

ŽILINSKÁ UNIVERZITA V ŽILINE

Fakulta riadenia a informatiky

AUTOREFERÁT

DIZERTAČNEJ PRÁCE

Žilina, 2024

Ing. Adam Mračko

Žilinská univerzita v Žiline
Fakulta riadenia a informatiky

Ing. Adam Mračko

Autoreferát dizertačnej práce

Klasifikácia mamografických nálezov konvolučnými neurónovými sieťami a ich interpretácia

na získanie akademického titulu „**philosophiae doctor**“ (v skratke **PhD.**)
v študijnom programe doktorandského štúdia
aplikovaná informatika

v študijnom odbore:
informatika

Žilina, apríl 2024

Dizertačná práca bola vypracovaná v dennej forme doktorandského štúdia na katedre softvérových technológií na Fakulte riadenia a informatiky Žilinskej univerzity v Žiline.

Predkladateľ: Ing. Adam Mračko
Katedra softvérových technológií
Fakulta riadenia a informatiky
Žilinská univerzita v Žiline

Školiteľ: prof. Mgr. Ivan Cimrák, Dr.
Katedra softvérových technológií
Fakulta riadenia a informatiky
Žilinská univerzita v Žiline

Externý školiteľ: MUDr. Lucia Vanovčanová, PhD.
II. rádiologická klinika
Lekárska fakulta
Univerzita Komenského v Bratislave

Oponent: prof. Ing. Vanda Benešová, CSc.
Ústav počítačového inžinierstva a aplikovanej informatiky
Fakulta informatiky a informačných technológií
Slovenská technická univerzita v Bratislave

Oponent: prof. Ing. Martin Klimo, PhD.
Katedra informačných sietí
Fakulta riadenia a informatiky
Žilinská univerzita v Žiline

Autoreferát bol rozoslaný dňa: 8. júla 2024

Obhajoba dizertačnej práce sa koná dňa 20.8.2024 o 10:30h. pred komisiou pre obhajobu dizertačnej práce schválenou pracovnou skupinou odborovej komisie v študijnom odbore **informatika** v študijnom programe **aplikovaná informatika**, vymenovanou dekanom Fakulty riadenia a informatiky Žilinskej univerzity v Žiline.

prof. Ing. Ľudmila Jánošíková, PhD.
predsedníčka pracovnej skupiny odborovej komisie
v študijnom odbore **informatika**
v študijnom programe **aplikovaná informatika**

Fakulta riadenia a informatiky
Žilinská univerzita
Univerzitná 8215/1
010 26 Žilina

Abstrakt

Dizertačná práca sa zaoberá prienikom mamografie a strojového učenia, so zameraním na tvorbu a hodnotenie konvolučných neurónových sietí (CNN) pre klasifikáciu mamografických výrezov. Prvotne sa zameriava na mamografický proces hodnotenia a analýzu štyroch databáz: MIAS, INbreast, CBIS-DDSM a OMI-DB. Popisuje tvorbu vlastnej databázy v spolupráci s Onkologickým ústavom Sv. Alžbety. Následne preskúma aplikáciu CNN v mamografii, kde bude pozorovať rôzne hypotézy pomocou experimentov pozorujúcich presnosť modelov. Medzi kľúčové pozorovania je možné zaradiť efektivitu použitia pretrénovaných váh, dôležitosť adresovania nevyváženosti tried a vplyv zmenšenia rozlíšenia výrezov na presnosť modelu. Štúdia tiež objaví limitácie modelov, ako napr. nesprávna klasifikácia nálezov pri prítomnosti viacerých typov abnormalít a prítomnosti šumu v interpretácii, hlavne pri benígnych nálezoch. Experiment s cieľom adresovať limitácie pomocou predspracovania databázy, zahrňujúci nahradenie okolitého prsného tkaniva čiernymi pixelmi, neprinieslo očakávané výsledky. V niektorých prípadoch došlo k zlepšeniu, avšak celkovo bolo pozorované zhoršenie presnosti modelov. Bol vykonaný pokus na zreplikovanie výsledkov z inej štúdie používajúcich SVM modely založené na príznakoch z CNN. Výsledky sa nepodarilo zreplikovať z dôvodu potenciálnej závislosti dát. Ďalšie experimenty využili kombináciu dátových množín CBIS-DDSM a OMI-DB na preskúmanie prenosu znalosti medzi databázami a preskúmanie rôznych CNN architektúr. Bolo preukázané, že kombinácia databáz zlepší presnosť modelu a boli identifikované výrezy, ktorých predikcia bola konzistentne zlá naprieč viacerými modelmi. Prochodom na troj-triedový klasifikátor sa nakoniec podarilo zlepšiť interpretačné výsledky za cenu zníženia počtu správne klasifikovaných malígnych nálezov.

Abstract

Dissertation delves into the intersection of mammography and machine learning, focusing on the development and evaluation of convolutional neural networks (CNN) for classifying mammography patches. The initially it details the mammography assessment process and analyze four databases: MIAS, INbreast, CBIS-DDSM, and OMI-DB. Own database creation in collaboration with Oncological Institute of St. Elizabeth was described. Subsequently, it explores the application of CNN in mammography, investigating various hypotheses and experimental setups to improve model accuracy. Key findings include the effectiveness of using pre-trained weights, the necessity of addressing class imbalance, and the impact of patch resolution reduction on model performance. The study also reveals limitations related to misclassification of abnormalities when multiple types coexist in a patch and the presence of noise in interpretation, particularly in the benign class. Efforts to address these limitations through database preprocessing, including the replacement of breast tissue with black pixels to reduce noise, resulted in mixed outcomes, with improvements in some cases but overall deterioration in model accuracy. Additionally, attempts to replicate superior accuracy results reported in other studies using SVM models based on CNN features were unsuccessful, highlighting potential issues with data dependency. Further experiments combine datasets from CBIS-DDSM and OMI-DB, exploring the transferability of knowledge across databases and the impact of different CNN architectures. The study finds that combining databases enhances model accuracy but also identifies challenging patches consistently misclassified across models. Finally, transitioning to three-class classification improves interpretation results but reduces correct malignant predictions.

Obsah

ÚVOD	6
1 MAMOGRAFIA A RAKOVINA PRSNÍKA.....	7
2 MAMOGRAFICKÉ DATABÁZY A ICH ANALÝZA	9
3 STROJOVÉ UČENIE V MAMOGRAFII	11
4 PREDPRÍPRAVA NA EXPERIMENTY	13
5 VÝSLEDKY VÝSKUMU Z EXPERIMENTÁLNYCH ŠTÚDIÍ.....	14
ZÁVER.....	17
ZOZNAM POUŽITEJ LITERATÚRY	19
ZOZNAM PUBLIKOVANÝCH PRÁC	20

Úvod

Do popredia sa dostávajú rôzne metódy strojového učenia. Typickým cieľom modelov umelej inteligencie je predpovedať výstupy na základe vstupných údajov. Na dosiahnutie presných predpovedí je potrebné takýto model najprv natrénovať. Základnou zložkou úspešného tréningu modelu je súbor údajov, na ktorom sa model učí identifikovať vzťahy medzi vstupmi a výstupmi. Kvalita a množstvo údajov priamo ovplyvňuje dosiahnuteľné výsledky. Práca bude zahŕňať hĺbkovú analýzu niekoľkých známych databáz.

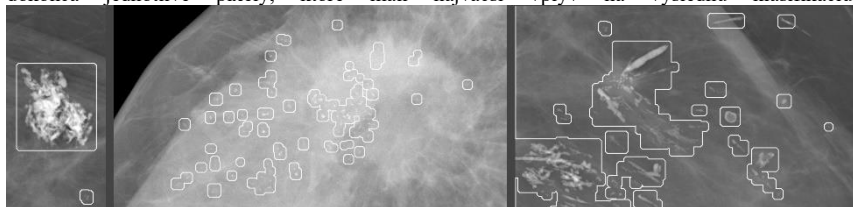
Využitie modelov strojového učenia je široké a rôznorodé a zahŕňa oblasti ako rozpoznávanie reči, počítačové videnie, preklad textu, umelú inteligencia v hrách, bioinformatiku a ďalšie. Ide predovšetkým o úlohy, v ktorých by bolo použitie bežných programovacích metód veľmi náročné alebo dokonca nemožné.

Nezriedkavé je aj použitie modelov v medicínskych oblastiach. Rozšírené sú najmä modely hlbokých konvolučných neurónových sietí, ktoré vynikajú pri analýze obrazových údajov. Možno ich použiť napríklad pri detekcii alebo klasifikácii nálezov. Ich využitie je veľmi vhodné v mamografii, najmä pri diagnostike rakoviny prsníka. Napriek využívaniu moderných technológií môže byť diagnostika tohto ochorenia pre špecializovaných lekárov stále veľmi náročná, čo má za následok vysoký pomer výsledkov biopsie nezhubného tkaniva.

Práca sa zameria predovšetkým na špecifické typy nálezov nazývaných kalcifikáty (Obrázok 1). Kalcifikáty sú bežnou abnormalitou, ktorá sa vyskytuje v mnohých prsníkoch, a delia sa do dvoch kategórií: makrokalcifikáty (väčšie benígne abnormality) a mikrokalcifikáty (malé benígne alebo maligne abnormality). Mikrokalcifikáty sa stávajú podozrivými, keď tvoria zhluky (niekoľko kalcifikátov na malej ploche). Správna klasifikácia môže byť náročná vzhľadom na ich variabilitu vlastností. Niektoré nálezy nemusia byť vizuálne klasifikovateľné odborníkmi, čo otvára potenciál pre konvolučné neurónové siete na odhalenie vzťahov na snímkach, ktoré nemusia byť viditeľné ľudským okom. Na lepšie pochopenie a určenie obmedzení modelov bude zahrnutý celý proces hodnotenia mamografických snímkov. Práca sa bude zaoberať aj priesečníkom oblastí mamografie a strojového učenia. Budú opísané niektoré obmedzenia modelov, ktoré nemusia byť zrejmé.

Rozhodnutie zamerať sa na kalcifikáty ovplyvnila spolupráca s Onkologickým ústavom svätej Alžbety. Jedným z cieľov tejto spolupráce je získať vlastný súbor údajov zameraný na kalcifikáty.

Samotné modely neurónových sietí fungujú ako čierne skrinky. V kontexte práce dostáva klasifikátor vstup v podobe obrázku a na jeho základe sa pokúša správne určiť jeho triedu. Určenie základu rozhodnutia modelu však nie je jednoduché. Metódy interpretácie sa zaoberajú vysvetlením rozhodovacieho procesu modelu. V prípade obrazového formátu môžu tieto metódy určiť zóny alebo dokonca jednotlivé pixely, ktoré mali najväčší vplyv na výslednú klasifikáciu.



Obrázok 1: Ukážka rôznych typov kalcifikátov

1 Mamografia a rakovina prsníka

Prvá kapitola práce sa venuje popisu témy a opiera sa o niekoľko štatistický informácií. Taktiež popisuje aj motiváciu tvorby CNN na klasifikáciu abnormalít.

Rakovina prsníka je najčastejším typom rakoviny u žien, s vysokou úmrtnosťou z dôvodu neskorej liečby a vysokej incidence. Kvôli vysokému výskytu sa zaviedol v mnohých krajinách mamografický skrining ženskej populácie. Včasná detekcia prostredníctvom mamografického skriningu výrazne zlepšuje šance na prežitie, pri detekcii choroby v lokálnom štádiu je pravdepodobnosť prežitia viac ako 5 rokov 99 % [1].

Intervaly skriningu sa líšia podľa rizikových faktorov (najčastejšie každé dva roky), pričom technológia snímania FFDM je preferovanou metódou kvôli svojej ostrosti v detailoch snímky. Jedná sa o technológiu, ktorá vytvára priamo digitálnu snímku. Práca sa zaoberá aj snímkami vzniknutými staršou SFM technológiou, ktorej výsledkom bol obraz prsníkov na filme, ktorý zachytil röntgenové žiarenie.

Dvojitá interpretácia mamografií radiológmi je bežná na zmiernenie chýb. Avšak tento proces je časovo náročný a stresujúci. Správna klasifikácia podozrivých abnormalít dokáže byť pre doktorov veľkou výzvou, čo sa odráža aj vo veľmi nízkej senzitivite, kde sa vykoná množstvo biopsii tkaniva, ktorých finálny výsledok je benígny. V USA je približne 55% až 85% takýchto biopsii [2]. Implementácia počítačových nástrojov na asistenciu, najmä založených na hlbokých konvolučných neurónových sieťach, by mohla zmierniť zaťaženie personálu nemocnice a zvýšiť efektívnosť.

Následne sa kapitola rozdeľuje do niekoľkých podkapitol, ktoré sa venujú opisu mamografického vyšetrenia. Pre vývojára AI modelov sú tieto znalosti nezanedbateľné, nakoľko dopomáhajú k správnej identifikácii limitácií modelov, ale aj k pochopeniu všetkých informácií, ktoré databázy s mamografickými snímkami poskytujú.

Každé mamografické vyšetrenie prechádza mamografických hodnotiacim procesom. Najpoužívanejším hodnotiacim štandardom je BI-RADS hodnotenie [3]. Výsledkom hodnotenia je numerická kategória (celkovo 6 kategórií), ktorá jasne udáva pravdepodobnosť zhubnosti nálezu. Nevýhodou hodnotenia je, že sa naprieč rozdielnymi doktorami môže výrazne líšiť, avšak počas dvojitého nezávislého čítania sa musia doktori zhodnúť na rovnakom hodnotení. Ak sa zhodnú na kategórii 4 (podozrivé – táto kategória ako jediná obsahuje navyše ešte tri ďalšie podkategórie) alebo 5 (veľmi vysoké podozrenie na malignitu) je odporúčané pacientovi, aby sa podrobil odberu tkaniva. Výsledkom je histo-patologický záver, ktorý nesie najobjektívnejšiu informáciu o benígnosti/malígnosti nálezu.

Štandardom každého skriningu sú štyri mamografické snímky. Dva snímky na každý prsník. Najprv je prsník zosnímaný zhora na dol (od hlavy dole) a potom je zosnímaný šikmo z boku. Vďaka týmto dvom pohľadom je možné presne určiť polohu podozrivej abnormality, ktorá je kľúčová pre vyšetrenia ostatnými modalitami alebo pri chirurgickom zákroku.

Po zhotovení snímok začne mamografický hodnotiaci proces, ktorý je rozdelený na nasledovné časti: indikácie, kompozícia prsníka, dôležité nálezy, porovnanie snímok so snímkami z predchádzajúcich vyšetrení a určenie finálnej hodnotiacej kategórie. Kroky sú vykonávané časovo v rovnakom poradí, ako boli definované.

Medzi indikácie spadajú informácie, ktoré sú často krát získané pomocou dotazníka. Prečo pacient prišiel? História pacienta, či už prekonal nejaké ochorenia prsníka. História blízkej rodiny, pomocou ktorej je možné stanoviť genetický vplyv. Má pacient nejaké hmatateľné hrčky a podobne.

Kompozícia prsníka je tvorená štyrmi kategóriami. Kategórie novšieho štandardu určujú stav fibroglandulárneho tkaniva ku schopnosti zakrývať dôležité abnormality. Tkanivo je tvorené aj mliečnymi žľazami. Mladšie pacientky majú prirodzene hustejšie množstvo mliečnych žliaz, ktoré môže zakryť dôležité abnormality. Postupom veku množstvo žliaz ubúda a je nahrádzané tukovým tkanivom.

Medzi najdôležitejšie nálezy patria ložiská, kalcifikáty, poruchy architektoniky a asymetrie. Najčastejšími sú ložiská a kalcifikáty. Práca sa bude zameriavať najviac na nálezy mikrokalcifikátov, ktoré je možné pre jednoduchosť popísať ako veľmi malé biele bodky v mamograme. Diagnostika mikrokalcifikátov je pre rádiológov veľkou výzvou, kvôli ich veľkej variabilite v tvare, veľkosti, počte, rozmiestnení a hustote. Zároveň podozrivé zhluky mikrokalcifikátov môžu viesť k diagnostike DCIS (ductal carcinoma in situ). Jedná sa o pre-invazívny typ rakoviny prsníka, ktorý je zatiaľ len v lokálnom štádiu. Včasná diagnostika tohto ochorenia by pre pacienta znamenala vysokú pravdepodobnosť úplného vyliečenia. Neliečené DCIS môže prerásť do nebezpečnejšieho invazívneho typu.

Porovnávanie snímok so snímkami predchádzajúcimi patrí podľa množstva doktorov za jedno z najdôležitejších častí vyšetrenia. Sleduje sa dynamika prsníka v čase. Je možné odhaliť predtým nevidené nálezy skryté pod mliečnymi žľazami, ale aj sledovať už objavené nálezy. Napríklad, ak sa nejaký zhluk mikrokalcifikátov razantne rozširuje, tak by sa malo výrazne zvýšiť podozrenie na prítomnosť rakovinotvorného tkaniva.

2 Mamografické databázy a ich analýza

Jedným z najdôležitejších prvkov vývoja modelov neurónových sietí sú kvalitné dáta, na ktorých sa bude vykonávať tréning a testovanie modelov. Kapitola sa zamerala na prieskum a analýzu známych databáz. Rozhodujúcim faktorom bolo, aby databáza obsahovala celý mamogram, informáciu o type abnormality a jej presnú polohu. Mamografické snímky sú veľmi vysokého rozlíšenia aby dokázali zachytiť aj jemné detaily. Pre neurónové siete je tak veľké rozlíšenie prekážkou, nakoľko pri malom množstve dát by bolo pre nich náročné identifikovať podstatnú oblasť s abnormalitou. Druhou limitáciou je pamäťová náročnosť na GPU hardvér pri tréningu modelov. Možným riešením by mohlo byť zmenšenie obrazu, avšak prieskumy preukázali, že pri zmenšení obrazu dochádza k strate dôležitých detailov, čo sa odráža na presnosti klasifikácie. Je teda bežnou praxou vývoj modelov, ktoré sa zameriavajú na klasifikáciu výrezov s abnormalitami z celého obrazu. Hlbšie analyzované boli štyri databázy: MIAS [4], CBIS-DDSM [5], INbreast [6] a OMI-DB [7].

MIAS databáza obsahuje mamogramy zhotovené starou SFM technológiou. Celková kvalita snímok je veľmi nízka v porovnaní s ostatnými databázami. Obsahuje iba 123 unikátnych nálezov. Snímky obsahujú množstvo artefaktov a redundantných elementov (napr. písomné informácie o strane a pohľade). Výhodou sú závery patologických výsledkov z biopsií.

CBIS-DDSM je veľmi známou používanou databázou. Podobne ako MIAS snímky sú zhotovené SFM technológiou, avšak sú digitalizované vo vysokom rozlíšení. Obsahuje len nálezy ložísk a kalcifikátov s histo-patologickými závermi. Nálezy je možné lokalizovať v mamograme pomocou priložených binárnych masiek. Celkovo obsahuje 3568 nálezov. Počas spracovania bolo odhalených niekoľko nevýhod. Niektoré masky nemali zhodné rozlíšenie s originálnym mamogramom, ku ktorému patrili. Iné masky ukazovali na oblasti tesne vedľa abnormality. Podobne ako pri MIAS databáze, boli prítomné aj ostatné nevýhody spájajúce sa s použitím SFM technológií.

INbreast obsahuje snímky FFDM technológie vysokého rozlíšenia. Ako jediná z databáz obsahuje individuálne označené kalcifikáty (ostatné databázy označujú celú skupinu/zhluk). Obsahuje všetky bežné typy abnormalít. Nevýhodami sú malé množstvo dát, konkrétne 501 nálezov a absencia histo-patologických výsledkov. Namiesto histo-patológie sú uvedené subjektívnejšie BI-RADS hodnotenia z vyšetrení.

OMI-DB databáza je veľmi komplexnou databázou s obmedzeným prístupom. Je potrebné vyžiadanie povolenia o prístup, kde je nutné splniť niekoľko podmienok a zaplatiť poplatok za sprostredkovanie dohody. Pre prácu s ňou je vhodné použitie oficiálnej voľne dostupnej python knižnice omidb. Pre spracovanie dát je potrebné najprv vytvoriť vlastný filter na získanie potrebných dát. Obsahuje histo-patologické výsledky. Jej výraznými výhodami v porovnaní s ostatnými databázami je to, že obsahuje aj pacientov, u ktorých nebol rádiológmi objavený žiadny podozrivý nález (nemali nikdy vykonanú biopsiu), či prítomnosť snímok pacienta z predchádzajúcich vyšetrení, predtým než mu bola objavená podozrivá abnormalita. Obsahuje všetky bežné abnormality (najpočetnejšie sú ložiská a podozrivé kalcifikáty).

Okrem vyššie spomenutých databáz sa jedna podkapitola venovala tvorbe vlastnej databázy v spolupráci s Onkologickým ústavom sv. Alžbety v Bratislave. V čase tvorby tohto dokumentu je doktormi vykonávaná manuálna anotácia podozrivých zhľukov mikrokalcifikátov, na ktorých bola vykonaná biopsia. Prvá fáza anotácie by potenciálne (horná hranica) mohla priniesť až 1552 nových výrezov, avšak skutočný výsledný počet nie je možné úplne odhadnúť.

Po vykonanej analýze databáz bolo možné zhodnotiť, že najvhodnejšími databázami pre tvorbu CNN budú CBIS-DDSM a OMI-DB. Hlavnými dôvodmi je vysoké rozlíšenie snímok a prítomnosť objektívnych histo-patologických výsledkov.

3 Strojové učenie v mamografii

Kapitola sa zaoberá prienikom oblasti strojového učenia a mamografie. Najprv sa zameriava na potenciálne použitie konvolučných neurónových sietí, pričom uvedie príklady jednotlivých štúdií.

Prvým popísaným použitím je predikcia budúcnosti. Takéto modely sú častokrát označované ako modely predikcie rizika. Ich cieľom je predikovať riziko vývinu rakoviny u pacienta. Pôvodné modely predikovali riziko čisto na základe veku, dedičnosti, hormonálnej liečby a podobne. Pacienti so zvýšeným rizikom mali odporúčené častejšie skríningové vyšetrenia. Nové modely používajúce aj mamogramy pacienta preukázali priaznivo lepšie výsledky v predikcii rizika.

Druhou popísanou úlohou je klasifikácia a detekcia (+ segmentácia) abnormalít. Jedná sa o najbežnejšiu úlohu pre konvolučné neurónové siete. Môže sa jednať o klasifikáciu výrezov s podozrivou abnormalitou do benígnej alebo malígnej triedy. Riešenie bežnej detekcie abnormalít v celom mamograme je náročnejšou úlohou. Databázy totiž to obsahujú len informácie o podozrivých abnormalitách (nepodozrivé abnormality sa nepodrobia biopsii). V prsníku sa zvyčajne nachádza viac abnormalít, ktoré by sieť mohla lokalizovať, avšak v databáze dané nálezy nie sú popísané. Validácia takéhoto modelu môže byť teda veľmi náročná.

Poslednou úlohou je lokalizácia nebezpečných zmien v čase, kde by doktor dokázal napr. z dvoch (alebo viacerých) snímok lokalizovať nebezpečné zmeny. Tento prístup je veľmi neprebádaný, nakoľko množstvo dát, s ktorými by bolo možné daný model vytvoriť, je veľmi limitujúce. Zároveň dve snímky toho istého prsníka nevyzerajú úplne zhodne. Prsník môže byť viac pootočený (rozdielna poloha medzi snímkami), zväčšený/zmenšený a podobne.

Po rozobratí potenciálneho použitia konvolučných sietí práca upozorní na niekoľko limitácií modelov. Niektoré limitácie nemusia byť hneď očividné. Môže sa napríklad jednať o nemožnosť objavenia abnormalít u mladších pacientiek v hustým fibroglandulárnom tkanivom, ale aj nemožnosť správnej klasifikácie iba na základe vizuálneho výstupu. Napríklad pri ložiskách (aj iných abnormalitách) nemusí byť možné správne klasifikovať nález len pomocou mamogramu. Keď je počas skríningového vyšetrenia objavené ložisko, je bežným postupom vykonanie doplnkového sonografického vyšetrenia pre zistenie pôvodu abnormality. Na mamograme môže ložisko pôsobiť nezhubne s pravidelnými okrajmi, zatiaľ čo sonografické vyšetrenie odhalí veľkú nepravidlosť okrajov, ktoré razantne zvyšuje podozrenie na prítomnosť rakoviny. V neposlednom rade je problémom aj to, že väčšina modelov je zaujatá. Databázy obsahujú len podozrivé nálezy, ktoré boli poslané na biopsiu tkaniva. Preto môžu mať modely problém pri klasifikácii úplne jednoduchých benígnych abnormalít (nakoľko ich nikdy nevideli). Veľmi dobrým príkladom je špeciálnym typ mikrokalcifikátov, ktoré sa volajú cievné calcifikáty. Táto limitácia bola objavená počas našej štúdie. Daný typ calcifikátov má množstvo spoločných znakov s malígnymi calcifikátmi, avšak pre doktora je triviálne ich správne kategorizovať, nakoľko dokáže rozlíšiť, že sa nachádzajú v cieve.

Nasledujúca kapitola sa zaoberá teoretickou stránkou architektúr konvolučných neurónových sietí. Cieľom nie je opísať úplnú hĺbku a základy konvolučných sietí. Skôr poukázať na prelomy v architektúrach, ich nové poznatky, ktoré dopomohli dosahovať všeobecne lepšie výsledky ako predchádzajúce architektúry. Moderejšie architektúry zvyčajne používajú poznatky, ktoré boli získané pri tvorbe starších architektúr. Popísané sú architektúry VGG, InceptionV3 ResNet, DenseNet a EfficientNet. Všetky popísané architektúry boli experimentálne použité počas jednotlivých experimentov.

Posledná kapitola popisuje interpretačnú metódu Grad-CAM [8], ktorá bola použitá na interpretáciu modelov. Metóda umožňuje zvýrazniť podstatné zóny vo vstupnom obraze, ktoré mali na výslednú klasifikáciu najväčší vplyv. Neurónové siete samé o sebe fungujú ako modely čiernej skrinky. Nie je možné určiť, ako vnútorné fungujú a aké majú pravidlá. Pomocou interpretácie je však možné túto limitáciu čiastočne obísť. Je možné definovať a potvrdiť, či sa model rozhoduje na základe dôležitých rádiologických príznakov alebo nie. Pri medicínskych modeloch je interpretácia povinnosťou.

4 Predpríprava na experimenty

Kapitola je rozdelená na štyri časti. Prvá sa venuje stáži v onkologickom ústave sv. Alžbety v Bratislave, ktorá vznikla v spolupráci s fakultou. Pomocou nej bolo možné sa zoznámiť s procesmi vyšetrenia a používanými systémami v ústave. Vďaka spolupráci boli objasnené niektoré limitácie modelov, ale aj ozrejmené dôležité rádiologické príznaky pri skriningových vyšetreniach.

Druhá kapitola sa venuje pracovnému prostrediu, a teda softvéru a hardvéru, ktorý sa počas práce používal. Pre neurónové siete bola použitá PyTorch knižnica pre jazyk Python. V kapitole sú zahrnuté aj všetky dôležité knižnice.

Tretia kapitola sa venuje rôznym predspracovaniám databázy. Zahrnuté sú len také predspracovania, na ktorých boli vykonané experimenty a sú zahrnuté v práci. Celkovo bolo vytvorených viacero predspracovaní, niektoré však nevedli k uspokojivým výsledkom alebo boli nahradené lepším predspracovaním. Z databázy CBIS-DDSM sú popísané predspracovania s menami CBIS-DDSM-3 a CBIS-DDSM-6. Neskôr sa podarilo získať prístup k databáze OMI-DB, vďaka ktorej boli vytvorené nové dátové množiny. Tento krát sa používali dáta z oboch databáz naraz. Vzniknuté množiny boli Combination 1, Combination 3 a Combination 5. Všetky množiny okrem Combination 5 slúžili na binárnu klasifikáciu do benígnej/malígnej triedy. Práca bola výhradne zameraná na nálezy mikrokalcifikátov, ktorých rozmery nepresahovali 674x674 pixelov. Combination 5 bola doplnená o tretiu triedu so zdravým tkanivom, ktorá vznikla prevažne z pacientov OMI-DB databázy, ktorí nemali žiadny histo-patologický záznam v žiadnom vyšetrení. Okrem zdravého tkaniva sa v novej triede nachádzali aj ostatné abnormality ako ložiská a makrokalcifikáty.

Posledná kapitola sa venovala použitým metrikám. Uvedené sú vzťahy na výpočet jednotlivých metrik. V práci sa je možné stretnúť s metrikami: presnosť, senzitivita, špecificita, F1 skóre, Matthewsov koleračný koeficient a matica zámen.

5 Výsledky výskumu z experimentálnych štúdií

Posledná kapitola sa venovala návrhu a tvorbe konvolučných modelov. Ešte pred experimentami bol popísaný všeobecný proces tvorby. Prvé experimenty sa venovali len množinám z databázy CBIS-DDSM. Neskôr sa podarilo získať prístup do databázy OMI-DB a experimenty sa začali venovať množine vzniknutej z kombinácie.

Prvý experiment preskúmaval výhodnosť použitia predtrénovaných váh v porovnaní s čerstvo inicializovanými váhami. Boli použité predtrénované váhy na ImageNet datasete. Jedná sa o jeden z najznámejších datasetov, ktorý pozostáva z celkovo 1000 rôznych výstupných tried. Triedy sú zamerané na rozličné objekty, môže sa jednať o zvieratá, rastliny, nástroje a podobne. Myšlienkou použitia predtrénovaných váh spočíva v tom, že model sa už naučil na rôznych objektoch extrahovať základné príznaky (napr. čiary, polkruhy, atď.), a teda sa nemusí všetko učiť od začiatku ako pri čerstvo inicializovaných váhach. Pri predtrénovaných váhach sa testovalo s dvoma variantami. V prvej boli všetky váhy prístupné počas tréningu. V druhej bola prístupná iba posledná klasifikačná vrstva. Experiment preukázal, že použitie predtrénovaných váh s plne odomknutými váhami dosiahlo najlepšie výsledky. Najlepší model dosiahol 78,8% presnosť. Použitie čerstvých váh zhoršilo presnosť v priemere o 3,2%.

Ďalší experiment sa zaoberal balansujúcimi metódami. Nakoľko v tréningovej vzorke databázy CBIS-DDSM prevládajú benígne nálezy, model ma bez nich tendenciu preferovať klasifikáciu do benígnej triedy. Boli testované dve bežné metódy „oversampling“ a „class-weights“. Okrem nich bolo testované aj balansovanie pomocou dátovej augmentácie, avšak táto metóda pridáva modelu počas tréningu nové umelo vygenerované snímky, a teda je očakávané, že si model poradiť lepšie. Pri oversamplingu sa modelu ukazujú snímky, ktoré už predtým mohol vidieť. Metóda si viedla lepšie ako model bez použitia metód na balansovanie, avšak v porovnaní s ostatnými dopadla najhoršie. Druhou najlepšou metódou boli class-weights, ktoré počas tréningu nastavujú vybraným triedami rozdielne ceny pre účelovú funkciu. Pri vhodnom nastavení veľkosti kroku dokázala metóda dosahovať rovnaké výsledky ako najlepšia metóda pomocou augmentácie.

Nasledoval experiment, kde bolo testované použitie menšieho rozlíšenia výrezov ako 674x674 pixelov, konkrétne sa jednalo o 224x224 pixelov (bežné rozlíšenie vstupov pri konvolučných sieťach). Veľké rozlíšenie prináša so sebou množstvo nevýhod. Napr. výrazne pomalší tréning a väčšie nároky na pamäť GPU. Primárne zameranie práce bolo na nálezy mikrokalcifikátov, ktoré sú veľmi malé. Zmenšenie výrezov viedlo k silnému zhoršeniu presnosti v priemere o 6%.

Znalosti z predchádzajúcich experimentov boli použité na tvorbu najlepšieho modelu databázy CBIS-DDSM. Bola testovaná regularizácia pomocou znižovania váh. Vo všeobecnosti modely, ktorých váhy si držia nižšiu veľkosť, majú tendenciu byť stabilnejšie a podávať lepšie výsledky. Pridanie regularizácie prispelo k zvýšeniu presnosti v priemere o 1,8%. Najlepší model dosiahol presnosť 80,5% (Obrázok 2). Súčasťou kapitoly sú aj Grad-CAM interpretácie.

Piaty experiment sa okrajovo zamerl na nálezy ložísk namiesto mikrokalcifikátov. Jednalo sa o jednoduchý experiment bez použitia pokročilejších metód (bez augmentácie, regularizácie, a pod.). Podarilo sa dosiahnuť presnosť 76,6%.

Doteraz sa práca zaoberala len nálezmi, ktorých rozmery nepresahovali 674x674 pixelov. V prípade menších nálezov ako dané rozmery bola do výrezov doplnená okolitá oblasť z mamogramu. Interpretácia z predošlých experimentov ukázala veľké množstvo šumu (pre model

boli podstatné aj zóny bez nálezu, resp. ďaleko od nálezu) pri niektorých klasifikáciách. Nasledoval experiment, ktorý namiesto dopĺňania okolitej oblasti z mamogramu doplnil čierne okraje, tak aby mal výsledný výrez taktiež rozlíšenie 674x674 pixelov. Cieľom bolo redukovať množstvo šumu v interpretáciách, nakoľko sa model bude môcť lepšie zamerať na to, čo je podstatné. Pridanie čiernych okrajov spôsobilo zhoršenie presnosti modelov. Interpretačné výsledky sa nepodarilo jednoznačne zlepšiť. V niektorých prípadoch boli interpretácie lepšie a v niektorých nie. Na základe tohto skúmania bola vo všetkých nasledujúcich experimentoch opäť použitá okolitá oblasť z mamogramu.

Siedma štúdia sa venovala použitiu konvolučných neurónových sietí ako extraktorov príznakov pre modely SVM. Motivácia vychádzala z jednej štúdie, ktorá dokázala dosiahnuť výrazne lepších klasifikačných výsledkov v porovnaní s neurónovými sieťami. Výsledky sa nepodarilo zreplikovať. Po hlbšom preskúmaní štúdie vzniklo podozrenie, že použité dáta neboli nezávislé, vďaka čomu sa jej podarilo dosiahnuť tak výrazného zlepšenia.

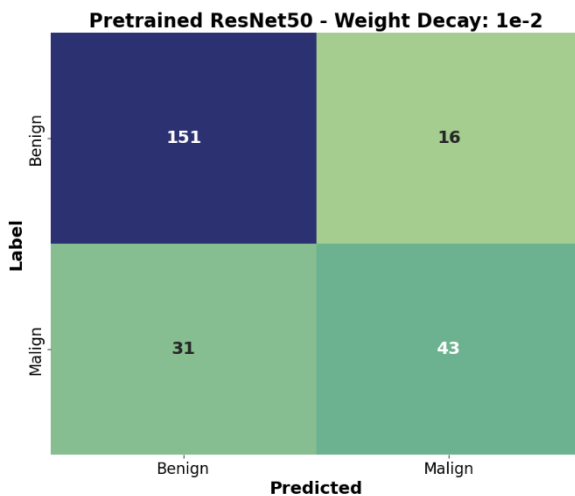
Nasledovali experimenty spojenia databáz CBIS-DDSM a OMI-DB. Prvým experimentom spojenia databáz bolo hľadanie najlepšej architektúry modelu. Boli otestované architektúry VGG, Inception-V3, ResNet, DenseNet a EfficientNet. Architektúra ResNet dosiahla najlepších výsledkov s presnosťou 77,2%, avšak ostatné architektúry si viedli veľmi porovnateľne.

Prístup k dvom vhodným databázam sprístupnilo možnosť experimentovania s prenositeľnosťou znalosti medzi rôznymi databázami. CBIS-DDSM je databáza pozostávajúca z mamogramov získaných staršou SFM technológiou, zatiaľ čo OMI-DB obsahuje mamogramy vytvorené modernou FFDM metódou. Cieľom bolo otestovať, ako dobre si povedie model natrénovaný na dátach CBIS-DDSM s dátami OMI-DB, a naopak. Testovalo sa aj spojenie oboch databáz. Výsledkom štúdie bolo, že znalosti natrénované na jednej alebo druhej databáze nie sú prenositeľné. Spojenie databáz zvýšilo celkovú presnosť modelov a ich interpretačné výsledky. Veľký dôraz bol kladený interpretáciu nesprávnych predikcií, hlavne na také výrezy, ktorých predikcie vychádzali zle naprieč viacerými modelmi naraz.

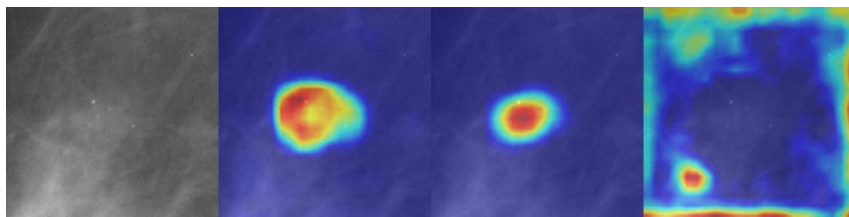
Posledný výskum sa zamerával na pridanie novej triedy, a teda namiesto binárnej klasifikácie rozhodoval klasifikátor o zaradení nálezu medzi tri triedy. Cieľom bolo odstránenie niektorých limitácií, ktoré boli identifikované pomocou Grad-CAM interpretácie z predchádzajúcich modelov. Konkrétne to boli:

- Nesprávna klasifikácia malígnych nálezov ako benígnych v dôsledku prítomnosti benígnych abnormalít (napr. makrokalcifikáty).
- Rozhodnutie o benígnosti nálezu na základe okolia nálezu (nemusí nevyhnutne znamenať nežiaduce správanie modelu).
- Väčší vplyv okolia (šum) ako samotného nálezu na rozhodovací proces.
- Niektoré predpovede boli založené na prítomnosti iných typov nálezov (napr. ložiská).
- Problémy s lokalizáciou mikrokalcifikátov, ktoré je ťažké vizuálne identifikovať aj ľudským okom.

Pridaním tretej triedy sa podarilo zlepšiť interpretačné výsledky a výrazne minimalizovať popísané limitácie. Veľkou pridanou hodnotou bolo, že sa model tento krát rozhodoval pre malígnu aj benígnu triedu na základe rovnakého zhľuku naraz (Obrázok 3). Pri binárnej klasifikácii bol zhľuk vždy lokalizovaný len v jednej triede.



Obrázok 2: Najlepší model na CBIS-DDSM databáze



Obrázok 3: Troj-triedový klasifikátor – (popis výrezov zľava doprava) vstupný výrez; čo malo najväčší vplyv pri benígnej triede; čo malo najväčší vplyv pri malignej triede; čo malo najväčší vplyv pri triede pozadia;

Záver

Úvodné kapitoly práce boli zamerané na opis procesu hodnotenia mamografie. Poznatky získané v týchto kapitolách sú kľúčové pre budovanie modelov a identifikáciu ich obmedzení. Na pochopenie informácií získateľných z mamografických databáz bolo zavedených niekoľko kľúčových pojmov.

V práci sa analyzovali databázy: MIAS, INbreast, CBIS-DDSM a OMI-DB. Boli definované výhody a nevýhody každej databázy spolu s podrobnými štatistickými údajmi o počte rôznych typov nálezov, ktoré obsahovali. Najvhodnejšie databázy na vytvorenie konvolučných sietí boli CBIS-DDSM a OMI-DB, pretože ako jediné spĺňali kritériá kvality, kvantity a prítomnosti histopatologických nálezov. Bol opísaný proces tvorby vlastnej databázy zameranej na mikrokalcifikačné nálezy v spolupráci s Onkologickým ústavom svätej Alžbety.

Následne sa v práci skúmal prienik oblastí mamografie a strojového učenia, pričom sa identifikovalo niekoľko potenciálnych využití modelov a ich obmedzení. Pozornosť bola venovaná aj niekoľkým publikáciám zameraným na mamografiu. Opísané bolo pracovné prostredie a spracovanie databázy do jednotlivých dátových množín. Súbor údajov sa zameriaval predovšetkým na nálezy mikrokalcifikácií nepresahujúce rozlíšenie 674 x 674 pixelov.

Záverečná kapitola bola zameraná na vytvorenie konvolučných modelov na klasifikáciu políčok. Každý experiment bol navrhnutý tak, aby overil konkrétnu hypotézu, pričom sa často sledoval vplyv zmien parametrov na presnosť modelu. Zistilo sa, že použitie predtrénovaných váh z dátovej množiny ImageNet s odomknutými váhami v konvolučných vrstvách prinieslo najlepšie výsledky v porovnaní s novo-inicializovanými váhami alebo predtrénovanými váhami so zamknutými konvolučnými vrstvami.

Experiment zaoberajúci sa nevyváženosťou tried ukázal, že je potrebné použiť metódu vyvažovania tried. Z bežných metód, ako sú oversampling a class-weights, sa lepšie osvedčili rozličné váhy pre triedy. Za predpokladu vhodne nastavenej rýchlosti učenia bola táto metóda porovnateľná s metódou vyvažovania pomocou augmentácie dát, pri ktorej boli umelé údaje generované v takom počte, aby sa triedy vyvážili.

Skúmal sa vplyv zníženia rozlíšenia výrezov z databázy CBIS-DDSM na 224 x 224 pixelov, ktoré je štandardné pre konvolučné siete. Potvrdila sa hypotéza, že zmenšenie políčok s už veľmi malými mikrokalcifikátmi by mohlo zhoršiť výsledky modelu.

Predchádzajúce znalosti sa použili na tvorbu najlepšieho modelu (80,5 % presnosť) databázy CBIS-DDSM, na ktorú sa použila metóda interpretácie Grad-CAM. Interpretácia ukázala, že model sa rozhodoval na základe dôležitých oblastí obsahujúcich nálezy kalcifikátov. Zistili sa aj nedostatky, keď prítomnosť benígnej abnormality a malígnej abnormality v tom istom výreze zavádzala model k nesprávnym predikciám. Prítomnosť benígnych abnormalít bola podmienená spôsobom spracovania databázy. Nevýhodou bola prítomnosť šumu, najmä v triede benígnych abnormalít, kde sa model rozhodoval aj na základe zón, ktoré boli vzdialené a bez prítomnosti abnormalít.

Na riešenie problému, keď sa v jednom výreze mohli vyskytovať dva typy abnormalít, sa vytvorilo nové predspracovanie databázy, ktoré pridávalo čierne pixely namiesto okolitej oblasti z mamogramu pre nálezy menšie ako 674 x 674 pixelov. Predspracovanie malo tiež pomôcť znížiť šum. Preukázalo sa zhoršenie presnosti modelov. Pri porovnávaní máp Grad-CAM sa jednoznačne nepreukázalo, že takýto model obsahuje menej šumu alebo lepšie interpretačné výsledky. V niektorých jednotlivých prípadoch došlo k zlepšeniu a v niektorých k zhoršeniu.

V jednej štúdií sa skúmalo aj použitie modelov SVM, pričom sa na klasifikáciu použili funkcie získané z konvolučných sietí. Štúdia bola motivovaná jednou z opísaných publikácií, v ktorej sa touto metódou dosiahli výrazne lepšie výsledky presnosti. Výsledky sa však nepodarilo replikovať, čo odhaľuje vysoko pravdepodobné podozrenie na závislosť tréningovej a validačnej množiny v publikácii, ktorá mohla spôsobiť také výrazné zlepšenie presnosti.

Následné štúdie sa zamerali na dátovú množinu vytvorenú z kombinácie databáz CBIS-DDSM a OMI-DB. Experimenty zamerané na rôzne konvolučné architektúry nepotvrdili, že by novšie architektúry priniesli lepšie výsledky. Najlepšou architektúrou bola ResNet101, ale ostatné architektúry poskytli veľmi porovnateľné výsledky.

Výhoda prístupu k dvom databázam, ktorých obrazy boli získané pomocou rôznych technológií, sa skúmala v štúdií o medzidatabázovej prenositeľnosti znalostí. Ukázalo sa, že model natrénovaný na jednej technológii nie je prenosný na inú technológiu. Použitie oboch databáz výrazne zlepšilo presnosť modelu v porovnaní s modelmi trénovanými len na jednej databáze. Zistilo sa, že niektoré výrezy boli konzistentne nesprávne klasifikované vo všetkých použitých modeloch. Výsledky boli konzultované s odborníkmi, ktorí potvrdili, že ide skutočne o náročné prípady, ktoré by bolo ťažké alebo až nemožné vizuálne klasifikovať.

V záverečnom experimente sa prešlo od používania binárnych klasifikátorov k trom triedam. Bola pridaná nová trieda pozadia, ktorá mala obsahovať zdravé tkanivo, abnormality s ložiskami a makrokalcifikáty. Pridanie tejto triedy viedlo k výrazne lepším výsledkom interpretácie. Šum sa výrazne znížil a model bol vedený tak, aby sa rozhodoval na základe rovnakej zóny pri malígnej aj benígnej triede. Takéto výsledky sú pre ľudí intuitívnejšie, vďaka čomu sa zlepšila dôveryhodnosť modelov. Pridanie triedy však viedlo k zníženiu správnych malígnych predpovedí v porovnaní s binárnymi klasifikátormi (presnosť: 91,7 %, MCC: 0,767). Bolo objavené nové obmedzenie modelov, v ktorých bol špeciálny typ mikrocalcifikátov nazývaný vaskulárne calcifikáty klasifikovaný do malígnej triedy. Dôvodom je, že tento typ je pre lekárov ťažko klasifikovateľný a z tohto dôvodu sa v databázach nenachádza. V prípade modelu však tento nález veľmi pripomína podozrivé zhluky mikrocalcifikátov.

Nadchádzajúca práca by sa mala zaoberať zlepšením troj-triedového klasifikátora a lepším vyčistením triedy so zdravým tkanivom. Po finalizácii databázy vznikajúcej v spolupráci s onkologickým ústavom sv. Alžbety bude možné dátové množiny doplniť o nové dáta, resp. sa ponúka možnosť testovania na úplne novej nezávislej vzorke.

Zoznam Použitej Literatúry

- [1] Survival Rates for Breast Cancer. <https://www.cancer.org/cancer/breast-cancer/understanding-a-breast-cancer-diagnosis/breast-cancer-survival-rates.html>. Accessed: 2022-05-31. Mar. 2022.
- [2] Jagpreet Chhatwal, Oguzhan Alagoz, and Elizabeth S Burnside. “Optimal breast biopsy decision-making based on mammographic features and demographic factors”. In: *Operations research* 58.6 (2010), pp. 1577–1591.
- [3] C.J. D’Orsi et al. *ACR BI-RADS atlas : breast imaging reporting and data system*. eng. 5th ed. Reston, VA: American College of Radiology, 2013. isbn: 9781559030168.
- [4] J. Suckling. “The Mammographic Image Analysis Society Digital Mammogram Database Exerpta Medica”. In: *Exerpta Medica International Congress* 1069 (1994), pp. 375–738.
- [5] Rebecca Lee et al. “A curated mammography data set for use in computer-aided detection and diagnosis research”. In: *Scientific Data* 4 (Dec. 2017), pp. 170–177. doi: 10.1038/sdata.2017.177.
- [6] Ines Moreira et al. “INbreast: Toward a Full-field Digital Mammographic Database”. In: *Academic radiology* 19 (Nov. 2011), pp. 236–48. doi: 10.1016/j.acra.2011.09.014.
- [7] Mark D. Halling-Brown et al. “OPTIMAM Mammography Image Database: A Large-Scale Resource of Mammography Images and Clinical Data”. In: *Radiology: Artificial Intelligence* 3.1 (2021). PMID: 33937853, e200103. doi: 10.1148/ryai.2020200103. eprint: <https://doi.org/10.1148/ryai.2020200103>. url: <https://doi.org/10.1148/ryai.2020200103>.
- [8] Ramprasaath R. Selvaraju et al. “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization”. In: *International Journal of Computer Vision* 128.2 (Oct. 2019), pp. 336–359. issn: 1573-1405. doi: 10.1007/s11263-019-01228-7. url: <http://dx.doi.org/10.1007/s11263-019-01228-7>.

Zoznam Publikovaných Prác

Články:

- Mračko, A.; Vanovčanová, L.; Cimrák, I. Mammography Datasets for Neural Networks—Survey. *J. Imaging* 2023, 9, 95. <https://doi.org/10.3390/jimaging9050095>
- Mračko, A.; Cimrák, I.; Vanovčanová, L.; Lehotská, V. (2024). Deep Learning in Breast Calcifications Classification: Analysis of Cross-Database Knowledge Transferability. In *Proceedings of the 17th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 1*, ISBN 978-989-758-688-0, ISSN 2184-4305, pages 527-535. DOI: 10.5220/0012535200003657
- Mračko, A.; Cimrák, I.; Vanovčanová, L. (2024). Enhancing Breast Microcalcification Classification: From Binary to Three-Class Classifier. In *Proceedings of the 35th conference of FRUCT association - Volume 1*, ISBN 978-952-65246-1-0, ISSN 2305-7254, pages 473-481.

Konferenčné prezentácie:

- Mračko, A.; Cimrák, I. Databases of Mammography Images for Deep Convolutional Neural Networks, In: *The Second International Workshop on Reliability Engineering and Computational Intelligence (RECI2022)*, 13 - 15 November 2022, Delft, The Netherlands
- Mračko, A.; Cimrák, I. Classification of Mammogram Patches Using Convolutional Neural Networks, In: *16th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2023)*, 16 - 18 February 2023, Lisbon, Portugal