

Universitatea POLITEHNICA București

Facultatea de Automatică și Calculatoare  
Departamentul de Calculatoare



# TEZĂ DE DOCTORAT

## Rezumat

Aplicații moderne ale analizei de imagini  
document și medicale

**Conducător științific:**  
Prof. Dr. Ing. Costin-Anton Boianțiu

**Autor:**  
Ing. Marcel Prodan

București, 2023

## Cuprins

Cuprins .....	ii
Lista abrevierilor .....	1
1 Introducere.....	3
1.1 Structura tezei.....	7
2 Lucrări Corelate .....	9
2.1 Prezentare generală a sistemelor de analiză a imaginilor documentelor .....	9
2.2 Prezentare generală a detectării cancerului la sân .....	9
2.3 Analiza și prelucrarea automată a imaginilor documentelor .....	10
2.4 Metrici pentru Asemănarea Imaginilor.....	11
3 Analiza imaginilor prin combinarea inteligentă a rezultatelor mai multor algoritmi.....	12
3.1 Procesul de Binarizare a Imaginilor Document.....	12
3.2 Segmentarea .....	15
3.3 OCR.....	16
3.4 Arhitectură Unificată de Procesare a Imaginilor: Analiză Documente, Imagistică Medicală și Recunoaștere a Indicatoarelor Rutiere.....	17
4 Studiu de Caz: Detectarea Cancerului de Sân .....	19
4.1 Rezultate.....	20
5 Concluzii.....	23
5.1 Contribuții .....	25
5.2 Lista publicațiilor .....	26
5.2.1 Journale .....	26
5.2.2 Conferințe.....	27
5.2.3 Postere .....	27
Bibliografie.....	28

## Lista abrevierilor

### Cases

ADIAP: Automatic Document Image Analysis and Processing .....	131
ADMANI: Annotated Digital Mammograms and Non-Image data.....	111
AI: Artificial Intelligence.....	22
ANN: Artificial Neural Network.....	30
AUC: Area Under the Curve.....	106
BAIR: Berkeley AI Research.....	79
BC: Breast Cancer.....	21
BI-RADS: Breast Imaging Reporting And Data System .....	107
CAM: Class Activation Map.....	118
CC: Cranio Caudal.....	106
CGAN: Conditional Generative Adversarial Network .....	110
CMD: Chinese Mammography Database .....	106
CNN: Convolutional Neural Networks.....	24
CT: Computerized Tomography.....	25
DDSM: Digital Database for Screening Mammography .....	106
DIA: Document Image Analysis.....	2, 11
DIBCO: Document Image Binarization Contest: Database of public images .....	60, 72
DIQM: Deep Image Quality Metric.....	42
DL: Deep Learning .....	20
DNN: Deep Neural Network .....	107
DRD: Distance Reciprocal Distortion.....	65
FCN: Fully Convolutional Networks.....	63
FID: Frechet Inception Distance.....	113
FN: False Negative .....	69
FP: False Positive.....	69
FSIM: Feature Similarity Index .....	32
FSIMc: Feature Similarity Index for Color Images .....	42
GAN: Generative Adversarial Networks .....	22
GPU: Graphics Processing Unit.....	74
Grad-CAM: Gradient-weighted Class Activation Mapping .....	118
GT: Ground Truth (image).....	65
ICR: Intelligent Character Recognition .....	13
IMRC: Intelligent Machines Research Corporation .....	12
INbreast: Full-field Digital Mammographic Database .....	107
IoU: Intersection Over Union .....	64
IWR: Intelligent Word Recognition.....	13
KNN: K-Nearest Neighbors Algorithm .....	109
KSONN: Kohonen Self-Organizing Neural Network.....	30
LUA: Computer programming language, meaning 'moon' in Portuguese .....	17
MAE: Mean Absolute Error.....	31
MBG: Meise Botanic Garden (Begium).....	29
METAe: Research endeavor financed by the European Commission .....	17
METS: Document file format .....	18
MIAS: Mammographic Image Analysis Society .....	107
MICR: Magnetic Ink Character Recognition.....	14
mIoU: mean Intersection Over Onion.....	64
MIR: Magnetic Ink Recognition.....	12

---

MLO: Mediolateral Oblique .....	106
MLP: MultiLayer Perceptrons .....	62
MO: Mammographically-Occult.....	110
MRI: Magnetic Resonance Imaging .....	25
MSE: Mean Squared Error.....	31
MS-SSIM: MultiScale Structural Similarity Index.....	32
NLP: Natural Language Processing.....	117
NVIDIA: The name of brand of video cards .....	77
OCR: Optical Character Recognition .....	11
OCROPUS: Open-source ptical recognition and document analysis program .....	17
OMR: Optical Mark Recognition .....	13
PASCAL: Dataset, in the context of this thesis.....	64
P-FM: Pseudo F-measure.....	65
PSNR: Peak Signal to Noise Ratio .....	65
RBGE: Royal Botanic Garden Edinburgh .....	28
RCA: Radio Communication of America .....	12
ROI: Regions Of Interest .....	106
RSNA: Radiological Society of North America .....	111
SGD: Stochastic Gradient Descent .....	65
SLIC: Simple Linear Iterative Clustering.....	80
SLN: Steerable Local Neighborhood.....	62
SOM: Self-Organizing Maps .....	62
SSIM: Structural Similarity Index .....	32
StyleGAN: Style Generative Adversarial Network - AI model .....	112
SVM: Support Vector Machines .....	62
TP: True Positive.....	69
TTCNN: Transferable Texture CNN.....	107
UK: United Kingdom.....	16
UMAP: Uniform Manifold Approximation and Projection .....	108
UQI: Universal Quality Index.....	32
USA: United States of America .....	12
USPS: US Postal System .....	13
VGG: Visual Geometry Group - AI classifier.....	63, 106
VIF: Visual Information Fidelity .....	32
ViT: Visual Transformer.....	19
XAI: Explainable AI - medical image analysis AI model.....	117

## 1 Introducere

Omul și natura au avut întotdeauna o relație profundă și complexă, o simbioză care a avut o influență majoră asupra dezvoltării noastre ca specie. Înțelegerea acestei relații ne poate oferi o perspectivă mai adâncă asupra modului în care interacționăm cu lumea din jurul nostru și a influenței pe care o avem asupra ei.

Relația complexă dintre oameni și natură este esențială pentru evoluția noastră ca specie, cuprinzând toate simțurile noastre. Totuși, vederea se distinge în mod deosebit, cu aproximativ 33% din creier dedicat acestei funcții. Procesarea realității de către creier prin organele senzoriale traduce în esență semnalele externe într-un limbaj electrochimic. Acest lucru indică faptul că creierul nostru creează percepția noastră asupra realității bazată pe modele interne formate din experiențe anterioare, fără a interacționa vreodată direct cu lumea exterioară. Diverse experimente au evidențiat capacitățile adaptive ale creierului. De exemplu, se adaptează la ochelari prismatici care răstoarnă perspectiva vizuală, subliniind flexibilitatea creierului în procesarea mai multor imagini ale aceleiași scene. În plus, sincronizarea semnalelor de la diferite simțuri joacă un rol pivotal, așa cum se arată prin diferențele de timp de reacție la atleți atunci când sunt semnalați de sunet față de lumină. Absența stimulilor externi, ca în cazurile de izolare extremă la Alcatraz, duce la generarea propriei realități de către creier. Acest lucru relevă rolul profund al modelului intern al creierului, care se îmbogățește cu fiecare experiență nouă, reducând nevoia de input extern continuu și permițând o procesare eficientă. [1]

Comparând funcționarea creierului uman cu modul în care imaginile sunt analizate folosind tehnici computaționale, putem extrage următoarele concluzii:

- utilizarea, dacă este posibil, a mai multor seturi de imagini pentru o evaluare mai bună a proprietăților lor și obținerea unor rezultate mai exacte.
- importanța sincronizării imaginilor capturate cu alte cantități asociate, cum ar fi timpul, distanțele, unghiurile, altitudinea, viteza sau alte caracteristici ale mediului.
- importanța utilizării filtrelor pentru a face procesul de analiză a imaginilor mai eficient.
- importanța utilizării bazei de cunoștințe oriunde există și poate fi folosită.
- importanța instruirii modelelor de inteligență artificială.

*Contextul sistemelor de analiză a imaginilor documentelor*

Cu apariția erei digitale, tranziția de la mijloacele tradiționale de comunicare, cum ar fi casetele și CD-urile, la formate digitale a fost esențială. Digitalizarea documentelor fizice abordează probleme legate de stocare, accesibilitate și sustenabilitatea mediului, păstrând în același timp înregistrările istorice pentru utilizare viitoare. Sistemele de analiză a imaginilor documentelor (DIA), în special Recunoașterea Optică a Caracterelor (OCR), sunt centrale în acest proces de digitalizare, permițând conversia diverselor documente, de la cărți la hârti, în formate digitale.

### ***Istoria Recunoașterii Optice a Caracterelor***

Dorința de a crea echipamente capabile să recunoască scrierea, fie că este de mână, scrisă la mașină sau tipărită, datează din secolul al 19-lea. Mai multe încercări timpurii de OCR au fost făcute prin intermediul cererilor de brevetare în anii 1920 și 1930. Cu toate acestea, limitările tehnologice ale epocii au împiedicat realizarea sistemelor OCR practice.

Progrese semnificative în sistemele de recunoaștere a caracterelor au fost realizate în anii 1950, motivate de necesitatea sectorului financiar de a procesa automat cecurile, cele mai comune documente tipărite la acea vreme. Prima mașină OCR comercială a fost achiziționată de Reader's Digest în 1955, ducând la dezvoltarea ulterioară și adoptarea sistemelor OCR în diverse aplicații, inclusiv procesarea documentelor, distribuția corespondenței și recunoașterea facturilor.

### ***Tehnologii alternative***

În plus față de OCR, au fost dezvoltate mai multe tehnologii alternative pentru a răspunde nevoilor specifice de procesare a documentelor. Acestea includ Sistemele de Recunoaștere Inteligentă a Caracterelor (ICR) și Sistemele de Recunoaștere Inteligentă a Cuvintelor (IWR) pentru scrisul cursiv, Sistemele de Recunoaștere Optică a Semnelor (OMR) pentru gestionarea structurilor de formulare, și Recunoașterea Caracterelor cu Cerneală Magnetică (MICR) utilizată în bănci și instituții financiare.

### ***Recunoașterea optică a Caracterelor***

Tehnologia OCR joacă un rol crucial în digitalizarea cărților și manuscriselor, păstrându-le semnificația culturală și istorică. Performanța OCR este evaluată în funcție de acuratețe și încredere în rezultate. Deși OCR a făcut progrese semnificative, intervenția umană este esențială pentru gestionarea cazurilor complexe și atingerea unor niveluri mai ridicate de acuratețe.

Procesul OCR implică mai multe etape, inclusiv preprocesare, îmbunătățirea imaginii, segmentare, recunoaștere și clasificare. Fiecare pas contribuie la acuratețea și calitatea generală a conținutului digitalizat.

### ***Metodologii aplicabile sistemelor de analiză a imaginilor documentelor***

Sistemele de analiză a imaginilor documentelor sunt de obicei concepute pentru a aborda sarcini specifice, având în vedere complexitatea analizei imaginilor la diferite niveluri. Aplicații adaptate sunt utilizate pentru a asigura o procesare eficientă a documentelor, luând în considerare factori precum structura documentului, suportul lingvistic și calitatea hârtiei. Feedback-ul utilizatorilor și intervenția umană sunt comune în astfel de sisteme, în special pentru manipularea documentelor deteriorate sau unice.

### ***Tehnici actuale de retro-conversie***

Importanța crescândă a digitalizării colecțiilor de cărți, ziare și înregistrări istorice a condus la apariția diverselor software-uri și inițiative de retro-conversie. OCRopus, METAe și Lib2Life sunt exemple ale unor astfel de sisteme menite să păstreze și să digitalizeze documente fizice valoroase. Viitorul sistemelor de analiză a imaginilor documentelor este concentrat pe atingerea unei acuratețe, flexibilități și eficiențe mai mari, reducând în același timp interacțiunea utilizatorului.

### ***Procesul de binarizare a imaginilor documentelor***

Binarizarea imaginilor documentelor este esențială pentru digitalizarea și conservarea documentelor istorice. În era digitală, există o nevoie crescândă de a conserva textele istorice. Această teză introduce o tehnică de binarizare folosind Rețeaua Complet Convoluțională (FCN) pentru digitalizare eficientă și precisă. Binarizarea convertește documentele într-un format binar, distingând pixelii importanți de cei irelevanți, oferind o lizibilitate și eficiență mai bune decât metodele mai vechi, precum scanarea. Scopul principal este de a conserva scrierile istorice, indiferent de starea lor, folosind un FCN pentru rezultate optime. Analizăm metodele existente de binarizare și introducem o arhitectură inovatoare bazată pe FCN. Sistemul folosește seturile de date DIBCO pentru instruire și evaluare, păstrând simplitatea și oferind o performanță solidă. Modelul nostru combină tehnici avansate folosind biblioteca Keras, având ca obiectiv îmbunătățirile în conservarea documentelor, recunoașterea textelor și imagistica medicală. Optimizând acest model, previzionăm avansări în analiza imaginilor și aplicații potențiale în domenii diverse.

## ***Segmentarea***

Introducerea tezei se apleacă asupra segmentării imaginilor, un component esențial al viziunii computerizate, cu aplicații care se întind de la conducerea autonomă la extragerea datelor. Scopul principal al cercetării este de a îmbunătăți rezultatele segmentării prin amalgamarea diferitelor algoritmi prin tehnici de votare. Deși sunt indispensabile, algoritmi de segmentare pot întâmpina probleme precum supra-segmentarea sau sub-segmentarea. Pentru a aborda acestea, studiul propune un model cu operații de convoluție, deconvoluție și concatenare. Se prevede că acest model va iniția avansări în sarcini precum binarizarea documentelor, segmentarea culorilor, OCR și detectarea cancerului. Teza evaluează trei metode cheie de segmentare: algoritmul lui Felzenszwalb, SLIC Superpixels și Watershed, discutându-le avantajele și dezavantajele. O tehnică de votare nouă combină rezultatele acestor algoritmi pentru a optimiza acuratețea segmentării, explorând atât votul democratic cât și cel bazat pe performanță. Eficiența acestei abordări este evaluată în raport cu rezultatele algoritmilor individuali, iar exemplele vizuale subliniază contrasturile și beneficiile. Acest studiu deschide calea pentru metode inovative de segmentare potrivite pentru o gamă de aplicații.

## ***OCR***

Recunoașterea Optică a Caracterelor (OCR) a devenit un instrument esențial în Computer Vision datorită capacității sale de a transforma textul din imagini în formate lizibile. Cu toate acestea, probleme precum zgomotul din imagine, iluminarea inconstantă și uzura documentului pot împiedica acuratețea OCR. Pentru a atenua aceste provocări, acest studiu introduce o metodă OCR inovatoare bazată pe votare. Această abordare combină rezultatele din multiple tehnici OCR folosind un proces de votare, având ca scop identificarea celor mai fiabile rezultate. Un pas inițial de preprocesare aplică diverse filtre asupra imaginii de intrare, îmbunătățind caracteristicile textuale. Imaginea este apoi supusă mai multor rulări OCR, cu alegerea rezultatului textual cel mai fiabil bazat pe evaluările de încredere la nivel de cuvânt. Testele au arătat că această metodă îmbunătățește semnificativ precizia OCR, deși necesită mai mult timp de procesare. Versatilitatea strategiei sugerează aplicații mai largi în procesarea nesupervizată a imaginilor și extragerea datelor. Ca parte a proiectului Lib2Life, se așteaptă ca metoda să fie esențială în îmbunătățirea conversiei digitale a documentelor din bibliotecile românești. În esență, această abordare OCR bazată pe votare oferă avansări promițătoare în extragerea datelor precise din medii de imagine provocatoare, propulsând și mai mult era digitală.



## ***Detecția cancerului de sân***

Cancerul de sân rămâne o preocupare dominantă pentru sănătate, necesitând detectarea timpurie pentru un tratament eficient. Deși mamografia a fost o piatră de temelie în screeningul pentru cancerul de sân, analiza sa manuală pune provocări în ceea ce privește timpul și subiectivitatea. Această cercetare se apleacă asupra soluțiilor de învățare profundă, în special folosind Rețele Neuronale Convoluționale (CNN) și Vision Transformers (ViT) pentru a analiza mamogramele. O metodă distinctă folosind generarea de imagini sintetice prin învățare automată îmbunătățește setul de date de instruire, ducând la o mai bună acuratețe a clasificării. Rolul preprocesării datelor, augmentării și tehnicilor de AI explicabil, precum hărțile de activare a clasei, sunt subliniate pentru transparența și înțelegerea modelului. Cu toate acestea, obstacole precum limitările setului de date și interpretarea modelului rămân. Acest studiu subliniază potențialul integrării AI în screeningul cancerului de sân, având ca obiectiv detectarea timpurie, îmbunătățirea succesului tratamentului și ameliorarea bunăstării pacientului.

### **1.1 Structura tezei**

Capitolul "Structura Tezei" servește ca un ghid pentru organizarea și conținutul acestei teze de doctorat, conturând principalele capitole și respectivele lor contribuții. Secțiunile următoare oferă o privire concisă asupra fiecărui capitol:

***Capitolul 1*** pune bazele cercetării prin introducerea contextului și obiectivelor tezei. Evidențiază semnificația procesării imaginilor și relevanța acestora în diverse scenarii tehnologice. De asemenea, schițează întrebările de cercetare și metodologiile utilizate de-a lungul studiului.

***Capitolul 2*** oferă un background cuprinzător și o prezentare generală a temei tezei. Acoperă cercetarea existentă în sistemele de analiză a imaginilor documentelor, metrici pentru asemănarea imaginilor și detectarea cancerului de sân, precum și procesarea automată a imaginilor documentelor. Capitolul stabilește fundația cercetării explorând cunoștințele de vârf în aceste domenii relevante.

***Capitolul 3*** prezintă conceptul de analiză inteligentă a imaginilor și fuziunea acestora. Include discuția despre analiza imaginii combinând rezultatele mai multor algoritmi. Capitolul explorează procesul de binarizare a imaginilor documentelor, tehnicile de segmentare și Recunoașterea Optică a Caracterelor. În plus, introduce o arhitectură de procesare a imaginilor

unificată care integrează analiza documentelor, imagistica medicală și recunoașterea semnelor de circulație. Acest capitol își propune să dezvolte o abordare cuprinzătoare și eficientă a analizei imaginilor, valorificând diverse tehnici și fuzionarea acestora.

**Capitolul 4** se concentrează pe un studiu de caz legat de detectarea cancerului de sân. Include o recenzie a tehnicilor de imagistică medicală utilizate frecvent pentru detectarea cancerului de sân. Capitolul prezintă apoi materialele și metodele utilizate în studiul de caz și rezultatele obținute din studiu. De asemenea, discută implicațiile și semnificația descoperirilor în secțiunea de discuție și rezumă punctele esențiale din studiul de caz, aruncând lumină asupra aplicațiilor și avansărilor potențiale în screeningul cancerului de sân.

**Capitolul 5** este capitolul de încheiere care oferă un rezumat complet al întregii teze. Reiterează principalele descoperiri și contribuții ale cercetării, precum și semnificația rezultatelor obținute. Acest capitol discută și implicațiile cercetării și impactul său potențial asupra domeniilor respective ale analizei imaginilor documentelor și detectării cancerului de sân. Se încheie cu sugestii pentru direcții viitoare de cercetare și aplicații potențiale ale descoperirilor. În plus, capitolul include o listă de publicații rezultate din cercetarea desfășurată în timpul programului de doctorat, recunoscând contribuțiile academice făcute în timpul călătoriei de cercetare.

## 2 Lucrări Corelate

### 2.1 Prezentare generală a sistemelor de analiză a imaginilor documentelor

Păstrarea informațiilor în siguranță este principalul motiv pentru care dorim să le stocăm și să le facem disponibile în format digital, având în vedere că trăim în era digitală. Un mod de a pune datele fizice la dispoziția lumii digitale este prin scanarea și stocarea documentelor fizice într-un format digital. Acest proces rezolvă multe probleme, precum spațiul de stocare, accesul și problemele de mediu. În plus, documentele digitale nu se deteriorează, ceea ce permite conservarea multor documente istorice semnificative. Traducerea documentelor fizice în format digital pare a fi mai dificilă decât pare, având în vedere numeroasele probleme care pot apărea, precum hârtia deteriorată, schimbările de culoare, problemele de font și așa mai departe. Scopul este de a extrage datele relevante din documente. Scopul sistemelor de „Analiză a Imaginilor Documentelor” (DIA) este de a procesa documentele într-un mod asemănător cu cel uman, recunoscând textul, imaginile și extrăgând informații. În ceea ce privește sistemele de analiză a imaginilor documentelor [3], o tehnologie numită recunoaștere optică a caracterelor (OCR) [4] este cea mai importantă. Acest lucru se datorează faptului că a fost utilizată mult timp în sistemul poștal din Statele Unite și pentru că utilizatorii folosesc frecvent OCR chiar și atunci când lucrează cu fișiere care nu conțin text. Cu toate acestea, această metodă reprezintă doar o mică parte din aplicațiile mai largi și creează o imagine stereotipă a ceea ce înseamnă cu adevărat un DIA, fiind utilizată pentru a recunoaște caracterele și pentru a converti imaginile tipărite în text codificat automat. Capacitatea de a recunoaște și digitaliza o gamă largă de documente (cărți, scrisori, desene și hărți) este unul dintre avantajele utilizării analizei imaginilor documentelor. Analiza imaginilor documentelor implică o varietate de etape și procesări la diferite niveluri de granularitate și este destul de complexă. În capitolele următoare, vom prezenta evoluția OCR și a sistemelor de analiză a imaginilor documentelor, împreună cu diferitele abordări și tehnologii utilizate în acest scop.

### 2.2 Prezentare generală a detectării cancerului la sân

Cancerul la sân rămâne o preocupare majoră de sănătate din cauza prevalenței sale și a nevoii de detectare timpurie pentru a asigura un tratament reușit [5]. Mamografia, o metodă recunoscută pe scară largă pentru depistarea cancerului la sân, are unele provocări inerente, în special în analiză, care poate fi atât consumatoare de timp, cât și subiectivă. Acest studiu investighează utilizarea tehnicilor de învățare profundă pentru a augmenta analiza mamografiei, folosind modele precum CNN și ViT pe un set de date public. O metodă nouă de

augmentare a datelor folosind generarea de imagini sintetice sporește în continuare eficiența modelului. Se subliniază valoarea preprocesării datelor și a augmentării în obținerea unei clasificări de vârf, în timp ce metodele de IA Explicabilă, inclusiv hărțile de activare a clasei și casetele delimitatoare centrate, aruncă lumină asupra procesului decizional al modelelor [6].

Statisticile globale subliniază gravitatea cancerului la sân, marcându-l ca tipul de cancer cel mai frecvent diagnosticat. A reprezentat peste 2,3 milioane de diagnostice și a condus la mai mult de 685.000 de decese numai în 2020 [5]. Populația feminină rămâne în mod special expusă riscului [7]. În fața unor astfel de provocări, învățarea automată (ML) și subsetul său avansat, învățarea profundă (DL), oferă speranță. Fundația DL pe rețele neuronale artificiale o desemnează ca un far în sarcini de recunoaștere și clasificare a imaginilor, reprezentând avangarda ML.

Importanța imagisticii medicale în detectarea timpurie a cancerului la sân nu poate fi supra-accentuată. Prin inițiative de screening și instrumente avansate de imagistică, detectarea timpurie este posibilă, ducând la planuri de tratament optime și șanse crescute de supraviețuire. Noutatea lucrării constă în contribuția sa dublă: oferind o recenzie discernătoare a metodelor de imagistică medicală pentru cancerul la sân și introducând un studiu empiric care folosește tehnici de învățare profundă pe mamograme pentru detectarea cancerului. Centrală acestei abordări este utilizarea imaginilor sintetice, create prin algoritmi de învățare automată, inspirându-se dintr-un set vast de mamograme autentice. Aceasta contribuie la acuratețea superioară a clasificării cancerului la sân.

Secțiunile următoare vor oferi o prezentare generală a metodologiilor medicale prevalente și a rolului în creștere al Inteligenței Artificiale în acest domeniu.

### **2.3 Analiza și prelucrarea automată a imaginilor documentelor**

Prelucrarea automată a imaginilor documentelor este esențială în peisajul tehnologic de astăzi, transformând volume vaste de documente digitale și scanate în date digitale acționabile. Versatilitatea sa se extinde în multiple sectoare, îmbunătățind eficiența operațională. Iată principalele sale aplicații: procesarea documentelor, care simplifică scanarea, recunoașterea și extragerea datelor din diverse materiale tipărite; imagistica medicală pentru a ajuta la analiza și diagnosticul imaginilor medicale; supravegherea video care îmbunătățește securitatea prin recunoașterea activității în timp real și alerte; vehicule autonome pentru detectarea obstacolelor și navigația sigură; analiza calității în industrie pentru a asigura calitatea produsului și detectarea defectelor; funcții militare pentru a asista în recunoaștere, supraveghere și evaluări

tactice; aeronautică pentru controlul zborului automatizat și navigație; operațiuni navale pentru a îmbunătăți supravegherea navelor și navigația; și aeronautică pentru a asista în controlul misiunilor spațiale și analiza datelor din explorările spațiale.

Evoluția și progresele în învățarea profundă și rețelele neuronale, în special CNN, au amplificat utilitatea și potențialul prelucrării automată a imaginilor documentelor [8].

## 2.4 Metrici pentru Asemănarea Imaginilor

Acest studiu explorează metricile utilizate pentru a evalua asemănarea imaginilor în știința calculatoarelor, în special în domenii precum procesarea digitală a imaginilor, viziunea computerizată, învățarea automată și imaginile de învățare profundă [10]–[14]. Printre metricile principale cercetate se numără măsurile tradiționale bazate pe pixeli, precum Eroarea Medie Absolută (MAE) și Eroarea Medie Pătrată (MSE), metrici de similaritate structurală precum Indexul de Similaritate Structurală (SSIM) și Indexul de Similaritate Structurală Multiscală (MS-SSIM), precum și măsuri avansate cum ar fi Indexul de Similaritate a Caracteristicilor (FSIM), Indexul Universal de Calitate (UQI) și Fidelitatea Informației Vizuale (VIF) pentru imagini [10], [15]–[19]. Aceste metrici sunt esențiale pentru compararea imaginilor, ghidând dezvoltarea algoritmilor și înțelegerea percepției vizuale umane. De asemenea, ele sunt esențiale în evaluarea eficacității algoritmilor în restaurarea, compresia și îmbunătățirea imaginilor, precum și în sarcini din învățarea profundă precum recunoașterea obiectelor și generarea de imagini cu Rețele Adversative Generative (GANs) [13], [15], [18], [20]. Evoluția acestor metrici de-a lungul timpului reflectă o căutare continuă de a alinia metodele computaționale cu percepția vizuală umană, asigurând atât eficiența computațională, cât și relevanța perceptuală [10], [11], [15], [17]–[21].

### 3 Analiza imaginilor prin combinarea inteligență a rezultatelor mai multor algoritmi

#### 3.1 Procesul de Binarizare a Imaginilor Document

În fața provocărilor legate de conservarea documentelor istorice datorită degenerării în timp, acest studiu recomandă o nouă abordare tehnologică, folosind inteligența artificială pentru digitalizare. În locul metodelor tradiționale precum scanarea, lucrarea sugerează binarizarea ca tehnică superioară. Binarizarea, care diferențiază pixelii utili de cei care cauzează imperfecțiuni vizuale, capturează eficient informația esențială a unui document în format digital, oferind o lizibilitate mai bună și utilizare eficientă a memoriei. Acest studiu își propune să ofere o prezentare cuprinzătoare a metodelor existente de binarizare și introduce o soluție inovatoare bazată pe rețele complet convoluționale (FCN). Rețeaua propusă, mai simplă în design, dar comparabilă în performanță cu metodele de segmentare semantică existente, este antrenată folosind seturile de date DIBCO deseori referențiate [12], [13]. Scopul este să se asigure că documentele istorice, bogate în semnificație culturală și tradițională, sunt păstrate cu o fidelitate ridicată pentru generațiile viitoare.

Studiul introduce o metodă de binarizare a documentelor folosind o rețea Fully Convolutional Network (FCN) inspirată de Long et al. [23] și Tensmeyer & Martinez [24]. Arhitectura rețelei este divizată în patru ramuri, fiecare diferind în adâncime și dimensiuni spațiale. Aceste ramuri extrag atât caracteristici locale, precum forme de bază, cât și caracteristici globale, semantice bogate, precum litere alungite. Tranzițiile între ramuri folosesc Max Pooling, iar operațiile de convoluție implică filtre care se dublează în număr pe măsură ce se avansează în jos pe ramuri. Ultimele trei ramuri trec printr-un proces suplimentar pentru a se potrivi cu hărțile de caracteristici și rezoluția ramurii 1 folosind deconvoluția, o abordare influențată de Long et al. [23]. După asigurarea uniformității între ramuri în ceea ce privește dimensiunile, ieșirile sunt concatenate, urmate de un strat de convoluție, iar în cele din urmă, predicțiile sunt făcute folosind funcția de activare Sigmoid, ideală pentru rezultate binare. Funcția de cost aleasă este Cross-Entropy Binomial, potrivită pentru clasificările binare.

Modelul a fost antrenat pe imagini de 128x128 pixeli din seturile de date DIBCO din perioada 2009-2019 [25], care cuprind documente scrise manual și tipărite mecanic, cu diferite niveluri de degradare. Aceste imagini au fost convertite în alb și negru, apoi normalizate. Antrenamentul a durat 300 de epoci, cu indicele mediu de intersecție peste uniune (mIoU) fiind metrica principală pentru evaluare datorită capacității sale de a ține cont de dezechilibrele de clasă.

Pentru binarizarea documentelor, procesarea inițială a redimensionat documentele pentru a permite împărțirea în segmente de 128x128 pixeli, asigurând cea mai bună rezoluție în funcție de antrenamentul modelului. Ulterior, predicțiile au atribuit valori pixelilor în funcție de un prag setat, rezultând documentul binarizat final cu dimensiunile asemănătoare celor ale originalului.

Modelul a fost antrenat pe parcursul a 300 de epoci, cu o rată de învățare de 0,1, folosind precizia, pierderea și mIoU ca metrici de performanță. În mod notabil, s-a observat o îmbunătățire semnificativă în precizie și pierdere, indicând competența modelului în precizarea clasei predominante a pixelilor de fundal din setul de date. Cu toate acestea, din cauza prezenței coplesitoare a pixelilor de fundal, predicțiile incorecte ale pixelilor din prim-plan au avut un impact minim asupra acestor metrici. Pentru a obține o înțelegere holistică a performanței modelului, este esențial să se ia în considerare metrici care evaluează atât precizia totală a setului de date, cât și precizia specifică a claselor. mIoU, care este sensibil la predicțiile pixelilor din prim-plan, a prezentat o rată mai lentă de creștere. A început să convergă către o valoare de 0,6 după aproximativ 250 de epoci, atât pentru seturile de antrenament, cât și pentru cele de validare.

#### ***Eficacitatea modelului pe setul de date de testare***

Utilizând setul de date de testare, modelul a reușit să obțină următoarele rezultate:

<b>mIoU</b>	<b>accuracy</b>	<b>loss</b>	<b>precision</b>	<b>recall</b>	<b>F-measure</b>
0.589	0.9823	0.0931	0.916	0.883	0.899

În tabelul de mai jos, am accentuat scorurile individuale pentru fiecare arhitectură. În ceea ce privește metricile folosite în cele două lucrări, FM și mIoU, am putut compara răspunsul sugerat cu soluțiile alternative. Așa cum se poate observa, scorurile din analiza proprie a modelului sunt comparabile cu cele din lucrarea lui Long et al. [23] și Tensmeyer & Martinez [24].

<b>Model</b>	<b>Dataset</b>	<b>FM</b>	<b>mIoU</b>	<b>Loss function</b>
<b>Tensmeyer</b>	HDIBCO 2016	90.2	-	Binary Cross-Entropy
<b>Long</b>	VOC2012	-	62.2	Cross-Entropy
<b>Proposed solution</b>	DIBCO	89.9	58.9	Binary Cross-Entropy

*Performanțele comparate între arhitecturi.*

Datorită variațiilor în seturile de date de testare și în metoda de segmentare utilizată, poziționarea soluției propuse față de cele două nu poate fi perfectă. Pentru a evalua corect cele trei modele, o comparație adecvată ar folosi seturi de date de testare standard și metrici.

<b>Model</b>	<b>GPU</b>	<b>Framework</b>
<b>Tensmeyer</b>	NVIDIA Tesla K40c	Caffe
<b>Long</b>	NVIDIA Tesla K40c	Caffe
<b>Proposed solution</b>	NVIDIA GTX 1050 Ti	Keras

*Cardurile video folosite în cele trei soluții.*

Tabelul de mai sus prezintă cardurile video folosite în cele trei experimente. Am folosit Nvidia GTX 1050Ti pentru soluția propusă, iar același model, Nvidia Tesla K40c, a fost folosit în celelalte două.

<b>Model</b>	<b>Long &amp; Tensmeyer</b>	<b>Proposed solution</b>
<b>Specification \ Video card</b>	<b>NVIDIA Tesla K40c</b>	<b>NVIDIA GTX 1050 Ti</b>
Memory type	GDDR5	GDDR5
Maximum RAM amount [GB]	<b>12</b>	4
Memory bus width [Bit]	<b>384</b>	128
Memory clock speed [MHz]	6008	<b>7008</b>
Memory bandwidth [GB/s]	<b>288</b>	112
Pipelines / CUDA cores	<b>2880</b>	768
Core clock speed [MHz]	745	<b>1291</b>
Boost clock speed [MHz]	876	<b>1392</b>
Number of transistors [million]	<b>7080</b>	3300
Power consumption [Watt]	245	<b>75</b>
Floating-point performance [gflops]	<b>4291</b>	2138
Launch price	\$7,699	<b>\$139</b>
Current price	\$360	<b>\$184</b>
Value for money	2.06	<b>11.36</b>

*Compararea specificațiilor: Tesla K40c vs GTX 1050Ti*

*Source: [26]*

Tabelul de mai sus compară specificațiile celor două tipuri de carduri video. Observați diferențele dintre modelul NVIDIA Tesla K40c și modelul NVIDIA GTX 1050Ti: memoria RAM (12 GB față de 4 GB), numărul de nuclee (2880 față de 788), prețul de lansare (\$7699 față de \$139) și prețul curent (\$360 față de \$184).

Concluzia este că, folosind resurse limitate în soluția propusă, s-au obținut rezultate foarte apropiate de cele obținute în celelalte două experimente în care resursele au fost superioare.



## 3.2 Segmentarea

Această studiu se concentrează pe îmbunătățirea segmentării imaginilor, un aspect cheie al viziunii artificiale folosit în domenii precum conducerea autonomă și extragerea de date. Strategia propusă combină rezultatele mai multor algoritmi folosind diferite metode de votare pentru a îmbunătăți precizia segmentării [27]. 'Segmentarea imaginilor' cuprinde algoritmi care își propun să categorizeze și să ordoneze pixelii în funcție de criterii specifice. Hoover et al. [28] au delimitat aceste provocări în categorii precum zgomot, detecții ratate, supra-segmentare și sub-segmentare. Interesant, sub-segmentarea adesea eclipsează supra-segmentarea, deoarece este mai simplu să rafinăm segmente mai mari decât să fragmentăm regiuni extinse [29].

Într-o încercare de a depăși rezultatele individuale ale algoritmilor de segmentare prevalenți, această studiu dezvăluie o metodă de votare care combină mai mulți algoritmi. Algoritmii selectați pentru această încercare includ "Segmentarea Felzenszwalb" [30], "Super-pixelii SLIC" [31], [32] și "Watershed" [33], [34], în principal datorită tendinței lor de a furniza rezultate distincte care sunt fie supra-segmentate, fie sub-segmentate.

Elucidând algoritmul bazat pe grafurile lui Felzenszwalb [70], fiecare pixel este vizualizat ca un nod în rețea în timpul segmentării. Interconexiunile între noduri tipifică perechile specifice de pixeli. Fiecare dintre aceste muchii vine cu o greutate asociată, reflectând caracteristicile grupurilor de pixeli. Tehnica lui Felzenszwalb [30] prezintă o versatilitate unică, cum ar fi capacitatea sa de a regla cu precizie criteriile de segmentare în conformitate cu particularitățile locale.

Super-pixelii SLIC introduc o perspectivă nouă prin utilizarea super-pixelilor. Prin gruparea pixelilor similari în secțiuni de pixeli, această metodă oferă o abordare inovatoare a grilei convenționale de pixeli. Achanta et al. [31] au evidențiat faptul că această tehnică nu doar simplifică complexitatea, ci folosește și primitive care modelează componente esențiale ale imaginii.

Al treilea algoritm, Watershed, privește imaginea ca pe un peisaj topografic, inspirându-se din conceptul original al lui Digabel și Lantuejoul [33]. Cu această perspectivă, valorile gri ale pixelilor sau intensitățile gradientului dictează contururile peisajului. Obiectivul principal al metodei Watershed este să delimiteze imaginea în zone focale variate prin discernerea intensităților pixelilor.

Discutând despre metodologiile de votare, există numeroase abordări pentru segmentarea imaginilor, fiecare având avantajele și dezavantajele sale în raport cu datele de intrare. Ideea de a combina decizii din diferite metode, adesea numite tehnici de votare, a câștigat teren în studiile recente.

Image	Number of segments			
	Felzenszwalb	SLIC Superpixels	Watershed	Voting method
Astronaut	194	190	256	198
Chelsea	86	237	247	117
Hubble deep field	181	212	255	188
Coffee	109	223	260	134
Retina	295	226	256	283

În ceea ce privește rezultatele, dată fiind absența unui standard universal pentru segmentarea imaginilor, este esențial să abordăm această operație cu o mentalitate "optimă" pentru un anumit context. Acest lucru necesită echilibrarea nivelului de segmentare dorit cu alocările de resurse, cum ar fi puterea de calcul și timpul. Prin urmare, optimizarea segmentării ar trebui să echilibreze atât calitatea necesară, cât și costurile corespunzătoare.

### 3.3 OCR

Această secțiune introduce o nouă tehnică OCR care combină rezultatele din diverse metode. Prin înțelegerea fiecărei abordări OCR, putem decide între rezultate. Metoda utilizează un sistem de votare și preprocesează imaginea documentului pentru a evidenția caracteristicile textului, având ca scop obținerea celei mai precise citiri OCR. Rezultatele arată că această abordare se descurcă constant bine în diferite scenarii, ceea ce o face potrivită pentru procesarea documentelor nesupravegheate și extragerea de date [40].

Metoda propusă este un sistem bazat pe votare care aplică diverse filtre pentru a îmbunătăți calitățile vizuale ale imaginilor. Aceste filtre ajustează contrastul, accentuarea și efectuează acțiuni morfologice precum dilatația și eroziunea utilizând diferite nuclee. Odată ce filtrele sunt aplicate, imaginea trece prin procesul de OCR. Nivelul de încredere al fiecărui cuvânt este apoi comparat, iar rezultatul cel mai încrezător este reținut. Acest proces se repetă pe imaginea originală pentru rezultate mai rafinate. În cele din urmă, rezultatele mai multor încercări de OCR sunt comparate și integrate folosind un mecanism de votare adecvat.

Pentru a îmbunătăți contrastul între elemente și a garanta recuperarea componentelor corelate fără ambiguitate, se utilizează binarizarea imaginii [41].

Studiul introduce o metodă OCR bazată pe votare care aplică diverse tehnici de preprocesare imaginilor pentru a îmbunătăți caracteristicile textului. În loc să folosească încrederea pe caracter comună, sistemul utilizează încrederea pe bază de cuvânt, calculată din litere individuale, pentru o precizie sporită. După procesare, cuvintele cu cea mai mare încredere sunt selectate. Tesseract, un motor OCR popular, se confruntă adesea cu probleme de imagine precum zgomot, iluminare slabă sau contrast inconsistent. Prin efectuarea mai multor etape de preprocesare, anumite zone ale imaginii sunt îmbunătățite, permițând obținerea unor rezultate OCR mai fiabile. Scopul este să se adune cele mai bune rezultate din fiecare rulare. Fiecare acțiune de preprocesare poate crește acuratețea detecției pentru unele cuvinte, dar poate scădea pentru altele. Cu toate că este intens din punct de vedere al timpului, această metodă minimizează eficient detectarea incorectă a cuvintelor.

Strategia implică activarea diferitelor atribute vizuale din imaginea de intrare utilizând filtre. Aceste procese ajustează factorii imaginii precum contrastul și aplică acțiuni morfologice. Binarizarea este utilizată pentru a îmbunătăți ulterior contrastul. Rezultatele demonstrează abilitatea metodei de a îmbunătăți citirea OCR în diferite scenarii, fiind ideală pentru procesarea documentelor nesupravegheate.

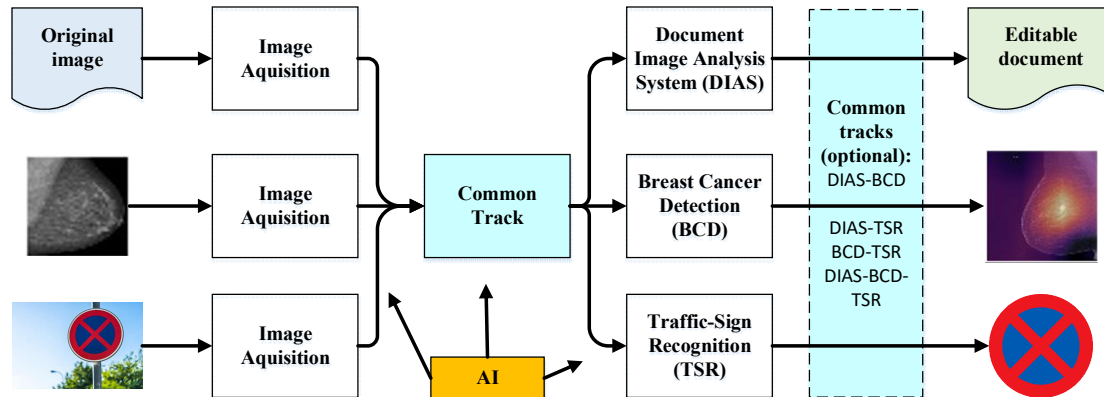
Istoric, tehnologia OCR a arătat importanța extragerii textului din imagini. Tesseract a înregistrat îmbunătățiri semnificative, dar se confruntă în continuare cu provocări datorită calității imaginii. Esența metodei propuse constă în tehnica bazată pe votare, în care se folosesc diferite filtre de imagine și se combină rezultatele pentru a optimiza rezultatul final.

Eficiența abordării în îmbunătățirea detecției textului, deși solicitantă din punct de vedere al timpului, este evidentă, făcând-o o unealtă promițătoare pentru proiectele de digitalizare pe scară largă, precum inițiativa Lib2Life, care vizează digitalizarea cuprinzătoare a documentelor din bibliotecile românești.

### **3.4 Arhitectură Unificată de Procesare a Imaginilor: Analiză Documente, Imagistică Medicală și Recunoaștere a Indicatoarelor Rutiere**

Acest capitol se adâncește într-o arhitectură unificată de procesare a imaginilor care integrează trei domenii specifice: analiza documentelor, imagistica medicală și recunoașterea indicatoarelor rutiere, construind pe procesele fundamentale discutate în capitolele 3.1 până la 3.3. Arhitectura inovatoare își propune să gestioneze în mod coerent și eficient datele de imagine în aceste categorii. Prin examinarea designului și funcționării sale, căutăm să

înțelegem beneficiile unei astfel de abordări integrate, impactul său asupra performanței sistemului și potențialul său în diverse aplicații practice. Scopul este de a sublinia avantajele combinării diferitelor tehnici de procesare a imaginilor în aceste domenii.



*Arhitectură Unificată de Procesare a Imaginilor*

În viitor, am putea vedea un dispozitiv care, cu perifericele potrivite, poate gestiona procesarea imaginilor pentru diverse domenii. Această arhitectură vizionată este modulară, ceea ce înseamnă că operațiile pot fi adăugate după cum este nevoie. Există module comune pentru toate fluxurile de date, iar imaginile sunt obținute prin echipamente specializate. Post-procesarea poate utiliza, de asemenea, aceste module comune, iar modelele de inteligență artificială pot fi integrate în diferite etape ale procesului.

Sistemele de prelucrare a imaginilor, esențiale în mai multe domenii, includ module comune de pre-procesare, post-procesare și module specifice domeniilor. Pre-procesarea implică standardizarea imaginilor, segmentarea acestora în părți relevante, extragerea caracteristicilor, aplicarea transformărilor avansate și reducerea zgomotului. Post-procesarea îmbunătățește claritatea imaginii și extrage informații valoroase. Modulele de inteligență artificială, adaptate la domenii precum documentele, imaginile medicale sau semnele de circulație, abordează provocări unice prin pre-procesarea datelor, recunoașterea obiectelor, segmentarea semantică și tehnici de învățare automată. În timp ce astfel de sisteme integrate oferă eficiență, avantaje în ceea ce privește costurile și adaptabilitatea, ar putea întâmpina și provocări legate de complexitate, specificitate și vulnerabilități de securitate.

## 4 Studiu de Caz: Detectarea Cancerului de Sân

Acest studiu se concentrează pe îmbunătățirea detecției cancerului de sân prin utilizarea metodelor de învățare adâncă, în special arhitecturile CNN și ViT. Prin includerea augmentării imaginilor sintetice, performanța modelelor este semnificativ îmbunătățită. Concluziile subliniază importanța preprocesării și augmentării datelor pentru clasificare optimă. În plus, cercetarea utilizează tehnici de inteligență artificială explicabilă pentru a oferi o înțelegere mai clară a modului în care modelele iau decizii [6].

Cancerul de sân (CS), cel mai comun tip de cancer la nivel mondial, a dus la peste 2,3 milioane de cazuri și peste 685.000 de decese în 2020 [5]. Acesta are un impact profund asupra numărului global de decese legate de cancer, în special printre femei [6].

Contribuția acestui articol în literatură este dublă: (a) furnizarea unei recenzii critice a tehnicilor de imagistică medicală pentru detectarea CS, precum și a metodelor de învățare adâncă pentru detectarea și clasificarea CS; (b) prezentarea studiului experimental propus pentru detectarea CS pe baza mai multor tehnici de învățare adâncă pe imagini de mamografie.

Abordarea inovatoare este reprezentată de generarea de imagini sintetice, care implică completarea setului de date cu imagini realiste ale sânilor digitale folosind algoritmi de învățare automată care pot imita aspectul mamografiilor reale.

Imagistica medicală este crucială pentru detectarea precoce a cancerului de sân, cu un accent din ce în ce mai mare pe tehnologiile digitale. Metodele principale de identificare a CS se concentrează pe detectarea calcificărilor, care sunt depozite de calciu în țesutul mamar [42]. Analizăm mai jos cele mai comune tipuri de tehnici de imagistică medicală pentru diagnosticarea CS, precum mamografia, ultrasunetele, RMN-ul, termografia și histopatologia, precum și metodele de învățare adâncă pentru detectarea și clasificarea CS, cum ar fi transferul de învățare și augmentarea datelor, extragerea de caracteristici și arhitecturile bazate pe mai multe modele și Generative Adversarial Networks.

În acest studiu, ne propunem să experimentăm cu tehnici bazate pe învățarea adâncă pentru detectarea cancerului în mamografii. Pentru aceasta, utilizăm setul de date furnizat de Societatea Radiologică din America de Nord (RSNA) într-o competiție Kaggle [72]. Scopul său a fost detectarea cazurilor de cancer de sân în mamografiile obținute în timpul examinărilor de screening.

Setul de date folosit în această competiție se bazează pe setul de date ADMANI [73], care conține mamografiile digitale și date non-imagini adnotate și este posibil cel mai vast și divers set de date pentru mamografiile documentat în literatură. Setul de date cuprinde un total de 28.911 cazuri de cancer de sân, din care 22.270 au fost detectate în timpul screening-ului și 6641 au fost cazuri de interval, mult mai multe decât oricare alt set de date publicat până acum. Setul de date conține, de asemenea, un număr mare de examinări (1.048.345) și pacienți (629.863), făcându-l unul dintre cele mai mari seturi de date raportate până în prezent [74]. Cu toate acestea, setul de antrenament pentru această competiție publică constă din 11.913 examinări, din care 486 sunt cazuri de cancer, ceea ce face ca setul de date să fie foarte dezechilibrat.

Am experimentat cu mai multe modele de învățare adâncă bazate pe arhitecturi CNN și ViT, cum ar fi ResNet [77] sau EfficientNet [78], și MaxViT [79].

XAI poate, de asemenea, ajuta la identificarea tendințelor sau erorilor din procesul decizional al modelului AI. De exemplu, poate ajuta la identificarea caracteristicilor din imagine care influențează disproporționat decizia, sau poate evidenția cazurile în care modelul poate lua decizii incorecte sau cu tendințe.

Fastai [80] oferă o soluție Grad-CAM (gradient-weighted class activation mapping) [81] pentru vizualizarea regiunilor dintr-o imagine de intrare care sunt cele mai importante pentru predicția unui model. Grad-CAM [82] este o tehnică populară pentru generarea de hărți de căldură care evidențiază regiunile dintr-o imagine care sunt cele mai importante pentru o anumită clasă, și a fost utilizată în diverse aplicații de viziune pe calculator.

Harta de activare a clasei utilizează ieșirea ultimului strat convoluțional împreună cu greutatea stratului complet conectat care corespunde clasei prezise. Aceasta calculează produsul scalar al acestor greutăți și a caracteristicilor pentru fiecare poziție din harta de trăsături, astfel încât să fie posibil să se obțină scorul caracteristicii folosit pentru predicție.

## 4.1 Rezultate

Pentru experimentare, imaginile din setul de date au fost mai întâi redimensionate la o dimensiune uniformă de  $256 \times 256$  și  $512 \times 512$  pixeli. Performanța modelelor a fost evaluată folosind acuratețea, AUC (area under the curve), precizia, recuperarea și scorul F1.

---

Mammograms with $256 \times 256$ resolution					
Model	Accuracy	AUC	Precision	Recall	F1

---

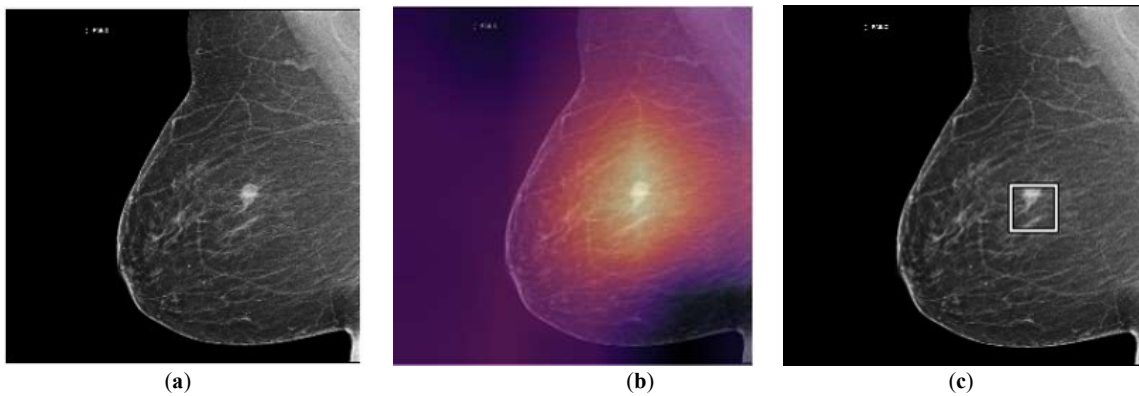
Original images					
ResNet18	<b>0.96</b>	0.61	0.06	0.04	0.04
ResNet34	<b>0.96</b>	0.63	0.08	0.06	0.06
ResNet152	0.89	0.58	0.03	<b>0.17</b>	0.06
EfficientNetB0	<b>0.96</b>	0.59	0.03	0.02	0.02
MaxViT	0.95	<b>0.66</b>	<b>0.09</b>	0.15	<b>0.11</b>
Processed images					
ResNet18	0.92	<b>0.67</b>	0.07	0.22	0.11
ResNet34	0.93	<b>0.67</b>	<b>0.09</b>	0.21	<b>0.13</b>
ResNet152	0.82	0.64	0.04	<b>0.33</b>	0.07
EfficientNetB0	<b>0.95</b>	0.63	0.08	0.11	0.09
MaxViT	0.93	<b>0.67</b>	<b>0.09</b>	0.21	<b>0.13</b>
Processed images + synthetic images					
ResNet18	0.92	0.83	0.26	0.58	0.36
ResNet34	0.94	0.82	0.34	0.56	0.42
ResNet152	0.84	0.8	0.15	<b>0.63</b>	0.24
EfficientNetB0	<b>0.95</b>	0.8	<b>0.39</b>	0.52	<b>0.45</b>
MaxViT	0.9	<b>0.85</b>	0.23	<b>0.63</b>	0.34
Mammograms with 512 × 512 resolution					
Model	Accuracy	AUC	Precision	Recall	F1
Original images					
ResNet18	<b>0.97</b>	0.66	<b>0.16</b>	0.09	0.12
ResNet34	<b>0.97</b>	0.66	<b>0.16</b>	0.09	0.12
ResNet152	0.83	0.59	0.03	<b>0.26</b>	0.06
EfficientNetB0	0.96	0.64	0.1	0.06	0.08
MaxViT	0.95	<b>0.69</b>	0.11	0.18	<b>0.14</b>
Processed images					
ResNet18	0.93	0.72	0.1	0.26	0.15
ResNet34	0.94	0.71	<b>0.13</b>	0.24	<b>0.16</b>
ResNet152	0.83	0.69	0.05	0.4	0.09
EfficientNetB0	<b>0.95</b>	0.69	0.12	0.18	0.14
MaxViT	0.89	<b>0.77</b>	0.08	<b>0.43</b>	0.14
Processed images + synthetic images					
ResNet18	0.94	0.85	0.35	0.59	0.44
ResNet34	<b>0.95</b>	0.84	0.4	0.58	0.47
ResNet152	0.83	0.82	0.14	0.66	0.23
EfficientNetB0	<b>0.95</b>	0.84	<b>0.47</b>	0.55	<b>0.51</b>
MaxViT	0.89	<b>0.88</b>	0.22	<b>0.72</b>	0.34

*Rezultate experimentale*

Performanța modelelor CNN relativ simple, cum ar fi ResNet18, a fost aproape la fel de bună ca a modelelor mai mari, cum ar fi ResNet152 sau EfficientNet B0. Mai mult, performanța modelelor CNN s-a dovedit a fi comparabilă cu cea a modelului bazat pe arhitectura transformatorului vizual. Cu toate acestea, ceea ce a făcut cea mai mare diferență a fost prelucrarea prealabilă a setului de date, în special augmentarea setului de date cu imagini

generate sintetic. Această tehnică de prelucrare prealabilă a contribuit la îmbunătățirea performanței generale a modelelor, rezultând în metrice de evaluare mai bune.

Studiul indică faptul că o prelucrare prealabilă meticuloasă și o augmentare a datelor pot îmbunătăți semnificativ performanța chiar și a modelelor de bază. Acest lucru subliniază necesitatea unei prelucrări prealabile optime și valoarea datelor sintetice în îmbunătățirea modelelor de viziune artificială. După clasificare, tehnicile de vizualizare au identificat zonele imaginii care influențează decizia modelului, cu o casetă încadrată în centrul care evidențiază regiunea cheie folosită pentru determinarea sa.



*Explicarea rezultatelor. (a) Imaginea originală. (b) Aplicarea hărților de activare a claselor. (c) Aplicarea unei casete încadrate*

Aceste tehnici de vizualizare au furnizat o modalitate utilă de înțelegere a modului în care modelul de învățare profundă a clasificat imaginile ca pozitive sau negative. Acestea au ajutat la identificarea zonelor specifice ale imaginii pe care modelul le-a utilizat pentru a lua decizia sa și au furnizat o modalitate mai intuitivă de interpretare a rezultatelor modelului.



## 5 Concluzii

### *Capitolul 2*

Sistemele de Analiză a Imaginilor Documentelor sunt esențiale pentru digitalizarea textelor istorice, având ca scop procesarea autonomă pentru a îmbunătăți viteza și precizia. Acest lucru deschide calea către o schimbare digitală în gestionarea vastelor colecții literare, îmbunătățind accesul și structura.

Detectarea Cancerului de Sân subliniază diagnosticul precoce și rolul Inteligenței Artificiale în îmbunătățirea evaluărilor mamografice. Cu diverse tehnici de imagistică, IA se remarcă prin capacitatea sa transformatoare de detecție. În ciuda provocărilor precum seturile limitate de date, inovațiile precum imaginile sintetice și rețelele antagoniste generative (GAN-uri) arată potențial în clasificarea cancerului, subliniind rolul vital al IA în îngrijirea pacientului.

Procesarea Automată a Imaginilor Documentelor, dat fiind volumul vast de documente din era digitală, explorează dezvoltarea sa și metodele de extragere de ultimă generație. Avansurile în IA, cum ar fi rețelele neuronale profunde, au catalizat avansuri în diverse domenii precum sănătatea și apărarea. Cu arhitecturi eficiente, aceste tehnologii convertesc imaginile în date digitale valoroase, promițând o precizie sporită a procesării datelor în diferite industrii. Acest capitol oferă o prezentare cuprinzătoare a semnificației și viitorului procesării imaginilor documentelor.

### *Capitolul 3*

Procesul de Binarizare a Imaginilor Documentelor: Acest studiu introduce o soluție promițătoare pentru binarizarea eficientă a imaginilor, folosind Python și biblioteci disponibile. Deși eficace, există un compromis în privința calității vizuale, în special atunci când defectele documentelor sunt evidențiate. Creșterea straturilor de convoluție și utilizarea unui set de date mai mare ar putea îmbunătăți rezultatele.

Segmentarea: Acest capitol se adâncește în dinamica votării în algoritmul de segmentare a imaginilor, punând accent pe algoritmi nesupervizați. Deși învățarea profundă promite, accentul actual este pus pe modul în care votarea influențează diverse caracteristici în rezultatele de segmentare. Scopul final este de a combina această metodă cu alte sisteme nesupervizate pentru un sistem eficient de procesare a imaginilor documentelor.

OCR: Prin introducerea unui nou sistem bazat pe votare, această secțiune discută îmbunătățirea preciziei OCR prin aplicarea unor operații de preprocesare multiple asupra documentelor cu imagini. Cu toate că poate necesita mai mult timp de procesare, robustețea sa îl face ideal pentru proiecte de digitalizare la scară largă, precum proiectul Lib2Life pentru bibliotecile române.

Arhitectură Unificată de Procesare a Imaginilor: Această arhitectură avansată integrează analiza documentelor, imagistica medicală și recunoașterea semnelor de circulație. Utilizând module de preprocesare comune și tehnici de IA, urmărește eficiența în mai multe domenii. În timp ce oferă mai multe avantaje, are și complexități și provocări de depășit. Dispozitivul viitor vizionat ar fi o unealtă multifuncțională pentru toate cele trei domenii, reprezentând un pas semnificativ în tehnologia de procesare a imaginilor.

#### ***Capitolul 4***

Acest capitol prezintă un studiu detaliat despre Detectarea Cancerului de Sân, evidențiind importanța diagnosticului precoce pentru tratamentul cu succes, mai ales în considerarea prevalenței cancerului de sân printre femei. Studiul explorează aplicarea tehnicilor de învățare profundă pentru îmbunătățirea acurateței și eficienței fluxurilor de screening pentru cancerul de sân, utilizând diverse metode de imagistică, cum ar fi mamografia, termografia, ultrasunetele și imagistica prin rezonanță magnetică.

Introducerea tehnologiei AI a demonstrat un mare potențial în automatizarea detectării cancerului de sân, cu algoritmi AI care demonstrează performanțe comparabile cu experții umani în seturile retrospective de date. Această zonă emergentă de screening pentru cancerul de sân alimentat de AI are un potențial semnificativ pentru îmbunătățirea detectării timpurii și a prognosticului, conducând în cele din urmă la rezultate mai bune pentru pacienți și la o reducere a deceselor legate de cancerul de sân.

Focalizarea principală a studiului a fost pe utilizarea metodelor de învățare profundă pentru analizarea mamografiei, iar rezultatele au subliniat importanța tehnicilor adecvate de preprocesare și augmentare a datelor. Prin utilizarea tehnicilor de generare a imaginilor sintetice, performanța de clasificare a modelelor de învățare profundă a fost îmbunătățită semnificativ, demonstrând potențialul acestor metode în sporirea acurateței sarcinilor de clasificare a imaginilor.

În rezumat, descoperirile studiului evidențiază impactul transformator al tehnicilor de învățare profundă alimentate de AI în detectarea cancerului de sân, arătând performanțe comparabile cu cele ale experților umani și potențialul pentru o acuratețe chiar mai mare cu tehnicile adecvate de preprocesare a datelor și generare de date sintetice. Integrarea AI în screening-ul cancerului de sân are potențialul de a îmbunătăți detectarea timpurie, prognosticul îmbunătățit și, în cele din urmă, contribuie la rezultate mai bune pentru pacienți în lupta împotriva cancerului de sân.

## 5.1 Contribuții

Una dintre principalele contribuții ale acestui studiu este evaluarea cuprinzătoare a aplicabilității prelucrării imaginilor în diverse medii tehnologice. Prin examinarea capabilităților și limitărilor diferitelor metode de prelucrare a imaginilor, am furnizat insight-uri valoroase cu privire la potrivirea lor pentru aplicații specifice. Această analiză servește ca bază pentru luarea deciziilor informate și optimizarea sarcinilor de prelucrare a imaginilor, deschizând calea pentru soluții de prelucrare mai eficiente și mai eficace.

O contribuție majoră a acestei cercetări constă în evidențierea rolului central al inteligenței artificiale (IA) în sistemele moderne de prelucrare a imaginilor. Prin prezentarea unei analize temeinice a tehnicilor de prelucrare a imaginilor ghidate de IA, am subliniat impactul lor transformator asupra analizei imaginilor, interpretării și extragerii caracteristicilor. Integrarea algoritmilor de IA conferă prelucrării imaginilor abilități cognitive avansate, sporindu-le capacitatea de a gestiona sarcini complexe și de a extrage informații semnificative din imagini intricate.

## 5.2 Lista publicațiilor

### 5.2.1 Journale

1. **Marcel PRODAN**; Elena PARASCHIV; Alexandru STANCIU. “*Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection*”. MDPI, Appl. Sci. 2023, 13, 4272. (GoogleScholar), <https://doi.org/10.3390/app13074272>
2. **Marcel PRODAN**, Costin-Anton BOIANGIU, “*Document Image Binarization Process*”, BRAIN. Broad Research in Artificial Intelligence and Neuroscience, <https://doi.org/10.18662/brain/14.2/446>
3. **Marcel PRODAN**, Giorgiana Violeta VLASCEANU, and Costin-Anton BOIANGIU; “Comprehensive Evaluation of Metrics for Image Resemblance”, The Journal of Information Systems & Operations Management, vol. 17, no. 1, May 2023, [Online], Available: [https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1\\_161-185.pdf](https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1_161-185.pdf)
4. Remus PETRESCU, Sergiu MANOLACHE, Costin-Anton BOIANGIU, Giorgiana Violeta VLASCEANU, Cristian AVATAVULUI, **Marcel PRODAN**, Ion BUCUR; “*Combining Tesseract and Asprise Results to Improve OCR Text Detection Accuracy*”, The Journal of Information Systems & Operations Management, Vol.13 No.1 - 2019, ISSN: 1843-4711, pp. 57-64 (CNCSIS B+/BDI: ProQuest, EBSCO, RePEc, Copernicus, QBE, GoogleScholar)
5. Nicolae TARBĂ, Daniel SCHMIDT, Anda - Elena POPOVICI, Eduard STĂNILOIU, Cristian AVATAVULUI, **Marcel PRODAN**; “*On Performing Skew Detection And Correction Using Multiple Experts’ Decision*”, The Journal of Information Systems & Operations Management, Vol.14 No.2 - 2020, ISSN: 1843-4711, pp. 188-195
6. Iulia-Cristina Stănică, Costin-Anton Boiangiu, Giorgiana Violeta Vlăsceanu, **Marcel PRODAN**, Cristian Avatavului, Răzvan-Adrian Deaconescu, Codrin Tăut, “A Survey on History, Present and Perspectives of Document Image Analysis Systems”, in New technologies and redesigning learning spaces Book of abstracts, page 36, 15th eLearning and Software for Education Conference, Bucharest, 2019, ISSN 2360-2198, doi: 10.12753/2066-026X-19-025 (ISI Proceedings, CEEOL, PROQUEST, EBSCO) [WOS: 000473322400025]
7. Giorgiana Violeta Vlăsceanu, Costin-Anton Boiangiu, Răzvan-Adrian Deaconescu, **Marcel PRODAN**, Cristian Avatavului, Răzvan Rughiniș, Irina Mocanu, “Designing a Document Image Analysis System on 3 Axis: Education, Research and Performance”, in New technologies and redesigning learning spaces Book of abstracts, page 37, 15th eLearning and Software for Education Conference, Bucharest, 2019, ISSN 2360-2198, doi: 10.12753/2066-026X-19-027 (ISI Proceedings, CEEOL, PROQUEST, EBSCO) [WOS: 000473322400027]
8. Ioana Monica DICHER, Ana-Georgia ȚURCUȘ, Eduard-Marius COJOCEA, Patricia-Steliana PENARIU, Ion BUCUR, **Marcel PRODAN**, Eduard STĂNILOIU; “*Unsupervised Merge of Optical Character Recognition Results*”, The Journal of Information Systems & Operations Management, Vol.14 No.1 - 2020, ISSN: 1843-4711, pp. 60-67
9. Alin-Florin MIHĂILĂ, Patricia-Steliana PENARIU, Giorgiana Violeta VLĂSCEANU, **Marcel PRODAN**; “*On Image Segmentation Using a Combination of Felzenszwalb, Slic and Watershed Methods*”, The Journal of Information Systems & Operations Management, Vol.14 No.1 - 2020, ISSN: 1843-4711, pp. 121-129
10. Cristian AVATAVULUI, **Marcel PRODAN**; “Evaluating Image Contrast: A Comprehensive Review and Comparison of Metrics”, Journal of Information Systems

& Operations Management, Vol. 17.1, May 2023, [Online], Available [https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1\\_143-160.pdf](https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1_143-160.pdf).

### 5.2.2 Conferințe

1. **Marcel PRODAN**, Elena Paraschiv, Alexandru Stanciu, “*Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection*”. at “9th International Conference on Physical Health, Public Health & Health Management”, Kington HR53DJ, UK, 2023 [online], <https://physicaltherapy.scientificmeditech.com/>
2. Iulia-Cristina Stănică, Costin-Anton Boiangiu, Giorgiana Violeta Vlăsceanu, **Marcel PRODAN**, Cristian Avatavului, Răzvan-Adrian Deaconescu, Codrin Tăut, “*A Survey on History, Present and Perspectives of Document Image Analysis Systems*”, in New technologies and redesigning learning spaces Book of abstracts, page 36, 15th eLearning and Software for Education Conference, Bucharest, 2019, ISSN 2360-2198, doi: 10.12753/2066-026X-19-025 (ISI Proceedings, CEEOL, PROQUEST, EBSCO) [WOS: 000473322400025]
3. Giorgiana Violeta Vlăsceanu, Costin-Anton Boiangiu, Răzvan-Adrian Deaconescu, **Marcel PRODAN**, Cristian Avatavului, Răzvan Rughiniș, Irina Mocanu, “*Designing a Document Image Analysis System on 3 Axis: Education, Research and Performance*”, in New technologies and redesigning learning spaces Book of abstracts, page 37, 15th eLearning and Software for Education Conference, Bucharest, 2019, ISSN 2360-2198, doi: 10.12753/2066-026X-19-027 (ISI Proceedings, CEEOL, PROQUEST, EBSCO) [WOS: 000473322400027]

### 5.2.3 Postere

1. **Marcel PRODAN**, “*Designing a Document Image Analysis System on 3 Axis: Education, Research and Performance*”, in Semicentennial Anniversary of the Department of Computers, University Politehnica of Bucharest, 2018

## Bibliografie

- [1] D. Eagleman, *The Brain: The Story of You*. Canongate Books, 2015. [Online]. Available: <https://books.google.ro/books?id=gUTGBwAAQBAJ>
- [2] I.-C. Stănică, C.-A. Boiangiu, G. V. Vlăsceanu, M. Prodan, C. Avatavului, R.-A. Deaconescu, and C. Tăut, "A Survey on History, Present and Perspectives of Document Image Analysis Systems", in *New technologies and redesigning learning spaces Book of abstracts*, Bucharest, 2019, p. 36., doi: 10.12753/2066-026X-19-025.
- [3] A. Tigora, "Designing A Flexible Document Image Analysis System – Part 1: The Architecture", *J. Inf. Syst. Oper. Manag.*, vol. 10, no. 1, pp. 235–245, May 2016.
- [4] R. Kasturi, L. O’Gorman, and V. Govindaraju, "Document image analysis: A primer", *Sadhana*, vol. 27, no. 1, Feb. 2002.
- [5] M. Arnold, E. Morgan, H. Rungay, A. Mafra, D. Singh, M. Laversanne, J. Vignat, J. R. Gralow, F. Cardoso, and S. Siesling, "Current and future burden of breast cancer: Global statistics for 2020 and 2040," *Breast*, vol. 66, pp. 15–23, 2022.
- [6] M. Prodan, E. Paraschiv, and A. Stanciu, "Applying Deep Learning Methods for Mammography Analysis and Breast Cancer Detection," *Appl Sci*, vol. 13, p. 4272, 2023, doi: 10.3390/app13074272.
- [7] WHO, "Breast cancer."
- [8] L. Chen, S. Li, Q. Bai, J. Yang, S. Jiang, and Y. Miao, "Review of Image Classification Algorithms Based on Convolutional Neural Networks," *Remote Sens.*, vol. 13, no. 22, 2021, doi: 10.3390/rs13224712.
- [9] M. PRODAN, G. V. VLASCEANU, and C.-A. BOIANGIU, "Comprehensive Evaluation of Metrics for Image Resemblance," *J. Inf. Syst. Oper. Manag.*, vol. 17, no. 1, May 2023, [Online]. Available: [https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1\\_161-185.pdf](https://web.rau.ro/websites/jisom/Vol.17%20No.1%20-%202023/JISOM%2017.1_161-185.pdf)
- [10] Z. WANG, A. C. BOVIK, H. R. SHEIKH, and E.-P. SIMONCELLI, "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600-612, Apr. 2004, doi: 10.1109/TIP.2003.819861.
- [11] R. GONZALEZ and R. WOODS, *Digital image processing*, 3rd ed. Upper Saddle River, N.J: Prentice Hall, 2002.
- [12] A. S. GOLESTANEH and D. M. CHANDLER, "No-reference quality assessment of jpeg images via a quality relevance map," *IEEE Signal Process. Lett.*, vol. 21, no. 2, pp. 155-158, 2004, doi: 10.1109/LSP.2013.2296038.
- [13] F. CHOLLET, *Deep learning with Python*. New York: Manning, 2007.
- [14] I. GOODFELLOW, Y. BENGIO, and A. COURVILLE, *Deep Learning*. MIT Press, 2016.
- [15] Z. WANG and A. C. BOVIK, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, 2009, doi: 10.1109/MSP.2008.930649.
- [16] Z. WANG, E. P. SIMONCELLI, and A. C. BOVIK, "Multiscale structural similarity for image quality assessment," in *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers*, Pacific Grove, CA, USA, 2003, pp. 1398–1402. doi: 10.1109/ACSSC.2003.1292216.
- [17] L. ZHANG, L. Zhang, and A. C. BOVIK, "A feature-enriched completely blind image quality evaluator," *IEEE Trans. Image Process.*, vol. 20, no. 8, pp. 2091-2100, 2011, doi: 10.1109/TIP.2015.2426416.

- [18] H. R. SHEIKH and A. C. BOVIK, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, 2006, doi: 10.1109/TIP.2005.859378.
- [19] Z. WANG, A. C. BOVIK, and B. L. EVANS, "Blind measurement of blocking artifacts in images," in *Proceedings 2000 International Conference on Image Processing (Cat. No.00CH37101, Canada, 2000*, pp. 981-984. doi: 10.1109/ICIP.2000.899622.
- [20] K. SIMONYAN and A. ZISSERMAN, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations (ICLR), 2015*, doi: 10.48550/arXiv.1409.1556.
- [21] H. R. SHEIKH, M. F. SABIR, and A. C. BOVIK, "A statistical evaluation of recent full reference image quality assessment algorithms," *IEEE Trans. Image Process.*, vol. 15, no. 11, pp. 3440-3451, 2006, doi: 10.1109/TIP.2006.881959.
- [22] M. PRODAN and C.-A. BOIANGIU, "Document Image Binarization Process," *BRAIN Broad Res. Artif. Intell. Neurosci.*, vol. 14, no. 2, pp. 93-114, Jun. 2023, doi: <https://doi.org/10.18662/brain/14.2/446>.
- [23] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR, Boston, MA, USA, 2015*, pp. 3431-3440. doi: 10.1109/CVPR.2015.7298965.
- [24] C. Tensmeyer and T. Martinez, "Document Image Binarization with Fully Convolutional Neural Networks," in *14th IAPR International Conference on Document Analysis and Recognition (ICDAR, Kyoto, Japan, 2017*, pp. 99-104. doi: 10.1109/ICDAR.2017.25.
- [25] D.I.B.C.O., "DIBCO dataset." DIBCO, 2023. [Online]. Available: <https://dib.cin.ufpe.br/#!/resources/dibco>
- [26] "Tesla K40c vs GeForce GTX 1050 Ti," *Technical City*, Aug. 01, 2023. <https://technical.city/en/video/GeForce-GTX-1050-Ti-vs-Tesla-K40c>
- [27] A.-F. MIHĂILĂ, P.-S. PENARIU, G. V. VLĂSCEANU, and M. PRODAN, "ON IMAGE SEGMENTATION USING A COMBINATION OF FELZENSZWALB, SLIC AND WATERSHED METHODS," *J. Inf. Syst. Oper. Manag.*, vol. 14, no. 1, pp. 121-129.
- [28] A. Hoover, G. Jean-Baptiste, X. Jiang, P. J. Flynn, H. Bunke, and D. B. G. R. B. Fisher, "An experimental comparison of range image segmentation algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* 18, no. ue 7, pp. 673-689, Jul. 1996, doi: 10.1109/34.506791.
- [29] J. Sigut, F. Fumero, and O. Nuñez, "Over-and under-segmentation evaluation based on the segmentation covering measure," in *23rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision 2015, in volume: Short papers proceedings, 2015*, pp. 83-89,.
- [30] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient graph-based image segmentation," *Int. J. Comput. Vis.*, vol. 59, no. ue 2, pp. 167-181, Sep. 2004, doi: 10.1023/B:VISI.0000022288.19776.77.
- [31] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLICsuperpixels compared to state-of-the-art superpixel methods," *IEEE Transactions Pattern Anal. Mach. Intell.*, vol. 34, no. ue11, pp. 2274-2282, Nov. 2012, doi: 10.1109/TPAMI.2012.120.
- [32] A. Levinshtein, A. Stere, K. N. Kutulakos, D. J. Fleet, S. J. Dickinson, and K. Siddiqi, "TurboPixels: Fast Superpixels Using Geometric Flows," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 12, pp. 2290-2297, Dec. 2009, doi: 10.1109/TPAMI.2009.96.
- [33] H. Digabel and C. Lantuejoul, "Iterative Algorithms," in *Proceedings of the 2nd European Symposium Quantitative Analysis of Microstructures in Material Science, Biology and Medicine, 1978*, pp. 85-89.

- [34] A. S. Kornilov and I. V. Safonov, "Review: An overview of watershed algorithm implementations in open source libraries," *J. Imaging*, vol. 4, no. 10, p. 123, Oct. 2018, doi: 10.3390/jimaging4100123.
- [35] M. Y. Lui, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy rate superpixel segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Providence, USA, Jun. 2011, pp. 2097-2104, doi: 10.1109/CVPR.2011.5995323.
- [36] A. Schick, M. Fischer, and R. Stiefelhagen, "Measuring and evaluating the compactness of superpixels," in *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, Nov. 2012, pp. 930-934.
- [37] D. Stutz, A. Hermans, and B. Leibe, "Superpixels: An evaluation of the state-of-the-art in Computer Vision and Image Understanding," vol. 166, pp. 1-27, Jan. 2018, doi: 10.1016/j.cviu.2017.03.007.
- [38] A. Fred, "Finding consistent clusters in data partitions," *Int. Workshop Mult. Classif. Syst. MCS2001*, pp. 309-318, Jul. 2001, doi: 10.1007/3-540-48219-9\_31.
- [39] C.-A. Boiangiu and R. Ioanimescu, "Voting-Based Image Segmentation," *J. Inf. Syst. Oper. Manag.*, vol. 7, no. 2/December, pp. 211-220, Dec. 2013.
- [40] I. M. DICHER, A.-G. ȚURCUȘ, E.-M. COJOCEA, P.-S. PENARIU, I. BUCUR, M. PRODAN, and E. STĂNILOIU, "UNSUPERVISED MERGE OF OPTICAL CHARACTER RECOGNITION RESULTS," *J. Inf. Syst. Oper. Manag.*, vol. 14, no. 1, pp. 60-67.
- [41] C.-A. Boiangiu, I. Bucur, and A. Tigora, "The Image Binarization Problem Revisited: Perspectives and Approaches," *J. Inf. Syst. Oper. Manag.*, vol. 6, no. 2, pp. 419-427, 2012.
- [42] M. Clinic, "Breast calcifications: When to see a doctor," *Mayo Clinic*.
- [43] A. Altameem, C. Mahanty, R. C. Poonia, A. K. J. Saudagar, and R. Kumar, "Breast Cancer Detection in Mammography Images Using Deep Convolutional Neural Networks and Fuzzy Ensemble Modeling Techniques," *Diagnostics*, vol. 12, no. 8, p. 1812, Jul. 2022, doi: 10.3390/diagnostics12081812.
- [44] U. Health, "Breast Cancer Diagnosis," *ucsfhealth.org*.
- [45] A. Case, "Differences Between Screening & Diagnostic Mammograms," *Midstate Radiology Associates*. Dec. 2020.
- [46] O. V. Michailovich and A. Tannenbaum, "Despeckling of medical ultrasound images," *IEEE Trans. Ultrason. Ferroelectr. Freq. Control*, vol. 53, no. 1, pp. 64-78, Jan. 2006, doi: 10.1109/TUFFC.2006.1588392.
- [47] R. Rulaningtyas, A. S. Hyperastuty, and A. S. Rahaju, "Histopathology Grading Identification of Breast Cancer Based on Texture Classification Using GLCM and Neural Network Method," *J. Phys. Conf. Ser.*, vol. 1120, p. 012050, Nov. 2018, doi: 10.1088/1742-6596/1120/1/012050.
- [48] S. S. Boudouh and M. Bouakkaz, "Breast Cancer: Breast Tumor Detection Using Deep Transfer Learning Techniques in Mammogram Images," in *2022 International Conference on Computer Science and Software Engineering (CSASE)*, Duhok, Iraq: IEEE, Mar. 2022, pp. 289-294, doi: 10.1109/CSASE51777.2022.9759702.
- [49] L. Shen, L. R. Margolies, J. H. Rothstein, E. Fluder, R. McBride, and W. Sieh, "Deep Learning to Improve Breast Cancer Detection on Screening Mammography," *Sci. Rep.*, vol. 9, no. 1, p. 12495, Aug. 2019, doi: 10.1038/s41598-019-48995-4.
- [50] I. C. Moreira, I. Amaral, I. Domingues, A. Cardoso, M. J. Cardoso, and J. S. Cardoso, "INbreast," *Acad. Radiol.*, vol. 19, no. 2, pp. 236-248, Feb. 2012, doi: 10.1016/j.acra.2011.09.014.



- [51] W. M. Salama and M. H. Aly, "Deep learning in mammography images segmentation and classification: Automated CNN approach," *Alex. Eng. J.*, vol. 60, no. 5, pp. 4701–4709, Oct. 2021, doi: 10.1016/j.aej.2021.03.048.
- [52] P. Oza, P. Sharma, S. Patel, F. Adedoyin, and A. Bruno, "Image Augmentation Techniques for Mammogram Analysis," *J. Imaging*, vol. 8, no. 5, p. 141, May 2022, doi: 10.3390/jimaging8050141.
- [53] S. Maqsood, R. Damaševičius, and R. Maskeliūnas, "TTCNN: A Breast Cancer Detection and Classification towards Computer-Aided Diagnosis Using Digital Mammography in Early Stages," *Appl. Sci.*, vol. 12, no. 7, p. 3273, Mar. 2022, doi: 10.3390/app12073273.
- [54] S. Montaha, S. Azam, A. K. M. R. H. Rafid, P. Ghosh, Md. Z. Hasan, M. Jonkman, and F. De Boer, "BreastNet18: A High Accuracy Fine-Tuned VGG16 Model Evaluated Using Ablation Study for Diagnosing Breast Cancer from Enhanced Mammography Images," *Biology*, vol. 10, no. 12, p. 1347, Dec. 2021, doi: 10.3390/biology10121347.
- [55] R. S. Lee, F. Gimenez, A. Hoogi, K. K. Miyake, M. Gorovoy, and D. L. Rubin, "A curated mammography data set for use in computer-aided detection and diagnosis research," *Sci. Data*, vol. 4, no. 1, p. 170177, Dec. 2017, doi: 10.1038/sdata.2017.177.
- [56] T. Mahmood, J. Li, Y. Pei, and F. Akhtar, "An Automated In-Depth Feature Learning Algorithm for Breast Abnormality Prognosis and Robust Characterization from Mammography Images Using Deep Transfer Learning," *Biology*, vol. 10, no. 9, p. 859, Sep. 2021, doi: 10.3390/biology10090859.
- [57] K.-J. Tsai, M.-C. Chou, H.-M. Li, S.-T. Liu, J.-H. Hsu, W.-C. Yeh, C.-M. Hung, C.-Y. Yeh, and S.-H. Hwang, "A High-Performance Deep Neural Network Model for BI-RADS Classification of Screening Mammography," *Sensors*, vol. 22, no. 3, p. 1160, Feb. 2022, doi: 10.3390/s22031160.
- [58] L.-A. Dang, E. Chazard, E. Poncelet, T. Serb, A. Rusu, X. Pauwels, C. Parsy, T. Poclet, H. Cauliez, C. Engelaere, *et al.*, "Impact of artificial intelligence in breast cancer screening with mammography," *Breast Cancer*, vol. 29, no. 6, pp. 967–977, Nov. 2022, doi: 10.1007/s12282-022-01375-9.
- [59] N. M. ud din, R. A. Dar, M. Rasool, and A. Assad, "Breast cancer detection using deep learning: Datasets, methods, and challenges ahead," *Comput. Biol. Med.*, vol. 149, p. 106073, Oct. 2022, doi: 10.1016/j.compbiomed.2022.106073.
- [60] Y. Wang, L. Zhang, X. Shu, Y. Feng, Z. Yi, and Q. Lv, "Feature-Sensitive Deep Convolutional Neural Network for Multi-Instance Breast Cancer Detection," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, vol. 19, no. 4, pp. 2241–2251, Jul. 2022, doi: 10.1109/TCBB.2021.3060183.
- [61] J. G. Melekoodappattu, A. S. Dhas, B. K. Kandathil, and K. S. Adarsh, "Breast cancer detection in mammogram: combining modified CNN and texture feature based approach," *J. Ambient Intell. Humaniz. Comput.*, Jan. 2022, doi: 10.1007/s12652-022-03713-3.
- [62] G. Altan, "Deep Learning-based Mammogram Classification for Breast Cancer," *Int. J. Intell. Syst. Appl. Eng.*, vol. 8, no. 4, pp. 171–176, Dec. 2020, doi: 10.18201/ijisae.2020466308.
- [63] H. M. Frazer, A. K. Qin, H. Pan, and P. Brothie, "Evaluation of deep learning-based artificial intelligence techniques for breast cancer detection on mammograms: Results from a retrospective study using a BreastScreen Victoria dataset," *J. Med. Imaging Radiat. Oncol.*, vol. 65, no. 5, pp. 529–537, Aug. 2021, doi: 10.1111/1754-9485.13278.
- [64] Y. Eroğlu, M. Yildirim, and A. Çinar, "Convolutional Neural Networks based classification of breast ultrasonography images by hybrid method with respect to benign,

- malignant, and normal using mRMR,” *Comput. Biol. Med.*, vol. 133, p. 104407, Jun. 2021, doi: 10.1016/j.combiomed.2021.104407.
- [65] D. Zhai, B. Hu, X. Gong, H. Zou, and J. Luo, “ASS-GAN: Asymmetric semi-supervised GAN for breast ultrasound image segmentation,” *Neurocomputing*, vol. 493, pp. 204–216, Jul. 2022, doi: 10.1016/j.neucom.2022.04.021.
- [66] I. U. Haq, H. Ali, H. Y. Wang, L. Cui, and J. Feng, “BTS-GAN: Computer-aided segmentation system for breast tumor using MRI and conditional adversarial networks,” *Eng. Sci. Technol. Int. J.*, vol. 36, p. 101154, Dec. 2022, doi: 10.1016/j.jestch.2022.101154.
- [67] J. Lee and R. M. Nishikawa, “Identifying Women With Mammographically- Occult Breast Cancer Leveraging GAN-Simulated Mammograms,” *IEEE Trans. Med. Imaging*, vol. 41, no. 1, pp. 225–236, Jan. 2022, doi: 10.1109/TMI.2021.3108949.
- [68] S. Guan, “Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks,” *J. Med. Imaging*, vol. 6, no. 03, p. 1, Mar. 2019, doi: 10.1117/1.JMI.6.3.031411.
- [69] O. N. Oyelade, A. E. Ezugwu, M. S. Almutairi, A. K. Saha, L. Abualigah, and H. Chiroma, “A generative adversarial network for synthetization of regions of interest based on digital mammograms,” *Sci. Rep.*, vol. 12, no. 1, p. 6166, Apr. 2022, doi: 10.1038/s41598-022-09929-9.
- [70] C. Zakka, G. Saheb, E. Najem, and G. Berjawi, “MammoGANesis: Controlled Generation of High-Resolution Mammograms for Radiology Education.” arXiv, Oct. 2020. Accessed: Mar. 18, 2023. [Online]. Available: <http://arxiv.org/abs/2010.05177>
- [71] E. Wu, K. Wu, D. Cox, and W. Lotter, “Conditional Infilling GANs for Data Augmentation in Mammogram Classification,” in *Image Analysis for Moving Organ, Breast, and Thoracic Images*, D. Stoyanov, Z. Taylor, B. Kainz, G. Maicas, R. R. Beichel, A. Martel, L. Maier-Hein, K. Bhatia, T. Vercauteren, O. Oktay, G. Carneiro, A. P. Bradley, J. Nascimento, H. Min, M. S. Brown, C. Jacobs, B. Lassen-Schmidt, K. Mori, J. Petersen, R. San José Estépar, A. Schmidt-Richberg, and C. Veiga, Eds., Cham: Springer International Publishing, 2018, pp. 98–106. doi: 10.1007/978-3-030-00946-5\_11.
- [72] C. Carr, F. Kitamura, J. Kalpathy-Cramer, J. Mongan, K. Andriole, M. Vazirabad, M. Riopel, R. Ball, and S. Dane, “RSNA Screening Mammography Breast Cancer Detection.” Kaggle, 2022.
- [73] H. M. L. Frazer, J. S. N. Tang, M. S. Elliott, K. M. Kunicki, B. Hill, R. Karthik, C. F. Kwok, C. A. Peña-Solorzano, Y. Chen, C. Wang, *et al.*, “ADMANI: Annotated Digital Mammograms and Associated Non-Image Datasets,” *Radiol. Artif. Intell.*, vol. 5, no. 2, p. e220072, Mar. 2023, doi: 10.1148/ryai.220072.
- [74] A. Cadrin-Chênevert, “Unleashing the Power of Deep Learning for Breast Cancer Detection through Open Mammography Datasets,” *Radiol. Artif. Intell.*, vol. 5, no. 2, p. e220294, Mar. 2023, doi: 10.1148/ryai.220294.
- [75] A. Sauer, K. Schwarz, and A. Geiger, “StyleGAN-XL: Scaling StyleGAN to Large Diverse Datasets.” arXiv, May 2022. Accessed: Feb. 28, 2023. [Online]. Available: <http://arxiv.org/abs/2202.00273>
- [76] V. Varkarakis, S. Bazrafkan, and P. Corcoran, “Re-training StyleGAN-A first step towards building large, scalable synthetic facial datasets,” *Pap. Present. 31st Ir. Signals Syst. Conf. ISSC Lett. Irel. 11-12 June, 2020*, doi: <http://dx.doi.org/10.1109/ISSC49989.2020.9180189>.
- [77] S. Targ, D. Almeida, and K. Lyman, “Resnet in Resnet: Generalizing Residual Architectures.” arXiv, Mar. 2016. Accessed: Feb. 28, 2023. [Online]. Available: <http://arxiv.org/abs/1603.08029>

- [78] M. Tan and Q. V. Le, “EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks.” arXiv, Sep. 2020. Accessed: Feb. 28, 2023. [Online]. Available: <http://arxiv.org/abs/1905.11946>
- [79] Z. Tu, H. Talebi, H. Zhang, F. Yang, P. Milanfar, A. Bovik, and Y. Li, “MaxViT: Multi-Axis Vision Transformer.” arXiv, Sep. 2022. Accessed: Feb. 28, 2023. [Online]. Available: <http://arxiv.org/abs/2204.01697>
- [80] J. Howard and S. Gugger, “Fastai: A Layered API for Deep Learning,” *Information*, vol. 11, no. 2, p. 108, Feb. 2020, doi: 10.3390/info11020108.
- [81] J. Howard and S. Gugger, “Deep Learning for Coders with Fastai and Pytorch: AI Applications Without a PhD.” O’Reilly Media, Incorporated, 2020.
- [82] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization,” *Int. J. Comput. Vis.*, vol. 128, no. 2, pp. 336–359, Feb. 2020, doi: 10.1007/s11263-019-01228-7.
- [83] R. Wightman, N. Raw, A. Soare, A. Arora, C. Ha, C. Reich, F. Guan, J. Kaczmazzyk, MrT23, Mike, *et al.*, “Pytorch Image Models.” Zenodo, Feb. 2023. doi: 10.5281/ZENODO.4414861.
- [84] A. Sriram, M. Muckley, K. Sinha, F. Shamout, J. Pineau, K. J. Geras, L. Azour, Y. Aphinyanaphongs, N. Yakubova, and W. Moore, “COVID-19 Prognosis via Self-Supervised Representation Learning and Multi-Image Prediction.” arXiv, Jan. 2021. Accessed: Mar. 18, 2023. [Online]. Available: <http://arxiv.org/abs/2101.04909>