

# Automatická strojová diagnostika - machine learning a precizní medicína. Pojmy, principy, perspektivy.

Friedecký B.

Ústav klinické biochemie a diagnostiky, FN Hradec Králové  
SEKK, spol. s r. o. Pardubice

## SOUHRN

Digitalizace laboratoří, aplikace big dat a automatizovaná strojová diagnostika („machine learning“) jsou nástroji pro vznik a fungování toho, co se označuje jako precizní medicína. Genomika, její dominantní metody (qPCR, dPCR, ddPCR, NGS), produkující obrovská kvanta dat (big data) a schopnosti počítačových systémů tyto soubory dat využívat v diagnostice a terapii za významného přispění „umělé inteligence“ se označují jako strojová automatizovaná diagnostika - machine learning respektive deep learning). Tyto postupy pronikají z průmyslu a výzkumu do rutinní medicíny včetně medicíny laboratorní. Zvládnutí technických a personálních problémů těchto změn bude stát značné úsilí, srovnatelné s před lety realizovanou přeměnou manuální laboratorní práce na automatizovanou činnost a s přeměnou papírové dokumentace výsledků na laboratorní a nemocniční informační systémy. Lze předpokládat nejen zásadní změny metod laboratorní práce, ale i změny požadavků na odbornost personálu laboratoří a rovněž lze předpokládat nevyhnutelnost radikálního ovlivnění činnosti klinických laboratoří. Etický rozměr nastávajících změn bude stejně závažný, jako ten technický a bude možné očekávat nejen významný progres v diagnostice a prognostice chorob, ale i vzestup rizika zdravotní péče v případech chyb a neprofesionality. Automatická strojová aplikace big dat a používání umělé inteligence jsou náročné, je s nimi v medicíně málo zkušeností, ale vyhnout se jim nebude možné.

*Klíčová slova:* big data, „machine learning“, umělá inteligence, precizní medicína.

## SUMMARY

**Friedecký B.: Automatic machine diagnostics - machine learning and precision medicine. Concepts, principles, perspectives.**

Digitalization of clinical laboratories, application of big data and methods of machine learning re contemporary tools for precision medicine. Precision medicine is based mainly on the genomic methods, namely of dominant PCR and NGS methods. These methods produces enormous number of dates (big data) and can be explored by means of artificial intelligence in processes called „machine learning“. Machine learning was primarily used in industry and research and now contemporary penetrates into medicine and also to laboratory medicine. Methods based on the big data and artificial intelligence with exploration of big data is certainly very important factor of future of medicine. It will be needs large requirements not only on high-technology equipment, but also for new type of young laboratory Professional used basically new methods of work and mind. Machine learning, part of precision medicine, necessary namely for oncology and prediction of patients state crettemeans also lot of new types of ethical problems. These ethical questions and problems should be soluted immediately, parallel with introduction of machine learning to laboratory practice.

*Keywords:* Big data, machine learning, artificial intelligence, precision medicine.

Digitalizace laboratorní medicíny je intenzivně probíhající proces, který v blízké budoucnosti silně změní její podobu, způsoby její práce a nároky na laboratorní pracovníky. V tomto sdělení jsou uvedeny základní informace o pojmech big (meta) dat a jejich počítačovém zpracování v diagnostice a hodnocení terapie („machine learning“), o podkladech precizní medicíny, spojené se zásadně novými přístupy zejména v onkologii, vyznačující se výrazným personalizovaným charakterem a důrazem na predikci chorob a jejich stavu.

## Big data

Pojem big (velkých, meta) dat v laboratoři nemá exaktní definici, lze je však chápat jako masivní počet, elektronicky zpracovatelných dat, použitelných s výhodou pro účely péče o pacienty [1].

Je to analog obecných databází typu Amazon, Google a ohromného množství dat, skladovaných s ne vždy přesvědčivým využitím v smartphonech, inteligentních hodinkách, tabletech a jinde. Mediálně dobře a vydatně zpracované náměty z oblastí špičkové vědy velmi názorně poodhalují možnosti současného získávání neuvěřitelného množství dat a schopnosti superpočítačových zařízení je zpracovávat (Tabulka 1).

Big data ve zdravotnictví by měla být chápána jako data s velkým potenciálem využití pro potřeby pacientů po jejich elektronickém zpracování. Jde o jeden z nástrojů vyvíjející se personalizované a prediktivní laboratorní medicíny, o základ digitalizace laboratorní činnosti, o tvorbu elektronické podoby dat pacientů EHR (Electronic Health Results).

Pregnantním příkladem metod, generujících a využívajících big dat jsou v klinických laboratořích metody NGS (Next-Generation Sequencing). Dnes se už vcel-

ku běžně k diagnostice využívají analýzy mutací sady cca 10 genů s 30-40 variantami například při stanovení kapalných biopsií u maligních chorob, k stanovení mikroRNA v biologických tekutinách a dalších. V genetickém výzkumu se rozvíjí NGS postupy k sekvenaci celých genomů/exomů, jmenovitě k studiu vzácnějších genetických chorob a poruch.

Dalším významným zdrojem big dat pro diagnostiku je radiomika, označení digitalizovaných zobrazovacích metod, kterou lze právem považovat za průkopníka digitalizace zdravotnictví, dnes již značně rozvinutá.

Základními požadovanými vlastnostmi big dat jsou rychlost (velocity), objem (volume), různorodost (variety) a věrohodnost (veracity).

K zpracování big dat jsou k dispozici a v řadě publikací použité softwarové nástroje typů Apache- Hadoop a dalších, ale využívá se i klasických nástrojů typu Excel. Obecně jde o to, aby byla relevantní data využívána, namísto aby byla bez užitku skladována, jak se děje v klinických laboratořích doposud. Big data zahrnují výsledky pacientů spolu s daty demografickými, klinickými, epidemiologickými.

Velkým problémem a současně úkolem je redukce závislosti kvality big dat na kvalitě dat dílčích. Zde hrají zásadní roli procesy standardizace a harmonizace analytických, preanalytických a postanalytických procesů. A samozřejmě schopnosti a dovednosti používání informačních technologií a umělé inteligence. Právě omezená dovednost v oblasti informačních technologií bude možná hlavním zdrojem problémů pokroků v digitalizaci rutinních klinických laboratoří. Nárůst publikací, věnujících se big datům má za poslední léta exponenciální charakter a nebylo zastaveno ani pandemií COVID-19 (tabulka 1).

Pojem big dat se v laboratorních metodách používá nespécifikovaně. Uvedeme pár aktuálních příkladů z oblasti mimo genomiku, kdy bylo využito pojmu big dat mimo oblast genomiky [2 - 5]:

- K tvorbě referenčních intervalů. Zde lze demonstrovat jejich základní vlastnosti: „volume“ (přes 7, 5 mil. dat), „variety“ (18 analytů, 16 laboratoří, 4 systémy měření) a „veracity“ dat (byly použity jen standardizované metody)
- Další oblastí přesahující do oblasti big dat je diabetologie, což je možné demonstrovat rozvojem CGM, produkujícím kvanta dat, získaných od jednoho subjektu. Systémy alarmů a možnost telemedicínského přístupu, možnost propojení s insulinovými pumpami, harmonizovaný algoritmus sledování hodnot a časů prohlubuje personalizovaný charakter. U CGM je vidět instruktivní význam grafického znázornění big dat.
- I oblast kontroly kvality s rozvíjející se metodou na bázi pohyblivých průměrů patientských vzorků v reálném čase (PBRTQC), doporučených experty IFCC jako realizace hodnocení a snižování rizika chyb měření, patří do této oblasti. Také zde je kvalita automaticky zpracovávaných dat supervizována pracovníky správnou volbou trunkačních limitů k výběru dat a respektováním tvaru jejich statistického rozdělení.

## Precizní medicína, strojová diagnostika (machine learning, deep learning)

„Machine learning“ je použití big dat, vyhodnocovacích prostředky umělé inteligence, k diagnostice v precizní medicíně [6]. A naopak, precizní medicína prosazující se zatím zejména v onkologii je produktem aplikace big dat a „machine learning“ v klinické laboratoři. Není mi znám obecně přijatý český slovní ekvivalent výrazu „machine learning. Počítačová diagnostika, strojová diagnostika nebo doslovný překlad strojové učení? **Pozornost zasluhuje, jak se v názvu zobrazuje poznávací charakter laboratorní diagnostiky (learning), výrazně překračující běžné pojetí rutinní analýzy jako prosté analýzy požadované dávky biologických vzorků (a jejího vyúčtování).** Četnost prací na toto téma v databázi PubMed® se změnil od nuly v roce 2002 na neuvěřitelných 10 800 v roce 2019 (Tabulka 2).

Nesupervidovaná (unsupervised) „machine learning“ pracuje s předem netříděnými vstupními daty a vyhodnocuje závislosti mezi nimi navzájem a klinickým stavem (například pomocí shlukové analýzy dat-cluster analysis).

Pokročilejší supervidovaná „machine learning“ studuje vztahy mezi různými a různě formulovanými vstupními daty (hormonů, genů a jiných) s daty výstupními po statistických operacích. Obvykle se zaměřuje na predikci klinického stavu jedinců při onkologických, ale i jiných chorobách [7]. Část postupů „machine learning“ patří spíše již do další kategorie „deep learning“ (hluboký přístup strojové diagnostiky) s aplikací umělé inteligence na bázi arteficiálních neuronových sítí. Ty pracují s modely arteficiálních biologických systémů, konstruovaných výzkumnými pracovníky na bázi sítí biomarkerů a genů [8].

Takové přístupy, umožněné použitím umělé inteligence jsou již dobře známé a využívané v řadě pozemských i vesmírných technologií, v dopravě i teoretických fyzikálních oborech, ale v klinických laboratořích jsou zatím nové. Počet prací o „deep learning“ ve zdravotní péči však není zdaleka zanedbatelný (Tabulka 2), i když se zatím soustřeďují na výzkum (farmakogenetika, výběry primerů pro genomiku, struktura a funkce buněk), ale v budoucnosti by mohla být jejich podstatným cílem medicínská predikce a stratifikace stavu chorob. S možností použití metod „deep learning“ k predikci a zhodnocení kritického stavu některých chorob se lze pro představu instruktivně seznámit v recentních pracích [9, 10].

Nedílnou součástí aplikace big dat a „machine learning“ je grafická vizualizace procesů a výsledků. Samotná big data, digitalizace, „machine learning“ by bez pochopení podstaty a bez nových algoritmů použití moc neznamenala. Jde o výzvu mladým laboratorním profesionálům k závažným změnám přístupu k práci.

Pro pracovníky laboratoří je k pochopení „machine learning“ klíčová diskuse panelu expertů, uveřejněná v časopise Clinical Chemistry [11]. Shrnutí aktuálních možností rychlé aplikace v rutinní klinické laboratoři, laboratorním pracovníkům dosud enigmatické oblasti, je zde prezentováno následovně:

**Table 1.** Demonstration of known, in media well popularized possibilities for working with big data in current high science

Equipment	Target
<b>ALMA-ESO</b> Atacama Large Millimeter/submillimeter Array	<b>Space studies</b>
<b>LHC-CERN</b> Large Hadron Collider	<b>Mass particle/subparticle study</b>
<b>LIGO</b> Laser Interferometer Gravitational Wave Observatory	<b>Detection of gravitational wave</b>

**Table 2.** Frequency of publications on big data in medicine up to July 20, 2020

Year	Number
1999	17
2020 July	2 949

Je na čase rozšířit standardní analytickou kontrolu kvality kontrolou pomocí (big) dat pacientů v reálném čase (patient-based-real time quality control-PBRTQC). Jde o starou metodu kontroly kvality, které strojové zpracování big dat dává nový rozměr, zejména jako nástroje hodnocení a snižování rizika chyb měření. S tím souvisí otázky volby truncačních limitů a jejich rozdílnosti podle typu statistického rozdělení (odlišné u substrátů a elektrolytů od enzymů, kreatininu) výsledků. Příslušné typy software jsou dostupné a testované (například <https://huvaros.com>). Postup je doporučovaný IFCC jako metoda, snižující míru rizika chyb laboratorních testů u pacientů. Paralelní s klasickou vnitřní kontrolou kvality a s důslednou komutabilitou kontrolního materiálu.

- Bez zásadních problémů lze aplikovat pomocí „machine learning“ detekci preanalyticky chybných vzorků namísto používání delta checků
- Lze dosáhnout zlepšení validačních postupů pomocí experimentů se soubory statistických až milionů dat o referenčních intervalech, rozhodovacích limitech, kritériích kvality, pocházejících z většího počtu zdrojů [2]
- Ukazují se možnosti realizace precizní medicíny i mimo oblast onkologie u klíčových chorob (diabetes, kardiovaskulární a další)
- Realizace ekonomické efektivity propojením zájmů a činnosti poskytovatelů péče, pacientů a plátců je více, než na místě a není bez digitalizace, big dat a „machine learning“ realizovatelná

V současnosti jsou největším a typickým zdrojem big dat metody NGS (next generation of sequencing). Potřeba nové generace laboratorních pracovníků, ovládajících metody sekvenace genomu s přístupem k potřebným databázím a biobankám je pro budoucnost laboratoří klíčová.

Stejně klíčová je nezbytnost pokračovat v harmonizaci analytiky, preanalytiky a interpretací výsledků k zajištění kvality big dat a aplikace „machine learning“ a deep learning“ postupů.

Směrem dalšího vývoje by měla být integrace jednotlivých diagnostických kroků do společného diagnostického systému, sestávajícího z laboratorních, zobrazovacích, demografických metod a podmíněného spoluprací jednotlivých specialistů [13]. Digitalizace je nutnou podmínkou této integrace.

Expertní komentáře a systémy diagnostických neuronových sítí, dávné cíle laboratorní diagnostiky, se v éře digitalizace, big dat a strojové diagnostiky už nejeví nedosažitelné. Odstranění bariér mezi diagnostickými obory, multidisciplinární přístupy, registry pacientů a národní i mezinárodní programy péče při zachování ochrany dat a integrity pacientů jsou dalšími cíli.

V tabulce 2 jsou uvedeny počty prací „machine learning“ podle zúžených témat a patrně je preferováno predikce chorob. Tabulka 3 ukazuje, že práce o „machine learning“ nejsou už jen doménou genomiky a onkologie, ale jsou používány i u jiných standardních chorob.

## Rizika aplikace umělé inteligence

Jsou závažná. Zásady jejich eliminaci rizik znějí obecně, ale měly by být neustále brány v potaz:

- Prospěch pacienta
- Nepřípustnost jeho poškození
- Ochrana osobnosti před zneužitím dat
- Právní záruky

Podrobněji jsou problémy situací, kdy aplikace genomiky při diagnostice nebezpečně hraničí s možností genové manipulace uvedeny v diskusi genetiků v Clinical Chemistry [14]. V ní se dost striktně zpochybňují genomické testy u zdravých jedinců a v dětském věku. Také experti AACCC (American Association of Clinical Chemistry) kladou explicitní důraz na důležitost etického přístupu laboratorních pracovníků k zábraně možných negativních vlivů aplikace umělé inteligence [15, 16], zejména v možnostech dehumanizace péče o pacienty. Ostatně v oblasti rizik digitalizace máme zejména v multimediální oblasti víc, než dostatek nedobrych zkušeností.

## Proč není možné ignorovat „machine learning“ v laboratořích klinické biochemie [11]?

- Potřeba respektovat nároky na precizní medicínu v onkologii, diabetologii, kardiovaskulárních chorobách je otázkou udržení odborného kreditu laboratoří

**Table 3.** Frequency of the machine learning/deep learning publications. PubMed® up to July 15, 2020

Advanced		Number of publications
<b>Machine learning total</b>		44 038
	prediction	19 445
	cancer	6 694
	CT	1 525
<b>Deep learning total</b>		14 897
	prediction	4 321
	cancer	2 250
	CT	989

**Table 4.** Number of machine learning publication for some disease (PubMed® July 15, 2020)

Disease	Number of publications
Diabetes type 1	183
Cardiac	2210
Kidney	484
COVID-19	114

- Potřeba důrazu na predikci a prevenci znamená rovněž podmínku odborného kreditu a zábranu přesunu klinických laboratoří mezi nezbytný, ale v propojení s pacienty oslabený klinický servis.
- Ignorování digitalizace v rutinních laboratořích by nutně znamenalo předat iniciativu podnikavým korporacím, bez zábran podnikajícím bez konzultací s laboratorními experty.
- Otázka dostupnosti přístrojových platforem a údajné náročnosti jejich obsluhy pro genomiku již nehraje zásadní roli ani u řady běžných klinických laboratoří, ne tak u laboratoří velkých zdravotních zařízení.

## Závěr

Bylo by naivní očekávat, že digitalizace klinických laboratoří, založená na exploataci big dat, automatizované (machine learning a deep learning) diagnostice, použití umělé inteligence, laboratorních aspektů telemedicíny se ještě netýká laboratorních profesionálů a rutinních klinických laboratoří. Vývoj probíhá významným tempem a je označován za epochální zlom ve vývoji laboratorní medicíny (laboratory disruption) ve všech směrech. Otázku, zdali jsme připraveni na tento směr vývoje, není možné odpovědět dostatečně pozitivně. Převrat v aplikaci, pojetí a pracovních procesech laboratorní medicíny lze považovat za hotovou věc a je třeba tuto skutečnost vzít na vědomí i u nás.

## Literatura

1. **Tolan, N., Parnas, M. L., Bandhuin, L. M. et al.** Big data in laboratory medicine, *Clin. Chem.*, 2015, 61, p. 1433-1440, DOI:10.1347/clinbiochem.2015.248591
2. **den Elzen, W. P. J., Brouwer, N., Thelen, M. H. et al.** NUMBER: Standardized reference interval in the Netherlands using a big data approach. *Clin. Chem. Lab. Med.*, 2018, 57, p. 42-56.
3. **Battelino, T., Danne, T., Bergenstal, R. M. et al.** Clinical Targets for Continuous Glucose Monitoring Data Interpretation, *Diabet. Care*, 2019, <https://doi.org/10.2337/dci19-0028>.
4. **Badrick, T., Bietenbeck, A., Cervinski, M. A. et al.** Patient-Based Real Time Quality Control: Review and Recommendations, *Clin. Chem.*, 2019, 65, p. 962-971.
5. **Smith, J. D., Badrick, T., Bowling, F.,** A direct comparison of patient. based real-time quality control techniques. The importance, of the analyte distribution. *Ann. Clin. Biochem.*, 2020, 57/3, 206-214.
6. **Telenti, A.** Machine Learning to Decode Genomics. *Clin. Chem.*, 2019, 66/1, 45-47. 308296.
7. **Liu, X., Chang, X., Liu, R. et al.** Quantifying clinical states of complex diseases using single-sample dynamic network biomarkers. *Plos Comput. Biol.*, 13 e1005633
8. **Weinberg, M., Merico, D., DeLong, A., Frey, B. J.** Deep learning in Biomedicine. *Nat. Biotechnol.*, 2018, 36, p. 829-838.
9. **Gal, J., Bailleux, C., Chardin, D. et al.** Comparison of unsupervised machine learning methods to identify metabolomic signatures in patients with localised breath cancer. *Comp. Struct. Biotechnol.*, 2020, 18, p. 1509-1521.

10. **Li, Q., Sun, Z. Y., Qi, L. et al.** A deep learning approach to characterize 2019 Coronavirus Disease (COVID-19) pneumonia in Chest CT images. *Eur. Radiol.*, 2020, p. 1-11.
11. **Badrick, T., Banfi, G., Bietenbeck, A. et al.** Machine learning for Clinical Chemists. *Clin. Chem.*, 2019, 65/11, 1350-1356.
12. **Mitani, T., Dai, S., Yokota, S. et al.** Highly accurate and explainable detection of specimen mix-up using a machine learning model. *Clin. Chem. Lab. Med.*, 2020, 58/3, 375-383
13. **Lippi, G., Plebani, M.** Integrated diagnostics-the future of laboratory medicine? *Biochem. Med.*, (Zagreb) 2020, 30/1, 010501
14. **Baudhuin, L. M., Biesecker, L. B., Burke, W. et al.** Predictional Precision Medicine with Genomic Data, 2020, 66/1, 33-41
15. AACC Ethics of AI and Big Data in Laboratory Medicine, *Clin. Lab. News* 2020 Data. *Clin. Chem.*, 2020, 66, p. 33-41
16. Laboratories must apply medical ethics to mitigate AI risks. *AACC Smart Brief* 2020.

Autor prohlašuje, že není ve střetu zájmů.

Do redakce došlo 21. 7. 2020

*Adresa pro korespondenci*  
*RNDr. Bedřich Friedecký, Ph.D.*  
*Střelnická 1680*  
*182 00 Praha 8*  
*e-mail: friedeky@sekk.cz*