ŽILINSKÁ UNIVERZITA V ŽILINE

FAKULTA RIADENIA A INFORMATIKY

DIPLOMOVÁ PRÁCA

TOMÁŠ ŠTULRAJTER

Klasifikácia mamografických nálezov pomocou umelej inteligencie

Vedúci práce: prof. Mgr. Ivan Cimrák, Dr.

Ministerské číslo práce: 28360020252124

Registračné číslo: 2647/2023

Žilina, 2025

ŽILINSKÁ UNIVERZITA V ŽILINE

FAKULTA RIADENIA A INFORMATIKY

DIPLOMOVÁ PRÁCA

ŠTUDIJNÝ ODBOR: BIOMEDICÍNSKA INFORMATIKA

TOMÁŠ ŠTULRAJTER

Klasifikácia mamografických nálezov pomocou umelej inteligencie

Žilinská univerzita v Žiline

Fakulta riadenia a informatiky

Katedra softvérových technológií

Žilina, 2025

ŽILINSKÁ UNIVERZITA V ŽILINE, FAKULTA RIADENIA A INFORMATIKY.

ZADANIE TÉMY DIPLOMOVEJ PRÁCE.

Študijný program: Biomedicínska informatika

Meno a priezvisko
Tomáš Štulrajter

Osobné číslo
561537

Názov práce v slovenskom aj anglickom jazyku

Klasifikácia mamografických nálezov pomocou umelej inteligencie

Classification of mammography images by artificial intelligence

Zadanie úlohy, ciele, pokyny pre vypracovanie (Ak je málo miesta, použite opačnú stranu)

Cieľ diplomovej práce:
Mamografia generuje rontgenové snímky zobrazujúce tkanivo, z ktorých lekári môžu identifikovať zhubné alebo nezhubné nálezy. V existujúcich databázach CBIS-DDSM a OPTIMAM sú obsiahnuté snímky s anotáciami nálezov a ich klasifikáciou.
Cieľom diplomovej práce je zostavenie neurónovej siete, ktorá bude brať do úvahy kontext, čo znamená širšie okolie nálezu a nielen samotné výrezy s bezprostredným okolím nálezu. Neurónová sieť bude mať za úlohu klasifikovať nález, prípadne klasifikovať celú snímku.

Obsah:
1. Zoznámenie sa s datasetom, ktorý je už k dispozícii, tvorba skriptov pracujúcich s datasetom
2. Zostavenie vhodnej architektúry neurónovej siete so zapojením kontextu
3. Príprava datasetu podľa vybranej architektúry
4. Trénovanie siete a následné doľaďovanie siete pre získanie najvyššej presnosti siete
5. Analýza výsledkov a návrh vhodných zmien v riešení vedúcim k potenciálnemu zlepšeniu.

Meno a pracovisko vedúceho DP: prof. Mgr. Ivan Cimrák, Dr., KST, ŽU
Meno a pracovisko tútora DP:

garant štud. programu
(dátum a podpis)

Zadanie zaregistrované dňa 31. 10. 2023 pod číslom 2647/2023 podpis _____

Čestné vyhlásenie

Prehlasujem, že diplomovú prácu som vypracoval samostatne s použitím uvedenej literatúry a vlastných vedomostí.

V Žiline, dňa 28.04.2025

Poďakovanie

Chcem sa poďakovať svojmu vedúcemu práce, prof. Mgr. Ivanovi Cimrákovi, Dr., za jeho podporu, cenné rady, trpezlivé vedenie počas celého priebehu práce a ľudskosť prejavenú v ťažkých časoch.

ABSTRAKT V ŠTÁTNOM JAZYKU

ŠTULRAJTER, Tomáš: *Klasifikácia mamografických nálezov pomocou umelej inteligencie* [Diplomová práca]. – Žilinská univerzita v Žiline. Fakulta riadenia a informatiky. – Školiteľ/Vedúci: prof. Mgr. Ivan Cimrák, Dr. – Stupeň odbornej kvalifikácie: inžinier. – Žilina: FRI UNIZA, 2025. Počet strán: 92

Práca sa zaoberá trénovaním reziduálnej konvolučnej neurónovej siete ResNet50 na výrezoch z mamografických snímok získaných z voľne dostupných databáz CBIS-DDSM a EMBED. Predstavili sme nový prístup k modifikácii vstupných dát spočívajúci v zahrnutí širšieho okolia nálezu (tzv. kontextu) priamo do pôvodnej obrazovej informácie pomocou manipulácie farebných kanálov výrezu. Dokázali sme, že táto úprava dokáže zlepšiť presnosť klasifikácie zhubných a nezhubných nálezov. Odhalili a popísali sme rôznorodé vplyvy kontextu na výsledný výstup trénovacieho procesu. Venovali sme sa taktiež vývoju univerzálne použiteľných skriptov pre prácu s mamografickými databázami.

Kľúčové slová: mamografia, databáza, CBIS-DDSM, EMBED, reziduálna konvolučná neurónová sieť, dátová augmentácia

ABSTRAKT V CUDZOM JAZYKU

The aim of this thesis is training a residual convolutional neural network ResNet50 on patches from mammographic images extracted from publicly available databases CBIS-DDSM and EMBED. We introduced a new approach to input data modification, involving the inclusion of a wider area around the region of interest („context“) directly into the original visual information by manipulating the color channels of the patch. We showed that this modification can improve the accuracy for benign and malignant sample classification. We described various effects of context on the overall results of the training process. We also developed generally applicable scripts for working with mammography databases.

Key words: mammography, database, CBIS-DDSM, EMBED, residual convolutional neural network, data augmentation

Obsah

Zoznam obrázkov

Zoznam skratiek

CBIS-DDSM	Curated Breast Imaging Subset of DDSM
CNN	Convolutional Neural Network
CSV	Comma-separated values
DICOM	Digital Imaging and Communications in Medicine
EMBED	Emory Breast Imaging Dataset
ROI	Region of Interest

1 Mamografia

Mamografické vyšetrenia zaujímajú v oblasti verejného zdravia stále dôležitú úlohu. Toto neinvazívne röntgenové vyšetrenie je aj v dnešnej dobe spoľahlivým nástrojom na zachytenie ranných štádií rakoviny prsníka, a tým pomáha pri prevencii a zvyšuje šance na úspešné vyliečenie ochorenia. Štatistické výskumy dokázali, že včasné odhalenie ochorenia výrazne zvyšuje šancu na prežitie, najmä v intervale päť rokov od diagnostiky[1].
Rakovina prsníka už dlhodobo predstavuje jednu z najčastejších foriem rakoviny u žien, pričom v Európskej únii je táto forma zodpovedná za takmer každý tretí nový prípad výskytu rakoviny u tejto demografickej skupiny[2]. Celosvetovo ide u žien o najčastejšiu formu rakoviny v 157 zo 185 štátov[3]. Príčiny vzniku tejto rakoviny sú rôzne a môžu súvisieť s vekom, obezitou, históriou výskytu v rodine aj genetickými faktormi.

Nemenej dôležitá než samotná metóda je aj potreba spoľahlivej interpretácie a ohodnotenia výsledkov vyšetrenia. Historicky a aj v súčasnej dobe túto úlohu zastávali a stále zastávajú lekári – odborníci. Skúsení rádiológovia sú v procese analýzy röntgenových snímok cvičení niekoľko rokov, a zostávajú nenahraditeľným prvkom celého systému diagnostiky.

S rozvojom strojového učenia a umelej inteligencie vo forme neurónových sietí však dostali lekári nového pomocníka. Súčasné modely dokážu vykonávať ako úlohu vyhľadávania nálezov na mamografickej snímke, tak aj klasifikáciu a diagnostiku nájdeného nálezu, čím šetria lekárom čas a pomáhajú im v rozhodovaní. Odhaľovacie a diagnostické schopnosti týchto nástrojov však stále nedosahujú požadovanej kvality, najmä na obmedzených a menej kvalitných datasetoch. Vynaliezanie akýchkoľvek nových metód a metodík na zlepšovanie ich efektivity je preto stále veľmi žiadúca a pre odbor dôležitá činnosť.

Mamografia používa tzv. „mäkké“ röntgenové žiarenie, teda jeho menej energetickú formu, na zobrazenie vnútorných tkanív a žliaz. Vykonáva sa tak, že sa prsné tkanivo umiestni medzi dve priliehajúce platne, a priestorom medzi nimi sa vyšle riadený röntgenový lúč. Rôzne typy tkaniva (ako je tukové, žľazové, svalové, rakovinové) pohlcujú toto žiarenie v rôznej miere, a spôsobujú výskyt tmavších a svetlejších miest na výslednom röntgenovom snímku. Riedke tkanivá (ako napr. tukové) pohlcujú žiarenie v menšej miere,

a prejavujú sa na snímke ako tmavšie oblasti. Naopak, hustejšie tkanivo (ako napr. žľazové a väzivové, ale typicky aj rakovinový nález) vstrebáva väčšiu porciu žiarenia, a javí sa jasnejšie. Na tomto mieste by sme chceli uviesť, že čitateľ môže v lekárskej literatúre naraziť na pri deskripcii snímok na pojmy „zatienenie" a „presvetlenie" ktoré môže mať v tomto kontexte kontradiktný význam. Presvetlením sa označujú tmavé regióny snímky, zatiaľ čo zatienením naopak svetlejšie oblasti.

So spôsobom snímania sa spájajú aj niektoré časté nedostatky tejto metódy, kde napr. veľmi husté prsné tkanivo (vyšší obsah väziva) býva často zdrojom falošne pozitívnych výsledkov, ktoré potom vedú k nepotrebným dodatočným vyšetreniam, zvyšujúc radiačnú záťaž pacienta.

Nálezy na mamografických snímkach môžeme vo všeobecnosti rozdeliť podľa ich pôvodu a habitusu, ktorý na snímke vykazujú, na dva typy:

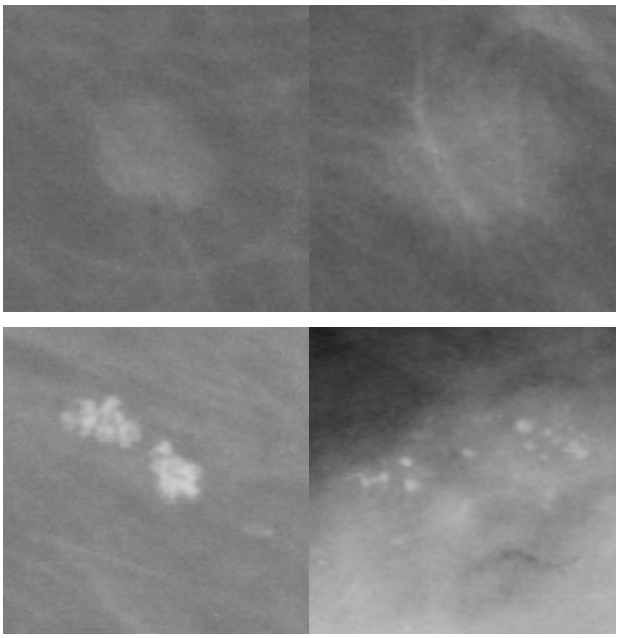
1. **Plošné lokálne zatienenie**
2. **Kalcifikát**

Nález lokálneho ohraničeného plošného zatienenia sa prejavuje na mamografickom snímku ako zosvetlená oblasť s mäkkým, niekedy až nejasným ohraničením a vyčlenením od okolitého tkaniva [4]. Úroveň zatienenia sa môže u jednotlivých nálezov významne líšiť. Tento druh ložiska môže vzniknúť z veľmi rôznorodých príčin, a preto je pri diagnostike dôležité objasniť ich pôvod dodatočnými vyšetreniami, a to hlavne preto, že nálezy rôzneho pôvodu sa môžu na mamografickom snímku javiť podobne až zameniteľne. V prípade benígnych zmien prsníkov, ktoré na snímkach vykazujú symptómy tejto kategórie, ide najčastejšie o nasledujúce javy:

- **Zápaly**
- **Cysty** – dutinové útvary vyplnené tekutinou, sú zväčša príznakom ochorenia fibrocystickej mastopatie.
- **Fibroadenóm** – zložený zo žľazových a väzivových buniek.
- **Lipómy** – tukový benígny nádor, vyskytujúci sa zväčša osamotene.

Druhý typ predstavuje skupinu nálezov, ktoré nazývame kalcifikáty. Jedná sa o telesá kryštalického vápnika, nahromadené v prsnom parenchýme, zväčša sa vyskytujúce v zhlukoch, kde sa javia ako biele bodky, zrná alebo lamely. Podľa veľkosti častíc v zhluku

rozlišujeme mikrokalcifikáty a makrokalcifikáty. Zatiaľ čo makrokalcifikáty nadobúdajú podobu väčších, pevných častíc, mikrokalcifikáty sa zobrazujú ako jemné, malé zrná, pripomínajúce čiastočky soli. S hľadiska diagnózy sú dôležité najmä mikrokalcifikáty, pretože práve tie môžu ukazovať na prítomnosť rakoviny, prípadne signalizovať zmeny v tkanive, ktoré budú pravdepodobne v budúcnosti viesť k rakovine.



Obrázok 1. Ukážka rôznych typov nálezov na prsníku. Vľavo hore - benígny plošný nález, vpravo hore - malígny plošný nález, vľavo dole - benígny kalcifikát, vpravo dole - malígny kalcifikát

Nádorové procesy môžu postihnúť prakticky každú štruktúru a typ tkaniva v prsníku. Svetová Zdravotnícka Organizácia (WHO) zaviedla štandardnú histopatologickú klasifikáciu nádorov, ktorá ich rozdeľuje do 7 základných kategórii, ďalej sa členiac do množstva podtypov v závislosti od konkrétneho typu zasiahnutej bunky. Medzinárodná klasifikácia chorôb (MKCH) zase rozoznáva 10 hlavných kategórii, rozčlenených podľa miesta umiestnenia nádoru.

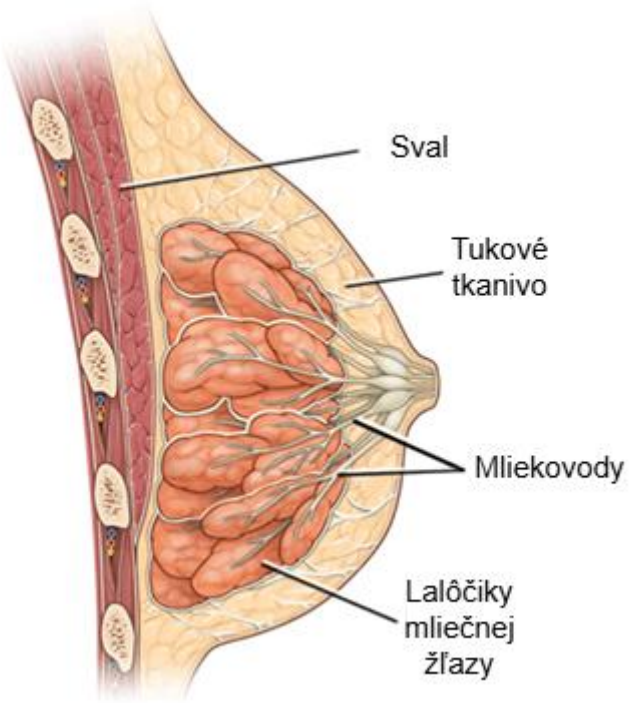
Najčastejšie sa vyskytujúce formy rakoviny sú:

- **Invazívny duktálny karcinóm** – je zodpovedný za približne 50% prípadov u žien. Napáda bunky vystieľajúce vnútro mliekovodov (ductus lactiferus), ktoré zbierajú

produkty mliečnej žľazy a odvádzajú ich k bradavkám. Ide taktiež o najčastejšie sa vyskytujúcu formu u mužských pacientov.

- **Invazívny lobulárny karcinóm** – vzniká v žľazových bunkách mliečnych lalôčikov, ktoré formujú mliečnu žľazu.

Na Obrázku 2 vidíme postranný prierez prsníka so znázornenými hlavnými štruktúrami (Obrázok bol prevzatý z webovej stránky Johns Hopkins Medicine [5], následne prispôsobený)



Obrázok 2. Postranný prierez prsníka

2 Konvolučné neurónové siete

Konvolučné neurónové siete („Convolutional Neural Networks“, ďalej len CNN) predstavujú modifikáciu neurónových sietí, vyvinutých špeciálne za účelom počítačového videnia, teda rozpoznávania a klasifikácie obrazu. Za týmto účelom zavádza oproti klasickým sieťam špecifické modifikácie architektúry. Klasická jednoduchá neurónová sieť sa skladá výhradne z vrstiev plne prepojených neurónov, konvergujúcich do výstupnej vrstvy. Príspevok konvolučných sietí spočíva v modifikácii architektúry, zakomponovaním dvoch dodatočných typov vrstiev:

- **Konvolučná vrstva**
- **„Pooling“ vrstva**

Vstupom do CNN je samotný obrázok, vo forme matice pixelových hodnôt. Dáta so vstupnej vrstvy následne opakovane prechádzajú systémom konvolučných a „poolingových“ vrstiev, ktoré sú zoradené tak, že po konvolučnej vrstve štandardne nasleduje poolingová [6].

Konvolučná vrstva predstavuje v abstraktnom zmysle slova detektor príznakov („feature detector“), teda detektor rôznych vizuálnych vzorcov a atribútov na danom obrázku, napr. hrán alebo okrajov. Detekcia prebieha za pomoci špeciálnych matíc, tzv. filtrov (ang. „filter“, „kernel“). Jedná sa o štvorcovú maticu reálnych čísel, ktorá sa postupne „prikladá“ na jednotlivé časti spracovávaného obrázka, a s korešpondujúcou maticou pixelových hodnôt danej časti obrázka vykoná operáciu konvolúcie - súčinu hodnôt matíc po prvkoch, zakončenú súčtom vzniknutých súčinov. Výsledkom operácie je reálne číslo (Obrázok 3, prevzatý z webovej stránky medium.com [7], upravený).

Obrázok 3. Konvolúcia pixelu [1, 1] pomocou zaostrujúceho filtra

Matica filtra sa presúva nad obrázkom pomocou systému plávajúceho okna, a nad každou sekciou, ktorú práve prekrýva, vykoná vyššie spomenutú operáciu. Výsledkom je výstupný obrázok obsahujúci zvýraznené príznaky, ktoré má filter zachytávať. Významným hyperparametrom konvolučnej vrstvy je veľkosť kroku („stride“), ktorá určuje, množstvo pixelov, o ktoré sa plávajúce okno filtra posunie po vykonaní aktuálnej konvolúcie. Veľkosť kroku sa uplatňuje na horizontálny aj vertikálny pohyb filtra. Platí, že pokiaľ je veľkosť kroku menšia než veľkosť strany matice filtra, tak dva po sebe nasledujúce pozície filtra spracujú úseky obrázka, ktoré sa budú čiastočne prekrývať. Proces konvolúcie uplatnený postupne nad celou šírkou aj dĺžkou obrázka má za následok zníženie rozlíšenia výstupného obrázka.

Filtre, ktoré sa v konvolučných vrstvách najčastejšie používajú, môžeme zaradiť do niekoľkých kategórii:

- **Hranové filtre** – slúžia na detekciu hrán – napr. Sobelov filter (operátor).
- **Zaostrujúce filtre** – zväčšujú rozdiely medzi hodnotami susedných pixelov, a tým zvýrazňujú zachytené príznaky.
- **Zmatňujúce filtre** – opak zaostrujúcich – používajú sa na potlačenie šumu a neželaných artefaktov – napr. Box filter.

V prípade, že má vstupný obrázok viacero farebných kanálov (napr. pri použití farebného modelu RGB), uplatňuje sa tradične na každý kanál odlišný filter.

„Poolingová“ vrstva slúži na ďalšiu redukciu rozlíšenia výstupného obrázka. Používa na to podobnú metódu ako konvolučná vrstva, avšak jej filtre nevykonávajú konvolúciu, ale zväčša nejakú formu agregačnej funkcie. Najčastejšie používanou je „Max Pooling“, ktorá vyberá maximálnu hodnotu pixelu v oblasti, ktorú filter aktuálne zasahuje.

Časté je aj použitie funkcie „Average Pooling“, teda spriemerovanie hodnôt všetkých pixelov nad aktuálnym filtrom.

Redukcia rozlíšenia je užitočným krokom pri trénovaní CNN, pretože prináša viacero výhod:

- Zníženie celkovej výpočtovej zložitosti trénovania – každá ďalšia dvojica konvolučnej a „poolingovej“ vrstvy je progresívne menej výpočtovo zložitá.
- Konvergencia a zahustenie priestorovej informácie do stále menšej pixelovej plochy umožňuje neurónom v hlbších vrstvách „vidieť“ väčšiu časť pôvodného obrazu.
- Zvyšuje invariantnosť CNN voči translácii – aj keď CNN nie sú vo všeobecnosti translačne invariantné[8], tak zníženie rozlíšenia pomáha zachovať najdôležitejšie príznaky, minimalizujúc vplyv drobných pozičných zmien objektov na obrázku.
- Redukcia „overfittingu” – odstránenie šumu a nedôležitých detailov s obrázka pomáha CNN sústrediť sa na všeobecnejšie a dôležité príznaky [9]. Táto vlastnosť je dôležitá najmä pre trénovanie na malých, resp. obmedzených datasetoch.

Následkom tejto špecifickej architektúry dochádza v CNN postupne k detekcii príznakov nižšej úrovne (hrany, rohy, krivky) a ich kombinovaním až k detekcii príznakov vyššej úrovne (konkrétne celistvé tvary, objekty). CNN vo väčšine prípadov vyúsťuje do niekoľkých plne prepojených vrstiev, končiacich výstupnou vrstvou, ktorá určuje výsledok klasifikácie.

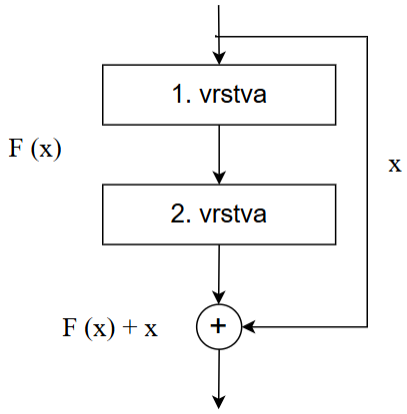
2.1 Reziduálne neurónové siete

Reziduálna neurónová sieť („Residual Neural Network” – skrátene ResNet) predstavuje nadstavbu and klasickou CNN, ktorú upravuje pridaním tzv. reziduálnych spojení („residual connections”, „skip connections”)[10]. Vývoj reziduálnych blokov bol motivovaný negatívnym javom tzv. miznúcich gradientov („Vanishing gradients“) [11], Ide o problém škálovania, ktorý sa prejavuje u sietí, ktoré majú veľké počty skrytých vrstiev („hidden layers“). Je spôsobený tým, že pri spätnej propagácii chyby sa v hlbších vrstvách postupne znižuje gradient, predstavujúci parciálnu deriváciu „loss“ funkcie, ktorá je zodpovedná za optimalizáciu váh medzi neurónmi, s cieľom minimalizovať chybu.

V pokročilých vrstvách môže gradient dosiahnuť veľmi malých hodnôt, čo vyústi do toho, že váhy už nie sú ďalej efektívne aktualizované. Konečným dôsledkom je sieť, ktorá dosahuje horšiu presnosť ako „plytšia“ sieť, teda sieť s menším počtom vrstiev.

Reziduálne konexie adresujú tento problém tým, že priamo prepájajú „skoršie“ vrstvy (teda vrstvy, ktorých poradie v lineárnej sekvencii vrstiev je nižšie) s vrstvami, ktoré sú ďalej v sieti. Výstup z jednej vrstvy je teda týmito skratkami priamo a v nezmenenej podobe (funkcia identity) prenášaný hlbšie do siete. Pri spätnej propagácii chyby potom gradient môže prúdiť aj do ďalekých častí siete, využívajúc práve tieto spojenia.

Na Obrázku 4 znázorňujeme princíp fungovania reziduálneho spojenia.



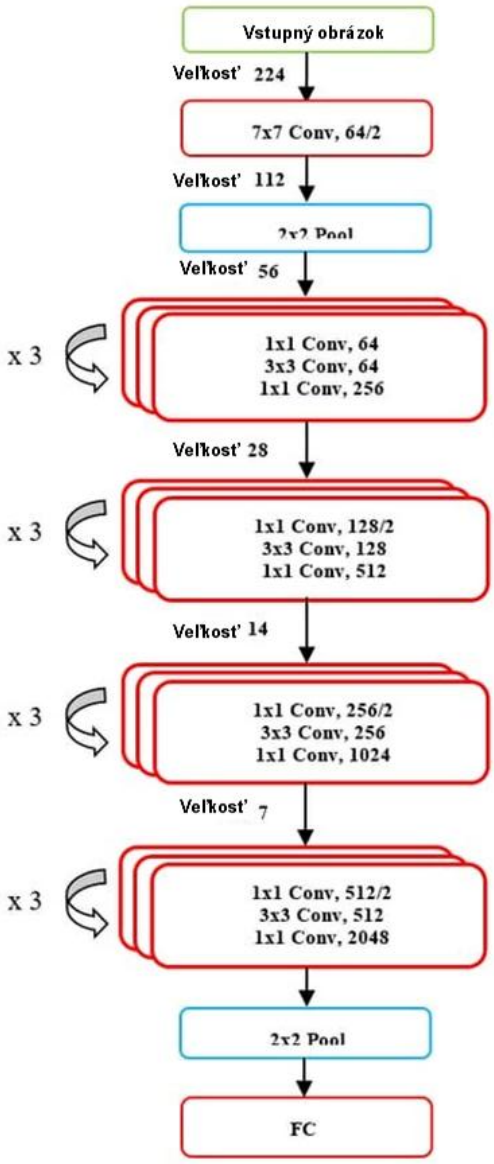
Obrázok 4. Schéma použitia reziduálneho spojenia, prenášajúceho výstup x z predošlej vrstvy do hlbších vrstiev

Postupným vývojom a experimentálnym testovaním sa ustanovili štandardné architektonické varianty reziduálnych sietí. Jednotlivé varianty sa odlišujú najmä počtom vrstiev a charakterom tzv. blokov. Bloky sú zoskupenia konvolučných vrstiev, v závislosti od typu zvyčajne dvoch alebo troch.

V tejto práci sme na trénovanie používali variantu zvanú ResNet50. Niektoré ďalšie verzie rodiny ResNet zahŕňajú ResNet18, ResNet34, ResNet101, ResNet152.

2.2 Architektúra ResNet50

Na Obrázku 5 (prevzatý zo stránky medium.com [12], následne upravený) môžeme vidieť kompletnú architektúru siete ResNet50.



Obrázok 5. Architekúra siete ResNet50 – skratka „Conv" predstavuje konvolučnú vrstvu, skratka „Pool" predstavuje „poolingovú" vrstvu, a „FC" vyjadruje plne prepojenú vrstvu

Vidíme, že ako vstup slúži obrázok veľkosti 224x224 (pri trénovaní môže byť použitý aj obrázok iného rozmeru (viď Kapitola 9.3, kde sa touto problematikou zaoberáme bližšie)), a po absolvovaní série blokov (v prípade ResNet50 pozostáva blok z troch konvolučných vrstiev, spolu s reziduálnym spojením medzi začiatkom bloku a koncom bloku, ktoré prenáša nezmenenú vstupnú informáciu do výstupu). Výstupom z modelu je plochý vektor veľkosti 2048.

Model ResNet50 bol pôvodne natrénovaný na datasete s názvom ImageNet. Ide o rozsiahlu kolekciu obrázkov, pozostávajúcu z 14 197 122 obrázkov, patriacich do 21 841

tried (čísla aktuálne v čase písania práce), ktoré sa v tejto databáze nazývajú „synsety". Model nebol natrénovaný na celom tomto objeme, ale len na podmnožine s názvom ImageNet Large Scale Visual Recognition Challenge (ILSVRC) (1 431 167 položiek, 1 000 tried) [13].

3 Ciele práce

Táto práca sa zaoberá novými prístupmi k vstupným dátam, ktoré sa používajú na trénovanie konvolučných neurónových sietí, s cieľom zlepšiť celkovú kvalitu a presnosť klasifikácie mamografických snímok. Nosnou myšlienkou je modifikácia vstupných dát vo forme zahrnutia tzv. „kontextu", kde sa okrem samotného snímku výrezu s nálezom vyšetrovania použije aj okolie daného nálezu, ako dodatočná vizuálna informácia, ktorú musí sieť pri svojom tréningu spracovať a zohľadniť. Zahrnutie kontextu má za cieľ zachovať detail štruktúry tkaniva, ale zároveň poskytnúť aj širší nadhľad nad nálezom a jeho okolím. Práca používa a experimentuje z rôznymi podobami kontextu, a porovnáva výstupy, v snahe identifikovať formu vstupných dát, ktorá prinesie optimálne výsledky na validačných a testovacích sadách.

Pracovali sme s voľne dostupnými databázami mamografických nálezov CBIS-DDSM a EMBED. Pretože práca s týmito databázami nie je triviálna, sme stanovili ďalšie dva sekundárne ciele: prehľadne a prakticky popísať ich štruktúry, a vypracovať skripty na zjednodušenie automatizácie spracovania ich dát, s prízvukom na naše ciele, ale s ohľadom na dostatočnú všeobecnosť a modularitu, ktorá by dovolila budúcim záujemcom a výskumníkom modifikovať a rozširovať ich pre vlastné potreby.

4 Aktuálny stav výskumu

V tejto kapitole sa budeme venovať aktuálnemu výskumu z oblasti klasifikácie mamografických snímok a tematike mamografických databáz vo všeobecnosti.

V súčasnosti môžeme v doméne klasifikácie pozorovať dva rozdielne prístupy, ktoré sa pri zlepšovaní pri pokusoch o zlepšovanie používajú:

1. **Technologický prístup** - použitie nových alebo špecificky modifikovaných druhov architektúr CNN, prípadne využívaním rôznych hybridných metód spojenia CNN a iných modelov strojového učenia.
2. **Dátový prístup** – skúmanie nových metód modifikácie alebo rozširovania vstupných dát, prípadne kombinovanie vstupných dát z viacerých zdrojov.

Väčšina prác, ktoré sa tejto doméne venujú, patria do prvej kategórie. To nie je prekvapivé, vzhľadom na prudký a rýchly rozvoj metód strojového učenia a umelej inteligencie, ponúkajúci na pravidelnej báze nové modely a architektúry, ktoré je možné na klasifikáciu použiť. Použitie rôznych typov modifikácie alebo augmentácie vstupných je síce bežnou praxou a štandardným postupom, zväčša sa ale jedná o ustanovené, overené metódy (normalizovanie, škálovanie, potlčenie šumu, a pod.). Časté je použitie sietí typu ResNet[14] (poznámka: nakoľko ResNet bol použitý aj v tejto práci, jeho architektúra je podrobnejšie priblížená v Kapitole 2.2) a DenseNet[15], čo sú rodiny príbuzných architektúr predstavujúcich deriváty CNN. Výskum ale prebieha aj na veľkom množstve ďalších modelov, ktoré dosahujú variabilné výsledky a štatistiky na rôznych testovaných databázach. Spomedzi populárnych modelov / rodín modelov uvedieme napr. EfficientNet[16], AmoebaNet[17], Inception-v3[18], Nasnet[19]. Časté je použitie predtrénovaných váh, ktoré vo všeobecnosti urýchľuje trénovanie a môže v niektorých prípadoch zlepšiť celkové výsledky. Databáza DDSM a jej derivácia CBIS-DDSM zaujíma popredné miesto v zozname populárnych a často využívaných databáz, pričom je používaná buď ako jediný zdroj dát alebo je zahrnutá v širšom datasete pozostávajúceho z viacerých databáz.

Relatívne novým trendom je spojenie algoritmov AI s cloudovými výpočtovými technikami a distribuovanými systémami. [20]. Tieto systémy umožňujú efektívne uchovávanie a paralelné spracovávanie veľkého množstva medicínskych dát, poskytujú

rýchly prístup k týmto dátam kdekoľvek na svete s internetovým pripojením, a zvyšujú efektívny výpočtový výkon pre beh klasifikačných algoritmov

Zvyšovanie výpočtových kapacít, spolu s postupným uvedomovaním si dôležitosti rozlíšenia snímku v mamografickej klasifikácii, prinieslo aj možnosti a rozvoj metód celosnímkovej klasifikácie, kedy sa spracováva a vyhodnocuje celá vstupná snímka (alebo jej väčšia časť), namiesto štandardného výrezu ROI. Tieto metódy je možné použiť ako na klasifikáciu, tak na segmentáciu snímok. Aktívna oblasť výskumu sa týka aj vývoja škálovateľných klasifikátorov [21], ktoré dokážu pracovať s výrezmi vyšších rozlíšení tak, aby minimalizovali zvýšenie výpočtovej zložitosti.

S nedávnym nástupom jazykových modelov typu „large language model“ (LLM), ako je napr. séria GPT, založených na transformerovej architektúre[22], dochádza taktiež k rozvoju tzv. vizuálnych transformerov, ktoré prevzali a prispôsobili ich koncepty pre úlohy počítačového videnia, kde sa namiesto malých blokov textu (tzv. „tokenov“) sa použijú malé časti obrázku. Výsledné modely sa potom dajú využiť aj pre klasifikáciu mamografických snímok[23].

Čo sa týka dátového prístupu k optimalizácii výkonu klasifikácie, tak tam sa snaženie sústreďuje na vývoj kvalitnejších a presnejších metód a algoritmov segmentácie obrazu, ktoré dokážu korektnejšie a spoľahlivejšie identifikovať presné obrysy ROI na snímke[24]. Podobný proces prispel aj k vzniku CBIS-DDSM databázy (ktorej sa venujeme aj v tejto práci) ako vylepšenej a štandardizovanej verzie DDSM databázy, kde sa pre detailnejšie upresnenie ROI použil Chan-Vese algoritmus[25]. Experimentuje sa tiež s rôznymi zdrojmi dát, kde sa pre trénovanie a klasifikáciu okrem röntgenových používajú aj ultrazvukové[26] a termografické snímky[27].

Počas zhromažďovania informácií a analýzy aktuálneho výskumu sme identifikovali dve hlavné prekážky, ktoré spomaľujú vývoj v tejto oblasti:

1. **Neprístupnosť dát** – Početné a kvalitné dáta sú základom pre vznik kvalitných diagnostických modelov. Nedostupnosť súkromných databáz a kolísavá kvalita verejne dostupných datasetov môžu prispievať k zhoršeniu konečných výsledkov experimentov. Neexistuje žiadny štandardizovaný formát pre verejné databázy, ktoré sa tak môžu výrazne odlišovať vo svojich veľkostiach, štruktúrach, kvalite a robustnosti metadát alebo metodiky vzniku dát, čo kladie zvýšené nároky na nutnosť spracovania a extrakcie dát pred ich samotným použitím. Častá je

prítomnosť chýb, nezrovnalostí a nekonzistentností v dátach (napr. pomiešanie dát z rôznych klasifikačných tried, rozdielnosť v jase a kontraste jednotlivých snímok), čo mnohokrát vedie k nutnosti aspoň čiastočnej manuálnej verifikácie a prečisteniu extrahovaných dát.

2. **Komplikované porovnávanie** – použitie rozličných metodík na získavanie, spracovávanie a rozširovanie dát z verejných databáz, výrazne komplikuje schopnosť replikovať zistenia rôznych štúdii, validovať ich modely a porovnávať výsledky medzi sebou.

Nedokonalosti v dátach viedli k vzniku samostatnej kategórie prác, venujúcej sa vývoju nástrojov a metód na detekciu abnormalít, ktorých účelom je práve zvýšenie ich kvality. Tieto procesy môžu zahŕňať korekciu snímok, odstránenie duplicitných alebo nepoužiteľných snímok, preusporiadanie datasetu alebo dokonca detekciu dosiaľ nepopísaných nálezov na existujúcich snímkach.

5 Databáza CBIS-DDSM

CBIS-DDSM (skratka pre „Curated Breast Imaging Subset of DDSM“) je databáza snímok mamografických nálezov, ktorá predstavuje updatovanú a štandardizovanú verziu databázy DDSM („Digital Database for Screening Mammography“)[28]. Táto databáza je voľne dostupná a obsahuje 2620 röntgenových snímok, pôvodne zhotovených na analogickom médiu a následne naskenovaných do digitalizovanej podoby. Celkovo sú k dispozícii záznamy od 1566 subjektov, pričom pre niektorých pacientov existujú viaceré záznamy, zodpovedajúce rôznym unikátnym vyšetreniam.

5.1 Logická štruktúra databázy

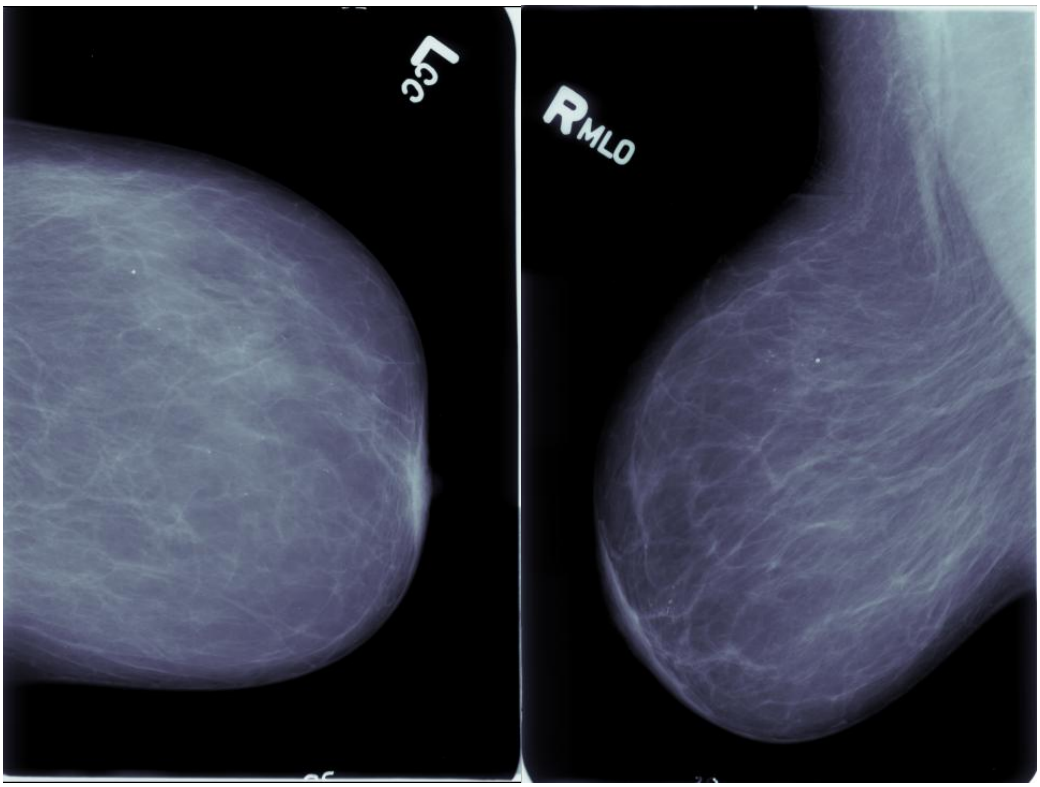
Dataset je štruktúrovaný a delený podľa viacerých kritérií. Základné logické delenie je podľa typu nálezu, kde rozlišujeme dve triedy, značené skrátene anglicky „mass“ a „calc“. Skupina „mass“ predstavuje skupinu plošných nálezov, a skupina „calc“ predstavuje triedu mikrokalcifikátov.

Obidva typy nálezov sú následne rozdelené do dvoch klasifikačných tried, a síce zhubné („malignant“) a nezhubné („benign“) nálezy. S týmito základnými triedami sme operovali aj v tejto práci. Dataset nad to obsahuje navyše ešte triedu „benign without callback“, ktorá sa oproti hlavnej triede nezhubných nálezov odlišuje kontextuálne a sémanticky. Táto trieda predstavuje vyšetrenia, ktoré boli diagnostikované ako nezhubné len na základe posúdenia röntgenového snímku [29]. Vyšetrujúci lekár v tomto prípade mohol označiť niektoré oblasti nálezu ako vhodné na ďalšie sledovanie, ale nenariadil žiadne ďalšie vyšetrenia. Pre účely práce, vzhľadom na zameranie na binárnu klasifikáciu, sme tieto dve triedy nezhubných nálezov zlúčili do jednotnej triedy.

Snímky nachádzajúce sa v databáze môžeme ďalej rozdeliť podľa typu zobrazenia (projekcie) do dvoch typov:

1. **CC** – Kraniokaudálne zobrazenie – röntgenové platne sú umiestnené zvrchu a zospodu prsníka, lúč je vysielaný vo vertikálnej rovine
2. **MLO** – Mediolaterálne šikmé zobrazenie - röntgenové platne sú umiestnené na prsník zľava a sprava, lúč je vysielaný horizontálnej rovine

Na Obrázku 6 vidíme príklady a porovnanie oboch typov snímok. Môžeme si všimnúť, že na projekcii typu MLO sú súčasťou záberu často aj tkanivá steny hrudného koša (zahustená oblasť v pravom hornom rohu MLO snímku).



Obrázok 6. Typy zobrazenia - vľavo CC, vpravo MLO

5.2 Fyzická štruktúra databázy

Aby sme mohli funkcionalitu automatizácie zrealizovať, museli sme sa najskôr podrobne oboznámiť s fyzickou štruktúrou databázy, ktorá sa realizuje na úrovni priečinkovej a súborovej hierarchie. Táto sa od tej logickej v mnohom odlišuje a zároveň nie je kompletne konzistentná a obsahuje výnimky, ktoré celý proces komplikujú. Špecifickými problémami vznikajúcimi pri snahe o automatizáciu procesov manipulácie dát v tejto databáze sa zase budeme zaoberať v Kapitole 5.3, kde predstavíme skripty, pomocou ktorých sme databázu sekvenčne spracovávali.

V nasledujúcej sekcii popíšeme fyzickú štruktúru CBIS-DDSM.

Konečným článkom hierarchického usporiadania jednotlivých datasetov je samotná mamografická snímka. Tá je vždy vo formáte DICOM, („Digital Imaging and Communications in Medicine“), ktorý sa od svojej publikácie stal najpoužívanejším formátom pre uchovávanie, prenos a spracovanie obrazových dát v medicíne.

Formát DICOM vznikol v roku 1993, a dnes je zahrnutý v zbierke štandardov Medzinárodnej organizácie pre normalizáciu („International Organization for Standardization“ – skratka ISO) pod kódom ISO 12052 [30]. V čase písania tejto práce sa štandard nachádzal vo svojej aktualizovanej druhej verzii, ktorá vyšla v roku 2017. Úplná väčšina súčasných obrazových diagnostických nástrojov na báze röntgenového žiarenia, CT, MRI a ultrazvuku má zabudovanú podporu pre tento formát. DICOM súbor obsahuje okrem samotnej obrazovej informácie aj hlavičku, obsahujúcu dodatočné informácie o snímke vo forme tzv. „tagov”. Tieto môžu niesť údaje o pacientovi, vykonávajúcom lekárovi, technické parametre vyšetrenia, vyšetrovacieho prístroja a pod. Detail hlavičky môžeme vidieť na Obrázku 7.

TAG	VR	SIZE	NAME	DATA
(0002,0000)	UL	4	Group size	194
(0002,0001)	OB	2	File Meta Information Version	0\1
(0002,0002)	UI	26	Media Storage SOP Class UID	1.2.840.10008.5.1.4.1.1.7
(0002,0003)	UI	64	Media Storage SOP Instance UID	1.3.6.1.4.1.9590.100.1.2.378776769711002686209961163603436313519
(0002,0010)	UI	18	Transfer Syntax UID	1.2.840.10008.1.2
(0002,0012)	UI	18	Implementation Class UID	1.2.40.0.13.1.1.1
(0002,0013)	SH	14	Implementation Version Name	dcm4che-1.4.35
(0008,0005)		10	Specific Character Set	ISO_IR 100
(0008,0016)		26	SOP Class UID	1.2.840.10008.5.1.4.1.1.7
(0008,0018)		64	SOP Instance UID	1.3.6.1.4.1.9590.100.1.2.378776769711002686209961163603436313519
(0008,0020)		8	Study Date	20160807
(0008,0023)		8	Content Date	20160503
(0008,0030)		6	Study Time	161037
(0008,0033)		14	Content Time	105435.478000
(0008,0050)		0	Accession Number	
(0008,0060)		2	Modality	MG
(0008,0064)		4	Conversion Type	WSD
(0008,0090)		0	Referring Physician's Name	
(0008,103E)		22	Series Description	full mammogram images
(0010,0010)		30	Patient's Name	Calc-Training_P_00005_RIGHT_CC
(0010,0020)		30	Patient ID	Calc-Training_P_00005_RIGHT_CC
(0010,0030)		0	Patient's Birth Date	
(0010,0040)		0	Patient's Sex	
(0013,0010)		4	[private]	CTP
(0013,1010)		10	[private]	CBIS-DDSM
(0013,1013)		8	[private]	43352602
(0018,0015)		6	Body Part Examined	BREAST
(0018,1016)		10	Secondary Capture Device Manufacturer	MathWorks
(0018,1018)		6	Secondary Capture Device Manufacturer'...	MATLAB
(0020,000D)		64	Study Instance UID	1.3.6.1.4.1.9590.100.1.2.408909860712120272633130274602115723157
(0020,000E)		64	Series Instance UID	1.3.6.1.4.1.9590.100.1.2.474143160103683865197403431727759385 48
(0020,0010)		4	Study ID	DDSM

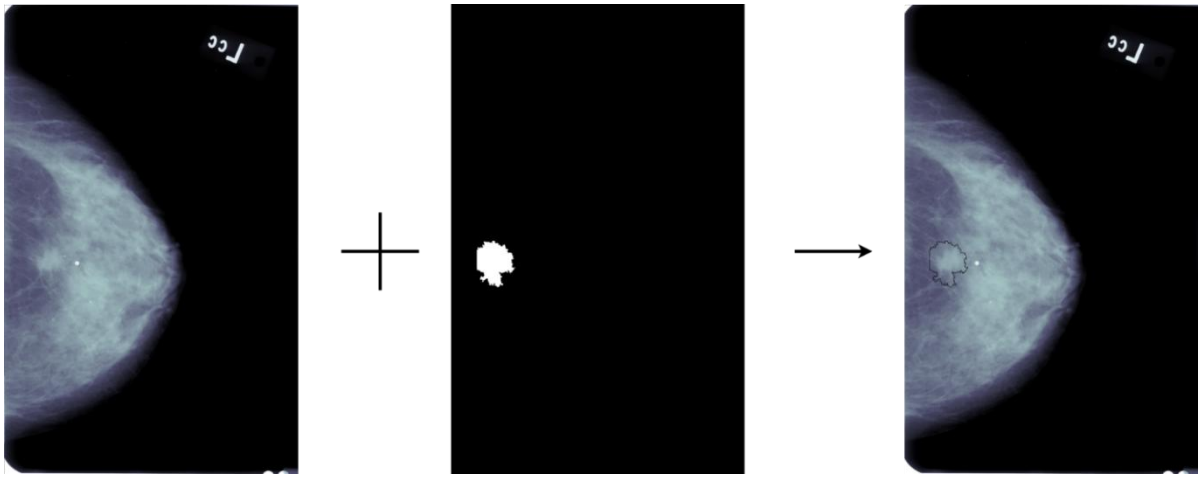
```
00000000   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000010   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000020   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000030   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000040   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000050   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000060   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
00000070   00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00   ................
```

Obrázok 7. Ukážka informácii obsiahnutých v hlavičke jednej zo snímok databázy CBIS-DDSM

Každá mamografická snímka obsahuje jeden alebo viacero nálezov („Region of Interest“ – ďalej len ROI), ktoré boli v rámci vyšetrenia v tkanive identifikované. Aby sme mohli tieto nálezy zo snímkou extrahovať, potrebujeme vedieť ich pozíciu. Na

uchovávanie pozície nálezov používa databáza schému binárnej masky. Binárna maska predstavuje dodatočný DICOM súbor, pričom každej snímke je priradených práve toľko súborov masiek, koľko je na nej nálezov. Binárna maska je obrázok, ktorý má vždy presne rovnaké rozlíšenie ako je mamografická snímka, ktorej prislúcha. Na rozdiel od pôvodnej snímky majú pixely len dve hodnoty – pre čiernu a bielu farbu (poznámka – snímky v CBIS-DDSM majú 16-bitový farebný formát, kde čiernej farbe prislúcha hodnota 0 a bielej farbe hodnota 65535 = 2^{16} - 1). Medzi pixelmi masky a pixelmi pôvodnej snímky existuje mapovanie, kde všetkým pixelom masky, ktorým v pôvodnej snímke prislúchajú pixely pokrývajúce oblasť ROI, je priradená biela farba, zatiaľ čo všetkým pixelom, ktoré v pôvodnej snímke ležia mimo ROI, je priradená farba čierna. Táto konštrukcia umožňuje vyhľadať na pôvodnej snímke presnú lokalitu nálezu a extrahovať len túto časť obrázka. Extrahovaniu ako aj použitej metodike sa budeme podrobnejšie venovať v Kapitole 5.3.

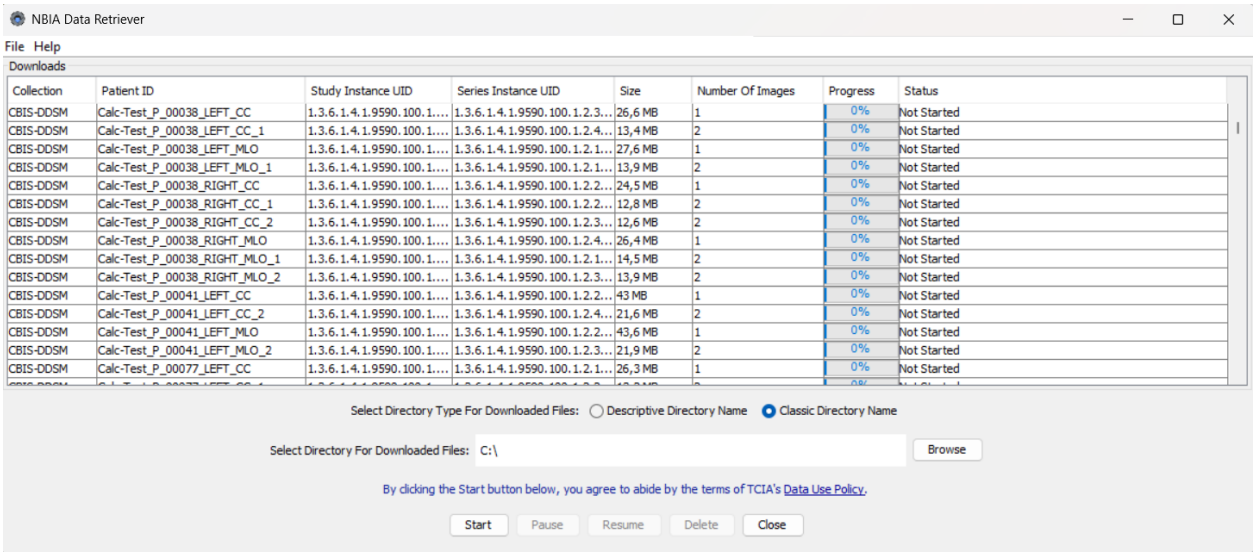
Na Obrázku 8 môžeme pozorovať ukážku uchovávania a identifikácie pozície nálezu pomocou binárnej masky.



Obrázok 8. Identifikácia pozície nálezu na mamografickej snímke pomocou binárnej masky

Aby sme mohli s databázou pracovať, je potrebné si ju najskôr stiahnuť na fyzické médium, pričom prevzatie je možné len prostredníctvom programu NBIA Data Retriever, ktorý číta súbory typu TCIA („The Cancer Imaging Archive“), čo je interný formát používaný službou Cancer Imaging Archive, ktorú prevádzkuje agentúra National Cancer Institute[31]. Ide o hlavnú federálnu agentúru Spojených štátov amerických zaoberajúcou

sa výskumom rakoviny. Agentúra spravuje aj stránky CBIS-DDSM databázy, ktoré poskytujú všetky potrebné zdroje a súbory. Upozorňujeme, že pre správne fungovanie našich skriptov je potrebné pred sťahovaním v okne programu NBIA Data Retriever zaškrtnúť pole „Classic Directory Name“. V prípade použitia možnosti „Descriptive Directory Name“, ktorá je predvolená, bude priečinková hierarchia používať iné názvoslovie, ktoré nie je zo sprievodnými CSV súbormi a s nami vytvorenými skriptami kompatibilné. Ukážku prostredia programu aj so správnym nakonfigurovaním uvádzame na Obrázku 9.



Obrázok 9. Prostredie programu NBIA Data Retriever, korektne nakonfigurované pre začatie sťahovania

Databáza sa ukladá do jedného koreňového priečinka, ďalej sa vetviaceho do troch hlbších úrovní, zakončených DICOM súbormi. Priečinky prvej úrovne prislúchajú jednotlivým záznamom. Záznam v kontexte databázy je výsledok jedného konkrétneho vyšetrenia na konkrétnom pacientovi, ktorý pozostáva s trojice súborov:

a) röntgenovej snímky (názov 1-1.dcm)
b) binárnej masky (zväčša názov 1-1.dcm)
c) hotového výrezu nálezu, uloženému taktiež v DICOM formáte (zväčša názov 1-2.dcm)

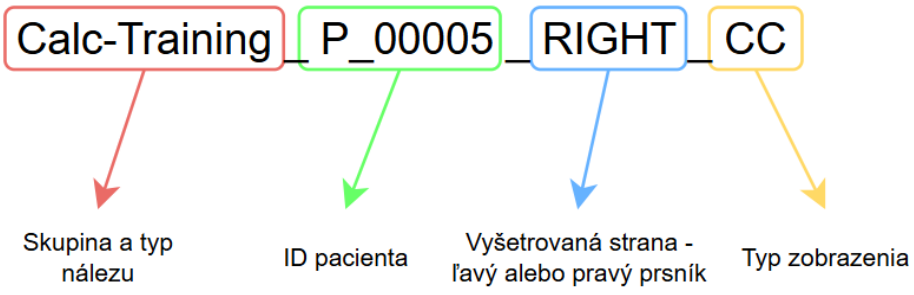
Ako poznámku uvádzame, že poskytnuté hotové výrezy sme v práci nepoužívali, nakoľko sme vzhľadom na odlišnú metodiku potrebovali vyrobiť vlastné. Úlohe generovania vlastných výrezov sa budeme venovať v kapitole 5.3.

Mimo samotné dáta databáza poskytuje aj štvoricu CSV súborov, slúžiacich ako metadáta, umožňujúce logické porozumenie a technické spracovanie databázy. Každý so CSV súborov slúži na popis niektorej navzájom medzi sebou disjunktnej podmnožiny záznamov, pričom dohromady pokrývajú celú databázu. Jedná sa o nasledujúce štyri podcelky:

a) „Calc Training“
b) „Calc Test“
c) „Mass Training“
d) „Mass Test“

Deskriptory „Mass“ a „Calc“ opisujú, či daná množina záznamov patrí do základnej kategórie plošných zatienení alebo mikrokalcifikátov. A keďže databáza implicitne predpokladá použitie dát na trénovanie modelov strojového učenia alebo neurónových sietí, ponúka už hotové riešenie rozdelenia záznamov do trénovacej a testovacej sady, k čomu slúžia popisy „Training“ a „Test“. Vzhľadom na to, že my sme si pri trénovaní delili dáta podľa vlastnej schémy pomocou skriptov, sme toto predurčené rozdelenie ignorovali, a tréningové aj testovacie sady sme pre oba typy nálezov zlúčili dohromady.

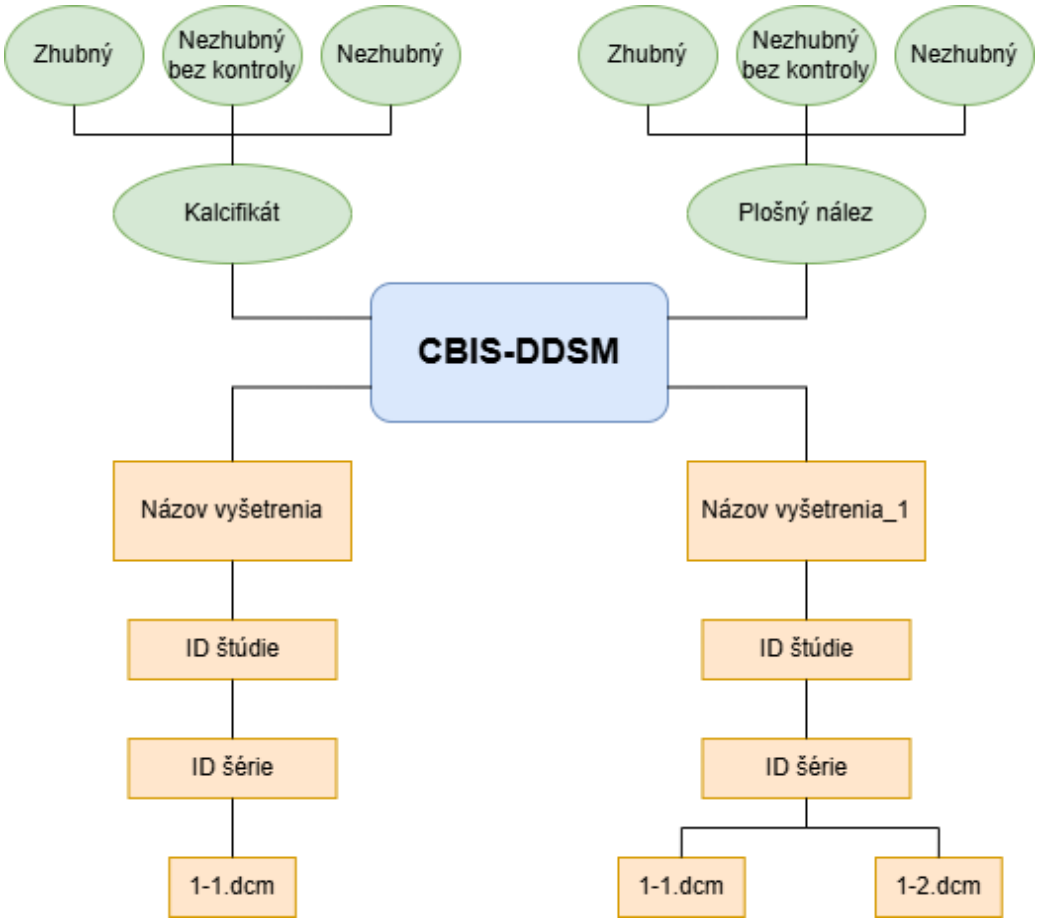
Na Obrázku 10 uvádzame príklad názvu záznamu (priečinka prvej úrovne), ktorý zjednodušene popisuje špecifiká vyšetrenia a objaveného nálezu.



Obrázok 10. Popis názvu záznamu v databáze CBIS-DDSM

CSV súbory poskytujú dodatočné informácie o okolnostiach a výsledkoch daného vyšetrenia. Vo všeobecnosti platí, že jeden riadok CSV súboru zodpovedá minimálne dvom primárnym priečinkom prvej úrovne v hierarchickej štruktúre. Prvý priečinok obsahuje samotnú röntgenovú snímku, zatiaľ čo druhý priečinok obsahuje binárnu masku a výrez nálezu. V prípade, že bolo počas jedného vyšetrenia identifikovaných viacero nálezov, existuje pre každý nález v rámci jedného vyšetrenia samostatný priečinok s vlastnou binárnou maskou a výrezom, ktorý zdieľa rovnaký identifikátor nálezu, ale pridáva k nemu číselnú príponu.

Na záver kapitoly uvádzame prehľadnú schému znázorňujúcu logickú a fyzickú štruktúru CBIS-DDSM a základných rozdielov medzi nimi.



Obrázok 11. Schéma logickej (zelená farba) a fyzickej (oranžová farba) štruktúry databázy

5.3 Skripty na prácu s databázou CBIS-DDSM

Pre účely práce s CBIS-DDSM databázou sme vytvorili vlastné, špecializované skripty, ktoré nám umožnili efektívne automatizovať procesy súvisiace so spracovaním dát, ako aj flexibilne, precízne a v kontexte narábať s jednotlivými snímkami. Automatizovanie spracovania dát spočíva v sekvenčnom prechádzaní všetkých alebo určitej množiny záznamov a aplikovaní vybraného procesu (funkcie) na jednotlivé záznamy. Procesy, ktoré sme v rámci tejto práce na záznamy uplatňovali, zahŕňajú:

- Identifikácia okrajových bodov ROI v pôvodnej röntgenovej snímke vo forme ohraničujúceho obdĺžnika („bounding box“)
- Extrakcia ROI výrezu v rozsahu ohraničujúceho obdĺžnika a jeho uloženie vo forme PNG obrázka
- Generovanie dodatočných modifikácii výrezu na základe daného záznamu za účelom rozšírenia datasetu
- Ukladanie súradníc ROI do pomocného CSV súboru
- Vykreslenie danej snímky spolu s grafickým ohraničením ROI a ohraničujúceho obdĺžnika
- Zbieranie štatistických údajov o datasete

Skripty, ktoré sme za účelom tejto práce vyvinuli, a pomocou ktorých sme s databázou pracovali, boli vytvorené v jazyku Python. Použitá bola verzia jazyka 3.12.9.

Základnou úlohou skriptov bolo automatizovanie sekvenčného prechádzania databázy, buď od počiatku do konca alebo na vybranej množine záznamov. Na každú jednotlivú snímku sa potom aplikuje želaná funkcia, najčastejšie vygenerovanie výrezu.

Za účelom prehľadnosti, modularity a jednoduchosti použitia sme pre skripty vytvorili jednoduchý objektový model a v nasledujúcich odstavcoch ho stručne popíšeme spolu s pracovným postupom, ktorý sme využívali.

Základom práce je vytvorenie inštancie iterátora databázy (objekt triedy CBIS_DDSM_Iterator), ktorej predáme všetky povinné parametre, s ktorých medzi najdôležitejšie patria nasledujúce 2:

- **data_address** – plná (absolútna) cesta ku koreňovému priečinku databázy
- **csv_address** – plná (absolútna) cesta ku CSV súboru s metadátami

Oba parametre nevyhnutné pre akúkoľvek ďalšiu prácu s databázou, pretože bez nich nie je možné napojiť sa na záznamy databázy na disku. Iterátor sa pri svojom vytvorení postará o pripravenie a načítanie všetkých potrebných dát a je ihneď pripravený k použitiu.

Ďalším dôležitým parametrom je „functions", ktorý predstavuje zoznam funkcií, ktoré chceme pri spracovávaní využívať. Iterátor potom vykoná dané funkcie na každej snímke v zadanej množine záznamov, pričom ich volá podľa poradia, v akom nasledujú v zozname. Množina záznamov, ktoré budú spracované je definovaná automaticky, pričom je možné ju navoliť manuálne pomocou poskytnutia parametrov „min_record" a „max_record", do ktorých vložíme počiatočný a koncový index záznamov, nad ktorými chceme pracovať. Pokiaľ tieto parametre nie sú zadané, iterátor prejde všetkými záznamami databázy.

Prvotnou funkciou volanou po inicializácii objektu iterátora a jeho spustení pomocou funkcie begin() je metóda __prepare_data(), ktorá sa postará o načítanie CSV súboru s metadátami do podoby vhodnej pre iteráciu. Proces iterácie nad jednotlivými záznamami prebieha v privátnej metóde __iterate(). Pred vykonaním zoznamu zadaných funkcií si iterátor automaticky uloží originálne (pochádzajúce z databázy) ROI koordináty vo forme ohraničujúceho obdĺžnika pre danú snímku do atribútu „bounding_box_coordinates" (funkcia __get_bounding_box_coordinates()). V prípade CBIS-EMBED databázy nie sú ROI súradnice explicitne zadané a musia sa získať z binárnej masky. Iterátor v jednom cykle prechádza celou maskou, zaznačujúc si začiatočný a koncový bod intervalu bielych pixelov v každom riadku. Na základe toho získa presnú pozíciu ohraničujúceho obdĺžnika.

V rámci našej implementácie triedy CBIS_DDSM_Iterator sme vytvorili niekoľko predpripravených funkcii, ktoré je možné použiť v procese iterácie. Ide o funkcie na zobrazovanie snímok, vytváranie výrezov a zbieranie niektorých štatistík o datasete. V tejto práci sme najčastejšie používali dvojicu funkcii, nasledujúcich po sebe:

recompute_bounding_box_coordinates() → create_roi_patch()

Prvá z menovaných funkcií má na starosti nastavenie nových koordinátov nálezu, ktoré budú použité pre extrakciu výrezu. Toto bolo potrebné z dôvodu našej metodiky, ktorá vyžadovala **vycentrované výrezy jednotnej veľkosti**. Keďže pôvodný ohraničujúci obdĺžnik nálezu mohol vo všeobecnosti u jednotlivých snímok nadobúdať variabilné

rozlíšenie (kvôli premenlivej veľkosti nálezu), bolo potrebné súradnice prepočítať. Nový ohraničujúci obdĺžnik sme zhotovili tak, že sme identifikovali stredový bod pôvodného obdĺžnika, okolo ktorého sme narysovali nový obdĺžnik tak, aby jeho výsledná veľkosť zodpovedala želanému rozmeru (zväčša násobkom rozlíšenia 224*224). Častý problem spočíval v tom, že pôvodný obdĺžnik mal menšie rozmery než tie, ktoré sme extrahovali. Akonáhle bol takýto malý nález umiestnený na okrajoch snímky, novo opísaný väčší obdĺžnik by už v jednom alebo oboch rozmeroch prekračoval hranice snímky. V takýchto prípadoch sme výsledný výrez ukotvili tak, aby sa dotýkal jej okraja, ale nepresiahol ho. Z toho dôvodu nie sú všetky výrezy dokonale vycentrované.

Najčastejším problémom, s ktorým sme sa pri spracovávaní dát z CBIS-EMBED stretávali, bola nekonzistentnosť metadát v CSV súboroch. Metadáta obsahujú pre každý záznam tri polia, ktoré udávajú relatívnu adresu mamografickej snímky (súbor 1-1.dcm), binárnej masky (súbor 1-1.dcm) a predpripraveného výrezu (súbor 1-2.dcm). Súbory masky a výrezu sa vždy nachádzali v spoločnom priečinku, oddelene od priečinku so súborom snímky. Často však dochádzalo k tomu, že mali prehodené názvy, čo bolo potrebné v kóde kontrolovať. Kontrolu sme vykonávali na základe veľkosti súborov. Ak bol súbor 1-1.dcm menší ako súbor 1-2.dcm, názvy boli prehodené. Adresy súborov a ich dĺžky sa taktiež mierne odlišovali medzi metadátami pre plošné nálezy a pre kalcifikáty, čo si vyžiadalo trochu rozdielne parsovanie pre obe skupiny dát.

Vytvorené výrezy sa ukladajú na disk, na adresu zadanú parametrom „output_address“. Výrezy sa ukladajú do troch priečinkov podľa vyhodnotenia nálezu:

- zhubné – priečinok „MALIGNANT“
- nezhubné – priečinky „BENIGN“ alebo „BENIGN_WITHOUT_CALLBACK“

Kompletný popis triedy CBIS_DDSM_Iterator, spolu s UML diagramom, a detailným rozborom funkcii a argumentov sa nachádza v Prílohe A. Táto príloha zároveň slúži aj ako užívateľská príručka. Predpokladáme, že používateľ skriptu je programátor alebo výskumník so základnou znalosťou jazyka Python. Našim cieľom bolo vytvoriť jednoduchý, ale univerzálny nástroj, ktorý ľahko bude konfigurovateľný, modifikovateľný a rozšíriteľný aj pre potreby mimo záber tejto práce.

6 Databáza EMBED

V záujme získania nových použiteľných, ktorými by sme mohli rozšíriť naše existujúce datasety extrahované z CBIS-DDSM sme prikročili ku spracovaniu ďalšej mamografickej databázy: EMBED[32]. Na tomto mieste je potrebné zmieniť, že pôvodným zámerom v zadaní práce bolo využiť na dodatočné dáta databázu OPTIMAM. Prístup k tejto databáze je však kontrolovaný a závisí od udelenia povolenia od vlastniacej entity. Keďže EMBED sme dokázali obstarať skôr a načas, rozhodli sme sa napokon použiť práve ten.

Databáza EMBED („EMory BrEast Imaging Dataset") je oproti CBIS-DDSM oveľa väčšia – obsahuje 3 383 659 záznamov od 116 000 žien, s rovnomernou reprezentáciou bielych a afro-amerických pacientok. Napriek veľkosti databázy je voľne dostupných len 20% jej objemu, čo zodpovedá prvým dvom vyšetrovaným kohortám (z celkového počtu 10). Tieto dáta sú distribuované prostredníctvom iniciatívy Amazon Web Services Open Data Program.

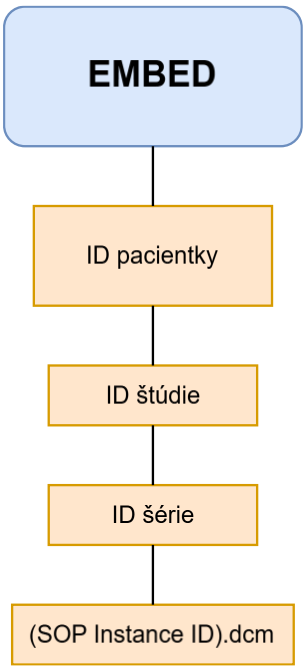
Ďalším významným rozdielom oproti CBIS-DDSM databáze je použitá metodika, ktorou boli mamografické snímky vytvorené. Snímky z EMBED databázy boli získané modernou technikou tzv. tomosyntézy, teda 3-rozmernej digitálnej mamografie. Zatiaľ čo dáta CBIS-DDSM pochádzajú z analógových snímok, pôvodne vyvolaných na špeciálnom fluorescenčnom papieri a až následne oskenované a digitalizované, EMBED ponúka natívne digitálne snímky.

6.1 Štruktúra databázy

V prípade databázy EMBED je logická štruktúra databázy podstatne zložitejšia – jednotlivé výsledky vyšetrení sa neobmedzujú len na jednoduché rozdelenie podľa typu nálezu alebo zhubnosti. Vďaka prítomnosti väčšieho počtu popisných veličín tu nachádzame omnoho vyššiu variabilitu špecifických stavov. Jednou z úloh, ktoré sme pri spracovávaní dát riešili, bolo práve mapovanie týchto rôznorodých inštancií do nami klasifikovaných kategórií.

Fyzická štruktúra databázy je naproti tomu pomerne priamočiara. Dáta sú rozdelené do jednotlivých kohort (priečinky prvej úrovne), ktoré sa ďalej vetvia podľa ID pacienta (priečinky druhej úrovne). Tie sa napokon delia podľa ID štúdie (tretia úroveň) a ID série

(štvrtá úroveň), zakončené vždy jedným DICOM súborom, ktorý je pomenovaný unikátnym identifikátorom DICOM súboru („SOP Instance UID“) . Toto ID vo všeobecnosti slúži na to, aby jednoznačne identifikovalo každú snímku v rámci jednej série. Fyzickú štruktúru databázy môžeme vidieť na Obrázku 12.



Obrázok 12. Fyzická štruktúra databázy EMBED

K dispozícii sú dva pomocné CSV súbory – metadáta a klinické dáta. Metadáta obsahujú údaje o jednotlivých DICOM snímkach – parametre vyšetrenia (napr. čas merania, použitá dávka röntgenového žiarenia, vyšetrovaná strana a pod.), technické parametre (napr. verzia softvéru zobrazovacieho prístroja) a informácie o snímke (názov, cesta k súboru, ROI koordináty). Klinické dáta potom obsahujú výsledky vyšetrenia. Z týchto boli pre nás najdôležitejšie dva parametre (stĺpce):

- **path_severity** – udáva najzávažnejší stupeň nálezu prítomný na danej snímke. Nadobúda hodnoty z množiny {0, 1, 2, 3, 4, 5, 6}, ktoré postupne indikujú klesajúcu závažnosť – od invazívneho nádoru (0) až po nenádorový nález (6).
- **asses** – obsahuje BI-RADS skóre vyšetrenia. BI-RADS („Breast Imaging-Reporting And Data System“) je diagnostická metrika mamografického

vyšetrenia, ktorá sa zaviedla kvôli potrebe štandardizovaného posudzovania výsledkov vyšetrenia [33]. Stanovuje ho rádiológ a môže nadobúdať 7 rôznych hodnôt, reprezentovaných buď číslami od 0 po 6 alebo veľkými písmenami abecedy (napr. N – negatívny výsledok, neprítomnosť abnormality).

Obidva CSV súbory je možné prepojiť na základe ID pacienta (stĺpec „empi_anon“) a ID vyšetrenia (stĺpec „acc_anon“).

6.2 Skripty na prácu s databázou EMBED

Skripty, pomocou ktorých sme spracovávali a extrahovali dáta z databázy EMBED vychádzajú z rovnakého základu ako tie pre databázu CBIS-DDSM – podstatou je trieda EMBED_Iterator, ktorá je potomkom abstraktnej triedy Database_iterator. Trieda implementuje rovnaké základné metódy ako jej ekvivalent pre CBIS-DDSM a definuje obdobné funkcie, ktoré vykonávajú porovnateľné úlohy – iterátor prechádza zadanú množinu záznamov a na každý z nich aplikuje definované funkcie, najmä extrakciu výrezov. V tejto kapitole sa teda zameriame na špecifiká práce s EMBEDom vychádzajúce z jeho formy, ako aj na unikátne problémy, na ktoré sme pri tejto databáze narazili.

Hlavné špecifikum oproti CBIS-DDSM spočívalo v rozdelení pomocných dát do dvoch samostatných CSV súborov – metadát a klinických dát. To viedlo k potrebe ich prepojenia, čo si vynútilo použitie trochu odlišného spôsobu práce a implementácie základných funkcií. Funkcia __prepare_data() už neslúži len na jednoduché načítanie jednotného CSV súboru (ako pri CBIS-DDSM), ale vykonáva špecifické spojenie oboch súborov, ku ktorým pristupuje ako k tabuľkám relačnej databázy. Používa pri tom knižnicu jazyka Python s názvom „sqlite3“, ktorá umožňuje vytvárať lokálne databázové objekty na disku. Na začiatku sa vytvorí nová databáza („embed.db“), pozostávajúca z dvoch tabuliek („metadata“ a „clinical_data“). Vzhľadom na to, že raz vzniknutá databáza perzistuje na disku, nie je potrebné ju opakovane vytvárať pri následnom použití nových inštancií iterátora. Pre toto je k dispozícii booleanový argument „create_database“, ktorý iterátor upozorní na potrebnosť či nepotrebnosť jej vytvorenia.

Po vytvorení databázy sa realizuje spojenie tabuliek prostredníctvom SQL príkazu typu „JOIN“, ktorý je potrebné taktiež dodať ako parameter (argument „database_query“). My sme za účelom tejto práce používali nasledujúci príkaz:

```
join_query = """
    SELECT *
    FROM metadata
    INNER JOIN clinical_data ON
    metadata.acc_anon = clinical_data.acc_anon AND
    metadata.empi_anon = clinical_data.empi_anon AND (
     (clinical_data.side = 'L' AND metadata.ImageLateralityFinal = 'L') OR
     (clinical_data.side = 'R' AND metadata.ImageLateralityFinal = 'R') OR
     (clinical_data.side = 'B')
     )
    WHERE (num_roi != 0 AND path_severity != '')
    OR (num_roi != 0 AND path_severity = '' AND asses IN ('B', 'P', 'K'))
    ORDER BY ROI_coords
    """
```

Obrázok 13. Použitý SQL príkaz na spojenie tabuliek metadát a klinických dát

Tento príkaz spája tabuľky na základe zhody v ID vyšetrenia („acc_anon“) a ID pacienta („empi_anon“). Zároveň vytvára mapovanie laterality nálezu (vyšetrovanej strany) medzi oboma tabuľkami:

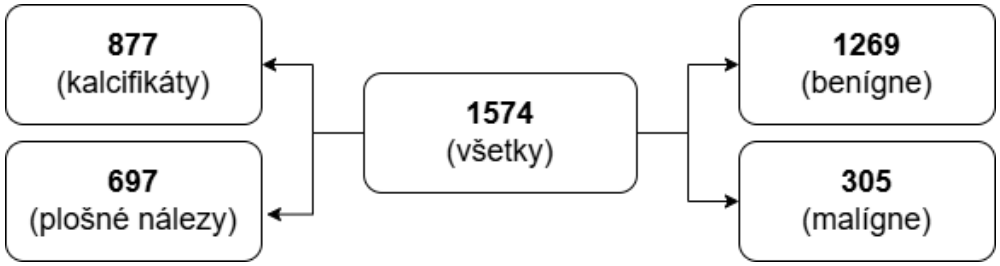
- Na vyšetrenie ľavej strany („L“) v klinických dátach sa namapuje snímok ľavej strany v metadátach.
- Na vyšetrenie pravej strany („R“) v klinických dátach sa namapuje snímok pravej strany v metadátach.
- Na obojstranné vyšetrenie („B“) v klinických dátach sa namapujú snímky oboch strán v metadátach.

Následne cez podmienku „WHERE“ filtrujeme len tie záznamy spojenej tabuľky, ktoré spĺňajú jednu z dvoch kritérií:

1. Obsahujú aspoň jednu sadu ROI súradníc a zároveň majú vyplnené pole „path_severity“ - obe polia potrebujeme na extrakciu a správne zaradenie výrezu
2. Obsahujú aspoň jednu sadu ROI koordinátov, nemajú vyplnené pole path_severity, ale ich BI-RADS skóre (pole „asses“) má hodnoty „B“ (benígne), „P“ (pravdepodobne benígne) alebo „K“ (nádor potvrdený biopsiou) - v tomto prípade

vieme pri chýbajúcej „path_severity“ nález zaradiť správne práve pomocou BI-RADS skóre.

Takto spojená a očistená tabuľka nám umožnila pre potreby trénovania extrahovať dodatočných 1574 záznamov, rozdelených nasledovne podľa Obrázka 14:



Obrázok 14. Schéma počtov výrezov získaných z EMBEDu

Z rozloženia výrezov môžeme pozorovať výrazný nepomer medzi zhubnými a nezhubnými nálezmi, so značnou prevahou tých benígnych. Môžeme konštatovať, že prvé dve kohorty dát EMBEDu dokážu poskytnúť len malé množstvo použiteľných malígnych výrezov.

Ako už príkaz vyššie naznačil, ďalším zásadným rozdielom EMBEDu oproti štruktúre CBIS-DDSM spočíval v spôsobe uchovávania ROI koordinátov. Tie sú k dispozícii priamo a nemusia sa vypočítavať z binárnej masky. Súradnice sú uložené v poli „ROI_coords“. Väčšina záznamov v prvých dvoch kohortách však toto pole nemá vyplnené, čo významnou mierou prispieva k celkovo malej výnosnosti tejto databázy.

Pri výrobe výrezov sme konzistentne narážali na pomerne zásadnú komplikáciu, spočívajúcu v nerozhodnosti stranovej orientácie ROI koordinátov. Súradnice pre x-ovú os (tzn. ľavý a pravý okraj ohraničujúceho obdĺžnika) v mnohých prípadoch nepodliehajú štandardnej súradnicovej sústave pre DICOM obrázky, kde počiatok (bod [0, 0]) sa nachádza v ľavom hornom rohu. Namiesto toho vychádzajú z pravostranného počiatku. Schopnosť detegovať, kedy sa x-ové súradnice nálezu sprava a kedy zľava sa ukázala byť problematická. Pokiaľ sme dokázali zistiť, nedokážeme na rozpoznanie využiť existujúce polia v metadátach alebo klinických dátach. Pre rozpoznávanie tohto problému sme použili vlastnosť výraznej laterality, ktorá je príznačná pre úplnú väčšinu snímok. Znamená to, že

prsník je na snímke situovaný vždy na niektorej strane, zatiaľ čo náprotivná strana ostáva tmavá a presvetlená. Na základe toho vieme pomerne spoľahlivo detegovať prevrátenú x-ovú os koordinátov podľa nasledujúceho algoritmu:

1. Predpokladaj počiatok súradnicovej sústavy **vľavo hore**. Extrahuj pixelové hodnoty regiónu v rámci zadaných ROI koordinátov.
2. Spočítaj pixely, ktorých hodnota ≤ 20 - tzn. veľmi **tmavé** pixely. Snímky v EMBEDe sú 8-bitové – ich pixelové hodnoty sú v rozmedzí [0-255].
3. Pokiaľ je podiel takýchto pixelov ≥ 0.6, pristál ohraničujúci obdĺžnik mimo prsného tkaniva ⇒ **x-ová os je prehodená**.
4. Ak je x-ová os prehodená, predpokladaj počiatok súradnicovej sústavy **vpravo hore**, extrahuj pixelové hodnoty znova a získaj korektný výrez

Získané výrezy sa opäť ukladajú na disk (na adresu danú argumentom „output_address“), pričom tentokrát sa ukladajú do priečinkov podľa typu nálezu (plošné nálezy do priečinka „MASS“, kalcifikáty do priečinka „CALC“), v rámci ktorých sú rozlíšiteľné na zhubné a nezhubné na základe názvu.

Podrobnému popisu triedy EMBED_Iterator sa opäť venujeme v Prílohe A.

7 Úloha klasifikácie mamografických snímok

Cieľom tejto práce bola optimalizácia klasifikačnej úlohy nad výrezmi z mamografických snímok, s cieľom zlepšiť presnosť klasifikácie pridaním „kontextu".

Vzhľadom na množstvo extrahovateľných dát bola prvotným zdrojom výrezov databáza CBIS-DDSM, dodatočne doplnená o databázu EMBED.

Úloha spočívala v klasifikácii výrezov (nálezov) na zhubné a nezhubné. Vzhľadom na to, že typ plošného nálezu je od typu mikrokalcifikátu rozlíšiteľný voľným okom, sme pristúpili k tomu, že sme tieto dva typy nemiešali, ale trénovali sme klasifikačnú úlohu nad oboma typmi osobitne.

Základom pre naše trénovanie bolo vytvorenie potrebných datasetov, ktorých položky slúžili ako vstupy do ResNetu. Datasetom nazveme skupinu obrazových výrezov zobrazujúcich ROI, získaných z originálnych mamografických snímok. Všetky výrezy v rámci datasetu sú rovnakého typu (plošné zatienenie alebo mikrokalcifikát), a každý dátový element je zaradený práve do jednej klasifikačnej skupiny (zhubný / nezhubný).

Vzhľadom na to, že model ResNet50 bol predtrénovaný na obrázkoch veľkosti 224*224, zvolili sme za základné rozlíšenie výrezu práve túto hodnotu.

Pod pojmom „pridanie kontextu" máme na mysli zahrnutie širšieho okolia nálezu do vstupných dát prichádzajúcich do neurónovej siete. Pri našich experimentoch sme vychádzali z faktu, že štandardný vstup do CNN, a teda aj do modelu ResNet50, je RGB obrázok, pozostávajúci z troch farebných kanálov – červeného, zeleného a modrého. Každý vstupný kanál obsahuje pixelové hodnoty pre danú farbu, pričom výsledný farebný obrázok vzniká zložením týchto troch kanálov. Každý kanál teda vo svojej podstate predstavuje samostatný obrázok. Vzhľadom na to, že mamografické DICOM snímky sa zväčša zhotovujú v čiernobielom formáte („grayscale"), tak pri ich konverzii na obrazový formát použiteľný na trénovanie (v našom prípade PNG) dochádza k tomu, že vznikne čiernobiely výrez, ktorý v každom farebnom kanáli obsahuje tie isté pixelové dáta, skopírované naprieč kanálmi trikrát. Takýto obrázok je potom použitý ako vstup do CNN.

Náš nápad, a centrálna myšlienka tejto práce, spočívala vo využití tých farbených kanálov, ktoré by normálne obsahovali redundantné dáta, na poskytnutie dodatočných informácií pre neurónovú sieť, ktoré by mohli zlepšiť výsledky klasifikácie. Ide teda o spôsob obohatenia vstupných dát bez akejkoľvek straty informácie z originálnych

vstupných dát. Keď do jedného farebného kanála umiestnime pixelové hodnoty pôvodného čiernobieleho obrázka, zostanú nám k dispozícii dva farebné kanály, kde môžeme uložiť nové dáta.

Pridanie kontextu teda v praxi znamená zakomponovanie výrezu väčšej veľkosti ako 224*224 do redundantných farebných kanálov. V rámci tejto práce sme vyrábali dva typy kontextových výrezov (**elementov**):

1. **Dvojnásobné rozlíšenie** (budeme ďalej skrátene označovať 2x) – štandardne rozlíšenie 448*448
2. **Trojnásobné rozlíšenie** (budeme ďalej skrátene označovať 3x) – štandardne rozlíšenie 672*672

Kontextový element 1x potom predstavuje výrez pôvodnej veľkosti (štandardne 224*224).

Vzhľadom na to, že z podstaty veci je nutné, aby všetky tri farebné kanály mali totožné rozlíšenie, tak kontextové výrezy určené do dodatočných kanálov sme po ich vygenerovaní dodatočne zoškálovali na základné rozlíšenie.

Na tomto mieste je nutné spomenúť, že v prípade klasifikácie plošných nálezov sme prikročili aj k použitiu výrezov základnej veľkosti 448 x 448 – tomuto rozhodnutiu sa budeme bližšie venovať v Kapitole 9.2. V tomto prípade boli odlišné aj rozlíšenia vyšších kontextových elementov, ale tak, aby sme zachovali konzistenciu metodiky (2x = 896 x 896, 3x = 1344 x 1344).

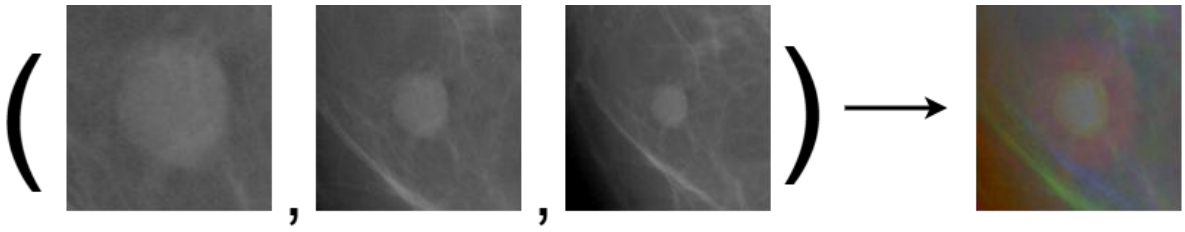
Pre jednotlivé testovacie scenáre sme zaviedli kódové označenia, ktoré jasne a zreteľne vyjadrujú, aké dáta sa nachádzajú v jednotlivých farebných kanáloch vstupných obrázkov pre daný experiment. Označenie mapuje farebné kanály R (červené pásmo), G (zelené pásmo) a B (modré pásmo) na veľkosť kontextového elementu, ktorý sa v danom kanáli nachádza: 1x (pôvodný výrez), 2x alebo 3x. Vzorec nižšie demonštruje na príklade konkrétnu inštanciu mapovania:

$$(R, G, B) \rightarrow (1x, 2x, 3x)$$

V prípade, že máme na výber z troch možností kontextovej veľkosti, a neberieme do úvahy poradie prvkov, existuje 10 kombinácii (usporiadaných trojíc), ktoré môžeme vytvoriť, a tieto zároveň predstavovali základ našich testovacích scenárov:

- (1x,1x,1x) (1x,1x,2x) (1x,2x,2x) (1x,1x,3x) (1x,3x,3x)
- (2x,2x,2x) (2x,2x,3x) (2x,3x,2x)
- (3x,3x,3x)

Na Obrázku 15 uvádzame príklad výrezu zo scenára (1x,2x,3x), na ktorom môžeme pozorovať, ako sa kontextové elementy umiestnené v jednotlivých farebných pásmach skombinovali do výsledného celku.



Obrázok 15. Ukážka procesu skombinovania kontextových elementov 1x, 2x, 3x do výsledného prvku kontextu (1x,2x,3x). Na príklade je zobrazený prípad malígneho plošného nálezu

V rámci jednotlivých klasifikačných úloh budeme konkrétny testovací scenár, venujúci sa konkrétnej kombinácii kontextových elementov, označovať zjednodušene ako **kontext**. Scenáre s totožnou informáciou vo všetkých troch kanáloch ((1x,1x,1x), (2x,2x,2x), (3x,3x,3x)) sú funkčne ekvivalentné so scenármi klasického testovania na nemodifikovaných čiernobielych obrázkoch (odstupňovaného rozlíšenia), kde každý farebný kanál obsahuje tie isté dáta. Náš záujem sa prirodzene prednostne sústreďoval na scenáre, kde kanály obsahovali rôznorodejšie kombinácie elementov.

8 Skripty pre trénovanie s modelom ResNet50

Pre prípravu dát, prácu s neurónovou sieťou, trénovanie a testovanie sme používali knižnicu Keras jazyka Python. Knižnica Keras [34] predstavuje API framework, ktorý slúži ako vysokoúrovňová nadstavba nad nízkoúrovňovými knižnicami, ako je napr. Tensorflow, ktorý sme práve použili ako „backend" aj v tejto práci. Keras zefektívňuje prácu s modelmi strojového učenia a umelej inteligencie tým, že abstrahuje niektoré štandardné procesy do praktických a použiteľných funkcii. Vzhľadom na to, že knižnice slúžiace na vývoj umelej inteligencie sa zvyknú časom vyvíjať a meniť (samotný Keras prešiel značnou evolúciou, kedy v rôznych fázach doterajšieho životného cyklu podporoval rôznu množinu „backendových" knižníc), uvádzame na tomto mieste verzie Kerasu a Tensorflowu, ktoré sme pri našich experimentoch používali:

- Keras – 3.8.0
- TensorFlow – 2.18.0

8.1 Príprava dát pred trénovaním

Základným predpokladom pre trénovanie neurónových sietí je existencia troch štandardných a samostatných datasetov, do ktorých je naše dáta potrebné rozdeliť:

- **Trénovacia sada** – používa sa na samotné trénovanie, kedy sa model učí vnímať vzorce a štruktúry v dátach – na základe prvkov tejto sady sa aktualizujú hodnoty trénovateľných parametrov – procesom spätnej propagácie podľa hodnoty „loss" funkcie.
- **Validačná sada** – používa sa na evaluáciu modelu po každej trénovacej sérii (epoche) – na základe výsledku validovania sa upravujú hyperparametre modelu.
- **Testovacia sada** – používa sa na finálnu verifikáciu modelu – obsahuje dáta, ktoré model predtým nevidel, pričom výsledok tejto verifikácie určuje, do akej miery dokáže model generalizovať, teda rozpoznávať všeobecné vzorce prítomné v dátach. Podľa tejto sady sa vyhodnocujú výsledky trénovania.

Základným problémom, ktorý sme v tejto fáze riešili, spočíval v korektnom rozdelení našich vygenerovaných výrezov medzi spomínané tri sady.

Trénovacie sady, ktoré sme v rámci rôznych úloh používali, sa prejavovali viac (mikrokalcifikáty) či menej (plošné nálezy) nevyváženým počtom dát v jednotlivých klasifikačných skupinách. Na základe toho sme pristupovali k ručnému vyvažovaniu datasetov. Dodatočné výrezy sme získali ako modifikácie existujúcich výrezov, pričom konkrétne sme použili osovo prevrátené varianty, podľa osi x aj y. Tento druh modifikácie predstavuje jeden z klasických a overených spôsobov rozširovania datasetov. Keďže sa ale nejedná o originálne, unikátne data, chceli sme zabezpečiť, aby sa originálny výrez spolu so všetkými modifikáciami, ktoré z neho pochádzajú, zaradili do rovnakej dátovej sady. Cieľom bolo predísť situácii, kedy by sa pôvodný výrez alebo ktorýkoľvek jeho variant nachádzali súčasne v trénovacej aj testovacej sade, čím sme chceli eliminovať prípadné skreslenia výsledkov verifikácie natrénovaného modelu, keďže testovacia sada by mala pozostávať len z dát, ktoré model nikdy nevidel.

Keďže knižnica Keras v základe túto funkcionalitu neponúka, dosiahli sme cieľ použitím vlastného systému rozčlenenia dát, skombinovaným s modifikovanými internými funkciami Kerasu na vytváranie hotových datasetov, vhodných ako vstupy do modelu.

Konkrétne rozdelenie datasetu medzi trénovaciu, validačnú a testovaciu sadu, je závislé od použitej celočíselnej násady do generátora pseudonáhodných čísel, ktorý používa funkcia train_test_split(), čo je interná funkcia Kerasu. Aby sme zachovali konzistenciu a mohli navzájom porovnávať výsledky trénovania, najmä pri rôznych scenároch v rámci jednej klasifikačnej úlohy, použili sme pri každom experimente totožné hodnoty násad, a to konkrétne:

- **42** - násada použitá pri rozdelení kompletného datasetu na trénovaciu sadu a zvyšok
- **84** – násada použitá pri rozdelení zvyšku na validačnú a testovaciu sadu

Hodnoty násad boli zvolené náhodne.

Vzhľadom na metodiku rozdelenia dát tak, aby sa varianty pôvodného obrázka nemiešali v jednotlivých sadách, vychádzajúcu zo spoločného základu názvu súborov, môže použitie rovnakých násad spôsobiť špecifické rozdiely v rozdelení aj pri použití datasetov totožných veľkostí. (viď drobné rozdiely pri koncových veľkostiach sád v úlohe klasifikácie plošných nálezov oproti sadám v klasifikačnej úlohe mikrokalcifikátov).

V rámci celej práce sme používali jednotný pomer množstva dát v jednotlivých sadách:

| trénovacia sada | : | validačná sada | : | testovacia sada | = 0.70 : 0.15 : 0.15

8.2 Príprava modelu

Na Obrázku 16 uvádzame konečnú podobu architektúry modelu, ktorá bola použitá na trénovanie vo všetkých klasifikačných úlohách. Model pozostáva z vlastných vrstiev modelu ResNet50, obohatených o ďalšie pridané vrstvy. Pre model ResNet50 sme sa rozhodli na základe výskumu kolegu Adama Mračka [35], ktorý porovnával výkonnosť rôznych reziduálnych architektúr na dátach z CBIS-DDSM. Tento model sa v ňom ukázal byť vhodným kompromisom medzi validačnou presnosťou a robustnosťou architektúry, keď dosiahol výsledky len o 1.3% nižšie ako ResNet101 s viac než dvojnásobným počtom vrstiev (blokov).

Layer (type)	Output Shape	Param #
resnet50 (Functional)	(None, 2048)	23,587,712
rescaling (Rescaling)	(None, 2048)	0
dense (Dense)	(None, 512)	1,049,088
dense_1 (Dense)	(None, 1)	513

Total params: 24,637,313 (93.98 MB)
Trainable params: 24,584,193 (93.78 MB)
Non-trainable params: 53,120 (207.50 KB)

Obrázok 16. Popis výsledného modelu použitého v experimentoch

Pre účely binárnej klasifikácie sú vlastné konvolučné vrstvy modelu ResNet50 napojené na nasledujúce dodatočné vrstvy:

- **„Rescaling" vrstva** – škálovacia vrstva, upravuje vstupné dáta (v tomto prípade pixelové hodnoty) z intervalu [0, 255] do intervalu [0, 1]
- **„Dense" vrstvy** – klasické vrstvy plne prepojených neurónov, zakončené výstupnou vrstvou, ktorá produkuje výsledok klasifikácie. Vzhľadom na to, že sme

použili ResNet50 architektúru bez záverečnej plne prepojenej vrstvy, museli sme tieto vrstvy dodať do modelu manuálne.

Na obrázku tiež môžeme vidieť celkový počet parametrov („Total params") a počet trénovateľných parametrov („Trainable params"). Trénovateľné parametre zahŕňajú parametre vnútri reziduálneho modelu, spolu s parametrami plne prepojených vrstiev.

Aj nami použitá verzia modelu bola inicializovaná so začiatočnými váhami pochádzajúcimi z trénovania na databáze ImageNet. Učinili sme tak preto, že náhodná počiatočná inicializácia váh by jednak zhoršila schopnosť porovnávania výsledkov jednotlivých scenárov medzi sebou, a jednak by značne spomalila tréning a vyžiadala si väčší počet trénovacích epoch. S nášho experimentovania s nastavením váh vyplynulo, že model pri nami zvolenom počte epoch rovnom 20 vykazoval na predtrénovaných váhach vyššie testovacie presnosti v rádoch niekoľkých prercent.

8.3 Počiatočná optimalizácia trénovacieho modelu

V rámci prípravy na tréningový proces sme najskôr uskutočnili optimalizáciu hyperparametrov nášho pripraveného modelu. Model sme optimalizovali s použitím základného kontextu (1x,1x,1x), teda na neupravených pôvodných čiernobielych snímkach. Tento kontext totiž slúžil pri našich experimentoch ako referenčný bod, voči ktorému sme porovnávali vplyv ostatných scenárov s pridaným kontextom.

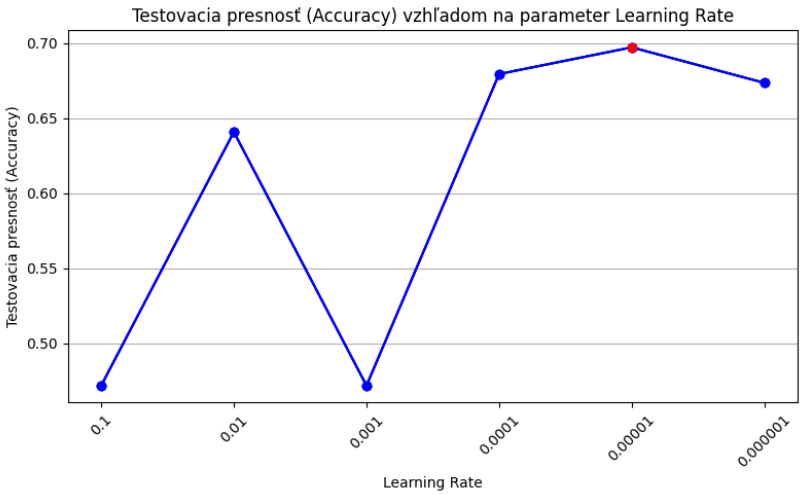
Zamerali sme sa na nasledujúcu dvojicu hyperparametrov:

- **learning_rate** – tento parameter určuje mieru aktualizácie váh modelu v odpovedi na aktuálnu chybu. Pri príliš nízkej hodnote model konverguje k optimálnym váham pomaly, čo môže viesť k dlhým a neefektívnym tréningom. Príliš nízka hodnota potom môže spôsobiť nestabilné chovanie modelu, kedy model nemusí byť schopný správne konvergovať do lokálnych miním stratovej funkcie.
- **batch_size** – určuje množstvo vzoriek dát, ktoré sa v jednom kroku použijú na tréning a aktualizovanie váh.

V prvej fáze sme položili „batch size“ rovnú 32 a optimalizovali sme „learning_rate“. Vyskúšali sme 6 štandardných hodnôt veličiny a porovnali sme výslednú presnosť. Testovali sme tieto hodnoty:

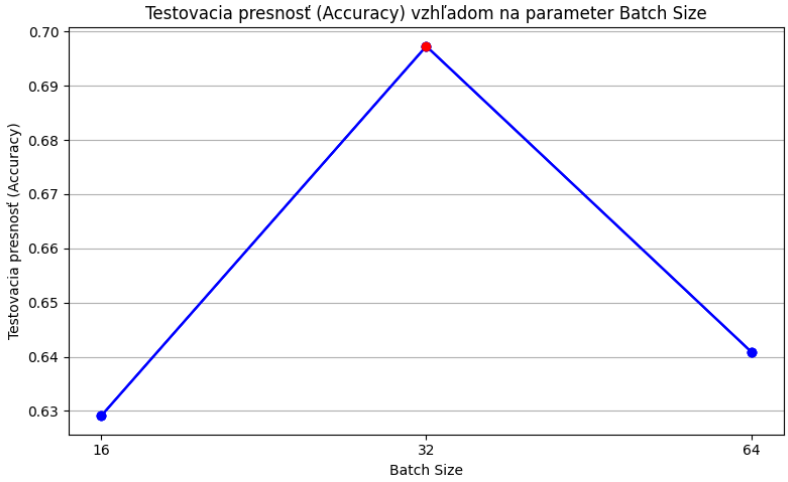
- 0.1, 0.01, 0.001, 0.0001, 0.00001, 0.000001

Na obrázku nižšie uvádzame porovnanie získaných presností („accuracy“) pri použití jednotlivých hodnôt.



Obrázok 17. Porovnanie testovacej presnosti pri použití rôznych hodnôt „learning rate“

Z výsledkov vidíme, že najväčšiu presnosť sme zaznamenali pri hodnote 0.00001, ktorú sme na základe toho nastavili pre náš model. S takto nastavenou a zafixovanou hodnotou „learning rate“ sme následne pokračovali do druhého kroku – optimalizácii hyperparametra „batch size“. Tu sme testovali trojicu hodnôt 16, 32 a 64 (hodnota 32 bola otestovaná v predošlom kroku, lebo sme ju ako strednú z možností zvolili ako základnú). Výsledky opäť uvádzame v grafe.



Obrázok 18. Porovnanie testovacej presnosti pri použití rôznych hodnôt „batch size“

Podľa získaných presností sme parameter „batch size“ modelu stanovili na 32.

9 Výsledky trénovania na dátach z CBIS-DDSM

V tejto kapitole predstavíme výsledky nášho testovania a experimentovania s pridávaním kontextu do trénovacích dát získaných z databázy CBIS-DDSM. Dôvodom na prednostné zameranie sa na túto databázu je ten, že predstavovala náš primárny zdroj dát, ku ktorým sme následne až ako sekundárnu úlohu dokladali dáta z EMBEDu (výsledky trénovania na zmiešaných dátach sú spracované v Kapitole 10).

Najskôr sa zameriame na výsledky klasifikačnej úlohy plošných nálezov, a následne popíšeme zistenia z úlohy klasifikácie mikrokalcifikátov. Popis výsledkov sa zameria na všeobecný trend vplyvov pridávania rôznych typov kontextu a zhrnutie pozorovaní. V rámci každého testovacieho scenára sme sledovali tieto metriky:

- Presnosť („Accuracy“) $= \frac{T_p + T_n}{T_p + F_p + T_n + F_n}$
- „Precision“ $= \frac{T_p}{T_p + F_p}$
- „Recall“ $= \frac{T_p}{T_p + F_n}$
- „Loss“ - hodnota stratovej funkcie („loss function“) na konci testovania – túto metriku meriame kvôli väčšej úplnosti výsledkov, ale v práci ju neanalyzujeme

Premenné vo vzorcoch definujeme nasledovne:

- T_p – počet **skutočne pozitívnych** odhadov modelu – počet nálezov, ktoré boli pozitívne, a model ich správne identifikoval ako pozitívne
- T_n - počet **skutočne negatívnych** odhadov modelu – počet nálezov, ktoré boli negatívne, a model ich správne identifikoval ako negatívne
- F_p – počet **falošne pozitívnych** odhadov modelu – počet nálezov, ktoré boli negatívne, ale model ich nesprávne identifikoval ako pozitívne
- F_n – počet **falošne negatívnych** odhadov modelu – počet nálezov, ktoré boli pozitívne, ale model ich nesprávne identifikoval ako negatívne

V našom kontexte sú pozitívne nálezy reprezentované zhubnými nálezmi, a negatívne nezhubnými nálezmi. Hodnoty metrík sa vždy vzťahujú na výsledky z testovacej sady.

Okrem týchto štatistík sledujeme pri každom pokuse aj tzv. maticu zámen („confusion matrix"), čo je štvorcová matica typu 2x2, ktorá má nasledujúci tvar:

T_p	F_n
F_p	T_n

Obrázok 19. Matica zámen

9.1 Kontext v klasifikácii mikrokalcifikátov

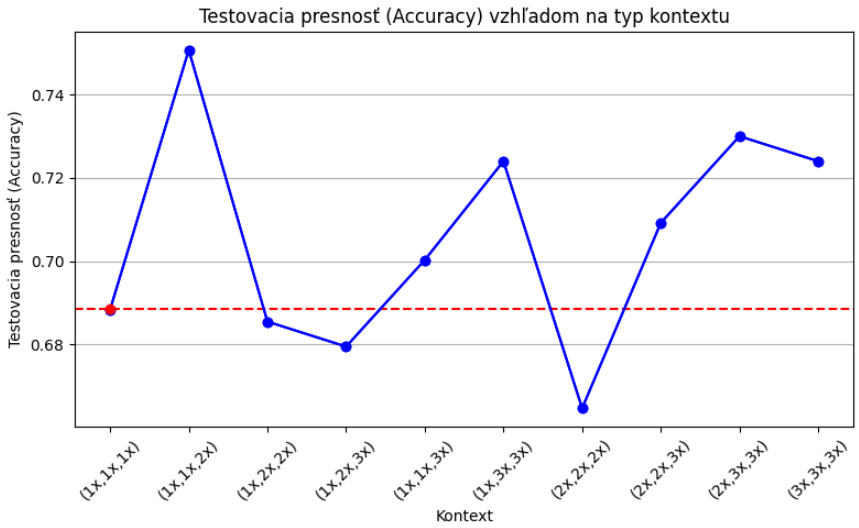
Plošné nálezy sme testovali na v rozlíšení výrezov 224 x 224, a pre každý testovací scenár sme vykonali 20 trénovacích a validačných epoch. Kvôli počiatočnej nevyváženosti datasetu vzhľadom na klasifikačné triedy (benígne nálezy – 765, malígne nálezy – 352), sme prikročili k manuálnemu vyváženiu oboch tried, pridaním augmentovaných, osovo prevrátených variant existujúcich výrezov. Vzhľadom na nutnosť delenia výsledného datasetu na trénovaciu, validačnú a testovaciu sadu, a snahu poskytnúť do všetkých troch sád dostatočné množstvo dát, sme použili augmentované výrezy nie len na dorovnanie, ale aj rozšírenie oboch tried. Výsledný dataset obsahoval 2118 výrezov, plne vyvážených medzi zhubnú a nezhubnú triedu, a rozdelených podľa nasledujúcej schémy:

- trénovacia sada – 1459 vzoriek
- validačná sada – 322 vzoriek
- testovacia sada – 337 vzoriek

Výsledky našich experimentov naznačujú, že zahrnutie kontextu má jednoznačný vplyv na výslednú presnosť klasifikácie mikrokalcifikátov (kategória „calc"). Počas

testovania sme zaznamenali pozitívne aj negatívne vplyvy kontextu, ktoré záviseli od použitia konkrétnej kombinácie kontextu.

Na Obrázku 20 vidíme vývoj presnosti („accuracy") klasifikácie na testovacích sadách, pri postupnom „zvyšovaní" kontextu, teda pridávaní vyšších kontextových elementov. Počiatočný červený bod vyjadruje kontext (1x,1x,1x), teda scenár, kedy trénujeme na pôvodnom snímku, bez prítomnosti akýchkoľvek dodatočných informácií. Tento scenár prirodzene predstavuje základný bod, voči ktorému sme porovnávali všetky ostatné scenáre.

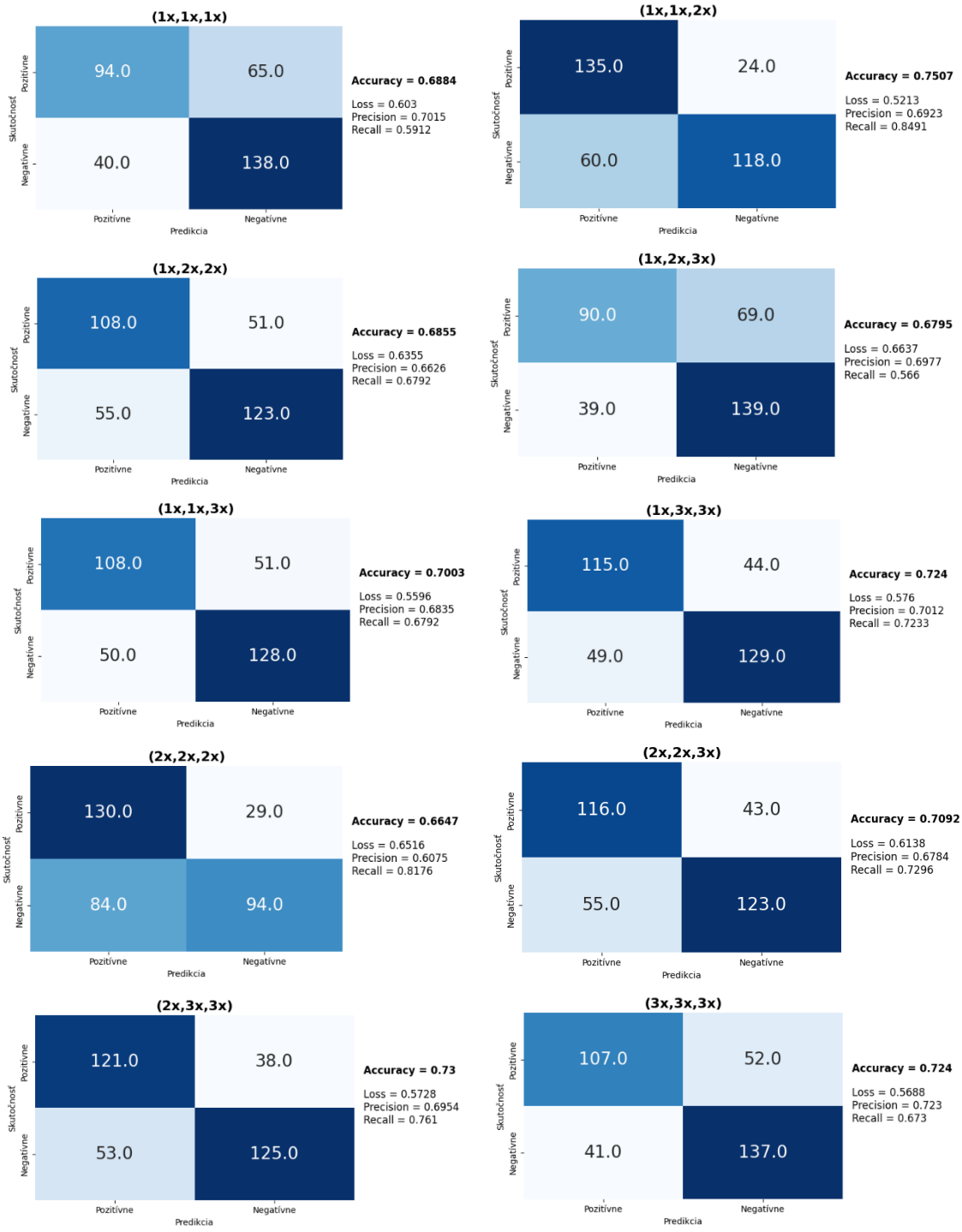


Obrázok 20. Vývoj presnosti vzhľadom na použitý kontext (mikrokalcifikáty)

Z grafu môžeme vyčítať, že použitie rozličného typu kontextu spôsobuje rôzne zmeny testovacej presnosti. V nadpolovičnej väčšine prípadov (6 z 9) spôsobuje pridanie kontextu zvýšenie presnosti, pričom výsledný efekt osciluje medzi zvýšením znížením. Zaujímavý jav nastáva pri kontexte (1x,1x,2x), kedy hodnota presnosti vyskočí na najvyššiu nameranú hodnotu (0.7507) oproti scenáru bez kontextu (0.6884), čo predstavuje zvýšenie o 6.23%. Tento výrazný skok sa podarilo zreplikovať aj pri rôznych nastaveniach počiatočných hodnôt násad pre rozdeľovanie dát do setov, ako aj pri opakovanom trénovaní z rovnakým hodnotami. Pri jednotlivých opakovaniach dosahuje tento scenár mierne variabilné hodnoty, ale všeobecný trend a strmosť nárastu ostávajú stabilné.

Na Obrázku 21 uvádzame porovnanie všetkých kontextov (testovacích scenárov) vo forme súhrnného zobrazenia ich výstupných štatistík z testovacej sady a

ich matice zámen. Pod obrázkom zhrnieme pozorovania, ktoré sme na základe týchto výsledkov uskutočnili.



Obrázok 21. Matice zámen pre všetky kontextové scenáre (mikrokalcifikáty)

Vidíme, že skokové zvýšenie celkovej presnosti kontextu (1x,1x,2x) voči kontextu (1x,1x,1x) je spôsobené významným zvýšením správnej klasifikácie malígnych nálezov,

zatiaľ čo v prípade benígnych nálezov nastal mierny pokles správnych zaradení. Taktiež vidíme, že zlepšenie súvisiace s malígnymi nálezmi (viac skutočne pozitívnych a menej falošne negatívnych predikcií) sa prejavilo aj na zvýšení štatistiky „recall".

Ďalším zaujímavým zistením je významnejší pokles presnosti v prípade kontextu (2x,2x,2x). Tu vidíme takmer totožné skokové zlepšenie klasifikácie malígnych nálezov oproti pôvodnému kontextu (1x,1x,1x) ako v predošlom prípade, ale zároveň aj ďalšie výraznejšie zhoršenie určenia benígnych nálezov, ktoré zráža celkovú presnosť. V tomto prípade sa schopnosť rozlišovať nezhubné nálezy blíži 50%, teda takmer náhodnému výberu.

Od tohto bodu, pri scenároch zahŕňajúcich kontextové elementy (2x,2x,3x) a (2x,3x,3x), dochádza k postupnému zlepšovaniu predikcie nezhubných nálezov, pri súčasnom zachovaní relatívne vysokej presnosti predikovania zhubných nálezov.

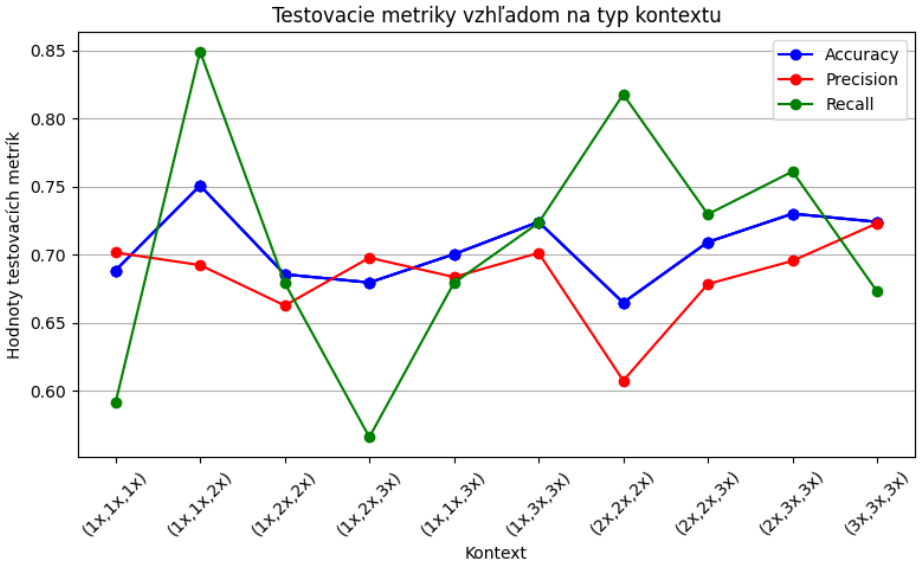
Ďalšie, a posledné zhoršenie výsledkov nastalo pri poslednom testovacom scenári – kontexte (3x,3x,3x), ktorý predstavuje pôvodný obrázok, len trojnásobne odzoomovaný, s rovnakými dátami vo všetkých troch farebných kanáloch.

Vo všeobecnosti z výsledkov môžeme pozorovať, že zahrnutie kontextu spôsobuje zlepšenie klasifikácie plošných nálezov, čo sa dosahuje najmä skvalitnením určovania zhubných nálezov. Tento jav je čiastočne viditeľný aj na maticiach zámen, kde podpriemerné výsledky vo vyhodnocovaní malignity vykazovali často práve tie scenáre, ktoré neobsahovali žiadny dodatočný kontext ((1x,1x,1x), (3x,3x,3x)). Výnimkou z tohto trendu je len kontext (2x,2x,2x). Ďalšou výnimkou hodnou pozornosti je kontext (1x,2x,3x), na ktorý sa teraz pozrieme bližšie.

Kontext 1x,2x,3x je zaujímavý v tom, že stabilne naprieč rôznymi klasifikačnými úlohami, dosahoval väčšinou priemerné až podpriemerné výsledky. Ide o jediný testovací scenár, v ktorom vstupné dáta pozostávajú z troch rôznych typov kontextových elementov, čím pôvodný obrázok prakticky získava najväčšie a najrozmanitejšie množstvo dodatočných dát (oproti ostatným skúmaným scenárom). Ponúka sa teda prirodzená, intuitívna hypotéza, že poskytnutie takejto formy dodatočnej informácie sa pozitívne premietne do výsledkov klasifikácie, k čomu ale nedošlo. Tento jav je možné objasniť charakteristickým efektom pridávania dodatočných informácií do čiernobielych vstupných dát (a teda aj pridávania kontextu k pôvodnému obrázku).

V prípade použitia nemodifikovaného čiernobieleho výrezu (kontexty (1x,1x,1x), (2x,2x,2x) a (3x,3x,3x)) je do siete na vstup privádzaná tá istá informácia vo všetkých farebných kanáloch, čo sa premietne do určitej formy symetrie a duplicity v hodnotách váh, do ktorých model pri trénovaní postupne konverguje. Pridaním čo i len jedného kontextového elementu tak zdvojnásobuje množstvo informácie, na ktorej sa model učí – je teda potrebné aktualizovať a optimalizovať oveľa väčší počet parametrov. Kontext (1x,2x,3x) potom predstavuje ešte silnejšie zvýraznenie tohto javu. Domnievame sa, že výsledky tohto kontextu sú teda limitované potrebou dlhšieho trénovania a taktiež malým množstvom dostupných tréningových dát.

Pre úplnosť uvádzame aj vývoj metrík „prescision" a „recall", kde pre porovnanie uvádzame aj pôvodnú krivku vývoja presnosti.



Obrázok 22. Porovnanie vývoja kriviek presnosti, „precision" a „recall" (mikrokalcifikáty)

Môžeme pozorovať, že trend krivky metriky „precision", hlavne pri vyšších kontextoch, pomerne blízko napodobňuje trend krivky presnosti. To môžeme v tomto prípade vysvetliť tým, že vo väčšine prípadov kolíšu počty skutočne negatívnych odhadov menej ako počty skutočne pozitívnych odhadov, čo znamená, že skutočné pozitíva majú aj väčší vplyv na výslednú presnosť. Znamená to tiež, že príspevok skutočne pozitívnych

odhadov do výpočtu „precision“ bude taktiež väčší než príspevok falošných pozitív, čím sa aj pohyb tejto metriky dostáva pod podobný vplyv ako v prípade presnosti.

Na záver zhrnieme, že zahrnutie kontextu pri klasifikácii mamografických mikrokalcifikátov má potenciál zlepšiť celkovú presnosť predikcie, k čomu prispieva prevažne zlepšená schopnosť identifikácie malígnych nádorov.

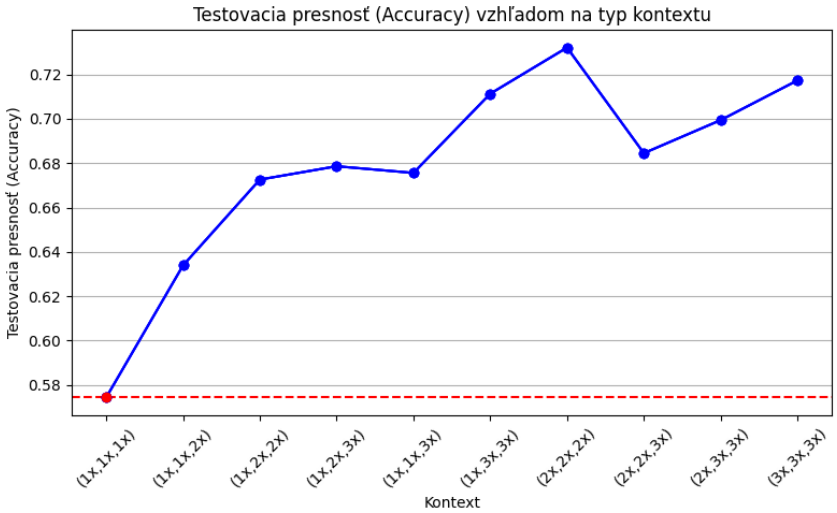
9.2 Kontext v klasifikácii plošných nálezov

V prípade vyšetrovania plošných nálezov sme začali na výrezoch rozlíšenia 224*224, a opäť sme pre každý kontextový scenár vykonali 20 trénovacích epoch. Počiatočná nevyváženosť počtu benígnych a malígnych výrezov nebola tak drastická ako v minulom prípade (853 zhubných nálezov a 747 nezhubných nálezov). Kvôli celkovej konzistencii naprieč úlohami sme obe sady znova rozšírili pomocou osovo prevrátených variant tak, aby výsledné počty súhlasili s počtami v úlohe klasifikácie kalcifikátov (celkovo 2118 výrezov, rovnomerne rozdelených pre obe klasifikačné triedy). Dátové sady pozostávajú z nasledovného počtu vzoriek:

- trénovacia sada – 1462 vzoriek
- validačná sada – 320 vzoriek
- testovacia sada – 336 vzoriek

Analýzu sme vykonali na výsledkoch z testovacej sady.

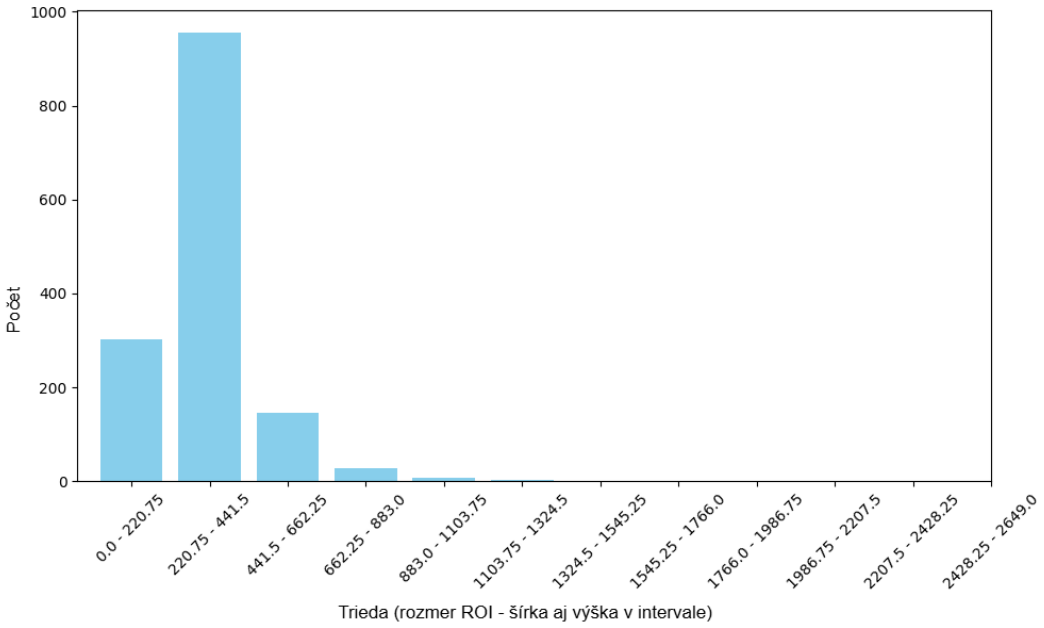
Na Obrázku 23 znázorňujeme vývoj celkovej presnosti („accuracy”) klasifikácie vzhľadom na jednotlivé testovacie scenáre, po vykonaní počiatočnej série experimentov.



Obrázok 23. Vývoj presnosti vzhľadom na použitý kontext (plošné nálezy)

Základným pozorovaním je to pri tejto úlohe to, že klasifikácia plošných nálezov je ťažšia úloha ako klasifikácia mikrokalcifikátov, čo vyplýva zo všeobecne nižších hodnôt presnosti. Domnievame sa, že zložitosť tejto úlohy spočíva najmä vo veľkej variabilite veľkostí, v kombinácii s často nezreteľným ohraničením okrajov nálezov.

Na obrázku môžeme pozorovať jasný primárny trend postupného stúpania presnosti vzhľadom na pribúdajúci kontext. Aby sme pochopili príčinu tohoto javu, spolu s anomálne nízkou presnosťou počiatočného kontextu (1x,1x,1x), uvádzame na Obrázku 24 histogram originálnych veľkostí plošných nálezov tak, ako sa nachádzajú v CBIS-DDSM, teda pred extrakciou, centrovaním alebo škálovaním.

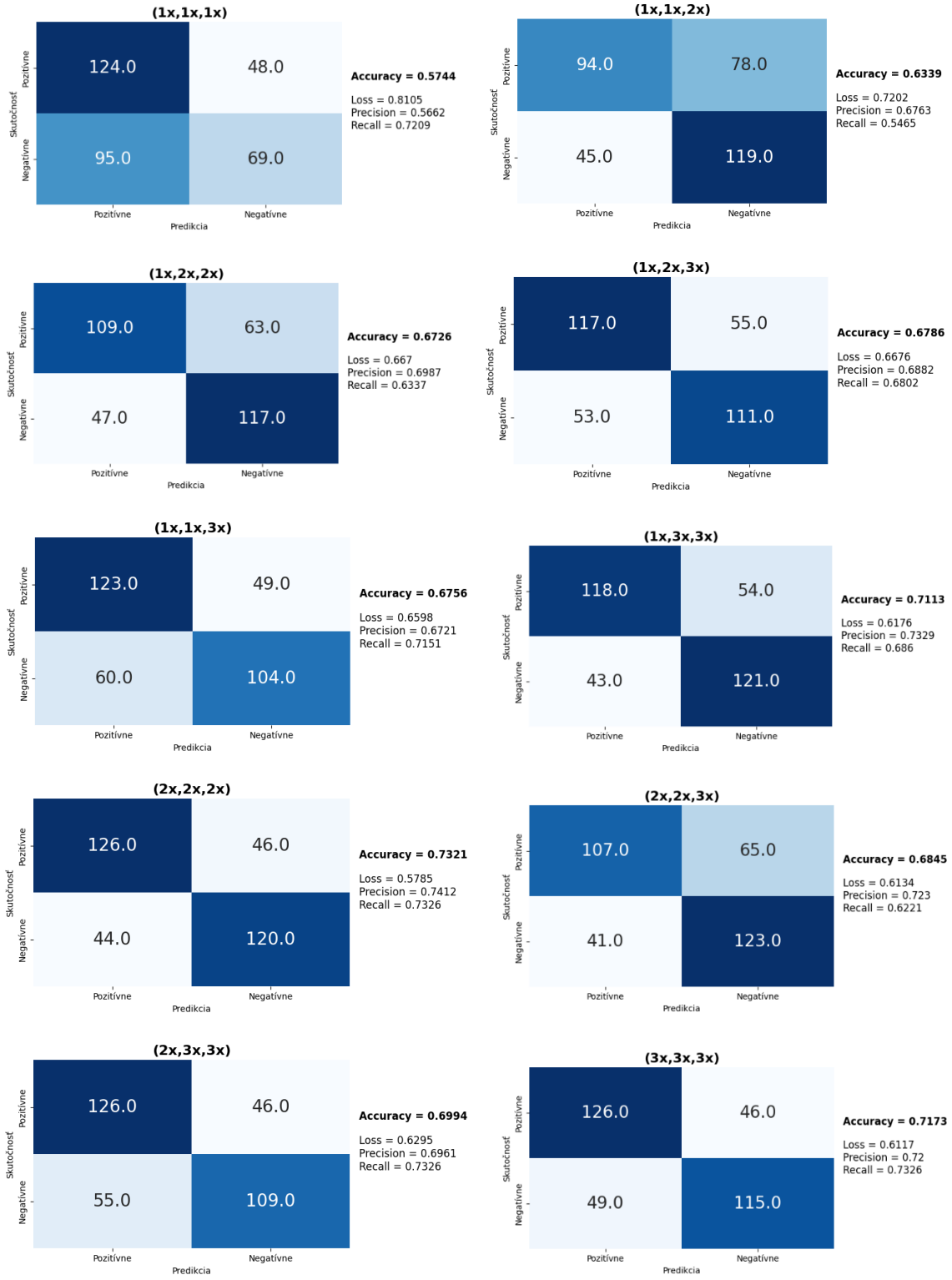


Obrázok 24. Histogram veľkostí plošných nálezov v CBIS-DDSM

Počty v jednotlivých intervaloch udávajú počty plošných nálezov, ktorých šírka aj výška sa nachádza v danom rozmedzí. Vidíme, že veľká väčšina nálezov spadá práve do intervalu 224 – 448 pixelov (čo predstavuje rozsah kontextového rozlíšenia medzi 1x a 2x). Toto vysvetľuje extrémne nízku presnosť pri kontexte (1x,1x,1x) – použitá veľkosť výrezu je príliš malá, aby pokryla priestor celého nálezu.

Taktiež vidíme, že okamžite po pridaní dodatočného kontextového elementu výsledná presnosť prudko stúpa hore, pretože ľubovoľný kontextový element (2x aj 3x) je schopný poskytnúť chýbajúce údaje, ktoré vo väčšine prípadov minimálne pokryjú celú plochu nálezu. Zaujímavý jav, ktorý si na grafe môžeme všimnúť, je aj to, že zahrnutie kontextu spolu s pôvodnými dátami (kontextový element 1x) nikdy nedosiahne takú úroveň presnosti, ako pri jednoduchom kontexte (2x,2x,2x), ktorý vo výsledkoch vychádza jednoznačne najlepšie. hoci pozostáva len z jedného typu kontextového elementu. Na tomto príklade vidíme, že pokiaľ pôvodné dáta nedokážu plne pokryť celú oblasť nálezu, tak ich zahrnutie môže mať na celkovú kvalitu klasifikácie negatívny dopad. Zmiešavanie kontextov nemusí teda priniesť len neutrálny alebo benefičný efekt. Ak teda niektorý kontextový element obsahuje neužitočné alebo negatívne interferujúce informácie, môže to výsledok trénovania zhoršiť.

Na Obrázku 25 uvádzame prehľad matíc zámen pre jednotlivé testovacie scenáre.



Obrázok 25. Matice zámen pre všetky kontextové scenáre (plošné nálezy)

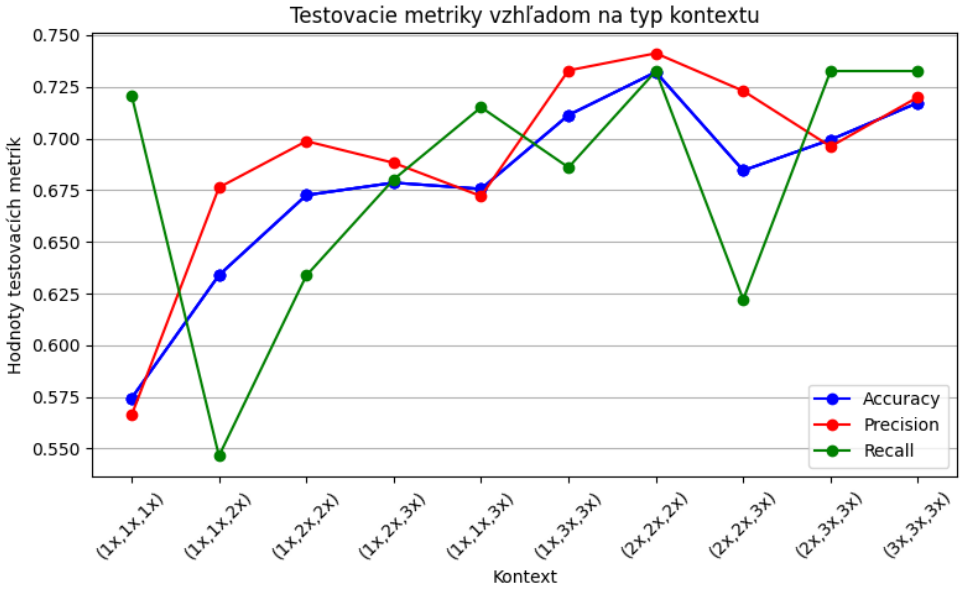
Na základe matíc môžeme pozorovať, že nízka presnosť kontextu (1x,1x,1x) je spôsobená hlavne zlými predpoveďami nezhubných nálezov, kde je celková úspešnosť len 69/164, čo predstavuje 42%. Vidíme, že pridanie len jediného kontextového elementu vyššej úrovne túto štatistiku výrazne zlepšilo (aj keď za súčasného zhoršenia výsledkov pri zhubných nádoroch). Pri ďalšom postupnom pridávaní a kombinovaní kontextu vidíme konsolidáciu kvality predikcií, keď nastáva postupné mierne narastanie celkovej presnosti, pričom úspešnosti pre jednotlivé triedy oscilujú na obe strany, ale nevychyľujú sa významne od seba navzájom. Celkovo teda môžeme povedať, že pridanie kontextu do obrázku plošného nálezu pomáha predikcii benígnych nálezov, pričom ale zlepšenie prebehne skokovo pri zahrnutí nejakého kontextového elementu, a následne sa už výrazne nezlepšuje, iba osciluje. Zároveň sú ale vyššie kombinácie kontextových elementov výhodnejšie, pretože pomáhajú zároveň udržiavať schopnosť klasifikácie zhubných nádorov, ktorá pri použití len nižších kontextových elementov nie je dobrá.

V rámci tejto úlohy sa nám podarilo ukázať vplyv zahrnutia kontextu na výsledné štatistiky klasifikácie plošných nálezov. Na základe zistení môžeme opísať dva osobitné, ale navzájom previazané javy:

- Pri nedostatočnej kvalite (v našom prípade veľkosti) vstupných dát, môže zahrnutie dodatočných kontextových informácii vyššej úrovne pomôcť zlepšiť schopnosť modelu rozpoznať klasifikovanú entitu. Tento efekt môžeme pozorovať pri pridaní elementov vyššieho rádu do základného kontextu (1x,1x,1x)
- Pri dostatočnej kvalite (v našom prípade veľkosti) vstupných dát môže zahrnutie kontextových informácii nižšej úrovne, ktoré neposkytujú užitočnú formu detailu, naopak trénovací proces sťažiť a výsledky zhoršiť. Tento jav môžeme pozorovať pri „spätnom“ pridávaní kontextu, kedy zahrnutie nižšieho kontextového elementu do kontextu (2x,2x,2x) vedie k zhoršeniu výsledkov.

Najlepšie výsledky vykazuje už spomínaný kontext (2x,2x,2x), u ktorého vidíme najlepšiu kombináciu správne pozitívnych a správne negatívnych nálezov. Vzhľadom na tento výsledok sme sa rozhodili vykonať túto testovaciu úlohu aj pre výrezy, ktorých základná veľkosť (tzn. veľkosť kontextového elementu 1x) činila 448*448, čo predstavuje pôvodnú (pred zoškálovaním na 224*224) veľkosť výrezov kontextu (2x,2x,2x).

Na záver podkapitoly uvádzame graf porovnania kriviek vývoja „accuracy“, „precision“ a „recall“:



Obrázok 26. Porovnanie vývoja kriviek presnosti, „precision" a „recall" (plošné nálezy)

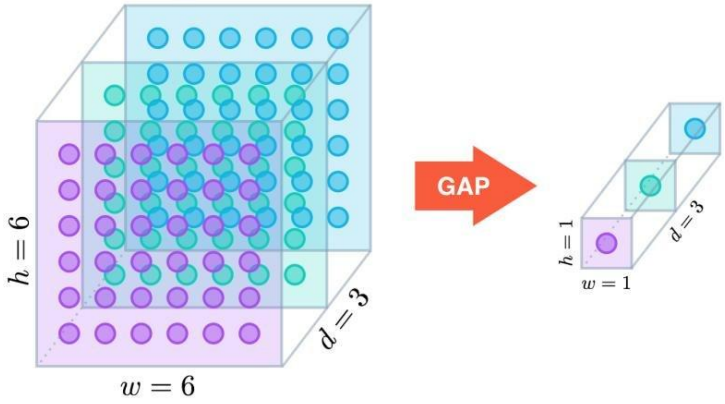
Na grafe môžeme tak ako v prípade klasifikačnej úlohy mikrokalcifikátov pozorovať blízku previazanosť kriviek presnosti a „precision“, zatiaľ čo vývoj metriky „recall“ je viac neusporiadaný a nepredvídateľný.

9.3 Kontext v klasifikácii zväčšených plošných nálezov

V tejto úlohe sme skúmali vplyv pridania kontextu na výrezy plošných nálezov veľkosti 448*448. Ostatné parametre trénovania sme oproti predošlej úlohe s veľkosťami 224*224 zachovali totožné (rovnaký model, počty trénovacích epoch, počty výrezov v oboch triedach a všetkých sadách).

Napriek tomu, že model ResNet50 bol predtrénovaný na obrázkoch veľkosti 224*224, je technicky aj prakticky možné použiť aj väčšie rozlíšenia. Dovoľuje to architektúra samotnej siete. Konvolučné vrstvy sú z princípu veľkostne nezaujaté smerom nahor – konvolučný filter sa jednoducho presúva nad dátami s vyšším počtom riadkov a stĺpcov, a výsledkom budú proporčne väčšie mapy príznakov. Možnosť napojiť výsledný

výstup modelu na plne prepojené vrstvy je zabezpečený blokom vykonávajúcim proces „Global Average Pooling" (GAP), ktorý je zapojený za poslednou konvolučnou vrstvou. Tento proces transformuje výstup z „poolingovej" vrstvy na plochý jednorozmerný vektor veľkosti 2048, ktorý je ďalej možné napojiť na ľubovoľnú plne prepojenú vrstvu. Na obrázku nižšie zobrazujeme príklad fungovania funkcie GAP, ktorá mapy príznakov konverguje do jednorozmerného vektora (obrázok bol prevzatý z článku[36]).



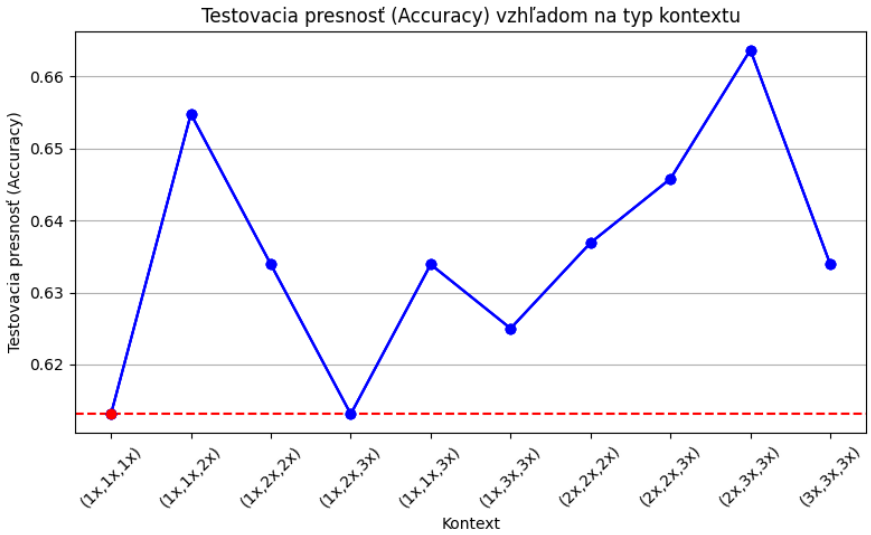
Obrázok 27. Vizualizácia funkcie „Global Average Pooling"

Jediným podstatným rozdielom oproti pôvodnej, „menšej" verzii úlohy je prispôsobenie veľkosti ostatných kontextových elementov vzhľadom na element 1x. Elementy v tejto úlohe mali nasledujúce rozlíšenia:

- 1x – 448*448
- 2x – 896*896 (zoškálované na 448*448)
- 3x – 1344*1344 (zoškálované na 448*448)

Toto hodnoty sme zvolili zámerne preto, aby sme zachovali konzistentnú metodiku oproti predošlým klasifikačným úlohám – cieľom bolo zachovať vzájomné veľkostné pomery medzi elementami (2x = 2*1x, 3x=3*1x). Sekundárnym cieľom bolo otestovať správanie a vplyv značne širokého kontextu.

Na Obrázku 28 opäť uvádzame graf vývoja celkovej presnosti naprieč scenármi s postupným pridávaním kontextových elementov, vzhľadom na referenčný bod základného kontextu (1x,1x,1x).



Obrázok 28. Vývoj presnosti vzhľadom na použitý kontext (zväčšené plošné nálezy)

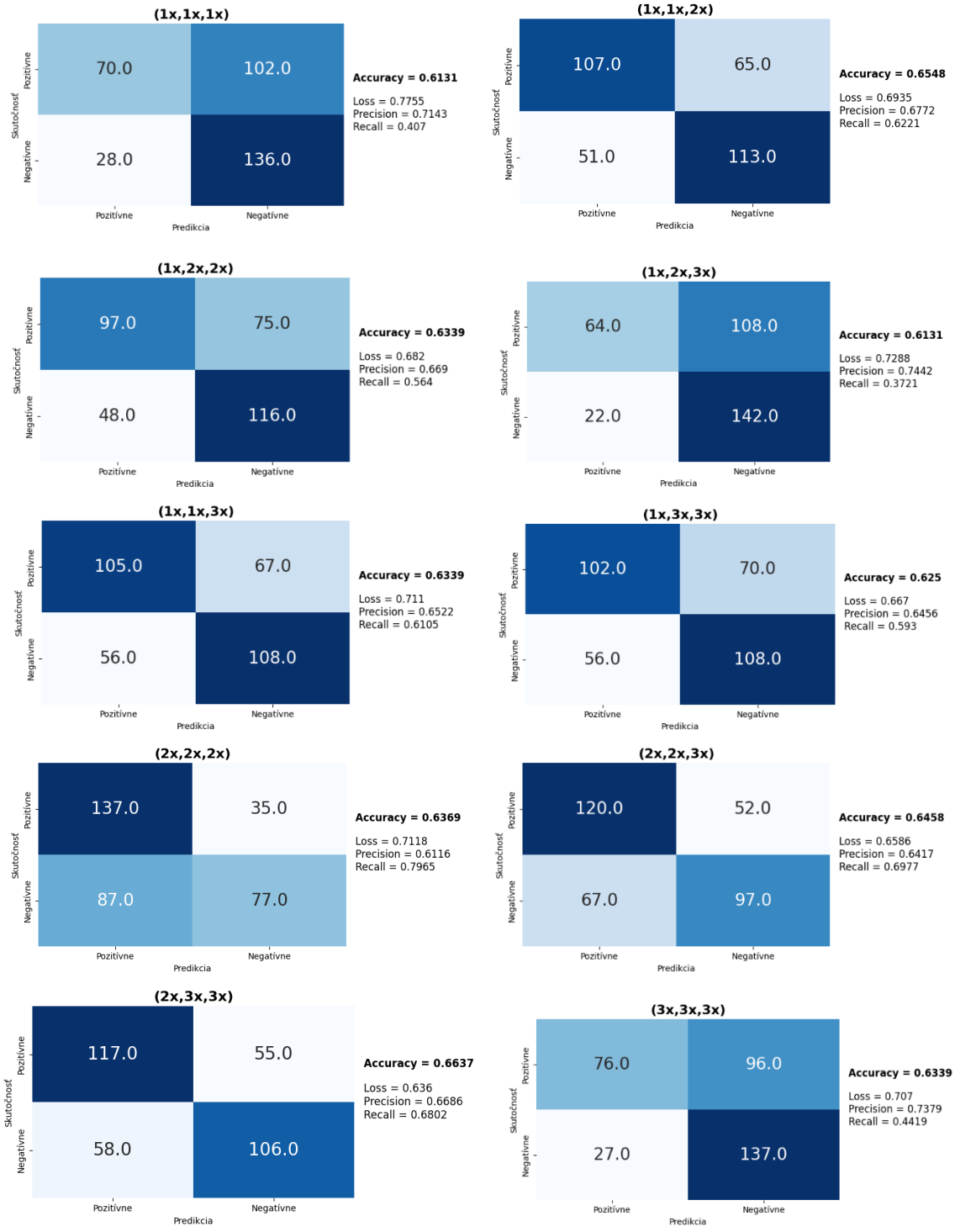
Na grafe je ihneď viditeľná pomerne nízka celková presnosť naprieč celým spektrom testovacích scenárov. Zaujímavé je porovnanie začiatočného kontextu (1x,1x,1x) voči kontextu (2x,2x,2x) v predchádzajúcej „menšej" úlohe. Dáta v týchto kontextoch pochádzajú z rovnakých výrezov veľkosti 448*448, s tým rozdielom, že v minulej úlohe boli zoškálované na 224*224. Vidíme, že v tomto prípade použitie výrezov natívnej veľkosti 448x448 prinieslo výrazne horšie výsledky (presnosť 0.7321) ako použitie škálovanej verzie (presnosť 0.6131).

U tejto úlohy môžeme taktiež pozorovať podobný jav ako pri testovaní kalcifikácii – pridanie jediného vyššieho kontextového elementu k základnému kontextu (1x,1x,2x) prinieslo relatívne významné zlepšenie.

Zaujímavý je interval medzi kontextami (1x,3x,3x) a (2x,3x,3x), na ktorom vidíme stabilný nárast presnosti v každom ďalšom štádiu. Konečný testovací scenár (3x,3x,3x) vykazuje znova zhoršenie oproti priamo predchádzajúcim scenárom. Tento jav môžeme v tomto prípade interpretovať tak, že príliš široký kontext bez zahrnutia nižšieho kontextového elementu, ktorý by poskytoval bližší detail, zhoršuje celkový výsledok. Zároveň to znamená, že v tejto konkrétnej úlohe poskytuje element 2x elementu 3x užitočné dodatočné informácie.

Aj v tejto klasifikačnej úlohe sa stretávame s efektom, kedy maximálne zmiešaný kontextový variant (1x,2x,3x) priniesol relatívne veľmi slabý výsledok. Príčinu si vysvetľujeme obdobne ako v predchádzajúcich prípadoch.

Na nasledujúcom obrázku (Obrázok 29) uvádzame súbor matíc zámen pre všetky scenáre tejto klasifikačnej úlohy.



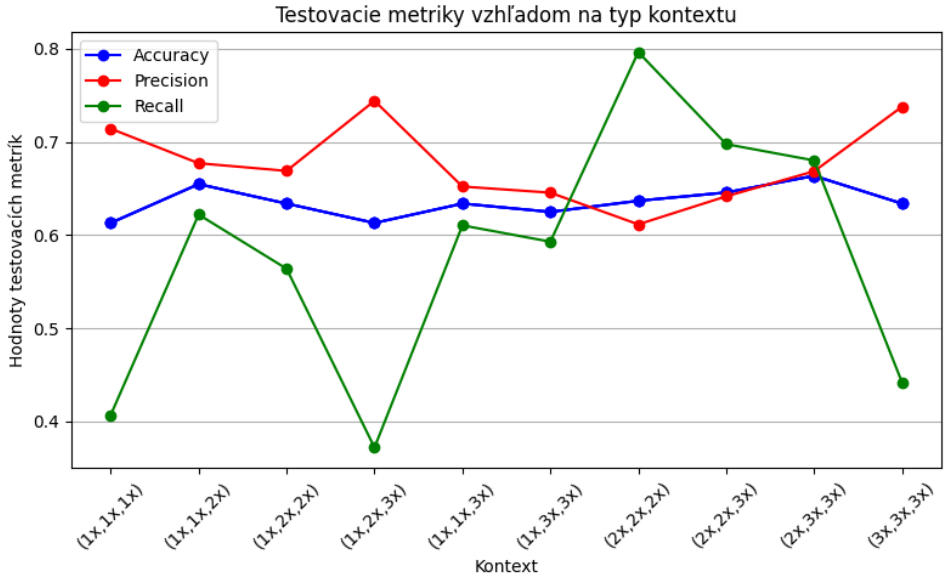
Obrázok 29. Matice zámen pre všetky kontextové scenáre (zväčšené plošné nálezy)

Celkovú zníženú kvalitu výsledkov v porovnaní s predošlou úlohou na menších výrezoch vidíme práve v použití zvýšeného rozlíšenia výrezov. Väčšia veľkosť výrezu dovoľuje uchovať viac nízko-úrovňovej informácie, čo môže spôsobiť, že model sa učí na „neužitočných" detailoch a prehliada podstatné vzorce vyššieho stupňa. Problémom

väčšieho rozlíšenia môže byť aj vyšší nárok na trénovanie – potreba robustnejšieho modelu, vyššieho počtu trénovacích epoch, **viac dát**.

Problém možného nedostatočného počtu dát sme sa rozhodli otestovať pridaním nových dát, ktoré sme získali z databázy EMBED. Trénovaniu na zmiešaných datasetoch sa budeme venovať v nasledujúcej kapitole.

Na záver ešte opäť uvádzame graf porovnania kriviek vývoja jednotlivých sledovaných metrík, na ktorom znova vidíme hlavne podobnosť kriviek presnosti a „precision“:



Obrázok 30. Porovnanie vývoja kriviek presnosti, „precision" a „recall" (zväčšené plošné nálezy)

Zároveň vidíme ešte výraznejšiu chaotickosť krivky veličiny „recall“, čo je spôsobené veľkým kolísaním kvality predikcie zhubných nálezov (tzn. množstva skutočne pozitívnych predikcií). Zvýšenie skutočne pozitívnych odhadov nezvyšuje „recall“ len samo o sebe, ale automaticky priamo úmerne vedie ku zníženiu falošne negatívnych klasifikácií, čo vedie k ďalšiemu zvýšeniu tejto metriky. Analogicky to potom platí aj pri poklese hodnoty skutočných pozitív.

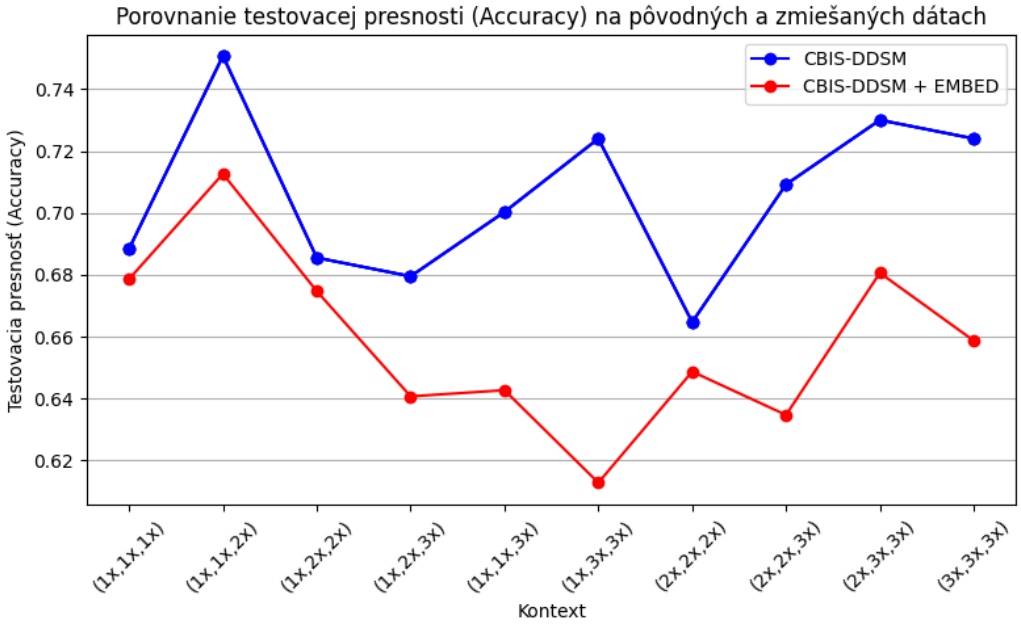
10 Výsledky trénovania na zmiešaných dátach

V tejto kapitole rozoberieme výsledky trénovania na dátach poskladaných z databáz CBIS-DDSM a EMBED. Nové, spoločné datasety sme vytvorili tak, že k pôvodnými, plne vybalansovaným dátam z CBIS-DDSM sme priložili dáta z EMBEDu, vyvážené do najvyššej možnej miery vzhľadom na našu metodiku. Budeme nasledovať príklad z predošlej kapitoly a ako prvej sa budeme venovať klasifikačnej úlohe mikrokalcifikátov, a plošné nálezy opíšeme ako druhé. Dôraz budeme klásť najmä na porovnanie výsledkov voči pôvodnej úlohe s menším počtom dát, s cieľom identifikovať vplyvy nových dát na výkonnosť trénovania. Všetky parametre tréningu sú totožné s parametrami originálneho testovania.

10.1 Kontext v klasifikácii mikrokalcifikátov

Z EMBEDu sa nám podarilo pre trénovanie kalcifikátov získať 877 nových záznamov, z ktorých 732 bolo benígnych a 145 malígnych. Dramatickú nevyváženosť nového datasetu sme vzhľadom na konzistentnosť metodiky dorovnali pridaním osovo prevrátených verzií zhubných nálezov. Kvôli ich veľmi malému počtu sme prikročili aj k použitiu variantu s invertovanými oboma osami – nový obrázok vznikol dvojstupňovým procesom ,kedy prevrátenie osi x bolo nasledované prevrátením osi y. Výsledný EMBED dataset napokon obsahoval 732 nezhubných a 580 zhubných nálezov. Tieto dáta sme zaniesli k pôvodným mikrokalcifikátom z CBIS-DDSM dátam a na ich kombinácii sme vykonali kompletnú sadu testovacích scenárov.

Na Obrázku 31 uvádzame porovnanie naučených presností medzi pôvodnými a novými dátami pre jednotlivé kontexty:



Obrázok 31. Porovnanie vývoja presností na pôvodných a zmiešaných dátach (mikrokalcifikáty)

Z výsledkov môžeme vidieť, že v prípade kalcifikátov spôsobilo pridanie dát z EMBEDu celkové a všeobecné zhoršenie výsledkov naprieč celým spektrom testovacích scenárov. Pre žiadny z použitých kontextov sa nepodarilo dosiahnuť vyššiu presnosť ako pri ekvivalentnom pokuse s pôvodnými CBIS-DDSM dátami. Lepši výkon zmiešaných dát môžeme pozorovať len pokiaľ budeme porovnávať viacero rôznych kontextov medzi sebou a medzi úlohami (napr. kontextový scenár (2x,3x,3x) dosahuje pri zmiešaných dátach lepšie výsledky ako pri použití pôvodných dát s kontextom (2x,2x,2x)).

Zaujímavá je perzistencia javu týkajúceho sa kontextu (1x,1x,2x), ktorý aj naďalej spôsobuje oproti základnému kontextu (1x,1x,1x) solídne zlepšenie, hoci nárast presnosti nie je v porovnaní s pôvodnou úlohou tak dramatický. Napriek tomu aj v tomto prípade poskytuje kontext (1x,1x,2x) najlepšiu celkovú presnosť.

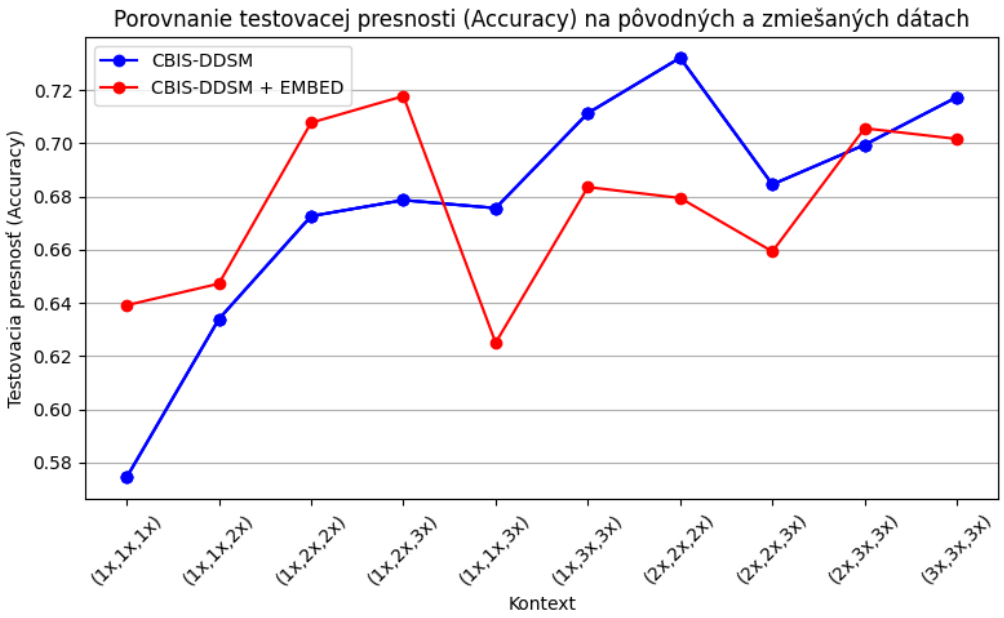
Najväčší rozkol medzi oboma krivkami nastal pri kontexte (1x,3x,3x), kde rozdiel medzi použitím pôvodných a zmiešaných dát činil až 11.12%.

Celkové plošné zhoršenie výsledkov môže byť spôsobené potenciálne nižšou kvalitou kalcifikátových dát v EMBED databáze, čo ale ako laici nie sme schopní posúdiť. Ďalšou príčinou môže byť práve efekt rozdielnej metodiky vytvárania snímok (natívne

digitálne vs. pôvodne analogické snímky) medzi oboma databázami, ktorá spôsobuje aj viditeľné rozdiely v celkovom vzhľade výrezov.

10.2 Kontext v klasifikácii plošných nálezov

Pre potreby úlohy s plošnými nálezmi sme boli schopní extrahovať 697 nových výrezov, z ktorých 537 bolo benígnych a 160 malígnych. Aj tu sme narazili na značnú nevyváženosť, hoci menšiu než v prípade kalcifikátov, takže sme boli schopní pomocou osovo prevrátených variant datasety pre obe klasifikačné triedy plne vyvážiť. Nové dáta sme opäť zakomponovali do originálnych CBIS-DDSM setov a zopakovali sme všetkých 10 testovacích scenárov za nezmenených podmienok. Výsledky a porovnanie presností ukazuje graf nižšie.



Obrázok 32. Porovnanie vývoja presností na pôvodných a zmiešaných dátach (plošné nálezy)

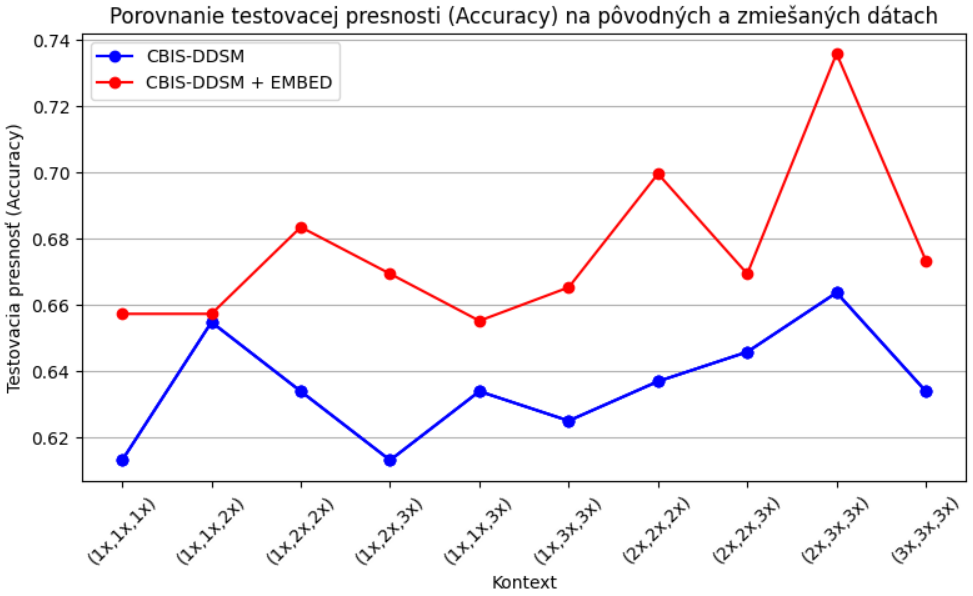
V tomto prípade pozorujeme, že situácia je oveľa rôznorodejšia ako v prípade mikrokalcifikátov. Vo všeobecnosti pozorujeme, že zlepšenie alebo zhoršenie výsledkov je výrazne závislé od použitého kontextu. Vidíme, že v prípade nižších kombinácii kontextových elementov sa zmiešaným dátam darí o dosť lepšie. To je najviac viditeľné

práve na základnom kontexte (1x,1x,1x), kde pridanie výrezov z EMBEDu zlepšilo presnosť o 6.47%. Situácia sa ale obráti na prelome kontextov (1x,2x,3x) a (1x,1x,3x), kedy sa presnosť zmiešaných dát dostane pod úroveň presnosti pôvodných dát, kde už prakticky ostáva. Na tejto úlohe vidíme teda pomerne zaujímavý jav, kedy zväčšenie počtu dát pomáha klasifikácii pri použití nižších kontextov, ale škodí jej pri použití vyšších kombinácií.

Zaujímavosťou je aj prekvapivo dobrý výsledok kontextu (1x,2x,3x), ktorý vo väčšine úloh dosahoval skôr podpriemerné výsledky, ale v tomto prípade vyprodukoval najvyššiu presnosť zo všetkých scenárov.

10.3 Kontext v klasifikácii zväčšených plošných nálezov

V úlohe klasifikácie plošných nálezov so zvýšeným základným rozlíšením 448*448 sa prvý krát naplno prejavil vplyv rozšíreného množstva dát, ktorý môžeme jasne pozorovať na grafe porovnania s pôvodnou úlohou.



Obrázok 33. Porovnanie vývoja presností na pôvodných a zmiešaných dátach (zväčšené plošné nálezy)

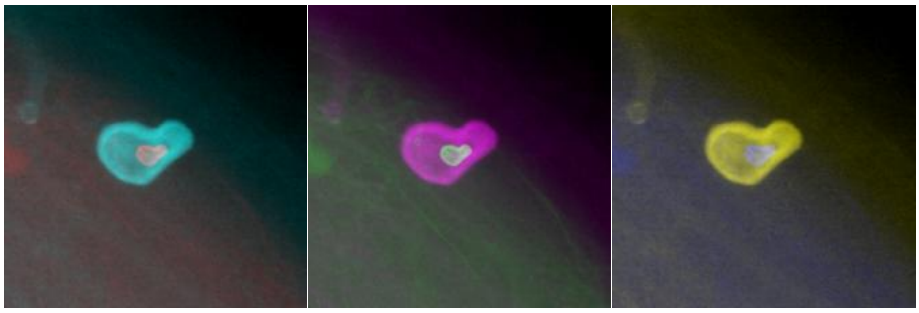
Graf ukazuje plošné zlepšenie presnosti u každého z testovacích scenárov, kde jedinou výnimkou je kontext (1x,1x,2x), u ktorého je zlepšenie zanedbateľné.

Vidíme, že u oboch porovnávaných kriviek dochádza k podobnému trendu postupne rastúcich presností v druhej polovici grafu, v teda v priestore kontextov využívajúcich vyššie kontextové elementy. Tento efekt sa zdá byť určitým spôsobom spoločný pre všetky úlohy nad plošnými nálezmi, pretože ho v určitej forme pozorujeme aj pri úlohe s plošnými nálezmi základnej veľkosti 224*224. Najlepší výsledok pri obidvoch úlohách na zväčšených výrezoch sme zaznamenali na kontexte (2x,3x,3x). Pri zmiešaných dátach sa ale jeho dobrý výkon zvýraznil do takej miery, že suverénne prekonal ostatné kontexty, pričom druhý najlepší kontext (2x,2x,2x) v presnosti prekonal o 3.63%.

Cenou za získané zlepšenie však bola výpočtová náročnosť trénovacieho procesu. Použitie výrezov základnej veľkosti 448*448, spolu so zavedením nových dát z databázy EMBED zvýšilo celkový čas na vykonanie jedného testovacieho scenára na nami použitých počítačoch z jednej hodiny (len CBIS-DDSM dáta, rozlíšenie 224*224) na takmer 7 hodín.

11 Experiment s umiestnením kontextového elementu

Na záver sme vykonali experiment, ktorého cieľom bolo otestovať potenciálny vplyv polohy kontextového elementu na výsledky trénovania. Pod polohou máme na mysli umiestnenie elementu do špecifického farebného kanála vstupného obrázka. Pre pokus sme použili testovací scenár (1x,1x,3x) pre klasifikačnú úlohu mikrokalcifikátov, v ktorom sme postupne menili pozíciu kontextového elementu 3x. Otestovali sme tri možné varianty: (3x,1x,1x), (1x,3x,1x) a (1x,1x,3x). Na Obrázku 34 môžeme pozorovať vizuálny vplyv, ktorý má zmena umiestnenia elementu do odlišných farebných kanálov.



Obrázok 34. Porovnanie vzhľadu výsledného výrezu v závislosti od umiestnenia kontextového elementu 3x

Parametre tréningu (počet epoch, rozloženie sád, použitý model) boli totožné s tými v úlohe trénovania kalcifikátov. V rámci výsledkov tréningu sme sledovali celkovú presnosť, ktorej porovnanie pre jednotlivé varianty uvádzame nižšie.

$$\text{Presnosť } (\ (3x,1x,1x)\ ) = 0{,}7122$$

$$\text{Presnosť } (\ (1x,3x,1x)\ ) = 0{,}7181$$

$$\text{Presnosť } (\ (1x,1x,3x)\ ) = 0.7092$$

Na základe výsledkov konštatujeme, že pozícia kontextového elementu neovplyvňuje výsledok trénovania. Malé rozdiely v presnosti sú prirodzené a vyplývajú z prítomnosti náhodnosti v trénovacom procese (napr. náhodné zamiešanie dát medzi jednotlivými epochami).

12 Záver

V tejto práci sme sa zaoberali tematikou mamografie, a to z viacerých rôznych hľadísk.

Prvým, najdôležitejším hľadiskom bolo trénovanie neurónových sietí pre klasifikovanie nálezov z mamografických snímok na zhubné a nezhubné. Našim cieľom bolo otestovať unikátny prístup k modifikácii vstupných dát použitých na trénovanie, ktorý spočíval v zahrnutí širšieho okolia nálezu vo forme tzv. kontextu.

Aby sme ale boli schopní experimenty vykonať, potrebovali sme najskôr získať a vytvoriť potrebné vstupné dáta, čím sa dostávame k druhému aspektu, ktorému sme sa v práci venovali – mamografickým databázam. Najskôr sme sa detailne oboznámili s dvoma konkrétnymi príkladmi verejných databáz – CBIS-DDSM a EMBED. Popísali sme ich štruktúru, spôsoby ako pracovať z ich dátami a čím sa navzájom odlišujú. Odpozorovali sme však taktiež aj spoločné znaky a najmä podobnosti v pracovných postupoch, ktoré sme u oboch využívali na pristupovanie k dátam, ich zobrazovanie a extrakciu. Tieto všeobecné vzorce sme následne premietli do vytvorenia jednotného objektového modelu, v ktorom sme dokázali abstrahovať a zadefinovať kľúčové procesy práce s mamografickou databázou, a ktorý nám umožnil na ne nahliadať pod zjednoteným rámcom, ale s dostatočnou flexibilitou na zabezpečenie databázovo-špecifických požiadaviek.

Následne sme vyvinuli metodiku a skripty na hromadnú extrakciu dát. Získané pôvodné dáta sme potom modifikovali manipuláciou obrazovej informácie v jednotlivých farebných kanáloch obrázkov, čím sme vytvorili hotové tréningové datasety so zahrnutým kontextom.

Po tejto fáze sme sa venovali extenzívnemu trénovaniu a testovaniu modelov reziduálnych konvolučných neurónových sietí s použitím týchto modifikovaných dát. Po počiatočnej optimalizácii hyperparametrov nami zvolenej architektúry siete sme vykonali 6 základných sád experimentov, každý z nich pozostávajúci z 10 testovacích scenárov, skúmajúcich vplyv rôznej kombinácie kontextových informácií na kvalitu a presnosť klasifikácie.

Na základe našich experimentov sa nám podarilo potvrdiť, že zahrnutie kontextu má jednoznačný a merateľný vplyv na výsledky trénovania. Zároveň z našich výsledkov vyplýva, že vplyv kontextu nie je vôbec priamočiary a jednoznačne popísateľný. Pozorovali sme rôznorodé efekty zahrnutia kontextu, od výrazného zlepšenia celkovej presnosti až po jej mierne zníženie. Vo všeobecnosti a vo väčšine testovacích scenárov naprieč všetkými klasifikačnými úlohami bol však efekt kontextu pozitívny, pričom ale miera zlepšenia (alebo prípadného zhoršenia) bola výrazne závislá od toho, na akom druhu dát sme trénovali (plošné nálezy alebo mikrokalcifikáty) a od použitej kombinácie kontextových elementov. Zaznamenali sme taktiež javy, ktoré sa stali viditeľnými až pri vzájomnom porovnaní rôznych kontextov medzi sebou – napr. variabilný vplyv zahrnutia konkrétneho kontextového elementu v závislosti od zvyšných elementov prítomných v kontexte.

Vychádzajúc z našich výsledkov navrhujeme v štúdiu kontextu na mamografických snímkach pokračovať. V rámci tejto práce sme sa vo veľkostiach kontextových elementov obmedzili len na skúmanie malých, celočíselných násobkov rozlíšení pôvodných výrezov, pričom ale toto rozhodnutie sme učinili čisto z praktického a kapacitného hľadiska. Medzi týmito pevne stanovenými rozmermi existuje celá škála rozlíšení, ktoré sme neotestovali a ktorých konkrétny vplyv nepoznáme. Ich rôznorodé kombinácie potom poskytujú dostatočný priestor na ďalšiu potenciálnu optimalizáciu kvality klasifikácie. Myslíme si, že by bolo zaujímavé zostaviť granulárnejší pohľad na vplyv kontextu, v ktorom by sme vedeli pozorovať vývoj vplyvu po menších skokoch ako len násobkoch pôvodného rozlíšenia, čím by sme možno dokázali postupne konvergovať k tej najoptimálnejšej hodnote.

Jedným s hlavných zistení, ktoré sme v rámci práce učinili, je že efekt zahrnutia kontextu je závislý od subjektov, ktoré klasifikujeme – tzn. konkrétny kontext sa mohol správať inak v klasifikačnej úlohe mikrokalcifikátov než pri úlohe s plošnými nálezmi. Na základe toho navrhujeme preskúmať a otestovať aj iné druhy úloh, potenciálne aj mimo oblasti mamografie. V prípade zotrvania v mamografickom odbore je potom možné vyskúšať iné klasifikačné schémy – napr. trénovať zvlášť na výrezoch pochádzajúcich z rôznych projekcií (CC a MLO). Ďalšou možnosťou je vymeniť predikciu zhubnosti nálezov za iné kritériá – napr. klasifikovať BI-RADS skóre.

V prípade dostupnosti ďalších mamografických databáz a dostatočnej výpočtovej kapacity by bolo zaujímavé zistiť, či dokáže kontext pomôcť aj v prípade už dobre zoptimalizovanej siete, ktorá sa učila na veľkom množstve dát, alebo by v takých prípadoch výhody kontextu oslabli alebo dokonca vymizli.

Na záver je možné navrhnúť otestovanie kontextu na iných architektúrach a modeloch umelej inteligencie a strojového učenia, kde by sme pri vzájomnom porovnaní mohli byť schopní bližšie porozumieť tomu, prečo a ako kontext funguje, prípadne tiež identifikovať špecifické architektúry, pre ktoré je jeho použitie najviac vhodné.

13 Zoznam použitej literatúry

[1] GINSBURG O., CHENG-HAR Y., BROOKS A. - Breast Cancer Early Detection: A Phased Approachto Implementation, *Cancer* , zv. 126, čís. S10, pp. 2379-2393., 2020 - https://doi.org/10.3390/diagnostics15030285

[2] OECD/EUROPEAN COMMISSION - EU Country Cancer Profiles Synthesis Report 2025, 2025 - https://doi.org/10.1787/20ef03e1-en

[3] WHO - Breast Cancer, 2024 – dostupné na: https://www.who.int/news-room/fact-sheets/detail/breast-cancer

[4] POHLODEK K., - Základy mamológie, 2014 – dostupné na: https://www.fmed.uniba.sk/fileadmin/lf/sluzby/akademicka_kniznica/PDF/Elektronicke_knihy_LF_UK/Zaklady_mamologie.pdf

[5] JOHN HOPKINS MEDICINE - Breast Ultrasound, 2025 – dostupné na: https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/breast-ultrasound

[6] LECUN Y., BENGIO Y, HINTON, G. - Deep learning. *Nature*, zv. 521, pp. 436–444., 2015 - https://doi.org/10.1038/nature14539

[7] ORMESHER I. – Convolution Filters, 2020 – dostupné na: https://medium.com/@ianormy/convolution-filters-4971820e851f

[8] KAUDERER-ABRAMS E. - Quantifying Translation-Invariance in Convolutional Neural Networks, 2017 – https://doi.org/10.48550/arXiv.1801.01450

[9] ZAFAR A, AAMIR M, MOHD NAWI N. - A Comparison of Pooling Methods for Convolutional Neural Networks, *Appl. Sci.*, zv. 12, čís. 17, 2022 – https://doi.org/10.3390/app12178643

[10] KAIMING H., XIANGYU Z., SHAOQING R., JIAN S. - Deep Residual Learning for Image Recognition, 2015 – https://doi.org/10.48550/arXiv.1512.03385

[11] PASCANU R., MIKOLOV T., BENGIO Y. - On the difficulty of training Recurrent Neural Networks, 2013 – https://doi.org/10.48550/arXiv.1211.5063

[12] MUKHERJEE S. – The Annotated ResNet-50, 2022 – dostupné na: https://medium.com/data-science/the-annotated-resnet-50-a6c536034758

[13] RUSSAKOVSKY O., DENG J., SU H., KRAUSE J., SATHEESH S., MA S., HUANG Z., KARPATHY A. - ImageNet Large Scale Visual Recognition Challenge, 2015 - dostupné na: https://www.image-net.org/challenges/LSVRC/index.php

[14] PATIL R. A., DIXIT V. V. - Deep Learning for Improved Breast Cancer Detection: ResNet-50 vs VGG16, *International Journal of Electronics and Computer Applications,* zv. 1, čís. 2, pp. 26-31., 2024 - https://doi.org/10.70968/ijeaca.v1i2.1

[15] GAONA Y. J., RODRIGUEZ-ALVAREZ M. J., MORATÓ H. E., MALLA D. C., LAKSHMINARAYANAN V. - DenseNet for Breast Tumor Classification in Mammographic Images, 2021 – https://doi.org/10.48550/arXiv.2101.09637

[16] GENTIGAN S., BING B., GUOYOU Z. - EfficientNet-Based Deep Learning Approach for Breast Cancer Detection With Mammography Images, 2023 – DOI: 10.1109/ICCCS57501.2023.10151156

[17] BANEERJEE S., KABIR H. - An Introductory Implementation of Breast Cancer Detection from Mammograms and Pixel Intensity with Efficient-Net Other Neural Nets, 2024 – https://doi.org/10.1101/2024.05.04.592536

[18] SLIMI H., ABID S., SAYADI M. - Advanced deep learning strategies for breast cancer image analysis, *Journal of Radiation Research and Applied Sciences*, zv. 17, čís. 4, 2024 – https://doi.org/10.1016/j.jrras.2024.101136

[19] KRISHNA G. V., THANGARAJ J. J. - Using convolutional neural networks and AlexNet, a novel technique for improved mammography image segmentation and classification, *Journal of Survey in Fisheries Sciences*, zv. 10, čís. 1S, 2023 – https://doi.org/10.17762/sfs.v10i1S.529

[20] SANDHU J. K., KAUR A., SHARMA C. - EfficientNet-based Cloud Network Infrastructure for Improved Early Detection ofBreast Cancer, *Cuestiones de Fisioterapia*, zv. 54, čís. 3, 2025 – https://doi.org/10.48047/2a4k2v93

[21] WEI T., AVILES-RIVERO A., WANG S., HUANG Y., GILBERT F. J., SCHÖNLIEB C. B., CHEN C. W. - Beyond fine-tuning: Classifying high resolution mammograms using function-preserving transformations, *Medical Image Analysis*, zv. 82, 2022 – https://doi.org/10.1016/j.media.2022.102618

[22] VASWANI A., SHAZEER N., PARMAR N., USZKOREIT J., JONES L., GOMEZ A. N., KAISER L., POLOSUKHIN I. - Attention Is All You Need, 2017 - https://doi.org/10.48550/arXiv.1706.03762

[23] MANSOUR A. G. M. A., ALSHOMRANI F., ALFAHAID A., ALMUTAIRI A. T. M. - MammoViT: A Custom Vision Transformer Architecture for Accurate BIRADS Classification in Mammogram Analysis, *Diagnostics*, zv. 15, čís. 3, 2024 - https://doi.org/10.3390/diagnostics15030285

[24] ZAHOOR S., SHOAIB U., LALI M. I. - Breast Cancer Mammograms Classification Using Deep Neural Network and Entropy-Controlled Whale Optimization Algorithm, *Diagnostics*, zv. 12, čís. 2, 2022 - DOI:10.3390/diagnostics12020557

[25] COHEN R. - The Chan-Vese Algorithm, 2011 - https://doi.org/10.48550/arXiv.1107.2782

[26] BALASUBRAMANIAM S., VELMURUGAN Y., JAGANATHAN D., DHANASEKARAN S. - A Modified LeNet CNN for Breast Cancer Diagnosis in Ultrasound Images, *Diagnostics (Basel)*, zv. 13, čís. 17, 2023 - DOI: 10.3390/diagnostics13172746

[27] ABDULLA M., HABAEBI M. H., GUNAWAN T. S. - Automatic Breast Cancer Detection Using Inception V3 in Thermography - DOI:10.1109/ICCCE50029.2021.9467231

[28] SAWYER-LEE R., GIMENEZ F., HOOGI A., RUBIN, D. - Curated Breast Imaging Subset of Digital Database for Screening Mammography (CBIS-DDSM) [Data set] - *The Cancer Imaging Archive*, 2016 - https://doi.org/10.7937/K9/TCIA.2016.7O02S9CY

[29] YI D., SAWYER R. L., COHN III D., DUNNMON J., LAM C., XIAO X., RUBIN D. - Optimizing and Visualizing Deep Learning for Benign/Malignant Classification in Breast Tumors, 2017 - https://doi.org/10.48550/arXiv.1705.06362

[30] ISO - ISO 12052:2017, 2017 – dostupné na: https://www.iso.org/standard/72941.html

[31] NATIONAL CANCER INSTITUTE - 2025 – dostupné na: https://www.cancer.gov/

[32] JEONG J. J., VEY B. L., BHIMIREDDY A., - The EMory BrEast imaging Dataset (EMBED): A Racially Diverse, Granular Dataset of 3.4 Million Screening and Diagnostic Mammographic Images, *Radiology: Artificial Intelligence*, zv. 5, čís. 1, 2023 - https://doi.org/10.1148/ryai.220047

[33] CLEVELAND CLINIC - BI-RADS, 2025 – dostupné na: https://my.clevelandclinic.org/health/articles/bi-rads-breast-imaging-reporting-and-data-system

[34] KERAS - A superpower for ML developers, 2025 – dostupné na: https://keras.io/

[35] MRAČKO A., CIMRÁK I., VANOVČANOVÁ L., LEHOTSKÁ V. - Deep Learning in Breast Calcifications Classification: Analysis of Cross-Database Knowledge Transferability, *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies*, zv. 1, 2024 - DOI: 10.5220/0012535200003657

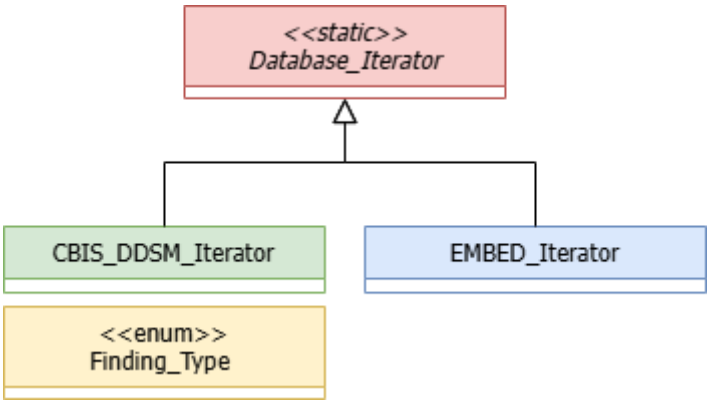
[36] SHUSTANOV A. V., YAKIMOV P. Y. - Modification of single-purpose CNN for creating multi-purpose CNN, *Journal of Physics: Conference Series*, zv. 1368, čís. 5, 2019 – DOI: 10.1088/1742-6596/1368/5/052036

Zoznam príloh

Prílohy

Príloha A: Objektový model skriptov pre prácu s databázami

UML diagram tried:



Ako základ sme zostrojili statickú triedu Database_Iterator, ktorá predstavuje abstrakciu funkcionality iterovania nad záznamami všeobecnej mamografickej databázy. Táto definuje základné atribúty a metódy nevyhnutné pre prácu s databázou. Pri jej tvorbe sme dbali na zachytenie a vystihnutie základných procesov, ktoré sa pravdepodobne budú opakovať pri práci s ľubovoľnou databázou tohto typu. Túto triedu potom dedia triedy CBIS_DDSM_Iterator a EMBED_Iterator, ktoré konkrétnym spôsobom implementujú dané metódy podľa špecifických potrieb a charakteristík danej databázy. V zdedených triedach je zároveň možné základnú funkcionalitu ďalej rozširovať podľa potrieb užívateľa.

Ďalej nasledujú popisy jednotlivých tried a ich atribútov a metód. U zdedených tried sa budeme venovať len nadstavbovým atribútom a funkcionalite doplnenej nad rámec implementácie statických metód z predka.

A1 Trieda Database_Iterator

Atribúty:

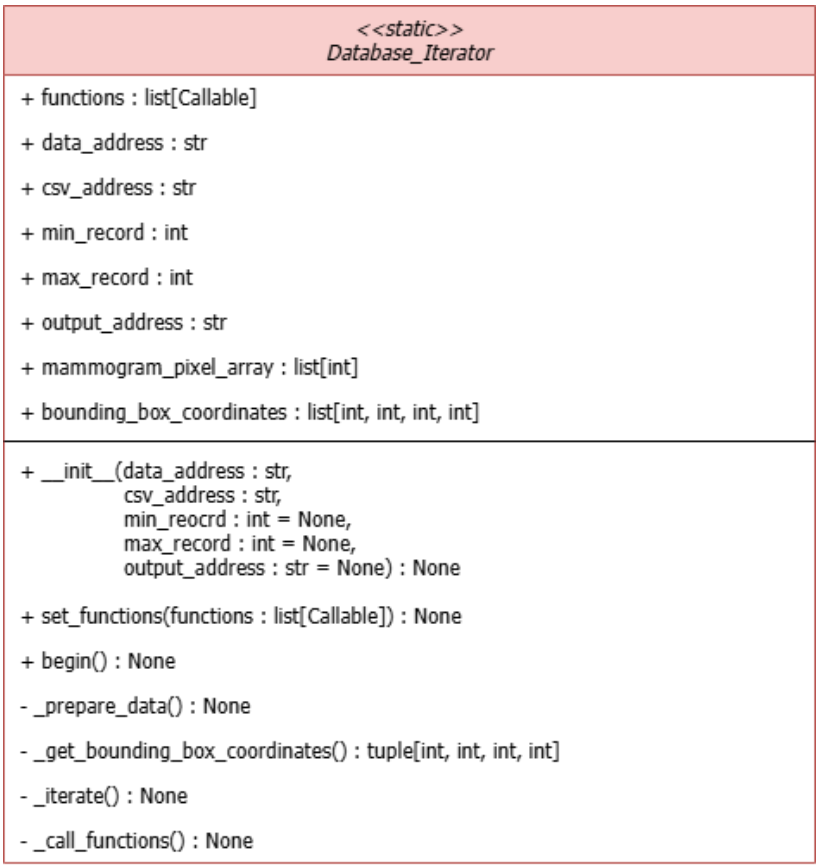
- **functions** : *list[Callable]* – uchováva zoznam funkcií, ktoré má iterátor vykonať v každom iteračnom kole (tzn. pre každú snímku). Funkcie sa vykonávajú v poradí, v akom sú uložené v zozname.
- **data_address** : *str* – absolútna adresa priečinku, v ktorom sa nachádza databáza, resp. miesto uloženia mamografických snímok.
- **csv_address** : *str* – absolútna adresa súboru s metadátami (typicky CSV súbor).
- **min_record** : *int* – index záznamu, od ktorého má iterátor započať spracovanie dát.
- **max_record** : *int* – index záznamu, na ktorom má iterátor ukončiť spracovanie dát.
- **output_address** : *str* – absolútna adresa priečinka, kam chce užívateľ uložiť vygenerované výrezy.
- **mammogram_pixel_array** : *list[int]* – dvojrozmerná matica pixelových hodnôt aktuálne vyšetrovanej snímky. Jej hodnota sa aktualizuje a prepisuje v každom iteračnom kole vo funkcii __iterate().
- **bounding_box_coordinates** : *list[int, int, int, int]* – zoznam súradníc ohraničujúceho obdĺžnika ROI nálezu (dolný okraj, horný okraj, ľavý okraj, pravý okraj) aktuálne vyšetrovanej snímky. Jej hodnota sa aktualizuje a prepisuje v každom iteračnom kole vo funkcii __get_bounding_box_coordinates(). Po tomto kroku k nemu môžu pristupovať ostatné funkcie a ďalej ho upravovať.

Adresy je potrebné zadávať v nasledujúcom formáte - „C:/Folder/Database/".

Argumenty metód, u ktorých je uvedená východzia hodnota sú nepovinné.

Za dvojbodkou je u atribútov uvedený ich typ, u funkcií typ návratovej hodnoty. Názvy typov uvádzame v súlade zo syntaxou jazyka Python.

Atribúty základnej aj odvodených tried sme z praktických dôvodov nastavili ako verejné.



Funkcie typu „public“:

- **__init()__** : *None* – konštruktor pre triedu, ktorý slúži na inicializáciu potrebných atribútov a premenných.
- **set_functions()** : None – funkcia na odovzdanie zoznamu funkcií, ktoré si užívateľ praje vykonať v každom iteračnom kroku, iterátoru - spolu s ich vyplnenými argumentami.
- **begin()** : *None* – pomocná funkcia na manuálne zahájenie iteračného procesu. Štandardne volá funkciu __prepare_data().

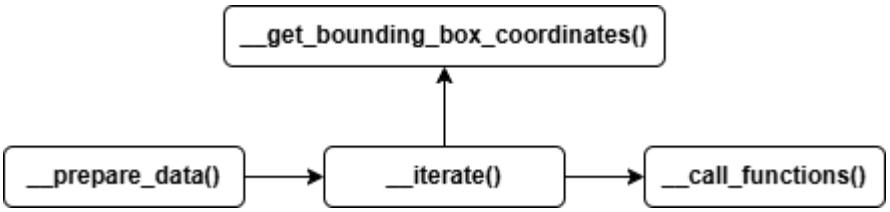
Funkcie typu „private“:

- **_prepare_data()** : *None* – funkcia zodpovedná za načítanie metadát a prípravu databázových súborov (mamografických snímok) na iteráciu. Na konci by mala volať funkciu __iterate().
- **_get_bounding_box_coordinates()** : *list[int, int, int, int]* – funkcia, ktorá vracia štvoricu hodnôt reprezentujúcich ROI koordináty (okraje ohraničujúceho

obdĺžnika) pre aktuálne spracovávanú snímku. Ak je potrebné súradnice dopočítavať nepriamo (napr. z binárnej masky), tu by sa mala nachádzať príslušná logika. Táto funkcia je typicky volaná vo vnútri funkcie __iterate(), najčastejšie pred vykonávaním funkcií zo zoznamu v atribúte "functions".

- **_iterate()** : *None* – funkcia vykonávajúca samotný proces iterácie nad zadanou množinou záznamov. Typicky pozostáva z „for" cyklu, v ktorom postupne volá funkcie v parametri „functions" pre každý jednotlivý záznam, k čomu použije funkciu __call_functions().
- **_call_functions()** : *None* – funkcia ktorá obaľuje funkcionalitu postupného volania funkcií v atribúte „functions".

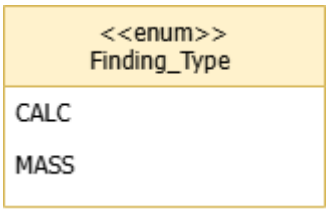
Na nasledujúcej schéme uvádzame štandardné poradie volania privátnych funkcií v iterátore, ktoré odráža typický pracovný postup, aký sme použili pre obe nami spracované databázy:



A2 Trieda CBIS_DDSM_Iterator

Atribúty:

- **finding_type** : *Finding_Type* – premenná, ktorá indikuje podmnožinu dát (mikrokalcifikáty alebo plošné nálezy), s ktorou chce užívateľ pracovať. Na účely tohto rozlíšenia sme zaviedli jednoduchý pomocný enumerátor Finding_Type s dvoma hodnotami – „CALC" a „MASS".



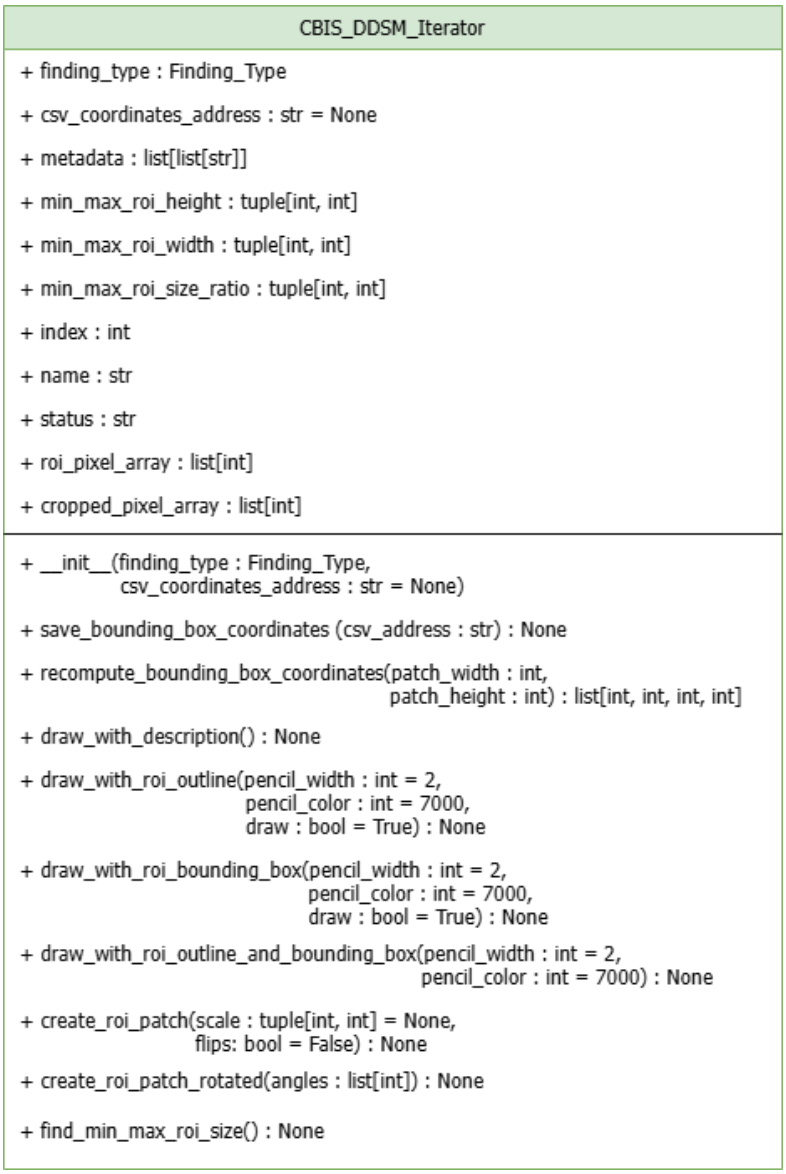
Pokiaľ chce užívateľ spracovať kalcifikáty, nastaví hodnotu na „CALC“, inak na „MASS“. Typ dát je potrebné špecifikovať z dôvodu drobných rozdielov v štruktúre metadát v CSV súboroch prislúchajúcim obom typom.

- **csv_coordinates_address** : *str* – absolútna adresa CSV súboru, v ktorom sú uložené ROI koordináty. Vzhľadom na nutnosť dopočítať súradnice ohraničujúceho obdĺžnika z binárnej masky, čo významne zvyšuje výpočtovú náročnosť určitých foriem spracovania dát (najmä generovania výrezov), má užívateľ možnosť počas iterácie ukladať koordináty pre aktuálnu snímku do zadaného CSV súboru pomocou funkcie save_bounding_box_coordinates(). V nasledovných procesoch iterácie ich potom môže iterátor čítať priamo z tohto súboru, čím sa celý proces výrazne urýchli.
- **metadata** : *list[list[str]]* – pomocná premenná, do ktorej iterátor ukladá všetky metadáta zo súboru na adrese danej základným atribútom „csv_address“. Každý záznam je uložený ako samostatný zoznam hodnôt typu str.
- **min_max_roi_height** : *tuple[int, int]* – ukladá minimálnu a maximálnu výšku ROI nálezu v doposiaľ spracovaných záznamoch.
- **min_max_roi_width** : *tuple[int, int]* – ukladá minimálnu a maximálnu šírku ROI nálezu v doposiaľ spracovaných záznamoch.
- **min_max_roi_size_ratio** : *tuple[int, int]* – ukladá minimálny a maximálny pomer strán ROI nálezu v doposiaľ spracovaných záznamoch.
- **index** : *int* – poradový index aktuálne vyšetrovaného záznamu. Ide o pomocnú premennú, slúžiacu na to, aby aktuálne číslo záznamu bolo k dispozícii vo všetkých funkciách volaných iterátorom. Nastavuje ho iterátor automaticky.
- **name** : *str* – názov aktuálneho záznamu. Nastavuje ho iterátor automaticky.
- **status** : str – indikuje zhubnosť alebo nezhubnosť nálezu z aktuálneho záznamu. Nastavuje ho iterátor automaticky.
- **roi_pixel_array** : *list[int]* – uchováva pixelové hodnoty binárnej masky

- **cropped_pixel_array** : *list[int]* – uchováva pixelové hodnoty predpripraveného výrezu ROI.

Funkcie typu „public“:

- **save_bounding_box_coordinates()** : None – pomocná funkcia na uloženie súradníc ohraničujúceho obdĺžnika ROI pre aktuálnu snímku do CSV súboru.
- **recompute_bounding_box_coordinates()** – *list[int, int, int. int]* – funkcia na základe zadanej výšky a šírky výrezu prepočíta aktuálne hodnoty ROI koordinátov práve spracovávanej snímky. Pokiaľ je to možné, súradnice sa vycentrujú na stred pôvodného ohraničenia nálezu.
- **draw_with_description** : *None* – vykreslí aktuálnu snímku do okna spolu s popisnými informáciami prevzatými z metadát pre daný záznam.
- **draw_with_roi_outline** : *None* – vykreslí aktuálnu snímku do okna so zvýrazneným presným obrysom nálezu, získaným z binárnej masky.
- **draw_with_bounding_box()** : *None* : vykreslí aktuálnu snímku do okna spolu s ohraničujúcim obdĺžnikom okolo miesta nálezu.
- **draw_with_roi_outline_and_bounding_box()** : *None* – kombinuje funkcionalitu oboch predošlých funkcií.
- **create_roi_patch()** : *None* – vygeneruje výrez podľa aktuálne nastavených ROI koordinátov a výsledok uloží ako PNG obrázok na adresu zadanú atribútom „output_address“. Prostredníctvom argumentu „scale“ je možné vytvorený obrázok pred uložením zoškálovať na ľubovoľné rozlíšenie. Pomocou argumentu „flips“ je možné iterátoru prikázať vytvoriť dve osovo prevrátené varianty – podľa osy x a y.
- **create_roi_patch_rotated()** : *None* – vygeneruje rotačné varianty výrezu daného aktuálne nastavenými ROI koordinátmi a výsledok uloží ako PNG obrázok na adresu zadanú atribútom „output_address“. Argument „angles“ obsahuje zoznam uhlov (udaných v stupňoch), podľa ktorých sa majú varianty vytvoriť.
- **find_min_max_roi_size()** : *None* – funkcia slúžiaca na detekciu najväčších a najmenších výšok, šírok a pomerov strán výrezov v rozsahu spracovanej množiny záznamov. Metóda vyšetrí ohraničujúci obdĺžnik aktuálnej snímky a ak nájde nového kandidáta na niektorú z kategórií, aktualizuje hodnoty atribútov „min_max_roi_height“, „min_max_roi_width“ alebo „min_max_roi_size_ratio“.



A3 Trieda EMBED_Iterator

Atribúty :

- **clinical_data_address** : *str* – absolútna cesta ku CSV súboru z klinickými dátami, ktorý je u EMBEDu sprievodným súborom k metadátam.
- **create_database** : *bool* – slúži iterátoru ako indikátor toho, či chce používateľ vytvárať novú lokálnu „sqlite3“ databázu obsahujúcu tabuľky metadát a klinických dát. Databázový objekt je vhodné vytvoriť pri inicializácii prvotnej inštancie

iterátora. Následne sa hodnota môže nastaviť na False, pretože databáza zostáva uložená na disku pre ďalšie použitie, až pokiaľ ju užívateľ manuálne nevymaže.

- **database_join_query** : *str* – SQL príkaz slúžiaci na komunikáciu užívateľa s vytvorenou databázou. Typicky príkaz typu „JOIN“ slúžiaci na spojenie tabuliek, čo predstavuje nevyhnutný krok pre umožnenie následného sekvenčného spracovania dát.
- **metadata** : *list[dict]* - pomocná premenná typu pole, do ktorej iterátor ukladá všetky riadky spojenej tabuľky. Každý riadok predstavuje asociatívne pole kľúčov a hodnôt („dictionary“).
- **index** : *int* – poradový index aktuálneho záznamu spojenej tabuľky.
- **metadata_row** : *dict* – pomocná premenná iterátora, ktorá uchováva hodnoty aktuálne spracovávaného riadku (záznamu) spojenej tabuľky vo forme asociatívneho poľa.

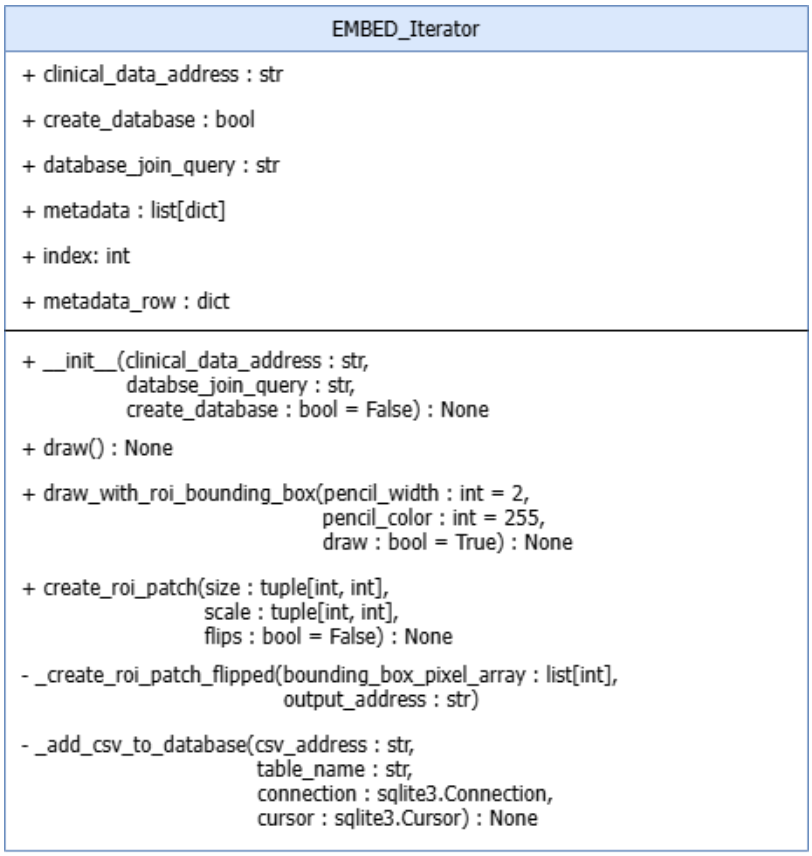
Funkcie typu „public“:

- **draw()** : None - vykreslí aktuálnu snímku do okna.
- **draw_with_bounding_box()** : *None* : vykreslí aktuálnu snímku do okna spolu s ohraničujúcim obdĺžnikom okolo miesta nálezu. Ohraničujúci obdĺžnik sa vykreslí vzhľadom na počiatok súradnicovej sústavy [0, 0] situovaný vľavo hore.
- **create_roi_patch()** : *None* – vygeneruje vycentrovaný výrez nárezu podľa zadanej veľkosti (argument „size“) a škálovania (argument „scale“) a výsledok uloží ako PNG obrázok na adresu zadanú parametrom „output_address“. Ak je indikátor „flips“ nastavený na True, vygenerujú sa spolu so základným výrezom aj tri osovo prevrátené varianty.

Funkcie typu „private“:

- **_create_roi_patch_flipped()** : *None* – pomocná metóda, ktorú volá interátor vo vnútri funkcie create_roi_patch() v prípade, že si užívateľ praje vygenerovať osovo prevrátené varianty výrezu.
- **_add_csv_to_database()** : *None* – pomocná funkcia, ktorá vytvára a pridáva novú tabuľku do lokálnej databázy „embed.db“ na základe CSV súboru z adresy danej

parametrom „csv_address“. Funkcia slúži iterátoru pri prvotnom vložení tabuliek metadát a klinických dát.



Postupy pri vytváraní datasetov vrátane správnych nastavení atribútov iterátorov sú uvedené v skriptoch.

Príloha B: Obsah DVD

Priložené DVD obsahuje:

- Práca v elektronickej podobe (formát PDF)
- Skripty na prácu s databázami a trénovanie neurónovej siete (projekt v jazyku Python)
- Výstupné štatistiky z testovania natrénovaného modelu pre všetky klasifikačné úlohy a testovacie scenáre