

Big data, strojové učení a umělá inteligence v klinické laboratoři. Pojmy a literatura k edukaci

Friedecký B.

Ústav klinické biochemie a diagnostiky, Fakultní nemocnice, Hradec Králové

SOUHRN

Práce s big daty vyžaduje použití prostředků umělé inteligence. Přináší to možnost transformace laboratorních výsledků do formy strojového učení-machine learning (ML). Od něho se očekává aktivace dat, přinášející zlepšení diagnostických možností laboratorních vyšetření. Jde o posuv od použití počítačů, sloužících z části jako skladiště mrtvých dat, k aktivnějšímu využití jejich potenciálu pro diagnostiku, management, edukaci, výzkum a další. Zejména pak k predikci stavu chorob a k precizní medicíně v onkologii i jinde. Důsledkem by měl být integrovaný mezioborový přístup k diagnostice a reálné dosažení efektivní personalizace při diagnostice a terapii pacientů. Sdělení je pokusem o pomoc při zavádění práce s big daty a umělou inteligencí v klinických laboratořích. Vychází z faktu obrovské akcelerace tohoto přístupu, zdaleka nejen pouze v laboratorní medicíně.

Klíčová slova: big data, umělá inteligence, machine learning (strojové učení).

SUMMARY

Friedecký B.: Big data, machine learning and artificial intelligence in clinical laboratory. Concepts and literature for education

Working the big data needs using of artificial intelligence tools. This approach introduced currently into practice by large velocity leads to machine learning. Machine learning should be a strong way namely for the prediction of patient's state, for precision medicine in oncology and many more cases. For example for aiming the real personalisation of patients in dese of their diagnosis and therapy. This work can be a helpful tool for the introduction of artificial intelligence in routine clinical laboratories.

Keywords: big data, artificial intelligence, machine learning.

Úvod

Vysvětlení pojmu big data získáme srozumitelně na stránkách <https://www.oracle.com/cz/big-data/what-is-big-data> Tam lze nalézt základní poznatky o pojmu, charakteru a aplikacích big dat.

Definice big dat je dobře vyjádřena zkratkou RRR (rozsah-různorodost, rychlost). V původní anglické verzi jde o pojem VVV (velocity, volume, variety). Laboratorní big data by měla v ideálním případě integrovat kromě výsledků analýz demografické údaje, medikaci, komorbiditu, genetická data a také data zobrazovacích metod. Nyní je také (konečně) silně zdůrazňovaná nutnost jejich věrohodnosti. V prostředí klinických laboratoří znamená věrohodnost standardizaci, harmonizaci, validaci a přiměřenou velikost hodnoty nejistoty měření.

Machine learning - strojové učení je systém, schopný automatického rozhodovacího procesu v laboratoři na podkladě big dat, hodnocených umělou inteligencí (dále ML/AI). Big data nelze ostatně vyhodnocovat jinak, než nástroji umělé inteligence. Webinar IFCC Live Science z 23. května 2022 shrnul stručně podmínky a důvody pro použití postupů ML v klinických laboratořích. Jde o postupující procesy ztrát zkušeností a dovedností zdravotního personálu (deskilling)

a o lepší využití obrovského množství dosud nepoužívaných, ale existujících dat, které mají prokázanou potenci existujícími nástroji umělé inteligence zlepšit efektivitu diagnostiky, managementu a rozhodovacích procesů. Za největší problémy nastupujících procesů ML jsou považovány nejistota platnosti diagnostických rozhodnutí, rozdíly při použití různých modelů ML, nedostatek infrastruktury k provádění ML (zejména software) a etické problémy, plynoucí z použití dat a z předpisů o jejich ochraně.

Doporučená literatura a její komentář

K demonstraci a komentování je vybrána co nejnovější literatura. Velkou většinou jde o práce, které jsou v plném znění volně přístupné, bez poplatků nebo požadování. (free PMC article) pouhým kliknutím na webovou adresu v seznamu literatury uvedenu. Tedy s ohledem na předpokládanou budoucí možnost elektronické verze časopisu. Pokud jsou uvedena jen abstrakta nebo zkrácené verze prací, lze je na požádání (nebo za platbu) snadno získat.

Big data jsou vhodným, ne-li již nezbytným nástrojem k metaanalýze laboratorních a klinických dat. Studií na toto téma je celá řada. Aktuální jsou v současnosti například data o důsledcích covido-

vé pandemie [1, 2]. Zde analýza big dat (pacienti z Veteran Affairs US) ukázala významný postkovidový nárůst rizika diabetu a závislost velikosti tohoto rizika na závažnosti onemocnění u kohorty stovek tisíc pacientů s detekovaným COVID-19 ve srovnání s kohortou cca pěti milionů osob v období před pandemií. Rozšířená analýza podobného typu big dat s velmi názornou vizualizací zvýšeného rizika diagnózy chorob 11 orgánových systémů ukazuje obzvláště zvýšené riziko nárůstu diagnóz akutního koronárního syndromu, poruch koagulace, centrálních mozkových příhod, ale i psychických poškození.

Shrnutí základních potřebných znalostí o big datech, umělé inteligenci, machine learningu v přístupné podobě pro laiky a možnostech získávání informací k hodnocení stavu pacientů vizualizací nalezených dat (data mining) umožní čtenáři další využívání literatury tohoto oboru [3].

Co očekávají od digitalizace laboratoří autoři, zabývající se klasickou "rutinní" klinickou biochemií je možno nalézt zejména v několika pracích italských autorů [4-6]. V laboratorní medicíně je k dispozici řada aplikací strojového učení - machine learningu (dále ML), jejich počet velmi rychle přibývá. Zejména jde o studie vztahů mezi některými analyty v jistých klinických situacích, dále z oblasti predikce velikosti rizika choroby. Zda, v jakých oblastech a nakolik se ML razantně a rutinně uplatní, prozatím jasné není. Favoritem praktického uplatnění je onkologie, kardiologie a septické stavy. Perspektivním polem pro ML jsou podle názorů klinických biochemiků rozhodně referenční intervaly a rozhodovací limity, autovalidace a také práce s výběrem, tříděním a prezentací odborné literatury a studií.

Určité rozpaky, nakolik ovlivní digitalizace charakter role klinických laboratoří, však přetrvávají. Digitalizace vede k integrované, interdisciplinární diagnostice, slučující data z laboratoří s daty zobrazovacích metod, demografickými údaji, daty o medikaci, komorbiditych a i dalších [6] a je nezbytným předpokladem EHR (electronic health record). Práce s big daty a umělou inteligencí je nebo aspoň by měla být rozhodně pro laboratorní profesionály výzvou [4].

Obtížný se jeví za dané situace problém s kontrolou a věrohodností dat POCT [5], neboť produkce POCT nezděrně kvantitativně a kvalitativně roste a proniká do všech odvětví medicíny a k pacientům všech onemocnění. Aktuální stav aplikace postupů ML je k nalezení v belgické práci [7]. Zajímavá demonstrace možných výhod postupů ML při hodnocení rizika akutního infarktu myokardu je publikovaná v práci [8]. Byl hodnocen soubor více než 11 000 pacientů. Postup ML má z cíl kvantifikovat formou indexu (MI3) míru individualizovaného rizika infarktu s uvážením vlivu věku, pohlaví a algoritmu odběru (doba mezi dvěma odběry, použití 99 percentilu nebo delta hodnoty, relativního nebo absolutního vyjádření hodnoty delta). Důvodů pro postup ML je mnoho (diference mezi metodikami měření, mezi precizností hodnot u mezí stanovitelnosti a další). Využití několika modelů ML k predikci hematologických chorob, kvalitu měřených

dat, porovnání různých modelů ML a odhad dalšího rozvoje v hematologii prezentuje další práce [9].

Turecké autory jsme vybrali jako představitele jedné z pracovních skupin, která se zabývala aplikací postupů ML k diagnostice COVID-19 [10]. Prezentace hodnotí a vizualizuje efektivitu rutinní analytiky při diagnostice COVID-19. Vyšší efektivita hematologických parametrů a indikátorů zánětu oproti parametrům biochemickým je zřejmá.

Management chronických chorob, podmíněných životním stylem, výživou i genetickými aspekty, je obecným předmětem funkce precizní medicíny. Souhrnná práce pojednává o nezbytnosti a možnostech využití umělé inteligence, kombinované s ML na tomto poli. Souhrn zásadních témat pojednaných v práci se týká zejména problémů zánětlivých reakcí, vlivu způsobu života, genomu, hlubšího studia fenotypu, významu mikrobiomu [11].

Vyhodnocení časných detekcí septických stavů pokročilými formami ML (deep ML, neuronové sítě) ukázal významné prediktivní schopnosti pokročilých forem ML při posuzování terapie septických stavů [12]. Na efektivitu terapie septických stavů byla zaměřena také velká studie [13], jejímž závěrem bylo konstatování podstatně vyšší efektivity terapie, nastavené nástroji AI, než klasickým způsobem.

Velká metaanalytická studie porovnává výsledky více než 8000 studií ML, z nichž měla řada z nich významné metodické nedostatky. Po jejich eliminaci pro hodnocení zůstala však řada kvalitních studií, u nichž je jasně prokazatelný pozitivní vliv na zvýšení efektivity diagnostických procesů. Nebezpečím překotných, nepřesvědčivě validovaných postupů ML jsou matoucí rozpory mezi závěry kliniků bez použití ML a s jeho použitím. Digitalizace diagnóz je velmi potřebná, ale jen pokud není odtržena od lidského faktoru a lidské zkušenosti [14].

Další klíčová práce studovala kritické použití ML v řadě oblastí: u problému referenčních intervalů, vnitřní kontroly kvality s využitím vzorků pacientů (PBRTQC), laboratorního managementu, diagnostiky a predikce rizika, EHK, analýzy zdrojů chybovosti, v epidemiologii [15]. Bylo nalezeno hodně problémů ve volbě modelů ML, což je výzva k dalším studiím. Reita a spol. se zabývali hodnocením efektů čtyř modelů ML při třídění (triage) péče o cca 135 000 pacientů na emergency odděleních podle mezinárodní klasifikace ESI (emergency severity index) [16]. Všechny modely zlepšovaly úroveň lékařské péče a byly ve srovnání s klasickým postupem (bez aplikace ML) přínosné.

Použití aplikace ML k autoverifikačním procesům je zatím nečetné. Lze však uvést práci [17]. Porovnání klasické autoverifikace validity nálezů s klasifikací pomocí ML ukázalo podle autorů i zde výhodu použití umělé inteligence - snížení četnosti neplatných nálezů.

Čerstvě publikované doporučení EFLM k časné detekci maligních chorob již explicitně předpokládá použití nového typu tumorových markerů ctDNA, cfDNA, integraci jejich laboratorních výsledků s výsledky zobrazovacích metod a hodnocením diagnostiky pomocí ML a umělé inteligence [18]. Toto doporučení by moh-

lo být impulsem k pohybu na cestě precizní medicíny z výzkumných ústavů do klinických laboratoří.

Když byl předložen řadě laboratoří dotazník o jejich názoru na potřebu a znalosti a užitečnost aplikace umělé inteligence zjistilo se, že vědomí významné role AI v blízké budoucnosti je sdílené obecně, znalosti této problematiky jsou však malé [19]. Potřeba edukace je naléhavá.

Nizozemská studie NUMBER využila výhody standardizace metod a komutability kontrolních materiálů pro ustanovení referenčních intervalů dobře standardizovaných metod stanovení rutinních sérových analytů nepřímou metodou (bez použití referenčního souboru pacientů). Veliký počet dat (přes 7, 6 milionu) a vysoký počet vzorků a účastnických laboratoří (provedení přímo v cyklech EHK) umožnil navíc aplikovat shlukovou analýzu a zohlednit skupiny podle věku a pohlaví u 19 rutinních analytů séra [20].

Podrobnější postup ustanovení u 22 standardizovaných rutinních analytů séra s použitím dat cca 150 000 probandů publikovali další autoři o něco později [21]. Navíc byly vyhodnoceny i vlivy věku (na dvě skupiny s diskriminátorem 65 let věku a sezónními vlivy).

Použití big dat by mohlo pomoci vyřešit nebo aspoň zlepšit těžký problém určení hodnot referenčních intervalů resp. rozhodovacích limitů u nestandardizovaných analytů, kterých je naprostá většina. Pokouší se o to některé skupiny odborníků se střídavými výsledky. Úsilí o řešení tohoto problému bude patrně výsledkem kombinace harmonizace měření, zpracování big dat a pokusů o personalizaci hodnot referenčních intervalů. Existuje celá řada dílčích publikací na téma nepřímých určení referenčních intervalů (rozhodovacích limitů) pro analyty s nedostatečnou harmonizací měření (kagulační faktory, řada hormonů, protilátky a další. Zde jako příklad uvádíme dvě práce, vykazující vyšší, obecněji platné ambice řešení tohoto zásadního problému laboratorní medicíny [22, 23].

Velmi zajímavé mohou být pro čtenáře výsledky dotazníku o umělé inteligenci [24]. Jen 4 % účastníků považovalo své znalosti AI jako "expertní", jen 24 % považovalo AI za schopnou redukovat počet chyb a ještě méně (16 %) za nástroj efektivity procesu laboratorní práce. Nicméně 64 % účastníků bylo srozuměno s jejím dalším vývojem. Asi první sdělení v českém jazyce na téma ML v klinické laboratoři bylo publikováno před cca dvěma roky [25].

Závěr

Protože nejsme na světě sami, nevyhneme se potřebě znalostí a později aplikaci big dat, umělé inteligence a machine learningu. Tlak na jejich aplikaci bude zesilovat a bude znásoben v okamžicích, kdy se programů ML chopí výrobci. Je tedy vhodné být na situaci připravený a počítat se skutečností průniku (to je kombinace přirozené a umělé) inteligence do klinických laboratoří. Skutečný rozsah tohoto průniku se těžko odhaduje. Některé zde uvedené kapitoly se zřejmě zařadí do standardní pracovní náplně, jiné zůstanou

kontroverzní a musí projít prověrkou času, jiné se ukáží být redundantní.

Umělá inteligence je občas vnímaná jako jedno z budoucích nebezpečí lidského rodu. V současném stavu a v současné laboratoři jde však zatím jen o efektivní využití zbytečně ztracených, za horentních sum získaných, dat.

Literatura

1. **Xie, Y., Al-Aly, Z.** Risks and burdens of incident diabetes in long COVID-19: a cohort study. *Lancet Diabet. Endocrinol.*, 2022, 10(5), 311-321.
2. **Al-Aly, Z., Xie, Y.** High-dimensional characterization of post-acute sequelae of COVID-19. *Nature*, 2021, 594, 259-264.
3. **Benke, K., Benke, G.** Artificial Intelligence and Big Data in Public Health. *Int. J. Environ. Res. Public. Health.*, 2018, 15(12), 2796.
4. **Cabitz, F., Banfi, G.** Machine learning in laboratory medicine-waiting for the flood? *Clin. Chem. Lab. Med.*, 2017, 56(4), 516-524.
5. **Cerriotti, F.** Is there a classical role for the clinical laboratory digital health? *Clin. Chem. Lab. Med.*, 2019, 57(3), 353-358.
6. **Lippi, G., Plebani, M.** Integrated diagnostics: the future of laboratory medicine? *Biochem. Med. (Zagreb)*, 2020, 30(1), 18-30.
7. **De Bruyne, S., Speeckaert, M. M., Van Biesen, W. M., Delanghe, J. R.** Recent evolutions of machine learning applications in clinical laboratory medicine. *Crit. Rev. Clin. Lab. Sci.*, 2020, 58(2), 131-152.
8. **Than, M. P., Pickering, J., Sandoval, Y. et al.** Machine learning to Predict the Likelihood of Acute Myocardial Infarction. *Circulation* 2019, 140, 899-909
9. **Gunčar, G., Kukan, M., Notar, M. et al.** An application of machine learning to haematological diagnosis. *Sci. Rep.*, 2018, 8, 411.
10. **Cubukcu, H. C., Topcu, D. I., Bayraktar, N. et al.** Detection of COVID-19 by Machine Learning Using Routine Laboratory Tests. *Am. J. Clin. Pathol.*, 2022, 157(5), 758-766.
11. **Subramanian, M., Wojtusciszyn, A., Favre, L. a spol.** Precision medicine in the era of artificial intelligence: implications in chronic disease management. *J. Transl. Med.*, 2020, 18, 472.
12. **Bedoya, A. P., Futoma, J., Clement, M. E. et al.** Machine learning for early detection of sepsis: an Internal and Temporal Validation study. *JAMIA Open*, 2020, 3(2), 252-260.
13. **Komorowski, M., Celi, L. A., Badavi, O. et al.** Artificial intelligence learns optimal treatment strategies for sepsis in intensive care. *Nat. Med.*, 2018, 24(11), 1716-1720.
14. **Vasey, B., Ursprung, S., Bedcloe, B. et al.** Association of Clinician Diagnostic Performance with Machine Learning Based Decision Support Systems. A Systematic Review. *JAMA Network Open*, 2021, 4, e211276
15. **Ma, Ch., Wang, X., Wu, J. et al.** Real-world-big data I laboratory medicine: Current status, application, and future considerations. *Clin. Biochem.*, 2020, 84, 21-30.
16. **Reita, J., Goto, T., Faridi, M. K. et al.** Emergency Department triage prediction of clinical outcomes using machine learning models. *Crit. Care*, 2019, 23(1), 64.
17. **Wang, H., Wang, H., Zhang, J. et al.** Using machine learning to develop an autoverification system in clinical biochemistry laboratory. *Clin. Chem. Lab. Med.*, 2020, 59(5), 883-891.
18. **Froelich, M. F., Capoluongo, E., Kovacs, Z. et al.**

- The value proposition of integrative diagnostics (early) detection of cancer. On behalf of the interdisciplinary Task and Finish Group "CNAPS/CTC". *Clin. Chem. Lab. Med.*, 2022, 60(6), 821-829.
19. **Paranjape, K., Schinkel, M., Hammer, D. et al.** The Value of Artificial Intelligence. *Am. J Clin. Pathol. Lab. Med.*, 2021, 155(6), 823-831.
 20. **den Elzen, W. P. J., Brouwer, N., Thelen, M. H. et al.** NUMBER:standardized reference intervals in the Netherlands using a big data approach. *Clin. Chem. Lab. Med.*, 2018, 57(1), 42-56.
 21. **Ma, Ch., Xia, L., Chen, X. et al.** Establishment of variation source and age-related reference interval model for 22 common biochemical analytes in older people assigned-world big data mining. *Age Ageing*, 2022, 49(6), 1062-1070.
 22. **Fleming, J. K., Katayev, A., Moorer, Ch. et al.** Development of nation-wide reference intervals using an indirect method and harmonized assays. *Clin. Biochem.*, 2022, 99, 20-59.
 23. **Yang, D., Su, Z., Zhao, M.** Big data and reference intervals. *Clin. Chim. Acta*, 2022, 527, 23-32.
 24. **Ardou, O., Schmidt, R. L.** Clinical laboratory employees attitudes toward artificial intelligence. *Lab. Med.*, 2020, 51(6), 649-654.
 25. **Friedecký, B.** Automatická strojová diagnostika-machine learning a precizní medicína. Pojmy, principy, perspektivy. *Klin. Biochem. Metab.*, 2020, 4, 161-165.

Autor prohlašuje, že není ve střetu zájmů.

Do redakce došlo 15. 6. 2022

Adresa pro korespondenci
RNDr. Bedřich Friedecký, Ph. D.
Střelníčná 1680
182 00 Praha 9
e-mail: friedecky@sekk.cz