

Prediktory výskytu cévní mozkové příhody

Teoretická východiska

Cévní mozková příhoda je akutní a život ohrožující stav vznikající při přerušení krevního zásobení mozku v důsledku uzávěru cévy nebo krvácení (WHO, 2025). V roce 2021 patřila mezi nejvýznamnější příčiny úmrtí na světě, přičemž nejdůležitějším ovlivnitelným rizikovým faktorem je arteriální hypertenze (Nussbaumerová, 2021). Mezi další významné **rizikové faktory** patří kouření, diabetes, vysoký cholesterol, obezita, fyzická neaktivita a nezdravý životní styl. Pro snížení následků je zásadní rychlá diagnostika a okamžitá léčba, zatímco v prevenci hraje klíčovou roli zejména kontrola krevního tlaku a celkové ovlivnění kardiovaskulárních rizik (WHO, 2025).

Výzkumy zároveň ukazují, že výskyt CMP výrazně souvisí se stárnutím populace a že s rostoucím podílem osob starších 65 let narůstá i celková zátěž tohoto onemocnění pro zdravotní systémy (Thayabaranathan et al., 2022). Další studie potvrzují, že vedle věku mají zásadní vliv na pravděpodobnost vzniku CMP také kardiovaskulární komorbidity a metabolické faktory, zejména hypertenze, srdeční onemocnění a zvýšená hladina glukózy v krvi, které významně přispívají k růstu morbidit i mortality spojené s tímto onemocněním (Pathan et al., 2020).

Pro účely statistické analýzy byl využit *Brain stroke prediction dataset* ($n = 4981$) z platformy Kaggle (Akbasli, 2021; Pathan et al., 2020). Soubor obsahuje 11 proměnných popisujících vybrané sociodemografické a zdravotní charakteristiky respondentů. Struktura dat tak umožňuje aplikaci statistických metod zaměřených na odhad pravděpodobnosti výskytu sledovaného jevu. Zároveň je důležité zmínit, že pouze zhruba 5 % respondentů CMP prodělalo, což naznačuje slabší predikční schopnost našeho modelu.

Logistická regrese

Pro analýzu dat jsme zvolili logistickou regresi, jelikož závislá proměnná *CMP* má dichotomický charakter. Tato metoda je určena pro modelování situací, kdy je cílem odhadnout přítomnost nebo nepřítomnost sledovaného jevu na základě souboru prediktorů. Její výhodou je možnost zahrnout do modelu jak spojité, tak kategoriální nezávislé proměnné (IBM, 2026).

Závislá proměnná

- *CMP* – výskyt cévní mozkové příhody
pacient cévní mozkovou příhodou neprodělal = 0
pacient prodělal cévní mozkovou příhodou = 1

Nezávislé proměnné

- *Pohlaví* – muž = 0, žena = 1
- *Věk*
- *Hypertenze* – výskyt hypertenze = 1
- *Srdeční onemocnění* – výskyt srdečního onemocnění = 1
- *Manželství* – pacient dříve nebo aktuálně v manželství = 1
- *Pracovní status* – kategorie: státní zaměstnanec (ref.), dítě, soukromý sektor, OSVČ
- *Bydliště* – venkov = 0, město = 1

- *Glukóza v krvi* – průměrná hladina glukózy v krvi
- *BMI* – Index tělesné hmotnosti
- *Kouření* – kategorie: kouří (ref.), dříve kouřil/a, nikdy nekouřil/a, neznámé

Výsledky

Tabulka 1: Výsledky logistické regrese

Faktor	β	SE	Odds Ratio	Wald	p
Intercept	-7,506	0,552		185,054	<0,001
vek	0,075	0,006	1,078	163,871	<0,001
glukoza_v_krvi	0,004	0,001	1,004	9,964	0,002
bmi	0,011	0,013	1,011	0,741	0,389
pohlavi	0,004	0,071	1,007	0,002	0,961
hypertenze	-0,208	0,083	0,659	6,367	0,012
srdecni_onemocneni	-0,136	0,096	0,762	2,030	0,154
pracovni_status_dite	0,801	0,620	2,796	1,672	0,196
pracovni_status_soukromy_sektor	-0,106	0,220	1,128	0,232	0,630
pracovni_status_OSVC	-0,468	0,245	0,785	3,660	0,056
manzelstvi	0,097	0,113	1,213	0,732	0,392
bydliste	-0,044	0,069	0,916	0,401	0,526
koureni_nezname	-0,022	0,135	0,837	0,026	0,872
koureni_nekouřil	-0,179	0,111	0,715	2,635	0,105
koureni_drive_kouřil	0,045	0,119	0,895	0,142	0,706

Pozn.: β = nestandardizovaný regresní koeficient, SE = směrodatná chyba, Odds Ratio = poměr šancí

Logistická regrese ukázala, že statisticky významným prediktorem výskytu cévní mozkové příhody jsou věk, hladina glukózy v krvi a hypertenze. S rostoucím věkem se šance na výskyt cévní mozkové příhody zvyšuje, konkrétně každý další rok zvyšuje poměr šancí výskytu o 7,8 %. Vyšší hladina glukózy v krvi rovněž zvyšuje pravděpodobnost výskytu cévní mozkové příhody, a to o 0,4 % při zvýšení o jednu jednotku. V případě hypertenze bylo zjištěno, že osoby bez hypertenze mají nižší šanci na výskyt cévní mozkové příhody než osoby s hypertenzí, což potvrzuje negativní vliv hypertenze na zdravotní stav respondentů. Ostatní proměnné zahrnuté do modelu nebyly statisticky významné, ale uvádíme je v tabulce, abychom poskytli transparentnost našich dat.

Na modelu byl proveden test jeho kvality pomocí Waldovy statistiky, na základě zjištěné p-hodnoty ($p < 0,001$) lze usuzovat, že predikce našeho modelu je nenáhodná. Pro posouzení kvality modelu byl dále sledován Nagelkerkeho R^2 , jeho hodnota dosáhla 0,239.

Klasifikace případů ukázala, že model velmi přesně identifikuje případy bez výskytu cévní mozkové příhody, u nichž dosáhl 100% úspěšnosti klasifikace. Naopak u případů s výskytem cévní mozkové příhody byla úspěšnost klasifikace pouze 0,4 %, protože model správně rozpoznal pouze 1 případ z celkových 248 pozitivních případů. Výsledky tedy naznačují, že model má velmi nízkou citlivost při identifikaci jedinců s cévní mozkovou příhodou a má tendenci zařazovat pozorování převážně do kategorie bez výskytu onemocnění. Z praktického hlediska to znamená, že model není příliš vhodný pro predikci výskytu cévní mozkové příhody, přestože některé prediktory byly statisticky významné.

Tabulka 2: Klasifikace případů

	Předpovězené: 1	Předpovězené: 0	Procento správných
Pozorované: 1	1	247	0,403
Pozorované: 0	0	4733	100

Závěr

Závěrem lze shrnout, že logistická regrese potvrdila jako statisticky významné prediktory výskytu cévní mozkové příhody věk, hladinu glukózy v krvi a hypertenzi. S rostoucím věkem a vyšší hladinou glukózy v krvi se zvyšovala pravděpodobnost výskytu onemocnění, významný byl také vliv hypertenze. Ostatní sledované proměnné se jako statisticky významné neprokázaly. Model však vykazoval jen omezenou vysvětlovací a velmi nízkou predikční schopnost při rozpoznávání pozitivních případů, a proto jej nelze považovat za dostatečně silný nástroj pro přesnou predikci výskytu cévní mozkové příhody.

Zdroje

Akbasli, I. T. (2021). *Brain stroke prediction dataset*. Kaggle. <https://www.kaggle.com/datasets/zzettrkalpakbal/full-filled-brain-stroke-dataset>

IBM. (2026, 20. ledna). *Logistic Regression*. <https://www.ibm.com/docs/en/spss-statistics/cd?topic=regression-logistic>

Nussbaumerová, B. (2021). Léčba hypertenze u cévních mozkových příhod. *Medicína pro praxi*, 17(5), 300-303. <https://solen.cz/pdfs/med/2020/05/03.pdf>

Pathan, M. S., Jianbiao, Z., John, D., Nag, A., & Dev, S. (2020). Identifying stroke indicators using rough sets. *IEEE Access*, 8, 210318-210327. doi: 10.1109/ACCESS.2020.3039439.

Thayabaranathan, T., Kim, J., Cadilhac, D. A., Thrift, A. G., Donnan, G. A., Howard, G., ... & Olaiya, M. T. (2022). Global stroke statistics 2022. *International Journal of Stroke*, 17(9), 946-956. doi:10.1177/17474930221123175.

TIBCO Software Inc. (2020). *Statistica (verze 14.0) [Software]*. TIBCO Software Inc. Retrieved from <https://www.tibco.com>

World Health Organization. (2025, 19. prosince). *Stroke*. <https://www.who.int/news-room/fact-sheets/detail/stroke>