snímka obrazovky, štvorec, diagram

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Špecifickejší vzťah s genetickým rizikom HH bol pozorovaný pri týchto kategóriách:

* Kapitola 3: **Krvné choroby a imunitne poruchy** (D50–D89)
  + Vyšší výskyt medzi pacientmi s predispozíciou; súvisí s metabolickými dôsledkami prebytku železa.

**Obrázok, na ktorom je text, snímka obrazovky, diagram, štvorec

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.**

* Kapitola 4: **Metabolizmus, žľazy** (E00–E90)
  + Zahŕňa samotnú HH a príbuzné poruchy (napr. diabetes mellitus typu 2, poruchy metabolizmu železa, ochorenia pečene).

Naopak, existuje viacero oblastí diagnóz, kde sa nevyskytol ani jeden pacient s genetickou predispozíciou, teda výskyt bol nulový. Tieto oblasti môžeme označiť ako diagnosticky nesúvisiace s HH, pričom výskyt absentoval úplne alebo bol u pacientov bez rizika alebo prenášačov. Patria sem:

* Kapitola 1: Infekcie
  + Diagnózy sa vyskytovali výhradne u pacientov bez rizika a prenášačov.
* Kapitola 2: Nádory
* Kapitola 5: Psychické poruchy
* Kapitola 6: Nervový systém
* Kapitola 12: Choroby kože

Tieto kategórie pomáhajú zvýrazniť oblasti, kde je genetická HH predispozícia významnejšia.

## 5.4 Zistenia z vizuálnej analýzy rozdelenia genotypov

**Rozdelenie genotypov**

Všetky tri mutácie majú dominantný výskyt normálneho genotypu (t. j. väčšina pacientov nemá mutáciu). Mutácie v heterozygotnej forme sú najčastejšie pri mutácii H63D, menej pri C282Y, a len minimálne pri S65C. Mutantný homozygotný genotyp je zriedkavý pri všetkých mutáciách a pri S65C sa neobjavuje vôbec.

**Vek vs. Genotyp**

Pre všetky tri mutácie je rozdelenie pacientov podľa veku nasledovné:

#### hfe\_c187g (H63D):

* Normálny genotyp má najvyšší výskyt medzi 30–70 rokmi.
* Heterozygoti majú nižší výskyt, ale podobné vekové rozloženie ako normálni až na zvýšenie počtu v 90. roku života.
* Mutanti majú minimálny výskyt a vekové rozloženie posunuté oproti ostatným. Prvý výskyt až v 25. roku života.

#### hfe\_a193t (S65C):

* Dominancia normálneho genotypu naprieč všetkými vekovými kategóriami.
* Heterozygoti sú len mierne zastúpení a roztrúsení naprieč vekom.

#### hfe\_g845a (C282Y):

* Normálny genotyp je najčastejší v strednom až vyššom veku.
* Heterozygoti aj mutanti sú prítomní v malom počte a ich výskyt je relatívne rovnomerný cez vekové spektrum.

**Pohlavie vs. Genotyp**

Naprieč všetkými mutáciami pozorujeme, že počet pacientiek (žien) je konzistentne nižší než počet pacientov (mužov) – rozdiel sa pohybuje približne medzi 200 až 250 osôb. Napriek tomuto rozdielu v absolútnych číslach sú pomery jednotlivých genotypov (normálny, heterozygotný, mutantný) medzi pohlaviami prakticky totožné pri každej z analyzovaných mutácií. Ako sme už zistili skôr, dominantný je normálny genotyp, nasleduje výrazne nižší výskyt heterozygotov a mutantný genotyp sa vyskytuje len ojedinele.

**Výskyt genotypov v rôznych diagnostických kategóriách**

Pri porovnaní výskytu genotypov naprieč rôznymi skupinami diagnóz možno pozorovať niekoľko spoločných trendov:

* Normálny genotyp dominuje vo všetkých diagnostických kategóriách, čo korešponduje s celkovým rozdelením v populácii.
* Heterozygotný genotyp sa vyskytuje v každej kategórii, avšak v nižších počtoch. Jeho výskyt sa mierne líši podľa typu ochorenia, no žiadna skupina diagnóz nevyčnieva zásadne vyšším výskytom.
* Mutantný genotyp je najzriedkavejší a jeho výskyt je vo väčšine kategórií len minimálny až zanedbateľný.

Celkovo sa však výskyt jednotlivých genotypov vo väčšine diagnostických skupín zhoduje so všeobecným zastúpením týchto genotypov v celej vzorke pacientov, teda dominantný je normálny genotyp, nasledovaný heterozygotným a mutantným.

## 5.5 Zistenia z analýzy diagnóz v čase

Väčšina skupín diagnóz vykazuje klesajúci trend v počte prípadov od roku 2010 do 2022. Najvýraznejší pokles je viditeľný v skupine Tráviaca sústava, ktorá bola dominantná najmä v rokoch 2011–2017, no potom prudko klesla. Od roku 2017 nastáva celoplošný pokles naprieč väčšinou skupín, môže to súvisieť s legislatívnymi zmenami, zmenou zberu dát alebo zmenou v prístupe k zdravotnej starostlivosti.

Najvyšší počet diagnóz v minulosti:

* **Tráviaca sústava** – až 120 prípadov (2015).
* **Infekcie** – kolísali, no miestami dosahovali až 50 prípadov (2013).
* **Koža a podkožie** a **Metabolizmus, žľazy** – tiež pomerne frekventované v niektorých obdobiach.

## 5.6 Záver

V rámci tejto práce bol vytvorený interaktívny prototyp formou webovej aplikácie na spracovanie a analýzu genetických a klinických údajov pacientov so zameraním na hereditárnu hemochromatózu (HH). Detailné spracovanie dát zabezpečilo ich očistenie, validáciu a správne typovanie, čo umožnilo kvalitnú a spoľahlivú analýzu.

Genetické testovanie sa sústredilo na tri kľúčové mutácie v géne HFE (H63D, S65C a C282Y), ktoré sú známe ako významné markery predispozície k HH. Na základe kombinácie týchto mutácií sme klasifikovali pacientov do troch kategórií rizika: bez rizika, prenášači a pacienti s predispozíciou. Najväčšiu skupinu tvorili prenášači (535) a 80 pacientov malo genetickú predispozíciu, ktorá môže zvýšiť riziko rozvoja ochorenia.

Testy Hardy-Weinbergovej rovnováhy potvrdili, že dve z troch analyzovaných mutácií (H63D a S65C) sú v genetickej populácii v rovnováhe, zatiaľ čo mutácia C282Y túto rovnováhu porušuje, čo môže naznačovať väčší výskyt pacientov, ktorí majú príznaky, abnormality v krvnom obraze v poskytnutej skupine pacientov ako v bežnej populácii.

Analýza vzťahu medzi genetickým rizikom a klinickými diagnózami ukázala, že pacienti s predispozíciou vykazujú zvýšený výskyt ochorení tráviacej sústavy, krvných a imunitných chorôb, ako aj porúch metabolizmu žliaz, čo korešponduje s klinickým obrazom HH. Naopak, niektoré skupiny diagnóz (napríklad nádory, psychické poruchy či ochorenia nervového systému) sa vyskytovali výhradne u pacientov bez genetického rizika alebo u prenášačov, čo poukazuje na ich nezávislosť od genetických faktorov HH.

Celkovo prototyp poskytuje efektívny nástroj pre odborníkov na identifikáciu a analýzu genetických a klinických faktorov ovplyvňujúcich HH, umožňujúc tak lepšie pochopenie ochorenia a podporu včasnej diagnostiky a liečby. Ďalším krokom môže byť rozšírenie dátového súboru, hlbšia štatistická analýza a integrácia ďalších klinických premenných na zlepšenie presnosti a užitočnosti aplikácie.