University POLITEHNICA of Bucharest

Faculty of Automatic Control and Computers
Computer Science and Engineering Department

PhD THESIS
Summary

Modern Applications of Document and Medical Image Analysis

Scientific Adviser:
Prof. PhD. Eng. Costin-Anton Boiangiu

Author:
Eng. Marcel Prodan

Bucharest, 2023

Contents

Contents ... ii
List of Abbreviations ... 1
1 Introduction ... 3
 1.1 Thesis Structure ... 7
2 Related Work ... 9
 2.1 Overview of Document Image Analysis Systems ... 9
 2.2 Overview of Breast Cancer Detection ... 9
 2.3 Automatic Document Image Analysis and Processing ... 10
 2.4 Metrics for Image Resemblance ... 11
3 Image analysis by intelligently combining the results of several algorithms ... 12
 3.1 Document Image Binarization Process ... 12
 3.2 Segmentation ... 15
 3.3 OCR ... 16
 3.4 Unified Image Processing Architecture: Document Analysis, Medical Imaging, and Traffic Sign Recognition ... 17
4 Case Study: Breast Cancer Detection ... 19
 4.1 Results ... 20
5 Conclusions ... 23
 5.1 Contributions ... 24
 5.2 List of publications ... 26
  5.2.1 Journals ... 26
  5.2.2 Conferences ... 27
  5.2.3 Posters ... 27
Bibliography ... 28

List of Abbreviations

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 Introduction

Man, and nature have always had a deep and complex relationship, a symbiosis that has had a major influence on our development as a species. Understanding this relationship can give us greater insight into how we interact with the world around us and the influence we have on it.

The intricate relationship between humans and nature is integral to our evolution as a species, spanning across all our senses. Yet, vision particularly stands out, with roughly 33% of the brain dedicated to this function. Our brain's processing of reality through sensory organs essentially translates external signals into an electrochemical language. This indicates our brain creates our perception of reality based on internal models formed from past experiences, without ever interacting directly with the external world. Various experiments have showcased the brain's adaptive capabilities. For instance, it adjusts to prismatic goggles which flip the visual perspective, emphasizing the brain's flexibility in processing multiple images of the same scene. Additionally, the synchronization of signals from different senses plays a pivotal role, as shown by reaction-time differences in athletes when signaled by sound versus light. The absence of external stimuli, like in extreme isolation cases at Alcatraz, results in the brain generating its reality. This reveals the profound role of the brain's internal model, which gets enriched with every new experience, reducing the need for continuous external input and allowing for efficient processing. [1]

By comparing the functioning of the human brain with the way images are analyzed using computational techniques, we can draw the following insights:

- the use, if possible, of multiple image sets for a better evaluation of their properties and obtaining more accurate results.
- the importance of synchronizing captured images with other associated quantities, such as time, distances, angles, altitude, speed, or other characteristics of the environment.
- the importance of using filters to make the image analysis process more efficient.
- the importance of using the knowledge base wherever it exists and can be used.
- the importance of training artificial intelligence models.

Background of Document Image Analysis Systems

With the rise of the digital era, the transition from traditional media, such as cassettes and CDs, to digital formats has been pivotal. The digitization of physical documents addresses issues related to storage, accessibility, and environmental sustainability while preserving historic records for future use. Document image analysis (DIA) systems, especially Optical Character Recognition (OCR), are central in this digitization process, enabling the conversion of varied documents, from books to maps, into digital formats.

History of Optical Character Recognition

The desire to create equipment capable of recognizing writing, whether handwritten, typewritten, or printed, dates back to the 19th century. Several early attempts at OCR were made through patent applications in the 1920s and 1930s. However, technological limitations at the time hindered the realization of practical OCR systems.

Significant advancements in character recognition systems were made during the 1950s, driven by the financial sector's need for automatic processing of checks, the most common printed documents at the time. The first commercial OCR machine was purchased by Reader's Digest in 1955, leading to further development and adoption of OCR systems in various applications, including document processing, mail distribution, and invoice recognition.

Alternative Technologies

In addition to OCR, several alternative technologies have been developed to cater to specific document processing needs. These include Intelligent Character Recognition (ICR) and Intelligent Word Recognition (IWR) systems for cursive writing, Optical Mark Recognition (OMR) systems for managing form structures, and Magnetic Ink Character Recognition (MICR) used in banks and financial institutions.

Optical Character Recognition

OCR technology plays a crucial role in digitizing books and manuscripts, preserving their cultural and historical significance. The performance of OCR is evaluated based on accuracy and confidence in the results. While OCR has made significant progress, human intervention is essential for handling complex cases and achieving higher accuracy levels.

The OCR process involves several stages, including preprocessing, image enhancement, segmentation, recognition, and classification. Each step contributes to the overall accuracy and quality of the digitized content.

Methodologies Applicable to Document Image Analysis Systems

Document image analysis systems are typically designed to address specific tasks, considering the complexity of image analysis at different levels. Tailored applications are used to ensure effective document processing, considering factors such as document structure, language support, and paper quality. User feedback and human intervention are common in such systems, especially for handling damaged or unique documents.

Present-day Retro-conversion Techniques

The increasing importance of digitizing collections of books, newspapers, and historical records has led to the emergence of various retro-conversion software and initiatives. OCRopus, METAe, and Lib2Life are examples of such systems aimed at preserving and digitizing valuable physical documents. The future of document image analysis systems is focused on achieving greater accuracy, flexibility, and efficiency while reducing user interaction.

Document Image Binarization Process

Document Image Binarization is vital for digitizing and preserving historical documents. In the digital era, there's a growing need to preserve historical texts. This thesis introduces a binarization technique using the Fully Convolutional Network (FCN) for efficient and accurate digitization. Binarization converts documents into a binary format, distinguishing important pixels from irrelevant ones, providing better legibility and efficiency than older methods like scanning. The primary aim is to preserve historical writings, regardless of their condition, using an FCN for optimal results. We analyze existing binarization methods and introduce an innovative FCN-based architecture. The system uses the DIBCO datasets for training and evaluation, maintaining simplicity and delivering strong performance. Our model combines advanced techniques using the Keras library, aiming for improvements in document preservation, text recognition, and medical imaging. By optimizing this model, we foresee advancements in image analysis and potential applications in diverse areas.

Segmentation

The thesis introduction delves into image segmentation, a vital component of computer vision, with applications spanning from autonomous driving to data extraction. The research's primary aim is to enhance segmentation results by amalgamating various algorithms through voting techniques. Though indispensable, segmentation algorithms can face issues like over- or under-segmentation. To address these, the study proposes a model with convolution, deconvolution, and concatenation operations. This model is envisioned to pioneer advancements in tasks like document binarization, color segmentation, OCR, and cancer detection. The thesis evaluates three key segmentation methods: Felzenszwalb's algorithm, SLIC Superpixels, and Watershed, discussing their merits and demerits. A novel voting technique combines these algorithms' results to optimize segmentation accuracy, exploring both democratic and performance-based weighted voting. The effectiveness of this approach is gauged against individual algorithm outputs, and visual examples emphasize the contrasts and benefits. This study paves the way for innovative segmentation methods suitable for a range of applications.

OCR

Optical Character Recognition (OCR) has become a pivotal tool in Computer Vision for its capability to transform image text into readable formats. Yet, issues like image noise, inconsistent illumination, and document wear can hinder OCR accuracy. To mitigate these challenges, this study introduces a novel voting-based OCR method. This approach amalgamates the results from multiple OCR techniques using a voting process, aiming to pinpoint the most trustworthy results. An initial preprocessing step employs various filters on the input image, enhancing textual features. The image then undergoes several OCR runs, with the most confident text outcome chosen based on word-level confidence ratings. Tests have shown that this method considerably enhances OCR precision, though it demands more processing time. The strategy's versatility suggests broader applications in unsupervised image processing and data extraction. As part of the Lib2Life project, the method is anticipated to be instrumental in enhancing the digital conversion of Romanian library documents. In essence, this pioneering voting-based OCR approach offers promising advancements in extracting accurate data from challenging image environments, further propelling the digital age.

Breast Cancer Detection

Breast cancer remains a dominant health concern, necessitating early detection for effective treatment. Though mammography has been a cornerstone in breast cancer screening, its manual analysis poses challenges in terms of time and subjectivity. This research delves into deep learning solutions, particularly employing Convolutional Neural Networks (CNN) and Vision Transformers (ViT) to analyze mammograms. A distinct method using synthetic image generation through machine learning enhances the training dataset, leading to better classification accuracy. The role of data preprocessing, augmentation, and Explainable AI techniques, like class activation maps, are emphasized for model transparency and understanding. However, obstacles such as dataset limitations and model interpretability remain. This study underscores the potential of integrating AI in breast cancer screening, aiming for early detection, enhanced treatment success, and improved patient well-being.

1.1 Thesis Structure

The Thesis Structure chapter serves as a guide to the organization and content of this PhD thesis, outlining the key chapters and their respective contributions. The following sections provide a concise overview of each chapter:

Chapter 1 lays the foundation for the research by introducing the context and objectives of the thesis. It highlights the significance of image processing and its relevance in various technological scenarios. Additionally, it outlines the research questions and methodologies employed throughout the study.

Chapter 2 provides a comprehensive background and overview of the thesis topic. It covers the existing research in document image analysis systems, metrics for image resemblance, and breast cancer detection, as well as automatic document image processing. The chapter establishes the foundation for the research by exploring the state-of-the-art knowledge in these relevant areas.

Chapter 3 presents the concept of intelligent image analysis and fusion. It includes the discussion of image analysis by combining the results of several algorithms. The chapter explore the document image binarization process, segmentation techniques, and Optical Character Recognition. Additionally, it introduces a unified image processing architecture that integrates document analysis, medical imaging, and traffic sign recognition. This chapter aims to develop a comprehensive and efficient approach to image analysis by leveraging various techniques and their fusion.

Chapter 4 focuses on a case study related to breast cancer detection. It includes a review of medical imaging techniques commonly used for breast cancer detection. The chapter then presents the materials and methods employed in the case study and the results obtained from the study. It also discusses the implications and significance of the findings in the discussion section and summarizes the key takeaways from the case study, shedding light on the potential applications and advancements in breast cancer screening.

Chapter 5 is the concluding chapter that provides a comprehensive summary of the entire thesis. It reiterates the main findings and contributions of the research, as well as the significance of the results obtained. This chapter also discusses the implications of the research and its potential impact on the respective fields of document image analysis and breast cancer detection. It concludes with suggestions for future research directions and potential applications of the findings. Additionally, the chapter includes a list of publications stemming from the research conducted during the Ph.D. program, acknowledging the scholarly contributions made during the research journey.

2 Related Work

2.1 Overview of Document Image Analysis Systems

This chapter is based on the article [2] to which we contributed as a co-author.

Keeping information safe is the vital reason why we want to be able to store and make it available in digital format, given that we live in a digital age. One way to make physical data available to the digital world is by scanning and storing physical documents in a digital format. This process solves many problems such as storage space, access, and environmental issues. In addition, digital documents do not deteriorate, which allows the preservation of many significant historical documents. Translating physical documents into digital format seems to be more difficult than it seems, considering the many problems that can arise such as damaged paper, color changes, font problems and so on. The goal is to extract the relevant documents. data. The goal of "Document Image Analysis" (DIA) systems is to process documents in a human-like manner, recognizing text, images, and extracting information. Regarding document image analysis systems [3], a technology called optical character recognition (OCR) [4] is the most important. This is because it has long been used in the United States postal system and because users frequently use OCR even when working with files that do not contain text. However, this method is only a small part of the wider applications and creates a stereotypical image of what a DIA really means, as it is used to recognize characters and convert printed images into automatically encoded text. The ability to recognize and digitize a wide range of documents (books, letters, drawings and maps) is one of the advantages of using document image analysis. Document image analysis involves a variety of steps and processing at different levels of granularity and is quite complex. In the following chapters, we will present the evolution of OCR and document image analysis systems, along with the different approaches and technologies used for these purposes.

2.2 Overview of Breast Cancer Detection

Breast cancer remains a severe health concern due to its prevalence and the need for early detection to ensure successful treatment [5]. Mammography, a widely recognized screening method for breast cancer, has some inherent challenges, particularly in analysis, which can be both time-consuming and subjective. This study delves into the utilization of deep learning techniques to augment mammogram analysis, leveraging models such as CNN and ViT on a public dataset. A novel method of data augmentation using synthetic image generation further boosts the model's efficiency. The value of data preprocessing and augmentation in attaining top-notch classification is emphasized, while Explainable AI methods, including class activation maps and centered bounding boxes, shed light on the models' decision-making [6].

Global statistics underscore the gravity of breast cancer, marking it as the most frequently diagnosed cancer type. It accounted for over 2.3 million diagnoses and led to more than 685,000 deaths in 2020 alone [5]. The female population remains particularly at risk [7]. In the face of such challenges, machine learning (ML) and its advanced subset, deep learning (DL), offer promise. DL's foundation on artificial neural networks earmarks it as a beacon in image recognition and classification tasks, representing ML's cutting-edge.

The salience of medical imaging in early breast cancer detection cannot be overemphasized. Through screening initiatives and advanced imaging tools, early detection is possible, leading to optimal treatment plans and heightened survival chances. The paper's novelty lies in its dual-fold contribution: offering a discerning review of breast cancer medical imaging methods and introducing an empirical study that employs deep learning techniques on mammograms for cancer detection. Central to this approach is the utilization of synthetic images, crafted through machine learning algorithms by drawing inspiration from an expansive dataset of authentic mammograms. This aids in the superior accuracy of breast cancer classification.

The subsequent sections will provide an overview of prevalent medical methodologies and the burgeoning role of Artificial Intelligence in this domain.

2.3 Automatic Document Image Analysis and Processing

Automatic document image processing is pivotal in today's technological landscape, turning vast volumes of digital and scanned documents into actionable digital data. Its versatility spans across multiple sectors, enhancing operational efficiency. Here are its key applications: document processing, which streamlines scanning, recognition, and data extraction from various printed materials; medical imaging for aiding in the analysis and diagnosis of medical images; video surveillance that enhances security through real-time activity recognition and alerts; autonomous vehicles for obstacle detection and safe navigation; quality analysis in the industry to ensure product quality and defect detection; military functions to assist in reconnaissance, surveillance, and tactical evaluations; aeronautics for automated flight control and navigation; naval operations to boost ship surveillance and navigation; and aerospace to assist in space mission control and data analysis from space explorations..

The evolution and advancements in deep learning and neural networks, especially CNN, have amplified the utility and potential of automatic document image processing [8].

2.4 Metrics for Image Resemblance

This chapter is based on the article [9] to which we contributed as first author.

This study delves into the metrics used to assess image resemblance in computer science, particularly in domains such as digital image processing, computer vision, machine learning, and deep learning images [10]–[14]. Key metrics under investigation include traditional pixel-based measures like Mean Absolute Error (MAE) and Mean Squared Error (MSE), structural similarity metrics such as Structural Similarity Index (SSIM) and MultiScale Structural Similarity Index (MS-SSIM), and advanced measures like Feature Similarity Index (FSIM), Universal Quality Index (UQI), and Visual Information Fidelity (VIF) images [10], [15]–[19]. These metrics are foundational for comparing images, guiding algorithm development, and understanding human visual perception. Moreover, they're pivotal in evaluating the efficacy of algorithms in image restoration, compression, enhancement, and tasks in deep learning like object recognition and image generation with Generative Adversarial Networks (GANs) [13], [15], [18], [20]. The evolution of these metrics over time reflects an ongoing pursuit to closely align computational methods with human visual perception, ensuring both computational efficiency and perceptual relevance [10], [11], [15], [17]–[21].

3 Image analysis by intelligently combining the results of several algorithms

3.1 Document Image Binarization Process

This chapter is based on article [22] to which we contributed as first author.

In the face of challenges in preserving historical documents due to degeneration over time, this study recommends a new technological approach leveraging artificial intelligence for digitization. Rather than traditional methods like scanning, the work suggests binarization as a superior technique. Binarization, which differentiates useful pixels from those causing visual imperfections, effectively captures the essential information of a document in a digital format, offering better legibility and efficient memory usage. This research aims to provide a comprehensive overview of existing binarization methods and introduces a novel solution rooted in fully convolutional networks. The proposed network, simpler in design yet comparable in performance to existing semantic segmentation methods, is trained using the frequently referenced DIBCO datasets [12], [13]. The goal is to ensure historical documents, rich in cultural and traditional significance, are preserved with high fidelity for future generations.

The study introduces a document binarization method using a Fully Convolutional Network (FCN) inspired by Long et al. [23] and Tensmeyer & Martinez [24]. The network's architecture splits into four branches, each differentiating in depth and spatial dimensions. These branches extract both local features, such as basic shapes, and global, semantically rich features like elongated letters. Transitions between branches use Max Pooling, and convolution operations involve filters that double in number as one moves downward through the branches. The last three branches undergo additional processing to match branch 1's feature maps and resolution using deconvolution, an approach influenced by Long et al. [23]. After ensuring uniformity across branches in terms of dimensions, outputs are concatenated, followed by a convolution layer, and finally, predictions are made using the Sigmoid activation function ideal for binary outcomes. The chosen cost function is Binary Cross-Entropy, apt for binary classifications.

The model was trained on 128×128-pixel images sourced from DIBCO datasets from 2009 to 2019 [25], comprising both handwritten and machine-printed documents with varying degradation levels. These images were converted to black and white, then normalized. The training spanned 300 epochs, with the mean intersection over union (mIoU) being the primary metric for evaluation due to its capability to account for class imbalances.

For document binarization, initial processing resized documents to allow division into 128×128 segments, ensuring the best resolution based on the model's training. Subsequent predictions assigned pixel values based on a set threshold, with the final binarized document resembling the original's dimensions.

The model was trained over 300 epochs with a learning rate of 0.1, using accuracy, loss, and mIoU as performance metrics. Notably, there was significant improvement in accuracy and loss, indicating the model's proficiency in predicting the predominant background pixel class of the dataset. However, due to the overwhelming presence of background pixels, incorrect predictions of the foreground pixels had minimal impact on these metrics. To get a holistic understanding of the model's performance, it's essential to consider metrics that evaluate both overall dataset accuracy and class-specific accuracy. mIoU, which is sensitive to the foreground pixel predictions, exhibited a slower growth rate. It began converging towards a value of 0.6 after approximately 250 epochs for both training and validation sets.

Model effectiveness on the test set of data

Using the test data set, the model was able to achieve the following results:

mIoU	**accuracy**	**loss**	**precision**	**recall**	**F-measure**
0.589	0.9823	0.0931	0.916	0.883	0.899

In the table below, we emphasized the individual scores for each architecture. In terms of the metrics employed in the two works, FM and mIoU, we were able to compare the suggested answer to the alternative solutions. As can be seen, the scores from the model's own analysis are comparable to those from Long et al. [23] and Tensmeyer & Martinez [24].

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

Compared performances between architectures.

Given that there are variations in the test data sets and the method of segmentation used, the proposed solution's placement with respect to the two cannot be perfect. In order to rate the three models fairly, a proper comparison would make use of standard test data sets and metrics.

Model	**GPU**	**Framework**
Tensmeyer	NVIDIA Tesla K40c	Caffe
Long	NVIDIA Tesla K40c	Caffe
Proposed solution	NVIDIA GTX 1050 Ti	Keras

Video cards used in the three solutions.

The table above shows the video cards used in the three experiments. We used Nvidia GTX 1050Ti for the proposed solution and the same model, Nvidia Tesla K40c, was used in the other two.

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

Specifications comparison: Tesla K40c vs GTX 1050Ti
Source: [26]

The table above compares the specifications of the two types of video cards. Note the differences between the NVIDIA Tesla K40c model and the NVIDIA GTX 1050Ti: RAM memory (12 GB versus 4 GB), number of cores (2880 versus 788), launch price ($7699 versus $139) and current price ($360 versus $184).

The conclusion is that using limited resources in the proposed solution, results were obtained very close to those obtained in the other two experiments in which the resources were superior.

3.2 Segmentation

This chapter is based on the article [27] to which we contributed as a co-author.

This study focuses on improving image segmentation, a key aspect of computer vision used in areas like autonomous driving and data extraction. The proposed strategy combines the results of various algorithms using different voting methods to enhance segmentation accuracy [27]. 'Image segmentation' encompasses algorithms that aim to categorize and order pixels according to specific benchmarks. Hoover et al. [28] demarcated these challenges into categories such as noise, missed detections, over-segmentation, and under-segmentation. Interestingly, under-segmentation often overshadows over-segmentation because it's simpler to refine larger segments than to fragment extensive regions [29].

In a bid to outperform the singular outcomes of prevalent segmentation algorithms, this study unveils a voting methodology that amalgamates several algorithms. The algorithms shortlisted for this endeavor include "Felzenszwalb segmentation" [30], "SLIC Superpixels" [31], [32] and "Watershed" [33], [34], primarily due to their propensity to render distinct outputs that are either over- or under-segmented.

Elucidating the Felzenszwalb graph-based algorithm [70], every pixel is visualized as a node in the network during segmentation. The interconnections between nodes typify specific pixel pairings. Each of these edges comes with an associated weight, reflecting pixel cluster characteristics. Felzenszwalb's technique [30] exhibits a unique versatility, such as its ability to fine-tune segmentation benchmarks in alignment with local nuances.

SLIC Superpixels, it introduces a novel perspective by leveraging superpixels. By clustering alike pixels into cohesive pixel sections, this method provides an innovative lens to the conventional pixel grid. Achanta et al. [31] highlighted that this technique not only simplifies complexities but also employs primitives that craft the image's essential components.

The third algorithm, Watershed, envisions the image as a topographical panorama, taking cues from the original concept by Digabel and Lantuejoul [33]. With this lens, pixel gray values or gradient intensities dictate the landscape's contours. The main objective of the watershed method is to demarcate the image into varied focal zones by discerning pixel intensities.

Discussing voting methodologies, numerous approaches exist for image segmentation, each with its pros and cons vis-à-vis input data. The idea of combining decisions from various methods, often termed voting techniques, has gained traction in recent studies.

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

In terms of results, given the absence of a universal standard for image segmentation, it's pivotal to approach this operation with an "optimal" mindset for a particular context. This requires balancing the desired outcome's segmentation level with resource allocations, such as computational power and time. Therefore, segmentation optimization should weigh both the requisite quality and corresponding costs.

3.3 OCR

This chapter is based on the article [40] to which we contributed as a co-author

This section introduces a new OCR technique that combines results from various methods. By understanding each OCR approach, we can decide between outcomes. The method uses a voting system and preprocesses the document image to highlight text features, aiming to achieve the most accurate OCR reading. The results show this approach performs consistently well across scenarios, making it suitable for unsupervised document processing and data extraction [40].

The proposed method is a voting-based system that applies various filters to enhance visual qualities in images. These filters adjust contrast, sharpening, and perform morphological actions like dilation and erosion using different kernels. Once the filters are applied, the image undergoes OCR processing. Each word's confidence level is then compared, and the most confident result is retained. This process is repeated on the original image for more refined results. Finally, the outcomes of multiple OCR attempts are compared and integrated using an appropriate voting mechanism.

To improve the contrast between elements and guarantee the retrieval of unambiguous related components, image binarization [41] is used.

The study introduces a voting-based OCR method that applies various preprocessing techniques to images to enhance text characteristics. Instead of using the common per-character confidence, the system uses word-based confidence, calculated from individual letters, for improved accuracy. After processing, words with the highest confidence are selected. Tesseract, a popular OCR engine, often struggles with image issues like noise, poor lighting, or inconsistent contrast. By conducting multiple preprocessing steps, certain image areas are improved, allowing for more reliable OCR results. The aim is to gather the best outcomes from each run. Each preprocessing action can increase detection accuracy for some words but might decrease it for others. Despite its time-consuming nature, this method effectively minimizes incorrect word detection.

The strategy involves activating different visual attributes in the input image using filters. These processes adjust image factors like contrast and apply morphological actions. Binarization is used to further enhance contrast. The results demonstrate the method's ability to boost OCR reading across different scenarios, ideal for unsupervised document processing.

Historically, OCR technology has shown the importance of text extraction from images. Tesseract has seen notable improvements, but still faces challenges due to image quality. The proposed method's essence lies in its voting-based technique, where it harnesses different image filters and amalgamates the outcomes to optimize the final result.

The approach's effectiveness in improving text detection, though time-intensive, is clear, making it a promising tool for large-scale digitization projects, like the Lib2Life initiative aiming for comprehensive digitization of Romanian library documents.

3.4 Unified Image Processing Architecture: Document Analysis, Medical Imaging, and Traffic Sign Recognition

This chapter delves into a unified image processing architecture that integrates three specific domains: document analysis, medical imaging, and traffic sign recognition, building on the foundational processes discussed in chapters 3.1 to 3.3. The innovative architecture aims to cohesively and efficiently handle image data across these categories. By examining its design and function, we seek to understand the benefits of such an integrated approach, its impact on system performance, and its potential in diverse practical applications. The goal is to underscore the advantages of combining various image processing techniques within these domains.

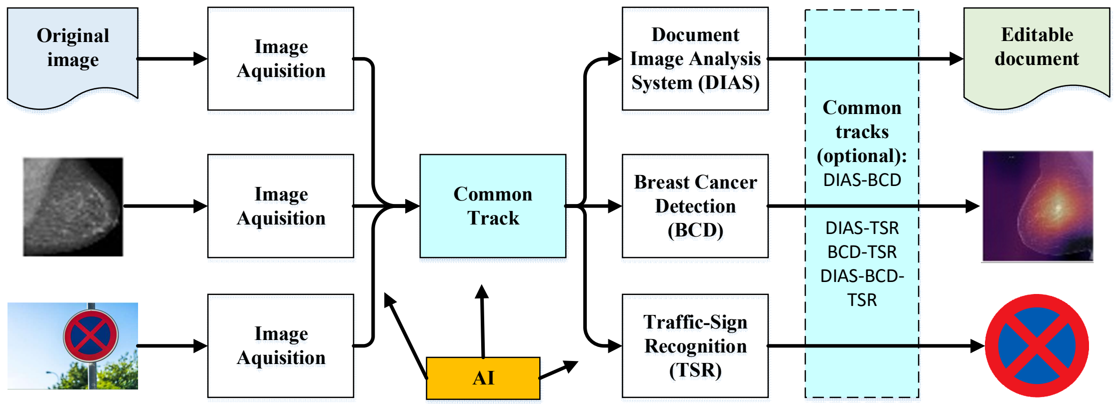


Figure 3-1 - Unified Image Processing Architecture

In the future, we could see a device that, with the right peripherals, can handle image processing for various domains. This envisioned architecture is modular, meaning operations can be added as needed. There are modules common to all data flows, and images are acquired through specialized equipment. Post-processing can also use these shared modules, and AI models can be integrated at various stages of the process.

Image processing systems, crucial across multiple domains, consist of common pre-processing, post-processing, and domain-specific AI modules. Pre-processing involves standardizing images, segmenting them into relevant parts, extracting features, applying advanced transformations, and reducing noise. Post-processing improves image clarity and extracts valuable information. AI modules, tailored to domains like documents, medical images, or traffic signs, address unique challenges through data pre-processing, object recognition, semantic segmentation, and machine learning techniques. While such integrated systems offer efficiency, cost benefits, and adaptability, they might also face challenges in complexity, specificity, and security vulnerabilities.

4 Case Study: Breast Cancer Detection

This chapter is based on the article [6] to which we contributed as first author.

This study focuses on improving breast cancer detection via mammography using deep learning methods, notably CNN and ViT architectures. By incorporating synthetic image augmentation, the models' performance is significantly enhanced. The findings stress the importance of data pre-processing and augmentation for optimal classification. Additionally, the research utilizes explainable AI techniques to offer a clearer understanding of how the models make decisions. [6].

Breast cancer (BC), the most common type of cancer worldwide, resulted in over 2.3 million cases and more than 685,000 fatalities in 2020 [5]. It has a profound impact on the global number of cancer-related deaths, particularly among women [6].

The contribution of this paper to the literature is twofold: (a) To provide a critical review of medical imagining techniques for BC detection, as well as the deep learning methods for BC detection and classification; (b) To present the experimental study proposed for the detection of BC based on several deep learning techniques on mammograms.

The innovative approach is represented by the generation of synthetic images which involves augmenting the dataset with realistic digital breast images using ML algorithms that can mimic the appearance of actual mammograms.

Medical imaging is crucial for early breast cancer (BC) detection, with increasing emphasis on digital technologies. The primary methods for identifying BC revolve around detecting calcifications, which are calcium deposits in the breast tissue. [42]. We analyze below the most common types of medical imaging techniques for BC diagnosis such as mammography, ultrasound, MRI, thermography and histopathology, as well as deep learning methods for BC detection and classification such as transfer learning and data augmentation, feature extraction and multiple model-based architectures and Generative Adversarial Networks.

In this study, we aim to experiment with deep learning-based techniques for detecting cancer in mammograms. For this, we use the dataset provided by the Radiological Society of North America (RSNA) in a Kaggle competition [72]. Its aim was to detect instances of breast cancer in mammograms obtained during screening examinations.

The dataset used in this competition was based on the ADMANI dataset [73], which contains annotated digital mammograms and non-image data, and is possibly the most extensive and diverse mammography dataset documented in the literature. The dataset incorporates a total of 28,911 instances of breast cancer, out of which 22,270 were detected during screening and 6641 were interval cases, significantly more than any other dataset published so far. The dataset also consists of a large number of examinations (1,048,345) and patients (629,863), making it one of the largest datasets reported to date [74]. However, the training set for this public competition consists of 11,913 examinations, out of which 486 are cancer cases, making it a very unbalanced dataset.

We experimented with several deep learning models based on CNN and ViT architectures, such as ResNet [77] or EfficientNet [78], and MaxViT [79].

XAI can also help to identify biases or errors in the AI model's decision-making process. For example, it can help identify features in the image that are disproportionately influencing the decision, or highlight cases where the model may be making incorrect or biased decisions.

Fastai [80] provides a Grad-CAM (gradient-weighted class activation mapping) solution [81] for visualizing the regions of an input image that are most important for a model's prediction. Grad-CAM [82] is a popular technique for generating heatmaps that highlight the regions of an image that are most important for a particular class, and has been used in a variety of computer vision applications.

The class activation map uses the output of the last convolutional layer together with the weights of the fully connected layer that corresponds to the predicted class. It calculates their dot product so that for each position in the feature map it is possible to get the score of the feature used for the prediction.

4.1 Results

For experimentation, the images from the dataset were first resized to a uniform size of 256×256 and 512×512 pixels. The performance of the models was evaluated using accuracy, AUC, precision, recall, and the F1 score.

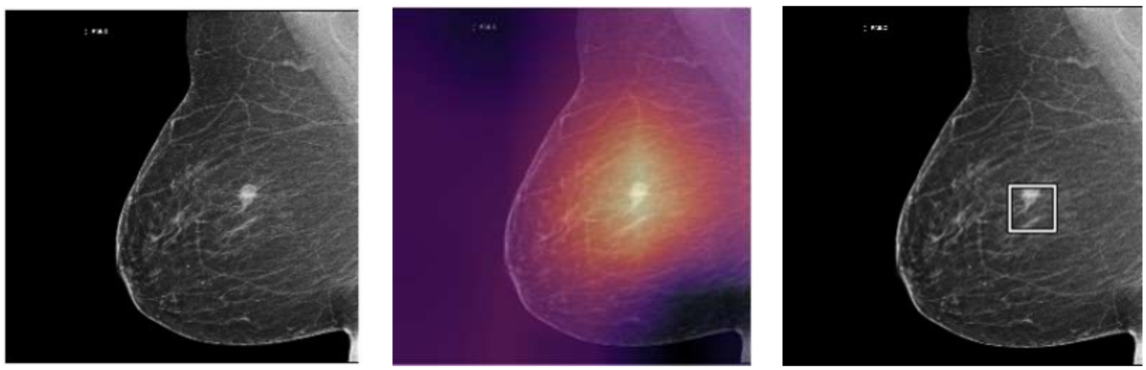
Mammograms with 256 × 256 resolution					
Model	**Accuracy**	**AUC**	**Precision**	**Recall**	**F1**
Original images					
ResNet18	**0.96**	0.61	0.06	0.04	0.04
ResNet34	**0.96**	0.63	0.08	0.06	0.06

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

Table 4-1 - Experimental results

The performance of relatively simple CNN models, such as ResNet18, was almost as good as larger models such as ResNet152 or EfficientNet B0. Furthermore, the performance of the CNN models was found to be comparable to that of the model based on visual transformer architecture. However, what made the most significant difference was the pre-processing of the dataset, specifically the augmentation of the dataset using synthetically generated images. This pre-processing technique helped to improve the overall performance of the models, resulting in better evaluation metrics.

The study indicates that meticulous pre-processing and data augmentation can significantly boost even basic models' performance. This emphasizes the need for optimal pre-processing and the value of synthetic data in enhancing computer vision models. Post-classification, visualization techniques identified the image areas influencing the model's decision, with a centered bounding box highlighting the key region used for its determination.



(a) **(b)** **(c)**

Explaining the results. (a) The original image. (b) The application of class activation maps. (c) The application of a bounding box

These visualization techniques provided a useful way of understanding how the deep learning model classified images as positive or negative. They helped to identify the specific areas of the image that the model used to make its decision and provided a more intuitive way of interpreting the model’s output.

5 Conclusions

Chapter 2

Document Image Analysis Systems are essential for digitizing historical texts, aiming for autonomous processing to boost speed and precision. This paves the way for a digital shift in managing vast literary collections, enhancing access and structure.

Breast Cancer Detection underscores early diagnosis and AI's role in enhancing mammogram assessments. With various imaging techniques, AI stands out for its transformative detection capabilities. Despite challenges like limited datasets, innovations like synthetic images and GANs show potential in cancer classification, underscoring AI's vital role in patient care.

Automatic Document Image Processing, given the digital age's vast document volume, explores its development and cutting-edge extraction methods. Advancements in AI, like deep neural networks, have catalyzed breakthroughs in diverse fields like healthcare and defense. With efficient architectures, these technologies convert images into valuable digital data, promising enhanced data processing accuracy across industries. This chapter offers a comprehensive overview of the significance and future of document image processing..

Chapter 3

Document Image Binarization Process: This study introduces a promising solution for efficient image binarization, using Python and available libraries. While effective, there's a compromise in visual quality, especially where document defects are highlighted. Increasing convolutional layers and utilizing a larger dataset could improve results.

Segmentation: This chapter delves into the dynamics of voting in image segmentation algorithms, emphasizing unsupervised algorithms. Though Deep Learning is promising, the current focus is on how voting impacts various features in segmentation results. The end-goal is to combine this method with other unsupervised systems for an efficient document image processing system.

OCR: Introducing a novel voting-based system, this section discusses improving OCR accuracy by applying multiple preprocessing operations on image documents. Even though it may require more processing time, its robustness makes it ideal for large-scale digitization projects like the Lib2Life project for Romanian libraries.

Unified Image Processing Architecture: This advanced architecture integrates document analysis, medical imaging, and traffic sign recognition. Using common preprocessing modules and AI techniques, it aims for efficiency across multiple domains. While offering multiple advantages, it also has complexities and challenges to overcome. The envisioned future device would be a multipurpose tool for all three domains, marking a significant step forward in image processing technology.

Chapter 4

Chapter 4 presents a comprehensive case study on Breast Cancer Detection, highlighting the significance of early detection for successful treatment, especially considering breast cancer's prevalence among women. The study explores the application of deep learning techniques to improve the accuracy and efficiency of breast cancer screening workflows using various imaging methods, such as mammography, thermography, ultrasound, and magnetic resonance imaging.

The introduction of AI technology has shown great promise in automating breast cancer detection, with AI algorithms demonstrating comparable performance to human experts in retrospective data sets. This emerging field of AI-powered breast cancer screening holds significant potential for improving early detection and prognosis, ultimately leading to better patient outcomes and a reduction in breast cancer-related deaths.

The main focus of the study was on the use of deep learning methods for analyzing mammograms, and the results underscored the importance of proper data preprocessing and augmentation techniques. By leveraging synthetic data generation techniques, the classification performance of deep learning models was significantly enhanced, demonstrating the potential of these methods in boosting the accuracy of image classification tasks.

In summary, the study's findings highlight the transformative impact of AI-powered deep learning techniques in breast cancer detection, showing comparable performance to human experts and the potential for even greater accuracy with proper data preprocessing and synthetic data generation. The integration of AI into breast cancer screening holds promise for early detection, improved prognosis, and ultimately contributing to better patient outcomes in the fight against breast cancer.

5.1 Contributions

One of the primary contributions of this study is the comprehensive evaluation of image processing’s applicability in various technological settings. By examining the capabilities and limitations of different image processing, we have provided valuable insights into their suitability for specific applications. This analysis serves as a foundation for informed decision-making and optimization of image processing tasks, paving the way for more efficient and effective processing solutions.

A major contribution of this research lies in highlighting the pivotal role of artificial intelligence (AI) in modern image processing systems. By presenting a thorough analysis of AI-driven image processing techniques, we have emphasized their transformative impact on image analysis, interpretation, and feature extraction. The integration of AI algorithms empowers image processing with advanced cognitive abilities, elevating their capacity to handle complex tasks and extract meaningful information from intricate images.

5.2 List of publications

5.2.1 Journals

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
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.

5.2.2 Conferences

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 Posters

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

Bibliography

[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. Rumgay, 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
[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 basedon 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, 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,” *IEEETransactions 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 PatternRecognition (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 InternationalConference on Pattern Recognition (ICPR2012*, Nov. 2012, pp. 930-934,.

[37] D. Stutz, A. Hermans, and B. Leibe, “Superpixels: An evaluation of the state-of-theart 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. Ioanitescu, “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. Brotchie, “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.compbiomed.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. Kaczmarzyk, 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