

# DIPLOMOVÁ PRÁCE

Umělá inteligence ve zdravotnictví

## STUDIJNÍ PROGRAM

Projektové řízení inovací

## STUDIJNÍ OBOR

## VEDOUCÍ PRÁCE

doc. RNDr. Bohumír Štědroň, CSc.

Dalibor

Čápek

2022

## I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Čápek** Jméno: **Dalibor** Osobní číslo: **499708**  
Fakulta/ústav: **Masarykův ústav vyšších studií**  
Zadávající katedra/ústav: **Institut ekonomických studií**  
Studijní program: **Projektové řízení inovací**

## II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

**Umělá inteligence ve zdravotnictví**

Název diplomové práce anglicky:

**Artificial Intelligence in Healthcare**

Pokyny pro vypracování:

Posluchač provede rozbor trendů v AI a aplikaci ve zdravotnictví.  
Výsledkem bude i prognóza do roku 2030.  
Výsledky mohou být publikovány v odborných časopisech.

Literatura:

Štědroň B. a kol.: Prognostické metody a jejich aplikace, C.H.BECK, Praha 2012  
Štědroň B. a kol.: Prognostika, C.H.BECK, Praha 2019  
Štědroň B. a kol.: Právo a umělá inteligence, A.Čeněk, Praha 2020

Seznam doporučené literatury:

Štědroň B. a kol.: Prognostické metody a jejich aplikace, C.H.BECK, Praha 2012  
Štědroň B. a kol.: Prognostika, C.H.BECK, Praha 2019  
Štědroň B. a kol.: Právo a umělá inteligence, A.Čeněk, Praha 2020

Jméno a pracoviště vedoucí(ho) diplomové práce:

**doc. RNDr. Bohumír Štědroň, CSc. Masarykův ústav vyšších studií ČVUT v Praze**

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **05.01.2022**

Termín odevzdání diplomové práce: **28.04.2022**

Platnost zadání diplomové práce: \_\_\_\_\_

doc. RNDr. Bohumír Štědroň, CSc.  
podpis vedoucí(ho) práce

Mgr. František Hřebík, Ph.D.  
podpis vedoucí(ho) ústavu/katedry

prof. PhDr. Vladimíra Dvořáková, CSc.  
podpis děkana(ky)

## III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací.  
Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

\_\_\_\_\_  
Datum převzetí zadání

\_\_\_\_\_  
Podpis studenta

ČÁPEK, Dalibor, 2022. *Umělá inteligence ve zdravotnictví*; Praha: ČVUT. Diplomová práce. České vysoké učení technické v Praze, Masarykův ústav vyšších studií.



**MASARYKŮV ÚSTAV  
VYŠŠÍCH STUDIÍ  
ČVUT V PRAZE**

## **Prohlášení**

Prohlašuji, že jsem svou diplomovou práci vypracoval samostatně. Dále prohlašuji, že jsem všechny použité zdroje správně a úplně citoval a uvádím je v příloženém seznamu použité literatury.

Nemám závažný důvod proti zpřístupnění této závěrečné práce v souladu se zákonem č. 121/200 Sb., o právu autorském, o právech souvisejících s právem autorským a o změně některých zákonů (autorský zákon) v platném znění.

V Praze dne: 25.04.2022

Podpis:

## **Poděkování**

Děkuji vedoucímu práce doc. RNDr. Bohumíru Štědroňovi, CSc. za odborné vedení, cenné rady, důvěru a trpělivost při vypracování mé diplomové práce. Rovněž děkuji kolektivu pracovníků Českého vysokého učení technického v Praze za vstřícnost a praktické rady během celého studia. V neposlední řadě pak patří můj dík mé rodině a přátelům za velkou podporu.

## **Abstrakt**

Diplomová práce se věnuje využití umělé inteligence ve zdravotnictví. Pojem umělá inteligence se objevuje stále více v souvislosti s dopady na různá odvětví. Cílem práce je zaměřit se na vývoj umělé inteligence ve zdravotnictví a její dopad na pracovní svět ve zdravotnictví. Teoretická část je věnována hlavním pojmům jako jsou data, strojové učení a v neposlední řadě etika umělé inteligence a strojového učení. Již v teoretické části je věnována pozornost praktickému využití umělé inteligence, respektive strojového učení v praxi. Druhá část práce se věnuje vybraným případovým studiím aplikace umělé inteligence v lékařské vědě a vývoji léků. Závěreční šestá kapitola je zaměřena na predikci dopadů umělé inteligence na vybrané profese ve zdravotnictví v roce 2030.

## **Klíčová slova**

Umělá inteligence, data, velká data, strojové učení, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, zdravotnictví.

# **Abstract**

This master thesis describes the use of artificial intelligence in healthcare. The concept of artificial intelligence is increasingly emerging in connection with the impact on various sectors. The aim of this work is to focus on the development of artificial intelligence in health care and its impact on the world of work in health care. The theoretical part is devoted to the main concepts such as data, machine learning and, finally, the ethics of artificial intelligence and machine learning. Already in the theoretical part, one section is devoted to the practical use of artificial intelligence, or machine learning in practice. The second part of the thesis describe and analyze selected case studies of the application of artificial intelligence in medical science and drug development. The final sixth chapter focuses on the prediction of the effects of artificial intelligence on selected health professions in 2030.

## **Key words**

Artificial Intelligence, data, big data, machine learning, supervised learning, unsupervised learning, semi-supervised learning, reinforcement learning, healthcare.

# Obsah

<b>Úvod</b>	<b>11</b>
<b>1. Data</b>	<b>14</b>
1.1. Co jsou data	14
1.1.1. Typy Dat	15
1.2. Big data	17
1.2.1. Objem dat	18
1.2.2. Rozmanitost dat	19
1.2.3. Rychlost dat	19
1.2.4. Hodnota dat	21
1.2.5. Věrohodnost dat	22
1.2.6. Platnost dat	23
1.2.7. Variabilita dat	23
1.2.8. Vizualizace dat	24
1.3. Využití dat ve zdravotnictví a farmacii	24
1.3.1. Předpověď doby čekání	24
1.3.2. Prediktivní analytika	24
1.3.3. Elektronické zdravotní záznamy	25
1.3.4. IoT (internet věcí ve zdravotnictví): notifikace v reálném čase, alerty, automatizace	26
1.4. Přeměna dat na informace	27
1.4.1. Deskriptivní analytika	28
1.4.2. Diagnostická analytika	29
1.4.3. Prediktivní analytika	29
1.4.4. Preskriptivní analytika	30
<b>2. Strojové učení</b>	<b>31</b>
2.1. Rozdíl strojového učení od tradičního softwarového inženýrství	31
2.2. Základy strojového učení	33
2.2.1. Supervised learning	33
2.2.2. Unsupervised learning	35
2.2.3. Semi-supervised learning	37



2.2.4.	Reinforcement learning .....	37
2.3.	Neuronové sítě.....	38
2.3.1.	Perceptrony.....	39
2.3.2.	Umělá neuronová síť.....	40
2.4.	Hluboké učení .....	41
2.5.	Dolování dat.....	41
2.6.	Parametrické a neparametrické algoritmy .....	43
2.7.	Jak fungují algoritmy strojového učení .....	44
<b>3.</b>	<b>Strojového učení v praxi .....</b>	<b>46</b>
3.1.	Rámování: Specifikace problému .....	47
3.1.1.	Sbírání příkladů.....	48
3.1.2.	Základní informace.....	48
3.1.3.	Chyby v datech .....	48
3.2.	Trénování modelu strojového učení .....	50
3.2.1.	Programovací jazyky.....	50
3.2.2.	Trénování a testování dat .....	52
3.3.	Vyhodnocení a optimalizace metod a výsledků .....	53
3.3.1.	Vyhodnocení přesnosti algoritmu .....	53
3.3.2.	Měření výkonnosti.....	54
3.3.3.	Optimalizace .....	54
3.3.4.	Zlepšování výsledků s lepšími daty.....	55
3.4.	Šíření výsledků .....	55
3.5.	Deployment.....	56
<b>4.</b>	<b>Etika umělé inteligence a strojového učení.....</b>	<b>58</b>
4.1.	Etika .....	58
4.2.	Etika datové vědy.....	58
4.2.1.	Etika dat .....	58
4.3.	Správci dat a zpracovatelé.....	59
4.4.	Politika, právo a regulace.....	60
4.5.	Správa dat a informací .....	61
4.6.	Etický kodex.....	61
4.7.	Etika umělé inteligence ve zdravotnictví .....	62
<b>5.</b>	<b>Aplikace umělé inteligence v lékařské vědě a vývoji léků.....</b>	<b>65</b>

5.1.	Umělá inteligence pro kvantitativní vztahy mezi strukturou a aktivitou při objevování léků .....	65
5.1.1.	Sítě hlubokého učení .....	66
5.1.2.	Strojové učení založené na podobnosti sítě.....	67
5.1.3.	Kernelova metoda a podpora vektorového stroje (SVM).....	68
5.1.4.	Metoda rozhodovacího stromu.....	69
5.2.	Umělá inteligence v predikci rakoviny pomocí microarray dat.....	70
5.2.1.	Detekce rakoviny z dat genové exprese .....	70
5.2.2.	Výběr funkcí .....	70
5.2.3.	Předpověď rakoviny .....	71
5.3.	Hluboké učení pro analýzu lékařských snímků .....	71
5.3.1.	Hluboké učení pro zpracování lékařských snímků .....	71
5.3.2.	Hluboké učení v mamografii.....	73
5.3.3.	Hluboké učení pro analýzu kardiologických snímků.....	74
<b>6.</b>	<b>Prognóza vývoje umělé inteligence ve zdravotnictví.....</b>	<b>76</b>
6.1.	Jak umělá inteligence změní pracovní svět ve zdravotnictví.....	76
6.2.	Změny v pracovních místech: Zdravotnictví v roce 2030 .....	77
6.2.1.	Dopad na počty zaměstnanců.....	79
6.3.	Jak změní umělá inteligence a automatizace činnost zdravotnických pracovníků?.....	82
6.3.1.	Méně administrativy, více péče o pacienty.....	82
6.4.	Nové činnosti a dovednosti ve zdravotnictví .....	85
6.4.1.	Nový způsob interakce s pacienty.....	85
6.4.2.	Posílení digitálních dovedností u širší pracovní síly ve zdravotnictví ...	86
6.5.	Představení nových odborníků ve zdravotnictví.....	89
<b>Závěr</b>	<b>.....</b>	<b>92</b>
	Seznam použité literatury.....	94
	Seznam obrázků .....	97
	Seznam tabulek .....	98
	Seznam grafů.....	98

# Úvod

Umělá inteligence je považována za jednu z nejvíce vzrušujících technologií dnešní doby.

Mnoho lidí bývá překvapeno, když zjistí, že umělá inteligence není nová technologie. Umělá inteligence je tady s námi již několik desítek let, ale teprve dnes se k ní upírá velká pozornost a dochází k obrovským pokrokům ve využití umělé inteligence napříč všemi oblastmi průmyslu.

V padesátých letech 20. století byl průkopníky v umělé inteligenci definován cíl: vytvořit stroj s inteligencí člověka. Tato zajímavá kombinace jasně definovaného cíle, avšak velmi složitého úkolu, přilákala do výzkumu umělé inteligence jedny z nejlepších mozků oboru informatiky jako jsou Marvin Minský, John McCarthy a Herbert Simon.

Opravdovou renesanci umělé inteligence, kterou zažíváme dnes, vyvolal vývoj dvou klíčových faktorů, díky kterým umělá inteligence dobře funguje a jeden významný technologický průlom, a to neuronové sítě. Neuronové sítě vyžadují obrovské množství výpočetního výkonu a dat, které dříve nebyly k dispozici. Zásadní průlom nastal, kdy Geoffrey Hinton objevil způsob, jak neuronové sítě účinně trénovat. Výkonnost v řešení úkolů, jako rozpoznávání řeči nebo objektů, se mnohonásobně zvýšila. Dnes tento průlom nazýváme hlubokým učením.

Tento průlom zajistil technologický boom v oblasti umělé inteligence a začalo se využívat schopností umělé inteligence k řešení celého spektra problémů reálného světa. Výzkumní pracovníci a ředitelové technologických firem se začali věnovat potenciálu umělé inteligence v dešifrování řeči, překládání textů, rozpoznávání obrazů předpovídání spotřebního chování, detekci podvodů, rozhodování o úvěrech, a dokonce i v řízení automobilů. Všechny potenciály vyžadují vydatné objemy relevantních dat a bezchybně nastavený algoritmus se správným cílem.

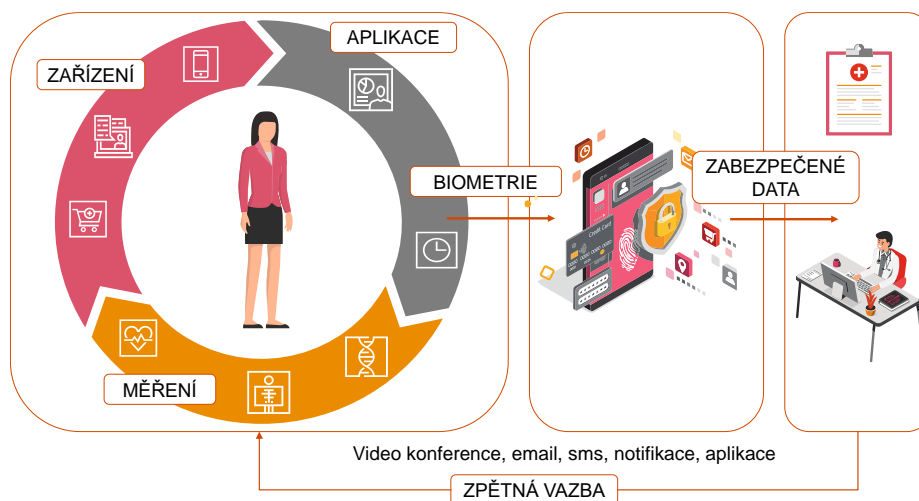
Hluboké učení je pouze tzv. úzká umělá inteligence, která využívá data z jedné specifické oblasti k dosažení jednoho konkrétního cíle. Hluboké učení přináší působivé výsledky, avšak je stále hodně vzdálena od tzv. obecné umělé inteligence, která bude umět vše, co dokážou lidé.

Hluboké učení se využívá především v oblastech jako je pojištění nebo poskytování úvěrů, kde jsou relevantní data k dispozici a cíl optimalizace je zřejmý. Mnoho

společností se zabývá již pokročilejší umělou inteligencí jako je řízení autonomních vozů. Výzkumníci pomáhají vozidlu vidět a vnímat svět kolem sebe pomocí rozpoznávání vzorů v pixelech snímací kamery, s tím korelují a používají se tyto informace k rozhodování. Nicméně čím dál častěji se začíná umělá inteligence využívat také ve zdravotnictví.

Je velmi nepravděpodobné, že by umělá inteligence plně nahradila doktory a zdravotní sestry, ale strojové učení a umělá inteligence transformují zdravotnictví, a především pomáhají zpřesnit a zlepšit výsledky. Strojové učení zlepšuje diagnostiku, předpovídá výsledku a začíná se prosazovat v oblasti personalizované zdravotní péče.

Obrázek 1: Řízený daty, vztah pacienta s lékařem



Vlastní zpracování, zdroj: *Kniha Machine Learning and AI for Healthcare, Arjun Panesar*

Představte si situaci, že jdete k lékaři (nebo se spojíte přes aplikaci) s bolestí v oblasti srdce. Po vyslechnutí symptomů, možná pomocí sdíleného videa nebo přes hodinky, či jiné zařízení, sledující zdravotní stav pacienta nebo lékař vloží všechna data do počítače. Počítač projde veškerou evidenci o pacientovi. Zjistí, že pacient byl v poslední době vyšetřen na magnetické rezonanci a počítačový asistent na základě nových skutečností detekuje malou anomálii na snímku, kterou lékař nemohl vidět. Systém bude schopný vygenerovat doporučení na léčbu a lékař během krátké doby může adekvátně rozhodnout o další léčbě pacienta.

Tohle je zatím hudba budoucnosti a tato práce zmapuje aktuální využití umělé inteligence ve zdravotnictví pomocí využití metody případových studií. Na základě historických dat a dostupných analýz bude v diplomové práci také predikován dopad vývoje umělé inteligence na zaměstnanost ve zdravotnictví v roce 2030.

# TEORETICKÁ ČÁST

# 1. Data

Data jsou všude kolem nás. Globální narušení a mezinárodní iniciativy jsou hybnou silou datové komunikace. Datafikace odkazuje na moderní trend digitalizace všech aspektů života (Panesar, 2020).

Toto vytváření dat umožňuje transformaci dat do nových a potenciálně cenných forem. Dnešní města jsou motivována, aby se stala chytřejšími. V nepříliš vzdálené budoucnosti budou naše města sbírat tisíce proměnných v reálném čase, aby optimalizovala, udržovala a zlepšovala kvalitu života pro celou populaci. Mohou být dokonce připojena k vašim účtům Amazon nebo Google a mít podvědomí o všech důležitých událostech související s naším zdravím a každodenním životem. To vše povede k prohloubení interakce mezi člověkem a počítačem, které vyústí v kvalitnější a chytřejší péči o naše zdraví.

Možnosti jsou nekonečné. Velká data z připojených zařízení, vestavěných senzorů a internetu věcí vedla k celosvětové potřebě analýzy, interpretace a vizualizace dat. COVID-19 se ukázal být skvělým příkladem sdílení dat mezi zeměmi a lékaři, zejména v souvislosti s rizikovými faktory, komorbiditami a komplikacemi nového koronaviru.

## 1.1. Co jsou data

Samotná data mohou mít mnoho podob – znak, text, slovo, číslo, obrázek, zvuk nebo video. Každá část dat spadá do dvou hlavních typů: strukturovaných a nestrukturovaných.

Data jsou souborem hodnot kvalitativních nebo kvantitativních proměnných. Aby se data stala informací, vyžadují interpretaci. Informace jsou organizovaná nebo utajovaná data, která mají pro příjemce nějakou smysluplnou hodnotu (nebo hodnoty). Informace jsou zpracovaná data, na nichž by měla být založena rozhodnutí a opatření.

Zdravotní péče prochází datovou revolucí, a to díky dvěma hlavním trendům: potřebě zkrotit rostoucí tlaky na náklady, které vytvářejí nové pobídky a struktury úhrad, a digitálnímu zdraví, které mění zdravotní péči prostřednictvím digitálních zařízení a inovativních technologií.

Klinické trendy se mění. Příchod sociálních médií a okamžitý přístup k informacím prostřednictvím internetu a zdravotnických komunit znamená, že pacienti mají přehled o

svém zdraví. Pacienti neustále generují data a chtějí svá data vlastnit a přenášet mezi službami. Současně se medicína založená na důkazech snaží o personalizovanou péči, vedenou přístupem k rozhodování založeným na důkazech. Pacienti i zdravotničtí pracovníci hlasitě apelují na všechna dostupná klinická data, aby se mohli účinněji rozhodovat o léčbě na základě současných důkazů.

Začlenění dat do moderní zdravotní péče poskytuje příležitost k výraznému zlepšení v celé řadě oblastí. Vzhledem k tomu, že se medicína vyvíjí, aby se stala více personalizovanou a založenou na důkazech, lze data použít ke zlepšení:

- pohodlí pacientů;
- vzdělávání pacientů;
- predikce rizik nemocí a péče;
- vzdálené monitorování;
- zvládání nemocí;
- individualizace a personalizace léčby a péče;
- finanční, transakční a environmentální předpovědi, plánování a přesnost (zvláště důležité pro jakékoli provozní a finanční týmy ve zdravotnických organizacích);
- zlepšení uživatelské zkušenosti, které lze použít k urychlení všech předchozích položek.

Data usnadňují informace, informace posilňují poznatky a poznatky vedou k lepšímu rozhodování (Panesar, 2021).

### **1.1.1. Typy Dat**

Strukturovaná data obvykle odkazují na něco uloženého v databázi – ve struktuře, která následuje po modelu nebo schématu. Téměř každá organizace bude s touto formou dat obeznána a možná ji již efektivně využívá. Většina organizací bude ukládat alespoň nějakou formu dat, například do tabulek v Excelu. V klinickém prostředí mohou mít elektronické zdravotní záznamy také podobnou strukturovanou formu.

Hodnoty z integrovaných senzorů, chytrých telefonů, chytrých hodinek a zařízení internetu věcí (IoT) jsou typické formy strukturovaných dat. Ať už se jedná o měření hodnot glykémie, kroků, spálených kalorií, srdeční frekvence nebo tlaku krve.

Strukturovaná data jsou podobná strojovému jazyku. Vysoce organizovaná, strukturovaná data usnadňují jednoduché a přímé operace vyhledávání a získávání informací. Strukturovaná data by typicky byla účelně uložena v relační databázi.

Nestrukturovaná data se týkají všeho ostatního. Nestrukturovaná data nemají předdefinovaný model ani schéma. Nestrukturovaná data v sobě nemají žádnou identifikovatelnou strukturu, což představuje problémy s dotazováním a vyhledáváním informací. Emaily, textové zprávy, příspěvky na Facebooku, tweety na Twitteru a další příspěvky na sociálních médiích jsou dobrým příkladem nestrukturovaných dat. Nestrukturovaná data mohou uvolnit poklad informací a správnosti. Zpráva společnosti Gartner naznačila, že objem dat má v příštích 5 letech vzrůst o 800 %, a 80 % těchto dat bude ve formě nestrukturovaných dat (McDonald, Rowsell-Jones, 2012).

Strukturovaná data jsou tradičně snadněji analyzována než nestrukturovaná data, a to kvůli nezpracované a neorganizované formě nestrukturovaných dat. Analytika nestrukturovaných, neformálních datových sad se však zlepšuje s využitím datové vědy a metod strojového učení, jako je zpracování přirozeného jazyka (NLP), které pomáhají porozumět a klasifikovat sentiment.

Polostrukturovaná data zapadají do prostoru mezi strukturovanými a nestrukturovanými daty. Semi-strukturovaná data nemusí nutně odpovídat formálním strukturám datových schémat spojených s relačními databázemi nebo datovými tabulkami. Polostrukturovaná data však mohou obsahovat značky nebo znaky k oddělení sémantických prvků a vynucení seskupení záznamů a polí v datech.

Data jsou palivem potřebným k učení v inteligentním systému a jsou nejdůležitější součástí úspěšné umělé inteligence (AI).

Data rozdělujeme do pěti zdrojů:

- data z webu a sociálních médií: kliknutí, historie, zdravotní fóra;
- data ze stroje na stroj: senzory, chytré zařízení jako hodinku, náramku a další;
- velké transakční údaje: údaje o zdravotních nárocích, fakturační údaje;
- biometrická data: otisky prstů, genetika, biomarkery pocházející z nositelných zařízení;
- údaje generované lidmi: e-mail, papírové dokumenty, elektronické lékařské záznamy.



## 1.2. Big data

Termín velké data je dán sběrem objemných, tradičních a digitálních dat, která jsou zdrojem pro objevování a analýzu.

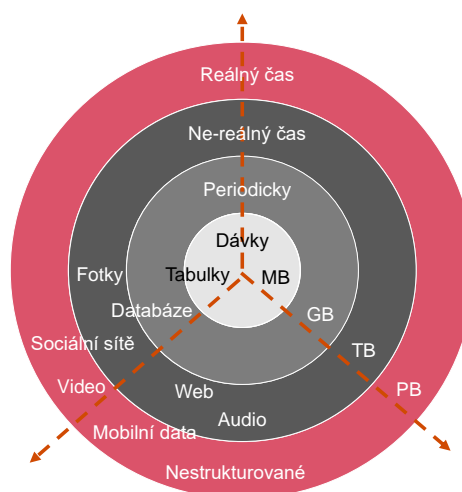
Big data je populární termín, který definuje datové sady, které jsou příliš velké na to, aby mohly být uloženy a zpracovány v konvenčním relačním databázovém systému. Tímto způsobem je termín big data vágní, zatímco velikost je bezpochyby součástí velkých dat, samotné měřítko neříká celý příběh o tom, co dělá velká data skutečně velkými.

Prostřednictvím analýzy velkých dat můžeme odhalit skryté vzorce, neznámé korelace, trendy, preference a další informace, které mohou zúčastněným stranám pomoci činit lepší a informovanější rozhodnutí. Strojové učení poskytuje soubor nástrojů technik, které lze použít pro účel datové sady.

Big data poprvé popsal Laney v roce 2001 a uvedl následující charakteristiky, které jsou známé také jako tři „V“ (obrázek 2) (Trifu, 2014):

- **Objem:** Množství generovaných a uložených dat. Obrovská velikost dat s sebou přináší jeho složitost v podobě ukládání, indexování a načítání.
- **Různost dat:** Big data se liší v typech a povaze dat, což vyžaduje efektivní ukládání a analýzu, jakož i systémy pro zpracování těchto dat.
- **Rychlost:** Velká data jsou přijímána obrovskou rychlostí, která přináší své požadavky a výzvy.

Obrázek 2: Tři „V“ dat



Vlastní zpracování, zdroj: *Kniha Machine Learning and AI for Healthcare*, Arjun Panesar

Datová věda již nějakou dobu diskutuje o třech „V“ big dat. Je však třeba vzít v úvahu dva společné aspekty dat, které jsou možná důležitější než základní tři. Jedná se o pravdivost dat a hodnotu dat. Ve skutečnosti existují různé podskupiny „V“ důležité pro organizace, které je třeba mít na paměti při vývoji datové strategie. Jedná se především o: objem, rozmanitost, rychlost, hodnotu, věrohodnost, dobu platnosti, variabilitu a vizualizaci.

### 1.2.1. Objem dat

Klíčovou setrvačností při přijímání velkých dat byla historicky související poptávka po úložišti a s tím spojenými náklady. Přesun datových center z vlastních data center do cloudu tuto obavu zažehnal. A co je ještě lepší, přináší to výhody. Především bezpečnost a škálovatelnost. To výrazně snížilo náklady na úložiště a také poskytlo flexibilitu při ukládání dat, spolupráci a obnově po havárii. Cloud nebo cloud computing označuje distribuované výpočty poskytované prostřednictvím internetu.

Moorův zákon neovlivnil pouze skladovací kapacitu, ale prospěl i nákladům. Moorův zákon je prohlášení, které učinil Gordon Moore, spoluzakladatel společnosti Intel, v roce 1965. Moore uvádí, že počet tranzistorů, které jsou schopné vejít se na integrovaný obvod, se přibližně každých 18 měsíců zdvojnásobí. V roce 1981 byla cena gigabajtu úložiště 300 000 dolarů. V roce 2004 to byl 1 dolar a v roce 2010 to bylo 0,1 dolaru. Dnes lze 1 GB pronajmout v cloudu za 0,023 dolaru měsíčně, přičemž první rok úložiště je zdarma.

Zdravotní péči, ať už je poskytována prostřednictvím veřejných nebo zdravotnických služeb, vlád, farmaceutických organizací nebo korporací, není pojem big data cizí. Data jsou nepřetržitě shromažďována digitálně i nedigitálně a pokrývají různé informace o pacientech, chování, epidemiologii, životním prostředí a zdravotnictví. Data se spotřebovávají, spravují a generují pro použití v péči o pacienty, transakčních záznamech, výzkumu, dodržování předpisů, procesech a regulačních požadavcích.

Pacienti a zdravotnické organizace ve svém prostředí přijímají chytré telefony, nositelná zařízení, jako jsou chytré hodinky a fitness trackery, domácí senzory, inteligentní osobní asistenti a sociální sítě. Data jsou ve velkém množství přístupná a spolehlivá. V současné době ji však většina poskytovatelů zdravotní péče nepoužívá a zdůrazňuje následující:

- generování údajů o pacientech probíhá mimo systém zdravotní péče;

- údaje generované pacienty je třeba integrovat a prozkoumat v rámci rozsáhlejších záznamů o pacientech nebo na časové ose zdraví;
- údaje o pacientech přesahují rámec lékařských markerů k holističtějším markerům včetně chování a zdraví životního prostředí;
- pro zdravotnické organizace je stále obrovský zájem, aby byly realizovány z důvodu výhod a nákladů.

### 1.2.2. Rozmanitost dat

Rozmanitost se netýká pouze variací v typech dat, ale také zdrojů a případů použití. Před dvaceti lety jsme dříve ukládali data ve formě tabulek a databází. Dnes mohou být data ukládána ve formě fotografií, dat ze senzorů, tweetů, narušených souborů atd. Tato rozmanitost nestrukturovaných dat vytváří problémy s ukládáním a analýzou dat. V této oblasti může strojové učení lidem výrazně pomoci.

Data jsou k dispozici ve formě strukturovaných dat (ve formě klinických záznamů a výsledků), nestrukturovaných dat (např. ve formě komunikačních a interakčních dat) a také polostrukturovaných dat (např. rentgenový snímek).

Řada big dat zahrnuje tradiční zdroje dat i novější zdroje strukturovaných i nestrukturovaných dat. Jak roste rozmanitost dat, roste i potenciál toho, čeho mohou dosáhnout algoritmy a nástroje strojového učení nasazené ve zdravotnictví. Různorodost big dat mohou mít podobu:

- datové typy: text, čísla, zvuk, video, obrázky atd.;
- funkce: případy použití a požadavky uživatelů;
- hodnota dat: odpovídají data účelu? Jednalo by se o zaměření na kvalitu v kontextu používání dat. Více dat nemusí nutně znamenat lepší data.

Stejně jako existuje celá řada zdrojů dat, hodnot, typů a případů použití, existuje také řada aplikací pro data. Patří sem přístupové body (tj. web, mobilní zařízení), SaaS [software jako služba], někdy také známé jako MDaaS [zdravotnické zařízení jako služba], API [rozhraní pro programování aplikací] a uživatelé (obvykle buď lidé nebo stroje).

### 1.2.3. Rychlost dat

Rychlosti míněno rychlost, s jakou jsou data vytvářena, ukládána a připravována pro analýzu a vizualizaci. Rychlost dat klade jedinečné požadavky na základní infrastrukturu

počítačového hardwaru. Výhodou cloud computingu byla schopnost rychle ukládat a zpracovávat objem a rozmanitost velkých dat, která by obvykle zahltila tradiční server. Cloud computing je upřednostňovanou cestou pro velké datové projekty kvůli flexibilitě týkající se úložiště a nákladů. Poskytovatelé cloudu mohou ukládat petabajty dat a škálovat tisíce serverů v reálném čase podle požadavků. Velmi cenné je, že výpočetní výkon je pro koncového uživatele levný a distribuovatelný.

V éře velkých dat se data vytvářejí v reálném čase nebo téměř v reálném čase. Všudypřítomnost kyberneticko-fyzických zařízení, vestavěných senzorů a dalších zařízení znamená, že k přenosu dat může dojít v okamžiku, kdy data vznikla.

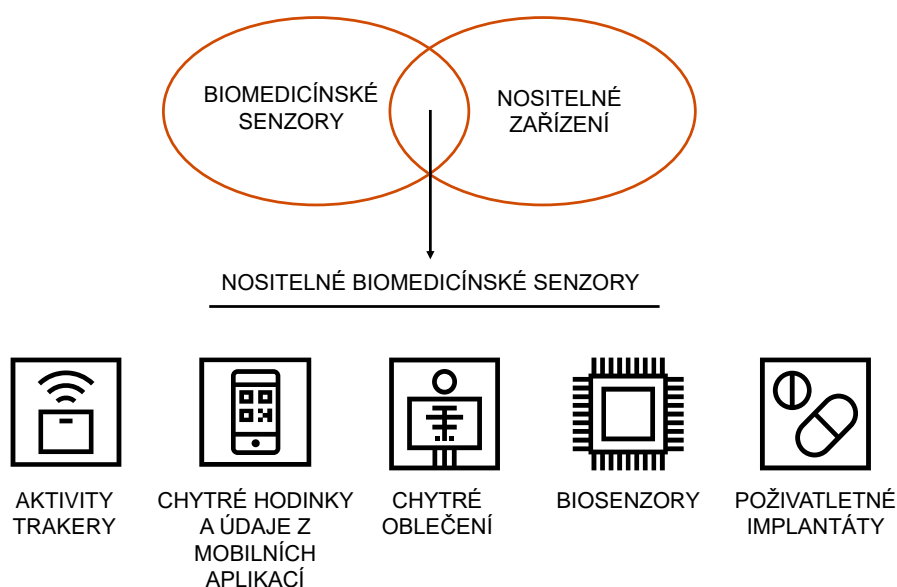
Rychlost vytváření dat je nepředstavitelná. Je všeobecně známo, že data generovaná za předchozí 2 roky jsou větší než data vytvořená od začátku času do té doby (Service Blog, 2020).

Klíčem k maximální klinické hodnotě je integrace různých a heterogenních zdrojů dat (tabulka 1, obrázek 3).

*Tabulka 1: Penetrace a druhy dat ve zdravotnictví*

Data	Příklad	Charakteristika
Chování a sentiment pacienta	Sociální média, smartphony, web, fóra	Většina dat je nestrukturovaná nebo polostrukturovaná. Data jsou velká a v reálném čase. Příležitosti pro učení, zapojení a porozumění. Typicky otevřený zdroj.
Údaje o zdravotním stavu pacienta	Senzory glukometrů, chytré telefony, fitness trackery, obrázky	Ve vlastnictví pacienta. Data jsou obvykle hlášena zařízením. Požadavky na úložiště jsou obvykle nízké. Požadavky na ukládání obrázků se liší v závislosti na datech požadovaných pro ukládání. Data ze senzorů mají vysokou rychlost. Strukturovaná data umožňují snadnou klasifikaci a detekci vzorů. Některé položky mohou vyžadovat digitalizaci.
Farmaceutická výzkumná a vývojová data	Údaje z klinických studií, údaje o populaci	Zahrnuje údaje o populaci a chorobách ve vlastnictví farmaceutických společností, výzkumu, akademické obce a vlády.
Zdravotní údaje na webu	Portály pacientů, digitální intervence (např. Nízkosacharidový program, hypo program)	Užitečná metadata. Zařízení je hlášeno spíše než klinicky.
Klinické údaje	Registry pacientů, snímky dat lékařů	Strukturovaná data. Ve vlastnictví poskytovatele.
Reklamace, náklady, administrativní data	Nároky za výdaje spojené s daty	Strukturovaná data. Zahrnuje všechny datové sady relevantní pro náhradu.
Transakční data	Výměny zdravotních informací	Může vyžadovat předběžné zpracování, aby se stalo užitečnými daty

Obrázek 3: Nositelné biomedicínské senzory



Vlastní zpracování, zdroj: Kniha *Machine Learning and AI for Healthcare*, Arjun Panesar

Tradiční tři „V“ poskytují pohled na rozsah dat a rychlost, s jakou tyto obrovské soubory dat rostou a množí se.

#### 1.2.4. Hodnota dat

Hodnota odkazuje na užitečnost dat. Klíčem k budování důvěry v data je zajištění jejich přesnosti a použitelnosti. Ve zdravotnictví by to bylo hodnoceno kvalitativní a kvantitativní analýzou. Hodnotu lze ohraničit řadou faktorů, zejména klinickým dopadem na výsledky pacientů nebo změnami v zapojení pacientů a zúčastněných stran, vlivem na procesy a pracovní toky a úsporami nákladů a dalšími nákladovými výhodami.

McKinsey uvedl, že potenciální roční hodnota velkých dat pro americké zdravotnictví je 300 miliard dolarů. Rovněž zmiňují, že velká data mají pro evropskou správu veřejného sektoru potenciální roční hodnotu 250 miliard dolarů (Kayyali a spol., 2013).

Hodnota spočívá v tom, jak organizace používají dostupná data a stávají se informačně zaměřenými a řízenými daty pro účely rozhodování. Hodnota nebývá nalezena, ale musí být vytvořena. Klíčem je zajistit, aby data byla pro uživatele a organizaci smysluplná tím, že budou dobře spravována.

Google odvedl perfektní práci při sledování vyhledávání příznaků chřipky v roce 2013. Nemohl však vědět, že jeho výsledky nebyly platné pro předpovídání prevalence chřipky. Google totiž nevěděl, že skutečné případy chřipky v reálném světě neodpovídají

distribuci a frekvenci vyhledávání. Takže s tím nikdy nemohl nic udělat. Pokud by měl být tento projekt prodloužen, mohla by být analýza příspěvků na sociálních médiích pro stanovení prevalence chřipky cennější. Sociální média a širší nestrukturované zdroje dat poskytují množství často neprozkoumaných informací.

Takovou jasnou a hodnotnou návratností investic může být identifikace populace s rizikem diabetu 2. typu, predikce fibrilace síní (AFIB), pomoc pacientům při zotavení po zdravotních událostech a doporučení pacientům nejnovější léčebné terapie založené na důkazech.

### 1.2.5. Věrohodnost dat

Bezcenná data, bez ohledu na to, v jakém množství, budou vždy bezcenná. Věrohodnost se týká pravdivosti údajů a toho, zda mají data optimální kvalitu a vhodnost v kontextu jejich použití souvisejícího se zkreslením, šumem a abnormalitou v datech.

Na pravdivost údajů mohou mít vliv různé faktory, včetně následujících:

- Vstupní data: Byla data zadána správně? Vyskytly se nějaké chyby? Existuje auditní záznam zadávání dat?
- Správa dat: Jaká je integrita dat pohybujících se systémem?
- Kvalita integrace: Jsou data jedinečná vhodně odkazována, komentována?
- Stabilita: Jsou data vhodná k použití? Jsou včasné a relevantní?
- Použití: Jsou data k dispozici způsobem, který lze použít? Budou data užitečná k dosažení obchodních cílů? Je etické používat tato data?

Je důležité zajistit, aby soubory dat byly čisté, úplné, aktuální, konzistentní a v souladu s regulacemi a předpisy.

- Čistá data: Vzniknou jako výsledek dobrých postupů kvality dat, což může být deduplikace, standardizace, ověřování, párování a zpracování. Čistá data umožňují robustní rozhodování o kvalitě.
- Aktuální data: Současná data jsou obvykle považována za důvěryhodnější než zastaralá data. Otázkou pak je, v jakém okamžiku nejsou data aktuální? Jak aktuální data rostou v reálném čase, dešifrování signálu od šumu se může stát náročnějším úkolem.
- Konzistentní data: Data musí být konzistentní (což umožňuje strojovou čitelnost). To je vyžadováno pro křížovou kompatibilitu systémů a platí to i pro metadata.

- Vyhovující data (v souladu s regulacemi a předpisy): Regulace dodržování předpisů mohou pocházet z různých zdrojů, jako jsou zúčastněné strany, klienti, právní legislativa, nebo v důsledku nové politiky. Ačkoli existuje mnoho datových infrastruktur a standardů, datová věda jako kategorie stále přichází s prvními problémy, pro které je vyžadována správa a legislativa. Soulad může pro interní a externí zúčastněné strany znamenat různé věci. Interně budou existovat standardy zajišťující shodu dat s postupy kvality, zabezpečení a ochrany osobních údajů. Všechny zúčastněné strany musí věřit, že k údajům bylo přistupováno a distribuováno interními a externími předpisy. Organizace obvykle mohou mít správní radu pro data a informace.

Na analýzy dat musí být zajištěná věrohodnost dat, aby data byly správná a akceptovatelná. Data musí být nejen důvěryhodná, ale také algoritmy a systémy, které je interpretují. Zadávaní dat může být rizikovým bodem, který bez dobré aplikované datové vědy zůstane bez povšimnutí. Obvykle se jedná o lidskou chybu např. v malých nastaveních dat. Aby byly údaje klinicky relevantní, vyžadují maximální pravdivost. Jak se pravdivost dat zlepšuje, tak strojové učení na těchto datech umožňuje poskytovat věrohodnější závěry.

#### **1.2.6. Platnost dat**

Podobně jako u věrohodnosti velkých dat existuje také problém platnosti dat s odkazem na to, zda jsou data správná a přesná pro zamýšlené použití. V klinických aplikacích lze za prioritní považovat platnost dat, aby se zajistilo, že budou použita pouze užitečná a relevantní data. Pravdivost dat nebo pravdivost údajů je absolutní, zatímco platnost je kontextová. Platná data představují skutečný svět bez předpojatosti.

#### **1.2.7. Variabilita dat**

Velká data jsou variabilní. Variabilita definuje data, jejichž význam se pravidelně mění. Variabilita je velmi důležitá při provádění analýzy sentimentu. Například v sérii tweetů může mít jedno slovo úplně jiný význam.

Variabilita je často zaměňována s rozmanitostí. Pro ilustraci, květinářství může prodávat pět druhů růží. To je pestrost neboli rozmanitost. Pokud tedy budete chodit po dobu dvou týdnů do květinářství a koupíte si každý den stejnou růži, bude mít každý den jemně jinou barvu, tvar a vůni. To je variabilita.

Aby bylo možné provést správnou analýzu sentimentu, musí být algoritmy schopné porozumět kontextu textů a být schopny rozluštit přesný význam slova v tomto konkrétním kontextu. To je stále velmi obtížné, dokonce i s pokrokem ve schopnostech zpracování přirozeného jazyka.

### **1.2.8. Vizualizace dat**

Vizualizace je poslední „V“, odkazující na příslušné analýzy a vizualizace požadované u velkých dat, aby byla čitelná, srozumitelná a použitelná.

Vizualizace nemusí znít složitě. Vizualizace komplexních datových sad je klíčová pro lepší porozumění a rozvoj zúčastněných stran. Poměrně často se právě vizualizace dat stává hlavní součástí přenosu informací a znalostí získaných z datových sad uživatelům.

## **1.3. Využití dat ve zdravotnictví a farmacii**

Stakeholderi ve zdravotnictví chápou, že jsou obklopeni množstvím údajů od pacientů, odborníků a transakcí. Je klíčové vědět, jak řídit a naplňovat hodnotu společnosti a plnit KPI (klíčové ukazatele výkonnosti). V následující části představím několik případových studií z praxe.

### **1.3.1. Předpověď doby čekání**

V Paříži se čtyři nemocnice, které tvoří asistenční službu Publique-Hopitaux de Paris (AP-HP), spojily s Intelem a použily data z interních a externích zdrojů, včetně 10letých záznamů o příjmu pacientů do nemocnic, aby předpověděli očekávaný počet pacientů, kteří vstoupí do jejich zařízení (Renaud a spol., 2009).

Techniky analýzy časových řad byly použity k predikci míry přijetí v různých časech. Tato data byla zpřístupněna všem ordinacím a klinikám a demonstrují okamžitý způsob, jak lze data využít ke zlepšení efektivity a posílení zainteresovaných stran.

Většina, ne-li všechny kliniky po celém světě mají přístup k podobným datům, což dokazuje, že zdravotnictví je teprve na začátku využití datových aplikací.

### **1.3.2. Prediktivní analytika**

Předchozí příklad by mohl velmi dobře používat statická data (tj. malá data, která nejsou v reálném čase) a být poměrně přesný při předpovídání čekacích dob a intervalů



zpětného příjmu. Stejný koncept analýzy dat lze použít v měřítku pro predikci onemocnění a demokratizaci péče.

Ve Spojených státech společnost Optum Labs shromáždila EHR (elektronické zdravotní záznamy) pro více než 30 milionů pacientů a vytvořila databázi pro nástroje prediktivní analýzy ke zlepšení poskytování péče. Záměrem je umožnit lékařům činit informovaná rozhodnutí na základě dat a zlepšit tak léčbu pacientů (Wallace a spol., 2014).

Robustnost, kterou poskytuje 30 milionů zdravotních záznamů, umožňuje trénovat a ověřovat modely za účelem nalezení lidí, kteří vyhovují prediktivním rizikovým trendům pro určité nemoci, jako je hypertenze, kardiovaskulární onemocnění, mrtvice, diabetes 2. typu, srdeční onemocnění a metabolický syndrom.

Analýzou údajů o pacientech, včetně věku, sociální a ekonomické demografie, fitness a dalších zdravotních biomarkerů, mohou poskytovatelé zlepšit péči na úrovni jednotlivce i populace, a to nejen prostřednictvím předvídání rizik, ale i poskytováním léčby pro optimální výsledky pacientů.

### **1.3.3. Elektronické zdravotní záznamy**

Myšlenka elektronických zdravotních záznamů (EHR) se zatím úplně nenaplnila. Myšlenka je teoreticky jednoduchá: že každý pacient má digitální zdravotní záznam sestávající z jeho informací, demografie, anamnézy, alergií, klinických výsledků a tak dále. Záznamy lze se souhlasem pacienta sdílet prostřednictvím zabezpečených počítačových systémů a jsou k dispozici poskytovatelům zdravotní péče z veřejného i soukromého sektoru. Každý záznam obsahuje jeden upravitelný soubor, což znamená, že lékaři mohou provádět změny v průběhu času bez nebezpečí replikace dat nebo nekonzistencí.

EHR dává perfektní smysl. Nicméně implementace v celé zemi představuje složitý úkol. Ve Spojených státech až 94 % způsobilých nemocnic používá EHR podle výzkumu HITECH. Evropa je mnohem více pozadu. Směrnice Evropské komise stanovila cíl vytvořit do roku 2020 centralizovaný evropský systém zdravotních záznamů (Evropská komise, 2020).

Ve Spojených státech společnost Kaiser Permanente zavedla systém, který sdílí data ve všech jejích zařízeních a usnadňuje používání EHR. Zpráva společnosti McKinsey

zdůraznila, jak systém sdílení dat dosáhl odhadovaných úspor ve výši 1 miliardy dolarů v důsledku snížení počtu návštěv v kanceláři a laboratorních testů. Systém sdílení dat zlepšil výsledky u kardiovaskulárních onemocnění (Quarterly a spol., 2009).

EHR se vyvíjí v blockchain, který se snaží decentralizovat a distribuovat přístup k datům.

#### **1.3.4. IoT (internet věcí ve zdravotnictví): notifikace v reálném čase, alerty, automatizace**

Milióny lidí používají zařízení, která měří jejich životní funkce směrem ke kvantifikovanému já. Zařízení připojená k internetu v současné době obsahují váhy, monitory aktivity (například Fitbit, Apple Watch, Microsoft Band), které měří srdeční frekvenci, pohyb a spánek, a také měřiče glykémie, které odesílají výsledky v reálném čase a sledují chování uživatelů. Dokonce existuje zařízení, které měří srdeční tep lidského plodu (Roham a spol., 2011).

Zaznamenaná data by mohla být použita k detekci rizika onemocnění, upozornění lékařů nebo vyžádání pohotovostních služeb v závislosti na přijatých biometrických údajích. Mnoho integrovaných zařízení jde dál, a kromě pulsu a pohybu měří i množství potu, hladiny oxidace, hladinu glukózy v krvi, spotřebu nikotinu a další.

Se sofistikovanými zařízeními přicházejí sofistikovaná řešení nových problémů. Například nyní, kdy je monitorování srdeční frekvence levnější a všudypřítomné, lze stavy, jako je fibrilace síní, detekovat mnohem snadněji a mnohem dříve než kdykoli předtím. Chvění srdeční frekvence nad 300 tepů za minutu (spíše než 60–80 tepů za minutu) může být příznakem fibrilace síní. U pacientů s diagnostikovaným stavem je o 33 % vyšší pravděpodobnost vzniku demence a více než 70 % pacientů zemře na cévní mozkovou příhodu (Henon a spol., 1997). Léčba může být často jednoduchá: antikoagulancia a ředidla krve, jejichž účinnost je až 80 %.

Poskytovatelé zdravotní péče se pomalu posouvají směrem k sofistikovaným sadám nástrojů, aby využili masivní datový tok vytvořený pacienty a reagovali pokaždé, když se výsledky zdají znepokojivé.

Nabízení uživatelům zařízení s hmatatelnými pobídkami a benefity, jako jsou slevy na zdravotní nebo životní pojištění, by mohlo vést k prevenci mnoha chronických

onemocnění souvisejících s životním stylem, podobně jako pojištění automobilů se zabudovanými senzory černé skříňky.

## 1.4. Přeměna dat na informace

Hodnota dat je demonstrována ve schopnosti je převést na informace, které mohou poskytnout užitečné informace pro řízení chování a pracovních postupů. Analýza dat je podmnožinou datové vědy, která se snaží čerpat poznatky ze zdrojů nezpracovaných dat.

Dnešní pokroky v analýze velkých dat umožňují výzkumníkům dekodovat lidskou DNA během několika minut, předvídat, kde teroristé plánují zaútočit, a určit, který gen je s největší pravděpodobností odpovědný za konkrétní nemoci, a samozřejmě, na které reklamy s největší pravděpodobností zareagujete na Facebooku.

Analýza velkých dat může organizacím pomoci porozumět informacím obsaženým v jejich datech a také identifikovat data, která jsou nejdůležitější pro obchodní cíle, výsledky a rozhodnutí.

Většina velkých dat je nestrukturovaná. IBM odhaduje, že 80 % světových velkých dat je nestrukturovaných, což zahrnuje fotografie, zvuk, video, dokumenty a interakce (Takmi, 2018).

Data musí být zpracována, aby měla nějakou hodnotu. K tomu jsou zapotřebí zdroje ve formě času, talentu a peněz. Při přijímání přístupů k velkým datům se klade následující otázka: Stojí zpracování dat a následná získaná informace za vynaložený kapitál a riziko?

V taxonomii analýzy datových věd jsou čtyři proudy (obrázek 4). Klasifikace jsou založeny na typech výsledků, které jsou vytvářeny.

Obrázek 4: Proudý analytiky



*Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar*

### 1.4.1. Deskriptivní analytika

Deskriptivní analytika přináší pohled do minulosti se zaměřením na otázku, co se stalo, prostřednictvím analýzy dat z historie. Popisná analytika využívá techniky jako je agregace dat a dolování dat, aby poskytla historické porozumění. Z deskriptivní analýzy se toho lze hodně naučit.

Běžnými příklady deskriptivní analýzy jsou zprávy, které poskytují odpovědi na otázky jako kolik pacientů bylo loni v červenci přijato do nemocnice, kolik bylo přijato zpět do 30 dnů a kolik z nich chytlo infekci. Deskriptivní analytika slouží jako první forma analýzy velkých dat.

Data se mohou stát cennějšími díky propojení datových sad. Například propojení návštěv lékaře s přijetím do nemocnice, s docházkou na vzdělávání, s dodržováním užívání léků a s přístupem ke službám může poskytnout podstatný vhled a pochopení pro optimalizaci poskytování zdraví a úsporu nákladů.

Omezení deskriptivní analýzy spočívá v tom, že poskytuje omezenou schopnost vést rozhodnutí, protože je založena na snímku minulosti. I když se jedná o metodiku užitečnou, ne vždy ukazuje na budoucnost.

### 1.4.2. Diagnostická analytika

Diagnostická analytika je forma analýzy, která zkoumá data, aby odpověděla na otázku, proč se něco stalo. Diagnostická analytika zahrnuje techniky, jako jsou rozhodovací stromy, dolování dat a korelace.

### 1.4.3. Prediktivní analytika

Prediktivní analytika nám umožňuje porozumět budoucnosti a předvídat pravděpodobnost budoucího výsledku. Prediktivní analytika využívá data, která jsou k dispozici, a pokouší se doplnit chybějící data nejlepšími odhady.

Prediktivní analytiku charakterizují techniky, jakými jsou regresní analýza, statistika s více proměnnými, dolování dat, párování vzorů, prediktivní modelování a strojové učení. Prediktivní analytika využívá historická a aktuální data k předpovídání pravděpodobnosti budoucích událostí nebo použitelných výsledků.

Dovednosti prediktivní analýzy jsou vyžadovány, protože poskytovatelé zdravotní péče hledají způsoby, jak snížit náklady, využít úhrady založené na hodnotě a vyhnout se sankcím spojeným se selháním kontroly nad chronickými nemocemi a nežádoucími účinky jejich léčiv, kterým lze předcházet. Posledních 5 let inovací přineslo významný pokrok v měřitelných dopadech na životy pacientů. Díky nositelným technologiím a chytrým aplikacím je nyní možné lokalizovat onemocnění, jako je astma, atriální fibrilace a COPD (Chronická obstrukční plicní nemoc). (Wooding, 2018).

Prediktivní analytika zůstává nepolapitelná, protože vyžaduje přístup k datům v reálném čase, což umožňuje rychlé rozhodování téměř ihned. Pro podporu prediktivní analytiky, musí být plně integrovány lékařské senzory a rovněž připojená zařízení, aby byly poskytovány aktuální informace o zdravotním stavu pacienta.

Kromě toho musí mít lékaři zkušenosti s používáním takových dat. Přesnější diagnózy a léčba musí kromě využití údajů o jednotlivých pacientech čerpat z co největšího množství údajů o pacientech, včetně údajů o úrovni populace.

Vývoj zdroje dat a technologie, pokročilá podpora rozhodování kognitivních výpočetních sil, zpracování přirozeného jazyka a prediktivní analytika mohou poskytovatelům pomoci určit diagnózy, na které by jinak nepřišli. Nástroje pro řízení zdraví populace mohou upozornit na ty, kteří jsou nejvíce ohroženi hospitalizací, rozvojem nákladného chronického onemocnění nebo nepříznivě reagují na léky.

#### 1.4.4. Preskriptivní analytika

Preskriptivní analytika se snaží přijímat rozhodnutí pro optimální výsledky, to znamená využívat všechna dostupná data a analýzy k informování a rozvíjení rozhodnutí o tom, jaké kroky podniknout, vedou tedy k chytřejším rozhodnutím. Preskriptivní analytika se pokouší kvantifikovat účinek budoucích rozhodnutí a poradit si s možnými výsledky předtím, než jsou učiněna rozhodnutí.

Preskriptivní analytika v nejlepším případě předpovídá nejen to, co se stane, ale také proč se to stane.

Preskriptivní analytika využívá kombinaci technik a nástrojů umělé inteligence, jakými jsou strojové učení, dolování dat a postupy výpočetního modelování. Tyto techniky jsou aplikovány proti vstupu z mnoha různých datových sad, včetně historických a transakčních dat, datových zdrojů v reálném čase a velkých datových sad.

Preskriptivní analytika provádí predikci v kontextu s přihlédnutím k dostupným důkazům. Například typická hospitalizační analytika zpětného přebírání zobrazuje prognózu pacientů, kteří se pravděpodobně vrátí v příštích 14–30 dnech. Užitečnější prediktor může integrovat stejný řídicí panel s projektovanými náklady, počty lůžek v reálném čase, dostupnými vzdělávacími materiály a následnou péčí. Díky dodatečným informacím mohou lékaři určit pacienty s vysokým rizikem a přijmout opatření k minimalizaci tohoto rizika, což povede ke zlepšení výsledků pacientů a snížení spotřeby nemocničních zdrojů. Farmaceutické společnosti mohou používat preskriptivní analytiku k urychlení vývoje léků tím, že určují cílové skupiny, které jsou pro klinické studie nejvhodnější. To zahrnuje pacienty, o kterých se předpokládá, že vyhovují, a pacienty, u kterých se neočekávají komplikace.

Preskriptivní analytika umožňuje optimalizaci řízení zdraví populace prostřednictvím identifikace vhodných intervenčních modelů pro jednotlivce v rizikové skupině populace, kombinováním klinických, patientských a širších dostupných zdravotních dat.

Nutno dodat, že analytika prochází další transformací, a to od preskriptivní analýzy až po kontextovou analýzu. Kontextová analýza bere v úvahu širší data, jakými jsou environmentální, lokalizační a situační data.

## 2. Strojové učení

Strojové učení je podmnožinou umělé inteligence, kde jsou počítačové modely trénovány, aby se v průběhu času učily ze svých akcí a prostředí se záměrem dosahovat lepších výkonů.

V roce 1959 definoval tvůrce Checker Arthur Samuel strojové učení jako „obor, který dává počítačům schopnost učit se, aniž by byly výslovně naprogramovány“ (Glauner a spol., 2018).

Strojové učení se zrodilo z rozpoznávání vzorů a teorie, že počítače se mohou učit, aniž by byly naprogramovány k provádění konkrétních úkolů. Tedy systémy, které se učí, aniž by byly explicitně naprogramovány. V důsledku toho je učení řízeno daty s inteligencí získanou prostřednictvím schopnosti činit efektivní rozhodnutí na základě povahy učebního signálu nebo zpětné vazby. Užitečnost těchto rozhodnutí se hodnotí porovnáním s cílem.

Strojové učení se zaměřuje na vývoj algoritmů, které se přizpůsobí prezentaci nových dat a objevům. Strojové učení je příkladem principů dolování dat, ale je také schopno odvodit korelace a naučit se z nich aplikovat na nové algoritmy. Cílem je napodobit schopnost učit se jako u člověka prostřednictvím zkušeností a dosažení zadaného úkolu bez nebo s minimální vnější (lidskou) pomocí.

Stejně jako učení u lidí se strojové učení skládá z mnoha přístupů. V první řadě se věci ukládají do paměti. Za druhé se učíme získáváním informací (čtením, nasloucháním, učením se novým věcem). A za třetí, učíme se z příkladů a získaných zkušeností.

Pokud bychom se například učili čtvercová čísla a studentům by byla ukázána sada čísel  $Y = \{1, 4, 9, 16, \dots\}$  a učili by se, že  $Y = n^2$ , obvykle by rozuměli pojmu čtvercová čísla, aniž by bylo výslovně označeno každé čtvercové číslo.

### 2.1. Rozdíl strojového učení od tradičního softwarového inženýrství

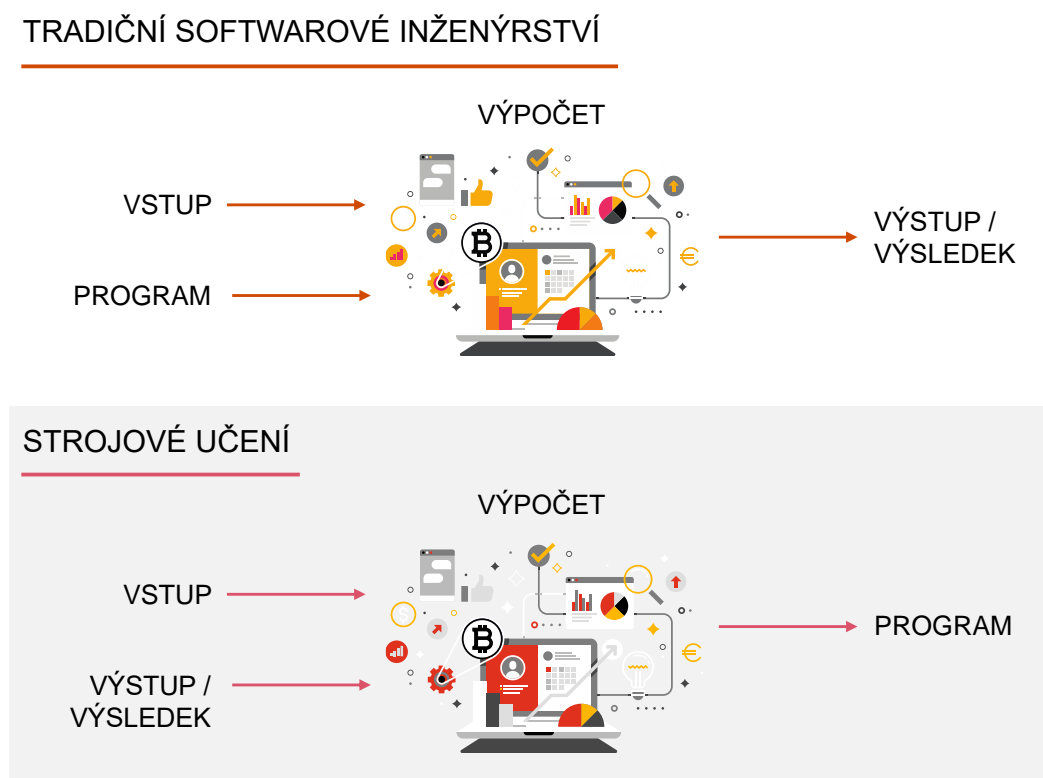
Tradiční softwarové inženýrství a strojové učení sdílejí společný cíl vyřešit problém nebo soubor problémů. Přístup k řešení problémů je to, co odlišuje tato dvě paradigmaty.

Tradiční softwarové inženýrství nebo programování odkazuje na úkol počítačového zpracování nebo automatizace úkolu tak, že funkce nebo program, daný vstup, poskytuje

výstup. Jinými slovy, zápis funkce  $f$ , je takový, že pro daný vstup  $x$ , výstup  $y = f(x)$ . To se provádí pomocí logiky, typicky příkazů if-else, while loops a boolean operators.

Strojové učení se liší od konvenčního programování v tom, že namísto poskytování instrukcí o funkci  $f$  má počítač k dispozici vstup  $x$  a výstup  $y$  a očekává se, že určí nebo předpoví funkci  $f$  (obrázek 5).

Obrázek 5: Jak strojové učení pracuje



Vlastní zpracování, zdroj: *Kniha Machine Learning and AI for Healthcare*, Arjun Panesar

Zatímco tradiční programy byly napsány lidmi, aby vyřešily problém, programy strojového učení se učí pomocí uvažování řešit problém z příkladů, pravidel a informací. Mnoho klasifikátorů ve strojovém učení pomáhá s predikcí.

Programy strojového učení se také mohou naučit zobecňovat a pomoci s otázkami nejistoty díky použití statistik a technik řízených pravděpodobností. Modely se mohou učit z předchozích výpočtů nebo zkušeností a vytvářet spolehlivá, opakovatelná rozhodnutí a výsledky.

Dostupnost dat, informací na internetu a skokový růst v množství digitálních dat, která jsou vytvářena, ukládána a zpřístupňována pro analýzu, umožňují inženýrům vytvořit stroje/kódy, které se mohou naučit, co preferujeme, a přizpůsobit je našim požadavkům.



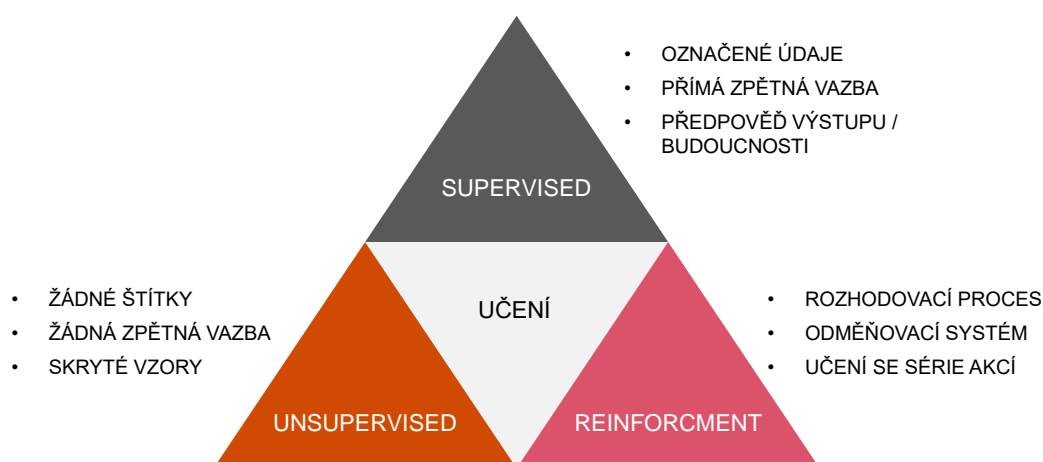
Strojové učení lze vidět všude, například hodnocení na webu, design léků, personalizaci preferencí a zdrojů sociálních sítí, doporučení hudby, logistiky, design webových stránek, samořídící auta, ledničky, pračky, televizory a virtuální asistenty.

## 2.2. Základy strojového učení

Úlohy pro algoritmy strojového učení jsou obecně rozděleny do následujících kategorií (obrázek 6):

- Supervised learning,
- Unsupervised learning,
- Semi-supervised learning,
- Reinforcement learning.

Obrázek 6: Metody strojového učení

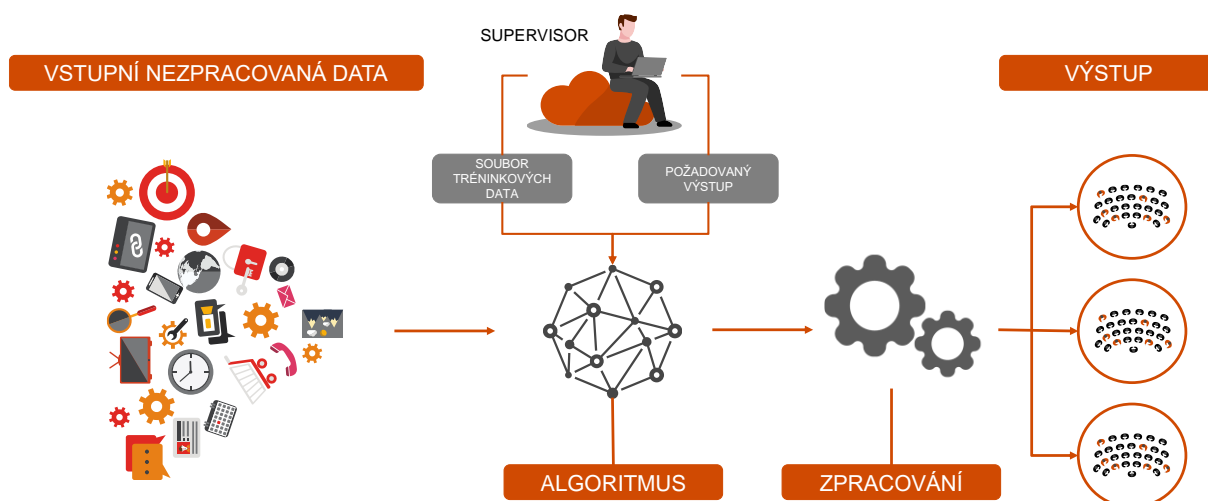


Vlastní zpracování, zdroj: *Kniha Machine Learning and AI for Healthcare, Arjun Panesar*

### 2.2.1. Supervised learning

Při učení pod dohledem jsou algoritmy prezentovány s příkladnými vstupy a následně požadovanými výstupy ve formě trénovacích dat s cílem naučit se obecná pravidla, která mapují vstupy na výstupy (obrázek 7). Vstupní data se nazývají tréninková data a je s nimi spojen známý výstup (výsledek).

Obrázek 7: Supervised learning



Vlastní zpracování, zdroj: *Kniha Machine Learning and AI for Healthcare*, Arjun Panesar

Tréninková data řídí vývoj algoritmu. Model je vytvořen tréninkovým procesem, ve kterém model vytváří předpovědi a je opraven, pokud jsou předpovědi nesprávné.

Trénink pokračuje, dokud model nedosáhne požadované úrovně přesnosti trénovacích dat. Mezi kontrolované algoritmy patří logistická regrese. Algoritmy strojového učení pod dohledem mohou aplikovat to, co se naučili v minulosti, na nová data pomocí označených příkladů dat k předpovědi budoucích událostí.

Učící algoritmus může také porovnávat výstup se správným výstupem, aby zjistil případné chyby a podle toho upravil model.

Problémy s učením pod dohledem lze využít v následujících formách.

## Klasifikace

Klasifikace neboli třídění má předpovídat výsledek na základě trénovací datové sady, kde je výstupní proměnná ve formě odlišných kategorií. Modely jsou vytvářeny prostřednictvím zadávání trénovacích dat ve formě předem označených dat.

Klasifikační techniky definují hranice rozhodování a zahrnují:

- Podpora vektorového stroje,
- Gaussův Bayes,
- k-Nejbližší sousedé (kNNs),
- Logistická regrese.

Příkladem klasifikace je označení (nebo v reálném světě diagnóza) někoho nemocného nebo nezdravého na základě souboru příznaků.

## Regrese

Regrese je velmi podobná klasifikaci. Jediný rozdíl mezi klasifikací a regresí je v tom, že regrese je výsledkem daného vzorku, kde výstupní proměnná je ve formě skutečných hodnot (tj. teplota jako hodnota, spíše než klasifikace horké nebo studené).

Příklady zahrnují výšku, tělesnou teplotu a váhu. Lineární regrese, polynomiální regrese, podpůrný vektorový stroj, soubory, rozhodovací stromy a neuronové sítě jsou příklady regresních modelů.

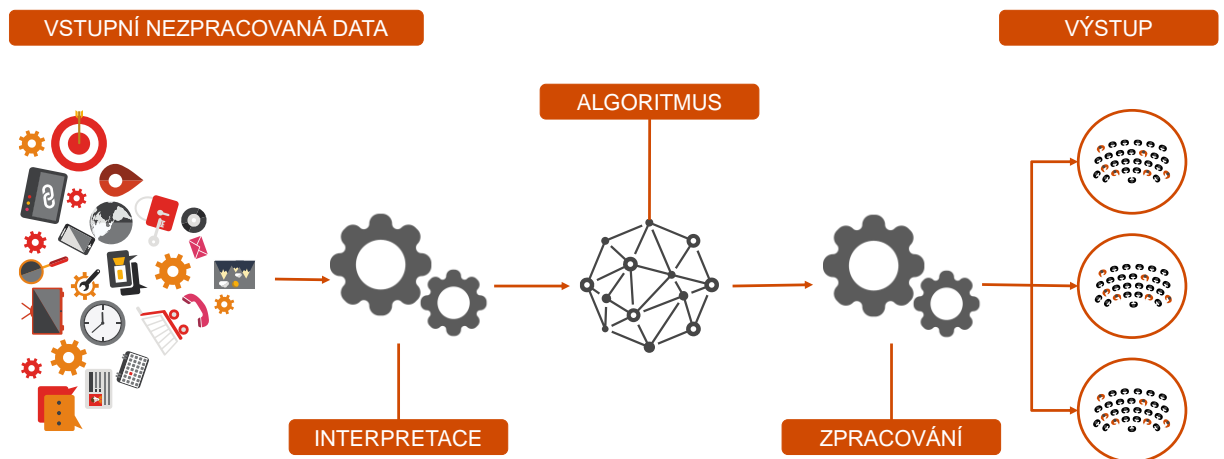
## Prognóza

Prognóza je metoda předpovědi na základě minulých a současných dat. Toto je také známé jako prognóza časových řad. Soubor je typ učení pod dohledem, který kombinuje více různých modelů strojového učení k předpovídání výsledku na novém vzorku. V sestavě se modely stávají funkcemi.

### 2.2.2. Unsupervised learning

Když lidé odkazují na systémy, které se mohou samy učit, ve skutečnosti odkazují na Unsupervised learning (obrázek 8). Při této metodě algoritmus učení nepřijímá označení pro data tzn., že algoritmus nachází strukturu přímo ze vstupu.

Obrázek 8: Unsupervised learning



Vlastní zpracování, zdroj: Kniha *Machine Learning and AI for Healthcare*, Arjun Panesar

Protože data nejsou označena, neexistuje žádné hodnocení přesnosti struktury, která je výstupem algoritmu.

V údajích může chybět klasifikace i označení. Model je tedy vyvíjen prostřednictvím interpretace: prostřednictvím hledání skrytých struktur a inferencí ve vstupních datech.

To může být způsobeno extrakcí pravidel, snížením redundance dat nebo jako výsledek organizace dat.

Zahrnuje shlukování, redukci dimenzionality a učení se pravidlům přidružení. Algoritmus nemusí nikdy najít správný výstup, ale místo toho modeluje základní strukturu dat.

Existují tři typy problémů s Unsupervised learning, jak je definováno níže.

### **Sdružení**

Sdružení neboli asociace je objev pravděpodobnosti společného výskytu předmětů ve sbírce dat. Tato metoda je široce používána v marketingu i ve zdravotnictví k potřebnému informování k rozhodování.

Například by to byla procentuální pravděpodobnost rozvoje jakékoli formy rakoviny, pokud by byla diagnostikována obezita. Asociace je podobná klasifikaci. Jakýkoli atribut však lze předpovědět v asociaci, zatímco klasifikace je binární.

### **Shlukování**

Shlukování neboli clustering se týká seskupování položek tak, že položky ve stejném clusteru jsou si více podobné než položky z jiného clusteru.

### **Zmenšení rozměrů**

Snížení rozměrů lze dosáhnout výběrem prvků a extrakcí prvků. Redukce rozměrů může matematicky znovu reprezentovat nebo transformovat data.

### **Extrakce funkcí**

Extrakce prvků provádí transformaci dat z vysokorozměrného do nízkorozměrného prostoru. To může zahrnovat snížení proměnné datové sady při zachování integrity dat a zajištění reprezentace nejdůležitějších informací.

Pro redukci rozměrů lze použít metody extrakce funkcí. Z původní sady funkcí se obvykle vytvoří nová sada funkcí. Příkladem může být sloučení všech výsledků klinických testů pacienta do skóre zdravotního rizika z výsledků, které prokazatelně ovlivňují úmrtnost.

Snížení počtu rozměrů umožňuje snazší vizualizaci, zejména ve dvou nebo třech rozměrech, a také zkrácení času a prostoru potřebného pro skladování. Algoritmy zahrnují:

- Apriori algoritmus,
- Růst FP (také známý jako častý růst vzoru),
- Skrytý Markovův model,
- Analýza hlavních komponent,
- Dekompozice singulární hodnoty,
- k-Means,
- Neuronové sítě,
- Hluboké učení.

Hluboké učení využívá architektury hlubokých neuronových sítí, což jsou typy algoritmů strojového učení.

### 2.2.3. Semi-supervised learning

Semi-supervised learning je hybridní metoda, kde je ve vstupu směs označených a neoznačených dat. Ačkoli může existovat požadovaný výsledek, model se také učí struktury pro organizaci dat a vytváření předpovědí. Příklady problémů jsou klasifikace a regrese.

### 2.2.4. Reinforcement learning

Reinforcement learning je tam, kde systémy interagují s dynamickým prostředím, ve kterém musí agent plnit konkrétní cíl. Reinforcement learning funguje jako průsečík mezi strojovým učením, psychologí chování, etikou a teorií informace.

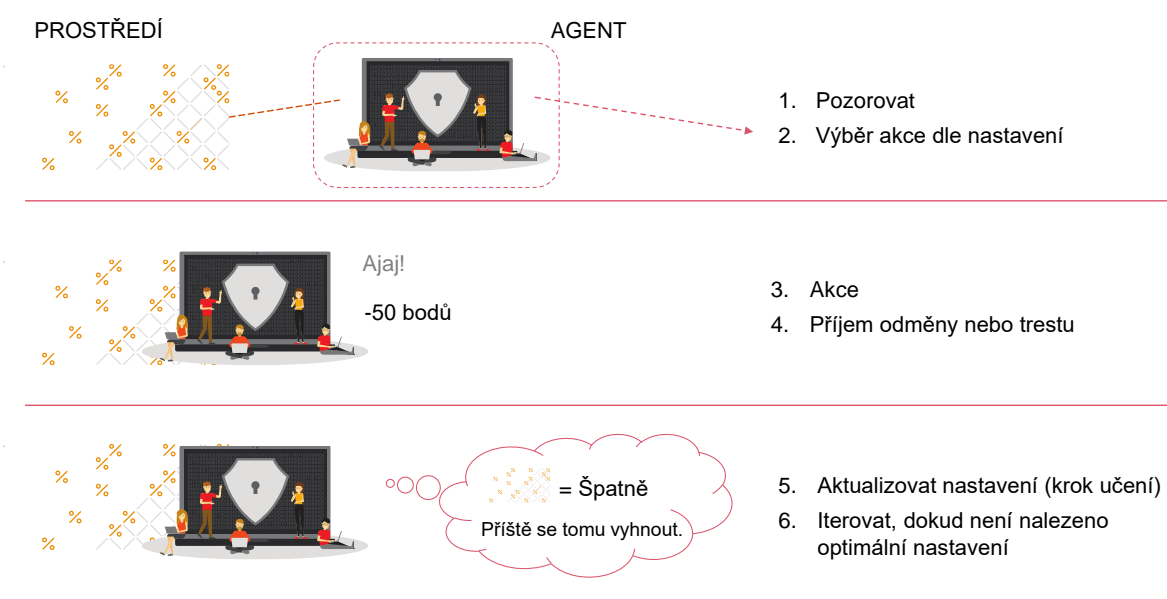
Algoritmus dostává zpětnou vazbu týkající se odměn a trestů, když se pohybuje v problému. Reinforcement learning umožňuje agentovi rozhodnout o další nejlepší akci na základě jeho aktuálního stavu a učením chování, které maximalizuje odměnu. Optimální akce (nebo optimální politika) se obvykle učí prostřednictvím pokusů a omylů a zpětné vazby. To umožňuje algoritmu určit ideální chování v kontextu.

Reinforcement learning se obvykle používá v robotice, například robotický vysavač, který se učí vyhýbat kolizím přijímáním negativní zpětné vazby prostřednictvím narážení do stolů, židlí a tak dále – i když dnes robotický vysavač zahrnuje také techniky počítačového vidění.

Reinforcement learning se liší od standardního supervised learningu v tom, že nikdy nejsou prezentovány správné příklady, ani nejsou výslovně opravena neoptimální rozhodnutí. Důraz je kladen na výkon v reálném světě.

Podobně jako se lidé učí, reinforcement předpokládá racionální chování, pokud jde o učení se z vlastní zkušenosti, a jako takové pomáhá pochopit, jak se lidé učí ze zkušeností, aby činili dobrá rozhodnutí (obrázek 9).

Obrázek 9: Reinforcement learning



Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar

## 2.3. Neuronové síť

Biologický mechanismus, ve kterém mozek funguje, inspiruje umělé neuronové síť, odlišné paradigma pro výpočetní techniku, založené na paralelní architektuře zvířecích mozků.

Neuronové síť jsou dobře vybaveny pro úlohy dolování dat díky své schopnosti modelovat vícerozměrná data a zvýšit účinnost při hledání skrytých vzorců mezi daty. Neuronové síť lze aplikovat na prediktivní a klasifikační problémy. Proces odhadování výsledku a jeho porovnávání se skutečným výstupem je známý jako dopředné šíření.

Neuronové síť jsou podmnožinou algoritmů strojového učení, které zahrnují perceptrony, plně propojené neuronové síť, konvoluční neuronové síť, rekurentní síť přesvědčení, generativní nepřátelské síť a mnoho dalších. Většina z nich je trénována pomocí algoritmu zvaného zpětné šíření.

Umělé neuronové sítě se používají v celé řadě úkolů:

- klasifikace nemocí – počítače se mohou naučit, jak vypadají obrazy nemocných orgánů, jako jsou ledviny, oči a játra, aby bylo možné předpovědět pravděpodobnost onemocnění;
- rozpoznávání řeči;
- překlad a digitalizace textu;
- rozpoznávání obličeje.

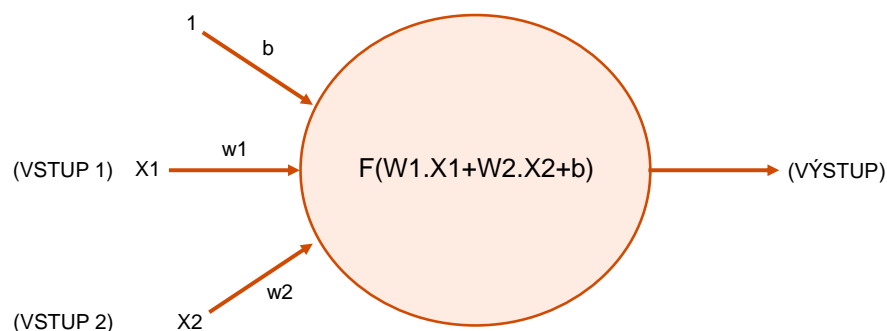
Neuronové sítě se mohou naučit, jak splnit úkol stejně jako to umí lidský mozek – pod dohledem, bez dozoru a prostřednictvím posilujícího učení.

### 2.3.1. Perceptrony

Perceptrony jsou základními jednotkami umělých neuronových sítí. V AI je perceptron uzel nebo jednotka, která přijímá více vstupů a má jeden výstup, obvykle 1 nebo  $-1$ . Perceptrony berou vážený součet vstupů „S“ a jednotková funkce vypočítá výstup pro uzel (obrázek 12). Funkce jednotky mohou být následující:

- lineární – např. vážený součet;
- funkce prahové hodnoty, která se spustí pouze v případě, že vážený součet překročí prahovou hodnotu;
- krokové funkce, které vydávají inverzní hodnotu, obvykle  $-1$ , pokud je S menší než prahová hodnota;
- funkce Sigma ( $1 / (1 + e^{-s})$ ), která umožňuje zpětné šíření.

Obrázek 10: Konfigurace umělého neuronu

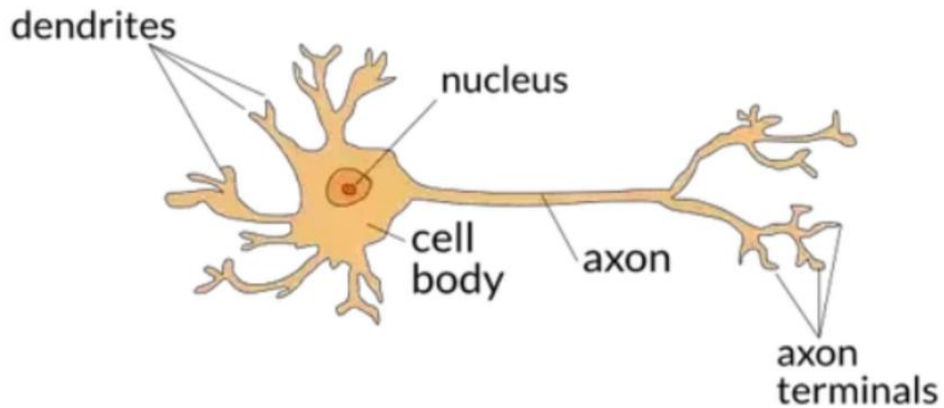


Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar

Perceptrony napodobují biologické neurony (obrázek 11), kde dendrity dostávají vstup. Pokud je signál dostatečně vysoký během nastavené doby, neuron vyše elektrický

impuls do terminálů, který pak obdrží dendrity jiných neuronů. Každý perceptron má odchylku podobnou b lineární funkce  $y = ax + b$ . Posouvá řádek nahoru a dolů, aby předpověď lépe odpovídala datům.

Obrázek 11: Konfigurace biologického neuronu

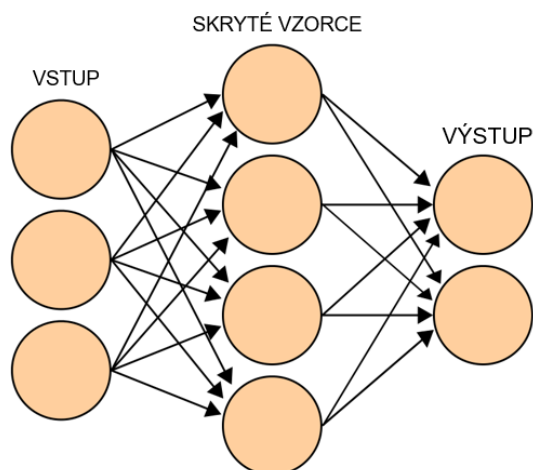


Zdroj: towardsdatascience.com

### 2.3.2. Umělá neuronová síť

Umělé neuronové sítě jsou tvořeny perceptrony a obsahují jednu nebo více skrytých vrstev (obrázek 12). Existuje několik topologií, přičemž nejjednodušší je dopředná síť.

Obrázek 12: Organizace umělé neuronové sítě



Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar



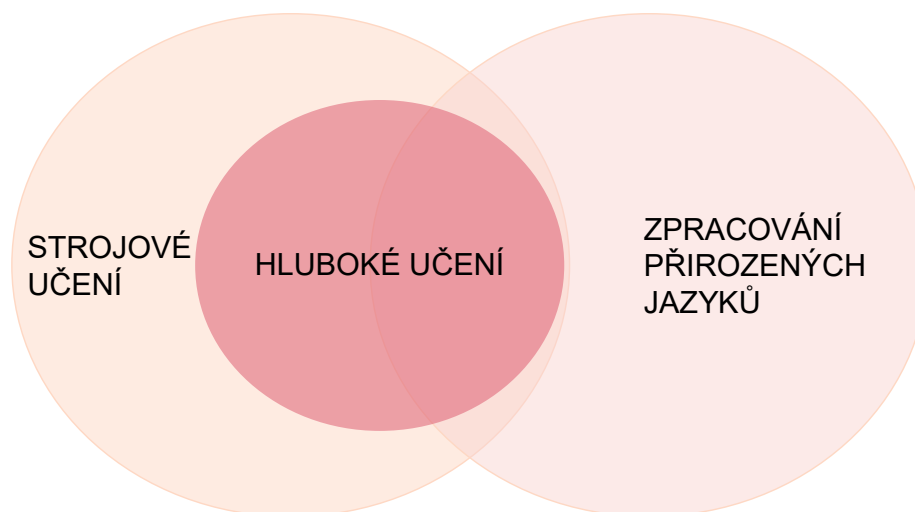
## 2.4. Hluboké učení

Hluboké učení je proces, který využívá hluboko vrstvé architektury neuronových sítí. Trénink modelu hluboké umělé neuronové sítě vyžaduje více času a vyšší spotřebu energie centrálních procesorových jednotek ve srovnání s jinými typy modelů. Je pozoruhodné, že výkon modelů umělých neuronových sítí nemusí být nutně lepší než typické supervised techniky.

Hluboké učení není nová technika. S hardwarovými pokroky, především s výkonem a cenou, se však stává dostupnější, a tudíž se častěji využívá.

Hluboké neuronové sítě použil Alphabet k vytvoření nejsilnějšího hráče Go v historii, když v roce 2017 porazil několik lidských šampionů Go (Holcomb a spol., 2018). Starověká hra Go byla historicky považována za výzvu pro umělou inteligenci a demonstruje schopnosti hlubokého učení (obrázek 13).

Obrázek 13: Hluboké učení



Vlastní zpracování, zdroj: Kniha *Machine Learning and AI for Healthcare*, Arjun Panesar

## 2.5. Dolování dat

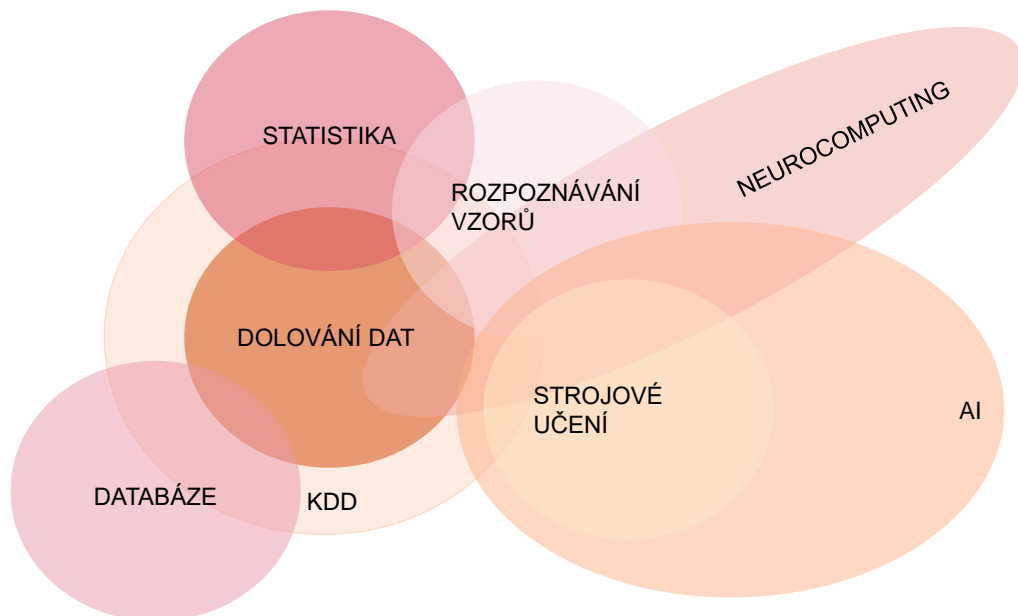
Dolování dat, také známé jako zjišťování znalostí v databázích, je definováno jako proces zkoumání dat za účelem extrahování nových informací (včetně vzorců a vztahů) pomocí sofistikovaných algoritmů. Dolování dat lze aplikovat na každý aspekt života a průmyslu a umožňuje následující:

- předpovědi vzorů založené na trendech a chování;
- předpověď založená na pravděpodobných výsledcích;

- analýza velkých datových sad (zejména nestrukturovaných);
- shlukování prostřednictvím identifikace skutečností dříve neznámých.

Data mining využívá sílu strojového učení, statistiky a databázové techniky a je jádrem analytického úsilí (obrázek 10). Dolování dat obvykle využívá velké množství dat a může také poskytnout jednodušší popisnou analýzu. Například ve zdravotnictví lze aplikace pro dolování dat použít k identifikaci stavů nebo lékařských postupů, které se často vyskytují společně.

Obrázek 14: Dolování dat a jejich synergie



*Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar*

Dolování dat je nejčastěji vidět v každodenní analýze sentimentu uživatelů, zákazníků nebo zpětné vazby od pacientů.

Data mining také umožňuje předpovídat pomocí identifikace vztahů v datech. Pojišťovně to může pomoci odhalit rizikové vzorce chování zákazníků.

Strojové učení lze využít pro dolování dat. Data mining bude informovat o rozhodnutích, která určí nejvhodnější druhy modelů učení pro řešení daného problému.

Technika dolování dat se liší od strojového učení tím, že při dolování dat je hlavním úkolem objevování neznámých znalostí, zatímco výkon strojového učení je hodnocen s ohledem na reprodukci známých znalostí.

Pokud máte například soubor údajů o krevním tlaku pacienta, můžete provést detekci anomálií, což se považuje za dolování dat k identifikaci dříve neznámých hodnot. Úkol

může využívat techniky strojového učení, možná algoritmus k-means pro shlukovou analýzu, k identifikaci anomálních dat a napomáhání algoritmu při učení.

Dolování dat není totéž jako zpracování přirozeného jazyka (NLP). NLP může být technika používaná při dolování dat, která pomáhá agentovi porozumět nebo „číst“ text.

Algoritmy dolování dat zahrnují:

- C4.5,
- k-Means,
- Podpůrné vektorové stroje,
- Apriori,
- EM (očekávání – maximalizace),
- PageRank,
- AdaBoost,
- kNN,
- Naivní Bayes,
- CART (klasifikační a regresní stromy).

## 2.6. Parametrické a neparametrické algoritmy

Algoritmy mohou být parametrické nebo neparametrické. Algoritmy, které lze zjednodušit na známou konečnou formu, jsou označeny jako parametrické, zatímco neparametrické algoritmy se učí funkční formu z trénovacích dat.

Jinými slovy, s neparametrickými algoritmy roste složitost modelu s množstvím tréninkových dat. U parametrických modelů existuje pevná struktura nebo sada parametrů, díky nimž jsou parametrické modely rychlejší než neparametrické modely.

Neparametrické modely mohou poskytovat lepší přesnost, když mají dostatek tréninkových zdrojů, pokud jde o tréninková data a čas (obrázek 11).

Obrázek 15: Parametrické vs. Neparametrické algoritmy

	Výhody	Nevýhody
Parametrické algoritmy	<p><b>JEDNODUŠŠÍ</b> Jednodušší na pochopení a výklad</p> <p><b>RYCHLEJŠÍ</b> Rychlé přizpůsobení datům</p> <p><b>MÉNĚ DAT</b> K dobrému výkonu vyžaduje méně dat</p>	<p><b>OMEZENÁ SLOŽITOST</b> Vhodné pro jednoduché problémy, kde lze odvodit strukturu dat</p>
Neparametrické algoritmy	<p><b>FLEXIBILITA</b> Vejde se do různých funkčních forem, které není třeba předpokládat</p> <p><b>VÝKON</b> Jakmile se datové struktury stanou složitými, výkon bude pravděpodobně vyšší než u parametrických algoritmů</p>	<p><b>POMALU</b> Výpočty jsou výrazně delší</p> <p><b>VÍCE DAT</b> K učení vyžaduje značná data</p> <p><b>OVERFITTING</b> Ovlivňuje výkon modelu</p>

*Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar*

Parametrické algoritmy zjednodušují funkci do známé formy a učí se funkční koeficienty z zpracovaných dat. Proto se o datech vytvářejí silné předpoklady. Parametrické algoritmy jsou užitečné v tom, že se rychle učí z dat a nevyžadují jich mnoho a lze je použít ve scénářích, které vyžadují snadno vysvětlitelné výsledky.

Neparametrické algoritmy používají různý počet parametrů. k-Nearest neighbors, neutrální sítě a rozhodovací stromy jsou neparametrické a hodí se pro scénáře, kde je více dat a vyžadován vyšší výkon.

## 2.7. Jak fungují algoritmy strojového učení

Supervised learning je počítačový úkon odvodit funkci z označených tréninkových dat.

$$Y=f(x),$$

které lze chápat jako

$$\text{Výstup} = \text{funkce}(\text{vstup}).$$

Vstupy a výstupy mohou být označovány jako proměnné a mají typicky vektorový formát: „f“ označuje funkci, kterou se snažíte odvodit ze vstupu.

Vezměte si příklad pacienta, kterému má být předepsána konkrétní léčba. V reálném životě může tento proces zahrnovat pacienta, který se dívá na výzkum popisující vlastnosti léků nebo léčby, které jsou doporučovány.

Pokud by například viděli, že výzkum se skládá ze slov jako „dobrý“, „skvělý“, „bezpečný“ a tak dále, pak bychom z pocitu vyvodili závěr, že léčba je dobrou léčbou a můžeme cítit jistotu v jeho přijetí. Zatímco pokud by se slova jako „špatný“, „nekvalitní“ a „nebezpečný“ objevovala pravidelně, pak bychom usoudili, že je asi lepší poohlédnout se po jiné léčbě. Výzkum nám pomáhá provést akci založenou na vzoru slov, která existují ve výzkumu léčby.

Strojové učení se pokouší porozumět tomuto lidskému rozhodovacímu procesu pomocí algoritmů. Zaměřuje se na vývoj algoritmů (nebo programů), které mohou přistupovat k datům a používat je k učení pro sebe. Strojové učení je velmi dobré pro klasifikační a prediktivní úlohy.

Strojové učení se nejčastěji používá pro prediktivní analytiku – učí se mapování  $Y = f(X)$ , aby bylo možné předpovídat „Y“ pro nové „X“.

Praktické příklady strojového učení jsou následující:

- detekce podvodu „x“ je vlastnost transakce  $f(x)$ , která určuje, zda je transakce podvodná;
- diagnóza onemocnění „x“ jsou vlastnosti pacienta,  $f(x)$  je, zda má pacient onemocnění;
- rozpoznávání řeči „x“ jsou vlastnosti řečového vstupu ( $x$ ), který je odezva daná instrukci obsažené v řečovém vstupu.

Proces strojového učení začíná vstupními daty poskytnutými jako příklady, přímá zkušenost nebo instrukce – k identifikaci vzorců v datech a přijímání lepších rozhodnutí v budoucnu na základě poskytnutých dat.

### 3. Strojového učení v praxi

„Strojové učení může rozšířit náš život, schopnosti a kapacity.“ (Harkrishan Panesar, 2020)

Strojové učení je rychle se vyvíjející a vzrušující disciplína v rámci počítačové vědy, která kombinuje řadu dovedností včetně průzkumu a objevování, inženýrství a analýzy. Strojové učení využívá velké množství faktických historických dat, která nám pomáhají lépe porozumět chování.

I když se to na první pohled může zdát zastrašující, postupný přístup ke strojovému učení a jeho předpokladům lze použít k vedení úspěšného projektu, aniž byste museli být zkušeným statistikem nebo programátorem. Přestože znalost matematiky a programování je jistě výhodou, je k dispozici mnoho algoritmů strojového učení pro vývoj nových algoritmů od nuly.

Přístup k projektu strojového učení je stejně jedinečný jako problém, který se snaží vyřešit. Pracovní postupy zahrnuté v životním cyklu projektu strojového učení nebo datové vědy jsou však podrobně popsány pro použití jako metodologie.

#### Vedení projektu strojového učení

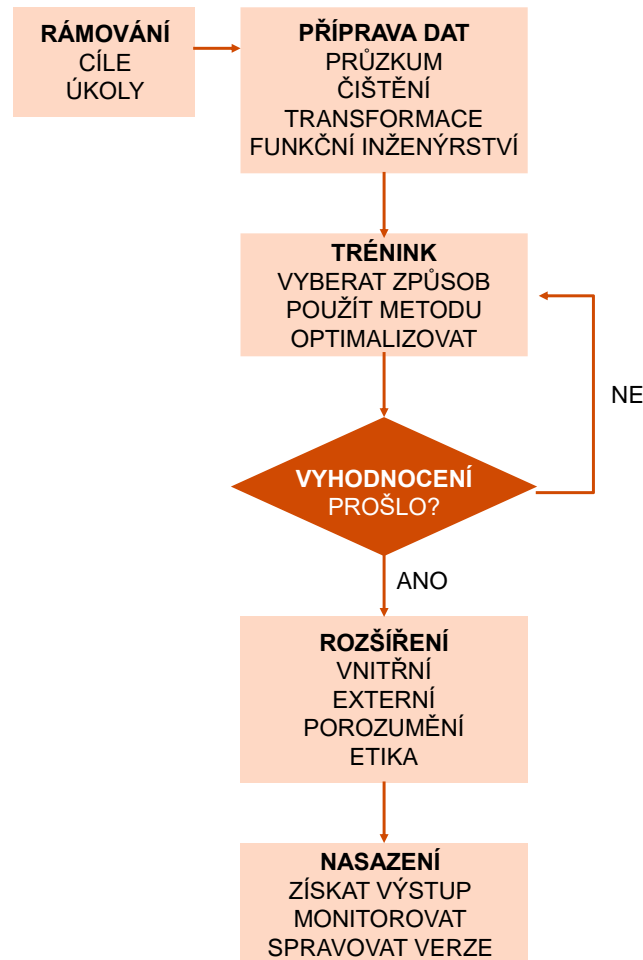
Cílem projektu strojového učení je vyvinout algoritmy, které se učí ze signálů v datech, na kterých jsou trénovány.

V metodice používané k vývoji modelů strojového učení je sedm definovaných pracovních postupů (obrázek 16). Každá z těchto fází může být prováděna iterativně, což znamená, že pracovní postupy lze opakovat během jakéhokoli kroku procesu.

- **Rámování:** Určení problému jako výukového úkolu (a identifikace, zda se jedná o problém strojového učení, než budete pokračovat).
- **Příprava dat:** Zahrnuje průzkum dat, analýzu, náhled a čištění.
- **Tréninkový model:** Výběr výukových metod, jejich aplikace k vytvoření modelu a pokus o optimalizaci vytvořených modelů.
- **Hodnocení:** Objektívni posouzení metody a výsledků.
- **Rozšíření:** Podávání zpráv o výsledcích hodnocení zainteresovaným stranám a zajištění vysvětlitelnosti.

- Nasazení: Uvolnění natrénovaného modelu strojového učení, sledování průběžného používání a přesnosti

Obrázek 16: Pracovní postupy spojené s projektem strojového učení



Vlastní zpracování, zdroj: Kniha Machine Learning and AI for Healthcare, Arjun Panesar

### 3.1. Rámování: Specifikace problému

První věcí, pokud jde o strojové učení, je specifikace problému. To zahrnuje pochopení, v čem je problém, cíle projektu strojového učení a jak lze vyhodnotit pokusy o dosažení cílů. Je užitečné pochopit, proč chcete problém vyřešit, zvláště pokud vytváříte rozpočtový případ pro projekt strojového učení. To zahrnuje hodnocení následujících věcí:

- Proč je to důležité?
- Čeho chceme dosáhnout? Jaké jsou naše největší naděje?
- Jaké jsou vstupy a výstupy pro daný úkol?
- Jsou potřebná data k dispozici?

- Jaký přínos budou mít výsledky?
- Je to proveditelné?
- Jaké jsou klíčové ukazatele výkonu (KPI)? Jak budeme výkon měřit?
- Jak bude vypadat úspěch?
- Máme dostatek talentu pokusit se o tento úkol?
- Jaká jsou další omezení daného úkolu? To může zahrnovat čas, finance, dovednosti, zkušenosti, znalosti domény, dostupnost dat atd.

### 3.1.1. Sbírání příkladů

Kategorizace je dovednost, kterou se lze naučit na příkladech. Může být použita pro prediktivní úkoly, jako je riziko onemocnění nebo predikce poptávky po zdrojích.

Některé problémy se vyskytují v tradičních zdravotnických zařízeních, jakými jsou údaje z lékařských schůzek a určité diagnózy s doplňkovými ručně psanými poznámkami. Špatné příklady mohou zavést chybu. Pro klasifikaci jsou příklady obvykle poskytovány ve formě pozitiv a negativ.

### 3.1.2. Základní informace

Základní informace se týkají znalostí a axiomů spojených s problémem strojového učení. To může zahrnovat metadata, atributy nebo vztahy mezi pojmy.

Například v nástroji pro predikci hypertenze by diagnóza diabetu 2. typu souvisela s pravděpodobností hypertenze na základě lékařských důkazů a údajů. Toto jsou informace, které se používají v naučených konceptech k dosažení kategorizací.

### 3.1.3. Chyby v datech

Chyby se mohou stát například z lidského zapříčinění, nesprávné klasifikace, chybějících dat, nesprávných informací na pozadí a opakovaných dat.

Chyby mohou nastat například z digitalizace papírových dokumentů a strojového čtení psaného textu. Je důležité mít na paměti, že model je jen tak dobrý, jak kvalitní jsou data, která poskytuje. Stojí za to věnovat čas ověřování dat, aby bylo zajištěno, že lidské chyby jsou minimalizovány.



## Příprava dat

Algoritmy strojového učení se učí z dat, na kterých jsou trénovány. Je velmi důležité poskytnout modelu platná a robustní data, na kterých se může model učit.

Data musí být připravena v použitelném formátu. V reálném světě by to zahrnovalo pochopení dat, která by byla použita k modelování problému, export dat o chybách a opravu jakýchkoli chybějících dat. Velikost datové sady může být větší, než je požadováno, a proto může být také vyžadováno vzorkování datové sady.

Každý kandidát na strojové učení musí získat zkušenosti a dovednosti v přípravě dat. Předzpracování dat je nezbytné, aby byla data vyčištěná a platná. Správná a platná data jsou klíčem k získání spolehlivých a pravdivých výsledků. Aby bylo možné vyvinout model strojového učení, který je vysoce použitelný v dané oblasti, techničtí odborníci, kteří efektivně implementují algoritmy, musí spolupracovat s odborníky na domény, kteří dokážou porozumět datům, klasifikovat je a zjistit trendy a vzorce. Příprava dat je zdaleka nejdůležitějším aspektem strojového učení.

## Kolik dat je potřeba?

Na tuto otázku bohužel neexistuje definitivní odpověď. To je něco, na co se přijde prostřednictvím školení algoritmu a jeho nasazení v reálném světě.

„Křivku učení“ může ovlivnit několik faktorů. Ty zahrnují složitost problému a složitost algoritmu. To zase ovlivňuje množství požadovaných dat. Zda model dokáže vytvářet výstupy, které jsou pravdivé v reálném světě, bude záviset na propracovanosti algoritmu a kvalitě poskytovaných dat.

Získat přístup k většímu množství dostupných a užitečných dat je složité především ve zdravotnickém prostředí. Randomizované kontrolované studie léků zřídka hodnotí více než 1000 pacientů v jakémkoli prostředí. V prostředí chirurgie je pravděpodobný přístup k větším datům, ale nemusí být vyžadován pro trénování modelu.

Nelineární, neparametrické algoritmy vyžadují více dat, aby se zlepšila jejich přesnost. Jedná se obvykle o výkonnější metody strojového učení. Vzhledem k platnosti dat ovlivňujících vychýlení i rozptyl je však klíčové trénovat a validovat model robustně na co největším množství dat, jak je to možné.

## 3.2. Trénování modelu strojového učení

Standardní otázka každého nováčka ve strojovém učení je „Který algoritmus bych měl použít?“ Volba algoritmu závisí na několika faktorech, včetně velikosti, kvality a povahy dat, termínu splnění úkolu, dostupných zdrojů a motivace pro použití dat. Technika učení se také označuje jako „reprezentace řešení“, protože každý přístup strojového učení představuje data jinak. Samotné algoritmy jsou běžně dostupné a nepotřebují kódování od nuly.

Díky tomu se některé algoritmy hodí k určitým problémům oproti jiným, což může pomoci při rozhodování, které algoritmy vyzkoušet:

- Klasifikace: Logistická regrese, podpůrné vektorové stroje, náhodný les, naivní Bayes;
- Regrese: Lineární regrese;
- Funkce redukce: PCA, LDA;
- Shlukování: k-Means, LDA;
- Kolaborativní filtrování: Střídání nejmenších čtverců.

Je téměř nemožné předpovědět, který přístup bude s daty fungovat nejlépe. A i když model funguje během trénování a ověřování dobře, není to žádná záruka, že bude dobře fungovat v reálné situaci.

Teorém „No Free Lunch“ ve strojovém učení říká, že neexistuje žádný algoritmus, který by fungoval nejlépe pro každý problém. To je zvláště důležité pro Supervised learning, protože existuje mnoho faktorů, které ovlivňují výkon algoritmu (Panesar, 2021).

### 3.2.1. Programovací jazyky

Úlohy strojového učení se obvykle provádějí v různých programovacích jazycích: převážně R, Python, Matlab a SQL. Java a C se také běžně používají.

- R se typicky používá pro statistickou analýzu. Umožňuje porozumět a prozkoumat data pomocí statistických metod a grafů, a obsahuje širokou škálu algoritmů strojového učení.
- Python je jazyk vhodný pro strojové učení. Rozšíření jako Numpy a SciPy jsou zvláště užitečná pro strojové učení a analýzu dat.

- Matlab je jazyk, se kterým začínají mnozí vysokoškoláci. Je užitečný pro rychlé prototypování, protože obsahuje velký depozitář strojového učení.
- SQL je jazyk používaný pro správu dat uložených v tradičním systému správy databází.

Je třeba učinit několik rozhodnutí z hlediska zdrojů a hodnot, která ovlivňují přístup. Projekt „proof of concept“ může být krátký, zatímco vysoce specifický a komplexní úkol bude obsahovat směs přístupů.

Je zcela možné spustit algoritmus strojového učení pomocí implementace hotových algoritmů poskytovaných výukovými knihovnamí, jako je Scikit-learn, SciPy, Pandas, Matplotlib, TensorFlow, Keras atd.

Tradiční přístupem k sestavování algoritmu je: nejprve sestavit data science tým, který obsahuje členy se zkušenostmi s jazyky jako jsou Python a R. Členové se zkušenostmi se strojovým učením budou mít obvykle větší dopad než ti, kteří s koncepty strojového učení začínají.

Inženýři si na základě zkušeností rozvíjejí znalosti o vhodných technikách strojového učení, které lze použít při daných problémech. Tato zkušenost se může ukázat jako zásadní pro zamezení plýtvání časem, zdroji a průzkumem.

Inženýři se základy informatiky budou obvykle obeznámeni s mnoha koncepty přítomnými ve strojovém učení. Výběr programovacího jazyka může ovlivnit aplikační programové rozhraní a standardní knihovny, které můžete použít ve své implementaci.

Cloudové služby, jako je Google AI, Microsoft Azure, IBM Watson a řada dalších poskytovatelů, umožňují poskytovateli postarat se o algoritmus a mnozí poskytují vizuální rozhraní, která uživatelé umožňují vyzkoušet konkrétní techniky strojového učení na nahraných datech.

Pro ty, kteří mají tradičnější zkušenosti s programováním, téměř všechny nabízejí rozhraní pro programování aplikací nebo zařízení pro vytváření webových služeb, což umožňuje vytvořit rozhraní pro zapojení do hostovaného modelu. Vše, co je potřeba k integraci s aplikačním programovacím rozhraním je efektivní tradiční programátor nebo webový vývojář.

Další zajímavou variantou je použití open source, procházení implementací, které již existují prostřednictvím GitHubu, Redditu a dalších fór strojového učení.

### 3.2.2. Trénování a testování dat

Z připravených dat se vybere testovací a tréninková sada. Algoritmus je natrénován na trénovací datové sadě a vyhodnocen oproti testovací datové sadě.

V mnoha případech se jedná o vyzkoušení několika metod strojového učení. Je užitečné porozumět dvěma termínům, pokud jde o modelování strojového učení:

- **Signál:** Skutečný základní vzor v datové sadě;
- **Šum:** Náhodné nebo irelevantní vzory v datové sadě.

Některé techniky strojového učení mohou vrátit jedno řešení, zatímco jiné mohou vytvořit několik. Často jde o shromažďování jejich výstupů (označovaných také jako hypotézy nebo naučené modely) a vyhodnocování jejich výstupů.

Posouzení hypotéz se provádí prostřednictvím hodnocení prediktivní přesnosti, srozumitelnosti nebo užitečnosti. Ve většině případů se používá Occamova břitva, kde je zvoleno nejjednodušší řešení, pokud jsou všechny ostatní stejné. Po dokončení hodnocení je vybrána finální hypotéza.

Existují různé způsoby, jak rozdělit data do trénovacích a testovacích datových sad pro trénování modelu strojového učení:

- **Prediktivní přesnost:** odkazuje na přesnost, s jakou agent plní úkol klasifikace.
- **Srozumitelnost:** označuje, jak dobře dokážeme porozumět výstupu.
- **Užitečnost:** odkazuje na míru hodnoty konkrétního problému. Například při syntéze léků může být nejlepší kombinací kombinace sloučenin, které nejsou bezpečné pro lidskou spotřebu.
- **Metoda zadržování:** jednoduchá trénovací metoda, kterou lze použít, která zahrnuje uchování části dat k testování, zbývající data ponechává pro trénování a ověření modelu.
- **n-Fold Cross validace:** zahrnuje rozdělení datové sady do stejně velkých skupin instancí. Model je poté trénován na všech datových sadách kromě jednoho a testován na poslední vynechané datové sadě. Tento proces se opakuje a každý záhyb je vynechán pro jednu iteraci.
- **Cross validace Monte Carlo:** podobná n-fold cross validaci. Zahrnuje náhodné rozdělení datové sady na trénovací a testovací data. Model je vyladěn

na trénovací data, a prediktivní přesnost je vyhodnocena pomocí testovacích dat. Výsledky jsou pak zprůměrovány.

### 3.3. Vyhodnocení a optimalizace metod a výsledků

Výkon úloh strojového učení se liší podle reprezentace daných dat. Záznam pacienta analyzovaný systémem umělé inteligence například nevyšetřuje pacienta přímo. Místo toho jsou data vkládána do systému, přičemž každá informace týkající se reprezentace pacienta je známá jako vlastnost. Pro vysoce spolehlivé výstupy není nutné vyžadovat kompletní sady funkcí jako součást reprezentací. Generalizace se týká modelu, který pracuje s maximální přesností na instancích, které nebyly během tréninku nasazeny.

#### 3.3.1. Vyhodnocení přesnosti algoritmu

Mnoho problémů se Supervised learningem má binární klasifikace, kde problémem je naučit se klasifikování neviditelných dat do jedné ze dvou kategorií. Tyto jsou známé jako pozitivní a negativní kategorie:

- Falešně pozitivum je přijato, pokud výstup z binárního klasifikačního agenta, kterému jsou poskytnuta nová data, měl být kategorizován jako negativní, ale byl klasifikován jako pozitivní.
- Falešně negativum je, když agent kategorizuje nová data jako negativní, ale jsou nesprávná.

V mnoha lékařských zařízeních není falešně pozitivum výsledek tak špatný jako falešně negativum. Například falešné pozitivum v tomto případě může zahrnovat diagnózu onemocnění pro pacienta na základě souboru konkrétních příznaků. Je evidentně lepší, když vám někdo řekne, že máte nějakou diagnózu, a ve skutečnosti ji nemáte, než aby vám řekli, že žádnou diagnózu nemáte, a pak vám řekli, že ve skutečnosti ano. Falešná negativa by znamenala, že někdo nebyl diagnostikován, což je možná více znepokojující.

Jednoduchým způsobem měření prediktivní přesnosti konkrétní hypotézy na testovacím souboru dat je výpočet počtu správně identifikovaných výsledků (pozitivních i negativních).

Pokud se například nástroj pro onemocnění srdce naučil hypotézu a dostal 250 nových kandidátů, aby je klasifikoval, zda mají onemocnění srdce a správně klasifikoval 133 ze 150 pozitivních a 98 ze 100 negativních, měl by přesnost:

$$(133 + 98) / 250 = 92,4 \%$$

To by byla šance 92 ze 100 na správné zařazení neviditelného příkladu.

Zobecnění se měří přesností. Pokud by model klasifikoval s 92% přesností při testování pouze s 30% na neviditelných datech, řekli bychom, že model dobře nezobecňuje z tréninku na reálné data.

Pojem „více dat dává lepší výsledky“ je zavádějící. V mnoha případech může pomoci více dat. Více údajů však neznamená dobré údaje. Pokud by se přidala data se šumem, nepomohlo by to k lepší přesnosti.

K úniku dat dochází, když data použitá k trénování modelu strojového učení obsahují informace, která se snažíte předvídat. To vede ke špatné generalizaci. K úniku dat může dojít v důsledku úniku testovacích dat do tréninkových dat nebo dalších funkcí. Triviálním příkladem úniku dat by byl model, který jako prediktor využívá samotnou proměnnou odezvy, čímž by například došlo k závěru, že „pacient se zánětem spojivek má zánět spojivek“.

Pokud je výkon algoritmu strojového učení příliš dobrý na to, aby to byla pravda, důvodem může být únik dat. N-fold cross validace může pomoci snížit únik dat.

### 3.3.2. Měření výkonnosti

Příklady metod pro vyhodnocování algoritmů zahrnují přesnost, předpověď a vyvolání, druhou mocninu chyby, pravděpodobnost, zpětnou pravděpodobnost, divergenci nákladů a entropie K-L.

### 3.3.3. Optimalizace

To, že model funguje dobře, ještě neznamená, že je nejlepší a jediný použitelný model. Klíčem je získat z výsledků maximální pravdivost. Techniky, jako je křížová validace, jsou užitečné pro stanovení spolehlivosti výsledků modelu. Algoritmy strojového učení však můžeme dále optimalizovat.

Kromě toho můžeme chtít zlepšit výkon, pokud jde o dobu do výstupu. Algoritmy lze optimalizovat pomocí následujících metod:

- **Ladění algoritmu:** lze chápat jako proces optimalizace parametrů, které ovlivňují model, aby algoritmus mohl vykonávat „nejlepší“ na základě toho, co je „nejlepší“ specifikováno.

- **Trénování a ověřování dat:** přiblížení se k problému strojového učení pomocí techniky n-fold cross validace nám umožní ověřit výsledky, zatímco metodologie zadržování má k učení vodopádový přístup.
- **Vyhodnocení řady metod:** vyzkoušení různých metod strojového učení je užitečné k určení, které mohou dosáhnout nejpřesnějších výsledků.

### 3.3.4. Zlepšování výsledků s lepšími daty

Existuje několik způsobů, jak zlepšit výkon pomocí dat, například:

- **Získání více dat:** může pomoci modelu zlepšit jeho výkon.
- **Získávání kvalitnějších dat:** spíše, než více dat je často lepší, pokud je to možné, získat kvalitnější data. Typicky by to zlepšilo signál v datech a snížilo šum.
- **Přerozdělení dat:** rozdělení ukázkových dat do různých velikostí nebo distribucí může lépe reprezentovat koncept nebo může pomoci zlepšit výkon omezením atributů (funkcí).
- **Reprezentace dat:** data mohou být reprezentována různými způsoby prostřednictvím různých metod strojového učení. Změnou typu použité metody strojového učení a přehodnocením problému lze optimalizovat model.
- **Výběr funkcí:** které jsou a nejsou důležité, může představovat nové závěry, které se model může naučit. Zvýšení počtu funkcí nemusí být vždy výhodné. V případě Unsupervised learning však může být výhodné zahrnout všechny funkce, aby se zajistilo, že nic nezůstane neprozkoumané.

## 3.4. Šíření výsledků

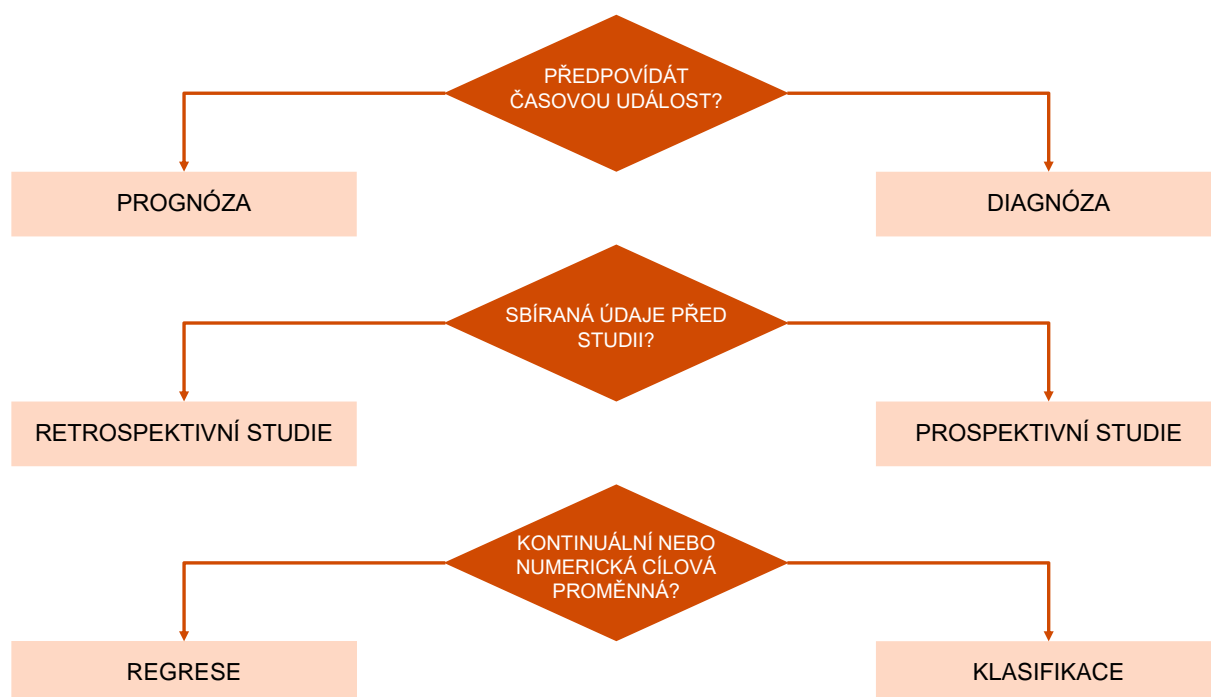
V mnoha prostředích je šíření výsledků nejdůležitějším aspektem pro zajištění podpory projektů strojového učení ze strany interních zúčastněných stran. Zda byl projekt strojového učení úspěšný nebo ne, je nepodstatné, pokud lidé nepochopí důvody událostí, může dojít ke ztrátě podpory pro přijetí inovativních technologií.

Existuje mnoho způsobů, jak hlásit výsledky a různé stupně hodnocení. Vzhledem k tomu, že zdravotní péče rychle implementuje big data a strojové učení, jsou zásadní důkazy k prokázání výhod a dopadu umělé inteligence, aby bylo zajištěno, že budou používány a postaveny pouze klinicky ověřené rámce a technologie založené na důkazech.

Nové zdroje informací, jako jsou nositelná zařízení, trackery, umísťují strojové učení do prostoru inovací a etického úžasu. Důsledky pro širší zdravotní péči a pojištění jsou obrovské a měly by být brány v úvahu u výzkumných projektů strojového učení.

V oblasti prediktivních modelů strojového učení došlo k pokroku, zejména v biomedicíně, kde byly zveřejněny pokyny pro podávání zpráv o modelech, včetně kroků pro identifikaci problémů s predikcí, zajištění n-fold cross validace a kalibrace výsledků (obrázek 17).

Obrázek 17: Porozumění úkolům ve strojovém učení



Vlastní zpracování, zdroj: Kniha *Machine Learning and AI for Healthcare*, Arjun Panesar

Kalibrace má význam v klinické praxi a týká se vztahu mezi předpověďmi provedenými modelem a pozorovanými výsledky.

Například, pokud predikční model „předpovídá 70% riziko úmrtnosti v příštím roce u pacienta s rakovinou plic, pak je model dobře kalibrován, pokud v souboru dat přibližně 70 % pacientů s rakovinou plic zemře během příštího roku.“

### 3.5. Deployment

Deployment model strojového učení je proces přesunu modelu z testovacího nebo přípravného prostředí do produkčního prostředí. K modelu pak mohou přistupovat ostatní a poskytovat výstup dalším agentům, ať už člověku nebo počítači.



Modely strojového učení poskytují hodnotu pouze tehdy, jsou-li zpřístupněny, takže nasazení je konečnou, klíčovou fází životního cyklu projektu strojového učení.

Deployment by měl také obsahovat mechanismus zaznamenávání všech pozorování a předpovědí a proces, při kterém je tento rostoucí záznam monitorován a pravidelně vyhodnocován, aby byla zajištěna relevance a nedostatek škod. Většina nasazených projektů bývá hostována na cloudových serverech, takže auditování bývá relativně jednoduché. Další výhodou je, že cloudové servery podporují škálovatelnost a zabezpečení.

Měl by být implementován proces monitorování a řízení používání nasazeného modelu a třídění dat.

## 4. Etika umělé inteligence a strojového učení

### 4.1. Etika

Etika nebo morální filozofie odkazuje na morální kodexy chování (nebo soubor morálních zásad), které utvářejí rozhodnutí lidí a jejich chování. Morálka označuje principy, které rozlišují mezi dobrým/správným a špatným/nesprávným chováním.

Etika na pracovišti je například často přenášena prostřednictvím profesionálních kodexů chování, které musí zaměstnanci dodržovat.

### 4.2. Etika datové vědy

Etika datové vědy je odvětví etiky, které se zabývá ochranou soukromí, rozhodováním a sdílením dat.

Etika datové vědy zahrnuje tři hlavní prvky:

#### **Etika dat**

- Tato oblast etiky datové vědy se zaměřuje na generování, shromažďování, používání, vlastnictví, zabezpečení a přenos dat.

#### **Etika inteligence**

- Tato oblast etiky datové vědy pokrývá výstupy nebo výsledky prediktivní analýzy, k jejichž vývoji se data používají.

#### **Etika praktik**

- Etika praktik byla navržena Floridi a Taddeo s odkazem na morálku inovací a systémů, které mají vést k nově vznikajícím obavám (Taddeo a spol., 2016).

#### 4.2.1. Etika dat

Na světě existuje více chytrých telefonů než lidí – a telefony, tablety a digitální zařízení spolu s aplikacemi, nositelnými zařízeními a senzory vytvářejí miliony datových bodů denně. Na světě se používá více než 7,2 miliardy telefonů, ročně se prodá 112 milionů nositelných zařízení a více než 100 000 zdravotnických aplikací dostupných ke stažení do vašeho mobilního telefonu (Banerjee a spol., 2018). Data jsou všude. Navíc jsou velmi cenná.

Téma datové etiky se dostalo do popředí pozornosti veřejnosti prostřednictvím významných fiasků, jako je skandál Facebook-Cambridge Analytica. Facebook, jedna z největších a nejdůvěryhodnějších organizací pro shromažďování dat na světě, nechal shromáždit uživatelská data prostřednictvím kvízu hostovaného na své platformě.

Behaviorální a demografická data 1,5 milionů lidí, kteří dokončili kvíz, byla prodána společnosti Cambridge Analytica. Z velké části se má za to, že údaje byly použity k zacílení a ovlivnění výsledku voleb v USA v roce 2017 (Cadwalladr a spol., 2018). Ještě více znepokojivé je, že toto narušení bezpečnosti bylo hlášeno více než 2 roky po prvotním úniku dat.

Nacházíme se v době, kdy se falešné zprávy mohou šířit rychleji než pravda. Společnost se nachází v kritickém bodě svého vývoje, kde je třeba kolektivně řešit používání, přijímání a spoléhání se na data, aby se rozvinuly rozhovory a hlavní zásady, jak s daty zacházet eticky. Lidé jsou si vědomi falešných zpráv a dezinformací. Umělá inteligence musí být založena na etice a musí být vysvětlitelná.

### **4.3. Správci dat a zpracovatelé**

Obavy o ochranu osobních údajů vyvolaly globální reakci. V květnu 2018 vstoupilo v celé Evropě v platnost GDPR (neboli obecné nařízení o ochraně osobních údajů). Legislativa GDPR řídí organizace v tom, jak používají a sdílejí uživatelská data (Voigt a spol., 2017).

Kvůli GDPR byly organizace nuceny shromažďovat souhlas se svým členstvím a prohlašovat uživatelům, jak budou používat uživatelská data a s kým by je měli sdílet. Legislativa GDPR vkládá kontrolu dat pevně do rukou uživatele.

Uživatelé systémů založených na údajích nyní mohou uplatnit svá práva na zobrazení údajů, které o nich jsou, a s kým jsou jejich údaje sdíleny a proč, právo být zapomenut a právo na výmaz. GDPR také definovalo štítky pro ty, kteří shromažďují a zpracovávají data.

#### **Správci dat**

Správce údajů se rozumí osoba nebo organizace, která řídí, ukládá a používá údaje.

## **Zpracovatelé dat**

Zpracovatelem údajů se rozumí osoba nebo organizace, která zpracovává údaje jménem správce údajů.

Pro ty, kdo pracují s daty, je užitečné porozumět rozdílu mezi zpracovatelem dat a správcem a povinnostem každé strany. Tyto odpovědnosti musí být identifikovány a jsou definovány v rámci GDPR.

GDPR ukazuje, jak zásadní se bezpečnost dat stala. GDPR zpřísňuje přístup k datům, zabezpečení a správu v ambiciózním pokusu chránit 500 milionů občanů EU. Maximální pokuta pro největší provinilce proti nařízení GDPR je stanovena buď na 20 milionů eur, nebo na 4 % celosvětových příjmů pro větší organizace.

## **4.4. Politika, právo a regulace**

Příští desetiletí bude klíčovým styčným bodem v etice umělé inteligence. Regulace ve formě zákonů, etických směrnic a průmyslových zásad ovlivní směr strojového učení. Stejně tak je ovlivněno strojové učení, pokud takové politiky neexistují a kde se má za to, že velká část rizika leží. Tím, že se nám nepodařilo zatím stanovit hranici přijatelnosti, otevíráme riziko širších nezamýšlených důsledků.

Je zapotřebí technické řízení, zejména pro aplikaci umělé inteligence. Je také nezbytné zapojit klíčové vůdce veřejného mínění, zúčastněné strany, zákonodárce a kritiky, aby vytvořili a pěstovali tyto zásady, aby bylo zajištěno, že potenciál strojového učení bude realizován bez zbytečných omezení. Aplikace umělé inteligence v nezdravotnických sektorech ukázaly, že existují potenciální etické problémy s algoritmickým učením, když je implementováno ve velkém měřítku.

Zákonodárci a kritické zúčastněné strany v procesu právního rozhodování musí porozumět složitosti umělé inteligence a přijmout výzvu, která přináší první problémy. Dále nelze předpokládat, že umělá inteligence bude využívána výhradně pro benevolentní účely. Umělá inteligence se již používá ve válce, přičemž Spojené státy a Čína vynakládají značné prostředky do vojenského využití (Lucas a spol., 2018). Spojené království se rozhodlo naopak zaměřit na etiku umělé inteligence (Winfield, 2016).

Zdravotníci používající umělou inteligenci musí rozumět tomu, jak jsou nasazené algoritmy umělé inteligence vytvářeny, vyhodnocovat data používaná pro modelování a rozumět tomu, jak může agent zajistit chyby. Regulátoři musí být připraveni řešit každý

problém individuálně, protože veškerá umělá inteligence je individuální, podobně jako u lidí.

Bezpečný a inkluzivní vývoj umělé inteligence bude dosažen pouze prostřednictvím různorodého, inkluzivního a multidisciplinárního týmu inženýrů, humanitních vědců a sociálních vědců.

Mezinárodní angažovanost mezi organizacemi a vládami na kritických tématech nastolených umělou inteligencí je nutností. Otázky etiky, odpovědnosti, zaměstnanosti, bezpečnosti, soukromého sektoru a akademického sektoru ke spolupráci při začleňování umělé inteligence do společnosti a zajištění demokratizované prosperity lidstva.

#### **4.5. Správa dat a informací**

Implementace umělé inteligence ve zdravotnictví vyžaduje řešení etických výzev, jako je potenciál pro neetické nebo podvodné algoritmy, algoritmy trénované s neúplnými nebo zkreslenými daty, nedostatečné pochopení omezení nebo rozsahu algoritmů a vliv umělé inteligence na základní vztah mezi lékaři a pacienty.

Zásady správy dat jsou zdokumentovaným souborem pokynů pro zajištění správné správy dat organizace – digitálních nebo jiných. Tyto pokyny mohou zahrnovat zásady pro řízení obchodních procesů (BPM) a plánování podnikových rizik (ERP), stejně jako zabezpečení, kvalitu dat a soukromí.

Etický kodex chování nebo politika etického řízení by pokrývaly směr a záměr organizace a řídily by řízení.

#### **4.6. Etický kodex**

Etický kodex je dokument, který se používá k řízení morálního chování organizace. Etický kodex prokazuje, že se organizace zavázala k odpovědnému podnikání a technologickému pokroku.

Etický kodex popisuje chování, které organizace podporuje, a chování, které je považováno za škodlivé pro vlastní morální kompas, pověst nebo klienty organizace. Nemusí se vztahovat na činnosti, které jsou nezákonné, ale obvykle uvádí důsledky nedodržení a způsob, jakým lze takové porušení nahlásit. Zaměstnanci by měli být upozorněni na položky, které jim nejsou zřejmé, a poté by jim mělo být umožněno vyhnout se neúmyslným, ale potenciálně škodlivým akcím. Etický kodex by měl také

obsahovat souhrn motivací pro používání dat a jejich účel v ambicích organizace – a odrážet ziskovost, integritu a pověst podniku.

Etický kodex by neměl obsahovat technický a filozofický žargon a měl by přímo sdělovat očekávání od nových zaměstnanců v době nástupu a rozvíjet organizační kulturu dodržování. Je důležité zajistit, aby všechny zúčastněné strany byly informovány o změnách a přezkumech. Při zkoumání etických rizik začlenění nových technologií do rozhodovacího procesu je důležité se řídit etickým kodexem.

Při vytváření kodexu etického chování organizace od nuly se doporučuje konzultovat příspěvky zaměstnanců a zainteresovaných stran. Otázky, které je třeba zvážit, zahrnují následující:

- Co znamená umělá inteligence pro organizaci nebo člověka?
- Jak může určitá organizace zlepšit lidstvo?
- Jak by se měla organizace chovat odpovědně ve svých ambicích...?
- Jaké jsou potenciální výhody toho, čeho by organizace chtěla dosáhnout?
- Jaké jsou potenciální nevýhody toho, čeho by organizace chtěla dosáhnout?
- Jak může organizace zlepšit etický kodex?
- Jsou v etickém kodexu položky, které jsou matoucí nebo vyžadují více vysvětlení?
- Je etický kodex užitečný při rozhodování?
- Je etika organizace v souladu s vlastními etickými názory zaměstnanců?

Jakmile jsou všechny výše uvedené otázky zodpovězeny různými segmenty organizace, lze dosáhnout konsensu o plánu implementace. Organizační etický kodex slouží jako referenční bod pro disciplinární řízení pro ty, kteří nedodržují normy. Dále etický kodex poskytuje pevný základ pro identifikaci a řešení etických výzev.

#### **4.7. Etika umělé inteligence ve zdravotnictví**

Smysluplné změny ve zdravotnictví pocházejí ze správné směsi inovací, etiky a úvah. S tím, jak se testují hranice objevů, dochází k realizaci potenciálu poskytovat pacientům přesnou medicínu a péči založenou na důkazech, a s tím je spojena povinnost zkoumat a předvídat sociální, ekonomické, politické a etické důsledky inovací.

Zavádění velkých dat a umělé inteligence ve zdravotnictví je nejlépe řídit jako postupný vývoj, což by bylo užitečné pro zúčastněné strany, aby měly čas zmapovat a

pochopit možné důsledky její implementace. Pro auditování a porozumění modelům umělé inteligence pro zajištění kvality a důvěry jsou vyžadovány robustní standardy.

Vývoj standardů umělé inteligence vyžaduje různorodé a veřejné fórum, které je akademicky robustní a pravidelně revidované v souladu s nejnovější důkazní základnou.

Je vyžadována informovanost v rámci lékařských odborníků a komunit pacientů. Zdravotníci potřebují základní znalosti o tom, jak aplikace umělé inteligence fungují v jejich příslušném lékařském prostředí, aby pochopili výhody, nevýhody a důsledky toho, jak jim umělá inteligence může usnadnit a pomoci v jejich každodenních úkolech.

Pacienti si musí zvyknout na zapojení do systémů umělé inteligence a prozkoumat jejich přínosy pro sebe a COVID-19 tento proces rozhodně urychlil. Prokazatelné příklady toho, jak umělá inteligence ve zdravotnictví optimalizuje efektivitu, optimalizuje zdroje a zlepšuje výsledky zdraví populace, jen posilují veřejné vnímání a přijímání umělé inteligence.

Organizace pracující s umělou inteligencí potřebují komunikovat a zapojovat širokou veřejnost do diskusí o potenciálních výhodách a rizicích používání umělé inteligence v medicíně. Organizace vyvíjející umělou inteligenci musí spolupracovat s akademickou obcí, aby podnikly nezbytné kroky, aby mohly nezávisle měřit úspěšnost a efektivitu systémů umělé inteligence. Je životně důležité podporovat cenově dostupná řešení umělé inteligence, aby byla umělá inteligence demokratizována jako stetoskop 21. století, a zveřejněné ověřené důkazy, pokyny a literatura jsou zásadní pro trvalou důvěru, přijetí a morálku umělé inteligence.

# **PRAKTICKÁ ČÁST**



## **5. Aplikace umělé inteligence v lékařské vědě a vývoji léků**

„Poučte se ze svých neúspěchů nebo se poučte z neúspěchů jiných lidí. Cokoli si vyberete, učte se.“ (Kirpa Gayatri Kaur, 2020)

Nové pokroky v medicíně a koncepty, které byly kdysi futuristickými tématy sci-fi, se postupně stávají realitou. Genové terapie a 3D tisk lidských orgánů jsou s přibývajícím časem stále sofistikovanější.

Pokrok v technologii ovlivňuje nejen praktikovanou medicínu, ale také vnímání a postoje veřejnosti ke zdraví, životnímu stylu a tomu, co to znamená být zdravý. Zdravotnictví musí zavádět inovace moudře, aby zapojilo co nejvíce pacientů.

Umělá inteligence zlepšuje zdravotní péči a přináší úspěch těm, kteří mohou využít a přizpůsobit se novému paradigmatu poskytování zdravotní péče. Praktická část práce poskytne přehled vývoje umělé inteligence z hlediska aplikačních oblastí: kvantitativní vztahy mezi strukturou a aktivitou při objevování léků, predikce rakoviny pomocí dat z mikročipů, hluboké učení pro analýzu lékařských obrazů, zdravotnictví, klinické studie a monitorování bezpečnosti léků.

Účelem této části práce je popsat a analyzovat reálné případy nasazení umělé inteligence a strojového učení v lékařské vědě a vývoji léků.

### **5.1. Umělá inteligence pro kvantitativní vztahy mezi strukturou a aktivitou při objevování léků**

Aplikace umělé inteligence, zejména hluboké učení při objevování léků, zahrnuje identifikaci nových molekul léčiv a proteinové inženýrství pomocí kvantitativních vztahů mezi strukturou a aktivitou, analýzu dat genové exprese a farmakokinetiku a farmakodynamické modelování. Příklady aplikací umělé inteligence jsou vývoj precizní medicíny, predikce sekvenční specifikace, geonomické modelování pro přeměnu léčiv a interakce s proteiny léčiv. Příklady umělé inteligence ve výpočetní biologii zahrnují předpověď, zda je protein sekretován z jeho aminokyselinové sekvence, nebo ne. Předpověď, zda je tkáň zdravá z experimentu s profilováním genů nebo predikce, zda se chemická sloučenina může ze své struktury vázat na daný cíl či nikoli. Tato část práce se

zaměří na aplikace umělé inteligence v kvantitativních vztazích mezi strukturou a aktivitou.

### 5.1.1. Síť hlubokého učení

Protokoly pro objevování léčiv ve farmaceutickém průmyslu se po mnoho let spoléhají hlavně na vysoce výkonné screeningové metody pro rychlé zjištění biologické nebo biochemické aktivity velkého počtu sloučenin podobných léčivům. Během procesu objevování se často setkáváme s různými problémy, včetně účinnosti, aktivity, toxicity a biologické dostupnosti navržených sloučenin. S vysoce výkonným virtuálním screeningem prokázaly ty, které zahrnují kvantitativní vztahy mezi strukturou a aktivitou svou použitelnost v moderních protokolech pro objevování léků. Při objevování vztahů mezi kvantitativní strukturou a aktivitou z biologických dat souvisejících s navrženými molekulami byly použity různé metody umělé inteligence.

Hiller v roce 1973 naznačil, že umělá neuronová síť by mohla být užitečná pro klasifikaci molekul do dvou kategorií: aktivní a neaktivní. Později Rose v roce 1991 aplikoval metodu nekontrolované neuronové sítě známou jako Kohonenova topologie zachovávající mapování v kvantitativní analýze vztahů mezi strukturou a aktivitou založenou na 2-D reprezentaci složených podobností. Hinton v roce 2006 navrhl algoritmus rychlého učení používaný pro síť hlubokého přesvědčení. Lowe v roce 2014 a později Xu v roce 2017 navrhli víceúlohové neuronové sítě pro kvantitativní předpovědi vztahů mezi strukturou a aktivitou. Eri v roce 2012 použila umělou neuronovou síť s úpravou relativní důležitosti deskriptorů k identifikaci lékových ligandů.

Soutěž Kaggle propagovaná společností Merck v roce 2012 vedla k nástupu hlubokého učení ve výzkumu léčiv. Výzkumníci zabývající se hloubkovým učením zvítězili v soutěži především díky použití hlubokého učení ve vztazích mezi kvantitativní strukturou a aktivitou k zachycení složitých statistických vzorců mezi tisíci deskriptory extrahovanými z mnoha sloučenin.

Gupta v roce 2015 použil rekurentní neuronové sítě a metody dlouhodobé paměti pro De Novo Drugs Design. Naučené pravděpodobnosti vzorů lze použít pro De Novo generování bez potřeby výčtu virtuální složené knihovny. Ghasemi v roce 2018 poskytl přehled neuronových sítí a algoritmů hlubokého učení používaných ve studiích kvantitativních strukturních aktivit. Zdůraznil, že problémy přetrvávají: tajemné

deskriptory mohou ovlivnit výsledky predikce nebo klasifikace biologické aktivity, ačkoli modely nelze interpretovat, zatímco jednodušší interpretovatelné deskriptory nemohou vytvořit dobré modely pro různé sady dat. Naproti tomu hlavními nevýhodami umělých neuronových sítí ve studiích kvantitativních vztahů mezi strukturou a aktivitou jsou problémy s nadměrným přizpůsobením a lokálním optimem, což činí předpověď nebo klasifikaci nespolehlivou. Navrhovaná nápravná opatření zahrnují regularizaci (zavedení penalizace za tzv. overfitting), konturační neuronové sítě s váhami výpadků, sítě hlubokého přesvědčení, autokodéry, opravené lineární jednotky místo sigmoidní funkce a podmíněně omezený Boltzmannův stroj.

Pereira v roce 2016 navrhl novou metodu virtuálního screeningu založenou na hlubokém učení, tzv. DeepVS. Provedl dokování se 40 receptory a 2950 ligandy a výsledky porovnal s 95316 návnadami. Dokovací výstupy receptoru DeepVS vykazoval povzbudivé výsledky. Hluboké učení bylo také použito k vytvoření knihoven zaměřených molekul od Seglera v roce 2017 nebo nových molekulárních otisků prstů od Kadurin v roce 2016.

### **5.1.2. Strojové učení založené na podobnosti sítě**

Bylo implementováno mnoho měřítek podobnosti s využitím biologických, chemických nebo topologických vlastností cílů, léků a známých interakcí. Hu a Agarwal v roce 2009 vytvořili sítě choroba-choroba, lék-lék, choroba-lék tím, že sladili molekulární profily nemocí a profily exprese léků. Výkonnost a předpovědní síla se lišily podle použitých měřítek podobnosti a dostupnosti dat.

Biologická relevance je stanovena, když je splněno alespoň jedno ze tří následujících kritérií: nové cíle přispívají k primární aktivitě drogy; zprostředkovávají nežádoucí účinky léků; nebo nesouvisí se sekvencí, strukturou a funkcí s kanonickými cíli. Guney v roce 2016 vyvinul síťovou metodu založenou na novém měření blízkosti, které se nazývá moduly nemoci. Modul nemoci je tvořen geny spojenými s danou nemocí. Autoři předpokládali, že lék je účinný proti onemocnění, pokud se zaměřuje na proteiny v těsné blízkosti šesti nejběžnějších podobností. Kromě toho je tato metoda schopna dosáhnout výsledku, není ovlivněna ani počtem cílů, které má lék, ani jejich stupněm. Toto zlepšení však vyžaduje přístup ke genům onemocnění, cílům léků a anotacím onemocnění souvisejících s léky.

Zhou v roce 2010 a Chen v roce 2011 navrhli metodu čistě založenou na topologických opatřeních k predikci asociací léky a nemocemi nebo k doporučení nemocí pro získávání dat o vlastnostech bipartitní sítě experimentálně ověřených asociací léků s chorobami. Existují dva různé způsoby, jak lze měření podobnosti sítě použít: pokud lék interaguje s cílovou skupinou, pak budou pro tuto cílovou skupinu doporučeny stejné či podobné léky, a pokud lék interaguje s cílem, pak se lék doporučí jiným cílovým skupinám s podobnými sekvencemi. Pro zlepšení predikce vyvinul Alaimoin v roce 2013 hybridní algoritmus využívající matici podobnosti k přímému začlenění biologických znalostí závislých na doméně do modelu. Matice podobnosti se získá jako lineární kombinace matice podobnosti struktury a matice cílové podobnosti. Tato metoda funguje lépe pro predikci biologicky významných interakcí a překonává metody.

### **5.1.3. Kernelova metoda a podpora vektorového stroje (SVM)**

Na rozdíl od malých molekul má znalost skládání proteinů hluboký dopad na pochopení heterogenity a molekulární funkce proteinů, což dále usnadňuje návrh léčiv. Předvídání 3D struktury proteinu je klíčovým problémem v molekulární biologii (Wei a Zou, 2016). Stanovení složení proteinu se spoléhá hlavně na molekulární experimentální metody, jako je rentgenová krystalografie a nukleární magnetická rezonanční spektroskopie. S rozvojem sekvenčních technik nové generace se objevování nových proteinových sekvencí rychle zvyšuje a převyšuje schopnost tradičních experimentálních technik určit skládání proteinů. Výpočtovými metodami pro rozpoznávání proteinových záhybů mohou být metody založené na šablonách nebo metody bez šablon (volná metoda). V metodách založených na šablonách se proteiny známých struktur získaných z veřejných databází proteinových struktur (např. proteinová databanka, UniProt, DSSP, SCOP, SCOP2 a CATH) používají jako vzorové proteiny pro dotazovanou proteinovou sekvenci. Aby predikce založená na šablonách byla rychlá a spolehlivá, obvykle se používá zjednodušená databáze, ve které je sekvenční podobnost menší než 50 % až 70 %. Algoritmy vícenásobného zarovnání jsou poté přijaty k využití evolučních informací kódováním aminokyselinových sekvencí do profilů. Konečně k určení optimálního srovnání se skórovací funkce používají jako měřítko pro hodnocení podobnosti mezi profily odvozenými z dotazovaného proteinu a profily proteinů se známými strukturami. Modely 3D struktury založené na optimálním zarovnání šablon dotazů lze dále optimalizovat pomocí minimalizace energie a modelování smyček.

Nedávné výzkumné úsilí se zaměřilo na vývoj volných metod, které se snaží předpovídat proteinové struktury pouze na základě aminokyselinových sekvencí spíše než na známých strukturních proteinech, jako u metody založené na šablonách. Metody bez šablon zahrnují skryté Markovovy modely, genetické algoritmy, neuronové sítě, podpůrné vektorové stroje (SVM) a souborové klasifikátory. Strojové učení si klade za cíl vytvořit predikční model tím, že se naučí rozdíly mezi různými kategoriemi skládání proteinů a pomocí naučeného modelu automaticky přiřadí dotazovaný protein ke konkrétní třídě skládání proteinu. Tento přístup je tedy efektivnější pro předpovědi velkého rozsahu a může prozkoumat velký počet nadějných kandidátů pro další experimentální ověření.

SVM jako speciální případ Kernelových metod byly použity při detekci vzdálené homologie proteinů, strukturní klasifikaci proteinů a predikci proteinu vázajícího DNA. Metody rozpoznávání záhybů proteinů jsou založené na SVM (Shamim, 2007, 2013; Damoulas a Girolami, 2008; Yang a Chen 2011). Hlavním rozdílem mezi těmito metodami založenými na SVM je jejich pokrok ve vlastnostech v metodách založených na strojovém učení pro rozpoznávání proteinových záhybů před rokem 2011. Poorinmohammad v roce 2014 zkombinoval přístup podpůrného vektorového stroje (SVM) s deskriptory pseudoaminokyselinového složení pro klasifikaci anti-HIV peptidů s přesností předpovědi 96,76 %.

#### **5.1.4. Metoda rozhodovacího stromu**

Rozhodovací stromy jsou transparentní a interpretovatelnou strategií strojového učení. Obecně existují dva základní kroky pro konstrukci rozhodovacích stromů: výběr atributů a prořezávání. Rozhodovací stromy mají problém s tím, že chyby v up-streamu se mohou rychle šířit do down-stream uzlů, což způsobuje selhání. Existuje několik řešení problému. Například použití náhodného lesa je přístup k modelování souboru, který funguje tak, že jako základ učení konstruuje více rozhodovacích stromů. Zavedením náhodného výběru vlastností a myšlenky „bagování“ (Breiman, 2001) každá báze dále zvětšuje „testovací“ uzly a je trénována náhodně vybranými podmnožinami namísto původního souboru dat. Konečným výsledkem je konsenzuální skóre ze všech jednotlivých výstupů rozhodovacího stromu. Ve srovnání s rozhodovacími stromy je u náhodného lesa méně pravděpodobné, že přeplní data. Náhodný les byl široce používán pro klasifikaci bioaktivity (Singh, 2015), modelování toxicity (Mistry, 2016), predikci vazebné afinity protein-ligand (Wang, 2015) a identifikaci cíle léčiva (Kumari, 2015).

Například Mistry poprvé použil náhodný les a rozhodovací stromy k modelování vztahu toxicity léčiv a „nosičů“. Jejich údaje zahrnovaly 227 093 potenciálních kandidátů na léky a 39 potenciálních „nosičů“. Wang použil metodu náhodného lesa k modelování vazebné afinity proteinového ligandu mezi 170 komplexy proteáz HIV-1, 110 komplexy trypsinu a 126 komplexy karbohydrázy. Kromě toho Kumari zkonstruoval vylepšený náhodný les integrací zaváděcích a rotačních prvků matrice, které úspěšně rozlišovaly lidské drogové cíle od nedrogových (Kumari, 2015).

## **5.2. Umělá inteligence v predikci rakoviny pomocí microarray dat**

### **5.2.1. Detekce rakoviny z dat genové exprese**

Rakovina je závažný celosvětový zdravotní problém obvykle spojený s genetickými abnormalitami. Tyto abnormality lze detekovat pomocí technik microarray, které měří expresi a aktivitu tisíců genů. Technologie microarray je běžně používaným nástrojem pro analýzu genetických chorob. Standardizovaná datová sada microarray se skládá z tisíců genových expresí a několika stovek vzorků. Každá exprese měří úroveň aktivity genů v dané tkáni, takže porovnání genů exprimovaných v abnormálních rakovinných tkáních s geny v normálních tkáních poskytuje dobrý přehled o patologii onemocnění a umožňuje lepší diagnostiku a předpovědi pro budoucí vzorky (Daouda a Mayo, 2019). Vzhledem k vysoké dimenzionalitě profilů genové exprese je však použití takových dat pro detekci rakoviny náročné. Metody umělé inteligence byly buď použity pro redukci rozměrů, jako je výběr nebo vytvoření nejrelevantnějších rozlišujících znaků a eliminace nerelevantních závislých znaků v předchozím kroku k predikci nebo predikci existence rakoviny, typu rakoviny nebo rizika přežití a shlukování pro neoznačené vzorky.

### **5.2.2. Výběr funkcí**

Mezi běžně používaná veřejná úložiště dat pro aplikace umělé inteligence v predikci a shlukování rakoviny patří databáze TCGA, UCI, NCBI Gene Expression Omnibus (CEO) a biomedicínské databáze Kentridge. Odstranění genů, které mají nízkou hodnotu exprese ve všech vzorcích, je jednou z nejjednodušších a nejpřímějších technik předběžného zpracování (Kumardeep, 2017). Principiální analýza komponent byla také použita pro výběr prvků (Zhang, 2018). Princip metody komponent analýzy transformuje prvky datové sady do prostoru nižších rozměrů. Metody filtrování neuronových sítí se

používají k extrakci reprezentací, které nejlépe popisují genové exprese bez přímého zohlednění cíle predikce jako učení bez dozoru (Tan, J., 2015). Konvoluční neuronové sítě byly použity pro extrakci příznaků skenováním sady váhových matic napříč vstupem (Steve, 1997). Generativní adversariální sítě sestávají z generativní sítě a diskriminační sítě, byly také použity pro výběr funkcí.

### **5.2.3. Předpověď rakoviny**

Predikce rakoviny pomocí umělé inteligence zahrnuje předpověď existence rakoviny, typu rakoviny a rizika přežití. Pro filtrování mikročipových genových expresí byly v poslední době použity různé typy autokodérů, jako jsou skládané odšumovací autokodéry (Jie, 2015), kontraktivní autokodéry (Macías-García, 2017), řídké autokodéry (Rasool, 2013), regulované autokodéry (2017ep, a variační autokodéry (Way and Green, 2017). Danae (2016) použil hlubokou architekturu čtyř vrstev s 15 000, 10 000, 2 000 a 5 000 neurony a algoritmus stochastické gradientové optimalizace sestupu. Nejlepší úroveň přesnosti (94,0 %) bylo dosaženo, když byl počet skrytých vrstev tři. Rajendra (2017) navrhl hluboký generativní model nazývaný hluboká rakovina pro binární (rakovinnou/nerakovinnou) klasifikaci. Pro klasifikaci typu rakoviny na základě somatických bodových mutací byl navržen model hlubokého genu, tzv. konvuluční dopředný model.

## **5.3. Hluboké učení pro analýzu lékařských snímků**

### **5.3.1. Hluboké učení pro zpracování lékařských snímků**

Lékařské zobrazování zahrnuje ty procesy, které poskytují vizuální informace o lidském těle. Účelem lékařského zobrazování je pomoci radiologům a lékařům zefektivnit diagnostický a léčebný proces. Lékařské zobrazování je převládající součástí diagnostiky a léčby nemocí (Qayyuma, 2018). Institut medicíny při Národní akademii vědy, inženýrství a medicíny uvádí, že „diagnostické chyby přispívají k přibližně 10 % úmrtí pacientů“ a představují 6 až 17 % nemocničních komplikací (Sennaar, 2018). Příčina chyb diagnostiky je připisována řadě faktorů, včetně informačních technologií.

Inovace v oblasti hlubokého učení a rostoucí dostupnost velkých anotovaných souborů lékařských obrazových dat vedou k dramatickému pokroku v automatizované detekci onemocnění a segmentaci orgánů a lézí. Počítačová diagnostika (CAD) v lékařském zobrazování za posledních několik desetiletí vzkvétala. Pro lékařské

zobrazování se CAD zaměřila především na radiologii, kardiologii a patologii. Příklady v radiologii zahrnují automatizovanou detekci mikrokalcifikací a mas na mamografii, plicních uzlů na RTG hrudníku a CT skenech a polypů tlustého střeva na CT kolonografii (Giger, 2008). V kardiologii patří mezi příklady CAD pro echokardiografii a angiografii (Willems, 1991). V digitální patologii patří mezi příklady detekce buněčných složek, jako jsou jádra a buňky, a onemocnění, jako je rakovina prsu, děložního čípku a prostaty (Xing a Yang, 2016).

U radiologických snímků přesné umístění a rozsah abnormality tradičně určují vyškolení odborníci. Taková ruční anotace odborníkem je časově náročná a nákladná. Například studie NLST zahrnovala více než 53 000 pacientů a stála více než 250 milionů dolarů. V komunitě zabývající se zpracováním lékařského obrazu existuje velká potřeba redukce falešně pozitivních (FP) pro automatizované detekční úlohy. První práce naznačují, že podstatné snížení FP nebo lepší vyhledávání obrázků je možné pomocí běžně dostupných sítí hlubokého učení (Roth, 2016). Přesto ani tato snížení nedosahují specifčnosti získané praktickými lékaři. Je zřejmé, že je zapotřebí další práce nejen na zlepšení strojového učení, ale také na tradičnějším inženýrství funkcí (Ranald a Summers, 2017). Mezi hlavní aplikace hlubokých neuronových sítí pro analýzu lékařského obrazu patří detekce, segmentace a klasifikace v heterogenních zobrazovacích modalitách.

Segmentace je proces rozdělování obrazu do více nepřekrývajících se oblastí na základě specifických kritérií, tj. sad pixelů s vnitřními vlastnostmi, jako je barva, kontrast a textura pro smysluplnou extrakci informací. Taková extrakce informací zahrnuje tvar, objem, relativní polohu orgánů a detekci abnormalit. Například algoritmus segmentace mozkových nádorů nebo gliomových tumorů může být použit v kaskádové hluboké konvoluční neuronové síti. Anwar (2018) představil nejnovější přehled v analýze lékařského obrazu pomocí konvoluční neuronové sítě. Oblast použití konvolučních neuronových sítí pokrývá celé spektrum analýzy medicínského obrazu včetně detekce, segmentace, klasifikace a počítačově podporované diagnostiky.

Konvoluční neuronové sítě se používají pro klasifikaci plicních onemocnění (Anthimopoulos, 2014) pomocí dvou databází intersticiálních plicních onemocnění a CT skenů, z nichž každá má rozměr 512 x 512. Po proceduře extrakce příznaků jsou obrazové záplaty reprezentovány příznakovými vektory, které jsou přiváděny do náhodného lesního klasifikátoru pro klasifikaci. Síť je trénována pomocí 14 696



obrazových polí odvozených z původního skenu s přesností klasifikace 85,5 %. Metoda hlubokého učení založená na Boltzmannově stroji omezeném konvoluční neuronovou sítí pro klasifikaci textury plic a detekci dýchacích cest v souborech dat CT plic. Síť je trénována na obrazových polích o velikosti 32 x 32 voxelů podél obvodu s překrytím 16 voxelů. Patch je zachován, pokud má 75 % voxelů patřících do stejné třídy. V obou aplikacích kombinace výukových cílů překonala čistě diskriminační nebo generativní učení a zvýšila například přesnost klasifikace plicní tkáně o 1 až 8 procentních bodů.

Kamnitsas (2016) navrhl dvojí cestu, 3D konvoluční neuronové sítě s hloubkou 11 vrstev pro náročný úkol segmentace mozkových lézí. Navržená architektura je výsledkem hlubkové analýzy omezení současných sítí navržených pro podobné aplikace. K překonání výpočetní zátěže při zpracování 3D lékařských skenů se používá efektivní husté školení. Pro následné zpracování měkké segmentace sítě autoři použili 3D plně propojené podmíněné náhodné pole, které efektivně odstraňuje falešné popluchy. Přístup hlubokého učení byl rozsáhle hodnocen pro tři náročné úkoly segmentace v datech pacientů s vícekanálovou MRI spojenou s traumatickými poraněními mozku, mozkovými nádory a ischemickou mrtvicí, což prokázalo jeho výpočetní účinnost.

### **5.3.2. Hluboké učení v mamografii**

Rakovina prsu je jedním z nejčastějších typů rakoviny, který ovlivňuje životy žen na celém světě. Nedávné statistické údaje zveřejněné Světovou zdravotnickou organizací (WHO) odhadují, že 23 % případů rakoviny a 14 % úmrtí na rakovinu u žen je způsobeno rakovinou prsu. Nedávné studie ukazují, že manuální analýza má senzitivitu 84 % a specifickou 91 %. Druhé čtení stejného mamografu buď od radiologů, nebo z počítačově podporovaných diagnostických (CAD) systémů může tento výkon zlepšit (Carnerio, 2017).

Systém umělé inteligence může analyzovat léze prsu pomocí mamografů, obvykle obsahuje tři typy:

1. Detekci lézí.
2. Segmentaci lézí.
3. Klasifikaci lézí.

Detekce lézí zahrnuje identifikaci velkého počtu kandidátských oblastí, obvykle na základě použití tradičních filtrů, jako jsou morfologické operátory. Tito kandidáti jsou

pak zpracování ve druhé fázi, jejímž cílem je odstranit falešně pozitivní výsledky pomocí přístupů strojového učení, např. klasifikátorů regionů (Kozegar, 2013). Segmentace lézí je užitečná, protože tvar lézí je důležitým znakem při klasifikaci lézí v konečné fázi analýzy. Problémy spojené s těmito kroky jsou způsobeny především nízkým poměrem signálu k šumu přítomnému při zobrazování léze a nedostatkem konzistentního umístění, tvaru a vzhledu lézí (Thang, 2009 a Oliver, 2010). Současný pokrok v hlubokém učení přináší naději, že tyto výzvy lze překonat. Hlavním problémem, kterému čelí metody hlubokého učení, je však přehnaná výbava: příliš mnoho dat s příliš malým počtem datových bodů. Potřeba velkých anotovaných trénovacích sad vzhledem k rozsahu sad parametrů má zřídka více než několik tisíc vzorků. Speciální konvoluční neuronové síť, nazývané regionové konvoluční neuronové síť, navrhl Dhungel (2015) pro metodologii detekce hmoty prsu, která se skládá z kaskády klasifikátorů. Metoda hromadné detekce založená na hlubokém učení sestávající z kaskády modelů hlubokého učení byla navržena Ertosunem a Rubinem (2015).

### **5.3.3. Hluboké učení pro analýzu kardiologických snímků**

Kardiovaskulární onemocnění jsou nejčastější příčinou úmrtí ve vyspělých zemích a každý rok si vyžádají více životů než dalších sedm hlavních příčin úmrtí dohromady. Různé zobrazovací modalitty, jako je počítačová tomografie (CT), magnetická rezonance (MRI), ultrazvuk a nukleární zobrazování, jsou široce dostupné technologie používané v klinické praxi k vytváření požadavků. Ultrazvuk je nejrozšířenější pro analýzu srdeční funkce kvůli jeho nízké ceně a nízké radiační dávce. Nukleární zobrazování a MRI se používají pro zobrazení perfuze myokardu a k měření životaschopnosti myokardu. CT odhaluje nejpodrobnější srdeční anatomické struktury a běžně se používá pro zobrazování koronárních tepen. Fluoroskopie je nejtěžší zobrazovací modalita pro srdeční intervence (Lloyd-Jones, 2009).

Kardiovaskulární struktury se skládají ze srdeční komory, chlopně, cév, tepen a žil. Analýza srdečního obrazu zahrnuje detekci, segmentaci, sledování pohybu, kvantifikaci a diagnostiku onemocnění (Gulsun, 2016).

Sítě hlubokého učení, jako je konvoluční neuronová síť (Avendi, 2016) a síť hlubokého přesvědčení (Carneiro, 2012), byly nedávno použity při zobrazování srdce, včetně segmentace levé/pravé komory (Zhen, 2016), segmentace cév sítnice (Wang, 2015 a Chandrakumar s Kathirvel, 2016).

Kniha, kterou vydal Lu (2017), pokrývá široké aplikace hlubokého učení v obrazových studiích s názvy jako jsou například:

- Efektivní falešně pozitivní snížení detekce počítačem pomocí konvolučních neuronových sítí a náhodné agregace zobrazení;
- Robustní detekce orientačních bodů v objemových datech s efektivním 3D hlubokým učením;
- Nová metoda detekce buněk využívající hlubokou konvoluční neuronovou síť a sadu nezávislých na maximální hmotnosti;
- Hluboké učení pro histopatologickou analýzu obrazu směrem k počítačové diagnostice rakoviny;
- Intersticiální plicní onemocnění prostřednictvím hlubokých konvolučních neuronových sítí;
- Tři aspekty použití konvolučních neuronových sítí pro počítačově podporovanou detekci v lékařském zobrazování;
- Detekce buněk pomocí hlubokého učení urychleného od Sparse Kernela;
- Plně konvoluční sítě v lékařském zobrazování, aplikace pro vylepšení a rozpoznávání obrazu;
- O nutnosti jemně vyladěných konvolučních neuronových sítí pro lékařské zobrazování;
- Plně automatizovaná segmentace využívající regulovanou vzdálenost a hluboce strukturované učení a vyvozování;
- Kombinace hlubokého učení a strukturované predikce pro segmentaci hmot v mamografech;
- Automatická segmentace patologické ledviny v CT založená na hlubokém učení: lokální versus globální obrazový kontext;
- Robustní buněčná detekce a segmentace v histopatologických snímcích pomocí řídké rekonstrukce a vrstvených odšumovacích autokodérů;
- Automatická segmentace slinivky břišní pomocí hrubého až jemného značení super pixelů.

## 6. Prognóza vývoje umělé inteligence ve zdravotnictví

Kapitola prognóza vývoje umělé inteligence je věnována analýze historických dat, dostupných analýz, informací a na jejich základě předikování dopadů umělé inteligence na pracovní místa ve zdravotnictví. Technologické změny nepřicházejí ze dne na den a neovlivňují stejně všechny pracovní místa. V této části práce se proto zaměřím na dopady umělé inteligence na zaměstnanost vybraných zaměstnání ve zdravotnictví. Bude zajímavé pozorovat, jak umělá inteligence bude ovlivňovat jednotlivé profese, jaké úkony bude umělá inteligence přebírat, a naopak jaké dovednosti budou vyžadovány od zdravotního personálu. Všechny následující prognózy budou predikovat dopad umělé inteligence na vybrané zdravotní profese v roce 2030.

### 6.1. Jak umělá inteligence změní pracovní svět ve zdravotnictví

Navzdory nepopíratelnému potenciálu různých technologií a aplikací zkoumaných v předchozích kapitolách, je stále mimo jak potenciální dopad automatizace a umělé inteligence jako spektra aplikací na praktické lékaře, tak pravděpodobnost přijetí ve velkém měřítku způsobem, který může způsobit revoluci v poskytování zdravotní péče. V této rané fázi vývoje umělé inteligence je těžké předvídat všechny způsoby, jak se změní poskytování zdravotní péče, a existuje několik možných provozních a obchodních modelů, které by se mohly objevit.

Aby se zdravotničtí pracovníci připravili na co nejlepší využití příležitostí, které nabízí umělá inteligence ke zlepšení péče o pacienty, museli by začít měnit svůj přístup ke vzdělávání, celoživotnímu učení, týmové práci, posílení postavení pacientů, vítat a integrovat nové kategorie profesionálů s hybridními „klinickými a datovými“ profily nebo zcela novými profily, jako jsou datoví vědci. Automatizace a umělá inteligence se často používají zaměnitelně, ale nejsou totéž. Automatizační řešení stále více zahrnují komponentu umělé inteligence, aby se rozšířila jejich flexibilita a efektivita při řešení problémů, kterým čelí zdravotnické organizace a jednotlivci. Vzhledem k tomu, že umělá inteligence představuje podmnožinu řešení obecně označovaných jako automatizace, v této části diplomové práce je čerpáno z práce McKinsey Global Institute o dopadu automatizace a umělé inteligence napříč průmyslovými odvětvími, aby bylo možné prozkoumat potenciální dopad umělé inteligence na zdravotnické pracovníky v

Evropě. Tento pohled je doplněn pohledem na činnosti na oblasti, kde umělá inteligence může změnit životy odborníků, a přehledem některých nových profilů, které by mohly být potřeba k tomu, aby bylo upozorněno na hlavní změny, které budou muset organizace řešit.

## 6.2. Změny v pracovních místech: Zdravotnictví v roce 2030

V sérii publikací McKinsey Global Institute studoval od roku 2017, jak automatizace pravděpodobně ovlivní pracovní sílu s důrazem na to, jak může automatizace snížit potřebu některých činností, a také kolik hnacích sil, jako je produktivita, stárnutí, investice a spotřeba by mohla zvýšit poptávku po určitých povoláních a pracovních činnostech (McKinsey, 2018).

„Umělá inteligence tu není od toho, aby převzala smysluplnou část práce lékaře – nikdy by nefungovala, kdyby přerušila lidské spojení mezi sestrami a jejich pacienty.“ (Dr. Kyra Bobinet, Fresh Tri, 2018)

Tato diplomová práce se zaměřuje konkrétně na zdravotní péči a na dopad automatizace a umělé inteligence na pracovní sílu v Evropě. Tato analýza se pomocí metodologie dřívější automatizace McKinsey Global Institute pokouší identifikovat činnosti a povolání ve zdravotnictví, která jsou nejvíce a nejméně náchylná k dopadu automatizace a zavádění umělé inteligence ve zdravotnictví.

Automatizace ovlivní většinu pracovních míst napříč sektory, ale stupeň ovlivnění se výrazně liší. Například až 58 % času stráveného ve výrobě je dnes ve vybraných zemích EU potenciálně automatizovatelné pomocí stávající technologie a mohlo by být uvolněno pro jiné úkony, ale ve zdravotnictví je tento čas pouze 30 % z celkového pracovního času – tím se řadí mezi sektory s nejmenším podílem potenciálně automatizovanými úkoly. Ve většině případů automatizace ovlivňuje pouze konkrétní činnosti v rámci povolání a uvolňuje čas, který lze využít jiným způsobem. Ve zdravotnictví by se to mohlo přenést do poskytování péče pacientům než na vyřizování administrativních úkolů. Analýzy McKinsey Global Institute naznačují, že potenciál pro automatizaci úkolů ve zdravotnictví se výrazně liší, přičemž větší potenciál pro automatizaci mají například činnosti vykonávané zdravotními asistenty nebo zubními laboranty, naproti tomu činnosti vykonávané chirurgy a psychiatry mají nejmenší potenciál pro automatizaci pracovních činností.

Kombinace činností vykonávaných zdravotnickou pracovní silou zahrnuje mnoho činností, které mají specifické dovednosti, které je obtížnější automatizovat, a tyto dovednosti jsou v tomto sektoru potřeba ve větší míře než v mnoha jiných pracovních povoláních. Povolání, která zahrnují nepředvídatelné fyzické aktivity, uplatnění odborných znalostí nebo sociálních dovedností, mají mnohem menší potenciál pro automatizaci než předvídatelné fyzické úkoly, rutinní sběr dat nebo úkoly zpracování. Ve zdravotnickém sektoru skutečně nedojde k poklesu fyzických a manuálních dovedností, očekává se, že procento času, kdy lékaři používají hrubou motoriku a sílu, nebo jemné motorické dovednosti k obsluze zařízení, poroste, zatímco jejich čas strávený kontrolou a sledováním se s příchodem digitálních a AI monitorovacích řešení znatelně sníží.

Potenciál pro automatizaci se zase liší od pravděpodobnosti přijetí. Zatímco technická proveditelnost je klíčovým prvkem automatizace, náklady na vývoj a implementaci automatizačních řešení, místní nabídka práce, poptávka a mzdové sazby, čisté ekonomické přínosy automatizace včetně zvýšené kvality, bezpečnosti a regulační i sociální akceptace jsou všechny faktory, které ovlivňují míru přijetí napříč zeměmi. Analýza McKinsey Global Institute například zdůrazňuje, že převládající místní mzdové sazby se mohou mezi zeměmi značně lišit, a to i pro podobná povolání, a že relativní mix povolání má také silnou místní složku. Analýza McKinsey Global Institute, ze které primárně vychází tato část diplomové práce využívá scénář středního bodu, který odhaduje, že do roku 2030 bude pravděpodobně automatizováno až 15 % současné pracovní doby ve zdravotnictví, ve srovnání s 30 %, které dnes mají technologické sektory jako potenciál pro automatizaci.

Graf 1 ukazuje podíl času, který by mohl být do roku 2030 uvolněn automatizací pro širokou škálu zdravotnických povolání v sedmi konkrétních evropských zemích (McKinsey, 2020). Jak již bylo v této práci poznamenáno, zkrácení pracovní doby se výrazně liší podle povolání. Například až polovina času stráveného na současných pracovních činnostech přípravářů lékařského vybavení by mohla být do roku 2030 automatizována ve scénáři midpoint automatizace, zatímco u porodní asistentky by to mohlo být jen 5 procent času nebo téměř nic pro chiropraktiky.

Graf 1: Podíl odpracovaných hodin, který by mohl být automatizován do roku 2030



Vlastní zpracování, zdroj: McKinsey Report – Transforming-healthcare-with-AI

### 6.2.1. Dopad na počty zaměstnanců

Někteří zdravotníci se mohou obávat, že automatizace a rozšiřování řešení umělé inteligence může ohrozit jejich zaměstnanost, ale skutečnost je taková, že evropský zdravotnický sektor čelí propasti pracovní síly, která odráží celosvětovou propast pracovní síly ve zdravotnictví – a očekává se, že se tato propast bude jen prohlubovat (Britnell, 2019). WHO odhaduje, že celosvětový rozdíl v počtu lékařů, sester a porodních asistentek bude do roku 2030 činit 9,9 milionů. V Evropě WHO očekává, že celková poptávka po lékařích (napříč profesemi) vzroste do roku 2030 na 18,2 milionů. (WHO, 2019). Podle britského zdravotnického think-tanku King's Fund, jen pro britské NHS je očekáván nárůst rozdílu mezi nabídkou a poptávkou po personálu na téměř 250 000 ekvivalentů plného úvazku (FTE), oproti dnešním více než 100 000. To znamená, že se musí nejen přilákat, školit a udržet zdravotníky i v budoucnu, ale také umožnit jim trávit čas se svými pacienty tam, kde to přináší největší přidanou hodnotu snížením zbytečných procesů, které mohou být automatizované.

Mohla by umělá inteligence a automatizace místo ohrožení pracovních míst pomoci zacelit mezeru? A do jaké míry? Tabulka 2 ukazuje nejnovější analýzu McKinsey Global Institute o změně počtu FTE potřebných do roku 2030, s přihlédnutím k dopadu

automatizace. Změna je způsobena demografickými a dalšími makroekonomickými faktory, včetně stárnutí populace, zvýšeným výskytem chronických onemocnění a komorbidit u starších náchylnějších pacientů a vstupem pečovatelských aktivit do tržní ekonomiky (např. monetizace péče poskytované dříve neformálně nebo v rámci rodiny). Analýza naznačuje výrazné zvýšení poptávky po konkrétních profesionálech, jako jsou ošetřovatelé, licencované praktické a odborné sestry, domácí zdravotní asistenti a další, se snížením poptávky po profesích, jako jsou přípraváři lékařského vybavení a technici lékařských záznamů a zdravotnických informačních techniků.

*Tabulka 2: Do roku 2030 budou různá povolání v různé míře ovlivněna demografickými faktory a zavedením automatizace a umělé inteligence ve zdravotnictví*

<b>Top 10 povolání dle růstu poptávky</b>	<b>FTEs vytvořené do r. 2030</b>	<b>% Rozdíl</b>
Zdravotní asistenti	1 172 365	73
Praktické a odborné zdravotní sestry	577 900	24
Domácí zdravotní asistenti	264 453	63
Anesteziologická sestra	200 232	22
Anesteziolog	103 968	27
Internisté	74 101	28
Lékárníci	49 224	15
Fyziologové	46 139	23
Záchranáři	38 067	21
Praktičtí lékaři	37 331	24

<b>Spodních 10 povolání dle růstu poptávky</b>	<b>FTEs ztracené do r. 2030</b>	<b>% Rozdíl</b>
Sociální a personální asistenti	348 957	-18
Příprava lékařského vybavení	70 497	-43
Oční lékařští technici	33 204	-11
Zubní asistenti	20 025	-21
Technici lékařských záznamů	17 493	-18
Lékařští asistenti	16 078	-6
Orální a maxilofaciální chirurgové	11 236	-5
Rehabilitační poradci	6 781	-11
Porodníci a gynekologové	6 139	-7
Poradci pro duševní zdraví	5 741	-8

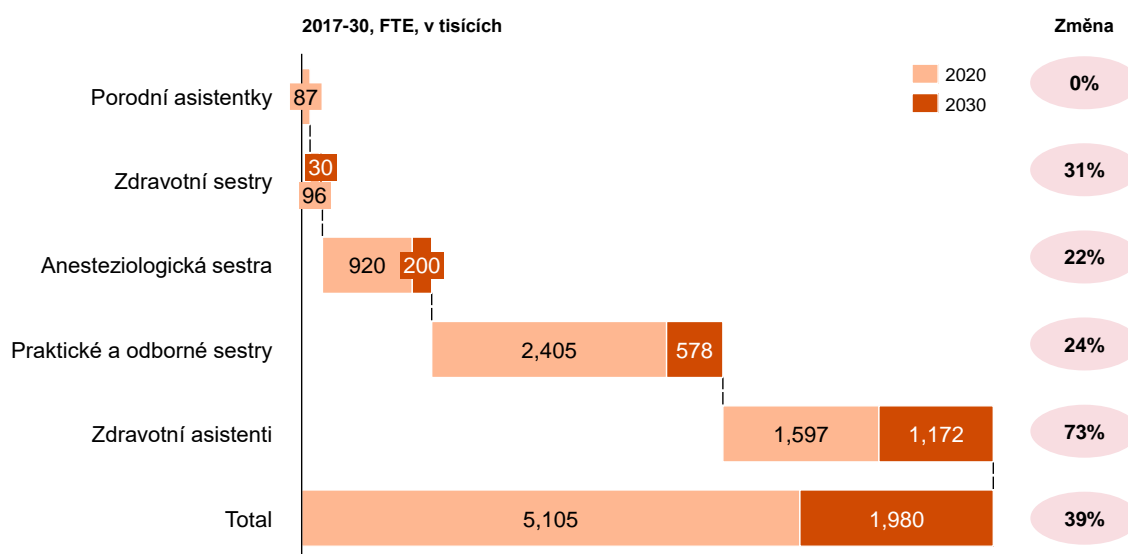
*Vlastní zpracování, zdroj: McKinsey Global Institute. Selected European countries: France, Germany, Hungary, Italy, Portugal, Sweden, UK*



Celkový závěr pro evropské i mezinárodní zdravotnické systémy by mohl být takový, že automatizace by mohla hrát významnou roli při zmírňování nedostatku pracovní síly ve zdravotnictví. Sestry například tráví méně než 40 % svého času přímou péčí o pacienty, ostatní čas často zabírá příprava a výdej léků, komunikace s ostatním personálem při předávání pacientů mezi směnami, kompletace dokumentace a koordinace provozu na oddělení (např. správa lůžek a personální zajištění). (Westbrook, 2011). Mnohé z těchto činností, zejména administrativní úkoly, by mohly být podporovány automatizací nebo a rozšířením umělé inteligence, aby se uvolnil čas sester a umožnilo jim trávit více času přímou péčí o pacienty, komunikací a koučováním nabízeným rodinám a pečovatelům.

Očekává se, že v evropských systémech zdravotní péče výrazně poroste poptávka po povoláních jako ošetrovatelství. Podle scénáře automatizace McKinsey Global Institute bychom například do roku 2030 mohli zaznamenat 39% nárůst ve všech ošetrovatelských profesích (Graf 2). K největšímu nárůstu v absolutním i relativním vyjádření by došlo u zdravotnických asistentů, ale zvýšila by se všechna ošetrovatelská povolání kromě porodních asistentek, jejichž počet by zůstal stabilní (částečně odrážející porodnost v celé EU). (Evropský parlament, 2019). Tato analýza zahrnuje odhad, že přibližně 10 % ošetrovatelských činností by bylo uvolněno automatizací, což ilustruje, jak by automatizace mohla ovlivnit to, jak zvládáme potenciální nedostatek personálu.

Graf 2: Počet hodin odpracovaných sestrami se plošně zvýší



Vlastní zpracování, zdroj: McKinsey Global Institute. Selected European countries: France, Germany, Hungary, Italy, Portugal, Sweden, UK

### **6.3. Jak změní umělá inteligence a automatizace činnost zdravotnických pracovníků?**

Po zvážení celkového dopadu na počty zaměstnanců ve zdravotnictví se nyní zaměříme na to, jak by se mohly každodenní aktivity zdravotnických pracovníků změnit se zavedením nebo škálováním umělé inteligence a automatizace.

Vzhledem ke spektru aplikací umělé inteligence popsanych v předchozí podkapitole není vždy možné oddělit, jak velký vliv automatizace na zdravotnictví, který byl popsán v části 6.1., lze připsat výhradně umělé inteligenci. Mnoho respondentů průzkumu společnosti McKinsey souhlasí s tím, že dopad umělé inteligence bude pocíťován napříč pracovní silou ve zdravotnictví. Shodují se také na tom, že jádrem každé změny je možnost přeorientovat se a zlepšit péči o pacienty, zejména v ošetrovatelství. „Inteligentní plánování a optimalizace počtu zaměstnanců prostřednictvím umělé inteligence by pro zdravotní sestry změnila hru.“ (Rebecca Love, 2020) Toto zlepšení by mohlo zahrnovat nejen poskytování lepší diagnostiky a účinnější léčby, ale také zásadní změny v interakcích mezi lékařem a pacientem a přehodnocení poskytování péče zaměřené na pacienta, jako je domácí péče s umělou inteligencí a management chronických onemocnění, z nichž každý má svůj vlastní účinek na kvalitu péče a klinické výsledky.

„V budoucnu umělá inteligence nenahradí lékaře. Lékaři, kteří používají umělou inteligenci, nahradí lékaře, kteří umělou inteligenci nepoužívají“ (Dr. Eyal Zimlichman, Sheba Medical Center).

Další část podkapitoly bude zaměřena na to, jak se pravděpodobně změní aktivity odborníků, protože umělá inteligence bude v těchto sektorech převládat. Celkově se očekává, že umělá inteligence umožní praktikům se zaměřit na ty činnosti, které pro pacienty přinášejí největší hodnotu. Neočekává se, že umělá inteligence nahradí lékaře, ale spíše rozšíří jejich schopnosti, aby měla dopad na pacienty a zdravotnické systémy, kterým slouží, (Stanford University, 2017).

#### **6.3.1. Méně administrativy, více péče o pacienty**

Administrativní úkoly pohlcují obrovské množství času zdravotnických pracovníků, přičemž některé studie naznačují, že mohou zabrat až 70 % času lékaře, zatímco sestry tráví pouze asi 40 % svého času přímou nebo nepřímou péčí (Telegraph, 2015). Velká

část této administrativní zátěže lze zmírnit dobře navrženými řešeními umělé inteligence, která budou znamenat nejen více času stráveného s pacienty, ale výrazně vyšší pracovní spokojenost. A samozřejmě usnadnění života klinického personálu zvyšuje pravděpodobnost, že tato řešení v rámci organizace přijmou a prosadí. „Umělá inteligence pomůže našemu ošetrovatelskému týmu vrátit více radosti. Nikdo nechce strávit půl dne zadáváním dat do počítače.“ (Robert Freeman, Mount Sinai, 2019)

Již můžeme vidět některé z aplikací, které mohou převzít administrativní úkoly, jako jsou klinické poznámky využívající schopnosti rozpoznávání řeči. Rozpoznávání řeči bylo zavedeno v několika klinických prostředích, aby se zabývalo velkým množstvím klinické dokumentace, a zde se zaměřujeme na potenciální dopad z hlediska skutečných zdrojů potřebných v péči. Například oddělení South Tees Hospital NHS Foundation Trust Accident & Emergency, které poskytuje zdravotnické služby přibližně 1,5 milionu lidem ve Spojeném království, uzavřelo partnerství s poskytovatelem ze soukromého sektoru s cílem zlepšit rychlost a kvalitu dokončování dokumentů (Digital Health, 2018). dramaticky se změnila aktivita napříč celým oddělením: dříve 95 % zaměstnanců psalo poznámky ručně nebo je psalo na stroji; nyní 69 % používá rozpoznávání řeči k vyplnění klinických poznámek, což ušetří v průměru 3,5 minuty na pacienta na hlavní jednotce, což je ekvivalent dvou lékařů na plný úvazek ročně.

Vzhledem k tomu, že je zapotřebí dalších 2 200 konzultantů na daném oddělení, aby bylo možné plně řešit personální potřeby ve Spojeném království, uvolnění času konzultantů by mohlo mít významný dopad na poskytování bezpečné a včasné péče a na zlepšení zkušeností pacientů (Hassan, 2017). Podobný software byl implementován v Homerton NHS Trust ve východním Londýně, kde se doba vyřízení klinických dopisů zkrátila ze 17 dnů na pouhé dva (Journal of mHealth, 2018).

Umělá inteligence může nejen zlepšit provoz v nemocnicích, ale také v primární a komunitní péči, včetně mimo nemocniční péče o starší a křehké pacienty. Například prediktivní plánování – inteligentně předpovídá, které schůzky pravděpodobně v ordinaci praktického lékaře překročí, a podle toho přizpůsobí plánování schůzek – a nástroje pro správu kapacity mohou pomoci zkrátit čekací doby pacientů, zlepšit jejich zkušenosti, výsledky a uvolnit zdroje v péči o seniory. To je důležité zejména tam, kde by zpoždění v přístupu k léčbě mohlo ohrozit klinické výsledky, například u specifických typů diagnóz rakoviny.

Dva pozoruhodné příklady managementu zdravotnické kapacity jsou management lůžek a management přípravy a využití operačního sálu (tzv. divadla). Tyto logistické výzvy zahrnují sladění dostupných zdrojů s očekávanou poptávkou nebo plánovaným harmonogramem, s ohledem na situaci pacienta, křehkost a komorbiditu, demografii, sezónnost, sníženou kapacitu v sousedních zařízeních atd., ale také specifický nemocniční kontext, jako je operující chirurg a chirurgický tým, klinické protokoly, úroveň zkušeností, předoperační procesy, protokoly týkající se anestezie a zotavení a tak dále. Na rozdíl od provozu leteckých společností, kde je často potřeba upravit přiřazování letadla k pravidelnému letu, aby reagovalo na narušení, by řešení za pomoci umělé inteligence mohlo umožnit plynulejší provoz nemocnice.

Umožnění personálu vidět dostupná lůžka na digitálních displejích v reálném čase místo toho, aby volal na jiná oddělení a potvrdil dostupnost, uvolňuje kapacitu od rutinních úkolů, které je třeba věnovat klinické péči. Postupem času však může klinický personál využít vygenerovaná data k informování o procesu učení rozšířeném o umělou inteligenci, který snižuje neefektivnost systému a zpoždění; například identifikace příčin klinicky zbytečných dlouhých pobytů. V roce 2016 spustila nemocnice Johnse Hopkinse v USA kapacitní velitelské centrum pro lepší předvídání a řízení toku pacientů. Pacientům je nyní přiděleno lůžko o 30 % rychleji, jakmile bylo přijato rozhodnutí je přijmout. Singapurská nemocnice Tan Tock Seng, jedna z největších multidisciplinárních nemocnic v zemi, používá podobný systém k nahrazení procesu, kdy zdravotnický personál musel projít více než 300 pravidel během příjmu, aby se správně určila priorita přidělování lůžek (Topol Review, 2017). Tato řešení mohou vést k vytvoření malých týmů centralizované správy lůžek, které veškerou námahu odeberou sestřám, lékařům a dalším praktikům.

Operační sály jsou ve zcela odlišné pozici, ale rovněž se nabízí pro řešení založeném na umělé inteligenci. Jsou to velmi složitá prostředí. Plánování je komplikované i bez neplánovaných mimořádných událostí a narušení. Určit, kolik času bude každý pacient potřebovat a jaké přesně zdroje potřebuje, je náročné. Procedury často trvají déle, než bylo plánováno, což vede ke zrušení procedur rezervovaných později během dne, nebo rezervační procesy počítají se značnými mezerami a operační sály zůstávají nevyužité. Umělá inteligence by mohla uvolnit drahocenný čas chirurgům a dalšímu personálu (např. anesteziologům, zdravotním sestřám) a předpovídat pravděpodobnou dobu, kterou by zákrok u konkrétního pacienta mohl zabrat. Nemocnice Frederick Memorial Hospital

v Marylandu používá právě takový nástroj a podobné aplikace vyvíjejí a implementují nemocnice i soukromý sektor. Celkově lze říct, že od umělé inteligence v chirurgii jsou velká očekávání. Jak říká profesor chirurgie z Cincinnati Children's Hospital Medical Center: „Tok pacientů přes rušný operační sál by mohl být ‚chytřejší‘. Použití algoritmů založených na umělé inteligenci by mohlo lépe předvídat časový průběh, což by mohlo podstatně zvýšit efektivitu a umožnit péči o více pacientů v daný den.“ (Todd Ponsky, Cincinnati, 2020)

Již z těchto dvou příkladů je zřejmé, že řešení s umělou inteligencí dokážou praktickému lékaři odebrat opakující se automatizační úkoly a mohou přinést provozní efektivitu, která lékařům, sestřám a dalšímu klinickému personálu umožní více času věnovat přímé péči o pacienty. Adopci by mělo bránit jen málo překážek. Opakovaným tématem našich rozhovorů bylo, že redukce administrativní práce není jen méně kontroverzní, ale je na prvním místě seznamu přání většiny lidí.

## **6.4. Nové činnosti a dovednosti ve zdravotnictví**

Nové technologie nevyhnutelně vedou k novým typům činností pro zdravotnické pracovníky – v neposlední řadě kvůli změnám ve způsobu interakce samotných pacientů se zdravotnickým systémem – a zcela novým dovednostem, které jim pomohou orientovat se v tomto digitálním světě.

„Náš pracovní profil se bude časem měnit. Umělá inteligence bude schopna provádět některé úkoly, které zaberou spoustu času. Jsme rádi, když to můžeme delegovat a soustředit se na diagnostické myšlení a rozhovory s pacienty. Nevidím žádnou hrozbu pro radiology ve využití umělé inteligence v příštích 20 až 30 letech“ (Felix Nensa, University Hospital Essen).

### **6.4.1. Nový způsob interakce s pacienty**

Pacienti již mají bezprecedentní přístup k lékařským znalostem. Jejich očekávání od digitálně dostupných služeb jsou vysoká, přenesená z odvětví více zaměřených na spotřebitele. Pacienti se při přístupu ke zdravotní péči stávají spíše spotřebiteli, kteří předem vyhodnocují možnosti a často přicházejí do konzultace s informacemi, které mají probrat se svým lékařem – i když jsou tyto informace často mylné.

Zavedení řešení umělé inteligence, zejména e-triage nebo autodiagnostických nástrojů, může umocnit trend, kdy pacienti navštěvují konzultaci s předem vytvořeným

názorem na pravděpodobnou diagnózu a léčbu. Kvalita těchto informací se může značně lišit, což může vést k frustraci, protože lékaři potřebují řešit otázky vyvolané neúplnými nebo někdy dokonce chybnými informacemi.

Ale kvalita informací, které pacienti mají, může být také výrazně lepší. Lékaři a další zdravotničtí pracovníci budou stále častěji žádáni, aby poskytli druhý názor – ten první byl odvozen algoritmem. To bude vyžadovat, aby lékař měl základní znalosti o tom, jak tyto aplikace fungují, byl schopen diskutovat o jejich omezeních a o tom, zda byly správně používány. Ale kromě toho to bude vyžadovat otevřenou mysl a ochotu zapojit se do doporučení umělé inteligence. Může dokonce vyžadovat další konzultační čas, který může být zmírněn zvýšenou digitální gramotností a umělou inteligencí gramotnosti širší populace pacientů, což může potenciálně snížit celkový počet pacientů, kteří vyhledávají péči v první linii.

Můžeme již dnes vidět, jak se stále více usiluje o umožnění předepisování a úhrady „digitální medicíny“. V Německu mohou být od roku 2020 pacientům předepisovány digitální aplikace, včetně digitálních deníků pro diabetes, aplikací na podporu fyzioterapie, psychoterapie nebo měření krevního tlaku, které budou hrazeny německým zákonným zdravotním pojištěním (GovInsider, 2021).

Aby lékaři takové aplikace předepisovali, musí je samozřejmě znát a rozumět tomu, pro koho jsou vhodné a za jakých podmínek. Musí porozumět relativním výhodám a omezením řešení umělé inteligence a také etickým důsledkům. Pokud by řešení s umělou inteligencí mohlo být pro pacienta možností, lékaři se budou muset rozhodnout nejen o tom, zda je technologie vhodná pro základní onemocnění, ale také o tom, zda ji pacient bude schopen správně používat. Budou muset trávit čas seznamováním pacientů a pečovatelů s technologií a radit jim o důsledcích generovaných klinických dat. Bude to svět transparentnosti, spolupráce a dynamické interakce mezi zdravotnickým pracovníkem, pacientem, pečovatelem a poskytovatelem umělé inteligence, a to nejen za účelem zlepšení výsledků, ale také za účelem preventivního řízení potenciálních rizik, včetně rizik týkajících se kalibrace toho, jak a kdy používat konkrétní řešení.

#### **6.4.2. Posílení digitálních dovedností u širší pracovní síly ve zdravotnictví**

Zdravotnický personál pracuje na své digitální gramotnosti posledních 15 let, a i nadále je digitální gramotnost uvedena jako hlavní priorita v The Topol Review (Healthcare IT News, 2019). Praktici musí nejen umět používat řešení založené na umělé

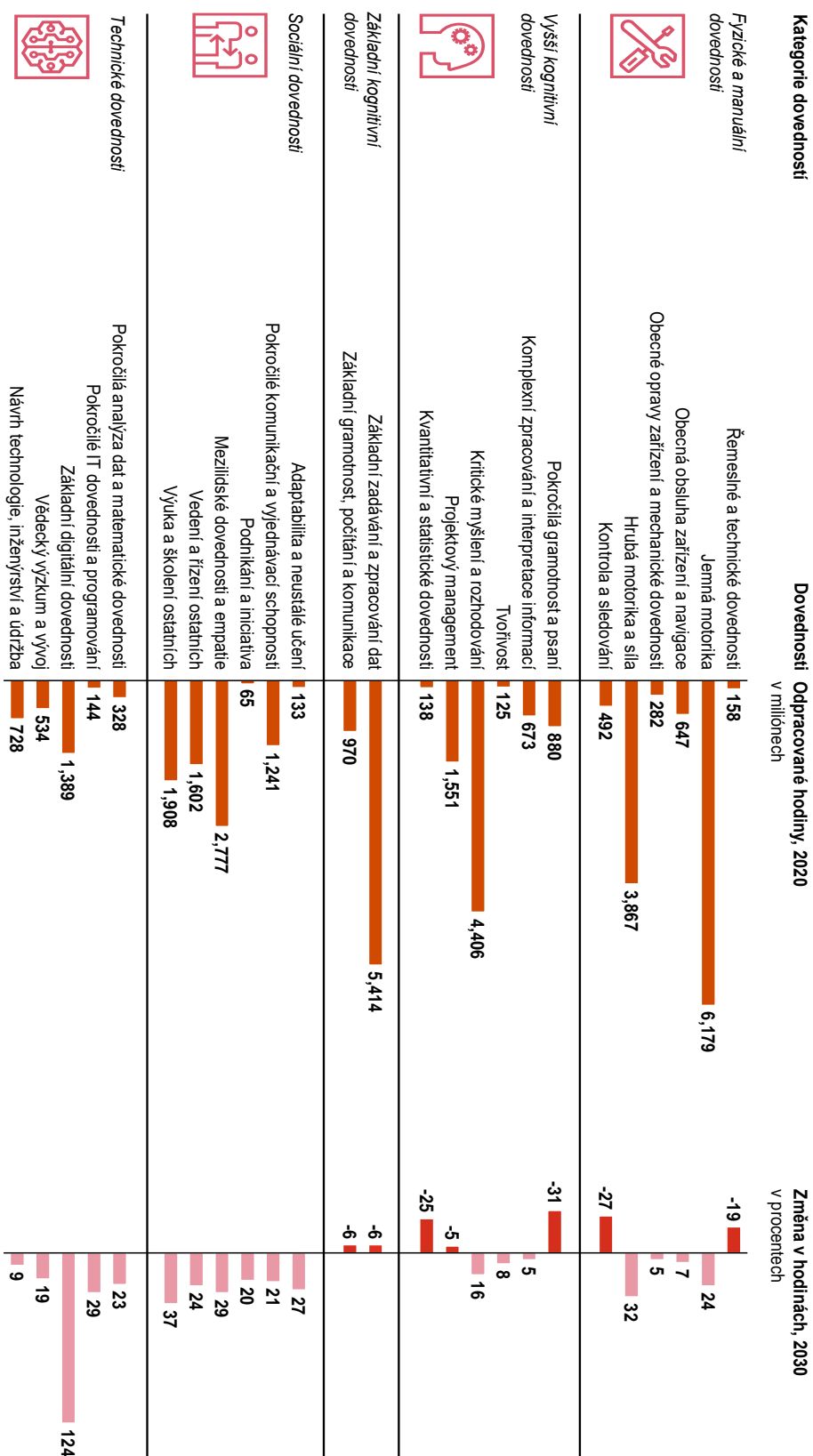
inteligenci, ale také posoudit, kdy jsou vhodná pro pacienty, a zdravotníci v první linii je musí pacientům vysvětlit. To vyžaduje kombinaci digitální gramotnosti a dovedností kritického myšlení spíše než specifické kvalifikace v oblasti statistiky – a tyto dovednosti potřebuje veškerý zdravotnický personál.

Dokument McKinsey Global Institute v roce 2018 nesoucí název „Skill shift: Automation and the future of the workforce“ analyzoval tyto měnící se požadavky, což přispělo k diskusi o změnách potřebných pro zdravotnickou pracovní sílu. Tato nová zpráva zaujímá výslovně evropskou perspektivu a jako výchozí bod používá sedm různých evropských zemí (Francie, Německo, Bulharsko, Itálie, Portugalsko, Švédsko a Velká Británie). Zdravotníci budou nepochybně potřebovat silnější technologické dovednosti, aby pochopili, jak umělá inteligence funguje na určité úrovni detailů, zatímco datoví vědci a inženýři budou potřebovat některé základní lékařské znalosti a silné komunikační dovednosti, které jim pomohou převést klinické požadavky do technologických řešení. Všichni zaměstnanci budou potřebovat jak digitální, tak genomickou gramotnost. To znamená, že praktici budou potřebovat průběžné odborné znalosti a pokyny k hodnocení nových technologií na základě skutečných důkazů a jejich pochopení platnosti a přesnosti dat.

Graf 3 ilustruje, kolik času budou muset zdravotničtí praktici strávit používáním technologických dovedností. Je pozoruhodné, že čas strávený používáním základních digitálních dovedností se více než zdvojnásobí.

Topol Review určil konkrétní oblasti pro vzdělávání a školení související s umělou inteligencí a robotikou, jako je původ zdravotních dat, kurátorství, integrace a správa, etika umělé inteligence a autonomních systémů a nástrojů a kritické hodnocení a interpretace umělé inteligence a robotických technologií (Muro a spol., 2017). Zdravotní systémy budou muset takový posun umožnit a vyvinout filozofii a program strukturovaného učení, protože pomáhají odborníkům přizpůsobit se rozšiřování nových technologií.

Graf 3: V sektoru zdravotnictví v Evropě dojde do roku 2030 k nejvyššímu růstu poptávky po dovednostech po základních digitálních dovednostech



Vlastní zpracování, zdroj: McKinsey Global Institute. Selected European countries: France, Germany, Hungary, Italy, Portugal, Sweden, UK



## 6.5. Představení nových odborníků ve zdravotnictví

Tato podkapitola se zaměřuje především na vztah mezi umělou inteligencí a dnešními zdravotnickými pracovníky, ale je poučné vzít v úvahu, že předpokládaný růst řešení umělé inteligence naznačuje, že do zdravotnické pracovní síly budou muset vstoupit zcela nové role.

Na průsečíku lékařské a datové vědy se objeví několik rolí. Ty budou vyžadovány pro vývoj algoritmů a jejich přeměnu na produkty, implementaci těchto produktů do klinických pracovních postupů, aplikaci nových přístupů ve velkém měřítku a vytvoření podmínek pro usnadnění aplikace umělé inteligence. Například vedoucí lékařští pracovníci budou muset vytvořit klinicky smysluplnou a vysvětlitelnou umělou inteligenci, která bude obsahovat poznatky a informace, které podpoří rozhodnutí zdravotnických pracovníků a prohloubí jejich porozumění pacientům. To by mohlo zahrnovat přizpůsobení algoritmů ke zvýraznění těch laboratorních výsledků, které mají největší vliv na pravděpodobnost diagnózy a sdílí srovnatelné případy, spíše než pouhé poskytování skóre rizika. Bude také vyžadováno klinické zapojení do vedení produktu, aby bylo možné určit přínos systémů pro podporu rozhodování založených na umělé inteligenci v rámci širších klinických protokolů. Adaptace algoritmů na produkty, které jsou použitelné a zapadají do klinických pracovních postupů, bude konkrétně vyžadovat vedení produktu od starších lékařů. Jsou to lidé, kteří dokážou definovat protokoly, které těží z poznatků umělé inteligence, ale zároveň je dokážou bezpečně a správně používat. Aby toho dosáhli, lídři v oblasti medicíny budou vyžadovat hluboké pochopení potenciálu a omezení výstupů algoritmů.

Návrháři specializující se na interakce člověk-stroj při klinickém rozhodování pomohou vytvořit nové pracovní postupy, které integrují umělou inteligenci. Datoví architekti budou mít zásadní význam při definování toho, jak zaznamenávat, ukládat a strukturovat klinická data, aby algoritmy mohly poskytovat přehled, zatímco lídři v oblasti správy dat a etiky dat budou také hrát zásadní roli. V jiných oblastech bohatých na data, jako je genomika, budou noví odborníci zahrnovat „hybridní“ role, jako jsou kliničtí bioinformatičtí, specialisté na genomickou medicínu a genomičtí poradci.

Instituce budou muset vytvořit týmy s odbornými znalostmi v oblasti partnerství, pořizování a implementace produktů umělé inteligence. Uspořádání zavádění nových specializací pocházejících z datové vědy a inženýrství v rámci poskytování zdravotní

péče se samo o sobě stane kritickou dovedností. Stále důležitější budou také odborníci na etiku a zákonnost používání dat, otázky duševního vlastnictví při práci se třetími stranami a osvědčené postupy v oblasti ochrany osobních údajů a správy dat. A konečně, ačkoli schopnost vyvíjet nové přístupy bude pro centra excelence důležitá, schopnost implementovat tyto přístupy ve velkém bude zásadní, pokud se mají přínosy dostat k většině pacientů. Implementace přístupů umělé inteligence ve velkém je ve skutečnosti stejně náročná jako jejich vývoj, a aby to bylo úspěšné, bude vyžadovat nové role ve školení, předělávání pracovních postupů a předávání výhod umělé inteligence.

Zpráva AHA z roku 2018 popisuje nové pozice, kompetence a sady dovedností, které by umělá inteligence mohla vytvořit ve zdravotnictví, se zaměřením na jednotlivé nemocniční organizace.

*Obrázek 18: Nové pozice, kompetence a dovednosti*



#### **Datový vědec**

Ví, jak funguje AI.

Dokáže navrhnout modely umělé inteligence pro provádění úkolů požadovaných v nemocnici nebo zdravotnickém systému.



#### **AI Engineer**

Vytváří modely umělé inteligence pro provádění úkolů požadovaných v nemocnici nebo zdravotnickém systému.



#### **Data-Governance Expert**

Zajišťuje, aby data byla čistá a přesná, a nastavuje zásady pro shromažďování dat.

Zajišťuje, aby zaměstnanci vykonávali svou práci eticky, chránili soukromí a bezpečnost dat pacientů a dodržovali zásady správy dat nemocnice nebo zdravotnického systému.



### **Data-Entry Expert**

Spravuje, čistí a strukturuje data z celé řady interních a externích zdrojů do systému, který dodává modelům umělé inteligence data, která potřebují k provádění úkolů požadovaných v nemocnici nebo zdravotnickém systému.



### **Data Engineer**

Vytváří systém, který pohání modely umělé inteligence daty potřebnými k provádění úkolů požadovaných v nemocnici nebo zdravotnickém systému.



### **Chief AI Officer**

Vede úsilí o prozkoumání potenciálních příležitostí, vývoj přesvědčivé strategie umělé inteligence a identifikaci finančních prostředků, odborníků, technologií a organizačních zdrojů k jejich implementaci.

Rozumí klinickému pracovnímu postupu, pracovní síle a kultuře, která řídí poskytování péče.

*Vlastní zpracování, zdroj: <https://www.aha.org/>*

Mnoho nových rolí bude obsazeno absolventy vědeckých oborů, matematiky a informatiky, kteří budou plnit role ve výpočetní genomice, datové vědě, informatice veřejného zdraví a bioinformatice. Tito odborníci budou zakotveni napříč lékařskými specializacemi prostřednictvím jmenování bioinformatiků na úrovni konzultantů, a také napříč vědeckými specializacemi a v prostředí primární péče. Topol Review naznačuje, že existuje naléhavá potřeba, aby zdravotnické systémy přilákaly a udržely tak vzácné a cenné talenty, například vytvořením flexibilních a vzrušujících kariérních cest a jasných cest k vedoucím rolím ve zdravotnictví.

## Závěr

Zdravotnictví se řadí k úspěšným příběhům naší doby. Na druhou stranu se aktuálně nachází v krizi. Obrovský pokrok v oblasti veřejného zdraví a prevence, biomedicínské znalosti, poskytování zdravotní péče a technologie přispěly ke zvýšení průměrné délky života a očekávání od zdravotní péče samotné. Náklady na zdravotní péči stále exponenciálně rostou a překračují růst HDP, a přesto musíme poskytovat lepší, rychlejší a nákladově efektivnější péči většímu počtu lidí.

Nejen, že nemáme finanční zdroje na splnění tohoto slibu, ale stále více chybí i lidské zdroje, a to především ve zdravotnictví. WHO odhaduje, že do roku 2030 bude světu chybět 9,9 milionu lékařů, sester a porodních asistentek.

Umělá inteligence a automatizace má potenciál změnit způsob poskytování péče – řeší potřebu lepší a nákladově efektivnější péči a pomáhají zaplnit část nedostatku odborných zdravotních zaměstnanců. Je tomu tak zejména z důvodu stárnutí populace a zvyšující se potřebě po zdravotní péči. Umělá inteligence a automatizace mají jedinečné postavení, aby pomohly pochopit tyto potřeby a složitou vzájemnou závislost mezi různými faktory, které ovlivňují zdraví populace. Mimořádný posun od lékařské praxe založené na symptomech k molekulární a buněčné medicíně navíc generuje stále větší množství dat. V takovém prostředí může umělá inteligence přidat hodnotu a pomoci urychlit nové biomedicínské objevy, diagnostické procesy a přístup k léčbě.

Umělá inteligence je stále ještě daleko od využití veškerého svého potenciálu, ale mezi předními zdravotnickými lídry, společnostmi zaměřujícími se na umělou inteligenci a investory, panuje shoda v tom, že potenciál umělé inteligence ve zdravotnictví zůstává významný.

Tato diplomová práce zhodnotila prostředí začínajících aplikací ve zdravotnictví s umělou inteligencí a diskutovala o dopadech, které by rozšířením umělé inteligence mohly nastat. Lze konstatovat, že potenciál umělé inteligence proměnit zdravotní péči platí napříč celým rámcem, ať už mluvíme o sebestarší a prevenci, třídění a diagnóze, podpoře klinického rozhodování, poskytování péče ve zdravotnických jednotkách nebo doma, řízení chronické péče, zdraví populace, provozní zlepšení nebo pokrok v inovacích ve zdravotnictví, například ve výzkumu a vývoji. Existují také oblasti, kde bude umělá inteligence ve zdravotnictví stále více dělat věci rychleji, levněji a v některých případech

i lépe, což by mohlo vést k významným změnám ve výsledcích péče, zkušenostech pacientů a přístupu k nim.

Důsledky využití umělé inteligence ve zdravotnických zařízeních na práci praktiků se začínají již objevovat. Dle provedeného výzkumu lze konstatovat, že zdravotnictví je jedním z odvětví nejméně náchylných k automatizaci.

Zatímco ve vybraných zemích EU je dnes 42 % času stráveného současnými pracovními činnostmi napříč všemi průmyslovými odvětvími technicky automatizovatelných, ve zdravotnictví je to jen 15 % času.

Analýza poptávky po zdrojích ve zdravotní péči pro různorodou skupinu evropských zemí, i když vezmeme v úvahu potenciální dopad automatizace, ukazuje, že do roku 2030 bude stále existovat značný nedostatek odborné zdravotní pracovní síly ve zdravotnictví.

Většinu lékařů umělá inteligence pravděpodobně nepřipraví o práci, ale jejich profesní život se změní.

Závěrem lze dodat, že navzdory rostoucí pozornosti věnované umělé inteligenci ve zdravotnictví zůstává její zavádění v Evropě pomalé: 44 % zdravotnických pracovníků v průzkumu, což je skupina, která se neúměrně zapojuje do inovací v rámci sítě EIT Health, se nikdy nepodílelo na vývoji ani zavádění umělé inteligence v jejich organizaci. Transparentnost a spolupráce inovátorů a odborníků z praxe bude klíčem k rozšíření umělé inteligence v evropském zdravotnictví. Zavedení umělé inteligence navazuje na digitální transformaci a existují významné paralely a poznatky z tohoto procesu, které nabízejí varovný příběh o tom, co je zapotřebí k přijetí a rozšíření inovací ve zdravotnictví.

# Seznam použité literatury

## Odborná literatura a odborné články:

1. „A future that works: Automation, employment and productivity”, McKinsey Global Institute, January 2017; “Jobs lost, jobs gained: Workforce transitions in a time of automation,” McKinsey Global Institute, December 2017; “Skill shift: Automation and the future of the workforce”, McKinsey Global Institute, May 2018.
2. Banerjee, S., Hemphill, T., & Longstreet, P. (2018). Wearable devices and Healthcare: Data sharing and Privacy. The information society.
3. Cadwalladr, C., & Graham-Harrison, E. (2018). Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. The guardian, 17, 22.
4. Dr. Taj Hassan, President of the Royal College of Emergency Medicine, quoted in “NHS England urgently needs 2,200 more A&E consultants”, The Guardian, July 19, 2017.
5. Glauner, P., & Meira, J. A. (2018). Machine Learning for Data Driven Smart Grid Applications.
6. Henon, H., Pasquier, F. Durieu, M. D., Godefroy, O., & Lucas, C., (1997). Preexisting dementia in stroke patients: baseline frequency, associated factors, and outcome. Stroke, 28 (12), 2429–2436.
7. Holcomb, S. D., Porter, W. K., Ault, S. V., Mao, G., & Wang, J. (2018 Březen). Overview on deepmind and its alphago zero ai. In Proceedings of the 2018 international conference on big data and Education (pp. 67–71).
8. Kayyali, B., Knott, D., & Van Kuiken, S. (2013). The big-data revolution in US Health care: Accelerating value and Innovation. McKinsey & Company, 2(8), 1–13.
9. Lucas, L., & Waters, R. (2018). The AI arms race: China and US compete to dominate big data.
10. Mark Britnell, Human: Solving the global workforce crisis in healthcare, Oxford University Press, 2019.
11. McDonald, M.P., & Rowsell-Jones, A. (2012). The digital edge. Gartner, Incorporated.

12. Muro, Mark et al., Digitization and the American workforce, Brookings Institute, November 2017.
13. PANESAR, A., & Panesar, H. *Artificial Intelligence and Machine Learning in Global Healthcare*. Handbook of Global Health, (2020).
14. Panesar, A., Machine Learning and AI for Healthcare, ISBN-13: 978-1-4842-6536-9, (2021).
15. PANESAR, Arjun, & Panesar, H. (2021). *Machine Learning and AI for Healthcare*. Second Edition, ISBN 978-1-4842-6536-9, (2021).
16. Quarterly, M. (2009). What Health Systems Can Learn From Kaiser Permanente: An Interview With Hal Wolff.
17. Renaud, B., Labarere, J. Coma, E., Santin, A., Hayon, J., Gurgui, M., .... & Salloum, M. (2009). Risk stratification of early admission to the intensive care unit of patients with no major criteria of severe community-acquired pneumonia: development of an international prediction rule. *Critical care*, 13(2), R54.
18. Roham, M., Saldivar, E., Raghavan, S., Zurcher, M., Mack, J., & Mehregany, M. (2011, March). A mobile wearable wireless fetal monitoring system. In 2011 5<sup>th</sup> International Symposium on Medical Information and Communication Technology (pp. 135-138). IEEE.
19. Štědroň B. a kol., Prognostické metody a jejich aplikace, C.H.BECK, Praha 2012.
20. Štědroň B. a kol., Prognostika, C.H.BECK, Praha 2019.
21. Štědroň B. a kol., Právo a umělá inteligence, A. Čeněk, Praha 2020.
22. Taddeo, M., & Floridi, L. (2016). The debate on the moral responsibilities of online service providers. *Science and Engineering Ethics*, 22(6), 1575–1603.
23. The European countries analysed are France, Germany, Hungary, Italy, Portugal, Sweden and the UK, McKinsey Global Institute.
24. The Topol Review: Preparing the healthcare workforce to deliver the digital future, NHS, February 2019.
25. The Topol Review: Preparing the healthcare workforce to deliver the digital future, NHS, February 2019.
26. Trifu, M.R. & Ivan, M.L. (2014). Big Data: present and future. *Database Systems Journal*, 5(1), 32–41.
27. Voigt, P., & Von dem Bussche, A. (2017). *The EU general data protection regulation (GDPR). A Practical Guide*, 1st Ed., Cham: Springer International Publishing.

28. Wallace, P.J. Shah, N.D., Dennen, T., Bleicher, P.A., & Crown, W.H. (2014). Optum Labs: building a novel node in the learning Health care system. *Health Affairs*, 33(7), 1187–1194.
29. Westbrook, Johanna I., et al., “How much time do nurses have for patients? a longitudinal study quantifying hospital nurses’ patterns of task time distribution and interactions with health professionals”, *BMC health Ser Res*, 2011, vol. 11.
30. Winfield, A.F. (2016). Written evidence submitted to the UK Parliamentary Select Committee on Science and Technology Inquiry on Robotics and Artificial Intelligence.

### **Elektronické zdroje**

1. Digital Health, 27 February 2018, <https://www.digitalhealth.net/2018/02/impact-clinical-speech-recognition-emergency-department-south-teeshospitals-nhs-foundation-trust/>.
2. EU 28 average birth rate: 1.6 children per woman in 2019, from 2.5 in 1960. See Demographic outlook for the European Union, 2019, European Parliament May 2019, [http://www.europarl.europa.eu/RegData/etudes/IDAN/2019/637955/EPRS\\_IDA\(2019\)637955\\_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/IDAN/2019/637955/EPRS_IDA(2019)637955_EN.pdf).
3. Exchange of Electronic Health Records across the EU. (2020, April 1). Shapping Europe’s Digital Future – European Commission.
4. Global strategy on human resources for health: Workforce 2030, World Health Organisation, 2016, [https://www.who.int/hrh/resources/pub\\_globstrathrh-2030/en/](https://www.who.int/hrh/resources/pub_globstrathrh-2030/en/), Healthcare personnel statistics – nursing and caring professionals, Eurostat, 2019, [https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare\\_personnel\\_statistics\\_-\\_nursing\\_and\\_caring\\_professionals#Healthcare\\_personnel](https://ec.europa.eu/eurostat/statistics-explained/index.php/Healthcare_personnel_statistics_-_nursing_and_caring_professionals#Healthcare_personnel).
5. GovInsider, <https://govinsider.asia/smart-gov/exclusive-tan-tock-seng-hospital-builds-artificial-brain-to-manage-services/>.
6. Health Monitoring | Swiss Re. (2020, April 8). Swiss Re.
7. Healthcare IT News, November 13, 2019, <https://www.healthcareitnews.com/news/europe/germany-introduces-digital-supply-act-digitalise-healthcare>.



8. <https://digital-strategy.ec.europa.eu/en/policies/electronic-health-records>.
9. [https://www.swissre.com/institute/research/library/health\\_monitoring\\_dave\\_wang.htm](https://www.swissre.com/institute/research/library/health_monitoring_dave_wang.htm)
10. IBM100 – TAKMI: Bringing Order to Unstructured Data. (2018). <https://www.ibm.com/ibm/history/ibm100/us/en/icons/takmi/>.
11. See also “Rads who use AI will replace rads who don’t”, Stanford University, November 2017, <https://aimi.stanford.edu/news/rsna-2017-rads-who-use-ai-will-replace-rads-who-don-t>.
12. Smarter Business Review – Services Blog. (2020, Srpen). Smarter Business Review: [www.ibm.com/block/services/](http://www.ibm.com/block/services/).
13. The Journal of mHealth, November 29, 2018, <https://thejournalofmhealth.com/nhs-trust-saves-150000-a-year-using-clinical-speech-recognition/>.
14. The Telegraph, 8 December 2015, <https://www.telegraph.co.uk/news/health/news/12037469/Junior-doctors-spend-up-to-70-per-cent-of-time-on-paperwork.html>, Westbrook, Johanna I.
15. Wooding, A. (2018). MTEP Publications. Work Health Economics Consortium. [https://yhec.co.uk/2020/?post\\_type=publication](https://yhec.co.uk/2020/?post_type=publication).

## Seznam obrázků

Obrázek 1 Řízený daty, vztah pacienta s lékařem

Obrázek 2 Tři „V“ dat

Obrázek 3: Nositelné biomedicínské senzory

Obrázek 4: Proudý analytiky

Obrázek 5: Jak strojové učení pracuje

Obrázek 6: Metody strojového učení

Obrázek 7: Supervised learning

Obrázek 8: Unsupervised learning

Obrázek 9: Reinforcement learning

Obrázek 10: Konfigurace umělého neuronu

Obrázek 11: Konfigurace biologického neuronu

Obrázek 12: Organizace umělé neuronové sítě

Obrázek 13: Hluboké učení

Obrázek 14: Dolování dat a jejich synergie

Obrázek 15: Parametrické vs. Neparametrické algoritmy

Obrázek 16: Pracovní postupy spojené s projektem strojového učení

Obrázek 17: Porozumění úkolům ve strojovém učení

Obrázek 18: Nové pozice, kompetence a dovednosti

## **Seznam tabulek**

Tabulka 1: Penetrace a druhy dat ve zdravotnictví

Tabulka 2: Do roku 2030 budou různá povolání v různé míře ovlivněna demografickými faktory a zavedením automatizace a umělé inteligence ve zdravotnictví

## **Seznam grafů**

Graf 1: Podíl odpracovaných hodin, který by mohl být automatizován do roku 2030

Graf 2: Počet hodin odpracovaných sestrami se plošně zvýší

Graf 3: V sektoru zdravotnictví v Evropě dojde do roku 2030 k nejvyššímu růstu poptávky po dovednostech po základních digitálních dovednostech

# Evidence výpůjček

Prohlášení:

Dávám svolení k půjčování této diplomové práce. Uživatel potvrzuje svým podpisem, že bude tuto práci řádně citovat v seznamu použité literatury.

Jméno a příjmení: Dalibor Čápek

V Praze dne: 25.04.2022

Podpis:

Jméno	Oddělení/Pracoviště	Datum	Podpis