**ŽILINSKÁ UNIVERZITA V ŽILINE**

FAKULTA RIADENIA A INFORMATIKY



SOFTVÉROVÉ SPRACOVANIE BIOMEDICÍNSKYCH ÚDAJOV

Študijný program: Biomedicínska informatika

Vypracovali: Alexandra Čižmárová Akademický rok: 2024/2025

Kristína Dostálová Študijná skupina: 5ZIB11

Šimon Meško

Matúš Rumíšek

Obsah

[Zoznam obrázkov 3](#_Toc197812986)

[Úvod 4](#_Toc197812987)

[1 Pochopenie údajov a problému 5](#_Toc197812988)

[2 Príprava a predspracovanie údajov 6](#_Toc197812989)

[3 Výber a použitie analytických nástrojov 7](#_Toc197812990)

[3.1 Programovací jazyk a knižnice 7](#_Toc197812991)

[3.2 Návrh Prototypu 7](#_Toc197812992)

[3.3 Výhody zvoleného prístupu 7](#_Toc197812993)

[4 Návrh a popis prototypu riešenia 8](#_Toc197812994)

[4.1 Architektúra riešenia 8](#_Toc197812995)

[4.2 Modulárna štruktúra 8](#_Toc197812996)

[4.3 Nasadenie aplikácie v praxi 8](#_Toc197812997)

[4.4 Používateľská príručka 9](#_Toc197812998)

[5 Analýza a vyhodnotenie výsledkov 10](#_Toc197812999)

[5.1 Hardy-Weinbergova rovnováha 10](#_Toc197813000)

[5.2 Percentuálne zastúpenie genotypov 11](#_Toc197813001)

[5.3 Súvislosť mutácií s pečeňovými diagnózami 11](#_Toc197813002)

[5.4 Demografická charakteristika pacientov a výskyt pečeňových ochorení 12](#_Toc197813003)

[5.5 Vývoj diagnóz podľa MKCH-10 13](#_Toc197813004)

[5.6 Limity analýzy 14](#_Toc197813005)

[Záver 16](#_Toc197813006)

# Zoznam obrázkov

[Obrázok 1: Rozdelenie genotypov mutácií 10](#_Toc197802291)

[Obrázok 2: Porovnanie genotypu podľa veku v mutáciach 12](#_Toc197802292)

[Obrázok 3: Porovnanie genotypu podľa veku v mutáciach 13](#_Toc197802293)

[Obrázok 4: Porovnanie genotypu podľa pečeňového ochorenia v mutáciach 13](#_Toc197802294)

[Obrázok 5: Vývoj diagnóz podľa MKCH-10 14](#_Toc197802295)

# ZOZNAM TABULIEK

[Tabuľka 1: Výsledky testu H–W rovnováhy pre mutácie HFE génu 11](#_Toc198073650)

[Tabuľka 2: Kontingenčná tabuľka pre mutáciu C282Y a prítomnosť pečeňovej diagnózy 11](#_Toc198073651)

[Tabuľka 3: Kontingenčná tabuľka pre mutáciu H63D a prítomnosť pečeňovej diagnózy 12](#_Toc198073652)

[Tabuľka 4: Kontingenčná tabuľka pre mutáciu S65C a prítomnosť pečeňovej diagnózy 12](#_Toc198073653)

# Úvod

Dedičná hemochromatóza je geneticky podmienené ochorenie charakterizované nadmerným vstrebávaním železa v tele, čo môže viesť k jeho ukladaniu v orgánoch – najmä v pečeni – a následnému poškodeniu ich funkcie. Najčastejšie sa spája s mutáciami v HFE géne, pričom za najvýznamnejšie sa považujú mutácie C282Y, H63D a S65C.

Cieľom tejto práce je analyzovať výskyt týchto mutácií v poskytnutom datasete pacientov a preskúmať ich súvis s vybranými demografickými a klinickými údajmi, ako sú vek, pohlavie a diagnózy podľa klasifikácie MKCH-10. Hlavným zameraním je overenie genetickej rovnováhy podľa Hardy-Weinbergovej rovnice, určenie frekvencií prenášačov a osôb s predispozíciou na hemochromatózu, a identifikácia možných súvislostí medzi mutáciami a výskytom ochorení pečene.

Riešenie bolo implementované formou interaktívnej aplikácie v prostredí Streamlit, ktorá umožňuje lekárom a analytikom jednoducho načítať genetické údaje, vykonávať základné štatistické analýzy a vizualizovať výsledky. Projekt zároveň automaticky generuje výstupný report vo forme dokumentu, čo zvyšuje jeho praktickú využiteľnosť v klinickej alebo výskumnej praxi.

# Pochopenie údajov a problému

Cieľom tejto semestrálnej práce je analyzovať genetické údaje týkajúce sa mutácií v HFE géne a ich súvislosť so zdravotným stavom pacientov, konkrétne s predispozíciou na ochorenie hereditárna hemochromatóza. Toto ochorenie je autozomálne recesívne a vzniká v dôsledku poruchy metabolizmu železa, čo môže viesť k jeho nadmernému hromadeniu v orgánoch, najmä v pečeni.

Získaný dataset obsahuje výsledky genetického testovania pacientov na tri mutácie HFE génu: C282Y, H63D a S65C. Súčasťou údajov sú aj demografické údaje (vek, pohlavie), dátumy testovania a zdravotné diagnózy pacientov podľa klasifikácie MKCH-10. V kontexte klinickej praxe sa práve tieto mutácie považujú za významné indikátory genetickej predispozície na vznik tohto ochorenia.

V rámci riešenia bolo potrebné pochopiť nasledovné aspekty problému:

* Genotypové výsledky sú klasifikované ako normal, heterozygot, alebo homozygot mutant, čo umožňuje ich transformáciu do formátu vhodného na ďalšiu analýzu (napr. wt/wt, wt/mut, mut/mut).
* Manažment pacienta s podozrením na hemochromatózu závisí nielen od genetického nálezu, ale aj od jeho veku, pohlavia a prítomnosti klinických symptómov, najmä ochorení pečene.
* Diagnózy sú uvedené v tvare kódov MKCH-10, ktoré si vyžadujú roztriedenie podľa skupín a overenie ich aktuálnosti (zastarané kódy sa môžu v súčasnosti klasifikovať inak).

Celkové spracovanie sa sústredilo na identifikáciu súvislostí medzi mutáciami a ochoreniami, posúdenie genetickej rovnováhy (Hardy-Weinbergov test), vyhodnotenie počtu prenášačov a predisponovaných pacientov, a vizualizáciu údajov na podporu rozhodovania v klinickej genetike.

# Príprava a predspracovanie údajov

Na začiatku projektu bol dataset poskytnutý vo formáte .xls a bol transformovaný do formátu .csv, s ktorým sa pracovalo prijateľnejšie. Niektoré stĺpce obsahovali nekonzistentné alebo duplicitné informácie a preto bolo nevyhnutné vykonať viacero krokov na jeho očistenie a úpravu do použiteľnej formy pre ďalšiu analýzu.

**Zjednotenie a odstránenie nepotrebných stĺpcov a záznamov**

Viaceré stĺpce (napr. s dátumami) boli rozdelené do dvoch samostatných stĺpcov, preto sa spojili do jedného (validovany\_vysledok, prijem\_vzorky). Okrem toho boli odstránené aj prázdne stĺpce. Boli odstránené záznamy s chýbajúcim identifikátorom pacienta (id) alebo nesprávnym vekom.

**Úprava stĺpca vek**

Stĺpec vek obsahoval hodnoty s čiarkami namiesto bodiek, preto sa normalizovali a pretypovali na číselný formát. Zároveň sa odfiltrovali extrémne hodnoty (vek nad 120 rokov), ktoré pravdepodobne predstavovali chyby.

**Štandardizácia genotypov**

Genotypy boli uvedené ako textové hodnoty: normal, heterozygot, homozygot, prípadne ako mutácia alebo patogénna. Na účely analýzy sa tieto hodnoty premapovali do štandardného genetického formátu: wt/wt, wt/mut, mut/mut.

**Dynamické vyhľadanie stĺpcov mutácií**

Stĺpce pre konkrétne mutácie (C282Y, H63D, S65C) boli identifikované na základe výskytu ich genetickej zmeny (napr. G845A pre C282Y) a názvu génu HFE. Tento prístup zabezpečil robustnosť voči zmenám názvu stĺpcov v datasete.

**Uloženie očisteného datasetu**

Finálny očistený dataset je stále uložený do súboru SSBU25\_dataset\_cleaned.csv, aby sa mohol opakovane použiť v iných častiach projektu (grafy, export, atď.).

# Výber a použitie analytických nástrojov

Na analýzu údajov a vizualizáciu výsledkov boli zvolené kombinované nástroje, ktoré umožňujú nielen rýchly vývoj, ale aj interaktivitu a generovanie výstupov vo formáte použiteľnom pre klinický výskum či zdravotnícku dokumentáciu.

## Programovací jazyk a knižnice

Hlavným programovacím jazykom bol Python, a to pre jeho rozsiahlu podporu v oblasti dátovej analýzy a biomedicíny.

Použité knižnice:

* **Pandas** – na načítanie a spracovanie dátového súboru, agregácie a transformácie údajov.
* **Matplotlib a Seaborn** – pre tvorbu vizuálne atraktívnych grafov.
* **Scipy** – na štatistické testy ako je napr. chi-kvadrát test.
* **Python-docx** – pre generovanie finálneho exportu analýzy do formátu .docx.
* **Streamlit** – na tvorbu interaktívneho používateľského rozhrania (aplikácie).

## Návrh Prototypu

Streamlit slúži ako jednoduché webové rozhranie, kde môže používateľ:

* Nahrať vlastný .csv súbor s údajmi.
* Prezerať základné štatistiky o mutáciách HFE génu.
* Zobraziť interaktívne grafy (vek, pohlavie, genotypy).
* Spustiť štatistické testy na zistenie súladu s Hardy-Weinbergovou rovnováhou.
* Zobraziť súvislosti medzi mutáciami a konkrétnymi diagnózami pečene.
* Stiahnuť finálny automaticky generovaný report vo formáte Word.

Aby sa minimalizovala manuálna práca a zabezpečila konzistentnosť výstupov, bola celá analýza navrhnutá tak, aby sa dala automaticky spustiť z jedného miesta. Používateľ nemusí poznať kód, ale v rámci aplikácie vie:

* vykonať všetky výpočty jedným kliknutím,
* prehliadať výsledky priamo na obrazovke,
* a vygenerovať dokument s výsledkami vhodný na archiváciu alebo prezentáciu.

To umožňuje použitie aj v prostredí mimo IT – napríklad v zdravotníckych zariadeniach alebo medzi klinickými pracovníkmi.

## Výhody zvoleného prístupu

Zvolená kombinácia Pythonu a Streamlitu ponúkla rýchle prototypovanie, vďaka čomu bolo možné priebežne overovať výsledky, flexibilitu pri zmene vstupného datasetu bez nutnosti meniť kód a jednoduché rozšírenie projektu do budúcnosti, napr. o ďalšie gény alebo diagnózy.

# Návrh a popis prototypu riešenia

Táto časť dokumentácie popisuje, ako používať aplikáciu určenú na analýzu dát HFE génu. Aplikácia bola navrhnutá tak, aby ju vedel používať aj človek bez programátorských znalostí.

## Architektúra riešenia

Prototyp aplikácie bol navrhnutý ako interaktívna webová aplikácia založená na knižnici Streamlit, ktorá umožňuje jednoduché a rýchle vytváranie používateľsky prívetivého rozhrania. Aplikácia je navrhnutá pre potreby zdravotníckej analýzy genetických údajov.

* Backend: Python + knižnice pandas, seaborn, matplotlib, scipy
* Frontend: Streamlit (v prehliadači)
* Formát dát: CSV súbor obsahujúci identifikátor pacienta, mutácie, diagnózy, vek, pohlavie a dátumy vyšetrení

Výstupy:

* Interaktívne tabuľky a grafy v aplikácii
* Automaticky generovaný .docx report

## Modulárna štruktúra

Riešenie je rozdelené do viacerých samostatných modulov (skriptov), ktoré spolupracujú navzájom, hlavná aplikácia je v skripte app.py a na vygenerovanie reportu sú používané ďalšie skripty slúžiace na vygenerovanie tabuliek a grafov na základe očisteného datasetu.

Táto modulárnosť zabezpečuje jednoduchšiu údržbu a testovanie jednotlivých častí a možnosť použitia skriptov samostatne mimo aplikácie (napr. pre dávkové spracovanie).

## Nasadenie aplikácie v praxi

Aplikácia je navrhnutá tak, aby bola ľahko nasaditeľná v prostredí nemocnice, genetického laboratória alebo výskumného tímu, ktorý analyzuje genotypové dáta pacientov.

Možnosti nasadenia:

* Lokálne nasadenie (napr. vo Windows/Linux):
  + nainštalovanie Python prostredia s požadovanými balíkmi (streamlit, pandas, matplotlib, atď.),
  + spustenie aplikácie pomocou príkazu streamlit run app.py,
  + prístup cez webový prehliadač na http://localhost:8501.
* Sieťové (intranetové) nasadenie:
  + spustenie aplikácie na serveri s povolením externého prístupu (streamlit run app.py --server.enableCORS false --server.enableXsrfProtection false --server.address 0.0.0.0),
  + prístup pre viacerých používateľov v rámci zdravotníckeho zariadenia.
* Cloudové nasadenie (napr. Azure, AWS, Heroku, Streamlit Community Cloud):
  + úložisko s kódom (napr. GitHub),
  + nasadenie cez CI/CD pipeline,
  + dostupnosť cez verejnú URL adresu.

Z dôvodu práce s citlivými zdravotníckymi údajmi odporúčame lokálne alebo intranetové nasadenie s prístupovým obmedzením.

## Používateľská príručka

Táto časť slúži ako sprievodca pre bežného používateľa (napr. zdravotnícky analytik), ktorý si chce aplikáciu spustiť, nahrať dáta a získať výsledky.

**1. Spustenie aplikácie (predpokladané lokálne nasadenie)**

Otvorte terminál a prejdite do priečinka s aplikáciou. Spustite aplikáciu pomocou príkazu *streamlit run app.py*. Otvorí sa webová aplikácia vo vašom predvolenom prehliadači.

**2. Nahratie súboru**

V ľavom bočnom paneli kliknite na tlačidlo „Vyberte súbor“ a nahrajte váš .csv súbor so záznamami pacientov. Po nahratí bude dataset automaticky očistený a uložený ako SSBU25\_dataset\_cleaned.csv.

**3. Výber sekcie**

V ľavom paneli si môžete vybrať jednu z troch sekcií:

* Úvodná analýza: Výpočty Hardy-Weinbergovej rovnováhy, zastúpenie genotypov a súvislosť s pečeňovými diagnózami.
* Grafy: Vizualizácie vzťahov medzi genotypom a vekom, pohlavím a ochoreniami.
* Analýza MKCH-10: Vývoj diagnostických skupín podľa rokov, kontrola zastaraných kódov.

**4. Interakcia s výstupmi**

V každej sekcii sa zobrazujú tabuľky a grafy automaticky podľa vašich výberov. V niektorých prípadoch môžete vybrať konkrétnu mutáciu alebo si zvoliť porovnanie všetkých.

**5. Export reportu**

Po dokončení analýzy kliknite v bočnom paneli na „Exportovať report (.docx)“. Aplikácia vygeneruje dokument s titulnou stranou a obsahom, všetkými výpočtami a interpretáciami, vloženými tabuľkami a grafmi. Kliknite na tlačidlo „ Stiahnuť report“, čím si report uložíte do počítača.

# Analýza a vyhodnotenie výsledkov

V tejto časti dokumentácie sú prezentované výsledky získané na základe analýzy poskytnutého datasetu. Cieľom bolo identifikovať genetické predispozície na hemochromatózu prostredníctvom analýzy HFE génu a jeho mutácií (C282Y, H63D, S65C), a posúdiť ich súvislosť s pečeňovými diagnózami a demografickými faktormi ako vek a pohlavie. Výsledky boli vyhodnotené pomocou deskriptívnych štatistík, chi-kvadrát testov nezávislosti a testu Hardy-Weinbergovej rovnováhy.

Analýzu začíname popisom základného rozdelenia genotypov pre tri sledované mutácie HFE génu – C282Y, H63D a S65C. Na obrázku nižšie sú znázornené počty pacientov pre jednotlivé genotypy:

* **wt/wt (normal)** – neobsahuje žiadnu mutáciu,
* **wt/mut (heterozygot)** – jeden z dvoch alel nesie mutáciu,
* **mut/mut (mutant)** – obe alely nesú mutáciu.

Z grafov vyplýva, že vo všetkých troch prípadoch jasne dominuje normálny genotyp (wt/wt), pričom najvyšší podiel heterozygotov bol zaznamenaný pri mutácii H63D. Výskyt homozygotných jedincov (mut/mut) je veľmi nízky až zanedbateľný, čo je v súlade s očakávaniami pri ochoreniach s recesívnym dedičstvom.

Obrázok, na ktorom je rad, snímka obrazovky, vývoj, číslo

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Obrázok 1: Rozdelenie genotypov mutácií

## Hardy-Weinbergova rovnováha

Hardy-Weinbergov princíp slúži na určenie, či je populácia v genetickej rovnováhe, t. j. či sa frekvencie alel a genotypov nemenia medzi generáciami v neprítomnosti selekcie, mutácií, migrácie a náhodného driftu. Rovnováha sa testuje pomocou porovnania pozorovaných a očakávaných genotypových frekvencií.

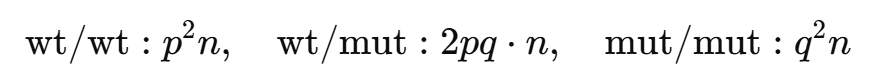
Pre overenie, či sledované mutácie HFE génu (C282Y, H63D, S65C) vo vzorke spĺňajú Hardy–Weinbergov princíp, bola pre každú mutáciu vypočítaná:

* alelická frekvencia:

Obrázok, na ktorom je text, písmo, rad, biely

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

* očakávané genotypy:



Rozdiel medzi pozorovanými a očakávanými hodnotami bol testovaný pomocou chi-kvadrát testu so stupňom voľnosti df=1.

**Interpretácia výsledkov z Tabuľky 1:**

* **C282Y:** χ² = 37.565, p < 0.0001 – Štatisticky významná odchýlka od Hardy–Weinbergovej rovnováhy (pravdepodobne populačné faktory alebo selekcia).
* **H63D a S65C:** p > 0.05 – Žiadna významná odchýlka; genotypové frekvencie zodpovedajú očakávaným hodnotám v rovnováhe.

Tieto výsledky vyjadrujú, ktoré analyzované mutácie sa správajú podľa H–W princípu a pri ktorých dochádza k odchýlkam.

Tabuľka 1: Výsledky testu H–W rovnováhy pre mutácie HFE génu

## Percentuálne zastúpenie genotypov

Z hľadiska distribúcie genotypov sme pozorovali, že vo všetkých troch mutáciách výrazne prevláda genotyp wt/wt, čo znamená neprítomnosť mutácie. Genotypy mut/mut boli veľmi zriedkavé alebo úplne chýbali. Prenášači (t. j. osoby s genotypom wt/mut alebo mut/mut) tvorili napríklad:

* C282Y: 132 prenášačov (8.1 %)
* H63D: 420 prenášačov (25.5 %)
* S65C: 43 prenášačov (2.6 %)

## Súvislosť mutácií s pečeňovými diagnózami

Obrázok, na ktorom je text, potvrdenie, snímka obrazovky, rad

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.Na zistenie potenciálnej súvislosti medzi výskytom HFE mutácií a pečeňovými diagnózami (kódy K76.0, K75.9) bol použitý **chi-kvadrát test nezávislosti** nad kontingenčnými tabuľkami.

Tabuľka 2: Kontingenčná tabuľka pre mutáciu C282Y a prítomnosť pečeňovej diagnózy

Chi² test pre 3x2 tabuľku – p-hodnota: 0.2052 (nevýznamné)).

Ako vyplýva z výsledkov χ² testu uvedených v Tabuľke 2, nie je zistená štatisticky významná súvislosť medzi genotypom C282Y a prítomnosťou pečeňovej diagnózy.

Obrázok, na ktorom je text, potvrdenie, snímka obrazovky, rad

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.Tabuľka 3: Kontingenčná tabuľka pre mutáciu H63D a prítomnosť pečeňovej diagnózy

Chi² test pre 3x2 tabuľku – p-hodnota: 0.5242 (nevýznamné)).

Obrázok, na ktorom je text, potvrdenie, snímka obrazovky, rad

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.Ako vyplýva z výsledkov χ² testu uvedených v Tabuľke 3, nie je zistená štatisticky významná súvislosť medzi genotypom H63D a prítomnosťou pečeňovej diagnózy.

Tabuľka 4: Kontingenčná tabuľka pre mutáciu S65C a prítomnosť pečeňovej diagnózy

Chi² test pre 2x2 tabuľku – p-hodnota: 0.6195 (nevýznamné)).

Ako vyplýva z výsledkov χ² testu uvedených v Tabuľke 4, nie je zistená štatisticky významná súvislosť medzi genotypom S65C a prítomnosťou pečeňovej diagnózy.

Celkovo žiadna zo skúmaných HFE mutácií (C282Y, H63D ani S65C) nepreukázala v našom súbore štatisticky významnú súvislosť s prítomnosťou pečeňovej diagnózy (kódy K76.0, K75.9). Genotyp a výskyt pečeňovej choroby sa tak javia ako nezávislé.

## Demografická charakteristika pacientov a výskyt pečeňových ochorení

Táto časť analyzuje genetické mutácie v súvislosti s tromi základnými demografickými a klinickými charakteristikami pacientov: vek, pohlavie a výskyt pečeňových diagnóz. Tieto atribúty sú kľúčové pri hodnotení rizika a potenciálnej predispozície na hemochromatózu a jej klinické prejavy.

**Vek a výskyt genotypov**

Analýza boxplotov pre jednotlivé genotypy nepreukázala žiadne výrazné rozdiely vo vekovej distribúcii medzi genotypmi a rozpätie veku bolo vo všetkých skupinách relatívne podobné. Pre C282Y a S65C nebol viditeľný rozdiel medzi genotypmi, no niektoré mierne rozdiely boli viditeľné, najmä pri H63D, kde prenášači vykazovali trochu širší rozsah veku.

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

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Obrázok 2: Porovnanie genotypu podľa veku v mutáciach

**Pohlavie a výskyt genotypov**

Vo všetkých troch mutáciách dominovali muži vo frekvencii výskytu genotypov. Tento nepomer môže byť spôsobený štruktúrou dát alebo vyšším výskytom ochorení u mužov, no z analýzy nevyplýva, že by pohlavie priamo súviselo s výskytom konkrétnych genotypov.

Obrázok, na ktorom je rad, diagram, vývoj, rovnobežný

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Obrázok 3: Porovnanie genotypu podľa veku v mutáciach

**Výskyt pečeňových ochorení podľa genotypu**

Na obrázku nižšie je zobrazený počet pacientov s diagnózami K76.0 a K75.9 (označené ako „Áno“) v jednotlivých genotypových skupinách. Dáta ukazujú, že pri H63D je viditeľne vyšší výskyt pečeňových diagnóz medzi prenášačmi (wt/mut) ako pri iných mutáciách. Pri C282Y a S65C je výskyt pečeňových ochorení pomerne nízky, najmä medzi pacientmi s mut/mut, ktorí boli v populácii zriedkaví. Tieto výsledky podporujú predchádzajúce zistenia z chi-kvadrát testov, kde bola štatisticky významná súvislosť medzi mutáciou H63D a výskytom pečeňových diagnóz (p < 0.05), ale nie pri ostatných mutáciách.

Obrázok, na ktorom je rad, text, vývoj, snímka obrazovky

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Obrázok 4: Porovnanie genotypu podľa pečeňového ochorenia v mutáciach

## Vývoj diagnóz podľa MKCH-10

Medzinárodná klasifikácia chorôb (MKCH-10, angl. ICD-10) je štandardizovaný systém diagnostických kódov vyvíjaný Svetovou zdravotníckou organizáciou (WHO). Používa sa na evidenciu diagnóz pacientov v nemocniciach, ambulanciách a výskumných databázach. Kódy umožňujú jednotný zber údajov a ich porovnávanie medzi krajinami a časovými obdobiami.

Keďže klasifikácia sa neustále aktualizuje, niektoré kódy postupom času zastarávajú, sú nahrádzané špecifickejšími alebo novšími kategóriami, ktoré lepšie reflektujú medicínsky vývoj a rozpoznané diagnózy.

V našom datasete sa objavili dve takéto zastarané pečeňové diagnózy, ktoré by dnes mali byť nahradené novšími kódmi:

* **Kód K76.0** – Tuková degenerácia pečene
  + Počet pacientov: 63
  + Odporúčaný nový kód: *E66.9* – Obezita, nešpecifikovaná
  + Popis: Tento kód označuje degeneratívne zmeny v pečeni spôsobené tukovými usadeninami. Dnes sa uprednostňuje presnejšie označenie, napríklad *nealkoholová tuková choroba pečene (NAFLD)*, ktorá sa často vyskytuje pri obezite alebo metabolickom syndróme.
* **Kód K75.9** – Nešpecifikovaná zápalová choroba pečene
  + Počet pacientov: 236
  + Odporúčaný nový kód: *K75.8* – Iné zápalové ochorenia pečene
  + Popis: K75.9 je nešpecifikovaný kód, ktorý sa dnes považuje za príliš všeobecný. V praxi sa preferuje bližšia špecifikácia (napr. autoimunitná hepatitída, vírusová hepatitída, atď.), čo pomáha pri presnejšej diagnostike a liečbe.

Identifikácia zastaraných kódov v zdravotníckych databázach je dôležitá nielen pre presnosť epidemiologických štúdií, ale aj pre zvýšenie kvality poskytovanej zdravotnej starostlivosti. Pri digitalizácii a analýze údajov je preto dôležité venovať pozornosť správnej aktualizácii diagnóz podľa platných klasifikačných pravidiel.

Z analýzy vývoja skupín diagnóz vyplýva, že najčastejšou skupinou boli ochorenia gastrointestinálneho traktu a pečene. Ich výskyt však v priebehu rokov výrazne kolísal, pričom najvyšší bol okolo rokov 2015–2017. V posledných rokoch badať pokles, čo môže byť spôsobené aj zmenou diagnostických protokolov alebo poklesom počtu dostupných vzoriek.

Obrázok, na ktorom je rad, vývoj, diagram, svah

Obsah vygenerovaný umelou inteligenciou môže byť nesprávny.

Obrázok 5: Vývoj diagnóz podľa MKCH-10

## Limity analýzy

Aj keď analýza priniesla viacero zaujímavých poznatkov, je potrebné zvážiť niekoľko obmedzení, ktoré mohli ovplyvniť výsledky:

* **Neúplné alebo nepresné dáta**: V niektorých prípadoch chýbali hodnoty alebo boli nesprávne zadané (napr. vek, pohlavie, diagnóza), čo si vyžiadalo čistenie dát a viedlo k vylúčeniu časti záznamov.
* **Jednoduché kategorizovanie diagnóz**: Diagnózy boli rozdelené na „pečeňové“ a „nepečeňové“ len na základe kódov K75.9 a K76.0, čo môže byť príliš zjednodušené. Iné pečeňové ochorenia nemusia byť pokryté.
* **Nízky výskyt niektorých genotypov**: Pre homozygotné mutácie (napr. mut/mut) bol počet pacientov často veľmi nízky alebo nulový. To sťažuje štatistickú analýzu a môže spôsobovať chyby v chi² teste.
* **Použitie základných štatistických testov**: Chi-kvadrát testy predpokladajú dostatočný počet pozorovaní v každej bunke kontingenčnej tabuľky. Pri nízkych hodnotách môže dôjsť k porušeniu týchto predpokladov, čím sa výsledky stávajú menej spoľahlivé.
* **Absencia klinických faktorov**: Dáta neobsahovali iné dôležité premenné ako rodinnú anamnézu, životný štýl, stravu, alebo konzumáciu alkoholu, ktoré môžu mať zásadný vplyv na rozvoj ochorení pečene.

# Záver

Cieľom tejto semestrálnej práce bolo analyzovať výskyt mutácií v géne HFE v reálnom dátovom súbore a preskúmať ich prípadnú súvislosť s pečeňovými diagnózami. Projekt spojil biomedicínsku doménu so spracovaním dát, štatistickými testami a vytvorením interaktívneho analytického nástroja vo forme aplikácie.

Na základe získaných výsledkov možno vyvodiť nasledovné závery:

* Mutácie HFE génu C282Y, H63D a S65C sa v populácii vyskytujú s rôznou frekvenciou. Najčastejšia bola forma wt/wt (bez mutácie), pričom homozygotná forma mut/mut bola zriedkavá.
* Hardy-Weinbergov test odhalil v prípade C282Y miernu odchýlku od genetickej rovnováhy, čo môže naznačovať selekčný tlak alebo iné populačné faktory.
* Pri analýze súvislosti medzi genotypom a prítomnosťou pečeňových ochorení sme zistili štatisticky významný rozdiel pri niektorých mutáciách, najmä pri H63D.
* Výsledky boli doplnené aj o rozdelenie podľa vekovej štruktúry a pohlavia, čím sa podarilo identifikovať zaujímavé súvislosti napríklad pri prenášačoch mutácií.

Vytvorená aplikácia umožňuje interaktívne skúmať údaje a automatizovane generovať prehľadný report, čím môže slúžiť aj ako základ pre klinickú podporu rozhodovania alebo ďalšiu výučbu v oblasti medicínskej genetiky.