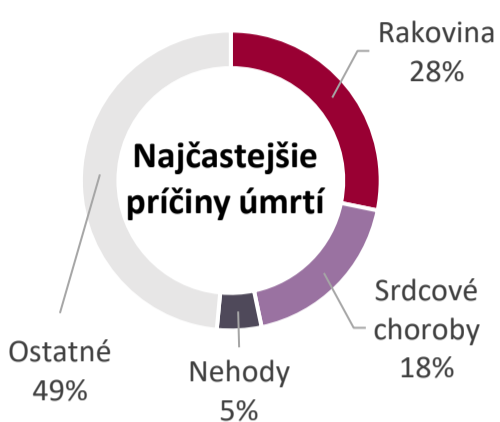


Využitie strojového učenia pri diagnostikovaní chorôb

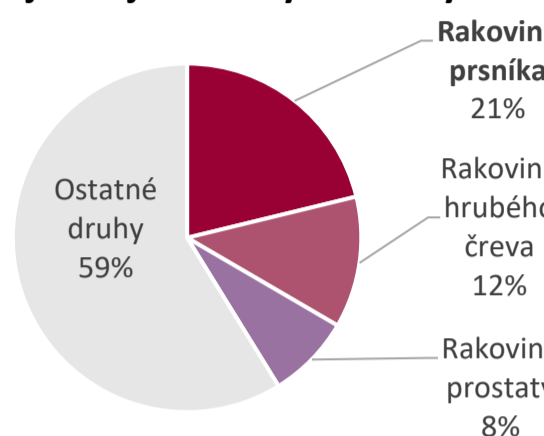


Motivácia

Rakovina ročne zabíja milióny ľudí oboch pohlaví. Každoročne sa viac ako 2,5 miliónom ženám diagnostikuje rakovina prsníka, z toho viac ako 680 tisíc žien na ňu za rok zomrie.

Včasná detekcia správneho typu rakoviny môže výrazne redukovať počet úmrtí a zvýšiť úspešnosť výberu správnej liečby.

Najčastejšie druhy rakoviny



Python TensorFlow Keras Google Colab Pro cross Validation

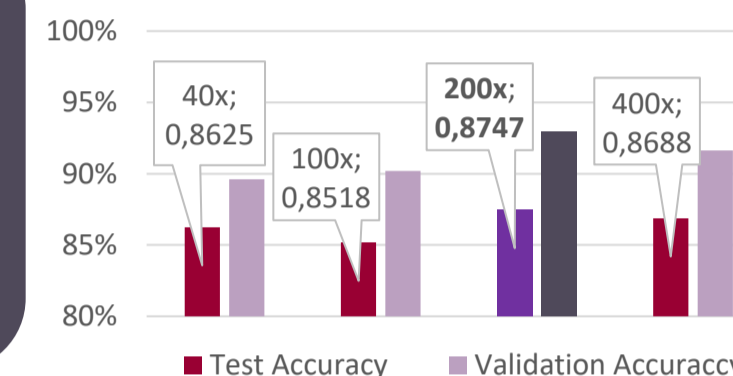
Klasifikácia a výsledky

Binárna klasifikácia

Prvá klasifikácia pozostávala z klasifikácie snímok do 2 tried - benígne a malígne snímky. Následne sme použili 27 krížovú validáciu (CV) ako dôkaz úspešnosti klasifikácie pre jednotlivé zväčšenia. 27 CV pozostávala z rozdelenia dát na tréningovú a testovaciu množinu. Tréningová množina sa rozdelila postupným spôsobom na 26 tréningových častí a 1 validačnú, podľa ktorej sa tréning zastavilo pomocou techniky "early stopping", aby nedošlo k pretrénovaniu modelu.

Po vykonaní experimentov sa ukázalo, že snímky s 200x zväčšením dávajú najlepšie výsledky spomedzi ostatných zväčšení. Tréningovanie modelov pre túto fázu trvalo viac ako 170 hodín.

Binárna klasifikácia



Ciel projektu

Našou snahou je predpovedať typ rakoviny zo snímok už v skorom štádiu pomocou umelých neurónových sietí, aby bolo možné sa vyhnúť invazívnym diagnostickým prístupom.

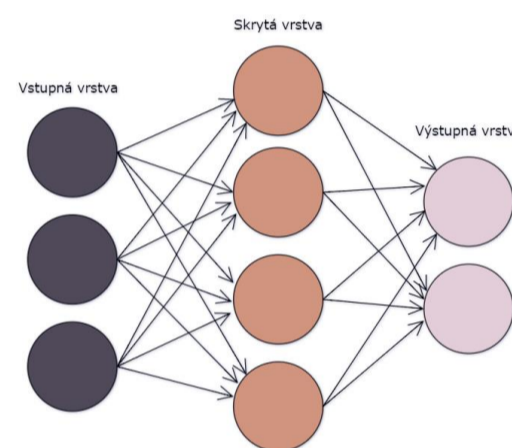
Metodológia

Neurónová sieť

Výpočtový model v oblasti umelej inteligencie zostavený na základe abstrakcie vlastností biologických nervových systémov. Obsahuje vstupnú vrstvu, skryté vrstvy a výstupnú vrstvu na predikciu výsledkov pre daný problém.

Konvolučná sieť

Najväčšie využitie je pri spracovaní obrazových dát. Jej základnou jednotkou je matematická operácia konvolúcie. Konvolúcia nám umožňuje zachytávať určité črty, ktoré abstrahujú kľúčové informácie pre finálnu predikciu. Pri konvolučnej sieti sa namiesto neurónov aktualizujú váhy filtračného jadra (kernelu).



Konvolúcia

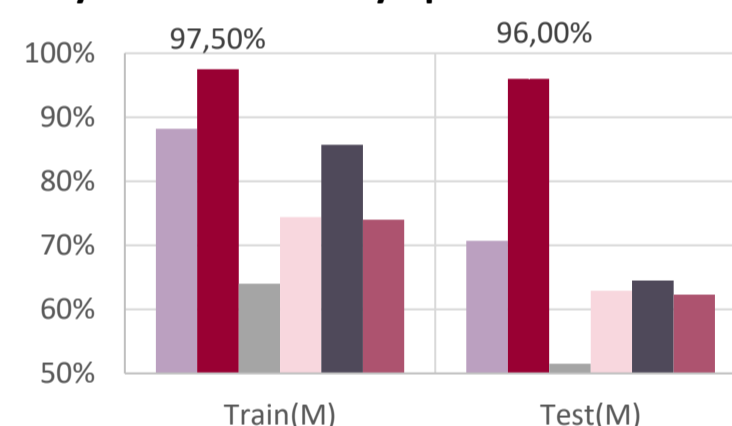
7	15	4
0	1	9
17	5	3

1	3
2	4

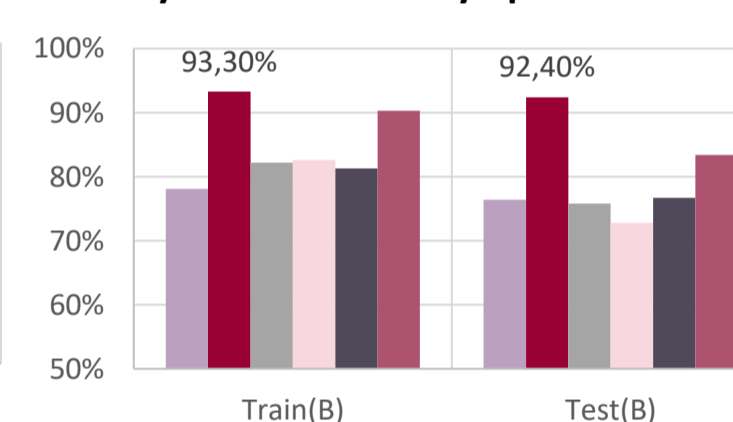
Filtračné jadro (kernel)

56	65
57	50

Klasifikácia MALÍGNÝCH vzoriek s využitím techník vylepšenia snímok



Klasifikácia BENÍGNÝCH vzoriek s využitím techník vylepšenia snímok



Klasifikácia podtypov

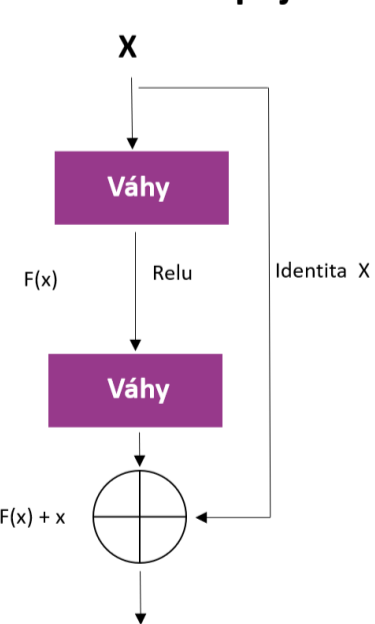
Na vzorkách s najlepším zväčšením pri binárnej klasifikácii sa vykonávali klasifikácie podtypov malígnych i benígnych vzoriek. Najlepšie výsledky boli dosiahnuté pomocou techniky posuvného okna.

Vykonávali sa experimenty s použitím techník vylepšenia snímok pomocou:

- augmentácie
- normalizácie (3 rôzne typy)
- posuvného okna (sliding window)

Tréningovanie 30 sietí pre určovanie podtypov bolo vykonané za 250 hodín.

Reziduálne spojenie



ResNet-152

Model zahŕňa 152 vrstiev. Je to dopredná konvolučná neurónová sieť, ktorá obsahuje reziduálny typ spojenia. Tento typ znamená funkciu preskočenia určitých vrstiev - zabráňuje sa stagnácii a urýchľuje proces učenia. Takýto model umožňuje tréningovať hlboké neurónové siete s menšou náchylnosťou k overfittingu.

Transfer learning

Spôsob použitia natrénovaného modelu na iný problém, ktorý transformujeme na naše použitie pridaním vlastných vrstiev. Z praktického hľadiska opätovné použitie alebo prenos informácií z predchádzajúcich naučených úloh má potenciál výrazne zlepšiť efektívnosť tréningovania nových modelov. Natrénovaný model často slúži na extrakciu určitých vlastností, ktoré sú vstupom do vlastných nadefinovaných vrstiev.

Dáta

Ako vstup do klasifikačných modelov je využitý súbor 7992 histopatologických snímok pacientiek s rakovinou prsníka v 4 zväčšeniach (40x, 100x, 200x a 400x). Vzorky sú rozdelené do 2 hlavných skupín a prislúchajúcich podtypov: malígne a benígne vzorky.

Použité techniky vylepšenia snímok

1. Augmentácia

Metóda zmeny snímok pre znásobenie počtu snímok v súbore. Využíva sa pridávanie kontrastu alebo jas, pridanie šumu, otočenie vertikálne alebo horizontálne, a ďalšie.

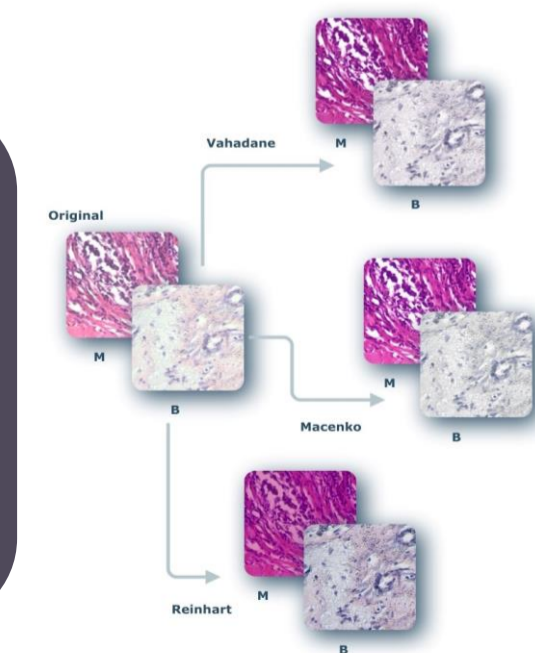
2. Normalizácia

Metódy snažiac sa prispôsobiť vlastnosti základnej vzorky na celý súbor snímok, a tak odstrániť veľké odlišnosti v celom súbore.

Použité metódy: Vahadane, Macenko a Reinhard.

3. Technika posuvného okna

Metóda zväčšujúca počet snímok súboru tak, že zoberie zo snímky určitú časť a posuvným efektom sa z nej snaží vytvárať ďalšie snímky.



Malígne podtypy		Benígne podtypy	
ductal carcinoma	lobular carcinoma	adenosis	fibroadenoma
mucinous carcinoma	papillary carcinoma	phyllodes tumor	tubular adenoma