

# Predikce rizika úmrtí u pacientů se srdečním selháním na základě klinických záznamů

## 1. Úvod

Kardiovaskulární onemocnění, mezi které se řadí i srdeční selhání, představují celosvětově nejčastější příčinu úmrtí (World Health Organization, 2025). Včasná identifikace rizikových faktorů a predikce vývoje nemoci jsou proto naprosto klíčové pro správné nastavení léčby a prodloužení života pacientů (Chicco & Jurman, 2020). Vzhledem ke komplexnosti tohoto syndromu se dnes pro odhad individuálního rizika úmrtí stále častěji uplatňují analytické a kvantitativní metody aplikované na klinická data (Kostić et al., 2025). Statistické nástroje, jako je právě logistická regrese, umožňují lékařům odhalit skryté vzorce v elektronických zdravotních záznamech a spolehlivě odhadnout pravděpodobnost fatální události na základě více klinických proměnných současně (Chicco & Jurman, 2020; Kostić et al., 2025).

## 2. Metodologie a příprava dat

Pro účely této práce byla použita volně dostupná data „*Heart failure clinical records dataset*.“ (UC Irvine Machine Learning Repository, 2020)<sup>1</sup>. Práce s daty byla provedena v programu TIBCO Statistica. Závislou (vysvětlovanou) proměnnou byl binární znak DEATH\_EVENT, který nabývá hodnot 1 (pacient zemřel) a 0 (pacient přežil). Kódování bylo záměrně nastaveno takto, protože v binomické logistické regresi hodnota 1 standardně označuje sledovanou událost, jejíž šanci se snažíme předpovědět. Jelikož je cílem této analýzy zkoumat mortalitu, definování úmrtí jako události „1“ usnadňuje a zpřehledňuje interpretaci výsledků. Kladné hodnoty odhadů u rizikových faktorů tak intuitivně znamenají zvyšující se riziko smrti.

Jako nezávislé proměnné (prediktory) bylo do modelu zařazeno celkem 11 klinických faktorů, zahrnujících jak spojité, tak kategoriální proměnné. Spojité proměnné představovaly věk (age), hladinu kreatininu (serum\_creatinine), sodíku (serum\_sodium) a kreatinkinázy (creatinine\_phosphokinase) v krvi, počet krevních destiček (platelets) a ejekční frakci (ejection\_fraction) určující procento krve, které srdce vypumpuje při každém stahu.

Kategoriální proměnné, respektive binární znaky, nabývající hodnot 0 a 1, zahrnovaly přítomnost anémie/chudokrevnosti (anaemia), přítomnost cukrovky (diabetes), výskyt vysokého krevního tlaku (high\_blood\_pressure), pohlaví pacienta (sex) a kuřáckou anamnézu (smoking). Referenční hodnotou (0) byla vždy nepřítomnost daného rizikového faktoru, v případě pohlaví pak 0 = žena, 1 = muž.

Z původního datového souboru byla před samotnou analýzou záměrně vyřazena proměnná určující dobu sledování pacienta (time). Důvodem je její metodologická nevhodnost pro tento typ

---

<sup>1</sup> Data a další informace o této zprávě jsou dostupné na adrese:  
<https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records>

predikce. Délka sledování přímo koreluje s cílovou událostí. Pacienti, kteří zemřeli dříve, mají logicky kratší dobu sledování. Zahrnutí této proměnné by vedlo k metodologické chybě a uměle by nadhodnotilo přesnost modelu na úkor skutečných fyziologických a biologických faktorů, jejichž vliv je primárním předmětem této práce.

Pro odhad vlivu těchto fyziologických faktorů na přežití pacientů byla využita metoda vícenásobné logistické regrese.

### 3. Výsledky

Jak ukazuje test globální nulové hypotézy (Likelihood Ratio Chi-Square LR  $\chi^2 = 81,07$ ;  $p < 0,001$ ), navržený model s 11 klinickými prediktory je jako celek statisticky vysoce významný. To znamená, že vybrané klinické proměnné skutečně dokážou společně predikovat úmrtí pacienta a nejedná se o náhodu.

Podle tabulky statistik shody modelu dosahuje pseudo-koefficient determinace (Nagelkerke  $R^2$ ) hodnoty 0,332. Tato hodnota indikuje přijatelnou, středně silnou, shodu modelu s daty (tzv. goodness-of-fit) a představuje prokazatelné zlepšení prediktivní schopnosti oproti nulovému modelu.

Pro zjednodušení výsledky shrnuje Tabulka 1.

**Tabulka 1**

*Hodnocení kvality a celkové prediktivní síly modelu*

LR $\chi^2$	p	Nagelkerke $R^2$
81,07	< 0,001	0,332

Z detailní analýzy odhadů parametrů pro jednotlivé proměnné vyplývá, že statisticky významný vliv ( $p < 0,05$ ) na úmrtí pacienta mají v našem modelu čtyři faktory:

- *Věk (age)*: ( $p < 0,001$ ). Věk je vysoce signifikantním rizikovým faktorem. Kladná hodnota regresního koeficientu ( $\beta = 0,056$ ) ukazuje, že s rostoucím věkem se přímo úměrně zvyšuje šance na úmrtí pacienta.
- *Ejekční frakce (ejection\_fraction)*: ( $p < 0,001$ ). Jde o nejsilnější klinický prediktor v modelu. Záporná hodnota regresního koeficientu ( $\beta = -0,070$ ) potvrzuje, že čím vyšší je procento krve, které srdce dokáže vypumpovat (lepší funkce srdce), tím nižší je šance úmrtí.
- *Hladina kreatininu v krvi (serum\_creatinine)*: ( $p < 0,001$ ). Kladná hodnota regresního koeficientu ( $\beta = 0,662$ ) ukazuje, že zvýšené hodnoty kreatininu, které poukazují na selhávající funkci ledvin, prokazatelně zvyšují šance smrti.
- *Kreatinkináza (creatinine\_phosphokinase)*: ( $p = 0,042$ ). Tento faktor vyšel jako statisticky významný na hranici hladiny významnosti  $\alpha = 0,05$ . Ačkoliv kladný regresní koeficient

( $\beta = 0,0003$ ) naznačuje zvyšující se šance, z praktického klinického hlediska je tento vliv oproti předchozím faktorům minimální.

Ostatní zkoumané faktory, včetně často zmiňovaných rizik jako je kouření ( $p = 0,697$ ), diabetes ( $p = 0,611$ ) nebo vysoký krevní tlak ( $p = 0,171$ ), se v přítomnosti výše zmíněných silnějších prediktorů (věk, srdce, ledviny) nepotvrdily jako statisticky významné pro přímočaré odhadnutí mortality. To ale pochopitelně neznamená, že tyto faktory nepředstavují pro pacienty zdravotní riziko, ale mohou se projevovat spíše nepřímo.

Hodnoty odhadnutých regresních parametrů  $\beta$  nejsou samy o sobě přímo interpretovatelné v měřítku pravděpodobnosti, protože v regresní funkci figurují jako exponenty. Mnohem názornějším ukazatelem je proto tzv. poměr šancí (Odds Ratio – OR). Ten vypočítáme umocněním Eulerova čísla konkrétním parametrem  $e^\beta$ , kde  $e$  je matematická konstanta rovna přibližně 2,718. Poměr šancí nám pak pro naše klíčové signifikantní proměnné říká následující:

- *Hladina kreatininu*: Pro parametr  $\beta = 0,662$  je poměr šancí roven  $e^{0,662} = 1,938$ . To znamená, že pokud se hladina kreatininu v krvi zvýší o jedničku (o 1 mg/dL), šance na úmrtí pacienta se zvýší téměř dvojnásobně (zhruba o 94 %).
- *Věk*: Pro parametr  $\beta = 0,056$  je poměr šancí roven  $e^{0,056} = 1,057$ . Každý jeden rok věku navíc tedy násobí šanci na úmrtí pacienta hodnotou 1,057. Šance úmrtí tedy roste přibližně o 5,7 % s každým dalším rokem života.
- *Ejekční frakce*: Pro parametr  $\beta = -0,070$  je poměr šancí roven  $e^{-0,070} = 0,932$ . Protože je beta záporná, poměr šancí je menší než 1. Znamená to, že pokud se ejekční frakce zlepší o 1 %, šance na úmrtí pacienta se násobí hodnotou 0,932. Šance se tedy s každým procentem výkonu srdce snižuje přibližně o 6,8 %. Podrobnosti statistické významnosti a poměru šancí shrnuje Tabulka 2.

## Tabulka 2

Odhady parametrů logistické regrese a poměry šancí (OR)

<i>Klinický faktor</i>	<i><math>\beta</math></i>	<i>OR</i>	<i>95% CI</i>	<i>p</i>
<i>Věk (age)</i>	0,056	1,057	[1,030 – 1,085]	< 0,001
<i>Ejekční frakce (ejection_fraction)</i>	-0,070	0,932	[0,905 – 0,960]	< 0,001
<i>Hladina kreatininu (serum_creatinine)</i>	0,662	1,938	[1,380 – 2,723]	< 0,001
<i>Kreatinkináza (creatinine_phosphokinase)</i>	0,0003	1,0003	[1,0000 – 1,0006]	0,042
<i>Krevní destičky (platelets)</i>	0,000	1,000	[1,000 – 1,000]	0,661
<i>Hladina sodíku (serum_sodium)</i>	-0,057	0,945	[0,885 – 1,009]	0,090
<i>Pohlaví (sex)</i>	0,399	1,490	[0,749 – 2,964]	0,255
<i>Diabetes (diabetes)</i>	-0,151	0,859	[0,480 – 1,539]	0,611
<i>Vysoký krevní tlak (high_blood_pressure)</i>	-0,419	0,658	[0,361 – 1,198]	0,171
<i>Kouření (smoking)</i>	-0,136	0,873	[0,441 – 1,729]	0,697
<i>Anémie (anaemia)</i>	-0,418	0,658	[0,365 – 1,188]	0,165

### 4. Závěr

Aplikací vícenásobné logistické regrese se podařilo úspěšně vytvořit explanační model, který na základě klinických dat hodnotí šanci úmrtí pacientů se srdečním selháním. Očištěním modelu o metodologicky nevhodnou proměnnou (dobu sledování pacienta) bylo dosaženo korektního odhadu vlivu čistě biologických a fyziologických faktorů.

Nejdůležitějším zjištěním je identifikace klíčových biomarkerů. Analýza prokázala, že primárním přímým hybatelem bezprostředního rizika úmrtí není životní styl pacienta (kouření, vysoký krevní tlak apod.), ale objektivní stav jeho orgánů, konkrétně funkce srdeční pumpy (ejekční frakce) a stav ledvin (hladina kreatininu v krvi), ruku v ruce s rostoucím věkem pacienta. Ačkoliv se jako statisticky významná ukázala i hladina kreatinkinázy, její reálný příspěvek k nárůstu rizika je v porovnání s výše zmíněnými faktory zanedbatelný.

Věk, ejekční frakce a hladina kreatininu by proto měly být pro ošetřující lékaře absolutní prioritou při určování bezprostřední prognózy. Zjištění, že faktory životního stylu nevyšly v tomto modelu jako statisticky významné, však nelze interpretovat tak, že pro pacienta nepředstavují riziko. Jejich vliv na mortalitu je pravděpodobně zprostředkovaný. Dlouhodobé kouření či hypertenze postupně zhoršují právě funkci srdce a ledvin, které se následně v modelu projevují jako hlavní a bezprostřední prediktory úmrtí.

## Zdroje

Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, 20(1). <https://doi.org/10.1186/s12911-020-1023-5>

Kostić, S., et al. (2025). A logistic regression-based model for predicting heart failure mortality. *Megatrend revija*, 22(1), 57–68. <https://scindeks-clanci.ceon.rs/data/pdf/2334-9638/2025/2334-96382501057K.pdf>

TIBCO Software Inc. (2020). *TIBCO Statistica* (Verze 14.0) [Počítačový software].

UC Irvine Machine Learning Repository. (2020). *Heart failure clinical records dataset*. <https://archive.ics.uci.edu/dataset/519/heart+failure+clinical+records>

World Health Organization. (2025). *Cardiovascular diseases (CVDs)*. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))