D 2024

U.PORTO
FEUP FACULDADE DE ENGENHARIA
UNIVERSIDADE DO PORTO

MULTIMODAL CERVICAL CANCER DIAGNOSIS:
DEEP LEARNING FOR AUTOMATIC DECISION SUPPORT

TOMÉ MENDES ALBUQUERQUE

DOCTORAL THESIS PRESENTED TO
FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO IN
BIOMEDICAL ENGINEERING

FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

Multimodal Cervical Cancer Diagnosis: Deep Learning for Automatic Decision Support

Tomé Mendes Albuquerque



Doctoral Program in Biomedical Engineering

Supervisor: Jaime dos Santos Cardoso, PhD

Co-Supervisor: Maria João Medeiros de Vasconcelos, PhD

March, 2024

© Tomé Mendes Albuquerque, 2024

Multimodal Cervical Cancer Diagnosis: Deep Learning for Automatic Decision Support

Tomé Mendes Albuquerque

Doctoral Program in Biomedical Engineering

Approved in public examination by the Jury:

President:	Professor Ana Maria Rodrigues de Sousa Faria de Mendonça
Referee:	Professor Adrian Galdran Cabello
Referee:	Professor Carlos Jorge Andrade Mariz Santiago
Referee:	Professor Francesco Renna
Referee:	Professor Miguel Fernando Paiva Velhote Correia
Referee:	Professor Jaime dos Santos Cardoso

Definitive version validated by the Supervisor:

_____

Jaime dos Santos Cardoso, PhD

March, 2024

Resumo

O cancro do colo do útero é a quarta principal causa de morte relacionada com cancro em mulheres, com a maioria dos casos a ocorrer em países de baixo e médio rendimento. Apesar do recente avanço científico, um tratamento totalmente eficaz, especialmente para diagnósticos em estágios avançados, permanece insuficente. Métodos de rastreio, como citologia, colposcopia e histologia, desempenharam um papel fundamental na redução significativa das mortes relacionadas com o cancro do colo do útero. A necessidade imperiosa agora é desenvolver sistemas inovadores de apoio à decisão para o rastreio do cancro do colo do útero que integrem perfeitamente múltiplos exames de rastreio. Neste contexto, os sistemas de Diagnóstico Assistido por Computador (DAC) têm assumido crescente importância, principalmente para efeitos de controlo de qualidade e avaliação do risco de cancro. Estes sistemas DAC tiram partido do poder dos algoritmos de aprendizagem computacional e de técnicas avançadas de análise de imagem para auxiliar os profissionais de saúde no rastreio eficiente e no diagnóstico do cancro. Ao oferecerem análise automatizada e apoio à decisão, os sistemas DAC elevam a precisão, a eficiência e a consistência dos diagnósticos do cancro. As suas capacidades estendem-se à avaliação da qualidade da imagem, à deteção de regiões anormais e à categorização do risco de cancro, desempenhando assim um papel fundamental na melhoria dos resultados para os pacientes e no auxílio à deteção precoce e tratamento do cancro do colo do útero. No entanto, apesar do progresso notável alcançado, permanece uma distância considerável a percorrer e uma multiplicidade de desafios a enfrentar para abordar de forma abrangente as complexidades deste campo. Esta tese embarca numa jornada para enfrentar vários desafios relacionados com o controlo de qualidade e a avaliação de risco, abrangendo domínios de pesquisa fundamentais e aplicados.

A primeira parte desta tese resulta em várias contribuições relacionadas com questões de controlo de qualidade em sistemas de apoio à decisão médica. Começa por focar na avaliação da qualidade da imagem, com ênfase particular na deteção de desfocagem em lâminas de patologia digital. A deteção de desfocagem em imagens médicas é de importância crucial para garantir decisões clínicas precisas. Imagens desfocadas ou distorcidas podem levar a interpretações erradas e diagnósticos potencialmente incorretos, o que pode ter graves consequências para o paciente. Posteriormente, esta parte explora também métodos para melhorar a qualidade da imagem em microscopia de citologia de baixo custo. Por fim, esta parte encerra abordando uma tarefa diária crucial num laboratório de patologia digital: a deteção automatizada e quantificação de fragmentos, com uma importância significativa em diagnósticos de rotina. A automatização do processo de contagem de fragmentos tem o potencial de aliviar a carga de trabalho dos clínicos, ao mesmo tempo que mitiga o risco de erros.

Na segunda parte desta tese, há um foco em fazer contribuições fundamentais para a classificação ordinal no âmbito da aprendizagem computacional. Estas contribuições envolvem a introdução de várias funções de perda ordinal adaptadas para a classificação do risco de cancro do colo do útero. O objetivo principal é aproveitar as características ordinais inerentes aos dados para melhorar o desempenho geral dos modelos de classificação.

Na parte final desta tese, apresentam-se novas contribuições relacionadas com a integração de dados multimodais em sistemas de apoio à decisão. Estas contribuições envolvem a exploração e aplicação de várias técnicas destinadas a combinar eficazmente dados de fontes e modalidades diversas. Este trabalho procura aproveitar as vantagens de cada modalidade e desbloquear o potencial sinérgico que surge quando diferentes modalidades de dados são integradas.

Em resumo, esta tese realça a importância crítica de abordagens inovadoras para o rastreio do cancro do colo do útero, especialmente em países de baixo a médio rendimento, onde continua a ser uma das principais causas de mortes relacionadas com o cancro entre as mulheres. Embora tenham sido feitos progressos através de métodos tradicionais de rastreio, a integração de sistemas de DAC, impulsionados por aprendizagem computacional e análise avançada de imagens, está a aprimorar significativamente a precisão e eficiência no diagnóstico do cancro do colo do útero. Esta tese concentra-se no controlo de qualidade, classificação ordinal e integração de dados multimodais como áreas-chave de melhoria. Estes esforços representam um passo em direção à redução do impacto do cancro do colo do útero e, em última análise, à melhoria dos resultados para os pacientes.

Abstract

Cervical cancer stands as the fourth leading cause of cancer-related fatalities among women, with the majority of cases occurring in low to middle-income countries. Despite the recent surge in scientific advances, a completely effective treatment, particularly for advanced-stage diagnoses, remains elusive. Screening methods, such as cytology, colposcopy, and histology, have played a pivotal role in significantly reducing cervical cancer-related deaths. The imperative now is to develop innovative decision support systems for cervical cancer screening that seamlessly integrate multiple screening exams. In this context, Computer-Aided Diagnosis (CAD) systems have assumed growing significance, primarily for the purposes of quality control and evaluating cancer risk. These sophisticated CAD systems harness the power of machine learning algorithms and advanced image analysis techniques to assist healthcare practitioners in the efficient screening and diagnosis of cervical cancer. By offering automated analysis and decision support, CAD systems elevate the precision, efficiency, and consistency of cervical cancer diagnoses. Their capabilities extend to assessing image quality, detecting abnormal regions, and categorizing cancer risk, thereby playing a pivotal role in improving patient outcomes and aiding in the early detection and treatment of cervical cancer. Nevertheless, despite the notable progress achieved, there remains a substantial distance to traverse and a multitude of challenges to confront in order to comprehensively address the complexities of this field. This thesis embarks on a journey to tackle various challenges associated with quality control and risk assessment, spanning both fundamental and applied research domains.

The first part of this thesis results in several contributions related to quality control issues in medical decision support systems. It begins by focusing on the evaluation of image quality, with a particular emphasis on the detection of blurriness in digital pathology slides. Detecting blurriness in medical images is of paramount importance for ensuring accurate clinical decisions. Blurry or distorted images can lead to misinterpretations and potentially incorrect diagnoses, which can have serious consequences for patient care. Subsequently, this part also explores methods for improving image quality in cost-effective cytology microscopy. Lastly, this part closes by addressing a crucial daily task in a digital pathology lab: the automated detection and quantification of fragments, which holds significant importance in routine diagnostics. Automating the process of fragment counting has the potential to alleviate the workload of clinicians while simultaneously mitigating the risk of errors.

In the second part of this thesis, there is a focus on making fundamental contributions to ordinal classification within the realm of machine learning. These advancements involve the introduction of various ordinal loss functions tailored for classifying cervical cancer risk. The primary objective is to harness the inherent ordinal characteristics of the data to improve the overall performance of classification models.

In the final part of this thesis, we presented novel contributions related to the integration of multimodal data within decision support systems. These contributions involve the exploration and application of various techniques aimed at effectively combining data from diverse sources and

modalities. This work seeks to leverage the strengths of each modality and harness the synergistic potential of integrating them.

In summary, this thesis underscores the critical importance of innovative approaches to cervical cancer screening, especially in low to middle-income countries where it remains a leading cause of cancer-related deaths among women. While progress has been made through traditional screening methods, integrating CAD systems, powered by machine learning and advanced image analysis, is enhancing precision and efficiency in cervical cancer diagnosis. This thesis focuses on quality control, ordinal classification, and multimodal data integration as key areas of improvement. These efforts represent a step toward reducing the impact of cervical cancer and ultimately improving patient outcomes.

Acknowledgments

I would like to express my sincere gratitude to the individuals who guided and inspired me throughout this research journey. Prof. Jaime Cardoso and Dr. Maria João Vasconcelos provided invaluable supervision and immense inspiration. Prof. Jaime Cardoso has consistently exemplified what it means to be a machine learning scientist, with an unwavering commitment to scientific and ethical rigor. Prof. Jaime represents the epitome of an ideal supervisor. Dr. Maria João Vasconcelos has consistently been a valuable presence, always available to engage in discussions on pertinent subjects. Dr. Maria skillfully bridges the worlds of academia and the research industry, effectively connecting these spheres. Furthermore, Dr. Maria exhibits commendable scientific and ethical rigor in her reviews and suggestions, exemplifying the highest standards in her work.

In addition to my supervisors, this work could not have been accomplished without the support of my two affiliated institutions, namely the Faculty of Engineering at the University of Porto and INESC TEC. Furthermore, I would like to express my gratitude for the financial backing received from the Portuguese Foundation for Science and Technology, made possible through the Ph.D. grant with the reference number 2021.05102.BD.

I also wish to express my deep gratitude to Prof. Peter Schüffler, Dr. Benedikt Wiestler, and Dr. Claire Delbridge who supervised me during my internship in Germany, motivating me, even more, to try to have an impact on computational pathology. A big thanks to my colleagues and friends from Munich for the wonderful moments we shared during my stay there.

When I joined INESC TEC towards the end of 2019, Dr. Ricardo Cruz who was also a Ph.D. student at the time and is now a former Ph.D., extended a warm welcome and imparted knowledge on machine learning, programming, and various helpful tools and tricks that proved invaluable throughout my doctoral journey. Dr. Ricardo Cruz was a *de facto* supervisor at the beginning of my Ph.D. He also actively engaged in our discussions, and we collaborated and still collaborating closely, particularly on the subject of ordinal classification. A huge thanks to my colleagues and friends in the VCMI group, from whom I have gained invaluable knowledge and insights. This learning experience occurred not only through formal collaborations but also during casual coffee breaks and informal discussions over snacks.

My thanks also go to Dr. Diana Montezuma and the rest of the IMP team for all their support, motivation, and supervision during the CADPath.AI project.

To all my friends from university and outside university, I extend my heartfelt gratitude for your support and encouragement throughout this journey. Your presence in my life has been an invaluable source of inspiration, motivation, and camaraderie.

I extend my immense gratitude to my parents, Goretti and Manuel, for their support, boundless affection, constant motivation, resilience, and enduring patience, even during my most challenging moments. The successful completion of this endeavor is also attributed to you, as you have consistently inspired me to pursue my studies, strive for excellence, and always give my best. My special thanks to my grandparents (Maria and Abílio) for all the affection and comfort they gave me. I would like also to extend my gratitude to the rest of my family. I want to extend a special

acknowledgment to Marta, who consistently showed great enthusiasm for my endeavors and remained by my side for the majority of this journey. Despite the divergent directions we have taken in life, you will forever hold a special place in my heart.

I would like to take a moment to express my heartfelt gratitude to my loyal companion, Charlie, who has been by my side throughout this research endeavor. Although a dog may not understand the complexities of academic pursuits, Charlie has provided support and companionship during the ups and downs of this journey.

Finally, to all of those who in some way contributed to my success during this chapter of my life, I am honestly thankful for your contribution.

Tomé Mendes Albuquerque



"Our intelligence is what makes us human,
and Artificial Intelligence is an extension of this quality."

Yann LeCun

Contents

1 Introduction **1**
- 1.1 Context and Motivation . . . 1
- 1.2 Objectives . . . 2
- 1.3 Dissemination . . . 3
- 1.4 Collaborations . . . 5
- 1.5 Thesis structure . . . 6

2 Cervical Cancer **9**
- 2.1 Epidemiology of Cervical Cancer . . . 9
- 2.2 Female Reproductive System Anatomy . . . 11
- 2.3 Cervix Cytology . . . 13
- 2.4 Signs and Symptoms . . . 15
- 2.5 Prevention and Treatment . . . 16
- 2.6 Pathophysiology . . . 17
- 2.7 Screening Methods . . . 18
  - 2.7.1 Cytology . . . 18
  - 2.7.2 Colposcopy . . . 19
  - 2.7.3 Histology . . . 20
- 2.8 Summary . . . 22

3 Machine Learning Fundamentals **25**
- 3.1 Machine Learning Algorithms . . . 25
  - 3.1.1 Feed-forward Neural Networks . . . 27
  - 3.1.2 Convolutional Neural Networks (CNNs) . . . 29
  - 3.1.3 Multimodal Machine Learning . . . 31
  - 3.1.4 Explainable Artificial Intelligence . . . 32
- 3.2 Datasets . . . 32
  - 3.2.1 Cytology . . . 32
  - 3.2.2 Colposcopy . . . 32
  - 3.2.3 Histology . . . 34

4 Quality Control in Medical Decision Support Systems **35**
- 4.1 Deep Ordinal Focus Assessment for Whole Slide Images . . . 36
  - 4.1.1 Context and Motivation . . . 36
  - 4.1.2 Related Work . . . 37
  - 4.1.3 Brief Summary . . . 38
  - 4.1.4 Methods . . . 38
  - 4.1.5 Results and Discussion . . . 43

4.1.6 Summary and Conclusions . . . 45
4.2 Image Enhancement using Deep Based Extended Depth of Field Methods . . . 45
4.2.1 Context and Motivation . . . 46
4.2.2 Related Work . . . 48
4.2.3 Proposed Methodology . . . 51
4.2.4 Experimental Details . . . 58
4.2.5 Results and Discussion . . . 61
4.2.6 Summary and Conclusions . . . 63
4.3 Quality Control in Digital Pathology: Automatic Fragment Detection and Counting 65
4.3.1 Context and Motivation . . . 65
4.3.2 Brief Summary . . . 67
4.3.3 Methods . . . 68
4.3.4 Experimental Details . . . 71
4.3.5 Results and Discussion . . . 73
4.3.6 Improvements in Automatic Fragment Counting . . . 74
4.3.7 Summary and Conclusions . . . 75

5 Classification of Cancer Risk 77
5.1 Ordinal Classification of Cervical Cancer Risk . . . 77
5.1.1 Motivation and Context . . . 79
5.1.2 Task definition and related work . . . 79
5.2 Non-parametric Unimodal Losses (CO, CO2, and HO2) . . . 84
5.2.1 Proposal . . . 84
5.2.2 Methods . . . 85
5.2.3 Experimental Details . . . 86
5.2.4 Results and Discussion . . . 89
5.2.5 Conclusion . . . 93
5.3 Non-parametric Quasi-Unimodal Losses (QULCE and QULHO) . . . 94
5.3.1 Proposal . . . 94
5.3.2 Methods . . . 97
5.3.3 Experimental Details . . . 98
5.3.4 Results and Discussion . . . 99
5.3.5 Summary and Conclusions . . . 104
5.4 UnimodalNet and Non-parametric Unimodal Losses (WU-KLDiv and WU-Wass) 104
5.4.1 Proposal . . . 104
5.4.2 Methods . . . 107
5.4.3 Experimental Details . . . 108
5.4.4 Results and Discussion . . . 108
5.4.5 Summary and Conclusions . . . 113

6 Multimodal Data Integration in Decision Support Systems 115
6.1 Embedded regularization for classification of colposcopic images . . . 115
6.1.1 Context and Motivation . . . 116
6.1.2 Methods . . . 117
6.1.3 Experimental Details . . . 119
6.1.4 Results and Discussion . . . 120
6.1.5 Summary and Conclusions . . . 121

7 Conclusion **123**
7.1 Production 124
7.2 Limitations 124
7.3 Future Work 125

References **127**

A Image Enhancement using Deep Based Extended Depth of Field Methods **147**

B Non-parametric Unimodal Losses (CO, CO2, and HO2) **148**

C Multimodal Data Integration **151**
C.1 Multimodal Context-Aware Detection of Glioma Biomarkers using MRI and WSI 151
C.1.1 Motivation and Context 152
C.1.2 Proposed Methodology 154
C.1.3 Experimental Details 155
C.1.4 Results and Discussion 158
C.1.5 Summary and Conclusions 159

List of Figures

2.1 Worldwide incidence and mortality of cervical cancer. Adapted from [111]. . . . 9
2.2 Bar chart containing the incidence and mortality of cervical cancer in each specific world region for 2018, the statistics are age-standardized. Adapted from [40]. . . 10
2.3 Incidence (a) and mortality (b) of cervical cancer by regions in Portugal, 2015. Adapted from [66] . . . 11
2.4 Coronal and sagittal view of the internal female reproductive system organs. Adapted from [188]. . . . 12
2.5 Columnar epithelium of the endocervix. Adapted from [185]. . . . 13
2.6 Squamous epithelium of the ectocervix. Adapted from [185]. . . . 14
2.7 Squamocolumnar junction (SCJ). Adapted from [185]. . . . 15
2.8 Cervical cancer diagnosis workflow in Portugal. . . . 18
2.9 Two examples of portable low-cost devices for cytology (a) and colposcopy (b). . 20
2.10 Digital Cervical histology workflow. . . . 23

3.1 A comparison scheme between the traditional programming paradigm (a) and machine learning (b). . . . 26

4.1 Examples of data augmentation on the FocusPath database. From left to right: original image and three examples of random transformations. . . . 38
4.2 Schematic representation of the model architecture. . . . 42
4.3 Confusion Matrix of our best model - MobileNet_v2 - using Ordinal Encoding as loss. . . . 44
4.4 Comparison between two EDoF images generated without alignment (a) and with the alignment (b) of the images stack. . . . 47
4.5 Most relevant fusion methods present in the literature. . . . 51
4.6 Representation of the proposed and tested prepossessing workflow method: acquisition of the stacks followed by chromatic, rigid, and elastic alignment of the stacks. 52
4.7 Comparison between two microscopy images without (a) and with (b) chromatic alignment . . . 53
4.8 Example of an alignment of a slide using elastic transformations where (a) represents the central/reference slide; (b) slide before alignment, (c) slide after the alignment, and (d) and (e) representing the deformation displacement in xx's and yy's respectively. . . . 55
4.9 Schematic representation of the EDoF-CNN-Fast Network. . . . 57
4.10 Schematic representation of the EDoF-CNN-Pairwise Network. . . . 57

4.11 Number of fragments assessment checkpoint. On the first workflow (A.), a manual assessment of the number of fragments on the slides is made and confronted with the macroscopic report; the slides are posteriorly scanned. In the proposed workflow (B.), the system automatically evaluates the adequacy of the number of fragments on the slides after scanning, giving an alert message if a discrepancy is found. 66
4.12 Schematic representation of the three different possibilities on the tissue slides preparation: (a) several sets of only one fragment; (b) one set of different fragments and (c) several equal sets of different fragments. 67
4.13 Considering the set identification task, pairs 1-2 and 3-4 should be assigned with label 1, while all the other possible pairs should be assigned with label 0, for this image. 69
4.14 Schematic representation of the methodology used for sets grouping. Each circle represents a component. The strong graphs are represented with black connection lines. The grey lines, together with the black lines, constitute the weak graph. The final red bounding boxes are the two possible results for this step. 69
4.15 Schematic representation of the proposed framework. 70
4.16 Three different examples of macro slide images present in the dataset. 72
4.17 Confusion Matrices of the three different models for the fragment/set counting task. 75
4.18 Schematic representation of the fusion model for fragment counting task. 76

5.1 Example of possible output probability distributions. Even if both outputs agree on the majority class, only the unimodal distribution is consistent with an ordinal regression task. 80
5.2 Summary of the current unimodal and quasi-unimodal approaches, where the axes represent the soft/hard constraint and parametric/non-parametric priorities. The proposed contributions are also mentioned in the right families. 82
5.3 Pictorial representation of unimodal and non-unimodal distributions in the 2D simplex. 83
5.4 Schematic representation of the model used for multi-class classification of Pap smear cells. 86
5.5 Schematic representation of the used and proposed ordinal losses. 87
5.6 Examples of data augmentation on the Herlev database. The original zero-padding image (A) and random transformations (B–D). 88
5.7 Comparison of state-of-the-art confusion matrix (7 classes) against WideResNet50 trained using the HO2 loss. 92
5.8 Probability distribution for WideResNet50 contrasting losses CE (solid line) and HO2 (dashed line). 93
5.9 Schematic representation of ordinal loss functions from the literature. The proposed loss function is underlined. 95
5.10 Schematic representation of an example where there is ordinal consistency in the decision, without a unimodal distribution in the classes. X axis corresponds to the input space (x) of the model, and Y axis corresponds to output probabilities ($p(\hat{y})$) of the model. Along x, an example of a decision region is represented where the model chooses between two different classes with a boundary region between this decision. 96

5.11 Probabilities produced by three different models for an observation n. HO2 forces the entire distribution curve to have decreasing probabilities. On the other hand, QULHO focuses on the top neighborhood, and is not as concerned about the more distant classes. In this example, $k_n^\star = 4$ is assumed to be the true class. 97
5.12 Schematic representation of the two QUL terms as in equation (5.10): the first term penalizes only the direct neighborhood, while the second term penalizes the remaining classes relative to the neighborhood. 97
5.13 Illustration for three classes of the proposed activation function, which forces the model output to be a unimodal distribution. 105
5.14 Results for different values of λ for the proposed WU-Wass. As λ increases, output probabilities tend to be more unimodal while sacrificing a little accuracy. . 112

6.1 Example of data augmentation on NCI/NHI database. (from left to right: original image and three examples of random transformations). 117
6.2 Gaussian matrix used for regularization of the model. 119
6.3 CNN model used in this work. 119
6.4 Heatmaps of baseline and Gaussian regularization models. A legend for each image is displayed under the latter. 121

C.1 Simplified overview of the classification pathway for the majority of adult-type diffuse gliomas based on WHO (2021) classification of CNS tumors. 153
C.2 Overview of the most relevant methods present in the literature for glioma biomarkers detection. 154
C.3 Schematic representation of the models proposed in this study: (a) Cascade w/ Fusion (b) Mid Fusion (c) Late Fusion (d) Mid context Fusion. 156
C.4 ROC plot of the 5-folds cross-validation results for the best model per modality. . 158

List of Tables

2.1 Scheme of the different types of transformation zones and respective examples of colposcopic images. Adapted from [25]. 14
2.2 Different Human Papillomavirus groups and types related to cervical cancer. Adapted from [85]. 17

3.1 Publicly available cervical cytology datasets in literature. 33
3.2 Available colposcopy datasets in literature. 34
3.3 Available histology datasets in literature. 34

4.1 Image examples of the 14 Different patch focus Classes in the FocusPath dataset. 41
4.2 Results in terms of Accuracy (%) for 12 class problem (higher is better). 43
4.3 Results in terms of Mean Absolute Error (MAE) for 12 class problem (lower is better). 43
4.4 Results in terms of Kendall's τ for 12 class problem (higher is better). 44
4.5 Comparison of our proposal with literature models 45
4.6 Summary of the number of trainable parameters across the different architectures(proposal architectures (P)) for a stack with five z-planes. 58
4.7 Example of several stacks and respective EDoF present in Cervix93 dataset. . . . 58
4.8 Example of a stack and respective EDoF present in EDoF Fraunhofer dataset. . . 59
4.9 Results for the different architectures on Cervix93 dataset. 61
4.10 Results for the ablation study across the different architectures and workflows on EDoF Fraunhofer dataset. 62
4.11 EDoF images obtained using the different methods with a 2x zoom for better visualization of the details. 64
4.12 Results of the different models for the fragment/set detection task. (GT - Ground Truth; CL - Component Labelling). 73
4.13 Results of the three different models for the fragment/set counting task. 74

5.1 Fraction of Unimodal Distributions 84
5.2 The seven different pap smear classes in the Herlev dataset. 87
5.3 Image examples of the seven different pap smear classes in the Herlev dataset. . . 88
5.4 Results in terms of Mean Absolute Error (MAE) for 7 class problem, averaged for 10 folds (lower is better). 90
5.5 Results in terms of Accuracy for 7 class problem, averaged for 10 folds (higher is better). 90
5.6 Results in terms of Uniform Ordinal Classification Index (UOC) for 7 class problem, averaged for 10 folds (lower is better). 90
5.7 Results in terms of Mean Absolute Error (MAE) for 4 class problem, averaged for 10 folds (lower is better). 91

5.8 Results in terms of Accuracy for 4 class problem, averaged for 10 folds (higher is better). 91
5.9 Results in terms of Uniform Ordinal Classification Index (UOC) for 4 class problem, averaged for 10 folds (lower is better). 91
5.10 Accuracy comparison of different models with literature for 7 and 4 classes. . . . 93
5.11 Examples of six different classes from the AFAD dataset. 99
5.12 Results for the Herlev dataset when considering 7 classes, averaged for 10 folds. . 100
5.13 Results for the Herlev dataset when considering 4 classes, averaged for 10 folds. . 101
5.14 Results for the FocusPath dataset, averaged for 10 folds. 102
5.15 Results for the AFAD dataset, averaged for 3 folds. 103
5.16 Datasets 108
5.17 Average Summarization of the Results 109
5.18 Results for the Tabular Datasets 110
5.19 Results for the Image Datasets 111

6.1 Ablation Study (averaged for 5-folds) for binary classification of colposcopic images ($Negative\ Predictive\ Value\ (NPV) = \frac{True\ Negative}{False\ Negative + True\ Negative}$). 120
6.2 Comparison of our proposed model with literature. 120

7.1 Paper production during the thesis, split by chapter.*Part of papers presented in Annex C 125

A.1 Results Fraunhofer dataset no chromatic align, only elastic. 147

B.1 Results in terms of Kendall's τ for 7 class problem, averaged for 10 folds (higher is better). 148
B.2 Results in terms of Kendall's τ for 4 class problem, averaged for 10 folds (higher is better). 149
B.3 Aggregate results for 7 class problem, averaged for 10 folds. 149
B.4 Aggregate results for 4 class problem, averaged for 10 folds. 150

C.1 Summary of TCGA dataset with targets encoding and frequency of samples per class. 157
C.2 Results for all modalities (5-fold cross validation, CE loss) - TCGA dataset. . . . 159

Abbreviations

ACC	Accuracy
AI	Artificial Intelligence
ANN	Artificial Neural Networks
AUC	Area Under the Curve
BU	Binomial Unimodal
CAD	Computer-Aided Diagnosis
CE	Cross-entropy
CL	Component Labelling
CNN	Convolutional Neural Network
CO2	Ordinal cross-Entropy Loss
DL	Deep Learning
DSS	Decision Support System
EDoF	Extended Depth of Field
ELM	Extreme Learning Machine
FQA	Focus Quality Assessment
FCM	Fuzzy c-Means Algorithm
FAST	Features from Accelerated Segments Test
GT	Ground Truth
HO2	Ordinal Entropy Loss
HPV	Human Papillomavirus
IQA	Image Quality Assessment
IoU	Intersection over Union
kNN	k-Nearest Neighbours
LDA	Linear Discriminant Analysis
LR	Learning Rate
MAE	Mean Average Error
MSE	Mean Squared Error
ML	Machine Learning
MLP	Multi-layer Perceptron
MLV	Maximum Local Variation

MRI	Magnetic Resonance Imaging
mAP	Mean Average Precision
NB	Naive Bayes classifiers
NCI	National Cancer Institute
NIH	National Institutes of Health
OE	Ordinal Encoding
ORB	Oriented FAST and Rotated BRIEF
PMF	Poisson Probability Mass Function
PLCC	Pearson Linear Correlation Coefficient
PSNR	Peak Signal-to-noise Ratio
PU	Poisson Unimodal
QDA	Quadratic Discriminant Analysis
QULCE	Quasi-unimodal Cross-entropy Loss
QULHO	Quasi-unimodal Entropy Loss
RMSE	Root Mean Squared Error
ROI	Region of Interest
R	Recall
SVM	Support Vector Machines
SRCC	Spearman's Rank Correlation Coefficient
SSIM	Structural Similarity Index
TCGA	The Cancer Genome Atlas
TBS	The Bethesda Classification System
UOC	Uniform Ordinal Classification Index
UN	UnimodalNet
WHO	World Health Organization
WSI	Whole Slide Image

Chapter 1

Introduction

1.1 Context and Motivation

Cervical cancer maintained its position as the fourth most prevalent cancer among women in 2020, and it continued to be the second leading cause of cancer-related deaths in women aged 20 to 39 globally. This age group witnessed over 10 deaths per week due to cervical cancer [207]. According to the World Health Organization (WHO), more than $604,000$ women were diagnosed with cervical cancer in 2020 alone, resulting in over $342,000$ deaths from this disease [111, 207]. The main factor for the high mortality rate is the asymptomatic characteristic of the condition in its initial stages [77].

Prevention and diagnostic methods play a crucial role in reducing the incidence and mortality rates [110]. In the majority of developed countries, screening programs have been established, which encompass a series of steps such as Human papillomavirus (HPV) testing, cytology testing (commonly known as Pap smear), colposcopy, and biopsy [32, 169]. HPV screening is of utmost importance due to the well-established association between HPV, a group of viruses, and increased cancer risk [111]. Cervical cytology testing is employed to detect any abnormal or potentially abnormal cells within the uterine cervix [197]. Subsequently, colposcopy is conducted to further assess the condition, utilizing a colposcope to capture images of the cervix. This procedure helps in clarifying inconclusive cytologic findings and provides an evaluation of the location, size, and extent of any identified lesion [91]. Following the colposcopy, the captured images are carefully examined by a healthcare professional to assess the severity of the lesion. If necessary, a biopsy is performed to obtain a tissue sample for further analysis and final diagnosis.

In low and middle-income countries, the availability of HPV testing and cytologic testing is limited. To address this issue, researchers have created low-cost microscopy devices. However, these devices have certain limitations that directly impact the quality of images. For instance, sample capture is performed using mobile phones, and cheaper lenses are often utilized, resulting in reduced physical properties. Overcoming these challenges is crucial for the research community in order to provide a more affordable solution for cytology examination in low-income countries. Alternatively, colposcopy devices can be utilized as a cost-effective method for diagnosing cervical

cancer, particularly in developing countries where the mortality rate is significantly higher [48]. To enhance the economic viability of this process, portable colposcopes have been developed. Furthermore, the implementation of an automatic Computer-Aided Diagnosis (CAD) system could serve as a feasible solution to reduce the time required for image analysis.

CAD systems have become increasingly important in the realm of cervical cancer, particularly for quality control and evaluating cancer risk. These sophisticated systems leverage machine learning algorithms and image analysis techniques to assist healthcare practitioners in efficiently screening and diagnosing cervical cancer. By offering automated analysis and decision support, CAD systems enhance the precision, effectiveness, and uniformity of cervical cancer diagnoses. With the capacity to assess image quality, detect abnormal regions, and classify cancer risk, CAD systems play a critical role in enhancing patient outcomes and aiding in the early identification and treatment of cervical cancer. However, despite the progress made, there is still a significant distance to cover and a multitude of challenges to be tackled in order to fully address the complexities of this field. This thesis addresses various challenges associated with quality control and risk assessment, encompassing both fundamental and applied research areas.

1.2 Objectives

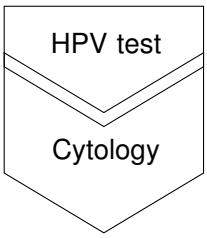
This thesis aims to contribute to the advancement of multimodal cervical cancer automatic detection by leveraging the information from cytology, colposcopy, histology, and clinical data of patients. The goal is to provide clinicians with a comprehensive set of tools that can be seamlessly integrated into their daily routine, ranging from quality control systems to diagnostic support. Recent and promising medical image analysis, machine learning techniques, and deep learning strategies will be applied and evaluated in this thesis and evaluated. Experimental evidence shows that humans take advantage of multimodal sensorial experiences even when tasks are meant to be unimodal [231]; furthermore, multiple modalities allow systems to build robustness to the loss of one modality. Thus, multiple types of information for diagnosis are clinically desirable. Therefore, it is our goal to build a method that can interpret predictions through visual attention, which is understandable to specialists. The interpretable aspect is essential for a robust decision support system (DSS). The main scientific and technical objectives of the work can be divided into five main groups:

- Research and study configurations for a real-time image quality assessment and enhancement system of cervical imaging;
- Promote a new technique for histology fragment counting and detection in national and international healthcare institutions;
- Include interpretability mechanisms in algorithms to decode learned representations into rich interpretable predictions, which are understandable by the clinicians;

- Propose new loss functions to take advantage of the ordinality present among cervical cancer stages;
- Develop automatic multimodal cervical cancer detection algorithms for a robust, secure, and confident DSS;
- Evaluate the algorithms using standard metrics used in medical systems.

1.3 Dissemination

The contributions of the doctoral research on deep learning strategies for automatic decision support systems in cervical cancer diagnosis described in this thesis have been disseminated as part of twelve scientific publications (three in international journals; five in international conferences; three in national conferences and one release of a public dataset). These are (clustered in reverse chronological order and following cervical cancer workflow):



[45] - Jaime S. Cardoso, Ricardo Cruz, and Tomé Albuquerque. Unimodal distributions for ordinal regression, 2023. (submitted to IEEE Transactions on Artificial Intelligence journal - waiting for decision) https://doi.org/10.48550/arXiv.2303.04547

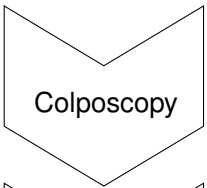
[14] - Tomé Albuquerque, Luís Rosado, Ricardo Cruz, Maria João M. Vasconcelos, Tiago Oliveira, and Jaime S. Cardoso. "Rethinking low-cost microscopy workflow: Image enhancement using deep based extended depth of field methods". Intelligent Systems with Applications, 2023. doi: https://doi.org/10.1016/j.iswa.2022.200170

[200] - Ana Sampaio, Tomé Albuquerque, Pedro Faria, Luís Rosado, and Maria João M. Vasconcelos. (2023). "Fraunhofer Portugal AICOS EDoF Dataset" [Dataset]. Zenodo. doi: https://doi.org/10.5281/zenodo.7473277

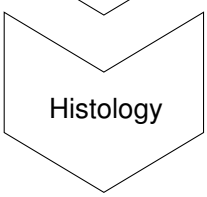
[8] - Tomé Albuquerque, Ricardo Cruz, and Jaime S. Cardoso. "Quasi-Unimodal Distributions for Ordinal Classification". Mathematics 2022, 10, 980. https://doi.org/10.3390/math10060980

[7] - Tomé Albuquerque, Ricardo Cruz, Jaime S. Cardoso. "Ordinal losses for classification of cervical cancer risk". PeerJ Computer Science 7:e457, 2021. https://doi.org/10.7717/peerj-cs.457

[15] - Tomé Albuquerque, Maria João M. Vasconcelos, and Jaime S. Cardoso. 2020. "Image Quality Assessment of Cytology Images using Deep Learning". 26th portuguese conference on pattern recognition (RECPAD), online, 2020. aprp.pt/wp-content/uploads/proceedings_recpad2020.pdf



[5] - Tomé Albuquerque and Jaime S. Cardoso, "Embedded Regularization For Classification Of Colposcopic Images," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1920-1923, doi: 10.1109/ISBI48211.2021.9433871.



[227] - Ana Beatriz Vieira, Tomé Albuquerque,Diana Montezuma, Liliana Ribeiro, Domingos Oliveira, Isabel Macedo Pinto, Jaime S. Cardoso, and Arlindo Oliveira. "To reject or

not to reject, that is the question: A new perspective on counting fragments in Whole-Slide-Images". 29th portuguese conference on pattern recognition (RECPAD), Coimbra, Portugal, 2023. [not covered]

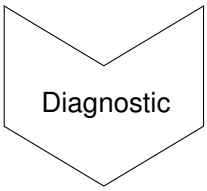
[228] - Ana Beatriz Vieira, Diana Montezuma, Tomé Albuquerque, Liliana Ribeiro, Filipa Rebolo, Domingos Oliveira, João Monteiro, Sofia Gonçalves, Isabel Macedo Pinto, Jaime S. Cardoso, and Arlindo Oliveira. "Automatic detection and counting of pathological specimens with a rejection option". 35th European Congress of Pathology (ECP), dublin, Ireland, 2023. [not covered]

[11] - Tomé Albuquerque, Ana Moreira, Beatriz Barros, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, Jaime S. Cardoso, "Quality Control in Digital Pathology: Automatic Fragment Detection and Counting," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 588-593, doi: 0.1109/EMBC48229.2022.9871208

[10] - Tomé Albuquerque, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, Jaime S. Cardoso, "Quality checkpoint in pathology specimens handling: an AI system to automate fragment detection and count". 18th European Congress on Digital Pathology (ECDP), Berlin, Germany, 2022.

[13] - Tomé Albuquerque, Ana Moreira, and Jaime S. Cardoso, "Deep Ordinal Focus Assessment for Whole Slide Images", 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 657-663, doi: https://10.1109/ICCVW54120.2021.00079

[12] - Tomé Albuquerque, Ana Moreira, and Jaime S. Cardoso. 2021. "Order is the key: Deep focus assessment in Whole Slide Images". 27th portuguese conference on pattern recognition (RECPAD), Évora, Portugal, 2021. recpad2021.uevora.pt/wp-content/uploads/2021/12/proceedings-recpad2021.pdf



Moreover, three additional publications in international and national conferences resulting from the exchange period of six months at the Technical University of Munich (TUM) regarding the use of multimodal data integration in the field of brain cancer were also produced:

[6] - Tomé Albuquerque, Beatriz Coutinho, João Rodrigo, Miguel Almeida, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Peter Schüffler, Jaime S. Cardoso. "Enhancing Glioma Detection through Multimodal Classifiers: Integrating MRI and WSI". 29th portuguese conference on pattern recognition (RECPAD), Coimbra, Portugal, 2023.

[9] - Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, Peter Schüffler. "Multimodal Identification of Brain Tumor Biomarkers using a Deep Fusion Strategy from MRI and Digital Pathology exams". 19th European Congress on Digital Pathology (ECDP), Budapest, Hungary, 2023.

[4] - Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, Peter Schüffler, "Multimodal Context-Aware Detection of Glioma Biomarkers using MRI and

WSI", 4th International Workshop on Multiscale Multimodal Medical Imaging (MMMI) - MICCAI, Vancouver, Canada, 8-12 October, 2023.

1.4 Collaborations

This thesis involved extensive collaborations with researchers and clinicians from various external institutions, specifically:

- **Dr. Maria João Vasconcelos** and **Dr. Luís Rosado** from Fraunhofer Portugal AICOS, who provided cytology image data crucial for the image quality assessment and enhancement in low-cost microscopy work, and also feedback for all the results, and articles we published. Furthermore, Dr. Maria João M. Vasconcelos co-supervised this thesis, providing inspiration and guidance.
- **Dr. Diana Montezuma**, **Dr. Domingos Oliveira**, **Dr. Ana Monteiro**, and **Dr. Isabel Macedo Pinto** from IMP Diagnostics, who provided image data regarding the automatic fragment detection and counting, slide annotations, and also feedback for all the results, and articles we published. Moreover, Dr. Diana Montezuma provided inspiration and clinical supervision.
- **Prof. Peter Schüffler** from the Technical University of Munich (TUM) who supervised me during my exchange period at the TUM, in Munich, Germany. Prof. Peter Schüffler provided scientific supervision related to computational pathology and multimodal data integration.
- **Dr. Benedikt Wiestler**, from the Technical University of Munich (TUM) co-supervised me during my exchange period at the TUM, in Munich, Germany, and also extensively supported my work with magnetic resonance imaging (MRI) knowledge.
- **Dr. Claire Delbridge** from the Technical University of Munich (TUM) who co-supervised me during my exchange period at the TUM, in Munich, Germany, and also extensively supported my work with neuropathology knowledge.

The author of this thesis also engaged in extensive collaborations with colleagues from the VCMI group at INESC TEC, namely Ricardo Cruz, Sara Oliveira, and Pedro Neto. These collaborations resulted in several publications that are presented within this thesis. During his Ph.D., he was involved in organizing several scientific events, such as a project committee member for VISUM Summer School in 2021 and 2022. The research work developed throughout the Ph.D. was related to or directly motivated by several scientific projects, namely:

- **CLARE**: Computer-aided cervical cancer screening [FCT Research Project]
- **TAMI**: Transparent Artificial Medical Intelligence [CMU Portugal Research Project]

- **CADPath.AI**: Computer-Aided Diagnosis in Pathology [FCT Research Project]

The author of this thesis collaborated and still collaborates as **co-supervisor**, alongside Prof. Jaime S. Cardoso (supervisor), in the supervision of the master dissertation of Maria Carneiro, work entitled "Quality Control in Digital Pathology: Improving fragment detection and counting" (2024), presented to FEUP as part of the Master's degree in Informatics and Computer Engineering.

The author of this thesis collaborated as **co-supervisor**, alongside Prof. Jaime S. Cardoso (supervisor), in the supervision of the master dissertation of Nuno Freitas, work entitled "Classification of Cervical Cancer Risk using Deep Learning" (2022), presented to FEUP as part of the Integrated Master's degree in Electrical and Computers Engineering.

Besides the above-mentioned dissertations, the author also collaborated in the supervision of eight more students on internships and summer internships related to medical image analysis in the cancer field (presented in reverse chronological order):

- **Supervisor** - summer internships - João Rodrigo, Beatriz Coutinho, and Miguel Almeida - "Deep Detection of Glioma Biomarkers using MRI and WSI", from 2023-06-26 to 2023-07-27, INESC TEC, Porto, Portugal
- **Supervisor** - internship - Mei Ling-Fang - Working on Multimodal Identification of Brain Tumor Biomarkers using a Deep Fusion Strategy from MRI and Digital Pathology exams, from 2022-11-15 to 2023-03-30, Technical University of Munich, Germany
- **Co-supervisor** - internship - Ana Moreira - Working on Image quality Assessment methods and Quality control in digital pathology - fragment counting using traditional machine learning approaches, from 2021-02-15 to 2022-07-30, INESC TEC, Porto, Portugal
- **Supervisor** - summer internships - Daniel Silva and Ana Romero - "Classification of Cervical Cancer Risk using Deep Learning", from 2021-06-28 to 2021-07-30, INESC TEC, Porto, Portugal
- **Supervisor** - internship - Ana Silva - "Cell Nuclei Detection and Segmentation from cytology images", from 2019-12-01 to 2020-02-28, INESC TEC, Porto, Portugal

1.5 Thesis structure

This document is structured into seven chapters, each presenting key contributions toward the aforementioned goals. The initial chapter serves as the introduction, providing an overview of the research objectives and setting the context for the subsequent chapters. The following chapters are organized as follows:

- Chapter 2: This chapter provides a comprehensive overview of cervical cancer, encompassing essential medical concepts necessary for a thorough understanding of the subsequent chapters in this thesis.

- Chapter 3: In this chapter, a high-level overview is presented regarding the foundations of machine learning methods and the existing cervical cancer datasets. Additionally, it highlights the availability and characteristics of various datasets specifically related to cervical cancer. This overview serves as a foundation for the subsequent chapters, enabling a deeper understanding of the applied methodologies and the context in which they are implemented.
- Chapter 4: This chapter comprises three main subchapters that delve into quality control problems and medical decision support systems. The chapter initiates by delving into the assessment of image quality, with a specific emphasis on detecting blur in digital pathology slides. It subsequently investigates approaches to enhance image quality in low-cost cytology microscopy. Lastly, the chapter concludes by addressing an essential daily task in a digital pathology lab: automatic fragment detection and counting, which plays a vital role in routine diagnostics.
- Chapter 5: This chapter proposes different ordinal approaches for the classification of cervical cancer risk, aiming to leverage the inherent ordinal nature of the data and enhance the performance of the classification models.
- Chapter 6: In this chapter, the topic of multimodal data integration in decision support systems is explored. The chapter aims to investigate and analyze the methods and techniques used to combine and integrate diverse types of data from multiple modalities.
- Chapter 7: The final chapter of this thesis serves as a culmination of the research, presenting comprehensive conclusions that encompass the advances made, limitations encountered, future perspectives, and potential avenues for further research.

Chapter 2

Cervical Cancer

In 2020, it was reported that cervical cancer remains the fourth most frequent cancer in women. This type of cancer is the second major cause of cancer-related deaths in women aged 20 to 39 years worldwide with more than 10 deaths per week in this age range [207]. According to the World Health Organization (WHO), in 2020 alone, more than $604,000$ women were diagnosed with cervical cancer and more than $342,000$ women died from this type of cancer [$111,207$]. Cervical cancer incidence and mortality can be reduced by prevention and diagnostic methods [110]. In this chapter, it is presented an overview of cervical cancer's main concepts including the epidemiology of cervical cancer, uterine cervix anatomy and cytology, signs and symptoms, prevention and treatment, pathophysiology, and the current workflow for screening methods.

2.1 Epidemiology of Cervical Cancer

More than $604,000$ women were diagnosed with cervical cancer in 2020, and this disease rates as the fourth most frequently diagnosed type of cancer and the fourth prominent cause of cancer death in women, just below breast, colorectal, and lung cancer (Figure 2.1) [$111,207$].

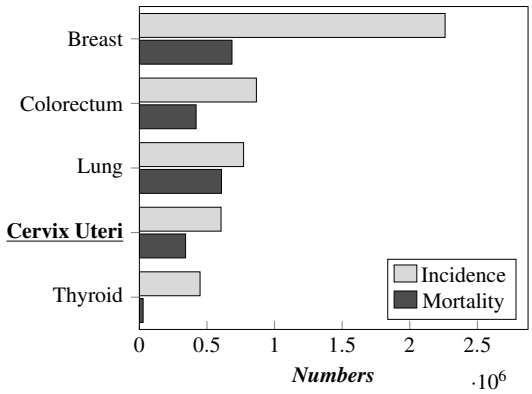


Figure 2.1: Worldwide incidence and mortality of cervical cancer. Adapted from [111].

The incidence of cervical cancer is higher in less-developed countries. Cervical cancer is the most frequently detected type of cancer in 28 different countries and the leading cause of

cancer-related death in 42 countries. Most of the cases occur in Sub-Saharan Africa and South-Eastern Asia. Consequently, the countries with the highest incidence rates are Swaziland (75 per 100,000 women (Southern Africa)), Malawi (73 per 100,000 women (Eastern Africa)), and Zambia (66 per 100,000 women (Eastern Africa)). In more relative terms, the rates are 7 to 10 times lower in more developed regions such as Australia/ New Zealand, North America, and Western Asia [40, 207]. Figure 2.2 shows a bar chart containing the incidence of cervical cancer in each Region, the statistics are age-standardized rates in 2018.

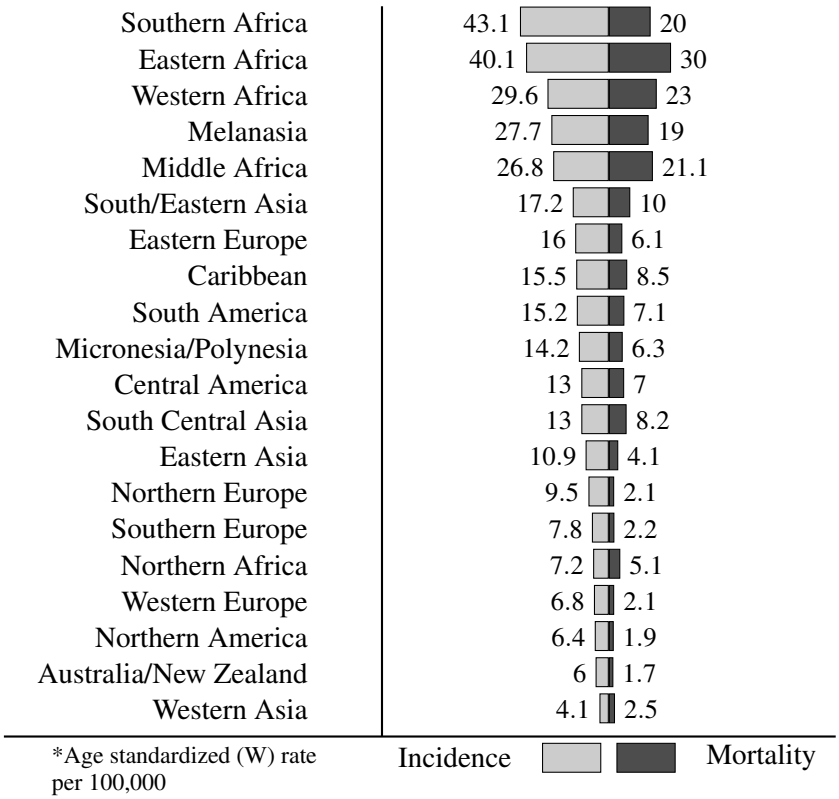


Figure 2.2: Bar chart containing the incidence and mortality of cervical cancer in each specific world region for 2018, the statistics are age-standardized. Adapted from [40].

The use and practice of screening methods have affected the incidence of cervical cancer. In more developed countries such as Nordic countries the age-standardized incidence percentages of cervical cancer have demonstrated a descending trend since the 1960s, demonstrating a decrease of the age-standardized incidence rate from 21.2 to 7.6 per 100,000 between 1966 to 2006 [110]. In contrast, an increase in cervical cancer rate occurred recently in several sub-Saharan African regions [159, 229]. The mortality rate is very high in less developed countries, with approximately 80% of deaths in all cervical cancer cases [169]. The highest mortality rates are reported in Malawi where 38 per 100,000 women die with cervical cancer. The age-standardized mortality rate in less developed countries is about three times higher than in developed regions. The mortality rate in the World reference population is 6.9 per 100,000 women, whereas the lowest mortality rates occur in Australia, Northern America, and Northern, Western, and South of Europe with rates of mortality

varying between 1.7 and 2.2 per 100,000 women (Figure 2.2) [40]. The existence of cervical cytology programs in developed countries promoted a declining trend in cervical cancer mortality. From 1993 to 2018, a regular decrease was reflected in the age-standardized mortality rate, from 3.4 per 100,000 women in 1993 to 2.1 per 100,000 in 2018 [207]. This high descent in mortality rates occurred in developed countries of Europe [111, 207]. The survival rate, in low to middle-income countries, in women affected by cervical cancer that can survive longer than 5 years is less than 50%, while in developed countries the 5-year survival is nearly 66% [154]. The improvement in diagnosis, and the better results of treatment, especially as a result of advances in radiotherapy methods, helped to improve the survival rate in the second half of the 20th century [180, 184, 216]. Despite the outburst of recent scientific advances, there is no effective treatment for cervical cancer being the diagnostic method that promotes an earlier diagnosis the main factor for the increase in survival rate [202].

In Portugal, cervical cancer remains the fifth most common cancer among women with the highest rate in Western Europe, which corresponds to an incidence rate of 13.5 per 100,000 women and mortality of 3.8 per 100,000 women in 2015 [66]. The increase in cervical cytology screening programs promoted a declining trend in cervical cancer incidence and mortality in Portugal. However, the south region remains the most affected due to less adherence to screening programs as Figure 2.3 shows.

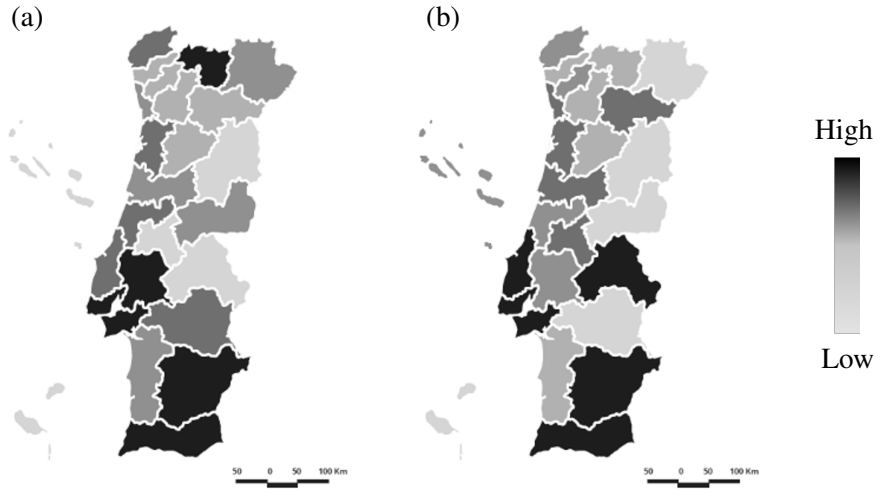
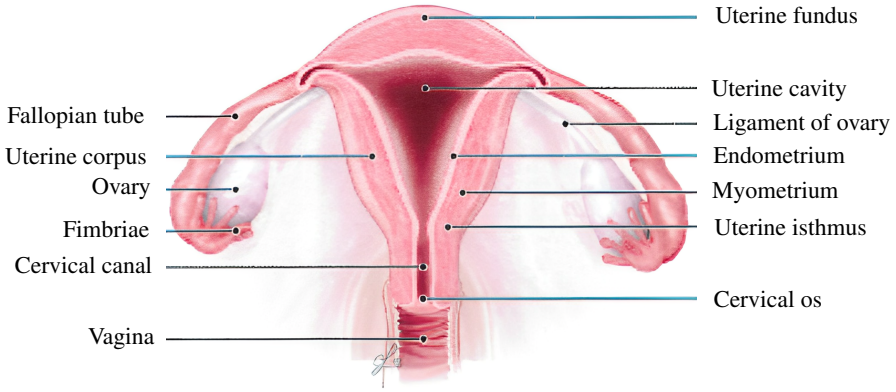


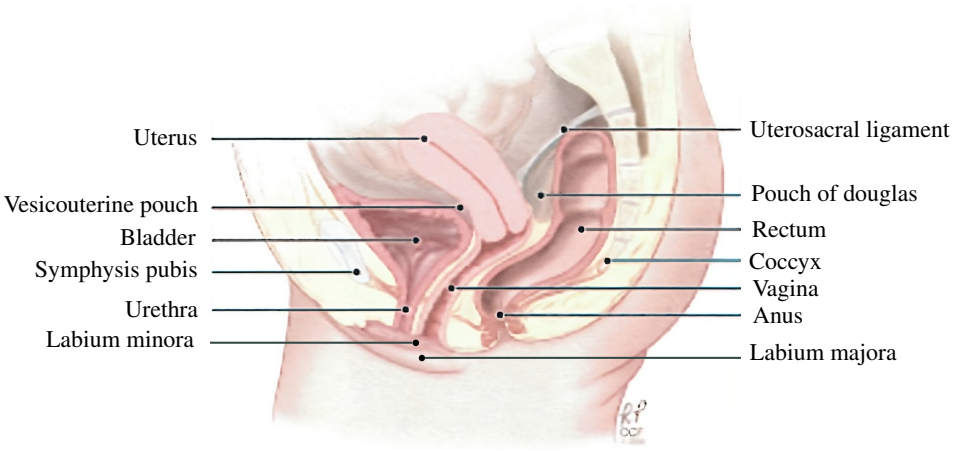
Figure 2.3: Incidence (a) and mortality (b) of cervical cancer by regions in Portugal, 2015. Adapted from [66]

2.2 Female Reproductive System Anatomy

The female genital system is located deep inside the pelvic cavity and it is composed of internal genitalia, which includes fallopian tubes (uterine tubes), ovaries, uterus, cervix, and vagina, and by external genitalia, which includes perineum, labium minora, labium majora, vulva, mons pubis, clitoris, urethral meatus, vestibule and urethral meatus (Figure 2.4) [188].



(a) Coronal view of the internal organs of the female reproductive system.



(b) Sagittal view of the internal organs of the female reproductive system.

Figure 2.4: Coronal and sagittal view of the internal female reproductive system organs. Adapted from [188].

The uterus, in a nonpregnant woman, is a muscular hollow, thick-walled with an inverted pear shape located within the pelvis in a position anterior to the rectum and next to the bladder. The uterus is composed of two parts: the body of the uterus (or corpus uteri) and the cervix at the lower end of the uterus. The body of the uterus is above the constriction behind the cervix and under the openings of the fallopian tubes. It is globe-shaped, and it can be retroverted (tilted back) or anteverted (tilted forward). In the case of anteverted, the cervix is bent forward, and the cervical external orifice is targeted to the posterior vaginal wall. On the other hand, when it is retroverted, the cervix goes into the vaginal through an anterior approach, being more complicated to detect during a colposcopy test. The body of the uterus has the major task of holding the fetus during pregnancy and its walls, during labor, contract to help the fetus to be expelled [156, 188].

The cervix or cervix uteri is the lower part of the uterus that divides the body of the uterus from the vagina. It usually measures 2 to 3 cm long, however, these dimensions change during pregnancy. It has a cylindrical shape that works like a canal that allows the access of semen into the uterine to fertilize an egg cell. There are two orifices in the cervix: the internal os and the

external os. The internal os opens into the endometrial cavity and is responsible for dilating during labor, allowing fetus delivery. The cervix is separated into endocervix, which is the canal between the internal os and external os, and ectocervix which consists of the lower part of the cervix, constituted by a portion of the cervix projected into the vagina which is visible during a colposcopy screening. The vagina is the lowest portion of the female reproductive system and consists in a fibromuscular canal that extends from the vulva to the cervix. It is responsible for expelling the uterine secretions and the fetus during childbirth. It receives the penis during intercourse. This canal connects internal genitalia with external genitalia, and it is situated between the urinary bladder and the rectum [188].

2.3 Cervix Cytology

As mentioned in the previous section, the cervix is the lower part of the uterus and it is divided into two portions endocervix (Figure 2.5) and ectocervix (Figure 2.6). The two portions are approximately of the same size, however, these portions differ in the lining. Endocervix is lined with columnar epithelium (glandular epithelium) and ectocervix is lined with squamous epithelium. The place where the squamous epithelium meets the glandular epithelium is named squamocolumnar junction (SCJ) [185] (Figure 2.7). This junction is dynamic, it changes in location through age from being more external to internal. Thus, before puberty, SCJ is located very close to the external os (original SCJ). Still, after the first menstruation, and during the first pregnancy, SCJ starts to move to the ectocervix which originates a new SCJ away from the external os. The exposure of columnar epithelium cells to the acidic vaginal environment, caused by SCJ transformation, destroys this epithelium. Thus, metaplastic squamous epithelium is freshly formed to repair the damaged portion of columnar epithelium tissue c. The area of epithelium between the original SCJ or the native and unaffected columnar epithelium of the endocervical canal and the new SCJ is called the transformation zone (TZ) which is also dynamic as SCJ [185].

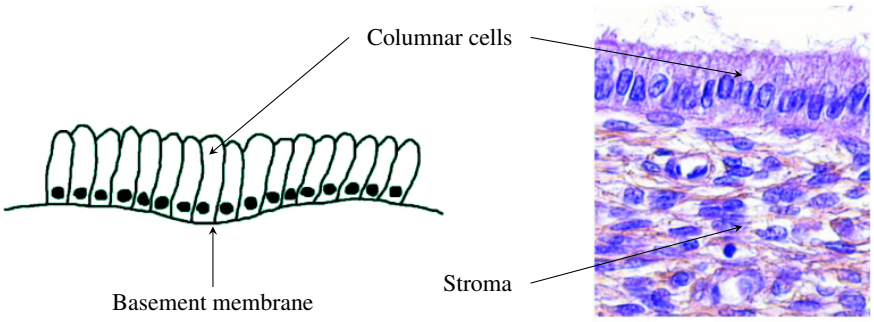


Figure 2.5: Columnar epithelium of the endocervix. Adapted from [185].

The TZ changes in size and position on the cervix. It can be partially or entirely in the endocervical canal. The visibility of TZ determines its type. During a colposcopic examination, it is essential to identify the TZ type, once squamous cervical cancer embodies most cervical cancers,

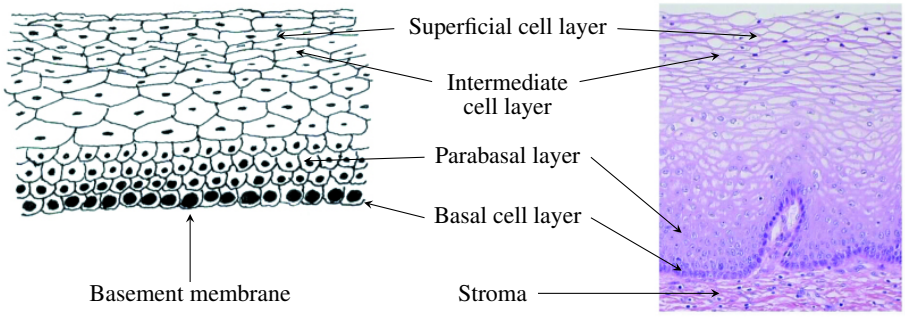
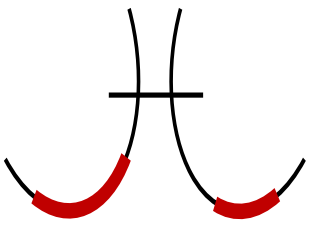
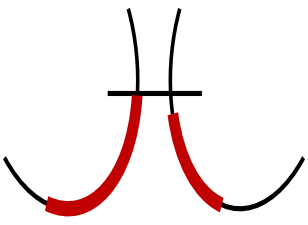
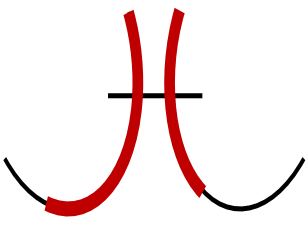
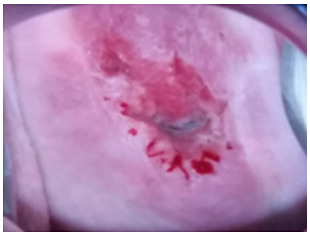
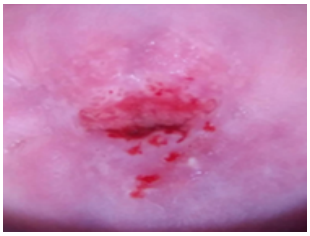


Figure 2.6: Squamous epithelium of the ectocervix. Adapted from [185].

which start in the TZ. Glandular cervical cancer can also be found in the columnar epithelium above this zone or in the TZ. The TZ is a dynamic change that occurs during a woman's life, and during premenopause the transformation zone is most often found on the ectocervix. The TZ after menopause moves into the cervical canal, and the cervix retracts; consequently, the visibility of the TZ is affected. There are 3 different types of transformation zones, as Table 2.1 shows:

- **Type I** – The TZ is fully visible and it is completely ectocervical.
- **Type II** – The TZ is fully visible and has an endocervical component.
- **Type III** – The TZ is not fully visible and has an endocervical component.

Table 2.1: Scheme of the different types of transformation zones and respective examples of colposcopic images. Adapted from [25].

Type I	**Type II**	**Type III**
		
		

Most cervical cancers arise at the TZ, mostly by stratified squamous epithelium metaplasia from a simple columnar epithelium. Thus, it is very important to identify the specific type of

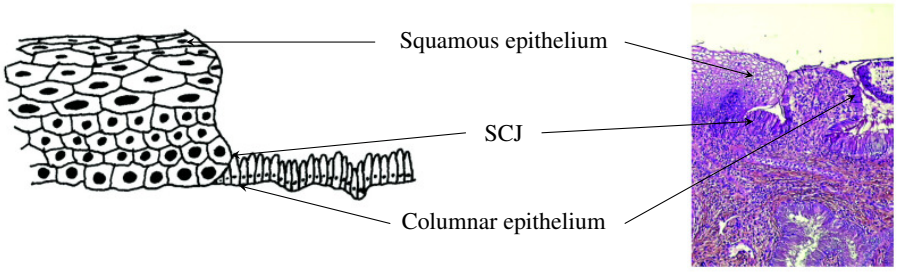


Figure 2.7: Squamocolumnar junction (SCJ). Adapted from [185].

TZ since each type requires different treatment approaches, so, cervix type detection prevents inadequate treatments [69].

2.4 Signs and Symptoms

Cervical cancer is a silent disease, most women do not have any signs or symptoms of pre-cancer. Thus, the appearance of signs and symptoms depends on the stage of the cancer. In most cases, just in the advanced stage of the disease, the patients have symptoms; in the early stage of cervical cancer, most women are asymptomatic. The symptoms are more severe in women with advanced and metastatic cancers, depending on the organs or tissues to which the cancer has spread. The most common signs and symptoms of cervical cancer are:

- Persistent pelvic and/or back pain;
- Menstrual bleeding longer and heavier than usual;
- Bleeding after douching or a pelvic examination;
- Blood spots or light bleeding between or following periods;
- Pain and burning during sexual intercourse;
- Bleeding after menopause;
- Urinary incontinence;
- Swelling of the legs;
- Increased vaginal discharge.

All the symptoms mentioned above depend on the stage of cancer [1]. Cervical cancer, as previously mentioned is a silent disease. Thus, prevention and screening programs are extremely important to reduce the mortality rate, especially in low- to middle-income countries, which have reduced healthcare services.

2.5 Prevention and Treatment

The risk of cervical cancer can be reduced in two different ways: by having regular screenings to find any precancers and by receiving the HPV vaccine. The best way to identify abnormal changes in cervical cells is by doing regular cervical screening tests. The different screening methods will be addressed in Section 2.7. Most of the worldwide vaccination programs use a vaccine called Gardasil 9. This HPV vaccine reduces the risk of some cancers and precancerous lesions both in men and women, covering different nine subtypes of HPV (6, 11, 16, 18, 31, 33, 45, 52, and 58), being the HPV-16 and HPV-18 carcinogenic types. [161, 230]. Although vaccination is an extremely important factor in preventing cervical cancer, it does not guarantee total immunity, so screening tests are vital to complement vaccination as a preventive measure.

The treatment of cervical cancer is strictly dependent on the lesion stage. The most commonly used methods for the early stages consist of the removal of the cervix and/or a part of the entire uterus. In advanced stages, radiotherapy and chemotherapy can be used. The removal procedure to be used depends on the stage, there are five main stages according to the International Federation of Gynecology and Obstetrics (FIGO), which provides a worldwide standardized classification that allows various clinical centers to compare results. The main categories of the FIGO classification are as follows [32]:

Stage 0: Carcinoma in situ with abnormal cells in the innermost lining of the cervix. This stage is treated with local ablative or excisional measures such as cryosurgery, laser ablation, and loop excision; surgical removal is preferred.

Stage I: In this stage occurs an Invasive carcinoma that is strictly confined to the cervix. Thus, the treatment in this stage is surgery, as in stage 0, which can be a total hysterectomy, radical hysterectomy, or a conization procedure.

Stage II: The cancer is locoregional spread beyond the uterus. The treatment comprises a combination of external beam radiation with brachytherapy and radical hysterectomy with bilateral pelvic lymphadenectomy.

Stage III: Cancerous disseminated to the pelvic sidewall or the lower third of the vagina, and/or due to invasion of the ureter occurs hydronephrosis or a non-functioning kidney. In this stage, cisplatin-based chemotherapy with radiation is the most used treatment.

Stage IV: Cancerous spread beyond the true pelvis or into the mucosa of the bladder or rectum. It requires individualized therapy on a palliative basis; radiation therapy is used only for control of bleeding and pain; systemic chemotherapy is used for disseminated disease.

Regarding the main stages presented above, there are several substages inside each stage [32].

2.6 Pathophysiology

The infection with human papillomavirus (HPV) is known to be an essential factor in promoting the development of cervical cancer. A sexually active woman has a high chance of developing HPV infection, however, only 5% of HPV infections will result in cervical intraepithelial neoplasia (CIN) grade 2 or 3 lesions. CIN 3, the precursor to cervical cancer, carries a substantial risk of progressing to cervical cancer, with a 20% likelihood of development within 5 years and a 40% likelihood within 30 years. More than 90% of HPV infections disappear totally in a few months to years without leaving sequels. [107, 166].

Despite the influence of HPV on the development of cervical cancer, other factors also promote it, such as poor immunity (HIV infection and deficient nutritional status), first intercourse in early age, environmental factors (smoking and vitamin deficiencies), genetic susceptibility, lack of access to routine screening and a high number of sexual partners [32]. Concerning genetic susceptibility, the activation of certain genes such as TNFa-857, TNFa-863, TNFa-8, TNFa-572, and TNF G-308A, which are tumor necrosis factors (TNF), increase the incidence of cervical cancer. These genes are responsible for initiating cell apoptosis [33]. The occurrence of Polymorphisms in Tp53, which are engaged in apoptosis and gene repair, have been linked with an expanded rate of HPV infection leading to progression to cervical cancer. Other genes such as human leukocyte antigen (HLA) genes, chemokine receptor-2 (CCR2), and the Fas are also involved in an increased risk of HPV infection progressing to cancer. In contrast, the CASP8 gene is associated with a decreased risk of cervical cancer [107, 166, 219].

There are several HPV subtypes, divided into four different groups. The first group contains the most carcinogenic subtypes, which are: 16, 18, 31, 33, 35, 45, 52, and 58. In Table 2.2 are represented by the four different groups and the respective subtypes with a description of the evidence of the gene for cervical cancer. The HPV alpha group 3 contains types 6 and 11, which have never been connected to cases of invasive cervical cancer.

Table 2.2: Different Human Papillomavirus groups and types related to cervical cancer. Adapted from [85].

HPV Alpha Group	Types	Cervical cancer relation
1	16, 18, 31, 33, 35, 39, 45, 51, 52, 56, 58, 59	Evidence of cervical cancer, being the 16 type the most carcinogenic.
2A	68	Reduced evidence in humans and robust mechanistic evidence.
2B	26, 53, 66, 67, 70, 73, 82, 30, 34, 69, 85, 97	Sufficient or limited evidence in humans.
3	6,11	Insufficient epidemiological evidence and lack of carcinogenic potential in mechanistic experiments.

The integration into the human genome by high-risk HPV DNA places the E6 and E7 genes in

a position of augmented replication. This integration opposes apoptosis, which will directly affect cell growth control. Thus, the growth of cells with damaged DNA will be promoted resulting in malignancy [254].

2.7 Screening Methods

Screening programs are implemented in most developed countries and the process follows a set of guidelines that differ according to the country where the screening is performed and the patient's age. They are responsible for a reduction in worldwide mortality by cervical cancer. According to FIGO guidelines, there are seven different screening procedures: the Papanicolaou (Pap) test (cytology), colposcopy, biopsy, conization of cervix, cystoscopy, proctosigmoidoscopy, and chest x-ray. However, the most used screening methods are the Papanicolaou test, colposcopy and biopsy [32, 169]. The screening pipeline for cervical cancer in Portugal starts with an HPV test; if the result is positive, the guidelines request a cytology examination. If the HPV test result is positive for types 16 and 18, even if cytology is negative the colposcopy is always performed. The biopsy and histological examination are the last steps if the outcome of the colposcopy is inconclusive. The patients with suspicious symptoms and/or inconclusive outcomes are in most cases followed up to prevent any malign progression. Figure 2.8 shows the most common workflow for cervical cancer diagnosis in Portugal, however, this can vary depending on the hospital.

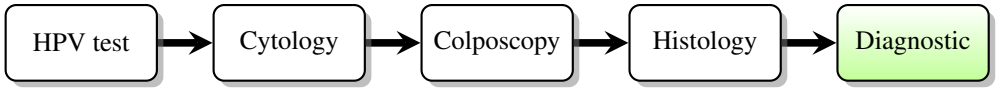


Figure 2.8: Cervical cancer diagnosis workflow in Portugal.

2.7.1 Cytology

Cervical cytology or Papanicolaou test or Pap smear is the oldest procedure used for cervical cancer screening, it is used to detect abnormal or potentially abnormal cells from the uterine cervix [169]. First, it is necessary to exfoliate cells from the transformation zone of the patient cervix to observe on the microscope. In the most traditional technique, the cells are transferred directly to the glass slide (smear) which may cause the inclusion of blood and other artifacts that make the task more difficult and perturb the slide image quality that is seen on the microscope. Nowadays, in many places it is used the liquid-base cytology (LBC) which involves the releasing of the cells into a preservative liquid, with the aim of uniformizing the sample and diminish the influence of external artifacts (e.g. blood, debris) [170]. LBC compared with the traditional method is more discriminative which promotes a better detection of abnormalities. The cervical cytology methods follow The Bethesda System (TBS) [212], which is a widely accepted reporting system among the medical society. According to TBS, squamous cell abnormalities detected can be separated into four groups: atypical squamous cells (ASC), low-grade squamous intraepithelial lesions (LSIL), high-grade squamous intraepithelial lesions (HSIL), and squamous cell carcinoma.

TBS details morphological variations related to each group, for instance, the nucleus becomes larger and more irregularly shaped or changes in the nuclear/cytoplasm size ratio.

Nowadays, there are portable and low-cost devices such as μSmartScope (Figure 2.9 (a)) from FhP-AICOS [1], allowing access to this screening method in low and middle-income countries with a vulnerable population. This 3D-printed device is portable and can easily be attached to smartphones to acquire microscopic images from cytology preparations. µSmartScope is used for Liquid-based cytology (LBC) smears analysis which has huge potential to avoid exhaustive and time-consuming activity of manual microscopic examination [194].

2.7.2 Colposcopy

Colposcopy consists in the use of a colposcope to capture images of the cervix to clarify inconclusive cytologic findings and give an assessment of the location, size, and extent of a lesion [91]. Posteriorly the images are analyzed by a professional to evaluate the severity of the lesion. As in cytology screening, nowadays, there are portable and low-cost devices such as EVA COLPO (Figure 2.9 (b)) from MobileODT [2], which allows the access of the vulnerable population to screening methods.

The colposcopy screening protocol, recommended by WHO, is divided into four different modalities: macroscopic observation, green filter, Hinselmann, and Schiller. In the first modality, a normal saline solution is applied to the cervix to highlight landmarks of the transformation zone for a macroscopic observation [205]. The external boundary of the TZ is defined by landmarks, which can be crypt openings and/or nabothian follicles. In the case of the crypt, the opening is substantially large and the acetowhite area is denser than normal, which may suggest an extension of the neoplasia into the crypt, which occurs for high-grade lesions. In this stage, the squamous epithelium has a pink tone and the columnar epithelium has a dark red tone with grape-like shapes [205].

The second modality consists of the application of a green light filter in order to enhance and see more clearly the blood vessels since cancer cells have a higher growth rate than usual which demands more vascularization of the tissue [24]. The Hinselmann modality, which is the third, involves the use of 5% acetic acid solution to enhance squamous and columnar epithelium, enabling the segmentation of these regions by the clinician expert. The use of acetic acid solution also facilitates the observation of precancerous lesions. The last modality is Schiller's test, which uses an iodine solution directly applied to the cervix. The color changes in this modality are very important, the normal squamous epithelium stain shows a black or brown tone, while the immature squamous metaplastic epithelium remains with the same color [205]. The staining velocity is also important since faster reactions suggest fragile tissues which may be related to neoplasias. Iodine does not stain some abnormal patterns such as cervical polyps, which allows the perception of the human eye of normal and abnormal regions in the TZ [185].

[1] https://www.aicos.fraunhofer.pt/en/our_work/portfolio/micron.html
[2] https://www.mobileodt.com

During colposcopy, the specialist pays attention to abnormal regions including acetowhite regions, abnormal vascularization, mosaic regions, and punctations. The presence of abnormal regions may be a sign of some neoplasia; consequently, the correct identification of these lesions is very important in cervical cancer diagnostic [24].



(a) μSmartScope [194]

(b) EVA COLPO developed by MobileODT.

Figure 2.9: Two examples of portable low-cost devices for cytology (a) and colposcopy (b).

2.7.3 Histology

Cervical histology remains the golden standard in cervical cancer diagnosis, playing a crucial role. It involves the examination of tissue samples taken from the cervix to study the cellular changes and patterns indicative of pre-cancerous or cancerous conditions. There are five key reasons why cervical histology is so important in cervical cancer diagnosis:

1. **Early Detection of Pre-cancerous Lesions** - Cervical Histology aids in locating abnormal alterations in cervix cells known as cervical dysplasia or cervical intraepithelial neoplasia (CIN). These alterations can be identified through a histological examination and are thought to be precursors to cervical cancer. These precancerous lesions can be identified early, allowing for prompt intervention and treatment to stop the development of invasive cervical cancer.

2. **Confirming the Presence of Cancer** - Cervical histology is essential to diagnose the presence of cervical cancer. A biopsy is frequently carried out to obtain a tissue sample when suspicious lesions are found during a cervical screening routine, such as a Pap smear or colposcopic exam. The presence and extent of cancerous cells in the cervical tissue are assessed using histological analysis of this sample.

3. **Determining Cancer Stage and Grade** - Histology determines the stage and grade of cervical cancer. The stage describes how far the cancer has gone, from being restricted to the

cervix (early stage) to including surrounding tissues or distant sites (advanced stage). The grade refers to the level of cellular abnormality and aggressiveness of the cancer cells.

4. **Guiding Treatment Decisions** - Results from cervical histology are crucial in determining the best strategy for treatment. Treatment options for cervical cancer can include surgery, radiation therapy, chemotherapy, or a combination of these modalities, depending on the stage and grade of the disease. Nevertheless, the choice of therapy cannot be based only on the results of the histology examination. Decisions must be made after analyzing the colposcopy and histology findings as well as the patient's clinical context [217].

5. **Monitoring Treatment Response and Recurrence** - Histology plays a role in monitoring the response to treatment and detecting cancer recurrence. Repeat histological examination of tissue samples may be performed after treatment to assess the presence of residual cancer cells. Additionally, follow-up histological evaluations can be necessary in post-treatment surveillance to identify any signs of recurrence or persistence of pre-cancerous lesions.

Cervical histology provides critical information that helps healthcare professionals tailor appropriate interventions and improve patient outcomes. Nowadays, pathology departments are increasingly gravitating towards digital pathology technologies for all or some of their diagnostic output [78]. Nevertheless, despite this revolutionary change, actual data indicate that only a few pathology laboratories globally have adopted a fully digital approach to the histology workflow. There have been several arguments that explain why they continue to employ the conventional "analog" workflow such as high costs, a lack of validation data, quality control issues, and the potential "threat" that this kind of implementation poses to pathologists [88]. On the other hand digital pathology presents major advantages such as safety, efficiency, easy access to expert pathologists, second opinions, implementation of automated algorithms, and also the research boost promoted by the increasing of scanned slides. Thus, to promote the dissemination of digital pathology among clinicians the industry and research community are tackling several challenges including technical aspects (speed, file sizes, storage management, interruptions, rescan rates, focus, and artifacts detection issues), workflow aspects (handling and usability, continuous scanning, integration in hospital network, and file format), and medical aspects (the quality of scans including special stains and the usability of the viewer) [61].

The general workflow of digital cervical histology can be divided into four steps, starting with the tissue sample collection, where a small piece of tissue is typically obtained using a process known as a biopsy. The biopsy can be carried out in a number of ways, such as a cone biopsy, an endocervical curettage, or a biopsy guided by colposcopy [98].

In the second step, the tissue processing/grossing encompasses several sub-steps including fixation, grossing, macroscopic report, chemical processing, sectioning, slide mounting, staining, and manual fragments assessment/counting. In the first sub-process, fixation, the tissue is immersed in formalin solution in order to keep the tissue sample's structure and cellular specifics after collection. To keep the tissue sample's structure and cellular specifics after collection, it must be

kept. Immersing the tissue in a formalin solution is the most typical fixation technique. The use of formalin fixation enables long-term storage, eliminates any potentially contagious pathogens, and assists in preventing tissue degeneration. After fixation, a macroscopic report must be done indicating the number of fragments presented in the tissue block. Then, the tissue starts the chemical processing where the tissue is dehydrated using a series of alcohol solutions to remove water and infiltrated with a clearing agent, such as xylene or similar substances, to make it transparent. The tissue is then embedded in a solid medium, usually paraffin wax, to enable the next sub-step which is the sectioning. Next, the embedded tissue block is cut into thin sections using a microtome, which produces slices of tissue that are usually around 4-6 micrometers thick. Next, the thin tissue sections are mounted onto glass slides, and any excess wax is removed, allowing for the enhancement of cellular details. Finally, before the slides are ready for scanning occurs the staining process where the tissue sections on the glass slides are stained to enhance cellular details and make them visible under a microscope. The most common staining method for cervical histology is the hematoxylin and eosin (H&E) staining technique. Hematoxylin stains the cell nuclei blue, while eosin stains the cytoplasm and other structures pink, allowing for better visualization and identification of abnormal cellular features. In some labs, before the scanning of the slide, a new quality control checkpoint is done, by verifying again the number of fragments to avoid any lost of valuable material during the previous sub-steps [98].

The third step consists of the digitization/ scanning of the tissue slides already processed in step two. In this step, the slide is digitized using advanced slide scanning technologies and creating the whole slide image (WSI) file [42].

In the last step, the diagnosis is done. Based on the detailed examination, the pathologist generates a comprehensive report that includes the diagnosis, stage, grade, and other relevant information. The report is then sent to the healthcare provider who requested the biopsy, who will use this information to guide further management and treatment decisions. Figure 2.10 summarizes the digital clinical workflow from the tissue sample collection to the diagnosis.

2.8 Summary

Although the prevalence of cervical cancer has been declining, middle- and low-income nations continue to see a significant number of cases and fatalities related to this disease. To enable the most vulnerable people to receive health care services, it is crucial to create low-cost screening techniques. Once they are affordable and portable, digital microscopy and colposcopy devices could potentially circumvent this restriction. However, the use of these kinds of devices raises some challenges that need to be solved, such as quality assurance and standardization: maintaining consistent quality in screening services is vital for accurate diagnosis and effective treatment; ethical and privacy considerations: the use of digital microscopy and colposcopy devices involves capturing and transmitting sensitive medical information. Ensuring patient privacy, informed consent, and data security are paramount. Implementing robust data protection measures, adhering to ethical guidelines, and obtaining informed consent from patients are essential steps to address

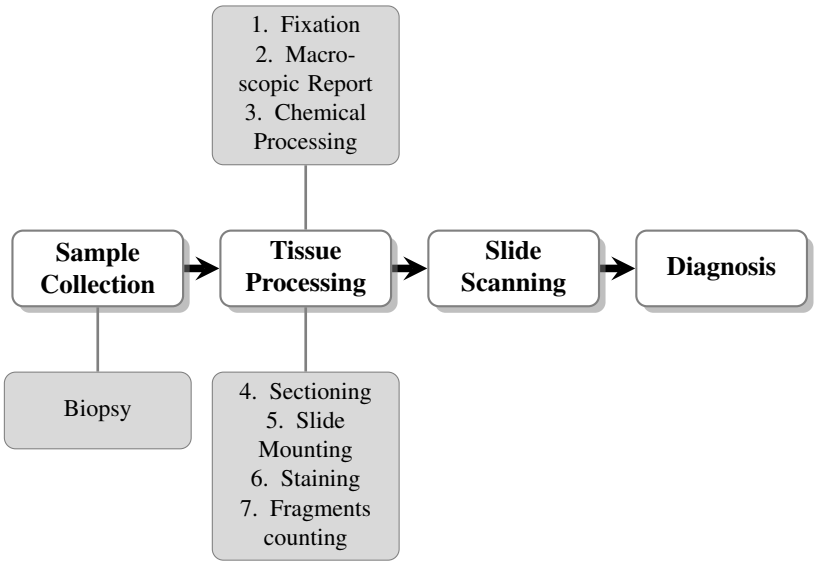


Figure 2.10: Digital Cervical histology workflow.

these concerns; integration into healthcare facilities: to effectively address the burden of cervical cancer, low-cost screening techniques should be integrated into existing healthcare systems. This integration requires coordination among various stakeholders, including policymakers, healthcare providers, and NGOs. Strategies should be developed to incorporate these technologies into national cancer screening programs, establish referral pathways for follow-up care, and ensure the availability of treatment facilities for diagnosed cases. By addressing these challenges, cervical cancer screening can be expanded in middle- and low-income countries by using low-cost screening methods such as digital microscopy and colposcopy equipment. In the context of histology, the shift to digital pathology facilitates remote consultations with experts and promotes standardized diagnosis. However, it necessitates integration into healthcare systems, appropriate training, and cost-effective implementation. Additionally, it generates valuable research data, presenting a comprehensive approach to enhance screening and alleviate the burden of cervical cancer.

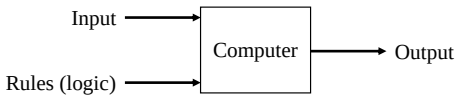
Chapter 3

Machine Learning Fundamentals

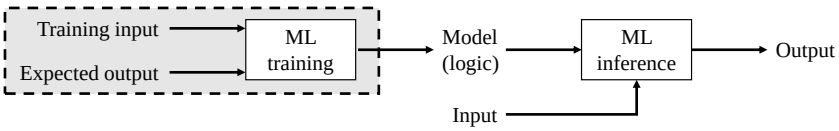
The use of Deep Learning (DL) has gained significant momentum in cervical screening, becoming the gold standard for Computer-Aided Diagnosis (CAD) systems utilizing screening images. These advanced CAD systems aim to provide valuable tools that assist healthcare professionals during cervical cancer screening. Within various screening modalities, the research community has focused on developing CAD systems for diverse tasks, including image quality assessment, image enhancement, segmentation of regions of interest (RoI), image registration, detection of abnormal regions or patterns, identification of cellularity abnormalities in cytology, classification of TZ (transformation zone) types, and assessing the risk of cancer. This chapter presents a concise overview of Machine Learning techniques, with a particular emphasis on DL. Furthermore, it addresses the different datasets available in the literature for CAD system development, encompassing cervical cytology, colposcopy, and histology.

3.1 Machine Learning Algorithms

In 1959, Arthur Samuel, while working on artificial intelligence for the game of checkers, introduced the term "machine learning." He described it as a "field of study that empowers computers to learn autonomously without requiring explicit programming." This distinction sets machine learning apart from pre-programmed agents that lack reasoning capabilities and simply follow predefined instructions [238]. Machine Learning (ML) is a sub-field of artificial intelligence (AI) dedicated to studying and proposing models capable of learning without explicit instruction, autonomously, being only dependent on the data and its patterns. On one hand, in the traditional programming paradigm, the programmer uses some data as input and then applies an algorithm written by her to that data and outputs and the results of the computations included in the computer program. On the other hand, machine learning systems take as input the data and also the desired output (ground truth) and learn a new model that represents the learned information from the input data [232]. Figure 3.1 shows the differences between both paradigms (traditional programming versus machine learning).



(a) Traditional programming.



(b) Machine learning.

Figure 3.1: A comparison scheme between the traditional programming paradigm (a) and machine learning (b).

ML systems can be divided into six different main branches: supervised, unsupervised, weakly-supervised, semi-supervised, self-supervised, and reinforcement learning. In supervised learning, the ML model is fed with input data and labels for each data instance. With the training data, X, and its respective labels, Y, the goal of the ML model is to learn the function that best estimates the unknown mapping function of X into Y, $Y = f(X)$. Thus, the model learns to map input features to the correct output by generalizing from the provided labeled examples. Unsupervised learning involves training models on unlabeled data without explicit target labels. The goal is to discover underlying patterns, structures, or relationships within the data. This can include tasks such as clustering, dimensionality reduction, and generative modeling. Weakly supervised learning falls between supervised and unsupervised learning. It deals with partially labeled data where the labels may be noisy, incomplete, or of lower granularity. The model learns from these imperfect labels to make predictions or infer patterns. Semi-supervised learning combines labeled and unlabeled data for training. It leverages a small amount of labeled data along with a larger amount of unlabeled data to improve learning performance. Self-supervised learning is a type of unsupervised learning that leverages supervised signals derived from the data itself. The model learns by solving pretext tasks, where it predicts certain aspects of the data, such as missing patches, rotations, or context. The learned representations can then be transferred to downstream tasks. The last paradigm is reinforcement learning, which involves an agent learning to interact with an environment through trial and error. The agent receives feedback in the form of rewards or penalties based on its actions. The objective is to find an optimal policy that maximizes the cumulative reward over time [87, 157].

The size and quality of the available data are pivotal factors in selecting suitable machine learning algorithms. In scenarios with limited datasets, classical machine learning algorithms typically outperform DL models. However, as these algorithms cannot directly process images, it becomes necessary to extract features from the images and employ them as inputs for the classifiers. In contrast, when a large volume of data is accessible, DL models often yield superior performance.

DL is a sub-field of machine learning that uses neural networks with many hidden layers to map a certain input to an output of predefined characteristics. The models have the intrinsic ability to learn useful features directly from the input data that are important for the task at hand. This may be classification, regression, detection, or segmentation. The networks are composed of several layers each with several nodes. The large numbers of hidden layers confer depth to the networks, and that is where the name comes from. Each node performs a weighted sum of its inputs and then feeds it to a non-linear function whose result is carried forward as input to the following layer and so on until the final layer where the network output is provided.

In the field of cervical cancer diagnosis, various machine-learning models have been employed in the literature. These include well-known algorithms such as K-Nearest Neighbors, Logistic Regression, Random Forest, Support Vector Machines, Multilayer Perceptron (MLP), and several others. These classical machine learning models have been utilized successfully in scenarios where the dataset size is limited and feature engineering is necessary. However, in recent years, with the advent of DL and the availability of substantial amounts of data, DL models have emerged as the gold standard. DL models, particularly Convolutional Neural Networks (CNNs), have demonstrated remarkable performance in cervical cancer diagnosis when provided with large-scale datasets. DL models have the ability to automatically learn complex patterns and features directly from raw data, mitigating the need for explicit feature engineering. Consequently, they have gained prominence in the field due to their capability to handle massive amounts of data and deliver superior performance in various medical image analysis tasks. In the upcoming subsection, we will provide a comprehensive definition of DL and present several illustrative examples of DL models [87, 157].

3.1.1 Feed-forward Neural Networks

The feed-forward neural network model has gained widespread popularity and extensive utilization, particularly in the realm of deep learning. In this section, we present a formal definition of the neural network model, accompanied by comprehensive insights into the training procedure and effective strategies for mitigating common challenges encountered in neural networks. An artificial neural network, referred to as a feed-forward artificial neural network or multi-layer perceptron (MLP), is a machine learning model that processes its input vector through a series of functional transformations [33]. To create an artificial neural network with a single hidden layer containing M nodes, we establish M linear combinations in the following manner:

$$a_j = \sum_{j=1}^{M} \theta_{ji}^{(1)} x_i + \theta_{j0}^{(1)} \tag{3.1}$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_D)$ is the D-dimensional input vector and $j = 1, \ldots, M$. The variables $\theta_{ji}^{(1)}$ and $\theta_{j0}^{(1)}$ are referred to as weights and biases, respectively. The superscript (1) indicates the weights that parameterize the first hidder layer and the suffix ji represents the direction from input i to hidden unit j. The total set of weights and biases is defined as $\Theta = \left\{\theta^{(1)}, \ldots, \theta^{(M)}\right\}$, where M

is the number of layers and $\theta^{(m)}$ contains the weights $\theta_{ji}^{(m)}$ and biases $\theta_{j0}^{(m)}$ for the layer m. The outcomes a_j obtained in a layer are referred to as its activations. These activations transform by applying a non-linear, differentiable activation function h. Among the commonly used activation functions are the sigmoid function, the tanh function, and the widely acclaimed Rectified Linear Unit (ReLU) function. The ReLU function, introduced by Nair and Hinton in 2010 [162], has gained significant popularity due to its superior performance and generalization abilities compared to the sigmoid and tanh functions. ReLU acts as a threshold function by setting negative values to zero while retaining positive values. Its mathematical definition is as follows:

$$f(x) = \max(x, 0) = \begin{cases} x, & \text{if } x \geq 0 \\ 0, & \text{otherwise} \end{cases} \tag{3.2}$$

There are also other activation functions available, such as the Leaky ReLU and ELU (Exponential Linear Unit), each with its own advantages and use cases. The purpose of applying an activation function is to introduce nonlinearity into the neural network, allowing it to learn and represent complex patterns in the data. The output activations of the hidden layer are determined by yet another linear combination referred to as:

$$a_k = \sum_{j=1}^{D} \theta_{kj}^{(2)} z_j + \theta_{k0}^{(2)} \tag{3.3}$$

where $k = 1, \ldots, K$, where K is the number of outputs of the neural network. Subsequently, the output activations a_k undergo a transformation by applying the last activation function suitable for the task at hand. The specific choice of the last activation function is contingent on the nature of the problem being addressed. The sigmoid function is commonly used in binary classification tasks because it maps the input to a value between 0 and 1, which can be interpreted as the probability of belonging to one of the two classes. It can be defined as follows:

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{3.4}$$

It is suitable for binary classification because it provides a clear separation between the two classes. On the other hand, the softmax function is typically used in multi-class classification problems where there are more than two classes. The softmax function normalizes the output values across all classes, ensuring that they sum up to 1. This allows the output values to be interpreted as class probabilities. In multi-class classification, each class is assigned a probability based on the softmax function, and the class with the highest probability is chosen as the predicted class. Softmax can be defined as:

$$\text{softmax}(x_i) = \frac{e^{x_i}}{\sum_{j=1}^{N} e^{x_j}} \tag{3.5}$$

The process of training neural networks involves finding the optimal weights of the model through the optimization of a loss or cost function. In the case of regression tasks with a single

target variable, a commonly used loss function is the sum-of-squares error function. This loss function quantifies the discrepancy between the predicted values and the actual target values. In classification tasks, the training procedure for neural networks also revolves around optimizing the model's weights by minimizing a loss or cost function. However, the choice of loss function differs from regression tasks. For classification tasks with multiple target classes, a standard loss function used is the cross-entropy loss. The cross-entropy loss measures the dissimilarity between the predicted class probabilities and the true class labels. It penalizes the model more when it assigns a high probability to incorrect classes and a low probability to the correct class. During the training process, the neural network's weights are adjusted iteratively using optimization algorithms such as gradient descent or its variants. These algorithms aim to find the set of weights that minimizes the loss function, resulting in a model that can accurately classify inputs into their respective classes.

Gradient descent is the prevailing optimization approach for training neural networks, utilizing the gradient information of the loss function. The weights are initialized following a specific strategy, and subsequent updates are made by moving in the direction opposite to the negative gradient, as represented by the equation:

$$\Theta^{(\tau+1)} = \Theta^{\tau} - \eta \nabla E\left(\Theta^{\tau}\right) \tag{3.6}$$

where τ denotes the time step, η is the learning rate hyperparameter, and $E\left(\Theta^{\tau}\right)$ refers to the error. The gradient of the error with respect to the weights is calculated at each step. Backpropagation, also known as backprop, enables the application of gradient descent in feedforward neural networks. It allows information to flow both forward and backward within the network. During the forward pass, the network calculates the output for a given input by applying linear combinations and non-linearities. The backward pass involves starting from the output layer and computing the gradient of the error function with respect to the local weights or activation functions. This gradient is then propagated backward through the network using the chain rule. Various optimizers have been developed to enhance the gradient descent approach. Examples include stochastic gradient descent (SGD), stochastic gradient descent with momentum, AdaGrad, and Adam.

3.1.2 Convolutional Neural Networks (CNNs)

In 2012, in a landmark moment in the history of machine learning and computer vision, researchers from the University of Toronto achieved an impressive result in the ImageNet competition conquering first place by a large margin to all other teams [124]. They used a Convolutional Neural Network (CNN) several layers deep and were able to nearly halve the error of the previous best result. CNNs had been first used over two decades before when LeCun et al. applied them to the task of recognizing hand-written digits in 1990 [43]. However, the outcomes and interest in the machine learning community for CNNs were not substantial and these were mostly abandoned until the introduction of AlexNet in 2012. Some reasons for this were the fact that learning feature extractors directly from data with little prior knowledge were considered infeasible and the belief

that simple gradient descent would get trapped in poor local minima [128]. Some advancements kept trickling in such as the discovery that Rectified Linear Units (ReLU) allow for much faster neural network training [162] and the development of very fast GPUs, which culminated in the success described above at the 2012 ILSVRC, possible by the use of other techniques such as data augmentation and dropout as regularization [218].

In a CNN, at least one layer of the network performs a convolution operation on that layer's inputs. This is performed by means of a filter or kernel, which is translated across and down the input matrix to produce a representation map of the original image. The filter is composed of learnable parameters, which serve as weights for a weighted sum of a patch of the input before each translation. These weights are updated through gradient descent, just like the parameters in a fully connected neural network. The difference is that the filter is the same for every part of the image (at each convolutional layer), such that any features can be extracted irrespective of their location in the image. This provides location invariance as well as a very efficient pattern recognition method because the number of parameters is greatly reduced when compared to a fully connected layer with the same capability. This operation is very successful because it can correlate a certain pixel's information with that contained by adjacent pixels. In subsequent layers, pooling operations are usually applied, which downsample the image to a smaller resolution, allowing for more convolution operations to capture an ever larger representation of the original image - the receptive field of the network is thus increased. If enough layers are added, at a certain point the whole image (or a large part thereof) can be captured by a single convolutional filter, and high-level information is properly processed by the network. The first layers' filters typically highlight basic lines in different orientations, and progressively deeper layers are activated by ever more abstract representations until the whole image's information is contained in a feature vector. For the task of image classification, this vector's points are connected to a fully connected layer and then to a final layer with one or more nodes (one for each output class) which acts as the classifier.

There are several described CNN architectures in the literature that have been highly influential and widely used in recent years, such as:

- **VGG (Visual Geometry Group)**: VGG networks first proposed by K. Simonyan et al. [211], particularly VGG16 and VGG19, are known for their simplicity and uniform architecture. They consist of multiple layers with small 3x3 convolutional filters, making them easier to understand and implement.

- **ResNet (Residual Network)**: ResNet introduced the concept of residual learning, allowing for the training of very deep neural networks and tackling the problem of vanishing gradients [95] by reusing activations from a previous layer. Variants such as ResNet50, ResNet101, and ResNet152 have been extensively used in various computer vision tasks. The authors explicitly reformulate the layers to learn the residual function with reference to the layer inputs, instead of learning unreferenced functions. It uses a simple building block with skip connections for that same reason.

- **Inception (GoogLeNet)**: Inception architectures, such as InceptionV3 and InceptionResNetV2, incorporate the Inception module, which uses parallel convolutions with different filter sizes. These architectures strike a balance between computational efficiency and high performance [220].

- **MobileNet**: MobileNet architectures are specifically designed for mobile and embedded devices. They utilize depth-wise separable convolutions (1x1 convolutions) to reduce the computational complexity while maintaining good accuracy. This network has lightweight architectural features. Variants like MobileNetV2 and MobileNetV3 have been developed with further improvements [104].

- **EfficientNet**: EfficientNet is a recent advancement that focuses on achieving high accuracy with efficient use of resources. It introduces a compound scaling method to balance the depth, width, and resolution of the network. A larger kernel works better when the information is globally distributed, although if the information is locally distributed, a smaller kernel is preferred. Consequently, choosing the size of the kernel can restrict the performance of the algorithm. Inception architectures try to overcome this challenge by employing different convolutions to the same object and concatenating the results in one single output. EfficientNet models, such as EfficientNet-B0 to EfficientNet-B7, have achieved state-of-the-art performance on various benchmark datasets [222].

The field of computer vision is constantly evolving, and new architectures and techniques continue to emerge. Researchers often propose novel architectures or adapt existing ones to address specific challenges or improve performance on specific tasks.

3.1.3 Multimodal Machine Learning

Experimental evidence shows that humans take advantage of multimodal sensorial experiences even when tasks are meant to be unimodal [141]; furthermore, multiple modalities allow systems to build robustness to the loss of one modality. Thus, multiple types of information for diagnosis are clinically desirable. Multimodal learning deals with different types of data (text, images, signals, etc.) to help in the model development [118]. When a dataset has also information about clinical data of the patient or other medical exams, this data can be incorporated into the model, constituting a Multimodal learning model. Regarding the use of multimodal approaches in cervical cancer diagnosis, there are few efforts in this area, some works use cervical images and additional clinical data such as Pap smear results, Human Papilloma Virus (HPV) test and patient age [50, 214, 243, 245]. Clinical data have a clear impact on screening decisions, which was confirmed by the performance boost that literature models demonstrate compared with classical approaches. There are clear limitations to these approaches, mainly regarding the datasets which have an absence of data and lack of interpretability in those models that may influence the confidence of medical staff when using them.

3.1.4 Explainable Artificial Intelligence

AI is present in our daily lives, from autonomous driving to film recommendation systems. However, this increasing importance of AI also raises a new hot topic, which is the Explainable AI (XAI). It is important to understand the reasons behind the decisions of these systems, so, XAI proposes a new paradigm that shifts towards transparent AI [3]. In the last ten years, crucial steps have been made to bring AI successfully into the medical field. Nevertheless, most clinicians are still far away from using AI in daily clinical practice [191]. The black boxes of AI models reduce their convincingness to clinicians. Thus, the interpretability and explainability of AI models are strictly necessary to increase the confidence of clinicians. In the case of cervical cancer, more details such as the severity of the cervical lesions, the place of the image where the model is focused, and even detailed diagnostic evidence should be provided by the AI models. This will make it simpler for medical and technician staff to recognize and accept these AI models in clinical practice and also in the near future, the AI models for diagnostics, could turn into a teaching tool in medical schools or hospitals [247]. However, the cooperation between medical AI experts and clinicians is vital to making substantial progress in both fields.

3.2 Datasets

To develop, validate, and measure the performance of any CAD system for Cervical Cancer diagnosis it is necessary to have a dataset, which in most cases contains labeled medical images and, in some cases, clinical information about the patients. The tiny number of samples in the datasets of cervical cancer presented in the literature, even meticulously reviewed, is not usually adequate for large validation of CAD systems. Thus, the good performance of a decision support system is strictly dependent on the availability of data. The dataset must contain a large diversity of abnormalities and acquisition settings in order to cover all the possible cases. The utility of each dataset also depends on the annotations provided by the authors who collected the data. The next sections will present the most relevant datasets for the three most important modalities of diagnosis in cervical cancer: cytology, colposcopy, and histology.

3.2.1 Cytology

Concerning cytology datasets there are nine different public datasets available in the literature: Herlev, ISBI14, ISBI15, Risk-Factors, CERVIX93, SIPaKMeD, Hussain, Cx22, and EDoF Fraunhofer dataset. A brief description of each dataset can be found in Table 3.1.

3.2.2 Colposcopy

Currently, there are five available datasets regarding colposcopy images. A brief description of each dataset can be found in Table 3.2 concerning the content, the purpose of the collection, and the annotations available. The most complete dataset present in the literature belongs to the American National Cancer Institute (NCI) from the National Institutes of Health (NIH), which has

Table 3.1: Publicly available cervical cytology datasets in literature.

Dataset	Year	Number of Instances	Type	Description
Herlev [113]	2005	917	Image	Single-cell images of 7 different classes with segmentation ground-truth.
ISBI14 [146]	2014	16 EDF + 945 Synthetic	Image	Extended depth field (EDF) images and synthetic images containing cells with different overlapping degrees. It is also included the segmentation of nuclei and cytoplasm.
ISBI15 [147]	2015	17 EDoF	Image	EDF images containing cells with different overlapping degrees and respective fields of view.
Risk-Factors [74]	2017	858	Text	Patients information and medical history.
CERVIX93 [178]	2018	93 EDoF	Image	Classification divided in seven classes. Only segmentation of nuclei points.
SIPaKMeD [182]	2018	4049	Image	Isolated cell images that have been manually cropped from 966 cluster cell images of pap smear slides.
Hussain [108]	2020	963	Image	Liquid-based cytology pap smear images with 4 classes according to TBS.
Cx22 [139]	2022	1320	Image	14,946 annotated cervical cellular instance labels.
EDoF Fraunhofer [200]	2023	144 EDoF	Image	144 EDoF cytology images with respective 5 stacks for each EDoF.

collected the dataset in several regions of North and South America. This dataset is composed of digital colposcopy images (cervigrams) from $12,686$ women with more than $76,000$ cervigrams. The dataset also includes information about the patient's age, HPV test, and histology results. Some of this data is missing, including the results from histologic examination, which should be the ground truth for cancer risk classification problems, resulting in a total of $74,820$ labeled cervigrams. The annotations for histology results correspond to the neoplasia progression level, being categorized as normal, abnormal, CIN2, CIN3, and cancer. More than 75% of the dataset are low-risk cervical cancer cases creating an imbalance in the data.

Table 3.2: Available colposcopy datasets in literature.

Dataset	Year	Number of Instances	Type	Description
Acosta-Mesa H. et al. [2]	2009	10	Video	Videos of colposcopy examination of 10 patients during Hinselmannn stage. This database has no annotations.
Fernandes et al. [74]	2017	287	Image	The dataset includes 62 predictive attributes for each image, including segmentation masks of cervix area, external orifice, vaginal walls, speculum, and artifacts, and 7 target variables.
NCI/NIH dataset [97]	1990s	76669	Image	The dataset includes cervigrams and information about the patient's age, HPV test, and histology results. The images are divided in 5 classes (normal, abnormal, CIN2, CIN3, and cancer)
Intel & MobileODT [176]	2017	8000	Image	The dataset includes the main colposcopy stages and has annotations about the cervix type regarding TZ location.
MSCI [255]	2021	4753	Image	The dataset includes over 4000 images from 679 patients. They present a multistate dataset, meaning a dataset with different stages of the colposcopy for each patient.

3.2.3 Histology

In the case of histology, there are only two publicly available datasets: MTCHI dataset [153] and Oliveira et al dataset [167]. A brif summary of each dataset can be found on Table 3.3.

Table 3.3: Available histology datasets in literature.

Dataset	Year	Number of Instances	Type	Description
MTCHI [153]	2021	100	WSI	Histology slides from 71 patients with pixel-level annotation of cervical cancer grading.
Oliveira et al [167]	2023	600	WSI	Histology slides without any annotations.

Chapter 4

Quality Control in Medical Decision Support Systems

Foreword on Author Contributions

The results of this work have been disseminated in the form of three papers at international conferences and one in an international journal. Additionally, a novel publicly available data set has been curated and made publicly accessible:

[13] - Tomé Albuquerque, Ana Moreira, and Jaime S. Cardoso, "Deep Ordinal Focus Assessment for Whole Slide Images", 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), Montreal, BC, Canada, 2021, pp. 657-663, doi: https://10.1109/ICCVW54120.2021.00079

[14] - Tomé Albuquerque, Luís Rosado, Ricardo Cruz, Maria João M. Vasconcelos, Tiago Oliveira, and Jaime S. Cardoso. "Rethinking low-cost microscopy workflow: Image enhancement using deep based extended depth of field methods". Intelligent Systems with Applications, 2023. doi: https://doi.org/10.1016/j.iswa.2022.200170

[200] - Ana Sampaio, Tomé Albuquerque, Pedro Faria, Luís Rosado, and Maria João M. Vasconcelos. (2023). "Fraunhofer Portugal AICOS EDoF Dataset" [Data set]. Zenodo. doi: https://doi.org/10.5281/zenodo.7473277

[11] - Tomé Albuquerque, Ana Moreira, Beatriz Barros, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, Jaime S. Cardoso, "Quality Control in Digital Pathology: Automatic Fragment Detection and Counting," 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Glasgow, Scotland, United Kingdom, 2022, pp. 588-593, doi: 0.1109/EMBC48229.2022.9871208

[10] - Tomé Albuquerque, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, Jaime S. Cardoso, "Quality checkpoint in pathology specimens handling: an AI system to automate fragment detection and count". 18th European Congress on Digital Pathology (ECDP), Berlin, Germany, 2022.

Medical Decision Support Systems have emerged as indispensable tools, offering healthcare professionals the means to harness sophisticated algorithms and clinical expertise to make informed decisions that can significantly impact patient outcomes. However, their potential is not

without its challenges and complexities. While these systems hold the promise of revolutionizing healthcare delivery, their effective implementation and utilization require meticulous attention to various critical factors, such as quality control concerns [26].

This chapter encompasses three key subchapters, each delving into intricate aspects of quality control in medical decision support systems. The chapter embarks on its journey by scrutinizing the assessment of image quality, with a pronounced focus on the detection of blur within digital pathology slides. It then proceeds into a comprehensive exploration of innovative approaches aimed to improve the quality of images obtained from low-cost cytological microscopy. In the final part, the chapter culminates by addressing an indispensable daily task in the pathology laboratories: the detection and quantification of histological fragments, a task of paramount importance in the context of routine diagnostics.

4.1 Deep Ordinal Focus Assessment for Whole Slide Images

Medical image quality assessment plays an important role not only in the design and manufacturing processes of image acquisition but also in the optimization of decision support systems. This work introduces a new deep ordinal learning approach for focus assessment in whole slide images. From the blurred image to the focused image, there is an ordinal progression that contains relevant knowledge for more robust learning of the models. With this new method, it is possible to infer quality without losing ordinal information about focus since ordinal losses were used instead of using the nominal cross-entropy loss for training. Our proposed model is compared with other state-of-the-art methods present in the literature. The first conclusion is the benefit of using data-driven methods instead of knowledge-based methods. Additionally, the proposed model is found to be the top performer in several metrics. The best-performing model scores an accuracy of 94.4% for a 12 classes classification problem in the FocusPath database.

4.1.1 Context and Motivation

In the last few decades, there has been a massive growth in the amount of digital color image content in the medical field due to the spread of advanced multimedia devices and digital services capable of acquiring, transmitting, and storage of digital data [57]. Digital Pathology images (Whole slide images (WSI)) are about 10X bigger than Radiology images, being over $>$1 GB in size in most cases. This type of image requires better storage management through their useful life cycle in clinical workflow [53, 168]. Thus, image quality assessment (IQA) methods are crucial, working as a filter in the first stage of acquisition, they can improve storage management. In the second stage, they can be used as an optimizer of the decision support systems.

In the medical field, image quality assessment methodologies are focused on two distinct processes: on the one hand, the low-level notion of quality, which includes the measurement of distortions at a signal level such as blur, noise, compression errors, and other types of distortions [67]; on the other hand, the semantic complex concepts such as the presence/absence of artifacts (e.g.

tissue folds or bubbles, the presence of coloration errors, among others) [27]. Thus, the medical image quality assessment is, in most cases, application-specific, which requires vast domain knowledge in the respective medical area.

Focus quality assessment (FQA) is fundamental in the normal WSIs acquisition workflow. Hence, it is essential to develop and improve FQA methods capable of improving the quality of acquired WSIs and reducing the acquisition time.

In the next Section 4.1.2, a literature review for focus quality assessment in WSIs is provided. A brief summary of the problem and the contributions of this work is made in subsection 4.1.3 The Section 4.1.4 describes the Data Pre-Processing steps, and it also presents a brief introduction to Convolutional Neural Networks, the used Network Architectures, and finally, the losses used during the training. In Section 5.3.3, the experimental details are described, followed by the results and discussion in Section 4.1.5. The study finished with a Conclusions Section 4.1.6.

4.1.2 Related Work

During the acquisition of WSI's by the scanning platforms, focus errors often occur. After the acquisition, manual inspection of the slides is required to infer if it has good quality to proceed with the analysis and diagnostic. The manual inspection of the slides is a time-consuming process and, in most cases, subjective to individual scores, which leads to inter/intra-variability issues between the experts [158]. This way, to automatize this process and improve clinical workflow, several focus quality assessment approaches have been developed. They can be divided into knowledge-based focus quality assessment methods and data-driven focus quality assessment methods.

Knowledge-based FQAs - These methods are based on domain knowledge, with a large presence of this type of algorithms in the literature. Knowledge-based FQAs can be divided into 7 different categories: derivative-based Algorithms [203, 253], Statistics-Based Algorithms [22, 89, 133], Histogram-Based Algorithms [75], Intuitive Algorithms (based on thresholds) [152], microscopic optic models [99], signal processing models [92, 127] and also human visual system models [102, 103]. Most of the methods previously mentioned methods require low computational power and are more interpretable; however, compared with the most recent data-driven methods, their performance is relatively low in terms of precision and computation time.

Maximum local variation method (MLV) [22] and FQPATH [99] are two state-of-the-art knowledge-based models in the literature. The MLV method quantifies the image quality, considering the maximum variation between each pixel and its eight neighbours. The final score is given by the standard deviation of the values found for each pixel, after applying a weighting function, which attributes more relevance to higher values, increasing the difference of scores between focus and defocus images. The FQPath method adopts a kernel with properties similar to the human visual system as a sum of even derivative filters. This kernel is then applied to the image, outputting a vectorized feature map. Then, only the values coding the most focus-relevant features are preserved, ignoring features related to noise artifacts. Lastly, the m-th central moment of the remaining features is calculated, and the final score is given by the negative logarithm of this value. The optimal value of m can be found through a grid search.

Data-Driven FQAs - In the last few years, there has been an advance in data-driven approaches for FQA in WSIs. These type of methods require a large number of data for the training, however, presents in most cases a boost in performance when compared to Knowledge-based FQAs. Inside these types of models, literature presents two different categories: network adapted with minor adjustments from pre-designed and well-known architectures [44, 122], and on the other hand, architectures tailored from scratch with application-specific features [181, 206, 234, 251]. There are some open-source software applications available in the literature that try to apply some approaches for FQA in WSIs derived from the methods mentioned above: HistoQC [112] and CellPro [151]. Data-Driven approaches present very good performances in WSIs focus assessment, however, the high computational costs to train these models represent its main drawbacks.

4.1.3 Brief Summary

Most of the current approaches to ordinal inference for neural networks are found to not adequately take advantage of the ordinal problem. In the case of FQA for WSIs, all the data-driven approaches present in the literature discard the ordinal information between the different focus classes.

In the present work, a novel model to infer focus quality in WSIs is proposed. This new model is developed on the FocusPath dataset [100] and uses a plethora of deep learning architectures, with several ordinal losses. Our proposal is also compared with the current state-of-the-art approaches and surpasses them in several metrics.

4.1.4 Methods

4.1.4.1 Data pre-processing

Firstly, the dataset was divided into train, validation, and test subsets (60-20-20%), maintaining the ratio among different classes. To feed the network, it was necessary to resize the patch images from 1024×1024 pixels (px) to 224×224 px. This process was done online during the training. It was also done normalization to speed up the training, which consists of a rescale of pixel values from the range of 0-255 to 0-1. To overcome the small amount of data in our dataset, data augmentation was used. Therefore, during the training of the models, a series of random transformations were done in each training epoch for every image. The transformations applied included 10% of width and height shift, 10% of zoom, horizontal and vertical flips, and image rotation. Figure 4.1 shows three different examples of random transformations that were done online during the training process of the models.

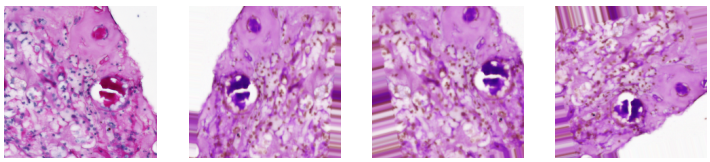


Figure 4.1: Examples of data augmentation on the FocusPath database. From left to right: original image and three examples of random transformations.

4.1.4.2 Convolutional Neural Networks

A convolutional neural network (CNN) is a class of deep learning algorithms that consecutively apply convolutions of filters to the image. These filters are learned and consist of quadrilateral patches that are convolved across the whole input image – unlike previous fully-connected networks, only local inputs are connected at each layer. Usually, each convolution is intertwined with downsampling operations, such as max-pooling, that progressively reduce the size of the original input image. The final layers are fully-connected and then the final output is processed by a softmax for the multi-class problems. Dropout was used to reduce overfitting by constraining these fully-connected layers [218]. This type of model has the intrinsic ability to learn useful features directly from the input images that are important for the task at hand.

4.1.4.3 Network Architectures

In this work, a model for classification of focus quality in WSIs was trained and tested with seven different CNN architectures: AlexNet [125], GoogLeNet [221], ResNet18 [96], MobileNet_V2 [105], ShuffleNet_V2_X1_0 [259], SqueezeNet1_0 [109], and VGG-16 [210]. The main goal of testing these different architectures is to compare the performance of a non-ordinal loss cross-entropy (CE) vs ordinal losses in a wide range of architectures. These seven different architectures were chosen as they are well known and often used in the literature. They came pre-trained with PyTorch on ImageNet[1].

The last block of each architecture was replaced by the following layers: dropout with p=20%, 512-unit dense layer with ReLU, dropout with p=20%, a 256-wide dense layer with ReLU, followed by 12 output neurons.

4.1.4.4 Losses

In this work, six different losses will be evaluated, Cross-Entropy (CE) for the baseline model and five different ordinal losses: Ordinal Encoding (OE), Binomial Unimodal (BU), CO, CO2, and also Ordinal Entropy Loss Function (HO2).

Cross-Entropy (CE): One of the most used losses in deep learning problems, where a model perform multi-class classification by minimizing cross-entropy (equation 4.1), averaged over the training set:

$$\mathrm{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = -\sum_{k=1}^{K} y_{nk} \log(\hat{y}_{nk}). \tag{4.1}$$

Cross-entropy is a suitable approach for nominal data. Nonetheless, for ordinal data, where the order between classes gives relevant insights, this information must be explored by the loss to further regularize learning.

Ordinal Encoding (OE): In ordinal encoding, which is an agnostic way to introduce ordinality, classes are encoded using a cumulative distribution – the indicator function used is $\mathbb{1}(k < k^{\star})$ so

[1] https://pytorch.org/docs/stable/torchvision/models.html

that y_{nm} is represented by 1 if $k < k_n^\star$ and 0 otherwise. Each output represents the incremental neighbor probability, and the inverse operation (during inference) is performed by summing up these outputs, $p_{nk} = \sum_{m=1}^{K-1} y_{nm}$ [52, 80].

Binomial Unimodal (BU): Constraining discrete ordinal probability distributions to be unimodal using binomial probability distributions is another technique to promote ordinality in classification problems. Thus, it is necessary to constrain the output of the network directly under a regression setting. This method predicts a single output (as the final layer) representing the probability along the classes, with $y_n = 0$ representing $k_n^\star = 1$ and $y_n = 1$ representing $k_n^\star = K$ [28, 58].

CO and CO2 Ordinal losses - In this loss, a regularization term that penalizes the deviations from the unimodal setting is added to CE [7].

Defining $\mathbb{1}(x)$ as the indicator function of x and $\text{ReLU}(x) = x\mathbb{1}(x > 0) = \max(0, x)$, a possible fix for an order-aware loss could be

$$\begin{aligned}\text{CO}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = \text{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \geq k_n^\star)\,\text{ReLU}(\hat{y}_{n(k+1)} - \hat{y}_{n(k)}) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \leq k_n^\star)\,\text{ReLU}(\hat{y}_{n(k)} - \hat{y}_{n(k+1)}),\end{aligned} \quad (4.2)$$

where $\lambda \geq 0$ controls the relative influence of the extra terms favoring unimodal distributions.

The additional terms, although promoting uni-modality, still allow flat distributions. A generalization of the previous idea is to add a margin of $\delta > 0$ to the ReLU, imposing that the difference between consecutive probabilities is at least δ [7]. This leads us to a second CE loss, **CO2**, suitable for ordinal classes:

$$\begin{aligned}\text{CO2}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = \text{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \geq k_n^\star)\,\text{ReLU}(\delta + \hat{y}_{n(k+1)} - \hat{y}_{n(k)}) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \leq k_n^\star)\,\text{ReLU}(\delta + \hat{y}_{n(k)} - \hat{y}_{n(k+1)}).\end{aligned} \quad (4.3)$$

A value of $\delta = 0.05$ has been empirically found to provide a sensible margin.

Ordinal Entropy Loss Function (HO2) This loss arises to mitigate some negative aspects of the CE in the CO and CO2 losses since the CE term by itself is only trying to maximize the probability estimated in the true output class (while ignoring the remaining probabilities). The ordinal terms in CO and CO2 promote unimodality but do not penalize (almost) flat distributions. A softer assumption is that the distribution should have a low entropy [7].

This leads us to the ordinal entropy loss, HO2, for ordinal data as

$$\begin{aligned}\mathrm{HO2}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = \mathrm{H}(\hat{\mathbf{y}}_n) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \geq k_n^\star)\,\mathrm{ReLU}(\delta + \hat{y}_{n(k+1)} - \hat{y}_{n(k)}) \\ + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \leq k_n^\star)\,\mathrm{ReLU}(\delta + \hat{y}_{n(k)} - \hat{y}_{n(k+1)}),\end{aligned} \quad (4.4)$$

where $\mathrm{H}(\mathbf{p})$ denotes the entropy of the distribution $\mathbf{p}$.

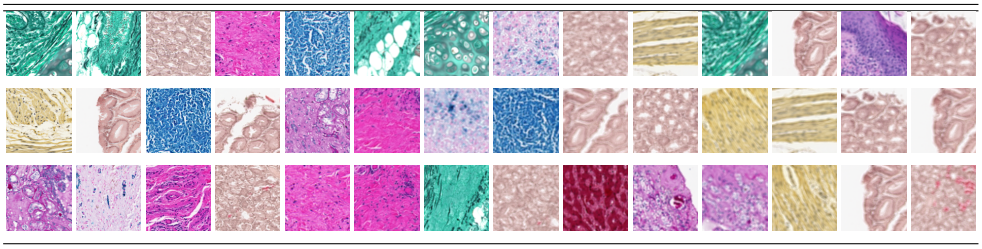
In Chapter 5 of this thesis, these loss functions will be explained in more detail.

4.1.4.5 Experimental Details

4.1.4.6 Dataset

In this work, it was used a public histopathological database named FocusPath dataset [2], which contains image quality annotations [101]. The FocusPath dataset contains 8640 patches of 1024x1024px images. These images were extracted from nine different stained slides from diverse human organs. The original Whole Slide Images were scanned by Huron TissueScope LE1.2. It uses a 40X optics lens at 0.25 μm/pixel resolution. There are 14 absolute z-level scores corresponding to the ground-truth class for the focus level. In Table 4.1 it is possible to see several examples from each class present in the FocusPath dataset. Due to the low number of examples of images in the more defocused classes (12 and 13), the label was changed to belong to class 11. This way, the dataset was grouped into 12 different focus classes.

Table 4.1: Image examples of the 14 Different patch focus Classes in the FocusPath dataset.



0	1	2	3	4	5	6	7	8	9	10	11	12	13

Focus Level - focus(0) $\rightarrow$ defocus(13)

4.1.4.7 Training

All the losses were tested on the CNN model represented on Figure 4.2. The model had as input the RGB WSIs patches from the FocusPath database. The model output consists of the multiclass classification (12 classes) of focus quality (0-focus patch to 11-defocus patch).

[2] https://zenodo.org/record/3926181#.YPFgluhKjIU

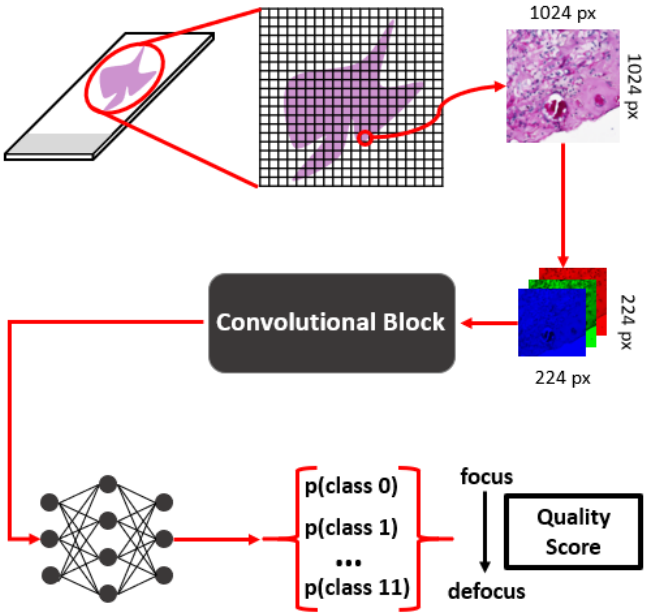


Figure 4.2: Schematic representation of the model architecture.

During the training initialization, the weights of the architectures previously mentioned were initialized based on ImageNet pre-training. The optimizer used was ADAM, starting with a learning rate of 10^{-4}. The learning rate is reduced by 10% whenever the loss is stagnant for 10 epochs using a specific scheduler. The training process is completed after 200 epochs. The training process was conducted on a single Nvidia GTX 1080ti (11GB) GPU.

In the case of the Ordinal Entropy Loss Function, the hyperparameter λ is tuned by doing nested k-fold cross-validating using the training set (with k=5) to create an unbiased validation set.

4.1.4.8 Evaluation Metrics

To evaluate the performance of the different models, three metrics were adopted: accuracy, mean absolute error, and also Kendall's τ rank correlation coefficient.

One of the most popular metrics used in classification problems is accuracy (Acc). For N observations, taking k_i and $\hat{k}_i$ to be the label and prediction of the n-th observation, respectively, then $\text{Acc} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}(\hat{k}_n^\star = k_n^\star)$, where $\mathbb{1}$ is the indicator function.

However, this metric treats all class errors as equal, whether the error is between adjacent classes or between classes in the extreme. If we have K classes represented by a set: $\mathscr{C} = \{\mathscr{C}^{(1)}, \mathscr{C}^{(2)}, \ldots, \mathscr{C}^{(K)}\}$, then accuracy will treat an error between $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(2)}$ with the same magnitude as an error between $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(K)}$ which is clearly worse. For that reason, a popular metric for ordinal classification is the Mean Absolute Error (MAE), $\text{MAE} = \frac{1}{N}\sum_i |k_i^\star - \hat{k}_i^\star|$. This metric is not perfect since it treats an ordinal variable as a cardinal variable. An error between classes $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(3)}$ will be treated as two times worse than an error between classes $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(2)}$. Naturally, the assumption of cardinality is not always warranted.

Kendall's τ rank correlation coefficient (Kendal Tau) is a non-parametric evaluation of relationships between columns of ranked data, so it is a measure of ordinal association between data. The τ correlation coefficient returns a value that ranges from -1 to 1, with 0 being no correlation and 1 perfect correlation.

4.1.5 Results and Discussion

The performance of the seven different architectures are presented in Tables 4.2–4.3, and 4.4, for 12-class classification problem, with the six different learning losses – conventional Cross-Entropy (CE), Binomial Unimodal (BU) [58], Ordinal Encoding (OE) [52], CO, CO2 and Ordinal Entropy Loss Function (HO2) [7]. The best models are shown in bold.

Table 4.2: Results in terms of Accuracy (%) for 12 class problem (higher is better).

	CE	BU	OE	CO	CO2	HO2
alexnet	**72.4**	69.7	66.8	64.9	62.7	62.9
googlenet	92.2	87.2	91.1	**92.4**	87.0	78.1
resnet18	**88.7**	86.8	87.2	83.3	82.7	86.6
mobilenet_v2	92.2	82.9	**94.4**	89.3	86.0	86.0
shufflenet_v2_x1_0	84.7	81.9	**86.0**	84.3	76.1	80.8
squeezenet1_0	73.4	71.5	**74.2**	31.1	63.9	62.9
vgg16	87.2	84.9	**89.1**	62.1	84.3	80.6
Avg	84.4	80.7	84.1	72.5	77.5	76.8
Winners	2	0	**4**	1	0	0

Table 4.3: Results in terms of Mean Absolute Error (MAE) for 12 class problem (lower is better).

	CE	BU	OE	CO	CO2	HO2
alexnet	**0.29**	0.32	0.34	0.37	0.42	0.40
googlenet	0.08	0.13	0.09	**0.08**	0.13	0.22
resnet18	**0.12**	0.13	0.13	0.19	0.19	0.14
mobilenet_v2	0.08	0.18	**0.06**	0.12	0.15	0.16
shufflenet_v2_x1_0	0.16	0.19	**0.16**	0.16	0.25	0.20
squeezenet1_0	0.30	0.29	**0.27**	1.04	0.43	0.42
vgg16	0.14	0.15	**0.11**	0.43	0.17	0.23
Avg	0.17	0.20	0.17	0.34	0.25	0.25
Winners	2	0	**4**	1	0	0

Table 4.2 shows the accuracy results for the 12-class focus classification problem. Regarding this metric, it is possible to infer the difference in the results between nominal and ordinal losses. OE loss performance was better for 4 different architectures against 2 for nominal CE. This can be explained by the lower role of ordinality in the CE loss, as also confirmed by the MAE results.

Table 4.4: Results in terms of Kendall's τ for 12 class problem (higher is better).

	CE	BU	OE	CO	CO2	HO2
alexnet	**93.8**	93.5	93.3	92.2	91.5	91.7
googlenet	**98.4**	97.2	98.0	98.3	97.3	95.7
resnet18	97.4	97.1	**97.4**	95.8	96.0	96.8
mobilenet_v2	98.3	96.6	**98.8**	97.5	96.9	96.2
shufflenet_v2_x1_0	96.3	95.8	96.3	**96.3**	95.0	96.0
squeezenet1_0	92.8	**94.2**	94.1	75.0	90.7	91.3
vgg16	97.2	97.1	**97.9**	89.9	96.2	94.8
Avg	96.3	95.9	96.5	92.1	94.8	94.7
Winners	2	1	**3**	1	0	0

This means that when misclassification occurs, ordinal losses tend to classify the focus quality in the WSI patches as being closer to the real class.

Table 4.3 presents the results for MAE, which confirm the influence of ordinal losses in promoting ordinality when compared to nominal loss (CE). OE loss achieved the best performance across the different architectures. Across the different architectures, ordinal losses won nominal CE in 5 against 2.

In Figure 4.3 is presented the confusion matrix of our best model (MobileNet_v2 trained with Ordinal Encoding). It is possible to infer that the misclassified images were very close to the main diagonal of the confusion matrix, which reveals that the ordinal loss helps to reduce the distance between the true class vs predicted class when the model classifies wrongly, as mentioned before.

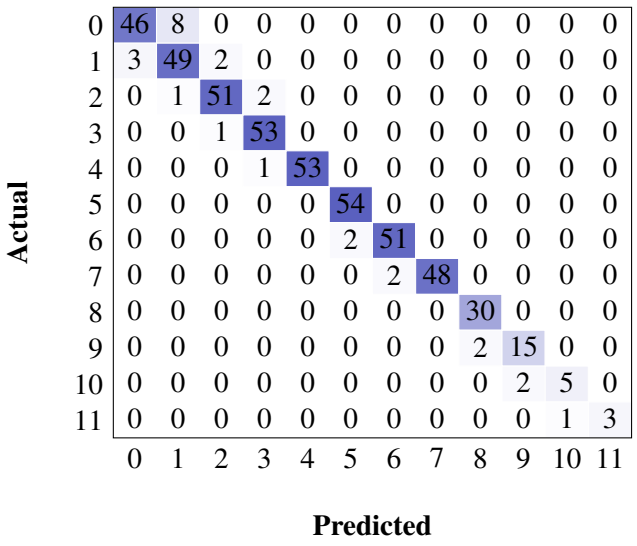


Figure 4.3: Confusion Matrix of our best model - MobileNet_v2 - using Ordinal Encoding as loss.

Going a step further and comparing the results with other works featuring focus quality assessment in the FocusPath database, it is clear that the evaluation metrics are above the state-of-the-art

results, especially on the Spearman's rank correlation coefficient (SRCC) and Pearson correlation coefficient (PLCC) metrics. Wang et.al [234] present a data-driven work with better results in all the metrics, in contrast to knowledge-based methods (MLV [22] and FQPATH [99]). However, in his approach, the model is fed with random crops (235 x 235 px) of the WSIs patches. Thus, it loses some relevant spatial information about the focus (e.g. when exists in the same patch focus and defocus zones).

This comparison of metrics between our proposed method and the literature works is presented in Table 4.5.

Table 4.5: Comparison of our proposal with literature models

	SRCC	**PLCC**	**Time (sec)**
MLV [22]	0.8623	0.8528	0.482
FQPATH [99]	0.8395	0.8295	0.269
FOCUSLITENN [234]	0.8931	0.8857	**0.019**
Proposal	**0.9969**	**0.9970**	0.047

4.1.6 Summary and Conclusions

Comparing the different deep learning approaches on WSIs patches, the models trained with ordinal losses achieved better results when compared with the nominal cross-entropy loss. Thus, a new model has been proposed for multi-class FQA in WSIs based on convolutional neural networks and respecting the ordinal progression among the different focus levels. The model was trained in a dataset with several color spectrums and tissue structures, these factors contributed to a robust training of the model. This new model demonstrated to be competitive with state-of-the-art results and surpasses them in some metrics.

Furthermore, another important outcome is that those data-driven methods obtained significantly better results than knowledge-based methods in terms of precision and performance.

For future works, since this model is a prototype of an FQA system, the architecture should suffer some changes and multi-processing must be used to reduce the processing time per patch in a real-time acquisition system for WSIs.

4.2 Image Enhancement using Deep Based Extended Depth of Field Methods

Microscopic techniques in low-to-middle-income countries are constrained by the lack of adequate equipment and trained operators. Since light microscopy delivers crucial methods for the diagnosis and screening of numerous diseases, several efforts have been made by the scientific community to develop low-cost devices such as 3D-printed portable microscopes. Nevertheless,

these devices present some drawbacks that directly affect image quality: the capture of the samples is done via mobile phones; more affordable lenses are usually used, leading to poorer physical properties and images with lower depth of field; misalignments in the microscopic set-up regarding optical, mechanical, and illumination components are frequent, causing image distortions such as chromatic aberrations. This work investigates several pre-processing methods to tackle the presented issues and proposes a new workflow for low-cost microscopy. Additionally, two new deep learning models based on Convolutional Neural Networks are proposed (EDoF-CNN-Fast and EDoF-CNN-Pairwise) to generate Extended Depth of Field (EDoF) images and compared against state-of-the-art approaches. The models were tested using two different datasets of cytology microscopic images: public Cervix93 and a new dataset that has been made publicly available containing images captured with μSmartScope. Experimental results demonstrate that the proposed workflow can achieve state-of-the-art performance when generating EDoF images from low-cost microscopes.

4.2.1 Context and Motivation

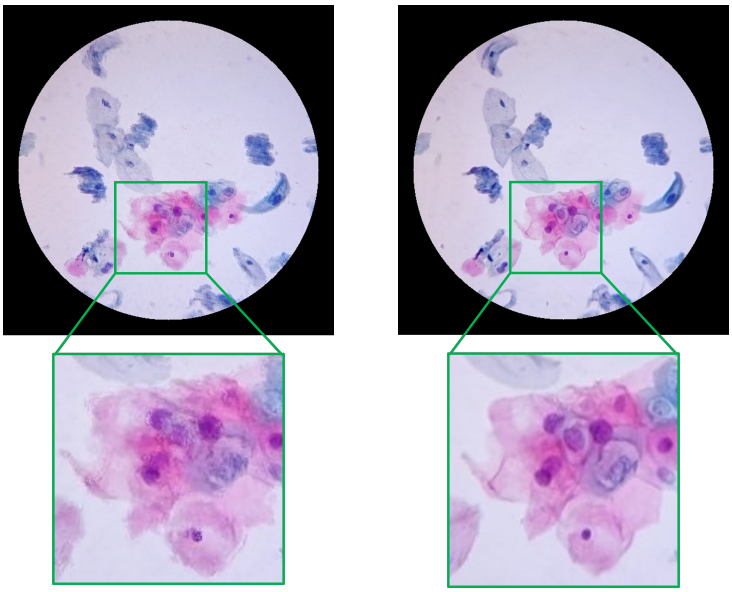
Optical microscopes remain a crucial tool for diagnosing several diseases that daily affect public health, such as cancer, parasitism, and infectious diseases, among others. In addition to its great usefulness in the medical diagnosis field, optical microscopy is also important for research, training, and knowledge dissemination within life sciences. However, the ability to perform microscopic-based diagnosis in medically underserved areas is still very limited due to the high costs of microscopic equipment and the lack of trained operators/technicians [194, 198]. Thus, its broad operationalization in real clinical contexts is difficult to implement, mainly to support extensive health plans aimed at serving populations in low and middle-income countries with a demarcated incidence of diseases.

To tackle this problem, several efforts were made by the scientific community to create low-cost devices such as 3D-printed portable microscopes [41, 83, 187, 194]. Compared with conventional microscopes, these devices present some drawbacks that directly affect image quality, such as the capture of the samples being done via mobile phones. Mobile-based image acquisition can lead to distortions like chromatic aberrations caused by misalignment in the RGB channels [226]. Another frequent source of chromatic aberrations is the physical misalignment in the microscopic set-up regarding optical, mechanical, and illumination components. Such image distortions can also be caused by more affordable lenses, usually incorporated in the optical systems of low-cost microscopes. These types of lenses have lower physical properties, leading to a lower depth of field.

Indeed, limited depth of field is a standard limitation in optical microscopy due to the 3D composition of slides containing biological material of variable volume, where some structures can appear focused at certain z-axis (optical axis) positions and others may not. To overcome this limitation, algorithms of extended depth of field (EDoF) can be applied, where a stack of slides that are taken at different focal positions (along the optical axis) are merged to form a single

and entirely focused composite image. This method can also be denominated as an image fusion algorithm.

Nonetheless, all image fusion algorithms rely on two conditions to work properly. The first condition is that the images must be aligned along the entire stack (the x and y axes are fixed, only the z-axis varies), and the second is that they must all have the same level of magnification [76]. However, in low-cost microscopes, the first requirement is usually not met because the motions are inaccurate due to the lower quality of the mechanical components (e.g. 3D-printed parts, stepper motors, technical gutters, etc.). Thus, when traditional EDoF fusion methods are applied to a sequence of images (stacks) acquired with such devices without any pre-processing technique to align the images through the optical axis, a very low-quality fusion image is produced. Figure 4.4 shows two fusion images generated with and without pre-processing.



(a) EDoF without alignment. (b) EDoF with alignment.

Figure 4.4: Comparison between two EDoF images generated without alignment (a) and with the alignment (b) of the images stack.

It is possible to see the differences between the generated fusion image with and without alignment. In the case of the image without alignment (Figure 4.4 (a)) the structures are poorly defined and blurred. Unlike the image with alignment (Figure 4.4 (b)), where in addition to seeing a highly focused image in all areas, it is also possible to see the well-delimited structures present on the slide (e.g. nucleus and cytoplasm). To improve the quality of low-cost microscopy images, it is necessary to consider the entire pre-processing workflow including the alignment of the stack images. Since they suffer not only from rigid but also elastic perturbations, causing the deformation of the structures and misalignment in the stacks, both alignment techniques must be tested and applied to improve the final EDoF quality.

Thus, in order to address all previous concerns, this work intends to develop, in a first step, a

novel pre-processing workflow that can be used across all low-cost microscopy devices. Furthermore, in a second step, new deep learning models based on Convolutional Neural Networks (CNN) are also proposed (EDoF-CNN-Fast and EDoF-CNN-Pairwise) to generate Extended Depth of Field images instead of traditional fusion methods, which are more time-consuming [148]. These new methods are compared with state-of-the-art approaches. From the acquisition of a series of images (stacks) at different focal planes (z-planes) per field of view (FOV) to the EDoF creation, this study aims to create an end-to-end method that can be applied in low-cost microscopy. All the proposed methods can be easily integrated into low-cost microscopy devices using a cloud-based computing system. To the best of our knowledge, this is the first study that explores the most important causes of image quality decrease in low-cost microscopy workflow and proposes a new strategy to surpass most of them. The code to reproduce this work is publicly available[3].

As demonstrated by [148], the use of EDoF images instead of limited FOV images improved the performance of the diagnostic models for malaria detection using cytology microscopic images. This indicates that the use of image quality enhancement by image fusion techniques can promote better results in diagnostic models, opening a new window for different areas that use low-cost microscopy and that are used by the medical community worldwide. The proposed EDoF models can also be used in different medical fields such as CT imaging of the brain, where several stacks are also captured [240].

The remainder of the article is as follows. Section 4.2.2 reviews other work on preprocessing, including image alignment in microscopy and image enhancement by image fusion algorithms. Section 4.2.3 describes firstly the proposed framework, followed by the data pre-processing steps, including the tested methods for chromatic, rigid, and elastic alignment of the images, and finally, it also presents a detailed description of the convolutional neural networks architectures proposed in this work. In Section 4.2.4 the experimental details, including the used datasets and the evaluation metrics, are described, followed by the results and discussion in Section 4.2.5. The study finishes with a conclusion (Section 4.2.6).

4.2.2 Related Work

This section is divided into two parts: the first part covers related work on image pre-processing in low-cost microscopy; while the second part focuses on the main image fusion methods for EDoF generation.

4.2.2.1 Low-cost microscopy images pre-processing

The simple optical systems of low-cost microscopes are suitable for the job; however, they are often made up of tiny and cheap lenses, which tends to sacrifice the quality of the obtained images. Additionally, the physical structure and the mechanical parts are often obtained using fast

[3] `https://github.com/tomealbuquerque/EDoF-CNN`

prototyping methods such as 3D printing. Despite being easy to get these components, their quality can affect the stability of the microscope, inducing misalignment and other distortions to the image [198].

Even so, most works in the literature using inexpensive microscopes do not consider these problems [163]. Image alignment, also known as image registration in microscopy, focuses on aligning images from various focal planes and images or volumes recorded at different time periods [186]. A low-cost automated digital microscope was used by [199] to identify Diatoms (microalgae). In this approach, after obtaining the images at different FOVs, a stitching method was applied to create a fusion image, using a rigid alignment. However, the chromatic aberrations and the elastic displacement between the different FOV images were not taken into account.

4.2.2.2 Image fusion methods for microscopy

The depth of field of optical systems has been increased by the use of image fusion methods [155]. Gathering all the crucial data from several images and combining them into fewer, typically just one, is known as the image fusion process. This new single image (EDoF) contains all the relevant information and is more accurate and informative than the other images from a single source. The goal of image fusion is to create images that are more relevant and intelligible for both human and machine perception while reducing the amount of data [17, 34]. Image fusion methods can be divided into three branches: spatial domain, transform domain, and more recently the use of convolutional neural networks.

Regarding spatial domain methods, there are two different approaches: pixel-based and block-based methods. In pixel-based methods, several approaches directly average the pixel values of all the source images. Novel pixel-based image fusion methods, including guided filtering [134] and dense SIFT [144], have recently been reported. After identifying the focused pixels in each source image using guided filtering and dense SIFT, the final fused image is created by choosing the focused pixels based on the changed decision map. On the other hand, block-based methods divide the input images into blocks [106, 260]. The block size has a significant influence on the fusion outcomes. The use of an artificially fixed block size in the earlier block-based algorithms has been replaced by several improved techniques [21, 23, 62].

Going a step further, in transform domain methods, the input images are often converted to other feature domains to promote an efficient image fusion. Multi-scale transform (MST) techniques are the cornerstone of the most well-known transform domain fusion algorithms. The non-subsampled contourlet transform (NSCT) [258], the complex wavelet transform (CWT) [76, 131], the discrete wavelet transform (DWT) [132], the morphological pyramid (MP) [224] and Laplacian pyramid (LP) [43] are a few examples of the most relevant algorithms present in literature. Decomposition, fusion, and reconstruction are the three steps that these techniques follow to merge the image [179]. Other relevant methods in the transformation domain are the sparse representation (SR) based methods. These fusion algorithms, as opposed to the MST-based method discussed above, use superior signal techniques to convert the image into a single-scale feature region. After the approximative translation invariant fusion phase, this kind of methodology often employs the

sliding window method. The exploration of a valid feature domain to produce the focus map is the key issue with these methods [136, 164, 250]. The discrete wavelet transform methods uses a selection rule based on maximum absolute coefficient values, which leads to good results; however, this method has been surpassed by the complex wavelet-based methods, which expand on this concept and demonstrate state-of-the-art performance for this class of algorithms [76]. This method first establishes an in-focus criterion before the increase of the depth of focus. An in-focus images has the most information that can be seen due to the high number of visible details. On the other hand, the microscope's point-spread function causes distortions, leading to unfocused images with low details. Thus, it is assumed by this method that the concentrated parts of an image have more high-frequency components than unfocused areas. In contrast to the Fourier technique used in traditional frequency analysis which does not offer any spatial localization, the wavelet transform appears to be the best method for high saliency identification since it enables a local study of the frequency content of the image. This kind of multi-resolution automatically adjusts to varying sizes of details, unlike windowed Fourier transforms, and it also offers flawless reconstruction and maintains the original size of the images [76].

Recently, deep learning methods based on CNNs have emerged in the scientific community, achieving top performance in multi-focus fusion of images of natural scenes. These methods can be separated into CNNs for image pair fusion and CNNs for image stack fusion. Similar to how segmentation is employed in typical image segmentation tasks, multi-focus image fusion serves a similar purpose. Therefore, it is theoretically possible to employ a CNN for image fusion [71,143]. The merging of two images with similar features and some degree of blur is easily achieved by the CNNs [71,143,233]. Despite that, it is still not known how well these techniques scale to combine more than two out-of-focus images. These networks are built on CNN models that resemble Siamese structures taught to produce the appropriate focus map for fusion.

The use of CNNs to increase the spatial resolution of images has also emerged in the scientific community [68]. Through the application of deep learning, [193] launched their usage in microscopy by quickly producing images with superior resolution, matching the performance of larger numerical aperture microscope lenses, and greatly outperforming their constrained field of view and depth of field. Spatial resolution enhancement can be formulated as a regression problem, since in a CNN model where a low-resolution image is used as input and passed through the network, the training objective is to minimize the pixel-wise difference between the network output and the corresponding high-resolution reference image. Thus, the multi-focus fusion problem can also be approached as a regression task to combine the various focal planes while improving the digital spatial resolution of the combined image [148]. In our work, the above approach is also followed by training CNNs for image fusion as a regression problem. [148] introduced EDoF-CNN-3D and EDoF-CNN-Max architectures, which are composed of three parts: encoder, residual layers, and decoder. EDoF-CNN-3D uses the z-stack as a 3D volume, which requires changing the two-dimensional convolutions in the CNN architecture's encoder section to three-dimensional ones. Prior to the residual layers of the network, the output tensor was flattened on the z-axis using average pooling. After the encoder, there are nine residual layers followed by

three 2D transposed convolutional layers (decoder) using a Hyperbolic tangent (tanh) as an activation function at the end of the network. EDoF-CNN-Max combines Siamese networks' and the complex wavelet-based multi-scale decompositions (MSD) EDoFs [76] theories. Before moving through the residual layers, each focus plane passes through the encoder part of the network, where the highest activation values are chosen using the max function. The remaining part of the network is the same as in EDoF-CNN-3D [148].

Figure 4.5 summarizes this section through a schematic diagram of the most relevant methods in the literature for image fusion.

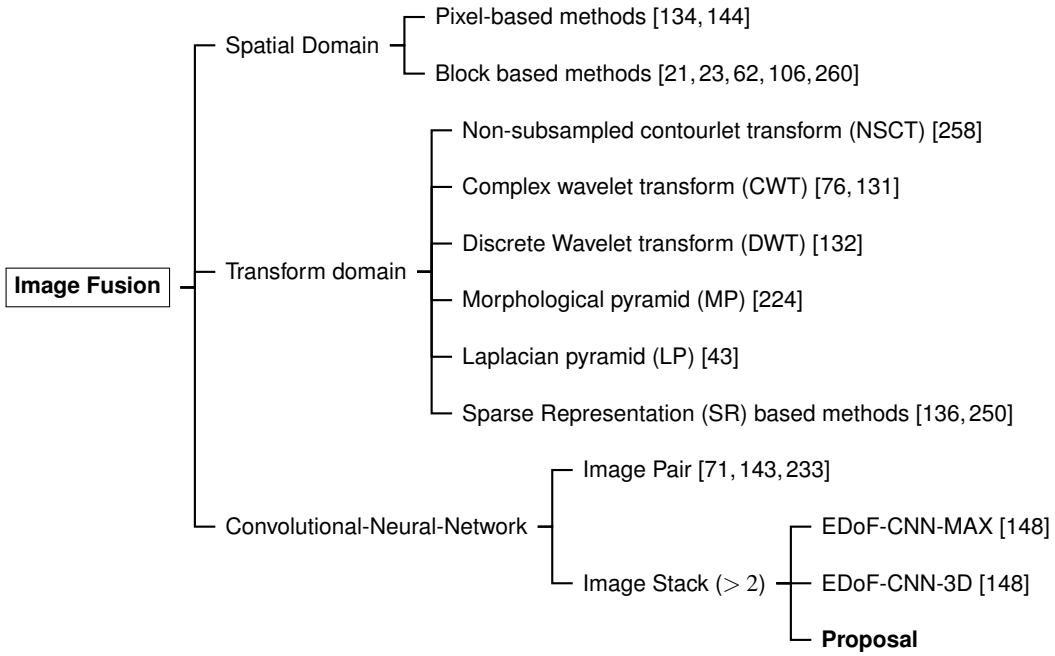


Figure 4.5: Most relevant fusion methods present in the literature.

4.2.3 Proposed Methodology

The first contribution of this work is a novel pre-processing methodology that can be applied to all low-cost microscopy equipment prior to the generation of EDoF images. By exploring three different pre-processing workflows after the z-planes capturing with the low-cost microscope, this work intends to provide the best workflow to generate better quality EDoF images. The three workflows studied are chromatic and rigid alignment; chromatic, rigid, and elastic alignment and finally chromatic, rigid, and elastic alignment of the stacks.

A second contribution of this work is to test and propose deep learning models based on Convolutional Neural Networks (CNN) to generate images with Extended Depth of Field from the z-planes. EDoF-CNN-Fast and EDoF-CNN-Pairwise are the proposed networks. These new methods are compared with state-of-the-art approaches and tested in a publicly available dataset

(Cervix93) and in a new publicly available dataset acquired with a low-cost microscope (EDoF Fraunhofer dataset).

This work intends to provide an end-to-end approach that may be used in low-cost microscopy, starting with the capture of a sequence of images (stacks) at distinct focal planes (z-planes) per field of view and ending with the development of the EDoF images using deep learning techniques.

4.2.3.1 Proposed Data pre-processing workflow

To investigate the best pre-processing workflow after the capturing of the z-planes, an ablation study was promoted by testing three different workflows: (A) chromatic and rigid alignment; (B) chromatic, rigid and elastic alignment, and finally (C) chromatic, rigid, and elastic alignment of the stacks. Figure 4.6 shows the pre-processing workflow with all the alignment techniques tested: chromatic, rigid, and elastic alignment of the stacks.

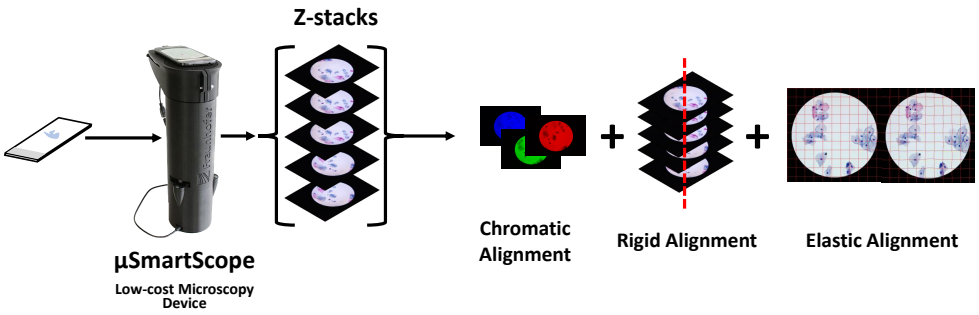


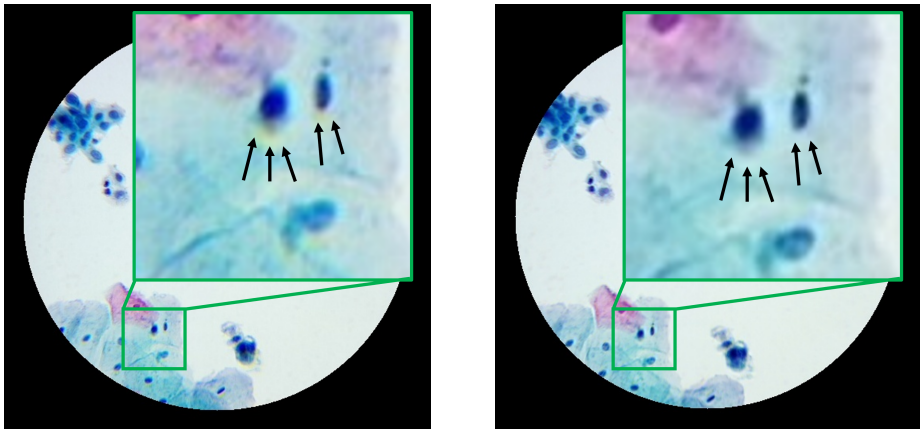
Figure 4.6: Representation of the proposed and tested prepossessing workflow method: acquisition of the stacks followed by chromatic, rigid, and elastic alignment of the stacks.

In the next subsections, each alignment methodology will be explained in more detail.

Chromatic Alignment The phenomenon known as chromatic aberration occurs when the light of various wavelengths fails to converge at the same position on the focal plane, leading to misalignment between the RGB channels. Mobile phone cameras with low-end plastic lenses rather than high-end, low-dispersion aspherical lenses exhibit more severe aberration. Because each camera model has a unique lens structure, chromatic aberration will vary somewhat depending on the model [226].

Figure 4.7 shows an example of a low-cost microscopy image before and after chromatic alignment. It is possible to visualize in Figure 4.7 that the chromatic aberration, marked with the black arrows, almost totally disappeared after the chromatic alignment. Thus, to reduce the impact of chromatic aberration on the quality of microscopy images, the alignment of the RGB channels must be done. Chromatic alignment or RGB alignment follows the same procedure as rigid alignment, however, instead of aligning the image using a reference image, the green channel is used as the reference image, and the other channels (Red and Blue) are aligned. After the alignment process, the final image is generated by reassembling the three channels into an

RGB image. More details about the techniques used for the alignment process are described in the next subsection.



(a) Without chromatic alignment. (b) With chromatic alignment.

Figure 4.7: Comparison between two microscopy images without (a) and with (b) chromatic alignment

Rigid Alignment When several images of the same slide are collected at different times or at different z-axis positions, it is mandatory to have them all properly aligned to generate high-quality fusion images (EDoF). Even if the images were taken during the same scanning session, the sample/ region of interest could have moved a little during each acquisition due to various mechanical perturbations. Rigid alignment is the simplest way to perform this alignment. Through exclusively translations and rotations, rigid registration tries to create the best alignment between two images [129]. It can be divided into four steps:

1. Selecting the reference image (central image from the stack) and the image that will be aligned that was captured in the same slide but at a slightly different viewing angle;
2. Running a Feature-based method to detect the keypoints in our two input images. Keypoints are intended to recognize prominent regions of an input picture;
3. Calculating the homography matrix automatically with the keypoints obtained previously;
4. Applying a perspective warp on that homography matrix to align the stack image.

To perform the slide alignment, several OpenCV functions[4] were used in this work.

After selecting by pairs the images that need alignment (step 1), a feature-based method named ORB (Oriented FAST and Rotated BRIEF) [195] is applied in step 2. Step 2 uses the FAST (Features from Accelerated Segments Test) algorithm to calculate the keypoints. This method considers the pixel brightness around a given area to calculate the keypoints. A feature vector for

[4]OpenCV is an image processing software available at `https://opencv.org/`

each keypoint is created by ORB. Supposing that kk is a keypoint in image xx and $k'k'$ is a keypoint in image $x'x'$, the feature vectors help to determine if keypoints k,k and k',k' are good matches. A suitable match occurs when: a keypoint from $x'x'$ is considerably more similar to the kk than any other keypoint in $x'x'$. A Brute-Force matcher from OpenCV is used to help with this matching process. A collection of potential inliers is created by the set of favorable matches.

The following step 3 consists in computing the homography matrix using RANSAC (Random sample consensus) algorithm [25]. The homography is computed using four excellent matches randomly chosen, which are then evaluated by how many of the good matches are consistent with the homography (e.g.if k,k and k',k' are good matches, the homography must map kk closer to $k'k'$). In the homography generation process, good matches that are compatible with the homography are referred to as inliers, while those that are not are referred to as outliers. After this first step in homography computation, the number of outliers is determined for this homography based on a threshold distance. Then, after repeating this process several times (selecting a new set of four good matches each time), the homography is generated and the number of outliers is counted. The homography that is saved is the one that contains the smallest number of outliers.

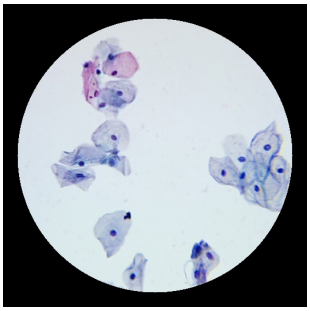
Finally, step 4 on rigid alignment is the perspective transformation, where a perspective warp is applied to the homography matrix to align the stack image. The perspective transformation consists of a change in the point of view of the stack that we want to align. In this type of transformation, it is not preserved the length, angle, and parallelism, but incidence and collinearity are preserved. The warp perspective function of OpenCV takes as input the stack that will be aligned and the homography matrix and returns the final image rigidly aligned.

Elastic Alignment During the stacks capturing along the z-axis, the microscope motion causes several nonrigid deformations that rigid or affine transformations alone are not capable to correct. Thus, the main goal of the elastic alignment given a reference image (central stack) and several stacks is to find the optimal transformation $\mathbf{T}$: $(x,y,z) \rightarrow (x',y',z')$ which maps any point in the dynamic image sequence along z-axis into its corresponding point in the reference image $\mathbf{I}(x',y',z')$.

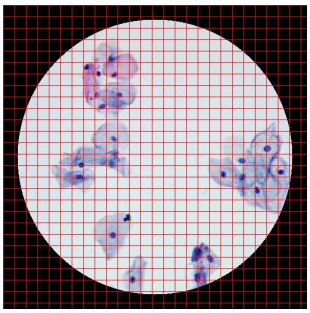
ELASTIX v5.0.1[5] was used to perform elastic alignment of the different stacks. ELASTIX is a publicly available software that can be integrated with Python to perform state-of-the-art elastic alignments [121]. The B-spline transformation developed in [196] is the foundation for ELASTIX, which is represented as a weighted sum of B-spline basis functions set on a uniform control point grid. The Local support property [39, 225] (each spline curve piece is only a function of the four closest control points) is provided by the B-spline basis functions, which is advantageous for fast computation. The resolution of the control point grid is set beforehand by a parameter file, which determines the flexibility of the deformation. This parameter file ('Elastix_parameter_Affine.txt') is provided in the GitHub repository. An example of a stack subjected to elastic alignment according to a reference image is represented in Figure 4.8. A red

[5] `https://elastix.lumc.nl/`

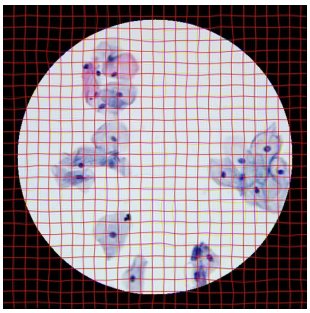
grid was added to the original stack before the alignment so that the image transformations were visible after the elastic alignment.



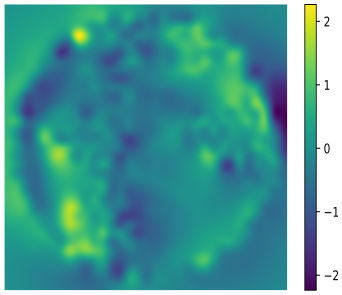
(a) Reference Image.



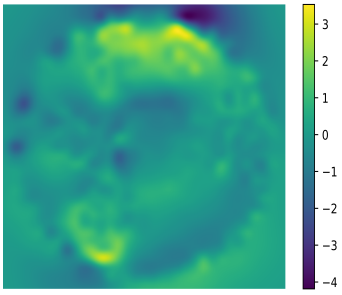
(b) Slide to align.



(c) Slide elastic aligned.



(d) xx's displacement.



(e) yy's displacement.

Figure 4.8: Example of an alignment of a slide using elastic transformations where (a) represents the central/reference slide; (b) slide before alignment, (c) slide after the alignment, and (d) and (e) representing the deformation displacement in xx's and yy's respectively.

Figure 4.8 (c) reflects on the red grid the deformations caused by elastic transformations to perfectly align the stack with respect to the reference image (Figure 4.8 (b)). The two hyperparameters used to run the algorithm for image alignment were the maximum number of iterations and the number of resolutions. As recommended in the ELASTIX documentation, the maximum number of iterations at each resolution used was 500 at 3 different resolutions[6]

[6] `https://elastix.lumc.nl/doxygen/parameter.html`

4.2.3.2 Proposed Image Fusion

Two types of image fusion methods are proposed (EDoF-CNN-Fast and EDoF-CNN-Pairwise). In the next subsection, both architectures will be explained.

CNN for EDoF generation The complex wavelet extended depth of field method remains the gold standard method for image fusion, however, it is a very slow method when compared with CNNs. The use of CNNs can improve the image's spatial resolution and also speed up to eleven times the fusion process when compared with complex wavelet EDoF method [148].

All four convolutional Neural Networks (CNN) tested in this work follow the encoder-decoder structure and have as input the multiple focal planes (z-stacks) and as output the EDoF image [117, 148]. Four different models are compared, two from the literature, the EDoF-CNN-3D, and EDoF-CNN-Max which are described in [148], and our proposal models which are EDoF-CNN-Fast and EDoF-CNN-Pairwise. A brief description of each proposed architecture is given below, explaining the three parts of each architecture: encoder, residual layers, and decoder.

EDoF-CNN-Fast This method is inspired by EDoF-CNN-Max with several modifications to the architecture. The encoder part remained almost the same, but with a reduction in the convolution channels depth. The residual part is also modified in terms of channel depth and reduced to five layers. The channels' depth is also changed in the decoder part and two more modifications were induced. The first one is in the upsampling technique of the decoder, which changed from nearest-neighbor to bicubic. In the nearest-neighbor technique, each value is just copied from the nearest pixel; however, with the bicubic approach, polynomial interpolations are employed to generate each pixel's values by taking into account all adjacent pixels. This last technique often creates a smoother surface than the other methods, but it demands more computational time, which in this case is compensated by the complexity reduction of the network when compared with EDoF-CNN-Max. The reduction of trainable parameters in EDoF-CNN-Fast improves the generalization capability of this network when compared with EDoF-CNN-Max. This improvement leads to a decrease in over-fitting during the training process and also an increase in model robustness. This complexity reduction in this architecture also tries to avoid the vanishing gradients problem. The last modification was in the last activation function, which was changed to linear instead of tanh. Figure 4.9 shows a schematic representation of EDoF-CNN-Fast Network detailing the used layers.

EDoF-CNN-Pairwise This architecture reuses the EDoF-CNN-Fast backbone in cascade, however, it is inspired by human perception for pairwise comparison of images. For humans, it is always easier to identify differences by the comparison between only two images instead of several [36]. Hence, the EDoF-CNN-Pairwise starts by receiving as input the first (Z_1) and second (Z_2) image from a stack to generate a new EDoF ($EDoF_{1/2}$). Then, this new $EDoF_{1/2}$ generated with only the first two images from the stack will be used as input for the network, however now the second input of the network will be the third (Z_3) image from the stack generating the $EDoF_{1/2/3}$. This process is repeated until the last image from the stack reaches the network. The last image (Z_n) and the $EDoF_{1/2/../n-1}$ is used to generate the final high quality EDoF image for

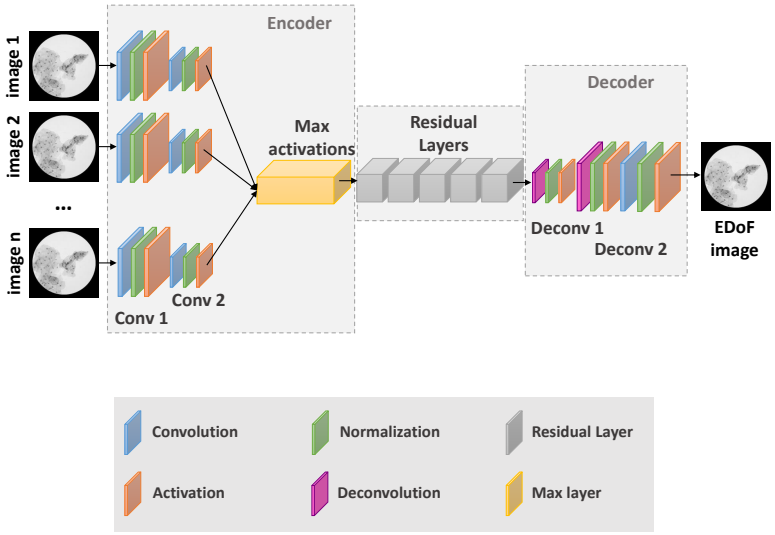


Figure 4.9: Schematic representation of the EDoF-CNN-Fast Network.

the stack. In Figure 4.10, it is possible to see a succinct representation of the EDoF-CNN-Pairwise network.

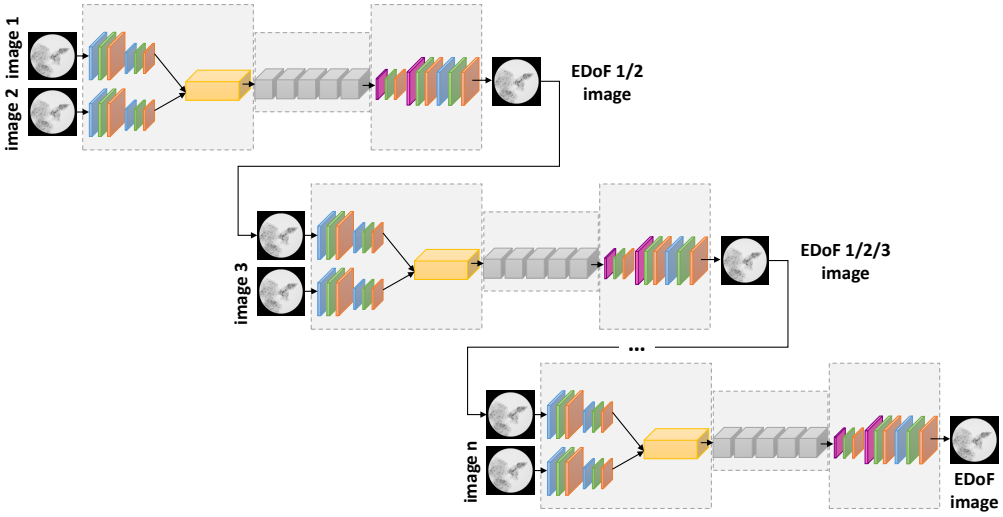


Figure 4.10: Schematic representation of the EDoF-CNN-Pairwise Network.

During the training process, the back-propagation is only done when all the images pass through the network, this way it is possible to reuse always the same network and save computational time and resources. As Table 4.6 shows, the EDoF-CNN-Fast network has the smallest number of parameters when compared with the others architectures.

Table 4.6: Summary of the number of trainable parameters across the different architectures(proposal architectures (P)) for a stack with five z-planes.

Architecture	Number of Parameters
EDOF-CNN-Max [148]	3410083
EDOF-CNN-3D [148]	2864035
EDOF-CNN-Fast (P)	526819
EDOF-CNN-Pairwise (P)	600931

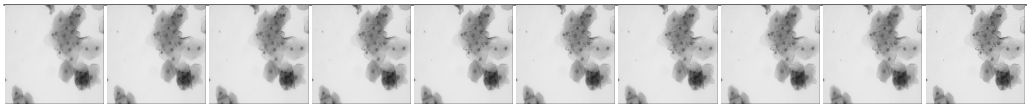
4.2.4 Experimental Details

4.2.4.1 Datasets

For this work two different datasets are used, one publicly available (Cervix93) and a new publicly available from Fraunhofer Portugal AICOS (EDoF Fraunhofer). A brief description of each dataset is presented in the next two subsections.

Cervix93 The Cervix93 cervical cytology image database is publicly available with annotations obtained from archived ThinPrep Pap-stained slides provided by the Moffitt Cancer Center (Tampa, FL) [178]. The dataset contains 93 different EDoF images obtained from 1602 image stacks at different z-axis positions. Each slide's top and bottom stack were manually determined at a manually chosen focal plane using the automated XYZ stepping motor and Stereologer software. The size of every image present in this database is 1280x960 pixels. The dataset was annotated according to the Bethesda System encompassing three different grades: Negative, Low-grade Squamous Intraepithelial Lesions (LSIL), and High-grade Squamous Intraepithelial Lesions (HSIL). For this work, it was used only 9 stacks for every EDoF image. Thus, a total of 837 stacks and 93 EDoF images are used. An example of stacks/EDoF is presented in Table 4.7.

Table 4.7: Example of several stacks and respective EDoF present in Cervix93 dataset.

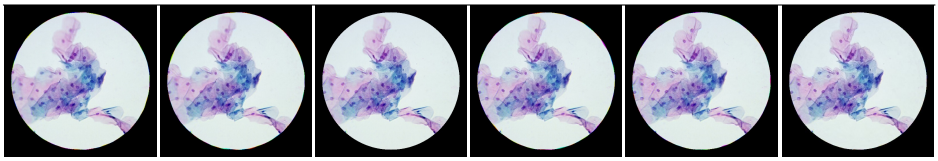
Stack									**EDoF**
									
1	2	3	4	5-(C)	6	7	8	9	-

EDoF Fraunhofer dataset The EDoF Fraunhofer dataset [7] is a new publicity available dataset composed of images of microscopic fields of view (FOV) of liquid-based cervical cytology (LBC) samples. A total of 15 LBC samples (ASC-US, LSIL, ASC-H, HSIL, and SCC) were supplied by the Pathology Anatomy Services from Hospital Fernando Fonseca and the Portuguese Oncology

[7] 10.5281/zenodo.7473277

Institute of Porto. For each LBC sample, a set of images were obtained using a version of the μSmartScope prototype [194,201] adapted to the cervical cytology use case [38,177]. This device is a portable 3D-printed prototype based on a smartphone developed by Fraunhofer AICOS that allows the fully automatic acquisition of microscopic images. The acquisition is done using the automated focus approach described in [194], where for each FOV, all images in the precise phase of the approach are stored as well as an indication of the image with the best focus metric (standard deviation of the Tenenbaum gradient [223]). The size of every image present in this database is 960x720 pixels. For this work, 144 EDoFs images generated with 5 aligned images per stack are used (where the central image is the one with the best focus metric (C). Table 4.8 shows an example of a stack and the respective EDoF.

Table 4.8: Example of a stack and respective EDoF present in EDoF Fraunhofer dataset.

Stack					**EDoF**
					
1	2	3-(C)	4	5	-

Complex wavelet transformation for EDof images generation To perform this work, it was necessary to generate EDoF images for EDoF Fraunhofer dataset to use as ground truth for CNNs training. Thus, the Complex wavelet extended depth of field method was used, which remains the traditional gold standard method for image fusion [76]. To perform EDoF generation with this method using 5 images per stack, we used ImageJ[8], a publicly available Java-based image processing program. This software was developed at the National Institutes of Health and the Laboratory for Optical and Computational Instrumentation (LOCI, University of Wisconsin) [204]. It was also necessary to add the Extended Depth of Field plugin[9] from the Biomedical Imaging Group (BIG) located in École Polytechnique Fédérale de Lausanne [76]. Thus, it was possible to generate high-resolution EDoF images for the EDoF Fraunhofer dataset.

4.2.4.2 Data pre-processing

The two datasets are partitioned into training and testing subsets (80-20%). The images from both datasets were resized to 520×520 pixels, this allows a faster computation without sacrificing too much the level of detail required for the task. Normalization was also carried out to scale the pixel values to 0–1, as is usual for deep learning models. This accelerates the training process and aids the network with regularization.

[8] `https://imagej.nih.gov/ij/index.html`
[9] `http://bigwww.epfl.ch/demo/edf/`

4.2.4.3 Evaluation Metrics

To evaluate the performance of the different models, four widely used metrics were adopted: mean squared error (MSE), root mean square error (RMSE), structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR).

The most used metrics for objective image quality assessment are the MSE and PSNR. The MSE consists of the measurement of the energy difference between the generated image (I^1) and the ideal/ reference one (I^2), where n is the number of pixels as shown in equation 4.5:

$$MSE(I^1, I^2) = \frac{1}{n} \sum_{i,j} \left(I^1_{i,j} - \hat{I^2_{i,j}} \right)^2 . \tag{4.5}$$

The RMSE metric is defined as the square root of MSE:

$$RMSE(I^1, I^2) = \sqrt{\frac{1}{n} \sum_{i,j} \left(I^1_{i,j} - \hat{I^2_{i,j}} \right)^2}. \tag{4.6}$$

The PSNR (equation 4.7) is a close relative of MSE, where L corresponds to the maximum dynamic range of the pixel number. In a standard 8-bit image L equals to 255 ($2^8 - 1 = 255$).

$$PSNR = 10 \times \log_{10} \left(\frac{L^2}{MSE} \right). \tag{4.7}$$

MSE has several good properties such as: it is easy to compute, it has a physically clear significance, it fulfills properties like symmetry, convexity, and differentiability and it is considered an outstanding measure in optimization applications. However, in both metrics (MSE/PSNR) there is a disregard for the location of the pixel, which is very important information for image quality assessment [236]. To overcome the weakness of MSE/PSNR, a new metric named structural similarity (SSIM) index has been developed to consider the location knowledge during a quality assessment experiment [235]. SSIM metric, present in equation 4.8, compares the luminance, contrast, and structural similarities as human visual perception do.

$$SSIM = \frac{1}{M} \sum_{i=1}^{M} l(r_i, d_i) \cdot \ c\,(r_i, d_i) \cdot s(r_i, d_i). \tag{4.8}$$

In the SSIM equation, M corresponds to the number of local windows in the image, and l, c, and s correspond to luminance, contrast, and structural similarity, respectively.

4.2.4.4 Training Details

The weights of the architectures previously mentioned are initialized randomly. Adam was used as the optimizer and started with a learning rate of 10^{-3}. The learning rate is reduced by 10% whenever the loss is stagnant for 10 epochs. The training process is completed after 200 epochs. The mean squared error (MSE) is used as a loss function. A fixed batch-size of 6 (6 x Z images of the stack) is used during the training.

The dataset was divided into five different k-folds. Therefore, the results are the average and standard deviation of these five folds. The complex wavelet EDoF method was run on an AMD Ryzen 7 2700x 3.7 GHZ CPU and the CNNs using an NVIDIA TESLA V100 GPU with 32 GB of memory. The code for EDoF-CNNs was written using the PyTorch framework.

4.2.5 Results and Discussion

The average performance for the 5-folds of the four different architectures are presented in Tables 4.9 and 4.10, for both the Cervix93 and EDoF Fraunhofer dataset, with the three different workflows for the last dataset: (A) RGB and rigid alignment; (B) RGB, rigid and elastic alignment and (C) RGB, rigid, and elastic alignment of the stacks. For Cervix93 dataset it was tested in four different configurations of z-planes of stacks: 3, 5, 7, and 9. The EDoF Fraunhofer dataset was only tested for 3 and 5 z-planes of stacks. The architectures performance was measured by mean squared error (MSE), root mean square error (RMSE), structural similarity index (SSIM), and peak signal-to-noise ratio (PSNR) as detailed in the previous section. The best models are shown in bold.

Table 4.9: Results for the different architectures on Cervix93 dataset.

EDoF-CNN	stack	MSE	RMSE	PSNR (dB)	SSIM (%)
Max [148]	3	0.01143 ± 0.00188	0.11913 ± 0.00861	22.547 ± 0.362	95.118 ± 0.549
	5	0.01139 ± 0.00186	0.11886 ± 0.00854	22.590 ± 0.375	95.243 ± 0.563
	7	0.01139 ± 0.00189	0.11875 ± 0.00865	22.621 ± 0.377	95.363 ± 0.554
	9	0.01152 ± 0.00185	0.11926 ± 0.00858	22.581 ± 0.375	95.268 ± 0.516
3D [148]	3	0.01144 ± 0.00188	0.11917 ± 0.00859	22.549 ± 0.372	95.061 ± 0.520
	5	0.01144 ± 0.00188	0.11914 ± 0.00863	22.554 ± 0.385	95.152 ± 0.527
	7	0.01143 ± 0.00188	0.11910 ± 0.00861	22.564 ± 0.379	95.219 ± 0.544
	9	0.01143 ± 0.00188	0.11901 ± 0.00862	22.582 ± 0.376	95.256 ± 0.513
Fast (P)	3	0.01137 ± 0.00188	0.11852 ± 0.00854	22.667 ± 0.392	95.746 ± 0.488
	5	0.01137 ± 0.00188	0.11839 ± 0.00855	22.697 ± 0.402	95.966 ± 0.499
	7	0.01135 ± 0.00190	$\mathbf{0.11816 \pm 0.00863}$	$\mathbf{22.751 \pm 0.401}$	96.193 ± 0.529
	9	0.01143 ± 0.00185	0.11855 ± 0.00869	22.701 ± 0.367	96.069 ± 0.629
Pairwise (P)	3	0.01136 ± 0.00184	0.11899 ± 0.00902	22.505 ± 0.530	95.392 ± 0.358
	5	0.01138 ± 0.00182	0.11894 ± 0.00895	22.520 ± 0.604	95.831 ± 0.343
	7	0.01142 ± 0.00184	0.11892 ± 0.00896	22.570 ± 0.559	96.218 ± 0.321
	9	$\mathbf{0.01130 \pm 0.00187}$	0.11819 ± 0.00902	22.670 ± 0.527	$\mathbf{96.323 \pm 0.311}$

Looking across the different architectures for Cervix93 dataset at Table 4.9, EDoF-CNN-Fast (P) and EDoF-CNN-Pairwise (P) achieved the best results. EDoF-CNN-Pairwise won in MSE and

Table 4.10: Results for the ablation study across the different architectures and workflows on EDoF Fraunhofer dataset.

		(A) Chromatic Alignment + Rigid Alignment			
EDoF-CNN	**stack**	**MSE**	**RMSE**	**PSNR (dB)**	**SSIM (%)**
Max [148]	3	0.00300 ± 0.00069	0.07253 ± 0.00879	28.254 ± 1.037	95.440 ± 0.811
	5	0.00300 ± 0.00069	0.07247 ± 0.00871	28.268 ± 1.013	95.746 ± 0.750
3D [148]	3	0.00303 ± 0.00068	0.07327 ± 0.00856	28.085 ± 0.940	95.384 ± 0.806
	5	0.00300 ± 0.00069	0.07259 ± 0.00874	28.231 ± 0.981	95.627 ± 0.772
Fast (P)	3	0.00293 ± 0.00069	0.07096 ± 0.00905	28.637 ± 1.123	95.868 ± 0.756
	5	$\mathbf{0.00292 \pm 0.00069}$	$\mathbf{0.07072 \pm 0.00899}$	$\mathbf{28.686 \pm 1.110}$	$\mathbf{96.052 \pm 0.776}$
Pairwise (P)	3	0.00295 ± 0.00069	0.07130 ± 0.00901	28.541 ± 1.100	95.817 ± 0.771
	5	0.00295 ± 0.00068	0.07134 ± 0.00862	28.514 ± 0.992	95.496 ± 0.804
		(B) Chromatic Alignment + Rigid Alignment + Elastic alignment			
EDoF-CNN	**stack**	**MSE**	**RMSE**	**PSNR (dB)**	**SSIM (%)**
Max [148]	3	0.00299 ± 0.00069	0.07233 ± 0.00882	28.359 ± 1.042	95.698 ± 0.753
	5	0.00298 ± 0.00069	0.07206 ± 0.00883	28.412 ± 1.027	95.648 ± 1.024
3D [148]	3	0.00299 ± 0.00069	0.07240 ± 0.00876	28.331 ± 0.983	95.668 ± 0.790
	5	0.00299 ± 0.00070	0.07242 ± 0.00890	28.328 ± 1.040	95.223 ± 0.953
Fast (P)	3	0.00291 ± 0.00069	0.07060 ± 0.00897	28.785 ± 1.108	95.833 ± 0.570
	5	$\mathbf{0.00290 \pm 0.00069}$	$\mathbf{0.07027 \pm 0.00907}$	$\mathbf{28.877 \pm 1.144}$	$\mathbf{96.187 \pm 0.659}$
Pairwise (P)	3	0.00293 ± 0.00069	0.07099 ± 0.00904	28.674 ± 1.125	95.725 ± 0.578
	5	0.00293 ± 0.00069	0.07099 ± 0.00884	28.650 ± 1.043	95.370 ± 1.244
		(C) Chromatic Alignment + Elastic Alignment			
EDoF-CNN	**stack**	**MSE**	**RMSE**	**PSNR (dB)**	**SSIM (%)**
Max [148]	3	0.00300 ± 0.00069	0.07259 ± 0.00870	28.282 ± 0.955	95.307 ± 0.865
	5	0.00299 ± 0.00069	0.07226 ± 0.00898	28.385 ± 1.104	95.378 ± 0.772
3D [148]	3	0.00301 ± 0.00071	0.07279 ± 0.00922	28.271 ± 1.097	95.609 ± 0.856
	5	0.00299 ± 0.00069	0.07248 ± 0.00880	28.321 ± 1.028	95.758 ± 0.788
Fast (P)	3	0.00291 ± 0.00069	0.07064 ± 0.00901	28.764 ± 1.128	95.929 ± 0.702
	5	$\mathbf{0.00289 \pm 0.00068}$	$\mathbf{0.07012 \pm 0.00887}$	$\mathbf{28.919 \pm 1.102}$	$\mathbf{96.193 \pm 0.748}$
Pairwise (P)	3	0.00293 ± 0.00069	0.07087 ± 0.00901	28.718 ± 1.108	95.874 ± 0.724
	5	0.00292 ± 0.00069	0.07074 ± 0.00911	28.759 ± 1.162	95.744 ± 1.325

SSIM metrics with 0.01130 and 96.323% respectively for 9 z-planes, on the other hand, EDoF-CNN-Fast obtained better results on RMSE and PSNR metrics with 0.011816 and 22.751dB respectively for 7 z-planes. Regarding the EDoF Fraunhofer dataset, the workflow (C) with only chromatic and elastic alignment achieved the best results using 5 z-planes with EDoF-CNN-Fast, getting an SSIM of 96.193%. In general, the results across the different proposed workflows are fairly close, however, it is possible to infer that elastic alignment promotes better results compared with rigid alignment. Thus, this fact indicates that elastic alignment can replace rigid alignment in low-cost microscopy pre-processing. Going a step further and comparing the results of literature models vs proposed models, it is possible to infer that the proposed algorithms surpass the literature models in all the metrics in both datasets. Even so, EdoF-CNN-Fast wins more times, EDoF-CNN-Pairwise remains very competitive in most of the metrics.

To understand the impact of chromatic alignment on the EDoF images quality Table A.1 is provided in the annex section A with results for the EDoF Fraunhofer dataset without chromatic alignment, only with elastic alignment as a pre-processing step. The results show a decrease in performance without the presence of chromatic alignment, with a drop of SSIM from 96.193 to 95.678 when compared with workflow (C) best model (EDoF-CNN-Fast 5 z-planes). The other metrics also decreased. These results support the use of chromatic alignment in the pre-processing workflow.

The use of chromatic and elastic alignment (workflow (C)), after the empirical assessment across the three different workflows, proved to be the best choice as pre-processing steps in low-cost microscopy before the generation of EDoF images as an image quality enhancement strategy.

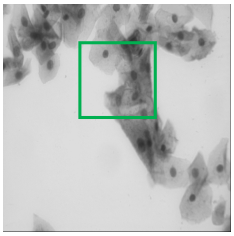
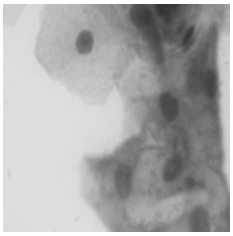
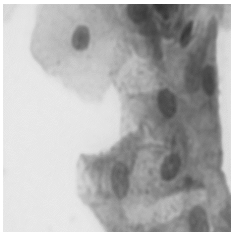
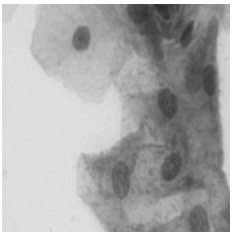
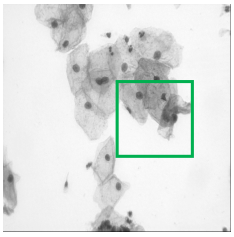
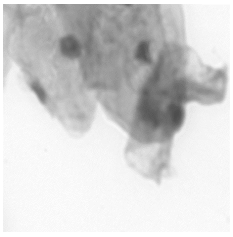
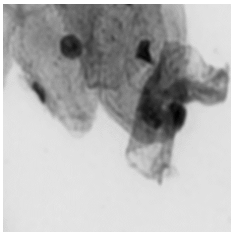
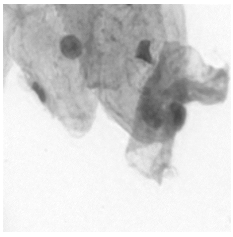
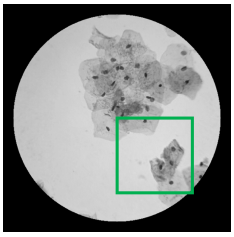
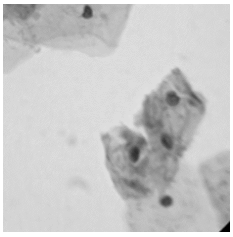
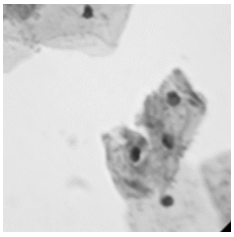
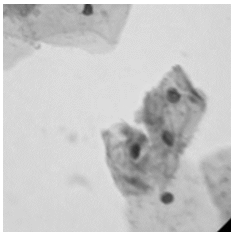
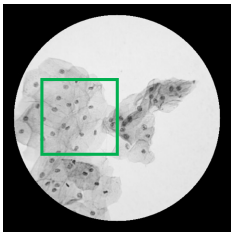
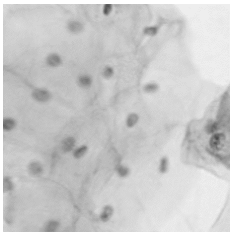
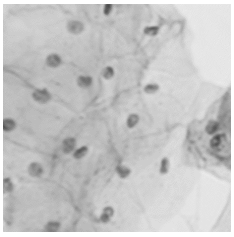
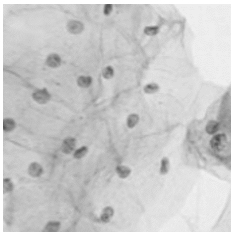
In Table 4.11, it is possible to visualize the EDoF images generated for Cervix93 and EDoF Fraunhofer datasets using EDoF-CNNs and complex wavelet transformation. The selection of the EDoF-CNNs used is based on the previous results mentioned above. The EDoF-CNN-Pairwise with 9 stacks is used for Cervix93 and the EDoF-CNN-Fast with 5 stacks including chromatic and elastic alignment as pre-processing steps is used for the EDoF Fraunhofer dataset. The central image from the z-planes stack is also included for comparison.

Comparing the different images in Table 4.11, it is possible to infer that both EDoF images obtained with EDoF-CNNs or the Complex wavelet fusion method present more spatial resolution and less blur compared with the central z-plane from the stacks. The differences between the fusion images obtained with the CNNs and the complex wavelet method are not so easy to visualize which indicates that deep-based methods are a good alternative for image enhancement in microscopy. Thus, Edof images obtained from EDoF-CNNs must be used in low-cost microscopy instead of using the central image from the stack as the reference image for diagnosis.

4.2.6 Summary and Conclusions

This work introduced a novel pre-processing workflow that can be used across all low-cost microscopy devices. The use of chromatic and elastic alignment proved to be the most efficient method in the pre-processing workflow. Furthermore, two novel deep-learning CNN-based EDoF networks (EDoF-CNN-Fast and EDoF-CNN-Pairwise) were proposed to generate Extended Depth

Table 4.11: EDoF images obtained using the different methods with a 2x zoom for better visualization of the details.

	Zoom Location	Central Image	EDoF-CNN	Complex Wavelet
Cervix93				
				
EDoF Fraunhofer				
				

of Field images. These new methods were compared with state-of-the-art approaches. Both proposed networks have surpassed the existing EDoF CNNs in literature in most of the experiments.

An end-to-end approach was presented that may be used in low-cost microscopy, starting with the collecting of the stacks with various focal planes and ending with the development of the EDoF images using a deep-based strategy. This approach can also be used to improve the performance of diagnosis models, that use low-cost microscopes to acquire the images. Thus, the proposed workflow can be used by the scientific community, the industry, and also medical audience.

As future work, an end-to-end CNN model should be explored to perform all the alignments necessary for the EDoF generation process. Thus, the network will eliminate pre-processing time, improving the speed of this method in real-time usage. One limitation of the proposed EDoF-CNN models is that they work only on grayscale stacks, they were also implemented for RGB images but they need to be tested and fine-tuned in future work. The embedded integration of these models on low-cost devices can also be explored.

4.3 Quality Control in Digital Pathology: Automatic Fragment Detection and Counting

Manual assessment of fragments during the processing of pathology specimens is critical to ensure that the material available for slide analysis matches that captured during grossing without losing valuable material during this process. However, this step is still performed manually, resulting in lost time and delays in making the complete case available for evaluation by the pathologist. To overcome this limitation, we developed an autonomous system that can detect and count the number of fragments contained on each slide. We applied and compared two different methods: conventional machine learning methods and deep convolutional network methods. For conventional machine learning methods, we tested a two-stage approach with a supervised classifier followed by unsupervised hierarchical clustering. In addition, Fast R-CNN and YOLOv5, two state-of-the-art deep learning models for detection, were used and compared. All experiments were performed on a dataset comprising 1276 images of colorectal biopsy and polypectomy specimens manually labeled for fragment/set detection. The best results were obtained with the YOLOv5 architecture with a $mAP@0.5$ of 0.977 for fragment/set detection.

4.3.1 Context and Motivation

When processing pathology specimens, it is critical to perform multiple checkpoints to ensure that the material available for further analysis is consistent with that recorded during gross preparation and that no valuable material is lost [78, 79]. This quality control can occur at different phases of the technical handling of the specimens and varies across different laboratories [93, 189]. One of the most important steps is to cross-check the number of fragments contained in the cassette with the number of fragments on the corresponding slide (after staining and mounting). This is performed manually by a biomedical scientist on the team, who compares the number of fragments

described in the macroscopic report with the number of fragments seen on the slide. In a digital pathology workflow, this manual step precludes taking full advantage of the lean system, since the slides have to be taken off the stainer racks, to verify the number of fragments and then the slides must be reassembled in the racks and fed into the scanners. This results in a loss of time and delays in obtaining the complete case ready for the pathologist's evaluation. A possible workaround would be to evaluate the fragments after scanning, but the manual comparison of the macroscopic report with the macro image retrieved from the scanner would also be very time-consuming. To overcome this handicap, we developed an automatic algorithm to detect and count the number of fragments on each slide and compare it against the macroscopic report information. The system then gives an alert if any slide has a discrepant number of fragments, allowing for subsequent review (Figure 4.11).

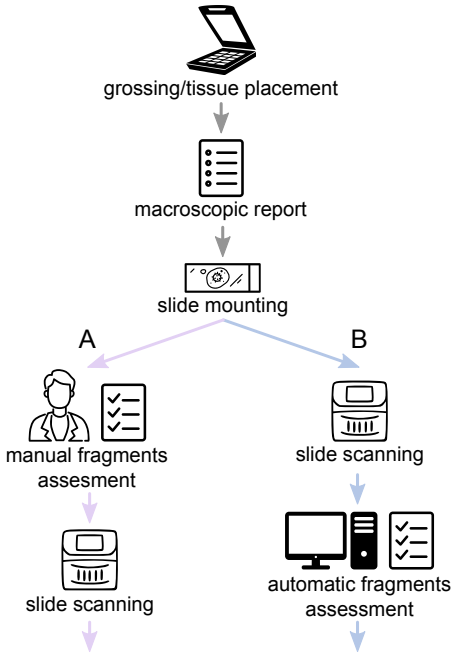


Figure 4.11: Number of fragments assessment checkpoint. On the first workflow (A.), a manual assessment of the number of fragments on the slides is made and confronted with the macroscopic report; the slides are posteriorly scanned. In the proposed workflow (B.), the system automatically evaluates the adequacy of the number of fragments on the slides after scanning, giving an alert message if a discrepancy is found.

To the best of our knowledge, this is the first study on this topic. Such an algorithm can be extremely valuable to pathology laboratories and departments. We expect that this system will reduce the workload of the team of biomedical scientists, allowing them to allocate more time to other, more important, tasks. Moreover, such analysis will also benefit the pre-processing stage of automatic diagnosis systems, by removing repeated fragments/sets from the region of interest to be fed to the models and also evaluate image quality per set [13]. This would avoid a manual selection step in case of repeated fragments/sets and also speed up the image analysis process.

In the next section, a brief description of the problem is presented, and in Section 4.3.3, we

detail the data pre-processing steps and the different models used in this work. The experimental details, including a detailed description of the dataset and used metrics, are stated in the Section 4.3.4, with the results being presented and discussed in Section 4.3.5. The study finishes with a conclusion section 4.3.7.

4.3.2 Brief Summary

The main goal of this work is to create models capable of effectively learning from the different tissue placement combinations, in order to detect and count the respective fragments and sets present in each Whole-slide image (WSI).

There are three possible cases during the slides mounting: one fragment repeated n times, where the number of sets is equal to the number of fragments (Figure 4.12.a); one set containing different fragments (Figure 4.12.b); and finally several copies of the same set of different fragments (Figure 4.12.c).

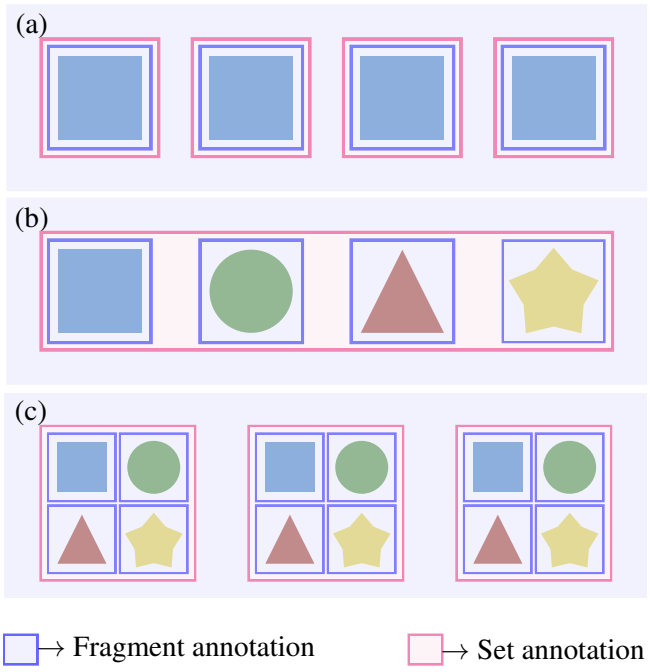


Figure 4.12: Schematic representation of the three different possibilities on the tissue slides preparation: (a) several sets of only one fragment; (b) one set of different fragments and (c) several equal sets of different fragments.

Fragment repetitions occur to ensure that there is no loss of slide material due to artifacts, poor focus, or other causes. Thus, even if one of the repetitions is affected, it will not affect the decision by the clinicians as they can support their decision in the other set of repetitions. The high variability that exists between the fragments, in the combinations of fragments, and their orientation/arrangement on the slides (Figure 4.13 and 4.16) makes this detection and counting task a big challenge.

Two different types of methodologies were adopted to solve the above-mentioned problem: conventional machine learning methods and deep convolutional network methods. In conventional machine learning methods, a two-stage approach was tested, with a supervised classifier, using Linear Discriminant Analysis (LDA), Quadratic Discriminant Analysis (QDA), Naive Bayes classifiers (NB), Logistic Regression Classifier (LR) and Support Vector Machines (SVM), followed by unsupervised hierarchical clustering. Regarding deep learning strategies, Fast R-CNN and YOLOv5, two state-of-the-art detection deep learning models, were used and compared.

4.3.3 Methods

4.3.3.1 Conventional Algorithms

The more conventional machine learning algorithm proposed was based on binarizing an image, identifying its connected components, and then grouping them into fragments or sets. A connected component of a binary image is an object where all pixels have a value of 1 and at least one pixel with a value of 1 in their 8-neighborhood. Note that two distinct connected components may actually belong to the same fragment object. This may be the case, for example, when a fragment is disrupted during the slide preparation step, resulting in multiple partial fragments. The proposed method can be divided into 3 parts: feature extraction, binary classification, and grouping.

Feature extraction: In this step, a set of representative features is extracted for each pair of connected components in each image to decide whether or not they are the same fragment or set. In order to further remove noise, we only considered the components whose areas were above 5% of the area of the biggest component. Then, for each remaining component, several features were extracted and used to calculate pairwise features: the distance between centroids, the 25th percentile distance between every point in the objects' contours, the normalized area of the entire bounding box, and the ratio of the circularity ratios of both components. These pairwise features were chosen based on theoretical knowledge of the task at hand.

Binary classification: Now, we want to predict whether a pair of connected components are part of the same group: we had to train each type of classifier twice, once to identify if two components are from the same fragment and another time to check if they are from the same set. For this purpose, each pair in the training and validation sets was labeled according to the ground truth bounding boxes (fragments or sets). Five classifiers (LDA, QDA, NB, LR, and SVM) were trained with cross-validation, ensuring that each image was represented in only one of a total of three folds. The best parameters for each model representation were selected and the entire training set was used to train the final models.

Grouping: The end goal is a robust algorithm that can group all connected components into the respective fragments and sets. To do this, we relied on graphs where the weight of the connection between two nodes (representing two connected components) is equal to the probability that this pair belongs to the same fragment or set. Thus, we created two graphs, one, called Strong Graph, where pairs are connected if the prediction (as given by the pairwise classifier) is higher than a restrictive threshold (0.9), and the other, Weak Graph, which is fully-connected, where all pairwise

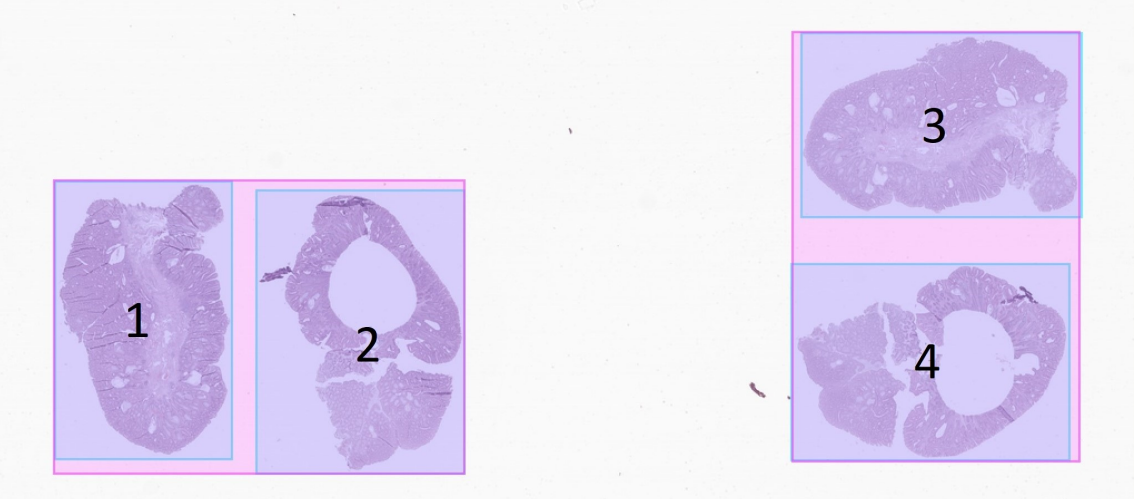


Figure 4.13: Considering the set identification task, pairs 1-2 and 3-4 should be assigned with label 1, while all the other possible pairs should be assigned with label 0, for this image.

predictions are stored. After identifying all connected subgraphs in the Strong Graph, we wanted to consider additional connections between pairs of strong subgraphs if they represented the same fragment/set. For a certain strong subgraph A, we analyzed all other strong subgraphs in pairs (A,B) by determining the mean of the predictions between every component of subgraph A to all components of subgraph B, represented in Fig.4 with grey lines. An additional connection between subgraphs was considered if the mean for the pair (A,B) was the highest mean found for subgraph A and if its value was higher than a permissive threshold (T=0.5). The process was repeated for all subgraphs until no more pair of subgraphs met the criterion.

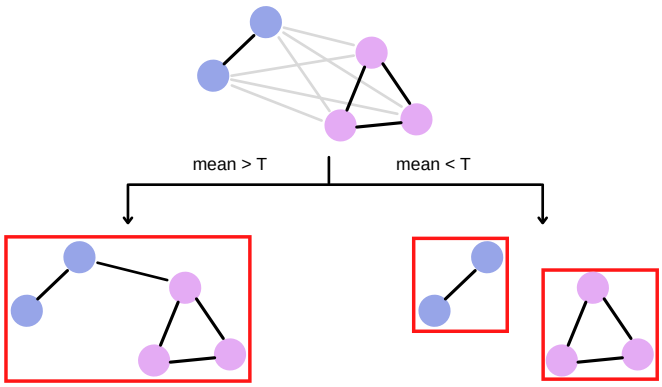


Figure 4.14: Schematic representation of the methodology used for sets grouping. Each circle represents a component. The strong graphs are represented with black connection lines. The grey lines, together with the black lines, constitute the weak graph. The final red bounding boxes are the two possible results for this step.

Additionally, we assessed the success of this proposal independently of the performance of the pre-processing step, since this step could either remove some of the tissues objects or keep some undesired noise, both problems that would decrease the performance of the method as a whole.

For this,instead of pre-processing and obtaining the connected components on the full image, we did this for each tile limited by the ground truth fragments' bounding boxes (blue boxes in Fig.3). With the identified components, the next steps remained as described above.

4.3.3.2 Deep Learning Models

To compare the more traditional approach described above with a deep learning model, two state-of-the-art detection architectures were used: the Fast R-CNN (Fast Region-based Convolutional Network method) [192] and the YOLOv5 (You Only Look Once) neural network [116]. These models were used to detect and classify fragments and sets, as shown in Figure 4.15.

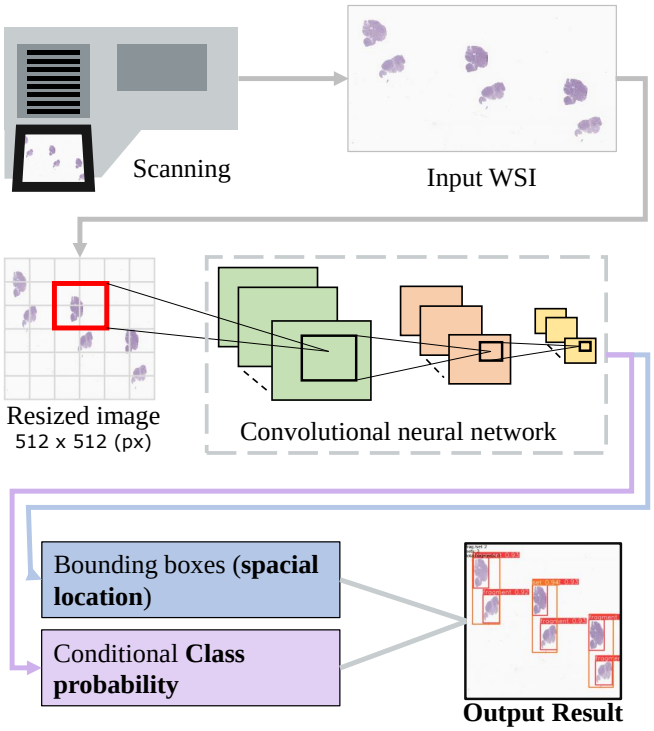


Figure 4.15: Schematic representation of the proposed framework.

The Faster R-CNN consists of two different modules. The first module is the Region Proposal Network (RPN), which detects the region of the object in the image, and the second module, in which a convolutional neural network that classifies the objects in the proposed region. Thus, the Faster R-CNN must first process the input image with a feature extractor to obtain feature maps and pass them to the RPN. The original Faster R-CNN uses the VGG-16 in the convolutional part. However, in this work, an improved version using MobileNetV3 is used, which allows this model to be used in mobile solutions [192].

The YOLOv5 neural network predicts a set of bounding boxes and class probabilities. First, it divides the entire image into several grids of different sizes and then creates anchor boxes in each grid of the input image using a predetermined scale and size. Unlike a two-stage detector like Faster R-CNN, each anchor box predicts the objectness score, box center offset x, box center

offset y, box width, box height, and class scores all at once. Thus, YOLO is a single-stage object detector that quickly detects objects.

4.3.4 Experimental Details

4.3.4.1 Data pre-processing

Due to the large size of each WSI (with thousands of pixels in both dimensions) and since object detection/ counting does require overly high tissue resolution, the slides were first reduced in size so that the largest dimension of the image was 1024 px. This enables faster computation without sacrificing the level of detail required for the task.

Regarding image pre-processing for conventional machine learning techniques, the original slides were transformed into binary images. For this purpose, firstly, any possible black spots were removed with a region-growing algorithm in the blue channel since it presents a higher contrast between tissues and these spots. Then the grayscale image was binarized through Otsu's adaptive thresholding. After morphological operations to remove noise, the image was reversed, resulting in white tissue objects on a dark background.

The deep learning models were fed with the colored images, resized to 512×512px, to match the squared input size of the chosen architectures. To speed up the training process, the input pixels were rescaled from the range of 0-255 to 0-1.

4.3.4.2 Dataset

This study includes 1276 samples from colorectal biopsies and polypectomies specimens and leverages our previous work on CAD systems for colorectal cancinoma [168]. All cases were retrieved from the data archive of the IMP Diagnostics Laboratory, in Portugal, and digitised with 2 Leica GT450 WSI scanners, at 40x equivalent magnification. Each slide could contain one or more fragments, as shown in the examples in Figure 4.16. Each set of fragments could also be repeated within the same slide (the same fragments could be cut more than once in order to increase the levels to be evaluated during subsequent examination). All cases were evaluated by one pathologist and two biomedical scientists and the number of fragments was compared with the original macroscopic report. The dataset contains two different types of annotations: fragments (11,300 annotations) and sets (3,517 annotations). The annotation process follows the annotation logic shown in Figure 4.12. If there was a discrepancy between the original report and the review, the case was decided between the three evaluators. The dataset was split into a training and test subset (80-20%), maintaining the ratio between the different cases.

4.3.4.3 Evaluation Metrics

To evaluate the performance of the different models, four different metrics were adopted: Intersection over Union (IoU), precision (P), recall (R) and mean Average Precision (mAP).

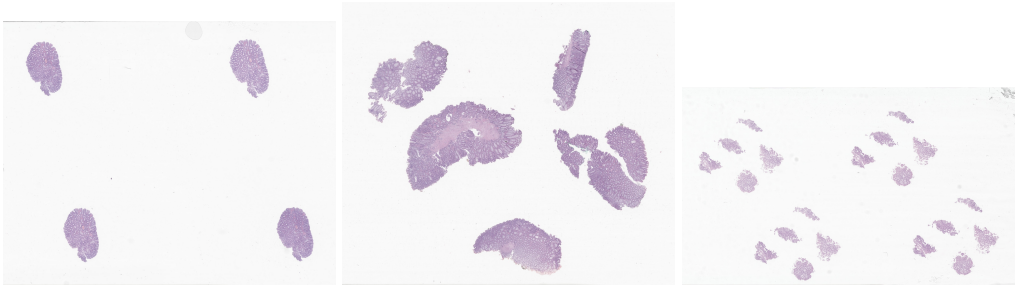


Figure 4.16: Three different examples of macro slide images present in the dataset.

The IoU, based on the Jaccard Index, measures the area of overlap between the predicted bounding box BB_P and the ground-truth bounding box BB_{GT} divided by the area of union between them, i.e.:

$$IoU(BB_P, BB_{GT}) = \frac{|BB_P \cap BB_{GT}|}{|BB_P \cup BB_{GT}|}. \tag{4.9}$$

It is possible to classify a detection as correct or wrong by comparing the IoU with a defined threshold t. If $IoU < t$ the detection is considered wrong.

The precision (P) and recall (R) metrics are also used to evaluate object detection algorithms, and they are defined as:

$$P = \frac{TP}{TP+FP}; R = \frac{TP}{TP+FN}. \tag{4.10}$$

Where TP refers to True Positives, FP to False Positives, and FN to False Negatives. Precision refers to the ability of the model to recognize only relevant objects. It is the percentage of positive predictions that are correct. The ability of a model to locate all relevant objects is called recall (all ground-truth bounding boxes). It is the proportion of positive predictions that are correct from all available ground truths.

The most common definition for the Average Precision (AP) is to look for the area under the precision-recall curve. The mean Average Precision (mAP) for object detection is the average of the AP calculated for all the classes. $mAP@0.5$ means that it is the mAP calculated at IoU with a threshold of 0.5. It can be defined as:

$$mAP = \frac{1}{N} \sum_{i=1}^{N} AP_{i'}. \tag{4.11}$$

where AP_i is the AP in the ith class and N is the total number of classes being evaluated.

4.3.4.4 Training Details

For the deep learning models, a grid search was first performed to find the best combination of hyperparameters, varying the image size (224x224, 512x512, and 640x640); batch size (8, 16, 32, and 64), and also the training epochs (100, 200, 300 and 400). For YOLOv5 the best parameters found for the fine-tuned model were an image size of 512x512, a batch size of 32, and 200 training

epochs. Concerning Faster R-CNN, the parameter that differs from YOLOv5 is the batch size, which turns out to be the best at 8. These hyper-parameters were chosen concerning the test subset. Stochastic Gradient Descent was used as the optimizer and started with a learning rate of 10^{-4} for both models. All the experiments are implemented in PyTorch.

4.3.5 Results and Discussion

Table 4.12: Results of the different models for the fragment/set detection task. (GT - Ground Truth; CL - Component Labelling).

	Fragment				**Set**				**All**			
	IoU	P	R	mAP@0.5	IoU	P	R	mAP@0.5	IoU	P	R	mAP@0.5
QDA (CL)	0.631	0.828	0.775	0.716	0.501	0.599	0.861	0.469	0.566	0.714	0.818	0.592
QDA (GT)	0.658	0.860	0.792	0.750	0.587	0.690	*0.883*	0.576	0.622	0.775	*0.837*	0.663
LDA (CL)	0.689	0.977	0.693	0.834	0.712	0.935	0.760	0.812	0.701	0.956	0.726	0.823
LDA (GT)	0.684	**1.000**	0.684	0.811	0.725	0.908	0.812	0.816	0.704	0.954	0.748	0.813
NB (CL)	0.631	0.808	0.789	0.696	0.537	0.691	0.811	0.532	0.584	0.750	0.800	0.614
NB (GT)	0.660	0.842	*0.807*	0.713	0.592	0.739	0.844	0.598	0.626	0.790	0.825	0.656
LR (CL)	0.699	0.977	0.703	0.848	0.733	0.935	0.775	0.837	0.716	0.956	0.739	0.842
LR (GT)	0.711	0.996	0.715	0.850	*0.766*	*0.936*	0.826	*0.875*	*0.738*	**0.966**	0.771	*0.863*
SVM (CL)	0.704	0.976	0.709	0.861	0.678	0.931	0.728	0.767	0.691	0.953	0.718	0.814
SVM (GT)	*0.717*	0.997	0.720	*0.864*	0.700	0.916	0.780	0.778	0.708	0.956	0.750	0.821
Faster R-CNN	0.853	0.915	0.589	0.942	0.884	0.828	0.667	0.878	0.858	0.872	0.628	0.910
YOLOv5	**0.930**	0.941	**0.940**	**0.970**	**0.920**	**0.957**	**0.969**	**0.985**	**0.925**	0.949	**0.955**	**0.977**

The performance of the conventional and deep learning algorithms is presented in Table 4.12, for the fragment/set detection and count problem. The best results per metric, for conventional algorithms, are shown in italic, while the best results for both deep learning models and conventional algorithms are shown in bold.

When evaluating the results from conventional algorithms, it is possible to note that metrics improve when using the ground truth fragments (GT) across all different classifiers used, compared to using component labeling on the full image (CL). However, this improvement is not as significant as expected, which indicates that the pre-processing algorithm has a good performance, removing noise without damaging the objects in the image. Although all traditional models used are very competitive, logistic regression (LR) is found to be the top performer when comparing GT or CL results for all class detection. The $mAP@0.5$ metric is one of the most relevant metrics in object detection and LR achieves values of 0.842 and 0.863 for CL and GT, respectively, with LDA and SVM's being close seconds. In fact, both LDA and LR are consistent with their performance, having always $mAP@0.5$ values of over 0.8 when considering fragment, set or both classes detection. This is not the case for SVM, although being the top performer for fragment detection, has lower $mAP@0.5$ values for set detection.

Going a step further and comparing the results of conventional vs. deep learning models, it is possible to infer that deep learning algorithms surpass conventional algorithms in most metrics. YOLOv5 achieved the best results with a $mAP@0.5$ of 0.977 for all classes. Fast R-CNN remains very competitive in several metrics such as precision and $mAP@0.5$, however, it is more difficult to find the best hyperparameters and requires more time to train than YOLOv5.

4.3.6 Improvements in Automatic Fragment Counting

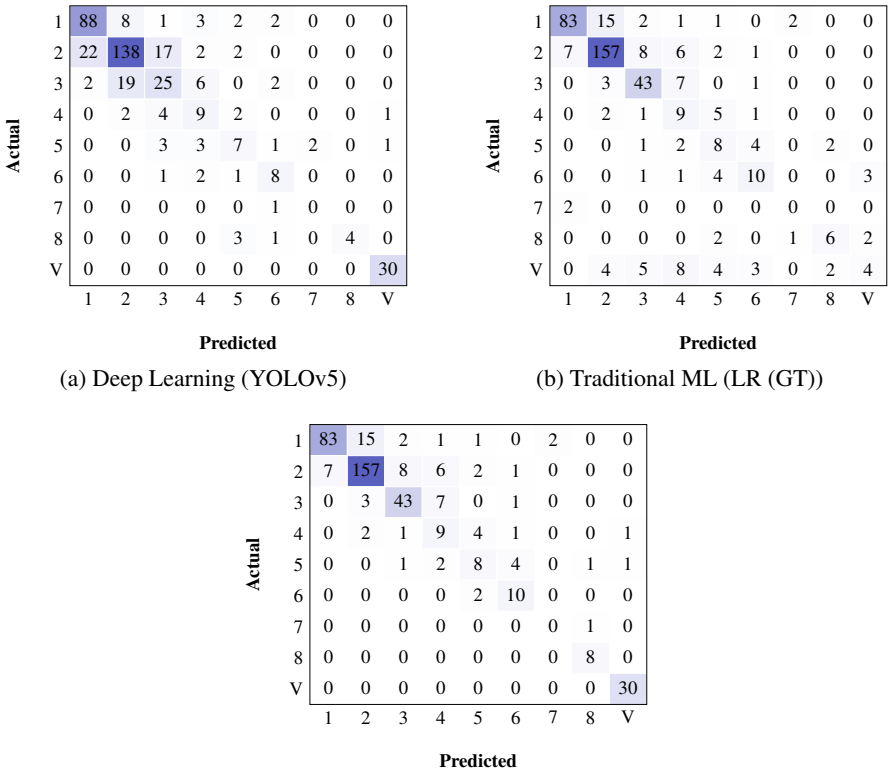
The principal focus of the previous subsections was on fragment/set detection; this section will explore the most important challenges and achievements in the counting process. To evaluate the performance of our algorithms, the best two models from the detection task were selected and evaluated in a new dataset containing 585 slides of different specimens of tissue (including cervical slides). Yolov5, from the deep learning models, and Logistic Regression (using the ground truth labels in the training process), from the traditional machine learning methods, were selected. The results for both approaches are presented in Table 4.13. The confusion matrices for each model are presented in Figure 4.17.

Table 4.13: Results of the three different models for the fragment/set counting task.

Metrics	**Deep Learning (YOLOv5)**	**Traditional ML (LR(GT))**	**Deep Learning + Traditional ML**
Accuracy (%)	73.27	75.29	82.67
MAE	0.409	0.583	0.287
MSE	0.861	2.230	0.687
Kendall Tau (%)	77.13	79.07	84.70
F1 (%)	72.95	75.14	82.42
Precision (%)	73.83	80.60	83.47
Recall (%)	72.71	75.29	81.88

After a detailed analysis of the results of Table 4.13 and Figure 4.17, it can be inferred that YOLOv5 works better in counting large amounts of fragments on the slides and Logistic Regression when there are few fragments per set. In this way, the idea of merging both models for counting the fragments arises. Joining both models increases the robustness of our approach against several issues such as tissue fragmentation, differences between tissue sizes across the different slides, and the intra-variability in the sizes of the fragments in the same slide. A schematic diagram of the proposed model is presented in Figure 4.18.

The join model starts with the slide as input for YOLOv5. After passing through the deep learning model, the counting decision is based on a threshold. If the deep model predicts more than seven fragments per set, the decision is accepted and the join model ends. On the other hand, if predicts less or equal to seven fragments, the counting decision is based on the Logistic Regression model.



(a) Deep Learning (YOLOv5)

(b) Traditional ML (LR (GT))

(c) Deep Learning + Traditional ML

Figure 4.17: Confusion Matrices of the three different models for the fragment/set counting task.

By merging the two models, all metrics evaluated in Table 4.13 were improved, showing favorable performance gains when using both machine learning methods. The cases where the model fails need to be studied to provide the model with more relevant information. The lack of prior knowledge in the models about tissue fragmentation leads to misclassification in several cases.

4.3.7 Summary and Conclusions

This study introduces a variety of both conventional and deep learning models designed for the purpose of detecting and quantifying fragments or sets within WSI. This novel image quality control method is recommended for implementation prior to employing AI algorithms for diagnosis, ensuring optimal algorithm performance. The newly developed fragment detection and counting system serves as a critical quality control step in this process. It automatically identifies instances where slide images exhibit fragment counts that deviate from the numbers reported in the macroscopic report, subsequently issuing warning messages as necessary. Notably, the YOLOv5 neural network consistently outperformed other detection models across multiple evaluation metrics, underscoring the substantial performance advantages associated with the application of deep learning techniques in the automated identification of slides with fragment count disparities in comparison to the macroscopic report.

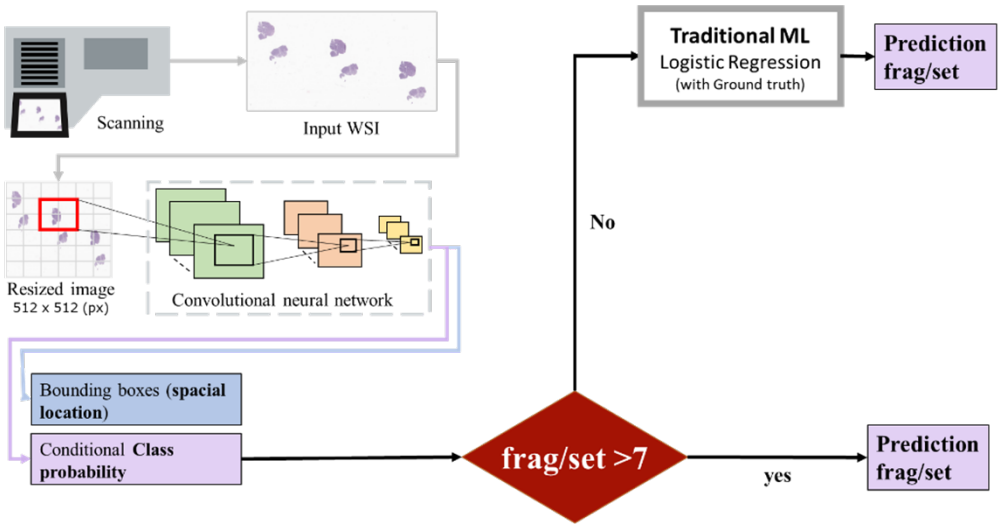


Figure 4.18: Schematic representation of the fusion model for fragment counting task.

Combining deep learning and traditional machine learning methods yielded superior outcomes for the task of fragment/set counting in comparison to using these methods individually.

In forthcoming research endeavors, it is essential to conduct a more extensive validation of the presented model. This validation should encompass a wider array of surgical specimen types and involve a larger sample size. This robust validation process will be instrumental in facilitating the seamless integration of the model into the routine operations of digital pathology workflows.

Chapter 5

Classification of Cancer Risk

Foreword on Author Contributions
The results of this work have been disseminated in the form of three papers in international journals, one in submission process.

[7] - Tomé Albuquerque, Ricardo Cruz, Jaime S. Cardoso. "Ordinal losses for classification of cervical cancer risk". PeerJ Computer Science 7:e457, 2021. https://doi.org/10.7717/peerj-cs.457

[8] - Tomé Albuquerque, Ricardo Cruz, and Jaime S. Cardoso. "Quasi-Unimodal Distributions for Ordinal Classification". Mathematics 2022, 10, 980. https://doi.org/10.3390/math10060980

[45] - Jaime S. Cardoso, Ricardo Cruz, and Tomé Albuquerque. Unimodal distributions for ordinal regression, 2023. (submitted to IEEE Transactions on Artificial Intelligence journal - waiting for decision) https://doi.org/10.48550/arXiv.2303.04547 - My contributions consisted of helping in the development of the work and paper writing process.

Cancer risk assessment is a critical component of modern healthcare, as early detection and precise risk stratification are paramount in improving patient outcomes. In recent years, the integration of deep learning techniques into cancer risk classification has sparked a transformative paradigm shift. Deep learning has demonstrated exceptional effectiveness in extracting intricate patterns and insights from complex medical data, thereby revolutionizing the field of oncology. However, current approaches do not take advantage of the ordinal information between the cancer stages. This chapter proposes four different ordinal approaches for the classification of cervical cancer risk, aiming to leverage the inherent ordinal nature of the data and enhance the performance of the classification models.

5.1 Ordinal Classification of Cervical Cancer Risk

Cervical cancer automatic screening via Pap smear is a highly valuable cell imaging-based detection tool, where cells must be classified as being within one of a multitude of ordinal classes, ranging from abnormal to normal. Current approaches to ordinal inference for neural networks

are found to not sufficiently take advantage of the ordinal problem or to be too uncompromising. This chapter is focused on the cervical cytology diagnostic method. A cervical cytology test is used to detect potentially abnormal cells from the uterine cervix. These premalignant dysplastic changes of cells are classified in progressive stages: seven stages by the World Health Organization classification (WHO) system or four stages by The Bethesda classification system (TBS) [65]. The risk of developing cancer is especially pronounced in the later stages. Therefore, distinguishing between the stages can be crucial for diagnosis. In most literature, the classification of Pap smear images consists of a binary separation between normal and abnormal cell (two classes), using different methodologies such as Support Vector Machines (SVM) [35, 49, 51, 119], k-Nearest Neighbours (kNN) [35, 49, 73, 149], Fuzzy c-Means Algorithm (FCM) [49, 239], k-Means clustering [175], Artificial Neural Networks (ANN) [49], and, more recently, Convolutional Neural Networks (CNN) [126, 138, 257].

However, all this work consists of binary classification, which is useful for screening, but not enough for a confident diagnosis. Fewer works explore the multi-class classification of cervical cells on the Herlev dataset.The authors of [49] proposed a multi-class automatic cervical cancer cell classification system using different classifiers, such as FCM, ANN, and kNN. Nevertheless, this system is based only on 9 cell-based features. The approach applies feature extraction from the nucleus and cytoplasm in each image and requires manual selection of the best threshold to minimize the error when applying the classifier to construct the cell mask. More recently, [126] performed feature extraction in a more autonomous way using a CNN. The use of a CNN simplifies the pre-processing steps that were necessary for the approach by Chankong et al. The authors of [86] proposed a new approach for multi-class cervical cancer cell detection and classification, using in the first step, CNNs to extract deep-learned features and in the second step, extreme learning machine (ELM)-based classifiers to classify the input cell images. [138] proposed a new CNN-based method that combines cell image appearance with cell morphology for multi-class classification of cervical cells in the Herlev dataset. In all these cases, cross-entropy is adopted for ordinal data classification.

Since the emergence of deep learning models, cross-entropy has been the most used loss function due to its simplicity, differentiability, and competence in dealing with many classes. When a model is trained with cross-entropy, there is no prior knowledge of the class labels, which means that all semantic relationships that might exist between the labels are ignored. However, several classification tasks expose a natural order/structure between the labels, which means that the labels present a specific order that may be useful as prior knowledge for the deep models [31]. Cancer grading is a typical example of an ordinal problem, wherein images show a natural progression from normal cells undergoing a gradual process to cancer cells and presenting different stages between the two terminal stages (normal and cancer) [7].

While cross-entropy focuses on each output probability in isolation, ordinal classification focuses on the relative order of the output probabilities. The rationale is that the error between two neighbor classes should be lower than the error between two distant classes. This makes ordinal classification more suitable in a variety of situations such as age estimation [140, 174, 263], movies

rating [59, 123], cancer grading [84], market bonds rating [160], emotion estimation [114, 242, 262], photographs dating [172], diabetic retinopathy grading [28], gene expression analysis [46], in addition to many others.

In ordinal regression, many works have incorporated ordinality into models and loss functions by promoting the probability output to be unimodal. However, current approaches rely on heuristics that lack a theoretical foundation. In the next subsections, we analyze the set of unimodal distributions in the probability simplex and establish fundamental properties.

5.1.1 Motivation and Context

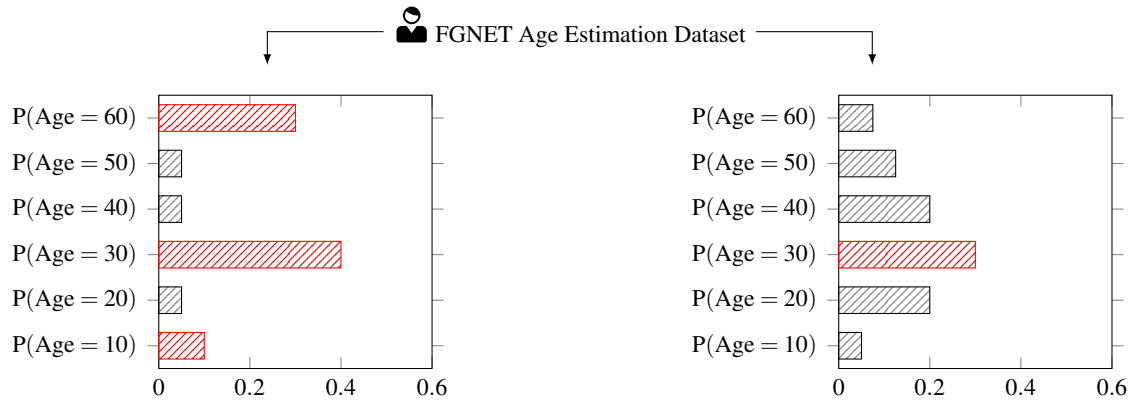
Ordinal regression (often called ordinal classification) is applied to data where the features of the n-th example $\mathbf{x}_n \in \mathscr{X}$ correspond to a label from a set of elements $\mathscr{C} := \{c_1, \ldots, c_K\}$ that have a well-defined ranking or order $c_1 \prec c_2 \prec \cdots \prec c_K$. However, unlike traditional metric regression, quantitative differences or distances cannot be assumed to exist between classes. The goal is to find a reliable rule or regression function $h\colon \mathscr{X} \to \mathscr{C}$ from the domain of features $\mathscr{X}$ to the domain of ordinal labels $\mathscr{C}$.

In the published literature for applied problems, it is still common to ignore the order of the labels and apply categorical algorithms to such data, often leading to the application of categorical cross-entropy loss (CE) in neural networks. Problematically, categorical loss assumes that all mislabelling by h is equally wrong. Although the problem with this practice has been known for more than 40 years [150], it is still common to implicitly or explicitly assume that ordinal data or labels exist on an interval or ratio scale [20].

Among methods recently developed specifically for ordinal regression, unimodal constraints on the probability distribution of the output have proven to be strong contenders. For a class c_ℓ for which the model outputs the highest posterior probability, $\ell = \arg_k \max \mathrm{P}(y = c_k \,|\, \mathbf{x})$, it would be expected for the second-highest posterior probability to be either $\mathrm{P}(y = c_{\ell-1} \,|\, \mathbf{x})$ or $\mathrm{P}(y = c_{\ell+1} \,|\, \mathbf{x})$, given the previously defined class order, $c_{\ell-1} \prec c_\ell \prec c_{\ell+1}$. The previous reasoning also applies to the ranking order from the third, fourth, et cetera posterior probabilities. For example, it makes no sense for a model to produce a high probability for “Hot” and for “Cold”, but a small probability for “Warm”. Furthermore, it is common for ordinal regression metrics to penalize errors more strongly when the distance between the predicted and the true class is farther apart than when it is closer. Some examples of output distributions are provided by Figure 5.1; only the unimodal is consistent with ordinal regression.

5.1.2 Task definition and related work

Assume that in a classification task, the instances have one of K classes, whose labels are c_1 to c_K, which correspond to the natural order of the ordinal classes.



(a) Multimodal output distribution.

(b) Unimodal output distribution.

Figure 5.1: Example of possible output probability distributions. Even if both outputs agree on the majority class, only the unimodal distribution is consistent with an ordinal regression task.

Typically, a neural network is trained to perform multi-class classification by minimizing the cross-entropy loss for the entire training set,

$$\mathrm{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = -\sum_{k=1}^{K} y_{nk} \log(\hat{y}_{nk}), \tag{5.1}$$

where, for the n-th observation, $\mathbf{y}_n = [y_{n1} \dots y_{nk} \dots y_{nK}]$, with $y_{nk} \in \{0,1\}$, represents the respective one-hot encoding of the corresponding label $c_n \in \mathscr{C}$ and $\hat{\mathbf{y}}_n = [\hat{y}_{n1} \dots \hat{y}_{nk} \dots \hat{y}_{nK}]$, with $\hat{y}_{nk} \in [0,1]$, being the respective vector of output probabilities assigned by the model for that n-th observation. Naturally, $\sum_{k=1}^{K} y_{nk} = \sum_{k=1}^{K} \hat{y}_{nk} = 1$, so $\mathbf{y}_n$ and $\hat{\mathbf{y}}_n$ are valid distributions in the $(K-1)$ dimensional probability simplex.

However, CE has limitations when applied to ordinal data. Defining $k_n^\star \in \{1, \dots, K\}$ as the index or rank of the true class of observation $\mathbf{x}_n$ (the position where $y_{nk} = 1$ in the one hot encoding vector $\mathbf{y}_n$), it is then clear that

$$\mathrm{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = -\log(\hat{y}_{nk_n^\star}). \tag{5.2}$$

5.1.2.1 Beyond Cross-entropy

Intuitively, CE just tries to maximize the probability in the output corresponding to the true class, ignoring all the other probabilities. For this loss, an error between classes c_1 and c_2 is treated as the same as an error between c_1 and c_K, which is undesirable for ordinal problems.

Furthermore, the loss does not constrain the model to produce unimodal probabilities, so inconsistencies can be produced such as $\hat{y}_{nj} > \hat{y}_{n\ell} < \hat{y}_{ni}$, even when $1 \leq j < \ell < i \leq K$. Cross-entropy is a fair approach for nominal data, where no additional information is available. By concentrating just on the mode of the distribution and disregarding all other values in the output probability vector, the ordinal information inherent in the data is ignored. However, for ordinal data, the order can be explored to further regularize learning.

Ordinal regression methods include a variety of techniques, which can be organized using different rationales. Gutiérrez et al. [90] proposed a taxonomy of ordinal binary decomposition approaches (decompose the ordinal problem into several binary ones, which are separately solved by multiple models or by one multiple-output model), naïve approaches (the model is obtained by using other standard machine learning prediction algorithms), and threshold methods (based on the general idea of approximating a real value predictor and then dividing the real line into intervals).

Ordinal Encoding (OE): Focusing only on probability-based approaches, cumulative link models were one of the first approaches to emerge [80, 150]. Classes are encoded using a cumulative distribution. Therefore, y_{nk} is represented by 1 if $k \leq k_n^\star$ and 0 otherwise. Notice that the model should produce class probabilities whose difference should tend to be incremental, since it represents a cumulative distribution. This way of introducing ordinality has the advantage of being independent of the way the model is optimized. During inference, to convert back the cumulative distribution to the individual probability of a class k, then the differences between classes should be produced $p_{nk} = y_{nk} - y_{nk-1}$, except for the first class [52, 80].

Another loss, CDW-CE, promotes ordinality by weighting CE with MAE (mean absolute error) [183].

More recently, several works have promoted or forced the output distribution to be unimodal. This can be achieved either by the construction of the model or by the use of an appropriate loss during training.

Binomial Unimodal (BU): A parametric approach to ensure this is by restricting the output of the model to that of a discrete probability distribution, such as the Binomial or Poisson. For the Binomial distribution [58], $\mathscr{B}(n,p)$ the support of the distribution is known, $n = K-1$, and the only parameter left to be estimated is the shape of the distribution, p. The model outputs a single output, the shape of the distribution, which is then converted into posterior probabilities using the Binomial probability mass function, $\mathrm{P}(y = c_k \,|\, \mathbf{x}) = \binom{n}{k-1} p^{k-1}(1-p)^{n-k-1}$, for each $k \in \{1,\ldots,K\}$, where p is the output produced by the model. This may be seen as an activation function. To train the model, a common loss such as cross-entropy may then be used.

Poisson Unimodal (PU): Similarly, the constraint for a discrete unimodal probability is enforced by a Poisson probability mass function [28]. On the final layer, the log Poisson probability mass function is applied, followed by a softmax activation function to normalize the output as a probability distribution. According to the authors, it tends to not work as well for more than eight labels [28]. Furthermore, [28] also proposes to control the variance of the distribution through a learnable softmax temperature term (τ).

Gaussian Distribution: In [18], the authors propose a diabetic retinopathy grading CAD system that provides an estimation of the uncertainty of the decision by imposing a Gaussian distribution centered on the predicted class.

These parametric approaches sometimes sacrifice accuracy to ensure the ordinality assumption. This sacrifice might sometimes prove too much, especially given the fact that modern deep-learning datasets are massive and have a significant number of mislabeled examples.

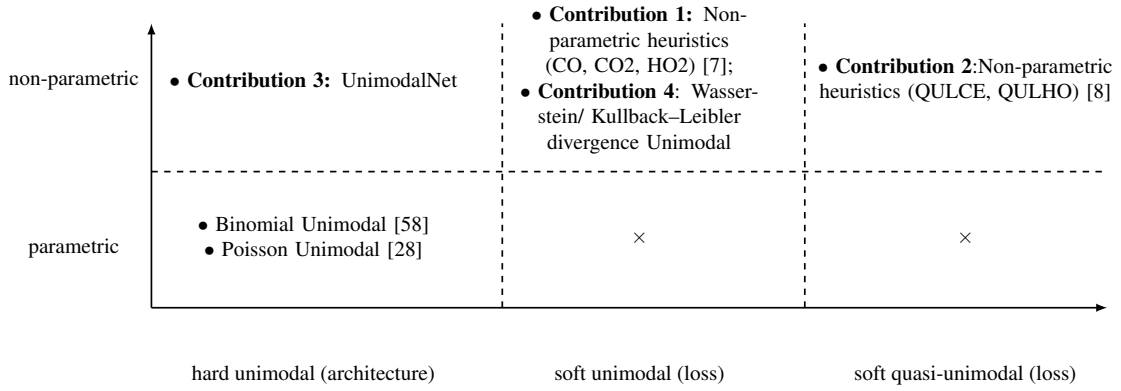


Figure 5.2: Summary of the current unimodal and quasi-unimodal approaches, where the axes represent the soft/hard constraint and parametric/non-parametric priorities. The proposed contributions are also mentioned in the right families.

Figure 5.2 summarizes the current unimodal and quasi-unimodal approaches concerning the soft/hard constraint and parametric/non-parametric priority axes. We also highlight the proposed contributions to be presented next.

5.1.2.2 Understanding the set of unimodal distributions

Before delving into the contributions sections, where the four main approaches proposed in this work are described, we start by contributing with a theoretical analysis of the subset of unimodal distributions in the $(K-1)$ dimensional probability simplex.

Let $\mathbf{p} = [p_1, \ldots, p_K]$ be a distribution in the $K-1$ dimensional probability simplex. We say $\mathbf{p}$ is unimodal with mode at $k^\star$ if $p_1 \leq p_2 \leq \cdots \leq p_{k^\star} \geq p_{k^\star+1} \geq \ldots p_K$. Note that the uniform distribution is considered unimodal (for any $k^\star$).

In the $(K-1)$ dimensional probability simplex, the set of unimodal distributions with the fixed mode is a connected set. Figure 5.3 illustrates the result for $K = 3$.

Let $\mathbf{p} = [p_1,\ p_2, \ldots, p_{k^\star}, \ldots, p_K]$ be a unimodal distribution with mode $k^\star$.

Consider

$$\mathbf{q}(\delta) = \Big[p_1 - \delta p_1,\ p_2 + \delta \frac{p_1}{k^\star - 1}, \ldots, p_{k^\star} + \delta \frac{p_1}{k^\star - 1}, \\ p_{k^\star+1}, \ldots, p_K\Big], \tag{5.3}$$

with $0 \leq \delta \leq 1$. Note that a) $\mathbf{q}(\delta)$ is a unimodal distribution with mode $k^\star$ since the order of the probability values is preserved, b) $\mathbf{q}(\delta)$ is a continuous function in δ, c) $q_1(1) = 0$, d) there is a path in the probability simplex connecting $\mathbf{p}$ to $\mathbf{q}(1)$ over unimodal distributions only.

Repeating the process sequentially for $p_2, \ldots p_{k^\star-1}, p_K, p_{K-1}, \ldots, p_{k^\star+1}$, one can continuously transform any unimodal distribution with mode in $k^\star$ in the distribution with $p_{k^\star} = 1$ and zero in all the other values. Therefore, the set of unimodal distributions with a fixed mode in $k^\star$ is connected.

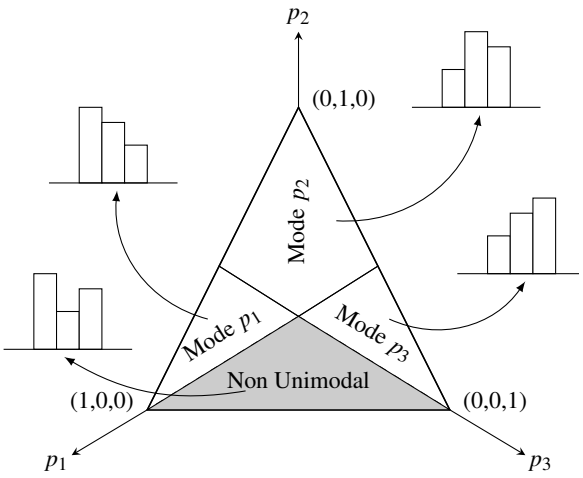


Figure 5.3: Pictorial representation of unimodal and non-unimodal distributions in the 2D simplex.

One can go further and confirm that the set of all unimodal distributions is indeed connected. It suffices to note that the uniform distribution is in the set of unimodal distributions for any $k^\star$. In the context of this property, it is trivially proved that the set of unimodal distributions with mode $k^\star$ is convex, but the set of all unimodal distributions is not convex.

Finally, it is interesting to discuss the "size" of the set of unimodal distribution within the $(K-1)$ dimensional probability simplex. Let $us(K)$ be the fraction of the distributions in $(K-1)$ simplex that corresponds to unimodal distributions and $ns(K)$ the fraction of non-unimodal distributions. When $K=3$, one-third of the distributions are not unimodal, corresponding to the distributions where p_2 is the smallest value of the three: $us(3)=2/3$ and $ns(K)=1/3$.

For general K, if $\min\{p_i\colon i=1,\ldots,K\} \notin \{p_1,p_K\}$ then the distribution is not unimodal. By symmetry, it is clear that $ns(K) \geq (K-2)/K$. It is possible to show that

$$ns(K) = \frac{(K-2)}{K} + \frac{2}{K} ns(K-1). \tag{5.4}$$

For $K=3$, the result is trivial, as already presented. For $K>3$, it was already stated that if $\min\{p_i\colon i=1,\ldots,K\} \notin \{p_1,p_K\}$ then the distribution is not unimodal. If $\min\{p_i\colon i=1,\ldots,K\} = p_1$, consider the distribution $\mathbf{q} = [p_2/(1-p_1)\ p_3/(1-p_1)\ \ldots p_K/(1-p_1)]$ in the $(K-2)$ simplex. Then $\mathbf{p}$ is non-unimodal $\iff$ $\mathbf{q}$ is non-unimodal. And this happens for $ns(K-1)$ of the $\mathbf{q}$ distributions. A similar argument applies when $\min\{p_i\colon i=1,\ldots,K\} = p_K$.

Table 5.1 presents the fraction of distributions that are (non-)unimodal, for several K values. Since the subset of unimodal distributions is considerably smaller than the complete set of probability distributions, even for small K, there is clear potential in exploring this knowledge in the learning of predictive models for ordinal regression.

Table 5.1: Fraction of Unimodal Distributions

K	3	4	5	6
unimodal	0.667	0.333	0.133	0.044
non-unimodal	0.333	0.667	0.867	0.956

5.2 Non-parametric Unimodal Losses (CO, CO2, and HO2)

The different stages of cervical cytology abnormalities are associated with different morphological changes in the cells, including the cytoplasm and nucleus. However, the small visual differences between some stages of cervical cells make the construction of a multi-class autonomous classification system a real challenge. Current approaches to ordinal inference for neural networks are found to not sufficiently take advantage of the ordinal problem or to be too uncompromising.

A non-parametric loss for ordinal data classification is proposed, whose goal is to promote unimodality in the prediction distributions produced by the neural network; e.g., it would be inconsistent to predict that stage 1 and stage 3 are both more likely than stage 2. Yet, this loss is more flexible than other losses from the literature that force a binomial distribution in the output [58]. This loss is also contrasted with the standard cross-entropy loss and networks that predict classes in the form of an ordinal encoding [52].

5.2.1 Proposal

As already explored, CE presents drawbacks when applied to ordinal data. By focusing only on the mode of the distribution and ignoring all the other values in the output probability vector, one is not leveraging the ordinal information intrinsic to the data.

5.2.1.1 Fixing CE with an Ordinal Loss Term

A possible fix for CE is to add a regularization term that penalizes the deviations from the unimodal setting. Defining $\mathbb{1}(x)$ as the indicator function of x and $\text{ReLU}(x) = x\mathbb{1}(x > 0) = \max(0,x)$, a tentative solution for an order-aware loss could be

$$\begin{aligned}\text{CO}(\mathbf{y}_n, \hat{\mathbf{y}}_n) &= \text{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) \\ &+ \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \geq k_n^\star)\,\text{ReLU}(\hat{y}_{n(k+1)} - \hat{y}_{n(k)}) + \lambda \sum_{k=1}^{K-1} \mathbb{1}(k \leq k_n^\star)\,\text{ReLU}(\hat{y}_{n(k)} - \hat{y}_{n(k+1)}),\end{aligned} \quad (5.5)$$

where $\lambda \geq 0$ controls the relative importance of the extra terms favoring unimodal distributions. Predicted probability values are expected to decrease monotonously as we depart left and right from the true class. The added terms penalize any deviation from this expected unimodal distribution, with a penalty proportional to the difference of the consecutive probabilities. The additional terms, although promoting uni-modality, still allow flat distributions. A generalization of the previous idea is to add a margin of $\delta > 0$ to the ReLU, imposing that the difference between

consecutive probabilities is at least δ. This leads us to a second CE loss, CO2, suitable for ordinal classes:

$$\begin{aligned}&\mathrm{CO2}(\mathbf{y}_n,\hat{\mathbf{y}}_n)=\mathrm{CE}(\mathbf{y}_n,\hat{\mathbf{y}}_n)\\&+\lambda\sum_{k=1}^{K-1}\mathbb{1}(k\geq k_n^\star)\,\mathrm{ReLU}(\delta+\hat{y}_{n(k+1)}-\hat{y}_{n(k)})+\lambda\sum_{k=1}^{K-1}\mathbb{1}(k\leq k_n^\star)\,\mathrm{ReLU}(\delta+\hat{y}_{n(k)}-\hat{y}_{n(k+1)}).\end{aligned}\quad(5.6)$$

A value of $\delta=0.05$ has been empirically found to provide a sensible margin. This loss is aligned with the proposal present in [30].

5.2.1.2 Beyond CO2: Ordinal Entropy Loss Function

In CO2, the CE term by itself is only trying to maximize the probability estimated in the true output class (while ignoring the remaining probabilities); the ordinal terms promote unimodality but do not penalize (almost) flat distributions. This also explains why the ordinal terms by themselves (especially the version without margin) are not enough to promote strong learning: the model could converge to solutions where the predicted probability in the true class is only slightly above the neighboring probabilities, which will not, most likely, provide a strong generalization for new observations.

However, the extreme nature of CE, ignoring almost everything in the predicted distribution, $\hat{\mathbf{y}}_n$ is equivalent to assuming that the perfect probability distribution is one on the true class and zero everywhere else. This assumes a strong belief and dependence on the chosen one-hot encoding, which is often a crude approximation to the true probability class distribution. Seldom, for a fixed observation $\mathbf{x}_n$, the class is deterministically known; rather, we expect a class distribution with a few non-zero values. This is particularly true for observations close to the boundaries between classes. A softer assumption is that the distribution should have a low entropy, only.

This leads us to propose the ordinal entropy loss, HO2, for ordinal data as

$$\begin{aligned}&\mathrm{HO2}(\mathbf{y}_n,\hat{\mathbf{y}}_n)=\mathrm{H}(\hat{\mathbf{y}}_n)\\&+\lambda\sum_{k=1}^{K-1}\mathbb{1}(k\geq k_n^\star)\,\mathrm{ReLU}(\delta+\hat{y}_{n(k+1)}-\hat{y}_{n(k)})+\lambda\sum_{k=1}^{K-1}\mathbb{1}(k\leq k_n^\star)\,\mathrm{ReLU}(\delta+\hat{y}_{n(k)}-\hat{y}_{n(k+1)}),\end{aligned}\quad(5.7)$$

where $\mathrm{H}(\mathbf{p})$ denotes the entropy of the distribution $\mathbf{p}$.

5.2.2 Methods

Network Architectures

Two different models were trained and tested in this work for multi-class (4-class and 7-class) classification of Pap smear cell images (Figure 5.4). Both models were trained and tested with nine different convolutional network architectures: AlexNet [125], GoogLeNet [221], MobileNet_V2 [105], ResNet18 [96], ResNeXt50_32X4D [241], ShuffleNet_V2_X1_0 [259], SqueezeNet1_0 [109],

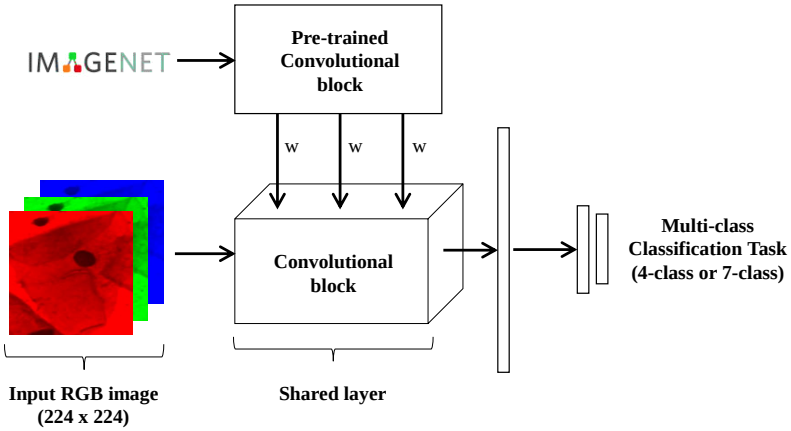


Figure 5.4: Schematic representation of the model used for multi-class classification of Pap smear cells.

VGG-16 [210], and Wide_ResNet50_2 [256]. All the architectures previously mentioned are already briefly explained in Chapter 3. The goal of testing these different architectures is to evaluate how well the proposed loss behaves in a wide range of architectures. These nine different architectures were chosen as they are often used in the literature and came pre-trained with PyTorch on ImageNet[1]. The last block of each architecture was replaced by the following layers: dropout with p=20%, 512-unit dense layer with ReLU, dropout with p=20%, a 256-wide dense layer with ReLU, followed by K output neurons.

5.2.2.1 Training

The weights of the architectures previously mentioned are already initialized by pre-training on ImageNet. Adam was used as the optimizer and started with a learning rate of 10^{-4}. The learning rate is reduced by 10% whenever the loss is stagnant for 10 epochs. The training process is completed after 100 epochs.

The dataset was divided into 10 different folds using stratified cross-validation, in order to maintain the class ratios. Therefore, the results are the average and deviation of these 10 folds. In the case of the proposed loss, the hyperparameter λ is tuned by doing nested k-fold cross-validating using the training set (with k=5) in order to create an unbiased validation set.

5.2.3 Experimental Details

Several neural network architectures are now trained using the aforementioned losses for the dataset at hand. In this work, it was also evaluated the performance differences between parametric and non-parametric losses for ordinal classification (Figure 5.5). All the experiments are implemented in PyTorch and are available online[2].

[1] `https://pytorch.org/docs/stable/torchvision/models.html`
[2] https://github.com/tomealbuquerque/ordinal-losses

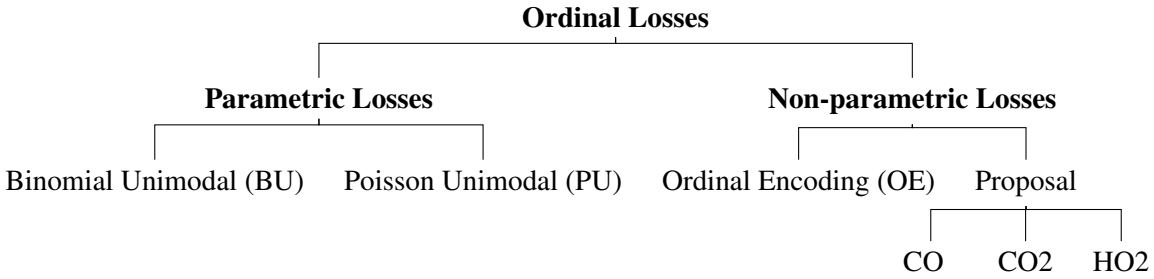


Figure 5.5: Schematic representation of the used and proposed ordinal losses.

5.2.3.1 Dataset

The dataset used in this work is the Herlev Dataset, which is a publicly available dataset[3] collected at the Herlev University Hospital (Denmark) using a digital camera and microscope with an image resolution of 0.201 μm per pixel [113]. The preparation of the specimens followed the traditional Pap smear and Pap staining. To amplify the certainty of diagnosis, two cytotechnicians and a doctor characterized the cervical images in the Herlev dataset into seven classes. The Herlev dataset is composed of a total of 917 images of individual cervical cells. Each image contains ground truth segmentation and classification labels. Table 5.2 shows the nomenclature of the 7 different classes from the dataset, wherein classes 1–3 correspond to types of normal cells and classes 4–7 to different levels of abnormal cells. Illustrations of these classes are then displayed in Table 5.3.

In most cases, the abnormal cells present a nucleus size bigger than healthy cells. However, the difference between the normal columnar nucleus and severe and/or carcinoma nucleus is not easy to differentiate, which makes the classification between these different types of cells a challenge.

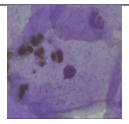
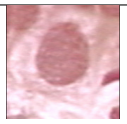
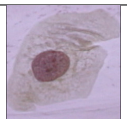
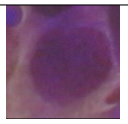
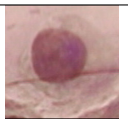
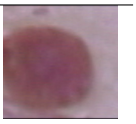
There is some imbalance in the class distribution of the dataset: 8%, 7%, 11%, 19%, 16%, 22%, and 17%, whereas 14% would be expected if the distribution was uniform.

[3] `http://mde-lab.aegean.gr/index.php/downloads`

Table 5.2: The seven different pap smear classes in the Herlev dataset.

	WHO	TBS	Type of cell	Quantity
Normal	1	1	Superficial squamous epithelial	74 cells
	2	1	Intermediate squamous epithelial	70 cells
	3	1	Columnar epithelial	98 cells
Abnormal	4	2	Mild squamous non-keratinizing dysplasia	182 cells
	5	3	Moderate squamous non-keratinizing dysplasia	146 cells
	6	3	Severe squamous non-keratinizing dysplasia	197 cells
	7	4	Squamous cell carcinoma in situ intermediate	150 cells

Table 5.3: Image examples of the seven different pap smear classes in the Herlev dataset.

	Normal			**Abnormal**			
							
WHO	$k=1$	$k=2$	$k=3$	$k=4$	$k=5$	$k=6$	$k=7$
TBS		$k=1$		$k=2$	$k=3$		$k=4$

5.2.3.2 Data Pre-processing

Given that all images from the Herlev dataset are of different sizes, all images were resized to 224×224 pixels; however, before the resize of cytological images, a zero-padding must be done to avoid the loss of essential information regarding cell shape. The last pre-processing step was to apply the same normalization as used by ImageNet [210].

Since the Herlev database has a relatively small number of observations (917), the training dataset was augmented by a series of random transformations: 10% of width and height shift, 10% of zoom, image rotation, horizontal and vertical flips, and color saturation. These transformations are illustrated in Figure 5.6.

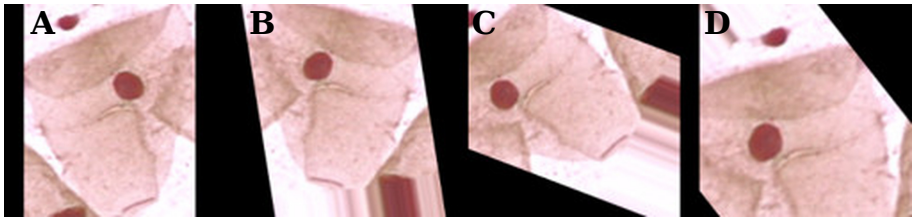


Figure 5.6: Examples of data augmentation on the Herlev database. The original zero-padding image (A) and random transformations (B–D).

5.2.3.3 Evaluation Metrics

The most popular classification metric is accuracy (Acc). For N observations, taking k_i and $\hat{k}_i$ to be the label and prediction of the n-th observation, respectively, then $\text{Acc} = \frac{1}{N}\sum_{n=1}^{N} \mathbb{1}(\hat{k}_n^\star = k_n^\star)$, where $\mathbb{1}$ is the indicator function.

However, this metric treats all class errors as the same, whether the error is between adjacent classes or between classes in the extreme. If we have K classes represented by a set $\mathscr{C} = \{\mathscr{C}^{(1)}, \mathscr{C}^{(2)}, \ldots, \mathscr{C}^{(K)}\}$, then accuracy will treat an error between $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(2)}$ with the same magnitude as an error between $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(K)}$ which is clearly worse. As an illustration, in a medical setting, a misdiagnosis between Stage II and Stage III of a disease, while bad, is not as bad as a misdiagnosis between Healthy and Stage III. For that reason, a popular metric for ordinal classification is the Mean Absolute Error (MAE), $\text{MAE} = \frac{1}{N}\sum_i |k_i^\star - \hat{k}_i^\star|$. This metric is not perfect

since it treats an ordinal variable as a cardinal variable. An error between classes $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(3)}$ will be treated as two times worse than an error between classes $\mathscr{C}^{(1)}$ and $\mathscr{C}^{(2)}$. Naturally, the assumption of cardinality is not always warranted.

To evaluate the models' performance, we also used a metric specific for ordinal classification, Uniform Ordinal Classification Index (UOC) which considers accuracy and ranking in the performance assessment and is also robust against imbalanced classes [209]. The better the performance, the lower the UOC.

By combining a quality assessment (accuracy) with a quantity assessment (MAE) and also with a specific metric for ordinality (UOC) we hope to provide a balanced view of the performance of the methods.

The two other metrics used are the AUC of ROC or AUROC (Area Under the Receiver Operating Characteristic) and Kendall's τ rank correlation coefficient. AUROC measures how well-calibrated are the probabilities produced by the model. This first metric is used in the binary classification context (two classes) and is extended for multi-class by comparing each class against the rest (one vs. rest strategy) and performing an overall average, known as macro averaging. On the other hand, Kendall's Tau is a non-parametric evaluation of relationships between columns of ranked data, so it is a measure of ordinal association between data. The τ correlation coefficient returns a value that ranges from -1 to 1, with 0 being no correlation and 1 perfect correlation.

5.2.4 Results and Discussion

The average performance for the 10-folds of nine different architectures are presented in Tables 5.4–5.9, B.1 and B.2, for both the 7-class and 4-class classification problems, with the seven different learning losses – conventional Cross-Entropy (CE), Binomial Unimodal (BU) [58], Poisson Unimodal (PU) [28], Ordinal Encoding (OE) [52] and our proposed losses (CO, CO2 and HO2), as measured by MAE, accuracy, UOC index and Kendall's coefficient detailed in the previous section. The best models are shown in bold, while italic is used to check for statistical similarity between the other models and the best one. A p-value of 0.1 is used with a two-sided paired t-test due to the small sample size (10 folds).

For the 7-class classification problem, Table 5.4 shows the results for MAE, which confirm the influence of ordinal losses in promoting ordinality when compared to nominal loss (CE). OE loss achieved the best performance across the different architectures, but it is also notable the good performance of our loss: in 67% of cases, the models trained with our proposed loss provide better MAE results. The MAE results present in Table 5.4 for 7-class classification are consistent with the 4-class Table 5.7, with ordinal losses winning over nominal CE.

Table 5.5 and Table 5.8 present the accuracy results for 7-class and 4-class classification problems, respectively. Regarding this metric, the results between nominal and ordinal losses are more balanced. CE loss performance is above ordinal losses in 11% for the 7-class problem and is tied for the 4-class problem. This can be explained by the lower role of ordinality in the CE loss, as also confirmed by the MAE results. This means that when misclassification occurs, ordinal losses tend to classify Pap smear images as being closer to the real class. Results for UOC index

Table 5.4: Results in terms of Mean Absolute Error (MAE) for 7 class problem, averaged for 10 folds (lower is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	0.46 ± 0.08	0.52 ± 0.09	0.50 ± 0.09	*0.44 ± 0.08*	0.90 ± 0.19	**0.41 ± 0.08**	0.45 ± 0.10
GoogLeNet	*0.39 ± 0.05*	0.41 ± 0.07	0.42 ± 0.08	*0.38 ± 0.09*	0.53 ± 0.10	*0.37 ± 0.07*	**0.36 ± 0.06**
MobileNet_v2	0.34 ± 0.05	0.36 ± 0.04	**0.31 ± 0.04**	*0.33 ± 0.05*	0.52 ± 0.26	0.34 ± 0.06	*0.34 ± 0.05*
ResNet18	*0.34 ± 0.09*	*0.36 ± 0.06*	*0.35 ± 0.06*	*0.35 ± 0.10*	0.49 ± 0.11	**0.34 ± 0.07**	*0.35 ± 0.10*
ResNeXt50_32x4d	0.34 ± 0.07	*0.33 ± 0.05*	*0.33 ± 0.03*	0.34 ± 0.06	0.41 ± 0.08	*0.33 ± 0.06*	**0.31 ± 0.07**
ShuffleNet_v2_x1_0	0.41 ± 0.07	0.49 ± 0.07	*0.41 ± 0.05*	*0.38 ± 0.07*	0.47 ± 0.08	*0.40 ± 0.05*	**0.38 ± 0.06**
SqueezeNet1_0	**0.38 ± 0.07**	0.45 ± 0.05	0.46 ± 0.07	*0.40 ± 0.09*	0.97 ± 0.31	0.41 ± 0.08	0.45 ± 0.09
VGG16	*0.37 ± 0.09*	0.44 ± 0.05	0.44 ± 0.10	*0.37 ± 0.06*	0.67 ± 0.15	*0.36 ± 0.06*	**0.36 ± 0.07**
Wide_ResNet50_2	0.33 ± 0.06	0.37 ± 0.05	0.32 ± 0.06	**0.30 ± 0.04**	0.45 ± 0.13	0.33 ± 0.06	0.35 ± 0.09
Avg	0.37	0.41	0.39	0.36	0.60	0.37	0.37
Winners	1	0	1	1	0	2	4

Table 5.5: Results in terms of Accuracy for 7 class problem, averaged for 10 folds (higher is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	**71.1 ± 5.1**	60.6 ± 3.7	64.8 ± 5.4	*70.1 ± 5.1*	44.2 ± 7.6	*70.8 ± 5.1*	67.9 ± 5.4
GoogLeNet	**72.5 ± 3.7**	66.1 ± 4.3	68.5 ± 4.5	*71.5 ± 5.3*	59.7 ± 8.2	*72.4 ± 4.9*	*72.4 ± 3.7*
MobileNet_v2	**75.0 ± 4.4**	69.0 ± 3.5	*74.2 ± 2.8*	*74.4 ± 3.8*	64.4 ± 16.5	73.1 ± 3.7	*74.1 ± 3.9*
ResNet18	**74.4 ± 6.1**	69.5 ± 3.7	*73.3 ± 4.3*	*73.6 ± 6.4*	64.6 ± 6.5	*73.3 ± 4.5*	*73.3 ± 6.4*
ResNeXt50_32x4d	*74.4 ± 3.7*	72.4 ± 4.3	72.8 ± 2.8	*74.0 ± 4.2*	68.0 ± 5.9	*75.5 ± 3.5*	**75.7 ± 5.3**
ShuffleNet_v2_x1_0	**71.9 ± 5.5**	61.0 ± 4.5	67.7 ± 4.6	70.7 ± 4.9	65.5 ± 4.5	*70.7 ± 3.1*	*71.3 ± 3.7*
SqueezeNet1_0	**73.0 ± 4.3**	63.3 ± 2.4	67.3 ± 3.6	*71.8 ± 5.3*	40.5 ± 13.3	70.8 ± 4.5	67.1 ± 5.0
VGG16	**73.1 ± 4.7**	63.9 ± 4.6	67.6 ± 6.2	*72.6 ± 3.8*	54.4 ± 8.5	*71.8 ± 3.3*	*72.0 ± 3.7*
Wide_ResNet50_2	*75.7 ± 3.2*	69.7 ± 3.1	74.5 ± 4.3	**76.8 ± 1.9**	66.1 ± 7.8	*75.6 ± 4.0*	74.3 ± 5.7
Avg	73.4	66.2	70.1	72.8	58.6	72.6	72.0
Winners	7	0	0	1	0	0	1

Table 5.6: Results in terms of Uniform Ordinal Classification Index (UOC) for 7 class problem, averaged for 10 folds (lower is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	*45.1 ± 6.5*	51.7 ± 5.7	49.8 ± 6.6	*44.0 ± 6.9*	70.3 ± 7.8	**42.8 ± 7.3**	46.4 ± 7.8
GoogLeNet	*38.9 ± 6.0*	44.2 ± 5.7	44.6 ± 7.3	*39.0 ± 7.2*	51.3 ± 9.1	*38.8 ± 6.9*	**38.1 ± 4.7**
MobileNet_v2	36.0 ± 5.7	39.7 ± 4.9	**33.6 ± 4.5**	35.4 ± 5.6	46.7 ± 15.0	36.2 ± 6.4	36.2 ± 6.1
ResNet18	**36.2 ± 9.3**	40.1 ± 5.7	*37.2 ± 6.3*	*37.3 ± 9.1*	46.9 ± 6.8	*37.1 ± 7.6*	*37.8 ± 8.7*
ResNeXt50_32x4d	36.9 ± 6.8	*37.0 ± 5.2*	37.6 ± 4.6	36.8 ± 6.1	42.2 ± 6.7	*35.3 ± 6.7*	**34.0 ± 7.2**
ShuffleNet_v2_x1_0	41.8 ± 7.1	49.6 ± 6.4	43.6 ± 4.9	**40.3 ± 6.3**	46.3 ± 6.0	42.4 ± 4.1	*40.3 ± 4.9*
SqueezeNet1_0	**40.4 ± 6.0**	47.9 ± 3.8	47.5 ± 4.8	*42.4 ± 8.1*	73.6 ± 13.6	42.7 ± 7.4	46.8 ± 7.0
VGG16	**38.5 ± 8.2**	47.2 ± 4.9	45.5 ± 8.6	*39.0 ± 6.4*	60.3 ± 10.0	*40.2 ± 6.1*	*39.6 ± 6.8*
Wide_ResNet50_2	35.7 ± 5.2	40.8 ± 5.4	35.6 ± 6.3	**33.5 ± 4.5**	44.2 ± 9.1	*34.8 ± 6.5*	36.6 ± 8.4
Avg	38.8	44.2	41.7	38.6	53.5	39.0	39.5
Winners	3	0	1	2	0	1	2

bold: best model, *italic:* statistically similar to best (paired t-test).

Table 5.7: Results in terms of Mean Absolute Error (MAE) for 4 class problem, averaged for 10 folds (lower is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	0.31±0.06	0.32±0.04	**0.28±0.04**	*0.29±0.06*	0.47±0.19	*0.29±0.05*	*0.31±0.06*
GoogLeNet	*0.24±0.04*	0.25±0.03	0.25±0.05	*0.24±0.05*	0.38±0.17	**0.22±0.05**	0.25±0.06
MobileNet_v2	*0.22±0.06*	**0.21±0.03**	0.24±0.05	*0.22±0.06*	*0.23±0.04*	0.24±0.05	*0.22±0.05*
ResNet18	0.24±0.03	0.26±0.05	*0.24±0.05*	**0.22±0.04**	0.29±0.11	*0.22±0.04*	0.26±0.06
ResNeXt50_32x4d	*0.21±0.03*	0.22±0.04	0.23±0.03	**0.20±0.04**	0.28±0.07	*0.21±0.03*	0.22±0.05
ShuffleNet_v2_x1_0	*0.28±0.05*	0.33±0.05	**0.27±0.05**	0.31±0.06	0.36±0.09	*0.28±0.06*	*0.28±0.04*
SqueezeNet1_0	*0.28±0.06*	0.30±0.05	0.30±0.06	**0.27±0.07**	0.66±0.17	*0.29±0.04*	0.31±0.05
VGG16	0.27±0.06	0.28±0.06	*0.26±0.05*	**0.24±0.03**	0.53±0.18	0.26±0.05	0.27±0.05
Wide_ResNet50_2	0.23±0.05	0.22±0.04	**0.20±0.06**	*0.22±0.05*	0.43±0.22	*0.21±0.05*	0.22±0.03
Avg	0.25	0.27	0.25	0.24	0.40	0.25	0.26
Winners	0	1	3	4	0	1	0

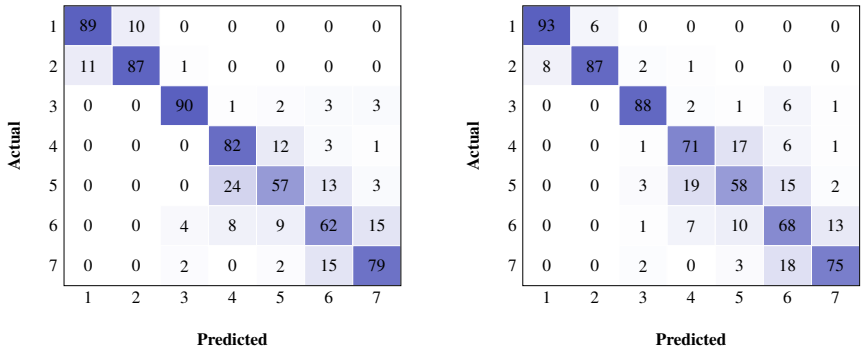
Table 5.8: Results in terms of Accuracy for 4 class problem, averaged for 10 folds (higher is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	*76.1±3.8*	72.8±2.7	*75.7±4.0*	**76.8±3.6**	63.9±12.5	*75.9±3.5*	74.9±3.9
GoogLeNet	*79.9±1.8*	*78.3±2.6*	77.3±3.1	*79.2±4.0*	69.4±12.0	**80.0±3.8**	*78.4±4.0*
MobileNet_v2	**81.8±4.3**	*80.7±2.5*	78.8±3.4	*81.2±4.9*	*79.8±3.7*	79.2±3.2	*80.8±3.7*
ResNet18	*79.8±2.6*	77.2±2.3	*78.5±4.1*	**80.7±4.1**	75.2±8.4	*80.4±3.8*	78.0±4.3
ResNeXt50_32x4d	*82.0±3.1*	80.0±3.5	79.5±3.2	**82.3±4.3**	76.2±5.1	80.8±2.8	79.9±3.9
ShuffleNet_v2_x1_0	**77.1±3.7**	72.1±3.5	*76.1±3.5*	75.0±4.4	70.4±6.6	*76.9±3.9*	*76.2±2.3*
SqueezeNet1_0	*77.2±4.2*	73.5±3.1	74.9±5.1	**77.3±5.3**	49.9±12.2	75.5±3.3	74.3±4.5
VGG16	*77.9±4.8*	74.4±4.7	*77.5±3.8*	**79.4±2.5**	58.1±11.8	77.0±3.9	*77.4±3.7*
Wide_ResNet50_2	80.8±3.2	79.3±3.3	**82.2±4.2**	*81.0±3.9*	64.0±15.3	*81.3±4.2*	80.6±2.6
Avg	79.2	76.5	77.8	79.2	67.4	78.5	77.8
Winners	2	0	1	5	0	1	0

Table 5.9: Results in terms of Uniform Ordinal Classification Index (UOC) for 4 class problem, averaged for 10 folds (lower is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	*38.2±5.1*	39.5±3.4	*37.1±4.3*	**37.0±4.9**	52.7±14.2	*37.4±5.8*	*38.9±5.8*
GoogLeNet	*31.6±3.1*	*31.7±3.6*	34.4±5.6	*32.5±5.7*	44.7±14.6	**30.8±5.5**	*32.9±6.3*
MobileNet_v2	*30.1±6.9*	**29.2±3.7**	32.8±5.2	*30.6±7.5*	*31.0±4.8*	32.5±5.5	*30.5±5.4*
ResNet18	31.4±4.6	33.1±3.7	*32.3±5.5*	**29.4±6.0**	36.7±11.0	*30.3±4.1*	33.2±6.7
ResNeXt50_32x4d	*28.7±4.7*	*29.8±4.9*	32.0±3.9	**27.5±5.3**	35.9±4.8	*28.8±4.6*	31.0±5.2
ShuffleNet_v2_x1_0	**35.8±5.3**	38.6±4.7	*36.7±4.4*	39.0±6.5	43.5±9.0	*36.4±6.9*	*35.9±4.7*
SqueezeNet1_0	*36.6±5.8*	*37.3±4.3*	38.2±6.8	**35.3±6.9**	65.1±9.4	37.6±4.1	39.6±4.6
VGG16	35.3±6.4	36.2±6.4	*34.6±4.7*	**32.3±3.8**	55.1±10.5	34.7±5.5	35.1±6.0
Wide_ResNet50_2	30.2±5.7	*29.9±4.9*	**28.2±5.0**	*30.5±6.2*	47.7±14.4	*29.1±5.6*	30.7±4.3
Avg	33.1	33.9	34.0	32.7	45.8	33.1	34.2
Winners	1	1	1	5	0	1	0

bold: best model, *italic:* statistically similar to best (paired t-test).



(a) [138] (b) Proposal

Figure 5.7: Comparison of state-of-the-art confusion matrix (7 classes) against WideResNet50 trained using the HO2 loss.

(Table 5.6 and 5.9) are also consistent with the MAE metric, with 78% of the models presenting a lowest UOC index when using the ordinal losses. Tables B.1 and Table B.2 in the appendix section B present the results for Kendall's τ coefficient test in 4-class and 7-class classification problems. These results are also aligned with the results of MAE and UOC metrics: the ordinal losses perform better advantage when compared with nominal CE.

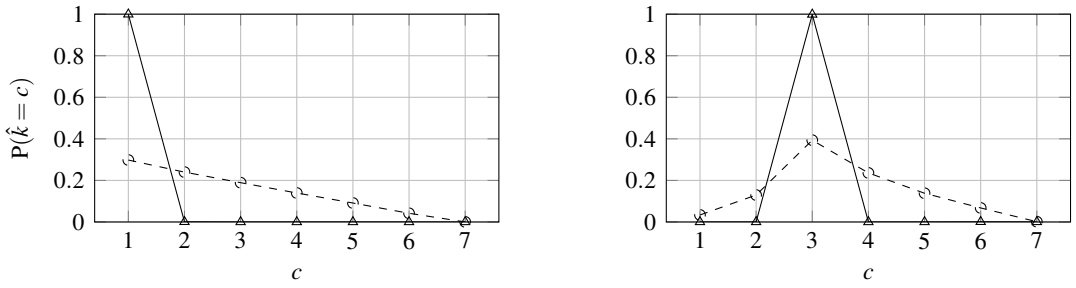
Adding the margin (CO $\rightarrow$ CO2) influences positively most of the metrics for 7 and 4 classes. Using entropy (CO2 or HO2), instead of cross-entropy, promotes better results on the metrics intrinsically connected with ordinality (MAE, UOC and Kendall's τ coefficient).

The average results for all losses across the nine different architectures for MAE, accuracy, UOC, AUROC, Kendall's τ coefficient and Gini index metrics are present Tables B.3 and B.4 in the appendix for 7 and 4 class classification, respectively. Both tables present the results using the classical mode (softmax) to aggregate the probabilities and also using mean (expectation trick) [28]. Concerning the sparsity of the prediction probabilities, as measured by the Gini index, it is notable that, as the loss is made more ordinal-aware, the predicted probabilities tend to be more spread across the classes. This can also be seen in Figure 5.8. Interestingly, the OE distribution is almost identical to the CE distribution and has been omitted from the figure for legibility.

On average, in most metrics, non-parametric losses outperformed parametric losses. This difference can be justified by the greater flexibility in boundary decisions provided by non-parametric losses. OE, CO2, and HO2 provided better results across the different metrics when compared to BU and PU.

Most work from the literature concerns the binary case using the Herlev dataset (normal vs. abnormal); only a couple concern themselves with the 7-class and 4-class ordinal classification problem. Table 5.10 contrasts the best-performing models from two recent works against the proposed method. In our case, the non-parametric loss (CO2) was able to beat the state-of-the-art nominal-class approaches by 11.1% (7 classes) and by 10% (4 classes) in the accuracy metric. Furthermore, the confusion matrices in Figure 5.7 contrast the proposal against [138].

There are classes of cells easier to classify than others, as shown by the confusion matrix in



(a) True $k = 1$ (b) True $k = 3$

Figure 5.8: Probability distribution for WideResNet50 contrasting losses CE (solid line) and HO2 (dashed line).

Figure 5.7 (b). Columnar cells are sometimes inappropriately classified as severe dysplasia cells since severe dysplasia cells have similar characteristics in appearance and morphology to columnar cells (e.g., small cytoplasm, and dark nuclei).

The main challenge occurs in the classification of abnormal cells (i.e., mild dysplasia, moderate dysplasia, severe dysplasia, and carcinoma) where the characteristics of these kinds of cells are very similar. The fact is that the abnormal classes correspond to different levels of evolution of structures, with a progressive change in their characteristics which leads them to present characteristics common to two levels, being a hard task even for cytopathologists to classify them correctly. Thus, the right multi-class classification of abnormal cells is highly desirable and with substantial clinical value.

Finally, the influence of the losses on the output probabilities is illustrated in Figure 5.8 when predicting two classes for the 7-class case. It is clear that the proposed loss tends to promote an unimodal distribution of probabilities relative to the CE loss, which tends to maximize the probability in the output corresponding to the true class and ignore all the other probabilities' distribution, even in contrast to OE.

5.2.5 Conclusion

Comparing ordinal deep learning approaches on cervical cancer data, non-parametric losses achieved better results when compared with parametric losses. This type of loss does not limit the learned

Table 5.10: Accuracy comparison of different models with literature for 7 and 4 classes.

	7 classes	**4 classes**
	Accuracy (%)	**Accuracy (%)**
Jantzen et al.	61.1	–
Lin et al.	64.5	71.3
Proposal	**75.6**	**81.3**

representation to a specific parametric model, which allows, during the training, to explore different and larger spaces of solutions avoiding ad hoc choices.

A new non-parametric loss is proposed for multi-class Pap smear cell-classification based on convolutional neural networks. This new loss demonstrated to be competitive with state-of-the-art results and more flexible than existing deep ordinal classification techniques that impose uni-modality in the probability distribution. The use of the proposed loss in training popular architectures from the literature outperforms the state-of-the-art nominal-class approaches by over 10%.

Furthermore, the proposed loss is a convenient way of introducing ordinality to the optimization problem without the major changes in architecture or data format required by other techniques from the literature. On the other hand, the proposed loss requires two new hyperparameters. However, the suggested values have been found to be robust. While motivated by this dataset, the proposed loss could potentially be used by other applications of ordinal classification.

In any case, there is a lot to improve in the multi-class classification of cervical cells to achieve better accuracy since results are still short of 75.6% accuracy. The Herlev data set is mainly composed of expert-selected "typical" cells, however, in real-life circumstances, data is more complex because a cytology image contains lots of cells and not only a single cropped cell, so further work is needed before moving the results of this work to practice. Another important detail is the effect of overlapping nuclei and cell clumps, which has not been taken into account in this work. The presence of artifacts on the images also interferes with classification accuracy.

5.3 Non-parametric Quasi-Unimodal Losses (QULCE and QULHO)

A large body of work in the literature proposes losses that promote an unimodal distribution in class labels $p(\mathbf{y}|\mathbf{x})$. This is often achieved with one of the two main classes of ordinal losses (as illustrated by Figure 5.9): parametric losses and non-parametric losses. Parametric losses try to impose a unimodality prior on the anticipated posterior distributions, using a single penalty on all the labels. This is done, for example, by assuming that the output function follows a Poisson distribution (PU) [28] or a Binomial distribution (BU) [58]. Regarding non-parametric losses, among several in the literature, Ordinal Encoding (OE) [80] or HO2 [7] try not to confine the learned representation to a single parametric model, so that a wider range of possible output functions can be explored, avoiding ad-hoc decisions. For neural networks, a novel non-parametric ordinal loss is presented that induces output probabilities to follow a quasi-unimodal distribution. This is accomplished by applying two different constraints: the first restriction is between the neighboring pair of output probabilities (relative to the true class). Then, a second constraint penalizes all remaining probabilities above that neighborhood, offering a more flexible decision boundary.

5.3.1 Proposal

Although the HO2 loss mitigates some negative aspects of using CE in the CO and CO2 loss function, it may be too restrictive to penalize probabilities associated with classes that are further

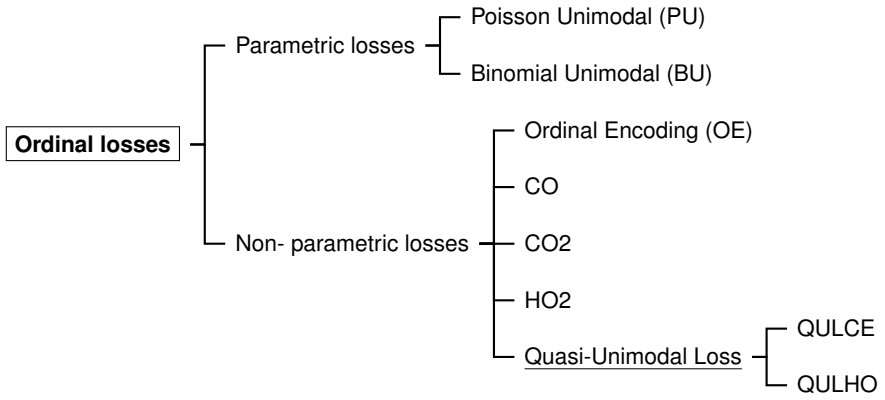


Figure 5.9: Schematic representation of ordinal loss functions from the literature. The proposed loss function is underlined.

from the true class. The motivation for our proposal (quasi-unimodal loss function (QUL)) stems from the definition of the ordinal setting in [47, 215], where it is suggested that in a model consistent with the ordinal setting, a small change in the input data should not lead to a "big jump" in the output decision. Assuming $f(x)$ as decision rule that assigns each value of x to the index $\in \{1,2,\ldots,K\}$ of the predicted class, the decision rule is said to be consistent with an ordinal data classification setting in a point $\mathbf{x}_0$ only if $\exists \varepsilon > 0 \; \forall \mathbf{x} \in \mathscr{V}_\varepsilon(\mathbf{x}_0) \; \max f(\mathbf{x}) - \min f(\mathbf{x}) \leq 1$, with $\mathscr{V}_\varepsilon$ representing a neighborhood with a small perturbation (ε). Equivalently, the decision boundaries in the input space $\mathbf{x}$ should be only between regions of consecutive classes. Note that the concept of consistency with the ordinal setting is independent of the type of model (probabilistic or not) and relies only on the decision regions output by the model.

It is easy to conclude that a continuous model $g(\mathbf{x})$ that outputs a unimodal distribution in the $K-1$ dimensional probability simplex is ordinal-consistent (as long as there are no ties between more than two values in the unimodal distribution). Although this is a sufficient condition, unimodality is not necessary to achieve ordinal consistency (Figure 5.10).

Let $y(k)$ be a discrete distribution over the set $\{1,\ldots,K\}$, with $k^\star = \arg_k \max y(k)$. We say that $y(k)$ is a quasi-unimodal distribution (QUD) if:

$$y(k^\star+1) \geq y(k), \qquad k \in \{k^\star+2,\ldots,K\}, \tag{5.8}$$

$$y(k) \leq y(k^\star-1), \qquad k \in \{1,\ldots,k^\star-2\}. \tag{5.9}$$

It should also be clear that a continuous model $g(\mathbf{x})$ outputting a QUD $y(k)$ in the $K-1$ dimensional probability simplex is ordinal-consistent (as long as there are no ties between more than two values in the quasi-unimodal distribution).

These observations suggest the adoption of a more flexible loss during the training of predictive models, promoting only QUD in the output and not the more restrictive unimodal distribution, as illustrated in Figure 5.11.

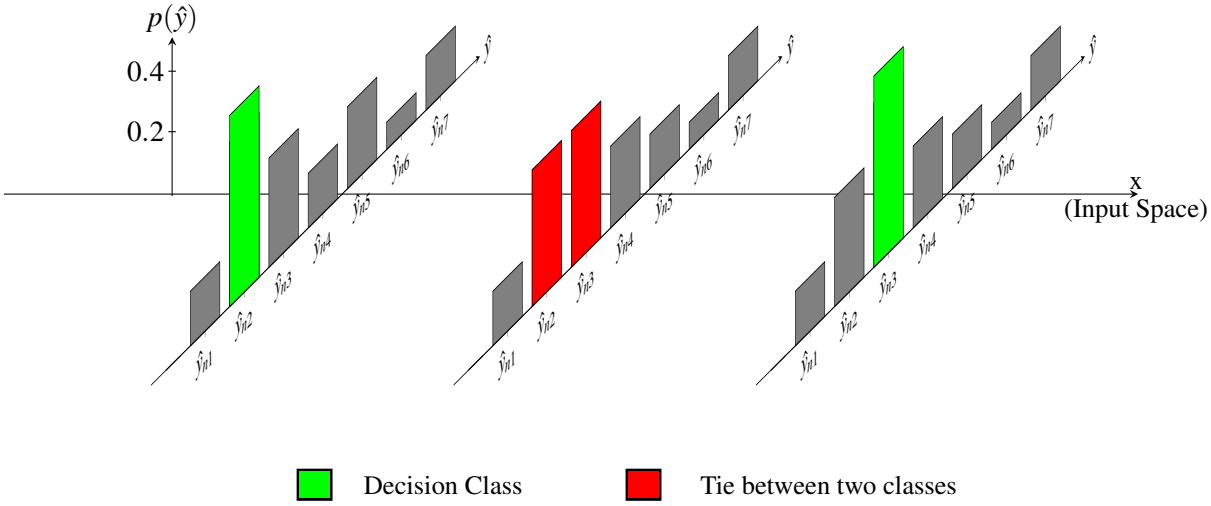


Figure 5.10: Schematic representation of an example where there is ordinal consistency in the decision, without a unimodal distribution in the classes. X axis corresponds to the input space (x) of the model, and Y axis corresponds to output probabilities ($p(\hat{y})$) of the model. Along x, an example of a decision region is represented where the model chooses between two different classes with a boundary region between this decision.

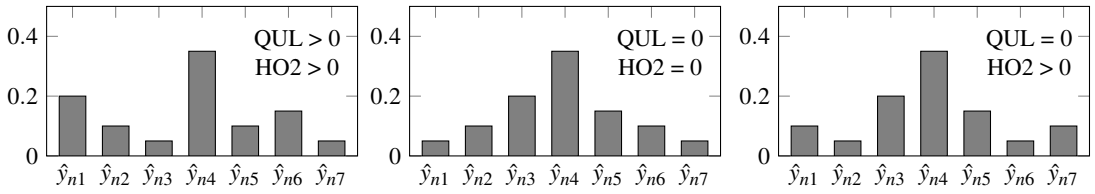
Quasi-unimodal Loss Function (QUL): Compared to HO2, the ordinal terms in the proposal loss promote unimodality only in the top neighborhood output, which allows for a more flexible curve:

$$\begin{aligned}\mathrm{QUL}(\delta, f, \mathbf{y}_n, \hat{\mathbf{y}}_n) = f(\mathbf{y}_n, \hat{\mathbf{y}}_n) \\ + \mathrm{ReLU}(\delta + y_{n(k_n^*-1)} - \hat{y}_{nk_n^*}) \\ + \mathrm{ReLU}(\delta + y_{n(k_n^*+1)} - \hat{y}_{nk_n^*}) \\ + \lambda \sum_{k=1}^{k_n^*-2} \mathrm{ReLU}(\delta + \hat{y}_{nk} - \hat{y}_{n(k_n^*-1)}) \\ + \lambda \sum_{k=k_n^*+2}^{K} \mathrm{ReLU}(\delta + \hat{y}_{nk} - \hat{y}_{n(k_n^*+1)}).\end{aligned} \tag{5.10}$$

The terms which constrain the top-3 neighborhood are in red (Figure 5.12 (a)), while the terms which constrain the remaining classes relative to the neighborhood are in blue (Figure 5.12 (b)). Function f may be defined as CE (like CO2) or as H (like HO2),

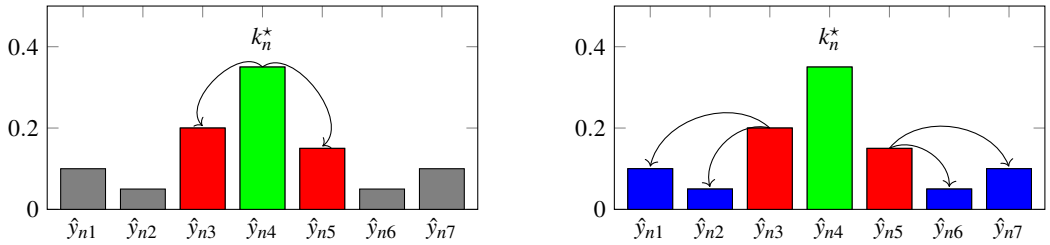
$$\mathrm{QULHO}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n) = \mathrm{QUL}(\mathrm{H}, \delta, \mathbf{y}_n, \hat{\mathbf{y}}_n). \tag{5.11}$$

$$\mathrm{QULCE}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n) = \mathrm{QUL}(\mathrm{CE}, \delta, \mathbf{y}_n, \hat{\mathbf{y}}_n). \tag{5.12}$$



(a) Non-unimodal (b) Unimodal (c) Quasi-unimodal

Figure 5.11: Probabilities produced by three different models for an observation n. HO2 forces the entire distribution curve to have decreasing probabilities. On the other hand, QULHO focuses on the top neighborhood, and is not as concerned about the more distant classes. In this example, $k_n^\star = 4$ is assumed to be the true class.



(a) First QUL term. (b) Second QUL term.

Figure 5.12: Schematic representation of the two QUL terms as in equation (5.10): the first term penalizes only the direct neighborhood, while the second term penalizes the remaining classes relative to the neighborhood.

5.3.2 Methods

5.3.2.1 Network Architectures

Convolutional neural networks (CNN) are used since all datasets consist of images. Filters are learnt in these models, which are quadrilateral patches that are convolved throughout the whole input image — unlike fully-connected networks, just a local neighborhood of inputs is coupled at each layer. Each convolution is usually combined with downsampling procedures such as max-pooling, which gradually lower the size of the original input image while increasing the receptive field.

Three traditional different CNN architectures are here used: MobileNet_V2 [105], ResNet18 [96] and VGG-16 [210]. These three architectures are widely used, well-known and frequently referenced in the literature. They were pre-trained on ImageNet using PyTorch [4].

The following layers were used to replace the last block of each architecture: dropout with p=20%, 512-unit dense layer with ReLU, dropout with p=20%, a 256-wide dense layer with ReLU, followed by K output neurons corresponding to the K output classes of each dataset.

[4] `https://pytorch.org/docs/stable/torchvision/models.html`

5.3.2.2 Losses

Eight different losses are evaluated in this work: Cross-Entropy (CE) for the baseline model and seven distinct ordinal losses: Ordinal Encoding (OE), Poisson Unimodal (PU), Binomial Unimodal (BU), CO2, Ordinal Entropy Loss Function (HO2), Quasi-unimodal entropy loss function (QULHO) and Quasi-unimodal cross-entropy loss function (QULCE).

5.3.2.3 Training

The weights of the aforementioned architectures were initialized during training using ImageNet pre-training. Adam was used as the optimizer, with a learning rate of 10^{-4}. When the loss is constant for 10 epochs using a specific scheduler, the learning rate is lowered by 10%. The training process is completed after 100 epochs for Herlev and AFAD dataset, however the FocusPath dataset required 300 epochs to finish the train. The training process was conducted on an Nvidia GTX 1080ti (11GB) GPU and on an Nvidia tesla v100 (32GB). In the case of the Ordinal Entropy Loss Function and Quasi-unimodal entropy loss function, the hyperparameter λ is tuned by conducting nested k-fold cross-validating using the training set (with k=5) to generate an unbiased validation set ($\lambda \in [0.00001; 0.0001; 0.001; 0.01; 0.1]$). It was used a fix value of $\omega = 0.05$ during the training.

As elaborated in the experimental section below, experiments were also performed in which the loss QULHO was initialized by pre-training on HO2 to further facilitate the optimization process.

5.3.3 Experimental Details

5.3.3.1 Datasets

Three different datasets are used: Herlev dataset, FocusPath dataset and Asian Face Age Dataset (AFAD), which will now be elaborated. All these datasets consist of RGB images.

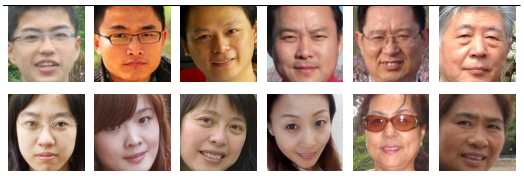
The **Herlev Dataset** and the **FocusPath** were already previously detailed (please see Section 5.2). The **Asian Face Age Dataset (AFAD)** [165] is a publicly available dataset[5] used to estimate age. Images smaller than 64×64 were discarded, resulting in 164,432 photos (63,680 of females and 100,752 of males). The ages range from 15 to 72 for a total of $K = 57$ classes.

5.3.3.2 Data pre-processing

The three datasets are partitioned into train, validation, and test subsets (60-20-20%), with the ratio between the different classes maintained. To feed the network and to use the pre-trained architectures on ImageNet it was necessary to resize the images to 224×224 pixels. As typically done in deep learning models, normalization was also performed to scale the pixel values to 0-1 which speeds up the training process and also helps the network in regularization.

[5] `https://github.com/afad-dataset/tarball`

Table 5.11: Examples of six different classes from the AFAD dataset.



15	25	35	45	55	66
Age (- $\rightarrow$ +)					

Furthermore, to tackle the small amount of data in the datasets, data augmentation is performed. During training, a series of random transformations are applied to each image: 10% of zoom, 10% of width and height shift, horizontal and vertical flips, and image rotation.

5.3.3.3 Evaluation Metrics

Four distinct metrics were used to assess the performance of the various models: accuracy, mean absolute error (MAE), mean squared error (MSE), and the Uniform Ordinal Classification Index (UOC).

By combining these four different evaluation metrics we hope to provide a balanced view of the performance of the methods.

5.3.4 Results and Discussion

The models' performance for the 10-folds (Herlev and FocusPath datasets) and 1-fold (AFAD dataset) using three distinct architectures and three different datasets are presented in Tables 5.12–5.15, with the eight different learning losses – conventional Cross-Entropy (CE), Binomial Unimodal (BU) [58], Poisson Unimodal (PU) [28], Ordinal Encoding (OE) [52], CO2, Ordinal Entropy Loss (HO2) [7], and our proposal losses QULCE and QULHO as measured by MAE, accuracy, MSE and UOC detailed in the previous section. The best models are shown in **bold**.

In general, results are fairly close – except that the proposed losses sometimes have a high variance due to them being harder to optimize. For that reason, we introduce a version that is pre-trained from the HO2 loss, in order to mitigate such optimization problems.

Even so, the proposed loss is quite competitive, especially on the Herlev dataset. Please notice that BU and PU fail to run as the number of classes increases after a certain point.

In the results for the AFAD dataset, due to a large number of classes, HO2 and QULHO losses are having optimization problems which suggest that a more exhaustive fine-tuning of hyper-parameters must be done.

The present work intends to promote the use of ordinal losses against nominal losses when the problem at hand has ordinal data. In the Herlev and FocusPath dataset in most of the cases ordinal losses, across the four different metrics, had better results than nominal cross-entropy loss.

Table 5.12: Results for the Herlev dataset when considering 7 classes, averaged for 10 folds.

	Mean Absolute Error (MAE)				Accuracy (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.34 ± 0.05	0.34 ± 0.09	0.37 ± 0.09	0.37	75.0 ± 4.4	74.4 ± 6.1	73.1 ± 4.7	73.4
BU	0.36 ± 0.04	0.36 ± 0.06	0.44 ± 0.05	0.41	69.0 ± 3.5	69.5 ± 3.7	63.9 ± 4.6	66.2
PU	$\mathbf{0.31 \pm 0.04}$	0.35 ± 0.06	0.44 ± 0.10	0.39	74.2 ± 2.8	73.3 ± 4.3	67.6 ± 6.2	70.1
OE	0.33 ± 0.05	0.35 ± 0.10	0.37 ± 0.06	0.36	74.4 ± 3.8	73.6 ± 6.4	72.6 ± 3.8	72.8
CO2	0.34 ± 0.06	0.34 ± 0.07	0.36 ± 0.06	0.37	73.1 ± 3.7	73.3 ± 4.5	71.8 ± 3.3	72.6
HO2	0.34 ± 0.05	0.35 ± 0.10	0.36 ± 0.07	0.37	74.1 ± 3.9	73.3 ± 6.4	72.0 ± 3.7	72.0
QULCE	0.34 ± 0.06	0.35 ± 0.06	0.36 ± 0.10	0.35	$\mathbf{75.2 \pm 3.6}$	73.8 ± 3.1	$\mathbf{73.5 \pm 4.9}$	74.2
QULHO	0.34 ± 0.07	$\mathbf{0.31 \pm 0.07}$	$\mathbf{0.35 \pm 0.06}$	0.33	73.2 ± 5.6	$\mathbf{74.5 \pm 5.0}$	73.2 ± 3.9	73.6
	Mean Squared Error (MSE)				UOC (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.59 ± 0.13	0.58 ± 0.21	0.64 ± 0.23	0.60	36.0 ± 5.7	36.2 ± 9.3	$\mathbf{38.5 \pm 8.2}$	36.9
BU	0.48 ± 0.07	0.47 ± 0.13	0.65 ± 0.12	0.53	39.7 ± 4.9	40.1 ± 5.7	47.2 ± 4.9	42.3
PU	$\mathbf{0.44 \pm 0.09}$	0.56 ± 0.14	0.74 ± 0.25	0.58	$\mathbf{33.6 \pm 4.5}$	37.2 ± 6.3	45.5 ± 8.6	38.8
OE	0.50 ± 0.10	0.57 ± 0.21	0.61 ± 0.15	0.56	35.4 ± 5.6	37.3 ± 9.1	39.0 ± 6.4	37.2
CO2	0.54 ± 0.15	0.55 ± 0.18	0.57 ± 0.15	0.55	36.2 ± 6.4	37.1 ± 7.6	40.2 ± 6.1	37.9
HO2	0.54 ± 0.12	0.57 ± 0.24	$\mathbf{0.56 \pm 0.18}$	0.55	36.2 ± 6.1	37.8 ± 8.7	39.6 ± 6.8	37.9
QULCE	0.58 ± 0.16	0.60 ± 0.17	0.62 ± 0.29	0.60	36.3 ± 5.1	37.1 ± 5.1	38.6 ± 8.4	37.3
QULHO	0.52 ± 0.16	$\mathbf{0.45 \pm 0.13}$	0.57 ± 0.14	0.51	36.5 ± 7.3	$\mathbf{34.7 \pm 6.9}$	38.8 ± 5.3	36.7

bold: best model.

Table 5.13: Results for the Herlev dataset when considering 4 classes, averaged for 10 folds.

	Mean Absolute Error (MAE)				Accuracy (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.22 ± 0.06	0.24 ± 0.03	0.27 ± 0.06	0.25	$\mathbf{81.8 \pm 4.3}$	79.8 ± 2.6	77.9 ± 4.8	79.2
BU	$\mathbf{0.21 \pm 0.03}$	0.26 ± 0.05	0.28 ± 0.06	0.27	80.7 ± 2.5	77.2 ± 2.3	74.4 ± 4.7	76.5
PU	0.24 ± 0.05	0.24 ± 0.05	0.26 ± 0.05	0.25	78.8 ± 3.4	78.5 ± 4.1	77.5 ± 3.8	77.8
OE	0.22 ± 0.06	$\mathbf{0.22 \pm 0.04}$	$\mathbf{0.24 \pm 0.03}$	0.24	81.2 ± 4.9	$\mathbf{80.7 \pm 4.1}$	$\mathbf{79.4 \pm 2.5}$	79.2
CO2	0.24 ± 0.05	0.22 ± 0.04	0.26 ± 0.05	0.25	79.2 ± 3.2	80.4 ± 3.8	77.0 ± 3.9	78.5
HO2	0.22 ± 0.05	0.26 ± 0.06	0.27 ± 0.05	0.26	80.8 ± 3.7	78.0 ± 4.3	77.4 ± 3.7	77.8
QULCE	0.24 ± 0.03	0.25 ± 0.05	0.28 ± 0.07	0.25	79.6 ± 2.8	79.2 ± 3.8	77.9 ± 4.8	78.9
QULHO	0.24 ± 0.06	0.24 ± 0.05	0.25 ± 0.03	0.25	79.2 ± 4.5	79.9 ± 4.3	78.6 ± 1.7	79.2
	Mean Squared Error (MSE)				UOC (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.33 ± 0.10	0.31 ± 0.05	0.37 ± 0.12	0.34	30.1 ± 6.9	31.4 ± 4.6	35.3 ± 6.4	32.2
BU	$\mathbf{0.26 \pm 0.05}$	0.33 ± 0.11	0.33 ± 0.08	0.31	$\mathbf{29.2 \pm 3.7}$	33.1 ± 3.7	36.2 ± 6.4	32.8
PU	0.30 ± 0.08	0.30 ± 0.08	0.32 ± 0.09	0.31	32.8 ± 5.2	32.3 ± 5.5	34.6 ± 4.7	33.2
OE	0.30 ± 0.11	$\mathbf{0.28 \pm 0.06}$	$\mathbf{0.30 \pm 0.06}$	0.29	30.6 ± 7.5	$\mathbf{29.4 \pm 6.0}$	$\mathbf{32.3 \pm 3.8}$	30.8
CO2	0.30 ± 0.11	0.28 ± 0.07	0.33 ± 0.08	0.31	32.5 ± 5.5	30.3 ± 4.1	34.7 ± 5.5	32.5
HO2	0.29 ± 0.10	0.34 ± 0.10	0.36 ± 0.11	0.33	30.5 ± 5.4	33.2 ± 6.7	35.1 ± 6.0	32.9
QULCE	0.32 ± 0.05	0.33 ± 0.08	0.41 ± 0.13	0.35	31.9 ± 3.8	31.8 ± 5.2	35.4 ± 7.8	33.0
QULHO	0.31 ± 0.10	0.34 ± 0.06	0.34 ± 0.06	0.33	32.5 ± 6.1	31.8 ± 6.0	33.6 ± 3.1	32.6

bold: best model.

Table 5.14: Results for the FocusPath dataset, averaged for 10 folds.

	Mean Absolute Error (MAE)				Accuracy (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.10 ± 0.02	$\mathbf{0.11 \pm 0.02}$	0.16 ± 0.02	0.12	91.0 ± 2.0	$\mathbf{89.3 \pm 1.4}$	85.4 ± 1.6	88.6
PU	–	–	–	–	–	–	–	–
BU	0.20 ± 0.03	0.18 ± 0.03	0.49 ± 0.96	0.29	80.6 ± 2.4	82.1 ± 3.2	75.6 ± 22.8	79.4
OE	$\mathbf{0.09 \pm 0.01}$	0.12 ± 0.02	$\mathbf{0.13 \pm 0.02}$	0.11	$\mathbf{91.4 \pm 1.4}$	88.6 ± 1.9	$\mathbf{87.5 \pm 1.2}$	89.2
CO2	0.11 ± 0.03	0.13 ± 0.02	0.18 ± 0.02	0.14	89.3 ± 2.2	87.3 ± 1.9	83.1 ± 1.5	86.6
HO2	0.13 ± 0.02	0.14 ± 0.04	0.21 ± 0.02	0.16	87.7 ± 2.1	86.8 ± 3.4	81.5 ± 1.6	85.3
QULCE	0.13 ± 0.01	0.14 ± 0.05	0.20 ± 0.02	0.16	87.8 ± 1.7	87.3 ± 3.7	81.5 ± 1.9	85.6
QULHO	0.12 ± 0.02	0.13 ± 0.04	0.19 ± 0.01	0.15	88.7 ± 1.6	87.2 ± 3.9	82.8 ± 1.4	86.3
	Mean Squared Error (MSE)				UOC (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.13 ± 0.09	0.15 ± 0.10	0.22 ± 0.10	0.17	17.5 ± 3.3	$\mathbf{19.4 \pm 3.4}$	29.1 ± 3.2	22.0
BU	0.20 ± 0.03	0.20 ± 0.06	0.18 ± 0.03	0.20	31.6 ± 3.5	27.1 ± 4.8	29.4 ± 2.5	29.4
PU	4.96 ± 9.70	0.15 ± 0.06	24.38 ± 0.06	9.83	36.6 ± 31.8	21.7 ± 4.8	99.8 ± 0.0	52.7
OE	$\mathbf{0.09 \pm 0.02}$	$\mathbf{0.13 \pm 0.03}$	$\mathbf{0.15 \pm 0.05}$	0.12	$\mathbf{15.9 \pm 3.6}$	21.4 ± 3.0	$\mathbf{27.0 \pm 4.1}$	21.4
CO2	0.13 ± 0.05	0.15 ± 0.04	0.22 ± 0.06	0.17	19.9 ± 4.7	25.0 ± 5.0	33.9 ± 6.4	26.2
HO2	0.15 ± 0.03	0.19 ± 0.14	0.28 ± 0.10	0.21	27.8 ± 6.0	26.0 ± 6.8	37.5 ± 6.3	30.4
QULCE	4.38 ± 1.66	1.35 ± 2.52	0.53 ± 0.32	2.09	84.1 ± 24.2	35.0 ± 29.0	44.7 ± 14.4	54.6
QULHO	0.14 ± 0.06	0.17 ± 0.11	0.24 ± 0.05	0.18	23.9 ± 4.9	25.2 ± 7.3	34.4 ± 4.9	27.8

bold: best model.

Table 5.15: Results for the AFAD dataset, averaged for 3 folds.

	Mean Absolute Error (MAE)				Accuracy (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.02 ± 0.01	$\mathbf{0.01 \pm 0.00}$	$\mathbf{0.02 \pm 0.00}$	0.02	99.2 ± 0.4	$\mathbf{99.7 \pm 0.0}$	$\mathbf{99.4 \pm 0.0}$	99.5
PU	–	–	–	–	–	–	–	–
BU	–	–	–	–	–	–	–	–
OE	$\mathbf{0.01 \pm 0.01}$	0.04 ± 0.03	0.03 ± 0.01	0.03	$\mathbf{99.3 \pm 0.5}$	96.8 ± 2.1	97.8 ± 0.3	98.0
CO2	0.42 ± 0.31	0.24 ± 0.24	0.16 ± 0.15	0.27	84.7 ± 7.4	95.6 ± 2.0	95.8 ± 2.1	92.0
HO2	1.84 ± 1.28	2.99 ± 0.01	3.01 ± 0.65	2.61	66.0 ± 16.1	55.3 ± 0.0	53.9 ± 6.6	58.4
QULCE	0.17 ± 0.02	0.03 ± 0.00	0.02 ± 0.00	0.07	97.0 ± 0.3	99.1 ± 0.0	99.3 ± 0.0	98.5
QULHO	1.62 ± 1.34	2.98 ± 0.01	2.94 ± 0.67	2.51	76.3 ± 19.9	55.4 ± 0.0	55.1 ± 6.0	62.3

	Mean Squared Error (MSE)				UOC (%)			
	MobileNet_v2	ResNet18	VGG16	Avg	MobileNet_v2	ResNet18	VGG16	Avg
CE	0.14 ± 0.08	$\mathbf{0.03 \pm 0.00}$	$\mathbf{0.09 \pm 0.01}$	0.09	$\mathbf{3.7 \pm 2.6}$	$\mathbf{0.4 \pm 0.0}$	$\mathbf{1.2 \pm 0.1}$	1.8
PU	–	–	–	–	–	–	–	–
BU	–	–	–	–	–	–	–	–
OE	$\mathbf{0.04 \pm 0.01}$	0.11 ± 0.06	0.11 ± 0.02	0.09	8.4 ± 5.2	41.8 ± 13.4	7.8 ± 2.9	19.3
CO2	3.80 ± 3.35	3.01 ± 3.08	1.81 ± 2.02	2.87	81.2 ± 22.7	93.3 ± 2.3	64.6 ± 38.3	79.7
HO2	19.10 ± 13.40	30.40 ± 0.09	30.49 ± 9.47	26.66	98.2 ± 0.9	99.3 ± 0.0	99.3 ± 0.1	98.9
QULCE	1.42 ± 0.22	0.27 ± 0.02	0.18 ± 0.01	0.62	12.0 ± 3.9	8.9 ± 8.4	10.0 ± 8.3	10.3
QULHO	17.65 ± 13.81	30.34 ± 0.07	29.86 ± 9.72	25.95	97.8 ± 1.2	99.3 ± 0.0	99.2 ± 0.1	98.8

bold: best model.

5.3.5 Summary and Conclusions

In ordinal classification, metrics are most concerned about the relative order of the predictions, rather than the absolute order. That is, an error between two distant classes is deemed worse than an error between two neighbor classes. For that reason, a wide variety of losses can be found in the literature.

In this study, we introduce two innovative non-parametric ordinal loss functions, namely QULCE and QULHO, designed to force output probabilities to follow a quasi-unimodal distribution. Other strategies in the literature involve large modifications in architecture or data format, therefore the suggested loss is a handy way to introduce ordinality in the classification task without requiring substantial changes in design or data format. It improves upon previous work by allowing for a more flexible distribution, since the unimodal constraint is relaxed for other classes outside the top-3 neighborhood. Furthermore, a review of existing ordinal losses is provided.

Several losses from the literature, plus the proposed one, are compared in an empirical assessment which comprises three different datasets using three popular neural network architectures. The parametric losses that were evaluated fail after the number of classes are increased beyond a certain point. The proposed loss seems competitive in terms of scalability and consistency.

5.4 UnimodalNet and Non-parametric Unimodal Losses (WU-KLDiv and WU-Wass)

This work identifies the limitations of current approaches that rely on unimodal distributions [7, 28, 29, 58], and addresses these issues through two novel solutions: (Contribution 3) A new architecture that by construction ensures the output is a unimodal distribution, and (Contribution 4) a new loss term that relies on the notion of projection in a set to promote unimodality. Experiments show the new architecture achieves top-2 performance, while the proposed new loss term is very competitive while maintaining high unimodality.

5.4.1 Proposal

5.4.1.1 Contribution 3: Non-parametric unimodal architecture

Let $\mathbf{z} \in R^K$ be the output of a certain (deep NN) model. Let $\mathbf{zz} \in (R_0^+)^K$, with $zz_i = f(z_i)$, being $f(.)$ a non-negative function (a non-negative function has a range of 0 to infinity). If z_i is already always non-negative ($\mathbf{z} \in (R_0^+)^K$) by construction (for instance, if the last activation unit is a ReLU or its smooth approximation Softplus), then one can simplify and make $\mathbf{zz} = \mathbf{z}$.

Let $\mathbf{zm} \in (R_0^+)^K$ with

$$\begin{aligned} zm_1 &= zz_1 \\ zm_i &= zm_{i-1} + zz_i, \qquad i = 2, \ldots, K \end{aligned} \tag{5.13}$$

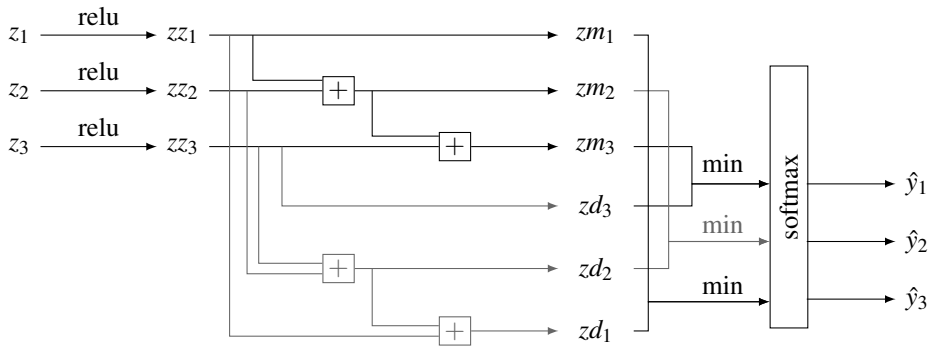


Figure 5.13: Illustration for three classes of the proposed activation function, which forces the model output to be a unimodal distribution.

Then $zm_i \geq zm_{i-1}$ by construction, for $i = 2,\ldots,K$. Similarly, one can construct **zd**, with $zd_i \leq zd_{i-1}$, see Figure 5.13 and Eq 5.14.

$$\begin{aligned} zd_K &= zz_K \\ zd_i &= zd_{i+1} + zz_i, \qquad i = K-1,\ldots,1 \end{aligned} \tag{5.14}$$

Setting $zo_i = \min(zm_i, zd_i)$, then, by construction, **zo** is unimodal. Finally, setting $\hat{\mathbf{y}} = \text{softmax}(\mathbf{zo})$ provides our unimodal distribution in the $(K-1)$ dimensional probability simplex. Let $\ell \in \{1,\ldots,K\}$ be the smallest index for which $zd_\ell \geq zm_\ell$. Then $zo_i = zm_i$ for $i < \ell$, $zo_i = zd_i$, otherwise, and clearly unimodal. If ℓ does not exist, then $zo_i = zm_i \forall i \in \{1,\ldots,K\}$ and also unimodal.

5.4.1.2 Contribution 4: Non-parametric unimodal loss

A regularization term $\text{u}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n)$ has been heuristically introduced in the literature to penalize, non-parametrically, deviations from the family of unimodal distributions. Let $k_n^\star$ be the true label. Defining $\text{ReLU}(x) = \max(0,x)$, a possible fix for an order-aware loss has been previously proposed [7] as

$$\text{CO2}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n) = \text{CE}(\mathbf{y}_n, \hat{\mathbf{y}}_n) + \lambda\, \text{u}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n). \tag{5.15}$$

where $\lambda \geq 0$ controls the relative influence of the extra term u which favours unimodal distributions and is defined as

$$\begin{aligned} \text{u}(\delta, \mathbf{y}_n, \hat{\mathbf{y}}_n) = &\sum_{k=1}^{k_n^\star} \text{ReLU}(\delta + \hat{y}_{nk} - \hat{y}_{n(k+1)}) + \\ &\sum_{k=k_n^\star}^{K-1} \text{ReLU}(\delta + \hat{y}_{n(k+1)} - \hat{y}_{nk}). \end{aligned} \tag{5.16}$$

Furthermore, a margin of $\delta \geq 0$ ensures that the difference between consecutive probabilities is at least δ [7]. As a special case, CO has been defined as the case when the margin is zero ($\delta = 0$),

$$\text{CO}(\mathbf{y}_n, \hat{\mathbf{y}}_n) = \text{CO2}(0, \mathbf{y}_n, \hat{\mathbf{y}}_n). \tag{5.17}$$

Other heuristics are possible. For instance, instead of penalizing the "errors" between consecutive values, one could penalize the errors between any pair of probability values for which an order relation is defined, Eq. (5.18).

$$\begin{aligned}\text{uu}(\delta,\mathbf{y}_n,\hat{\mathbf{y}}_n) = &\sum_{\ell=1}^{k_n^\star-1}\sum_{k=\ell+1}^{k_n^\star}\text{ReLU}(\delta+\hat{y}_{n\ell}-\hat{y}_{nk})+\\ &\sum_{\ell=k_n^\star}^{K-1}\sum_{k=\ell+1}^{K}\text{ReLU}(\delta+\hat{y}_{nk}-\hat{y}_{n\ell}).\end{aligned} \tag{5.18}$$

As an illustrative example, if $\delta = 0$, $\hat{\mathbf{y}} = [2/6\ 3/6\ 0\ 1/6]^t$, and $k^\star = 4$ then $\text{u}(\delta,\mathbf{y}_n,\hat{\mathbf{y}}_n) = 3/6$ and $\text{uu}(\delta,\mathbf{y}_n,\hat{\mathbf{y}}_n) = 8/6$. For $\hat{\mathbf{y}} = [2/6\ 3/6\ 1/6\ 0]^t$, and $k^\star = 4$ then $\text{u}(\delta,\mathbf{y}_n,\hat{\mathbf{y}}_n) = 3/6$ and $\text{uu}(\delta,\mathbf{y}_n,\hat{\mathbf{y}}_n) = 9/6$.

Departing from these heuristics, we now follow a more theoretically-grounded approach, supported by the notion of projection in a set.

Let $\mathscr{S}$ be the set of all unimodal distributions with mode in $k^\star$. Let also $\mathscr{D}_1(.,.)$ and $\mathscr{D}_2(.,.)$ be suitable measures of dissimilarity between two distributions (for instance, Kullback-Leibler (KL) divergence or Wasserstein Distance). The projection in $\mathscr{S}$ of a distribution $\hat{\mathbf{y}}_n$ is defined as

$$\hat{\mathbf{y}}_n^P = \arg\min_{\mathbf{y}\in\mathscr{S}} \mathscr{D}_1(\mathbf{y},\hat{\mathbf{y}}_n). \tag{5.19}$$

$\hat{\mathbf{y}}_n^P$ is the unimodal distribution more similar to $\hat{\mathbf{y}}_n$. The loss term during the learning stage is then defined as

$$\mathscr{D}_2(\hat{\mathbf{y}}_n^P,\hat{\mathbf{y}}_n). \tag{5.20}$$

When $\mathscr{D}_1 = \mathscr{D}_2 = \mathscr{D}$, the loss term simplifies to $\mathscr{D}(\hat{\mathbf{y}}_n^P,\hat{\mathbf{y}}_n) = \min_{\mathbf{y}\in\mathscr{S}}\mathscr{D}(\mathbf{y},\hat{\mathbf{y}}_n)$. Nevertheless, in practice, for analytical and numerical reasons, we found it useful to also consider $\mathscr{D}_1 \neq \mathscr{D}_2$. We analyze next the computation of the projection when $\mathscr{D}_1$ is mapped to the Wasserstein Distance.

Wasserstein Projection: The projection of distribution $\hat{\mathbf{y}}$ in $\mathscr{S}$ (Eq. (5.19)) is easily computed by modifying the computation of the Wasserstein Distance between two distributions.

Giving the transportation cost $c_{ij} \geq 0$ between positions i and j, the Wasserstein Distance between the two distributions $\mathbf{p}$ and $\mathbf{q}$ is computed as

$$\begin{aligned}&\min_{t_{ij}} \sum_{ij} t_{ij}c_{ij}\\ &s.t.\\ &\qquad \textstyle\sum_j t_{ij} = p_j, \forall j \in \{1,\ldots,K\}\\ &\qquad \textstyle\sum_i t_{ij} = q_i, \forall i \in \{1,\ldots,K\}\\ &\qquad t_{ij} \geq 0, \forall i,j\end{aligned} \tag{5.21}$$

The distance of $\mathbf{q}$ to the set of unimodal distributions is trivially computed as

$$\begin{aligned} \min_{t_{ij}} & \sum_{ij} t_{ij} c_{ij} \\ s.t. & \\ & \textstyle\sum_i t_{ij} = q_j, \forall j \in \{1,\dots,K\} \\ & \textstyle\sum_j t_{ij} \leq \sum_{j+1} t_{ij}, \forall i \in \{1,\dots,k^\star - 1\} \\ & \textstyle\sum_j t_{ij} \geq \sum_{j+1} t_{ij}, \forall i \in \{k^\star,\dots,K-1\} \\ & t_{ij} \geq 0, \forall i,j \end{aligned} \tag{5.22}$$

The projection of $\mathbf{q}$ in $\mathscr{S}$ results as $\mathbf{q}_i^P = \sum_j t_{ij}$.

Training Loss: The penalty given by Eq. (5.20) cannot be used standalone to train a (deep neural network) model. The learning would most likely converge to the "constant model", always outputting a uniform distribution. The penalty (5.20) penalizes non-uniform distributions but gives zero penalties to any unimodal distribution.

For that reason, a CE term is added. The weighted sum of the CE with the unimodal penalty simultaneously gives preference to unimodal distributions and, in the set of unimodal distributions, gives preference to the zero-entropy distribution, with $\hat{y}_{k^\star} = 1$.

5.4.2 Methods

5.4.2.1 Network Architectures

The following neural network architectures were used. For the tabular data, an MLP with 128 hidden neurons and a ReLU activation function, was trained for 1,000 epochs. While for images, a ResNet-18 [96] pre-trained in ImageNet, trained for 100 epochs. Adam was used as the optimizer with a learning rate of 10^{-4}.

For each loss with a regularization term (WU-KLDiv, WU-Wass, CO2, and HO2), the regularization coefficient (λ) was validated across the linear space $\lambda \in 10^{-3}, 10^{-2}, \dots, 10^{3}$. For all other hyperparameters, the default values suggested by the authors were used.

5.4.2.2 Models

Eight baseline methods are considered: a model that outputs always the majority class (Dummy), cross-entropy (CE), ordinal encoding (OE) [52], CDW-CE [183], Binomial unimodal (BU) [58], Poisson unimodal (PU) [28], and two non-parametric unimodal regularizers: CO2 and HO2 [7].

Three proposals are considered: the hard UnimodalNet architecture from section 5.4.1.1 (UN), and two proposed soft regularizations from section 5.4.1.2 (WU-KLDiv and WU-Wass). The difference between the two regularizers is how regularization is performed relative to the projection using the Wasserstein distance ($\mathscr{D}_1$): WU-KLDiv uses Kullback–Leibler divergence for $\mathscr{D}_2$, while WU-Wass uses the Wasserstein distance for $\mathscr{D}_2$.

Table 5.16: Datasets

Tabular datasets			
Name	**Description**	**N**	**K**
Abalone5 [72]	Predict the age of abalone from physical measurements (discretized in 5 classes)	4,177	5
Abalone10 [72]	*idem* (discretized in 10 classes)	4,177	10
Balance scale [72]	Balance scale weight & distance database	625	3
Car [72]	Car qualitative evaluation	1,728	4
New thyroid [72]	Normal/hyper/hypo-thyroidism	215	3
Image datasets			
Name	**Description**	**N**	**K**
BACH [19]	Breast histology microscopy	400	4
DHCI [173]	Dating historical color images	1,325	5
FGNET [81]	Age estimation	1,002	70
PAP-SMEAR [70]	Pap Smear from DTU/Herlev	570	5
FocusPath [100]	Estimate focus level for whole-slide images	8,640	12

5.4.3 Experimental Details

The proposed models are contrasted against previously mentioned methods from the literature across ten datasets, both tabular and image datasets.

5.4.3.1 Datasets used for the Experiments

Five tabular and five image datasets were used for evaluation, as detailed in Table 5.16. For evaluation purposes, a stratified 5-fold split was performed: the first fold was used for hyperparameter validation, while the remaining four folds were used to estimate the final results.

For the tabular datasets, the exogenous variables were z-normalized and categorical variables were one-hot encoded. Images were normalized as ImageNet and resized to 268x268; furthermore, the following augmentations were performed: random crop to 256x256, brightness and contrast jitter of 0.1, and horizontal flipping.

5.4.3.2 Evaluation Metrics

Classical metrics for ordinality are used: Accuracy and Mean Absolute Error (MAE). While accuracy is categorical and MAE is a regression metric, they are both often used in tandem for ordinal problems since an ordinal problem is a mix of both problems. Furthermore, a "Unimodality" column shows the fraction of times that the probability distribution produced by the model is unimodal.

5.4.4 Results and Discussion

Table 5.17: Average Summarization of the Results

Loss	**%Accuracy**	**MAE**	**%Unimodality**	**Acc Rank**	**MAE Rank**
Dummy	36.6	2.16	**100.0**	10.5	10.6
CE	66.1	0.92	73.5	3.4	4.6
OE	**67.0**	**0.63**	61.8	3.7	**2.7**
CDW-CE	48.9	1.34	91.3	9.3	8.4
BU	59.3	0.80	**100.0**	7.3	6.5
PU	49.1	1.18	**100.0**	8.9	9.0
UN*	66.6	0.67	**100.0**	3.6	3.5
WU-KLDiv*	65.7	0.78	79.1	4.2	5.4
WU-Wass*	65.1	0.71	83.0	5.2	5.1
CO2	66.4	0.76	87.1	**3.2**	3.7
HO2	61.1	0.82	92.6	6.2	5.9

* proposed model

Table 5.18: Results for the Tabular Datasets

Dataset	**Methods**										
Abalone5	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	72.70 ± 0.08	**78.87 ± 0.68**	78.54 ± 0.69	71.71 ± 0.47	77.40 ± 1.14	56.06 ± 2.18	78.69 ± 0.57	77.88 ± 0.92	76.89 ± 1.40	78.57 ± 1.26	78.21 ± 0.69
MAE	0.307 ± 0.002	**0.225 ± 0.006**	0.228 ± 0.007	0.289 ± 0.006	0.236 ± 0.010	0.465 ± 0.026	0.226 ± 0.004	0.239 ± 0.011	0.244 ± 0.012	0.227 ± 0.013	0.233 ± 0.007
%Unimodal	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	99.91 ± 0.11	99.01 ± 0.23	99.91 ± 0.11	**100.00 ± 0.00**
Abalone10	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	39.45 ± 0.08	**58.64 ± 2.16**	57.74 ± 2.03	37.74 ± 1.93	55.70 ± 2.08	45.11 ± 0.53	58.04 ± 2.32	57.71 ± 2.39	54.62 ± 1.38	58.40 ± 2.31	55.49 ± 0.84
MAE	0.905 ± 0.006	0.534 ± 0.021	**0.520 ± 0.02**	3 0.724 ± 0.022	0.533 ± 0.026	0.688 ± 0.013	0.527 ± 0.029	0.538 ± 0.028	0.593 ± 0.031	0.529 ± 0.022	0.566 ± 0.009
%Unimodal	**100.00 ± 0.00**	95.96 ± 1.46	96.68 ± 1.55	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	99.10 ± 0.57	96.83 ± 2.27	99.55 ± 0.30	**100.00 ± 0.00**
Balance scale	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	20.00 ± 0.00	24.20 ± 1.65	30.20 ± 4.25	24.80 ± 3.64	**34.20 ± 5.71**	29.80 ± 4.54	32.00 ± 5.73	28.40 ± 3.55	22.00 ± 3.55	26.40 ± 1.96	28.60 ± 4.10
MAE	1.200 ± 0.000	1.188 ± 0.071	0.952 ± 0.093	0.996 ± 0.036	**0.910 ± 0.11**	1.010 ± 0.133	1.062 ± 0.079	1.130 ± 0.141	1.224 ± 0.089	1.062 ± 0.073	0.950 ± 0.073
%Unimodal	**100.00 ± 0.00**	82.40 ± 5.95	64.60 ± 2.88	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	98.00 ± 1.67	98.20 ± 2.10	98.60 ± 1.65	**100.00 ± 0.00**
Car	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	70.04 ± 0.12	98.99 ± 0.17	98.77 ± 0.68	91.61 ± 1.22	89.00 ± 1.46	79.52 ± 1.21	**99.35 ± 0.49**	95.01 ± 1.71	95.23 ± 0.98	97.32 ± 1.19	96.31 ± 1.48
MAE	0.415 ± 0.002	0.012 ± 0.003	0.013 ± 0.007	0.084 ± 0.012	0.110 ± 0.015	0.210 ± 0.019	**0.007 ± 0.00**	5 0.059 ± 0.017	0.056 ± 0.013	0.029 ± 0.013	0.041 ± 0.016
%Unimodal	**100.00 ± 0.00**	92.26 ± 1.10	91.17 ± 1.74	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	98.26 ± 0.41	98.55 ± 1.55	99.71 ± 0.24	**100.00 ± 0.00**
New thyroid	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	69.77 ± 0.00	92.44 ± 2.93	93.60 ± 3.97	44.77 ± 7.68	86.05 ± 2.69	78.49 ± 7.45	92.44 ± 2.93	**94.19 ± 4.03**	**94.19 ± 4.03**	**94.19 ± 4.03**	58.72 ± 28.63
MAE	0.442 ± 0.000	0.116 ± 0.063	0.099 ± 0.079	0.552 ± 0.077	0.163 ± 0.057	0.256 ± 0.109	0.116 ± 0.063	**0.093 ± 0.083**	**0.093 ± 0.083**	**0.093 ± 0.083**	0.494 ± 0.245
%Unimodal	**100.00 ± 0.00**	58.72 ± 7.20	29.07 ± 4.84	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	62.21 ± 6.94	66.86 ± 7.91	86.63 ± 2.23	84.30 ± 31.40

* proposed models

Table 5.19: Results for the Image Datasets

Dataset					**Methods**						
BACH	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	25.00 ± 0.00	**88.75 ± 3.06**	85.00 ± 6.12	66.25 ± 27.63	72.81 ± 4.72	62.19 ± 4.83	85.31 ± 4.13	87.50 ± 2.70	88.44 ± 4.38	**88.75 ± 2.04**	79.38 ± 2.39
MAE	1.500 ± 0.000	0.175 ± 0.031	0.184 ± 0.072	0.550 ± 0.634	0.316 ± 0.053	0.444 ± 0.078	0.188 ± 0.051	0.184 ± 0.021	**0.153 ± 0.049**	0.159 ± 0.031	0.272 ± 0.028
%Unimodal	**100.00 ± 0.00**	76.25 ± 2.70	63.44 ± 2.13	99.06 ± 1.88	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	76.56 ± 2.95	77.50 ± 3.68	85.31 ± 2.77	97.81 ± 1.57
DHCI	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	20.00 ± 0.00	57.55 ± 1.91	57.08 ± 2.58	48.96 ± 3.41	46.89 ± 2.01	43.77 ± 1.48	55.66 ± 2.76	**58.30 ± 2.19**	56.23 ± 1.87	56.98 ± 1.45	52.08 ± 3.03
MAE	2.000 ± 0.000	0.670 ± 0.035	**0.596 ± 0.050**	0.686 ± 0.069	0.724 ± 0.056	0.866 ± 0.025	0.634 ± 0.041	0.687 ± 0.066	0.689 ± 0.029	0.722 ± 0.026	0.700 ± 0.057
%Unimodal	**100.00 ± 0.00**	51.23 ± 2.10	56.79 ± 5.43	98.30 ± 1.46	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	53.96 ± 2.92	52.92 ± 2.41	50.00 ± 4.23	91.79 ± 2.58
FGNET	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	4.62 ± 0.25	11.49 ± 2.86	12.86 ± 1.59	4.74 ± 2.09	8.86 ± 1.59	9.12 ± 1.45	12.74 ± 2.88	13.36 ± 1.46	**13.74 ± 2.42**	12.61 ± 1.79	11.11 ± 2.88
MAE	10.518 ± 0.082	5.697 ± 0.276	**3.201 ± 0.282**	6.868 ± 0.800	4.166 ± 0.252	6.657 ± 0.408	3.451 ± 0.110	4.227 ± 0.517	3.522 ± 0.404	4.221 ± 0.215	4.371 ± 0.224
%Unimodal	**100.00 ± 0.00**	1.75 ± 0.29	0.50 ± 0.71	36.21 ± 8.46	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	20.98 ± 6.00	54.43 ± 1.48	61.68 ± 6.20	57.56 ± 2.66
PAP-SMEAR	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	32.02 ± 0.51	80.48 ± 4.38	78.51 ± 2.73	71.93 ± 4.47	61.62 ± 4.61	50.44 ± 3.89	79.17 ± 2.81	80.26 ± 4.61	80.04 ± 4.26	**80.70 ± 4.11**	76.32 ± 5.77
MAE	1.061 ± 0.007	0.219 ± 0.062	0.228 ± 0.039	0.292 ± 0.048	0.404 ± 0.036	0.533 ± 0.053	0.230 ± 0.037	0.217 ± 0.061	0.215 ± 0.039	**0.211 ± 0.053**	0.270 ± 0.065
%Unimodal	**100.00 ± 0.00**	82.89 ± 4.15	18.64 ± 7.21	96.71 ± 3.82	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	85.96 ± 3.04	88.60 ± 2.48	89.47 ± 2.15	94.96 ± 2.62
FocusPath	Dummy	CE	OE	CDW-CE	BU	PU	UN*	WU-KLDiv*	WU-Wass*	CO2	HO2
%Accuracy	12.50 ± 0.00	69.70 ± 2.30	**77.34 ± 1.95**	26.03 ± 22.85	60.94 ± 12.04	36.31 ± 3.99	72.44 ± 5.17	64.31 ± 3.22	69.21 ± 5.54	69.76 ± 6.00	74.45 ± 2.05
MAE	3.245 ± 0.002	0.341 ± 0.020	**0.234 ± 0.018**	2.362 ± 2.031	0.403 ± 0.129	0.682 ± 0.041	0.282 ± 0.053	0.382 ± 0.032	0.325 ± 0.060	0.319 ± 0.062	0.276 ± 0.021
%Unimodal	**100.00 ± 0.00**	93.58 ± 0.68	97.28 ± 0.68	83.04 ± 22.82	**100.00 ± 0.00**	**100.00 ± 0.00**	**100.00 ± 0.00**	95.89 ± 1.10	97.57 ± 0.60	99.88 ± 0.07	99.96 ± 0.03

* proposed models

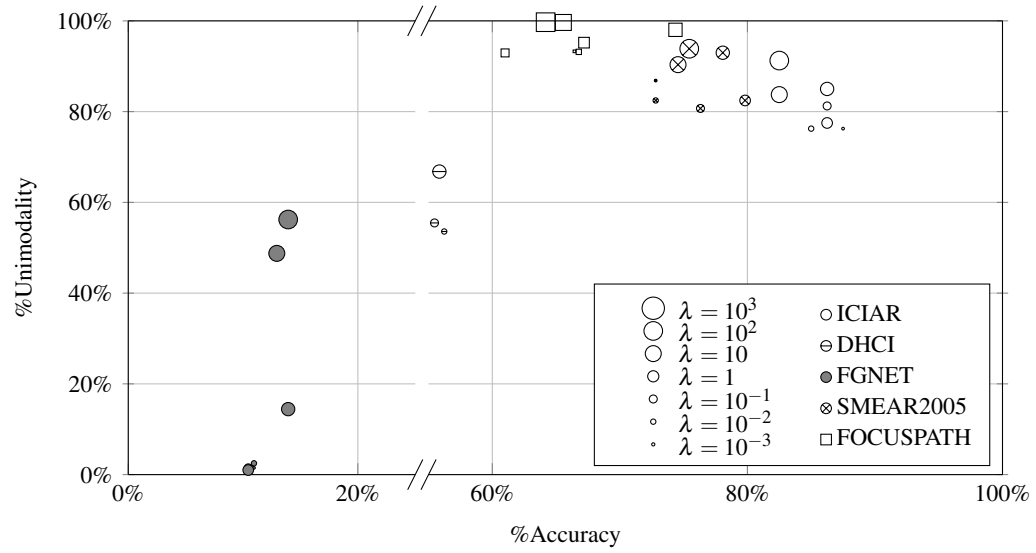


Figure 5.14: Results for different values of λ for the proposed WU-Wass. As λ increases, output probabilities tend to be more unimodal while sacrificing a little accuracy.

Results are shown in Tables 5.18 and 5.19, across several datasets for the various models. The methods are divided into three groups by the gray highlight: the first group is non-unimodal strategies, the second group is the hard unimodal models, and finally, the soft unimodal losses. To make results easier to read and to measure consistency, Table 5.17 summarizes results by averaging the previous results. In addition, two extra columns are shown, representing the "rank" of each loss across the datasets when compared against the other losses (1-11, lower is better).

On average, the losses have obtained an accuracy of 61.5%, with a maximum of 67.0% (OE). Our proposed methods (UN*, WU-KLDiv*, and WU-Wass*) have reached close to the maximum, 65.1–66.6%, with the advantage of the output probabilities being more unimodal. For mean absolute error (MAE), the results are similar. In both of these metrics, the proposed UnimodalNet (UN*) was only surpassed by the Ordinal Encoding (OE), but OE performed poorly in having consistent output probabilities – OE averaged 62% on %Unimodality, while UN* scored 100% on %Unimodality as expected. When considering the average rank position along the datasets, the proposed methods performed fairly well, with UnimodalNet (UN*) showing to be more consistent by scoring second place in both accuracy and MAE. The tables also show that the proposed losses consistently produce probability distributions that follow a unimodal distribution, as would be expected for an ordinal model.

Since WU-KLDiv is the sum of CE and the proposed regularization, weighted by λ, an additional study illustrating the effects of the λ coefficient is shown in Figure 5.14. As λ is increased, and Wasserstein Unimodality is exerted, the trend is towards higher unimodal probability outputs. The drop in accuracy is only slight and even non-existent in some datasets.

A package containing the proposed losses, and many ordinal losses, has been produced at `https://github.com/rpmcruz/unimodal-ordinal-regression`.

5.4.5 Summary and Conclusions

A vast literature on ordinal losses exists. Beyond promoting ordinality, many losses focus on having the model probability outputs follow an unimodal distribution, a property expected by an ordinal model. This unimodality may be either forced into the model (hard-unimodality) or promoted through regularization (soft-unimodality).

This work identifies and addresses two missing parts of the literature. Firstly, some *parametric methods* exist that ensure hard-unimodality, yet no *non-parametric method* to ensure hard-unimodality has been proposed, as far as we are aware. The UnimodalNet is proposed to ensure that the architecture outputs are unimodal.

Furthermore, existing soft-unimodal proposals are only heuristically motivated, without a clear mathematical foundation. For that purpose, a regularization term is proposed that identifies the closest unimodal distribution to the output produced by the model through the Wasserstein Distance. It is shown that this guides the optimizer toward a unimodal solution.

In the end, experiments are performed using ten datasets, with UnimodalNet consistently performing second-best on both accuracy and MAE, while always ensuring unimodality. The proposed regularization terms are also fairly competitive while providing a mathematical foundation and empirically offering high levels of unimodality.

Chapter 6

Multimodal Data Integration in Decision Support Systems

Foreword on Author Contributions
The results of this work have been disseminated in the form of one paper at an international conference.

[5] - Tomé Albuquerque and Jaime S. Cardoso, "Embedded Regularization For Classification Of Colposcopic Images," 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI), Nice, France, 2021, pp. 1920-1923, doi: 10.1109/ISBI48211.2021.9433871.

Within the context of contemporary healthcare, the integration of multimodal data holds the promise of transforming medical decision support systems into powerful tools for clinicians and healthcare providers. Multimodal integration represents a paradigm shift, where diverse streams of patient information, including imaging, clinical data, genetic profiles, and more, converge to provide a comprehensive view of an individual's health.

In the present chapter, the topic of multimodal data integration in decision support systems is explored in the context of colposcopic images for cervical cancer risk assessment. In Appendix C, the work resulting from the exchange period of six months at the Technical University of Munich (TUM) regarding the use of multimodal data integration in the field of brain cancer was presented. This work explores different multimodal integration strategies that could also be used in cervical cancer.

6.1 Embedded regularization for classification of colposcopic images

Significant progress in the realm of artificial intelligence, particularly in neural networks and deep learning helps physicians to diagnose cervical cancer more accurately. In this work, we address a classification problem with the widely used VGG16 architecture. In addition to classification error, our model considers a regularization part during the tuning of the weights, acting as prior

knowledge of the colposcopic image. This embedded regularization approach, using a 2D Gaussian kernel, has enabled the model to learn which sections of the medical images are more crucial for the classification task. The experimental results show an improvement compared with standard transfer learning and multimodal approaches of cervical cancer classification in literature.

6.1.1 Context and Motivation

Cervical cancer remains one of the most common causes of cancer death for women, where most cases occur in the least developed countries [208]. Despite the efforts of recent scientific advances, there is no effective treatment, especially when diagnosed in an advanced stage. Screening tests, such as cytology or colposcopy, have been responsible for a strong decrease in cervical cancer deaths. The main purpose of colposcopy examination is the detection of cervical intraepithelial neoplasia and the evaluation of cervical cancer risk. Regarding the use of multimodal approaches, there are few efforts in this area; some works use cervical images and additional clinical data such as Pap smear results, Human Papilloma Virus (HPV) test and patient age [50, 213, 244, 246]. Clinical data have a clear impact on screening decisions, which was confirmed by the performance boost that literature models demonstrate compared with classical approaches. The use of deep learning approaches for classification raises new challenges. There are many parameters in deep structures, but do we need all of them for a particular task? The existence of these large numbers of parameters makes the structures overfit on the training dataset. One of the fundamental methods to attack the overfitting of a structure is regularization. The term regularization has various definitions, and regularization methods can affect data, network architectures, error terms, etc. Generally, regularization is any supplementary technique that aims at making the model generalize better.

During the last decades, several approaches to overcome the overfitting problem have been proposed. This work is focused on developing Convolutional Neural Networks (CNN) models to classify the risk of cervical cancer on colposcopic images. To promote better generalization, during the training, a 2D Gaussian kernel is used as a new regularization technique. This new regularization term encodes prior knowledge about the region of interest (ROI) of colposcopic images during the training of CNN models.

6.1.1.1 Brief Summary

In our work, a new model to classify cervical cancer risk (binary decision – low risk/high risk) on colposcopic images is proposed. This new model is developed on the NCI dataset and uses a VGG16 architecture, with a loss extended with a 2D Gaussian Kernel to regularize and force the CNN to focus more in the center of the image, the ROI of colposcopic images. Our proposal is compared with state-of-the-art approaches. In the next section 6.1.2, the methods are presented. The experimental details are described in section 6.1.3 and results and discussion are presented in section 6.1.4. The study finishes with some conclusions and several future works in section 6.1.5.

6.1.2 Methods

6.1.2.1 Data set

The American National Cancer Institute (NCI) from the National Institutes of Health (NIH) collected a dataset in several regions of North and South America [97]. This dataset is composed of digital images (cervigrams) from 12,686 women. A subset of this dataset containing $76,669$ cervigrams is available for scientific works. The dataset also includes information about the patient's age, HPV test, and histology results. Some data are missing, including the results from histologic examination, which should be the ground truth for cancer risk classification problems. This resulted in a total of $74,820$ labeled cervigrams. The annotations for histology results correspond to the neoplasia progression level, categorized as normal, CIN1, CIN2, CIN3, and cancer. In an examination, each case is classified as high risk or low risk. To convert the categories of the dataset to these two classes, normal and CIN1 cases are considered low risk, and CIN2, CIN3, and cancer cases are considered high risk. More than 75% of the dataset are low-risk cervical cancer cases, unbalancing the data.

6.1.2.2 Data Pre-Processing

Firstly, the NCI dataset was divided into train, validation, and test subsets (60-20-20%), maintaining the ratio among different classes. To feed the network, it is necessary to resize the colposcopic image to 224×224. In order to speed up the training, the last pre-processing that must be done is the normalization, which consists of a rescale of pixel values from the range of 0-255 to 0-1. To overcome the small amount of data in our dataset, data augmentation was used. Therefore, during the training of the models, a series of random transformations were done in each training epoch for every image. The transformations applied included 10% of width and height shift, 10% of zoom, horizontal and vertical flips, and image rotation. All the applied transformations are slight in order to not affect the regularization prior. Figure 6.1 shows three different examples of random transformations that were done online during the training process of the models.

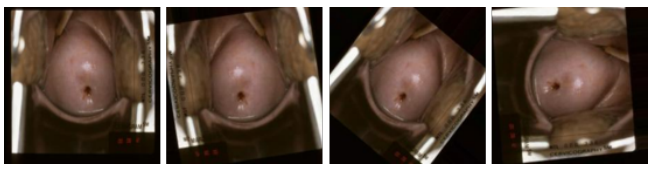


Figure 6.1: Example of data augmentation on NCI/NHI database. (from left to right: original image and three examples of random transformations).

6.1.2.3 Network training and testing

To start the training, initial weights, which can be attributed to random Gaussian distributions or pre-defined values, are required. Transfer learning was used to initiate the training of new models

by transferring the weights of a CNN model pre-trained on the ImageNet dataset. This process speeds up the training of the model.

Multimodal data was also used to train the model. Multimodal learning deals with different types of data (text, images, signals, etc.) to support the model development. As previously mentioned, the dataset also has information about clinical data relevant to the decision. Therefore, this data will be incorporated into the training and concatenated to the ante-penultimate fully connected layer after VGG16 convolutional blocks. In this work, four different clinical features were included as input in the classification task (age, time point, HPV test, and worst histology diagnosis (normal, CIN1, CIN2, CIN3, and cancer)).

During training, the model weights are iteratively adjusted to minimize the classification error (loss function) on the training set using the back-propagation algorithm. The weights are updated using the gradients of the loss function, computed via stochastic gradient descent (SGD). The optimizer used was Adam, an extension of the Stochastic gradient descent optimizer. Instead of performing computations on the whole dataset, which is redundant and inefficient, Adam optimizer only computes on a small subset of a random selection of data examples. The training process is completed after a pre-determined number of epochs, using a TensorFlow callback named "Early Stopping". This function avoids overfitting and unnecessary training time. Only the model with the lowest validation loss value is saved as the final network using "Model Checkpoint", which is also a TensorFlow callback.

6.1.2.4 Losses

In this work, binary cross-entropy was used for binary classification tasks. The binary cross-entropy (bt) is displayed on eq. 6.1 $\mathbf{y}$ is a binary indicator and therefore takes either the value of 0 or 1. $\mathbf{y}$ takes the value 1 if class label is the correct classification for observation. $\hat{\mathbf{y}}$ is the predicted probability of observation.

$$\mathrm{L}_{bt}(y,\hat{y}) = -(y\log(\hat{y}) + (1-y)\log(1-\hat{y}). \tag{6.1}$$

For the model with 2D Gaussian regularization, a new term was added to binary cross-entropy loss; this new loss is represented on eq. 6.2. This new term consists of an element-wise multiplication of all weights in a specific layer ("block 3 pool") of the VGG16 model by a 2D Gaussian matrix with a standard deviation equal to 3. This matrix consists of a 2D Gaussian matrix with the same size as the "block 3 pool" layer of VGG16 (28x28); however, it was necessary to subtract 1 minus the matrix to penalize activations at the edges of the matrix and not in the center. It was necessary to add weight (λ) to this new term on the loss L, in order to control the influence of the regularization. W corresponds to the weights of the "block 3 pool" layer and M to the values of the 2D Gaussian Matrix.

$$\mathrm{L}(y,\hat{y}) = \mathrm{L}_{bt} + \lambda \sum_i \sum_j W_{i,j} M_{i,j}. \tag{6.2}$$

The matrix is represented on Figure 2, where it is possible to see the variation of the values inside the matrix.

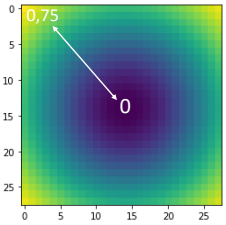


Figure 6.2: Gaussian matrix used for regularization of the model.

6.1.3 Experimental Details

6.1.3.1 Network Architecture

Three different approaches were tested on the CNN model represented on Figure 6.3. The model had two different inputs, the RGB colposcopic images and the clinical data. The model output consists of the classification of risk of cervical cancer (low/high).

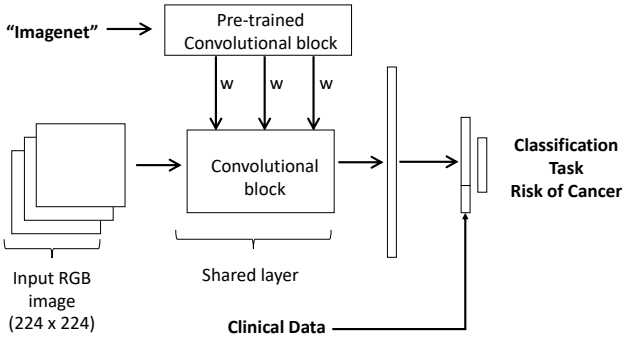


Figure 6.3: CNN model used in this work.

The convolutional block represents the VGG16 CNN, which is initialized with the weights obtained with the 'ImageNet' database. Four different experiments with the CNN were done: a first baseline experiment only using binary cross-entropy loss without multimodal data; a second experiment using multimodal data; a third using multimodal data and oversampling to reduce the imbalance of the dataset by generating offline image copies of the minor class to guarantee an equal number of images from both classes to feed the network and a fourth approach which includes the multimodal data and the use of the 2D Gaussian matrix kernel for regularization. Finding the best hyperparameters for a CNN model is a time-consuming process. For this reason, the hyperparameters were adjusted just for the first approach (baseline) by grid search. The tuned hyperparameters were learning rate (LR $\in$ [0.001; 0.0001]), batch size (BS $\in$ [16; 32]), and dropout rate (Dt $\in$ [0.2; 0.5]). During 150 epochs (using ModelCheckPoint and EarlyStopping

TensorFlow callbacks), the model was tested with all these hyperparameters. The results of this grid search are shown in the next section.

6.1.4 Results and Discussion

The best combination of hyperparameters found for the fine-tuned model was LR of 0.0001, BS of 32, Dt of 0.2, and a λ in loss regularization term of 0.2. These hyperparameters were chosen concerning the test subset on the first approach (VGG16 – baseline). To evaluate the performance of the models, several metrics were taken into consideration. In Table 6.1, the results for the four different models tested are presented in an ablation study. The first approach (VGG16 – Baseline) demonstrated a slightly higher specificity when compared with the others, however, a very low recall and F1 are registered. The use of multimodal data increases almost all metrics. The model with our proposed 2D Gaussian regularization method presented the best accuracy, AUC, recall, NPV, and F1 metric. The VGG16 is a recent and deep architecture, which appears to be able to extract the most compelling features of colposcopic images. This results in a good performance when classifying colposcopic images.

Table 6.1: Ablation Study (averaged for 5-folds) for binary classification of colposcopic images ($Negative\ Predictive\ Value\ (NPV) = \frac{True\ Negative}{False\ Negative + True\ Negative}$).

	Accuracy (%)	**AUC (%)**	**Recall (%)**	**Specificity (%)**	**Precision (%)**	**NPV (%)**	**F1 (%)**
VGG16 baseline	83.00	57.66	6.03	**99.24**	62.62	83.34	10.99
VGG16 Multimodal	86.73	93.22	40.74	96.71	**71.96**	88.73	52.02
VGG16 Multimodal + Oversampling	87.26	93.11	82.72	88.25	61.39	96.42	67.79
VGG16 Multimodal + Gaussian Reg.	**90.79**	**95.54**	**85.60**	91.88	69.02	**96.80**	**76.41**

Table 6.2: Comparison of our proposed model with literature.

	Accuracy (%)	**AUC (%)**	**Recall (%)**	**Specificity (%)**
Song et al. [213]	87.79	-	82.79	**92.82**
Xu et al. [246]	88.91	94.00	**87.83**	90.00
Proposal	**90.79**	**95.54**	85.60	91.88

Going a step further and comparing the results to other works featuring Multimodal learning, it is clear that the evaluation metrics are above the state-of-the-art results, especially on the accuracy and AUC metrics. Xu et al. [246] present a work with a slightly better Recall than our proposal,

in contrast to Song et al. [213] which only shows the specificity above ours. This comparison of metrics between our proposed method and the literature works is presented in Table 6.2.

Regarding our best model, VGG16 with 2D Gaussian regularization, and in order to understand the changes in the decision process induced by regularization, the heatmaps on the "block 3 pool" were plotted. Before regularization, the network solely focused its decision on characters outside the ROI, and after the application of the 2D Gaussian regularization, the network moved its focus to the ROI. Thereby, it is possible to validate the positive effect of regularization. Two different examples of heatmaps and original colposcopic images are represented in Figure 6.4. From a more semantic and medical point of view, this improvement in the focus decision zone supports the enhanced performance of the model. Comparing the baseline model, where several images are classified mainly focusing on the characters in the right bottom corner of the cervigram, our proposal avoids this. Thus, the introduction of prior knowledge by 2D Gaussian regularization helps the model to detect the relevant area of interest for decision-making.

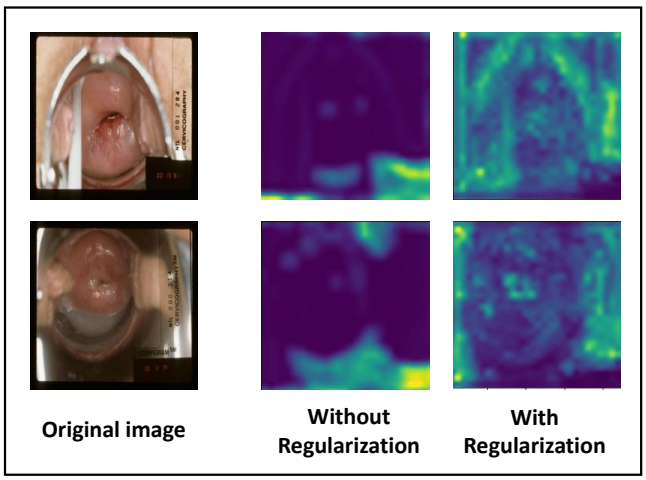


Figure 6.4: Heatmaps of baseline and Gaussian regularization models. A legend for each image is displayed under the latter.

6.1.5 Summary and Conclusions

This work proposes different convolutional neural network architectures to classify colposcopic images. The last model with Gaussian regularization presented the best results in several metrics, highlighting the favorable performance gains introduced by the usage of this newly proposed method of embedded regularization. The present work outperforms some different works present in the literature, nevertheless, there is a lot to improve in the binary classification of cervical cancer risk using multimodal data in accuracy performance. The heatmaps of our proposed model presented improvements in the region of the decision on the colposcopic images. In the baseline and literature models, in several images, the CNNs were focusing the decision on artifacts and disregarding the ROI, which in most cases is in the center of the image.

For future works, several tests using different weights in the λ, which controls the influence of the regularization, should be performed for a better understanding of regularization effects.

Chapter 7

Conclusion

Cervical cancer remains a significant global health concern and continues to exert a profound impact on women's lives worldwide. This malignancy, predominantly caused by persistent infection with certain types of HPV, poses substantial morbidity and mortality risks. Despite notable advancements in medical science and increased awareness, cervical cancer remains a leading cause of cancer-related deaths among women, particularly in regions with limited access to healthcare resources and preventive measures. Early detection through regular cervical screening and HPV vaccination has proven pivotal in improving patient outcomes.

Guided by a primary clinical application, this thesis explores both fundamental and practical aspects. We dedicated significant attention to fundamental machine learning concepts, particularly focusing on multimodal (Chapter 6) and ordinal classification learning (Chapter 5). On the other hand, our applied research places a stronger emphasis on quality control (Chapter 4) within the context of cervical cancer. In most instances, the fundamental topics overlap with the practical questions, demonstrating their interconnectedness and relevance within the applied context.

A substantial portion of our research efforts was dedicated to making fundamental contributions to the fields of Machine Learning and Computer Vision. Our key achievements in these fundamental areas included:

- The proposal of seven novel non-parametric loss functions (CO, CO2, HO2, QULCE, QULHO, WUKLDiv, and WU-Wass loss) for the ordinal classification of cervical cancer in multiclass scenarios using convolutional neural networks. This study illustrates that these recently introduced loss functions can effectively rival state-of-the-art methods while providing enhanced flexibility when compared to existing deep ordinal classification approaches, which frequently impose restrictions on the probability distributions. Moreover, the proposed losses offer a convenient means of introducing ordinality into the optimization process without necessitating significant alterations in architecture or data format, as often seen in other existing techniques in the literature.

- The proposal of a truly non-parametric ordinal network (UnimodalNet), which ensures the architecture outputs are unimodal without the use of any heuristics.

Besides our contributions to ordinal learning, we also explored more application-specific tasks. Our three key achievements in these applied areas included:

- a deep ordinal model to assess the image quality, with specific emphasis on blur in digital pathology slides;
- an end-to-end workflow, starting with the capture of the stacks and ending with the generation of EDoF images using CNNs for efficient processing;
- revolutionizing fragment detection and counting: the work found in this thesis pioneered the development of an automated system for fragment detection and counting, a pivotal tool for routine diagnostics in pathology laboratories.
- a model capable of classifying in low and high-risk colposcopic exams with the integration of prior knowledge about the ROI and also clinical data from the patient.

Moreover, the work resulting from the exchange period of six months at the Technical University of Munich (TUM) regarding the use of multimodal data integration in the field of brain cancer also resulted in key achievements focused on multimodal integration strategies that could also be used in cervical cancer.

7.1 Production

In total, my Ph.D. journey resulted in the publication of fifteen papers, which includes the period I spent on an exchange program at TUM for six months. Among these, three have already been published in reputable international journals, with one more paper currently awaiting a decision. Additionally, the remaining papers found their place in various conferences, with three of them gaining recognition at top-tier conferences such as IEEE ISBI, IEEE ICCVw, and MICCAIw. The journals in which my work was published present impact factors, ranging from 2.6 to 9.29.

To evaluate the caliber of the publication venues, I relied on the widely recognized CORE, ERA, and MSAR rankings [1], except for ECDP, ECP, and RECPAD, which were not included in these lists (Table 7.1). Along my research journey, there were several projects that were initiated but, for various reasons, remained unfinished or were deemed impractical. For the sake of ensuring reproducibility and facilitating access to my research, I have compiled most of the work I conducted during my Ph.D. into a dedicated GitHub repository [2] You can explore this repository and access the research materials.

7.2 Limitations

One common limitation in deep learning experiments is the tendency to become overly optimistic about one's own ideas and proposals. This optimism can arise because researchers often invest

[1] `http://www.conferenceranks.com/` and `http://portal.core.edu.au/conf-ranks/`
[2] `https://github.com/tomealbuquerque`.

Table 7.1: Paper production during the thesis, split by chapter.*Part of papers presented in Annex C

Chapter	–	**1**	**2**	**3**	**4**	**5**	**6**	**∑**
Journal	–	0	0	0	1	2	0	3
Conference	Rank A	0	0	0	1	0	2*	3
	Rank B	0	0	0	1	0	0	1
	Rank C	0	0	0	0	0	0	0
	National	0	0	4	0	0	1*	5
Abstract	Unranked	0	0	0	2	0	1*	3

a significant amount of time refining their approaches. In several works, we tried to mitigate this limitation using several varied datasets, and in some cases, such as in ordinal classification, by performing statistical tests. Yet, another constraint in this research is the scarcity of medical datasets related to cervical cancer. As far as we are aware, there is not a dataset available that includes a minimum of two diagnostic modalities, such as cytology and colposcopy, suitable for the purposes of multimodal learning; only datasets that include one imaging diagnostic method and clinical data.

Another limitation of the presented work is the lack of extensive clinical validation to establish the safety standards essential for routine adoption by clinicians.

7.3 Future Work

We believe that certain chapters within the thesis have the potential to inspire the exploration of new research lines. Chapter 4 shows three different problems in quality control for medical imaging that still have the potential for further improvement. Starting with the blur assessment, which is a problem partially solved, when we are dealing with images with more than 1GB size (WSI), the efficiency of processing becomes paramount in order to sustain smooth laboratory operations. Consequently, the challenge lies in maintaining stringent quality control standards without introducing delays in processing time. Thus, further research could be done to promote even more efficient models without compromising the quality. Another avenue for future research that warrants exploration involves developing an end-to-end Convolutional Neural Network (CNN) model with the capability to handle all the alignment tasks required for the Extended Depth of Field (EDoF) generation process in low-cost microscopy. By doing so, this network would eliminate the need for pre-processing steps, thereby enhancing the method's real-time speed and efficiency. Furthermore, there is potential for investigating the seamless integration of these models into affordable, low-cost devices. Finally, in chapter 4, the challenge of fragment detection and enumeration remains partially unresolved. To address this, there is a need for the exploration of novel detection strategies and the comprehensive validation of these methods across a wide array of datasets acquired from various scanners and involving different specimens of tissue.

Moreover, in Chapters 5 and 6, the task of classifying cervical cancer risk based on diagnostic exams remains a persistent challenge, which leaves room for further exploration in research. Although the ordinal classification of cancer risk has shown potential, it has yet to achieve flawless performance. Exploring ordinal classification within alternative diagnostic modalities for cervical cancer should also be considered.

References

[1] Symptoms of cervical cancer — nhs.uk. https://www.nhs.uk/conditions/cervical-cancer/symptoms/. [Accessed 29-Jun-2023].

[2] Héctor Gabriel Acosta-Mesa, Nicandro Cruz-Ramírez, and Rodolfo Hernández-Jiménez. Aceto-white temporal pattern classification using k-NN to identify precancerous cervical lesion in colposcopic images. *Computers in Biology and Medicine*, 2009.

[3] Amina Adadi and Mohammed Berrada. Peeking inside the black-box: A survey on explainable artificial intelligence (xai). *IEEE Access*, 6:52138–52160, 2018.

[4] Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, and Peter Schüffler. Multimodal context-aware detection of glioma biomarkers using mri and wsi. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023 Workshops*, pages 157–167, Cham, 2023. Springer Nature Switzerland.

[5] Tomé Albuquerque and Jaime S. Cardoso. Embedded regularization for classification of colposcopic images. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1920–1923, 2021.

[6] Tomé Albuquerque, Beatriz Coutinho, João Rodrigo, Miguel Almeida, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Peter Schüffler, and Jaime S. Cardoso. Enhancing glioma detection through multimodal classifiers:
integrating mri and wsi. In *29th Portuguese Conference on Pattern Recognition (RECPAD), Coimbra, Portugal, 27 October*, 10 2023.

[7] Tomé Albuquerque, Ricardo Cruz, and Jaime S. Cardoso. Ordinal losses for classification of cervical cancer risk. *PeerJ Computer Science*, 7:e457, 4 2021.

[8] Tomé Albuquerque, Ricardo Cruz, and Jaime S. Cardoso. Quasi-unimodal distributions for ordinal classification. *Mathematics*, 10(6), 2022.

[9] Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, and Peter Schüffler. Quality checkpoint in pathology specimens handling: an ai system to automate fragment detection and count. In *19th European Congress on Digital Pathology (ECDP), Budapest, Hungary, 14-17 June*, 2023.

[10] Tomé Albuquerque, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, Joao Monteiro, Liliana Ribeiro, Goncalves Sofia, Ana Monteiro, Isabel M. Pinto, and Jaime S. Cardoso. "quality checkpoint in pathology specimens handling: an ai system to automate fragment detection and count. In *18th European Congress on Digital Pathology (ECDP), Berlin, Germany, 15-18 June*, 2022.

[11] Tomé Albuquerque, Ana Moreira, Beatriz Barros, Diana Montezuma, Sara P. Oliveira, Pedro C. Neto, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Ana Monteiro, Isabel M. Pinto, and Jaime S. Cardoso. Quality control in digital pathology: Automatic fragment detection and counting. In *2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 588–593, 2022.

[12] Tomé Albuquerque, Ana Moreira, and Jaime Cardoso. Order is the key: Deep focus assessment in whole slide images. In *27th Portuguese Conference on Pattern Recognition (RECPAD), Évora, Portugal, 5 November*, 11 2021.

[13] Tomé Albuquerque, Ana Moreira, and Jaime S. Cardoso. Deep ordinal focus assessment for whole slide images. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 657–663, 2021.

[14] Tomé Albuquerque, Luís Rosado, Ricardo Cruz, Maria João M. Vasconcelos, Tiago Oliveira, and Jaime S. Cardoso. Rethinking low-cost microscopy workflow: Image enhancement using deep based extended depth of field methods. *Intelligent Systems with Applications*, 17:200170, 2023.

[15] Tomé Albuquerque, Maria Vasconcelos, and Jaime Cardoso. Image quality assessment of cytology images using deep learning. In *26th Portuguese Conference on Pattern Recognition (RECPAD), Évora, Portugal, 30 October*, 10 2020.

[16] Kaitlyn Alleman, Erik Knecht, Jonathan Huang, Lu Zhang, Sandi Lam, and Michael DeCuypere. Multimodal deep learning-based prognostication in glioma patients: A systematic review. *Cancers*, 15(2), 2023.

[17] M. Amin-Naji and A. Aghagolzadeh. Multi-focus image fusion in DCT domain using variance and energy of Laplacian and correlation coefficient for visual sensor networks. *Journal of AI and Data Mining*, 6(2):233–250, 2018.

[18] Teresa Araújo, Guilherme Aresta, Luís Mendonça, Susana Penas, Carolina Maia, Ângela Carneiro, Ana Maria Mendonça, and Aurélio Campilho. DR|GRADUATE: Uncertainty-aware deep learning-based diabetic retinopathy grading in eye fundus images. *Medical Image Analysis*, 63:101715, 2020.

[19] Guilherme Aresta and Teresa Araújo et al. BACH: Grand challenge on breast cancer histology images. *Medical Image Analysis*, 56:122–139, 2019.

[20] Murtaza Ashraf, Willmer Quiñones Robles, Mujin Kim, Young Ko, and Mun Yi. A loss-based patch label denoising method for improving whole-slide image analysis using a convolutional neural network. *Scientific Reports*, 12, 01 2022.

[21] Veysel Aslantas and R. Kurban. Fusion of multi-focus images using differential evolution algorithm. *Expert Syst. Appl.*, 37:8861–8870, 01 2010.

[22] Khosro Bahrami and Alex C. Kot. A fast approach for no-reference image sharpness assessment based on maximum local variation. *IEEE Signal Processing Letters*, 21(6):751–755, 2014.

[23] Xiangzhi Bai, Yu Zhang, Fugen Zhou, and Bindang Xue. Quadtree-based multi-focus image fusion using a weighted focus-measure. *Information Fusion*, 22:105–118, 03 2015.

[24] B Baliga. *Principles and Practice of Colposcopy*. 2011.

[25] Daniel Barath and Jiří Matas. Graph-cut ransac. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[26] Richard Bartels, Jeroen Dudink, Saskia Haitjema, Daniel Oberski, and Annemarie van 't Veen. A perspective on a quality management system for ai/ml-based clinical decision support in hospital care. *Frontiers in Digital Health*, 4, 2022.

[27] Pinky A. Bautista and Yukako Yagi. Detection of tissue folds in whole slide images. In *Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society: Engineering the Future of Biomedicine, EMBC 2009*, 2009.

[28] Christopher Beckham and Christopher Pal. Unimodal probability distributions for deep ordinal classification. In Doina Precup and Yee Whye Teh, editors, *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pages 411–419, International Convention Centre, Sydney, Australia, 06–11 Aug 2017. PMLR.

[29] Soufiane Belharbi, Ismail Ben Ayed, Luke McCaffrey, and Éric Granger. Deep ordinal classification with inequality constraints. *ArXiv*, abs/1911.10720, 2019.

[30] Soufiane Belharbi, Ismail Ben Ayed, Luke McCaffrey, and Eric Granger. Non-parametric uni-modality constraints for deep ordinal classification, 2019.

[31] Soufiane Belharbi, Ismail Ben Ayed, Luke McCaffrey, and Eric Granger. Non-parametric uni-modality constraints for deep ordinal classification, 2020.

[32] Neerja Bhatla, Jonathan S. Berek, Mauricio Cuello Fredes, Lynette A. Denny, Seija Grenman, Kanishka Karunaratne, Sean T. Kehoe, Ikuo Konishi, Alexander B. Olawaiye, Jaime Prat, and Rengaswamy Sankaranarayanan. Revised figo staging for carcinoma of the cervix uteri. *International Journal of Gynecology & Obstetrics*, 145(1):129–135, 2019.

[33] Christopher Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. 10 2007.

[34] Erik Blasch, Yufeng Zheng, and Zheng Liu. *Multispectral Image Fusion and Colorization*. SPIE Press, 03 2018.

[35] Kangkana Bora, Manish Chowdhury, Lipi Mahanta, Malay Kundu, and Anup Das. Automated classification of pap smear image to detect cervical dysplasia. *Computer Methods and Programs in Biomedicine*, 138:31–47, 01 2017.

[36] N.G. Bourbakis. Emulating human visual perception for measuring difference in images using an spn graph approach. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 32(2):191–201, 2002.

[37] Nathaniel Braman, Jacob W. H. Gordon, Emery T. Goossens, C Steadman Willis, Martin C. Stumpe, and Jagadish Venkataraman. Deep orthogonal fusion: Multimodal prognostic biomarker discovery integrating radiology, pathology, genomic, and clinical data. *ArXiv*, abs/2107.00648, 2021.

[38] Pedro Brandão, Paulo T Silva, Marco Parente, and Luís Rosado. µsmartscope: Towards a low-cost microscopic medical device for cervical cancer screening using additive manufacturing and optimization. *Proceedings of the Institution of Mechanical Engineers, Part L: Journal of Materials: Design and Applications*, 2021.

[39] Kristin Branson. A practical review of uniform B-splines, 2004.

[40] Freddie Bray, Jacques Ferlay, Isabelle Soerjomataram, Rebecca L. Siegel, Lindsey A. Torre, and Ahmedin Jemal. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: A Cancer Journal for Clinicians*, 2018.

[41] David N. Breslauer, Robi N. Maamari, Neil A. Switz, Wilbur A. Lam, and Daniel A. Fletcher. Mobile phone based clinical microscopy for global health applications. *PLOS ONE*, 4(7):1–7, 07 2009.

[42] Romain Brixtel, Sébastien Bougleux, Olivier Lézoray, Yann Caillot, Benoît Lemoine, Mathieu Fontaine, Dalal Nebati, and Arnaud Renouf. Whole slide image quality in digital pathology: Review and perspectives. *IEEE Access*, 10:131005–131035, 2022.

[43] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4):532–540, 1983.

[44] Gabriele Campanella, Arjun R Rajanna, Lorraine Corsale, Peter J Schüffler, Yukako Yagi, and Thomas J Fuchs. Towards machine learned quality control: A benchmark for sharpness quantification in digital pathology. *Computerized medical imaging and graphics : the official journal of the Computerized Medical Imaging Society*, 65:142–151, apr 2018.

[45] Jaime S. Cardoso, Ricardo Cruz, and Tomé Albuquerque. Unimodal distributions for ordinal regression, 2023.

[46] Jaime S. Cardoso and Joaquim F. Pinto da Costa. Learning to classify ordinal data: The data replication method. *J. Mach. Learn. Res.*, 8:1393–1429, dec 2007.

[47] Jaime S. Cardoso and Ricardo Sousa. Classification models with global constraints for ordinal data. In *2010 Ninth International Conference on Machine Learning and Applications*, pages 71–77, 2010.

[48] Carmen Phang Romero Casas, Rita de Cássia Ribeiro de Albuquerque, Rafaela Borge Loureiro, Angela Maria Gollner, Marina Gonçalves de Freitas, Graciela Paula do Nascimento Duque, and Juliana Yukari Kodaira Viscondi. Cervical cancer screening in low- and middle-income countries: A systematic review of economic evaluation studies. *Clinics*, 77:100080, 2022.

[49] Thanatip Chankong, Nipon Theera-Umpon, and Sansanee Auephanwiriyakul. Automatic cervical cell segmentation and classification in PAP smears. *Computer methods and programs in biomedicine*, 113, 01 2014.

[50] Tingting Chen, Xinjun Ma, Xingde Ying, Wenzhe Wang, Chunnv Yuan, Weiguo Lu, Danny Z. Chen, and Jian Wu. Multi-modal fusion learning for cervical dysplasia diagnosis. In *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pages 1505–1509, 2019.

[51] Yung-Fu Chen, Po-Chi Huang, Ker-Cheng Lin, Hsuan-Hung Lin, Li-En Wang, Chung-Chuan Cheng, Tsung-Po Chen, Yung-Kuan Chan, and John Y. Chiang. Semi-automatic segmentation and classification of pap smear cells. *IEEE journal of biomedical and health informatics*, 18:94–108, 01 2014.

[52] Jianlin Cheng, Zheng Wang, and Gianluca Pollastri. A neural network approach to ordinal regression. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, pages 1279–1284. IEEE, 2008.

[53] Elizabeth Chlipala, Jesus Elin, Ole Eichhorn, Richard E Long, Bikash Sabata, and Matthew Smith. Archival and Retrieval in Digital Pathology Systems. *Digital Pathology Association*, 2011.

[54] Kyu Sung Choi, Seung Hong Choi, and Bumseok Jeong. Prediction of IDH genotype in gliomas with dynamic susceptibility contrast perfusion MR imaging using an explainable recurrent neural network. *Neuro-Oncology*, 21(9):1197–1209, 06 2019.

[55] Yoon Seong Choi, Sohi Bae, Jong-Hee Chang, Seok Gu Kang, Se Hoon Kim, Jinna Kim, Tyler Hyungtaek Rim, Seung Hong Choi, Rajan Jain, and Seung-Koo Lee. Fully automated hybrid approach to predict the idh mutation status of gliomas via deep learning and radiomics. *Neuro-oncology*, 2020.

[56] Pranathi Chunduru, Joanna J. Phillips, and Annette M. Molinaro. Prognostic risk stratification of gliomas using deep learning in digital pathology images. *Neuro-Oncology Advances*, 4, 2022.

[57] Cisco. Cisco VNI forecast and methodology, 2014-2019. *Cisco Visual Networking Index*, 2015.

[58] Joaquim Costa and Jaime Cardoso. Classification of ordinal data using neural networks. In *European Conference on Machine Learning*, pages 690–697, 2005.

[59] Koby Crammer and Yoram Singer. Pranking with ranking. In T. Dietterich, S. Becker, and Z. Ghahramani, editors, *Advances in Neural Information Processing Systems*, volume 14. MIT Press, 2002.

[60] Danni Cui, Yingying Liu, Gang Liu, and Lei Liu. A multiple-instance learning-based convolutional neural network model to detect the idh1 mutation in the histopathology images of glioma tissues. *Journal of Computational Biology*, 27, 01 2020.

[61] Heather Dawson. Digital pathology – rising to the challenge. *Frontiers in Medicine*, 9:888896, 07 2022.

[62] Ishita de ghosh and Bhabatosh Chanda. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Information Fusion*, 14:136–146, 04 2013.

[63] Milan Decuyper, Stijn Bonte, Karel Deblaere, and Roel Van Holen. Automated mri based pipeline for segmentation and prediction of grade, idh mutation and 1p19q co-deletion in glioma. *Computerized Medical Imaging and Graphics*, 88:101831, 2021.

[64] Elise Deluche, Barbara Bessette, Stéphanie Durand, François Caire, Valérie Rigau, Sandrine Robert, Alain Chaunavel, Lionel Forestier, François Labrousse, Marie-Odile Jauberteau, Karine Durand, and Fabrice Lalloué. Chi3l1, ntrk2, 1p/19q and idh status predicts prognosis in glioma. *Cancers*, 11, 2019.

[65] R M DeMay. *Practical Principles of Cytopathology*. American Society for Clinical Pathology Press, 2007.

[66] DGS. Doenças Oncológicas em Números 2015: Programa Nacional para as Doenças Oncológicas. Technical report, 2016.

[67] Yong Ding. *Visual Quality Assessment for Natural and Medical Image*. Springer, 03 2018.

[68] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks, 2015.

[69] John Doorbar and Heather Griffin. Refining our understanding of cervical neoplasia and its cellular origins, 2019.

[70] George Dounias, B Bjerregaard, Jan Jantzen, A Tsakonas, N Ampazis, G Panagi, and E Panourgias. Automated identification of cancerous smears using various competitive intelligent techniques. *Oncology reports*, 15(4):1001–1006, 2006.

[71] Chaoben Du and Shesheng Gao. Image segmentation-based multi-focus image fusion through multi-scale convolutional neural network. *IEEE Access*, 5:15750–15761, 2017.

[72] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017.

[73] Shervan Fekri Ershad. Pap smear classification using combination of global significant value, texture statistical features and time series features. *Multimedia Tools and Applications*, 07 2019.

[74] Kelwin Fernandes, Jaime S. Cardoso, and Jessica C. Fernandes. Transfer learning with partial observability applied to cervical cancer screening. In *Iberian Conference on Pattern Recognition and Image Analysis*, 2017.

[75] Lawrence Firestone, Kitty Cook, Kevin Culp, Neil Talsania, and Kendall Preston. Comparison of autofocus methods for automated microscopy. *Cytometry*, 1991.

[76] Brigitte Forster-Heinlein, Dimitri Van De Ville, Jesse Berent, Daniel Sage, and Michael Unser. Complex wavelets for extended depth-of-field: A new method for the fusion of multichannel microscopy images. *Microscopy research and technique*, 65:33–42, 09 2004.

[77] John R. Fowler, Eyad V. Maani, Charles J. Dunton, et al. Cervical cancer. StatPearls [Internet], Jan 2023. [Updated 2022 Nov 2].

[78] Filippo Fraggetta, Alessandro Caputo, Rosa Guglielmino, Maria Pellegrino, Giampaolo Runza, and Vincenzo L'Imperio. A survival guide for the rapid transition to a fully digital workflow: The "caltagirone example". *Diagnostics*, 11:1916, 10 2021.

[79] Filippo Fraggetta, Yukako Yagi, Marcial Garcia-Rojo, AndrewJ Evans, JMark Tuthill, Alexi Baidoshvili, DouglasJ Hartman, Junya Fukuoka, and Liron Pantanowitz. The importance of eslide macro images for primary diagnosis with whole slide imaging. *Journal of Pathology Informatics*, 9:46, 12 2018.

[80] Eibe Frank and Mark Hall. A simple approach to ordinal classification. In *European Conference on Machine Learning*, pages 145–156, 2001.

[81] Yanwei Fu, Timothy M. Hospedales, Tao Xiang, Jiechao Xiong, Shaogang Gong, Yizhou Wang, and Yuan Yao. Robust subjective visual property prediction from crowdsourced pairwise labels. In *IEEE TPAMI*, 2016.

[82] Frank Gaillard. Diffuse glioma classification (WHO 5th edition, 2021), October 2021.

[83] Jaime García-Villena, Jose Emilio Torres, Carlos Aguilar, Lin Lin, David Bermejo-Peláez, Elena Dacal, Adriana Mousa, María del Pilar Ortega, Alvaro Martínez, Alexander Vladimirov, Daniel Cuadrado, María Postigo, Jaume Ordi, Quique Bassat, Javier Salamanca, Jose Luis Rodriguez-Peralto, María Linares, Juan Enrique Ortuño, María Jesús Ledesma-Carbayo, Andres Santos, and Miguel Luengo-Oroz. 3d-printed portable robotic mobile microscope for remote diagnosis of global health diseases. *Electronics*, 10(19), 2021.

[84] Amanda Gentry, Colleen Jackson-Cook, Debra Lyon, and Kellie Archer. Penalized ordinal regression methods for predicting stage of cancer in high-dimensional covariate spaces. *Cancer informatics*, 14:201–8, 05 2015.

[85] Raffaella Ghittoni. The role of human papillomaviruses in carcinogenesis. *ecancermedicalscience*, 9, 04 2015.

[86] Ahmed Ghoneim, Ghulam Muhammad, and M. Shamim Hossain. Cervical cancer classification using convolutional neural networks and extreme learning machines. *Future Generation Computer Systems*, 102, 09 2019.

[87] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. Adaptive computation and machine learning. MIT Press, 2016.

[88] Jon Griffin and Darren Treanor. Digital pathology in clinical use: where are we now and what is holding us back? *Histopathology*, 70(1):134–145, 2017.

[89] Frans C.A. Groen, Ian T. Young, and Guido Ligthart. A comparison of different focus functions for use in autofocus algorithms. *Cytometry*, 1985.

[90] Pedro Antonio Gutiérrez, María Pérez-Ortiz, Javier Sánchez-Monedero, Francisco Fernández-Navarro, and César Hervás-Martínez. Ordinal regression methods: Survey and experimental study. *IEEE Transactions on Knowledge and Data Engineering*, 28(1):127–146, 2016.

[91] Maureen L Harmon and Kumarasen Cooper. Chapter 5 - Cervical Neoplasia. In Marisa R Nucci, Esther Oliva, and John R Goldblum, editors, *Gynecologic Pathology*, pages 141–196. Content Repository Only, Edinburgh, 2009.

[92] Rania Hassen, Zhou Wang, and Magdy M. A. Salama. Image sharpness assessment based on local phase coherence. *IEEE Transactions on Image Processing*, 22(7):2798–2810, 2013.

[93] Robert Hawkins. Managing the pre- and post-analytical phases of the total testing process. *Annals of laboratory medicine*, 32:5–16, 01 2012.

[94] Nisreen Haydar, Khatoun Alyousef, Usama Alanan, Rana Issa, Fawaz Baddour, Zuheir Alshehabi, and Moatasem Hussein Al-janabi. Role of magnetic resonance imaging (mri) in grading gliomas comparable with pathology: A cross-sectional study from syria. *Annals of Medicine and Surgery*, 82:104679, 2022.

[95] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[96] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[97] R Herrero, M H Schiffman, C Bratti, A Hildesheim, I Balmaceda, M E Sherman, M Greenberg, F Cárdenas, V Gómez, K Helgesen, J Morales, M Hutchinson, L Mango, M Alfaro, N W Potischman, S Wacholder, C Swanson, and L A Brinton. Design and methods of a population-based natural history study of cervical neoplasia in a rural province of Costa Rica: the Guanacaste Project. *Revista panamericana de salud publica = Pan American journal of public health*, 1(5):362–375, may 1997.

[98] Lynn Hirschowitz, Alex Albus, Laurence Brown, Nagindra Das, Raji Ganesan, C Simon Herrington, Tito Lopes, Glenn McCluggage, David Meechan, Naveena Singh, John HF Smith, Alexandra Thackeray, and Sue Wild. Histopathology reporting in cervical screening – an integrated approach, 2012.

[99] M. S. Hosseini, Jasper A. Z. Brawley-Hayes, Y. Zhang, Lyndon Chan, K. Plataniotis, and S. Damaskinos. Focus quality assessment of high-throughput whole slide imaging in digital pathology. *IEEE Transactions on Medical Imaging*, 39:62–74, 2020.

[100] M. S. Hosseini, Y. Zhang, and K. N. Plataniotis. Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment. *IEEE Transactions on Image Processing*, 28(9):4510–4525, Sep. 2019.

[101] M. S. Hosseini, Y. Zhang, and K. N. Plataniotis. Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment. *IEEE Transactions on Image Processing*, 28(9):4510–4525, Sep. 2019.

[102] Mahdi S. Hosseini and Konstantinos N. Plataniotis. Image sharpness metric based on maxpol convolution kernels. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 296–300, 2018.

[103] Mahdi S. Hosseini, Yueyang Zhang, and Konstantinos N. Plataniotis. Encoding visual sensitivity by maxpol convolution filters for image sharpness assessment. *IEEE Transactions on Image Processing*, 28(9):4510–4525, 2019.

[104] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications, 2017.

[105] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. MobileNets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[106] Wei Huang and Zhongliang Jing. Evaluation of focus measures in multi-focus image fusion. *Pattern Recognition Letters*, 28:493–500, 03 2007.

[107] Johannes Huber, Anna Mueller, Manuela Sailer, and Pedro-Antonio Regidor. Human papillomavirus persistence or clearance after infection in reproductive age. what is the status? review of the literature and new data of a vaginal gel containing silicate dioxide, citric acid, and selenite. *Women's Health*, 17:17455065211020702, 2021. PMID: 34096424.

[108] Elima Hussain, Lipi B. Mahanta, Himakshi Borah, and Chandana Ray Das. Liquid based-cytology pap smear dataset for automated multi-class diagnosis of pre-cancerous and cervical cancer lesions. *Data in Brief*, 30:105589, 2020.

[109] Forrest N Iandola, Song Han, Matthew W Moskewicz, Khalid Ashraf, William J Dally, and Kurt Keutzer. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size. *arXiv preprint arXiv:1602.07360*, 2016.

[110] IARC. IARC Handbooks of Cancer Prevention Cervix Cancer Screening. Technical report, 2005.

[111] IARC International Agency for Research on Cancer. World Health Organization. Cervical Cancer Statistics - Global Cancer Observatory, 2020. Accessed on 30.05.2023.

[112] Andrew Janowczyk, Ren Zuo, Hannah Gilmore, Michael Feldman, and Anant Madabhushi. HistoQC: An Open-Source Quality Control Tool for Digital Pathology Slides. *JCO clinical cancer informatics*, 3:1–7, apr 2019.

[113] Jan Jantzen and George Dounias. Analysis of Pap-smear image data. In *Proceedings of the Nature-Inspired Smart Information Systems 2nd Annual Symposium*, volume 10, 2006.

[114] Xiuyi Jia, Xiang Zheng, Weiwei Li, Changqing Zhang, and Zechao Li. Facial emotion distribution learning by exploiting low-rank label correlations locally. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9833–9842, 2019.

[115] Shuai Jiang, George J Zanazzi, and Saeed Hassanpour. Predicting prognosis and idh mutation status for patients with lower-grade gliomas using whole slide images. *Scientific Reports*, 11, 2021.

[116] Glenn Jocher, Alex Stoken, Ayush Chaurasia, Jirka Borovec, NanoCode012, TaoXie, Yonghye Kwon, Kalen Michael, Liu Changyu, Jiacong Fang, Abhiram V, Laughing, tkianai, yxNONG, Piotr Skalski, Adam Hogan, Jebastin Nadar, imyhxy, Lorenzo Mammana, AlexWang1900, Cristi Fati, Diego Montes, Jan Hajek, Laurentiu Diaconu, Mai Thanh Minh, Marc, albinxavi, fatih, oleg, and wanghaoyang0106. ultralytics/yolov5: v6.0 - yolov5n 'nano' models, roboflow integration, tensorflow export, opencv dnn support. Zenodo, October 2021.

[117] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In Bastian Leibe, Jiri Matas, Nicu Sebe, and Max Welling, editors, *Computer Vision – ECCV 2016*, pages 694–711, Cham, 2016. Springer International Publishing.

[118] Lukasz Kaiser, Aidan N. Gomez, Noam Shazeer, Ashish Vaswani, Niki Parmar, Llion Jones, and Jakob Uszkoreit. One model to learn them all, 2017.

[119] Debashree Kashyap, Abhishek Somani, Jatin Shekhar, Anupama Bhan, Malay Kishore Dutta, Radim Burget, and Kamil Riha. Cervical cancer detection and classification using independent level sets and multi SVMs. In *2016 39th international conference on telecommunications and signal processing (TSP)*, pages 523–528. IEEE, 2016.

[120] Donnie Kim, Nicholas Wang, Viswesh Ravikumar, D. R. Raghuram, Jinju Li, Ankit Patel, Richard E. Wendt, Ganesh Rao, and Arvind Rao. Prediction of 1p/19q codeletion in diffuse glioma patients using pre-operative multiparametric magnetic resonance imaging. *Frontiers in Computational Neuroscience*, 13, 2019.

[121] Stefan Klein, Marius Staring, Keelin Murphy, Max Viergever, and Josien Pluim. Elastix: A toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging*, 29:196–205, 11 2009.

[122] Timo Kohlberger, Yun Liu, Melissa Moran, Po-HsuanCameron Chen, Trissia Brown, JasonD Hipp, CraigH Mermel, and MartinC Stumpe. Whole-slide image focus quality: Automatic assessment and impact on ai cancer detection. *Journal of Pathology Informatics*, 10(1):39, 2019.

[123] Yehuda Koren and Joe Sill. Ordrec: An ordinal model for predicting personalized item rating distributions. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 117–124, 10 2011.

[124] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In F. Pereira, C.J. Burges, L. Bottou, and K.Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, volume 25. Curran Associates, Inc., 2012.

[125] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.

[126] Kurnianingsih Kurnianingsih, Khalid Allehaibi, Lukito Nugroho, Widyawan Widyawan, Lutfan Lazuardi, Anton Satria Prabuwono, and Teddy Mantoro. Segmentation and classification of cervical cells using deep learning. *IEEE Access*, PP:1–1, 08 2019.

[127] Arthur Leclaire and Lionel Moisan. No-reference image quality assessment and blind deblurring with sharpness metrics exploiting fourier phase information. *Journal of Mathematical Imaging and Vision*, 52:145–172, 05 2015.

[128] Yann LeCun, Y. Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521:436–44, 05 2015.

[129] C. Ledig and D. Rueckert. Chapter 14 - semantic parsing of brain mr images. In S. Kevin Zhou, editor, *Medical Image Recognition, Segmentation and Parsing*, The Elsevier and MICCAI Society Book Series, pages 307–335. Academic Press, 2016.

[130] Min Kyoung Lee, Ji Eun Park, Young-Seung Jo, Seo Young Park, Sang Joon Kim, and Ho Sung Kim. Advanced imaging parameters improve the prediction of diffuse lower-grade gliomas subtype, idh mutant with no 1p19q codeletion: added value to the t2/flair mismatch sign. *European Radiology*, 30:844–854, 2019.

[131] John J. Lewis, Robert J. O'Callaghan, Stavri G. Nikolov, David R. Bull, and Nishan Canagarajah. Pixel- and region-based image fusion with complex wavelets. *Inf. Fusion*, 8(2):119–130, apr 2007.

[132] H. Li, B.S. Manjunath, and S.K. Mitra. Multisensor image fusion using the wavelet transform. *Graphical Models and Image Processing*, 57(3):235–245, 1995.

[133] Leida Li, Dong Wu, Jinjian Wu, Haoliang Li, Weisi Lin, and Alex C. Kot. Image sharpness assessment by sparse representation. *IEEE Transactions on Multimedia*, 18(6):1085–1097, 2016.

[134] Shutao Li, Xudong Kang, and Jianwen Hu. Image fusion with guided filtering. *IEEE Transactions on Image Processing*, 22(7):2864–2875, 2013.

[135] Yingping Li, Samy Ammari, Littisha Lawrance, Arnaud Quillent, Tarek Assi, Nathalie Lassau, and Émilie Chouzenoux. Radiomics-based method for predicting the glioma subtype as defined by tumor grade, idh mutation, and 1p/19q codeletion. *Cancers*, 14, 2022.

[136] Yuanyuan Li, Yanjing Sun, Xinhua Huang, Guanqiu Qi, Zheng Mingyao, and Zhi-Qin Zhu. An image fusion method based on sparse representation and sum modified-laplacian in nsct domain. *Entropy*, 20:522, 07 2018.

[137] Benjamin Liechty, Zhuoran Xu, Zhilu Zhang, Cheyanne Slocum, Cagla Deniz Bahadir, Mert Rory Sabuncu, and David J. Pisapia. Machine learning can aid in prediction of idh mutation from h&e-stained histology slides in infiltrating gliomas. *Scientific Reports*, 12, 2022.

[138] Haoming Lin, Yuyang Hu, Siping Chen, Jianhua Yao, and Ling Zhang. Fine-grained classification of cervical cells using morphological and appearance based convolutional neural networks. *IEEE Access*, PP:1–1, 05 2019.

[139] Guangqi Liu, Qinghai Ding, Haibo Luo, Min Sha, Xiang Li, and Moran Ju. Cx22: A new publicly available dataset for deep learning-based segmentation of cervical cytology images. *Computers in Biology and Medicine*, 150:106194, 2022.

[140] Hao Liu, Jiwen Lu, Jianjiang Feng, and Jie Zhou. Ordinal deep feature learning for facial age estimation. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 157–164, 2017.

[141] Kuan Liu, Yanen Li, Ning Xu, and Prem Natarajan. Learn to combine modalities in multimodal deep learning, 2018.

[142] Sidong Liu, Zubair Shah, Aydin Sav, Carlo Russo, Shlomo Berkovsky, Yi Qian, Enrico Coiera, and Antonio Di Ieva. Isocitrate dehydrogenase (idh) status prediction in histopathology images of gliomas using deep learning. *Scientific Reports*, 10:7733, 05 2020.

[143] Yu Liu, Xun Chen, Hu Peng, and Z. Wang. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion*, 36, 12 2016.

[144] Yu Liu, Shuping Liu, and Zengfu Wang. Multi-focus image fusion with dense SIFT. *Information Fusion*, 23:139–155, 05 2015.

[145] David Louis, Arie Perry, Pieter Wesseling, Daniel Brat, Ian Cree, Dominique Figarella-Branger, Cynthia Hawkins, H Ng, Stefan Pfister, Guido Reifenberger, Riccardo Soffietti, Andreas Deimling, and David Ellison. The 2021 who classification of tumors of the central nervous system: a summary. *Neuro-Oncology*, 23, 06 2021.

[146] Zhi Lu, Gustavo Carneiro, and Andrew P. Bradley. An improved joint optimization of multiple level set functions for the segmentation of overlapping cervical cells. *Trans. Img. Proc.*, 24(4):1261–1272, apr 2015.

[147] Zhi Lu, Gustavo Carneiro, Andrew P. Bradley, Daniela Ushizima, Masoud S. Nosrati, Andrea G. C. Bianchi, Claudia M. Carneiro, and Ghassan Hamarneh. Evaluation of three algorithms for the segmentation of overlapping cervical cells. *IEEE Journal of Biomedical and Health Informatics*, 21(2):441–450, 2017.

[148] Petru Manescu, Lydia Neary Zajiczek, Michael J. Shaw, Muna Elmi, Remy Claveau, Vijay Pawar, John Shawe-Taylor, Iasonas Kokkinos, Mandayam A. Srinivasan, Ikeoluwa Lagunju, Olugbemiro Sodeinde, Biobele J. Brown, and Delmiro Fernandez-Reyes. Deep learning enhanced extended depth-of-field for thick blood-film malaria high-throughput microscopy, 2019.

[149] Yannis Marinakis, Georgios Dounias, and Jan Jantzen. Pap smear diagnosis using a hybrid intelligent scheme focusing on genetic algorithm based feature selection and nearest neighbor classification. *Computers in biology and medicine*, 39:69–78, 02 2009.

[150] Peter McCullagh. Regression models for ordinal data. *Journal of the Royal Statistical Society. Series B (Methodological)*, 42(2):109–142, 1980.

[151] Claire McQuin, Allen Goodman, Vasiliy Chernyshev, Lee Kamentsky, Beth A. Cimini, Kyle W. Karhohs, Minh Doan, Liya Ding, Susanne M. Rafelski, Derek Thirstrup, Winfried Wiegraebe, Shantanu Singh, Tim Becker, Juan C. Caicedo, and Anne E. Carpenter. Cellprofiler 3.0: Next-generation image processing for biology. *PLOS Biology*, 16(7):1–17, 07 2018.

[152] Mortimer L. Mendelsohn and Brian H. Mayall. Computer-oriented analysis of human chromosomes—iii. focus. *Computers in Biology and Medicine*, 2(2):137–150, 1972. Chromosome Analysis.

[153] Zhu Meng, Zhicheng Zhao, Bingyang Li, Fei Su, and Limei Guo. A cervical histopathology dataset for computer aided diagnosis of precancerous lesions. *IEEE Transactions on Medical Imaging*, 40(6):1531–1541, 2021.

[154] Ray Merrill. Cancer precursors: Epidemiology, detection, and prevention. *American Journal of Epidemiology*, 156, 08 2002.

[155] Benjamin Milgrom, Roy Avrahamy, Tal David, Avi Caspi, Yosef Golovachev, and Shlomo Engelberg. Extended depth-of-field imaging employing integrated binary phase pupil mask and principal component analysis image fusion. *Opt. Express*, 28(16):23862–23873, Aug 2020.

[156] Aurora M Miranda. Female Reproductive Organ Anatomy: Overview, Gross Anatomy, Microscopic Anatomy, 2018.

[157] Mehryar Mohri, Afshin Rostamizadeh, and Ameet Talwalkar. *Foundations of Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, 2 edition, 2018.

[158] Xavier Moles Lopez, Etienne D'Andrea, Paul Barbot, Anne-Sophie Bridoux, Sandrine Rorive, Isabelle Salmon, Olivier Debeir, and Christine Decaestecker. An automated blur detection method for histological whole slide imaging. *PloS one*, 8:e82710, 12 2013.

[159] M. Moodley and K. Moodley. Trends of histopathology of cervical cancer among women in Durban, South Africa. *European Journal of Gynaecological Oncology*, 2009.

[160] John E. Moody and Joachim Utans. Architecture selection strategies for neural networks: Application to corporate bond rating predicti. In *NIPS 1995*, 1995.

[161] Raúl Murillo and Camila Ordóñez-Reyes. Human papillomavirus (HPV) vaccination: From clinical studies to immunization programs, 2019.

[162] Vinod Nair and Geoffrey Hinton. Rectified linear units improve restricted boltzmann machines vinod nair. volume 27, pages 807–814, 06 2010.

[163] Ali Naqvi, Niti Manglik, Ellen Dudrey, Cynthia Perry, Zuber Mulla, and Jorge Cervantes. Evaluating the performance of a low-cost mobile phone attachable microscope in cervical cytology. *BMC Women's Health*, 20, 03 2020.

[164] Mansour Nejati, Shadrokh Samavi, and Shahram Shirani. Multi-focus image fusion using dictionary-based sparse representation. *Information Fusion*, 25:72–84, 2015.

[165] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, and Gang Hua. Ordinal regression with multiple output cnn for age estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[166] Kehinde Sharafadeen Okunade. Human papillomavirus and cervical cancer. *Journal of Obstetrics and Gynaecology*, 40(5):602–608, 2020. PMID: 31500479.

[167] Sara Oliveira, Diana Montezuma Felizardo, Ana Moreira, Domingos Oliveira, Pedro Neto, Ana Monteiro, João Monteiro, Liliana Ribeiro, Sofia Gonçalves, Isabel Pinto, and Jaime Cardoso. A cad system for automatic dysplasia grading on h&e cervical whole-slide images. *Scientific Reports*, 13, 03 2023.

[168] Sara Oliveira, Pedro C. Neto, Joao Fraga, Diana Montezuma, Ana Monteiro, Joao Monteiro, Liliana Ribeiro, Sofia Goncalves, Isabel M. Pinto, and Jaime S. Cardoso. Cad systems for colorectal cancer from wsi are still not ready for clinical acceptance. *Scientific Reports (Nature)*, 2021.

[169] World Health Organization. Cervical cancer. `https://www.who.int/health-topics/cervical-cancer#tab=tab_1`. (Accessed on 06/30/2023).

[170] World Health Organization. Un joint global programme on cervical cancer prevention and control. `https://www.who.int/publications/m/item/un-joint-global-programme-on-cervical-cancer-prevention-and-control`, November 2016. (Accessed on 06/30/2023).

[171] Quinn T Ostrom, Gino Cioffi, Haley Gittleman, Nirav Patil, Kristin Waite, Carol Kruchko, and Jill S Barnholtz-Sloan. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012–2016. *Neuro-Oncology*, 21(Supplement_5):v1–v100, 11 2019.

[172] Frank Palermo, James Hays, and Alexei Efros. Dating historical color images. In *European Conference on Computer Vision*, volume 7577, pages 499–512, 10 2012.

[173] Frank Palermo, James Hays, and Alexei A. Efros. Dating historical color images. In *ECCV (6)*, pages 499–512, 2012.

[174] Hongyu Pan, Hu Han, Shiguang Shan, and Xilin Chen. Mean-variance loss for deep age estimation from a face. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5285–5294, 2018.

[175] Payel Rudra Paul, Mrinal Kanti Bhowmik, and Debotosh Bhattacharjee. *Automated cervical cancer detection using Pap smear images*, pages 267–278. Springer, 2015.

[176] J. G. Payette. Intel and mobileodt cervical cancer screening kaggle competition : Cervix type classification using deep learning and image classification. 2017.

[177] Carlos Pereira, Paulo T. Silva, Luís Rosado, Luís Mota, and João Martins. The design thinking process in the development of an intelligent microscopic equipment. In *Advances in Design and Digital Communication II*, pages 170–182, Cham, 2022. Springer International Publishing.

[178] Hady Ahmady Phoulady and Peter R. Mouton. A new cervical cytology dataset for nucleus detection and image classification (Cervix93) and methods for cervical nucleus detection. *CoRR*, abs/1811.09651, 2018.

[179] Gemma Piella. A general framework for multiresolution image fusion: From pixels to regions. *Information Fusion*, 4:259–280, 12 2003.

[180] Sharmila Pimple and Awadhesh Mishra. Global strategies for cervical cancer prevention and screening. *Minerva Ginecologica*, 71, 07 2019.

[181] Henry Pinkard, Zachary Phillips, Arman Babakhani, Daniel A. Fletcher, and Laura Waller. Deep learning for single-shot autofocus microscopy. *Optica*, 6(6):794–797, Jun 2019.

[182] Marina E. Plissiti, P. Dimitrakopoulos, G. Sfikas, Christophoros Nikou, O. Krikoni, and A. Charchanti. Sipakmed: A new dataset for feature and image based classification of normal and pathological cervical cells in pap smear images. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3144–3148, 2018.

[183] Gorkem Polat, Ilkay Ergenc, Haluk Tarik Kani, Yesim Ozen Alahdab, Ozlen Atug, and Alptekin Temizel. Class distance weighted cross-entropy loss for ulcerative colitis severity estimation. *arXiv preprint arXiv:2202.05167*, 2022.

[184] Jan Pontén, Hans-Oiov Adami, Reinhold Bergström, Joakim Dillner, Lars-Gösta Friberg, Leif Gustafsson, Anthony B. Miller, D. Max Parkin, Par Sparén, and Dimitrios Trichopoulos. Strategies for global control of cervical cancer. *International Journal of Cancer*, 60(1):1–26, 1995.

[185] W Prendiville and R Sankaranarayanan. *Colposcopy and Treatment of Cervical Precancer*. IARC Technical Publication No. 45, 2017.

[186] Lei Qu, Fuhui Long, and Hanchuan Peng. 3-D registration of biological images and models: Registration of microscopic images and its uses in segmentation and annotation. *IEEE Signal Processing Magazine*, 32(1):70–77, 2015.

[187] Diganta Rabha, A. SARMAH, and Pabitra Nath. Design of a 3d printed smartphone microscopic system with enhanced imaging ability for biomedical applications. *Journal of Microscopy*, 276, 09 2019.

[188] Juan Andrés Ramírez-González, Ricardo Vaamonde-Lemos, Joao Cunha Filho, Alex Varghese, and R James Swanson. Overview of the Female Reproductive System. In *Exercise and Human Reproduction: Induced Fertility Disorders and Possible Therapies*, pages 19–46. 2016.

[189] Shalinee Rao, Suresh Masilamani, Sandhya Sundaram, Duvuru Prathiba, and Swaminathan Rajendiran. Quality measures in pre-analytical phase of tissue processing: Understanding

its value in histopathology. *Journal of clinical and diagnostic research : JCDR*, 10:EC07–EC11, 02 2016.

[190] Saima Rathore, Tamim Niazi, Aksam Iftikhar, and Ahmad Chaddad. Glioma grading via analysis of digital pathology images using machine learning. *Cancers*, 12:578, 03 2020.

[191] Peter Regitnig, Heimo Müller, and Andreas Holzinger. *Expectations of Artificial Intelligence for Pathology*, pages 1–15. 06 2020.

[192] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1*, NIPS'15, page 91–99, Cambridge, MA, USA, 2015. MIT Press.

[193] Yair Rivenson, Zoltán Göröcs, Harun Günaydin, Yibo Zhang, Hongda Wang, and Aydogan Ozcan. Deep learning microscopy. *Optica*, 4(11):1437–1443, Nov 2017.

[194] Luís Rosado, Paulo Silva, José Faria, João Oliveira, Maria Vasconcelos, Dirk Elias, José Manuel Correia da Costa, and Jaime Cardoso. *µSmartScope: Towards a Fully Automated 3D-Printed Smartphone Microscope with Motorized Stage*, pages 19–44. Springer International Publishing, 07 2018.

[195] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. ORB: An efficient alternative to SIFT or SURF. In *2011 International Conference on Computer Vision*, pages 2564–2571, 2011.

[196] D. Rueckert, L.I. Sonoda, C. Hayes, D.L.G. Hill, M.O. Leach, and D.J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, 1999.

[197] PushpLata Sachan, Meenakshi Singh, Munna Patel, and Rekha Sachan. A study on cervical cancer screening using pap smear test and clinical correlation. *Asia-Pacific Journal of Oncology Nursing*, 5:337, 07 2018.

[198] Jesus Salido, Gloria Bueno, Jesus Ruiz-Santaquiteria, and Gabriel Cristobal. A review on low-cost microscopes for open science. *Microscopy Research and Technique*, 2022.

[199] Jesús Salido, Carlos Sánchez, Jesús Ruiz-Santaquiteria, Gabriel Cristóbal, Saul Blanco, and Gloria Bueno. A low-cost automated digital microscopy platform for automatic identification of diatoms. *Applied Sciences*, 10(17), 2020.

[200] Ana Sampaio, Tomé Albuquerque, Pedro Faria, Luís Rosado, and Maria Vasconcelos. Fraunhofer portugal aicos edof dataset, January 2023. Funded by European Regional Fund (ERDF) through the North Portugal Regional Operational Program- NORTE 2020 and by the Portuguese Foundation for Science and Technology-FCT under the CMU-Portugal International Partnership: TAMI-Transparent Artificial Medical Intelligence (NORTE-01-0247-FEDER-045905).

[201] Ana Filipa Sampaio, Luís Rosado, and Maria João M. Vasconcelos. Towards the mobile detection of cervical lesions: A region-based approach for the analysis of microscopic images. *IEEE Access*, 9:152188–152205, 2021.

[202] Milena Sant, Claudia Allemani, Mariano Santaquilani, Arnold Knijn, Francesca Marchesi, and Riccardo Capocaccia. Eurocare-4. survival of cancer patients diagnosed in 1995–1999. results and commentary. *European Journal of Cancer*, 45(6):931–991, 2009. Survival of cancer patients in Europe, 1995–2002: The EUROCARE 4 study.

[203] A Santos, Carlos Ortiz-de Solorzano, Juan Jose Vaquero, J Peña, Norberto Malpica, and Francisco Del Pozo Guerrero. Evaluation of autofocus functions in molecular cytogenetic analysis. *Journal of microscopy*, 188:264–72, 01 1998.

[204] Caroline Schneider, Wayne Rasband, and Kevin Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature Methods*, 9, 07 2012.

[205] J W Sellors and R Sankaranarayanan. Colposcopy and Treatment of Cervical Intraepithelial Neoplasia. A Beginner's Manual. *IARC Press (Lyon)*, 2013.

[206] Caglar Senaras, M Khalid Khan Niazi, Gerard Lozanski, and Metin N Gurcan. DeepFocus: Detection of out-of-focus regions in whole slide digital images using deep learning. *PloS one*, 13(10):e0205387–e0205387, oct 2018.

[207] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 70(1):7–30, 2020.

[208] Rebecca L. Siegel, Kimberly D. Miller, and Ahmedin Jemal. Cancer statistics, 2020. *CA: A Cancer Journal for Clinicians*, 2020.

[209] Wilson Silva, João Ribeiro Pinto, and Jaime S Cardoso. A uniform performance index for ordinal classification with imbalanced classes. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2018.

[210] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[211] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[212] Diane Solomon, Diane Davey, Robert Kurman, Ann Moriarty, Dennis O'Connor, Marianne Prey, Stephen Raab, Mark Sherman, David Wilbur, Thomas Wright, and Nancy Young. The 2001 bethesda system: Terminology for reporting results of cervical cytology. *Journal of the American Medical Association*, 2002.

[213] Dezhao Song, Edward Kim, Xiaolei Huang, Joseph Patruno, Hector Munoz-Avila, Jeff Heflin, L. Long, and Sameer Antani. Multimodal entity coreference for cervical dysplasia diagnosis. *IEEE transactions on medical imaging*, 34, 08 2014.

[214] Dezhao Song, Edward Kim, Xiaolei Huang, Joseph Patruno, Hector Munoz-Avila, Jeff Heflin, L. Long, and Sameer Antani. Multimodal entity coreference for cervical dysplasia diagnosis. *IEEE Transactions on Medical Imaging*, 34:229 – 245, 01 2015.

[215] Ricardo Sousa and Jaime S. Cardoso. Ensemble of decision trees with global constraints for ordinal classification. In *2011 11th International Conference on Intelligent Systems Design and Applications*, pages 1164–1169, 2011.

[216] P Sparén, L Gustafsson, L G Friberg, J Pontén, R Bergström, and H O Adami. Improved control of invasive cervical cancer in sweden over six decades by earlier clinical detection and better treatment. *Journal of Clinical Oncology*, 13(3):715–725, 1995. PMID: 7884431.

[217] S Spuhler. Role of the histologic examination in the diagnosis of cervical intraepithelial cancer. *Journal de gynecologie, obstetrique et biologie de la reproduction*, 25(2):124–127, 1996.

[218] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 06 2014.

[219] Tong Sun, Yang Gao, Wenjie Tan, Sufang Ma, Yuankai Shi, Jiarui Yao, Yongli Guo, Ming Yang, Xuemei Zhang, Qingrun Zhang, Changqing Zeng, and Dongxin Lin. A six-nucleotide insertion-deletion polymorphism in the casp8 promoter is associated with susceptibility to multiple cancers. *Nature genetics*, 39:605–13, 06 2007.

[220] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014.

[221] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[222] Mingxing Tan and Quoc Le. EfficientNet: Rethinking model scaling for convolutional neural networks. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6105–6114. PMLR, 09–15 Jun 2019.

[223] JM Tenenbaum. *Accommodation in Computer Vision Stanford University Ph. D*. PhD thesis, Thesis, 1970.

[224] A. Toet. A morphological pyramidal image decomposition. *Pattern Recognition Letters*, 9(4):255–261, 1989.

[225] Michael A. Unser. Splines: a perfect fit for signal and image processing. *IEEE Signal Process. Mag.*, 16:22–38, 1999.

[226] Lanh Tran Van, Sabu Emmanuel, and Mohan S Kankanhalli. Identifying source cell phone using chromatic aberration. In *2007 IEEE International Conference on Multimedia and Expo*, pages 883–886, 2007.

[227] Ana Beatriz Vieira, Tomé Albuquerque, Diana Montezum, Liliana Ribeiro, Domingos Oliveira, Isabel Macedo Pinto, Jaime Cardoso, and Arlindo Oliveira. To reject or not to reject, that is the question: A new perspective on counting fragments in whole-slide-images. In *29th Portuguese Conference on Pattern Recognition (RECPAD), Coimbra, Portugal, 27 October*, 10 2023.

[228] Ana Beatriz Vieira, Diana Montezuma, Tomé Albuquerque, Liliana Ribeiro, Filipa Rebolo, Domingos Oliveira, João Monteiro, Sofia Gonçalves, Isabel Macedo Pinto, Jaime Cardoso, and Arlindo Oliveira. Automatic detection and counting of pathological specimens with

a rejection option. In *35th European Congress of Pathology (ECP), Dublin, Ireland, 9-13 September*, 2023.

[229] Henry R. Wabinga, Sarah Nambooze, Phoebe Mary Amulen, Catherine Okello, Louise Mbus, and Donald Maxwell Parkin. Trends in the incidence of cancer in Kampala, Uganda 1991-2010. *International Journal of Cancer*, 2014.

[230] Renjie Wang, Wei Pan, Lei Jin, Weiming Huang, Yuehan Li, Di Wu, Chun Gao, Ding Ma, and Shujie Liao. Human papillomavirus vaccine against cervical cancer: Opportunity and challenge. *Cancer Letters*, 471:88–102, 2020.

[231] Shouyi Wang, Wanpracha Chaovalitwongse, and Robert Babuska. Machine learning algorithms in bipedal robot control. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(5):728–743, 2012.

[232] Shouyi Wang, Wanpracha Chaovalitwongse, and Robert Babuska. Machine Learning Algorithms in Bipedal Robot Control. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 42:728–743, 2012.

[233] Zhaobin Wang, Yide Ma, and Jason Gu. Multi-focus image fusion using pcnn. *Pattern Recognition*, 43(6):2003–2016, 2010.

[234] Zhongling Wang, Mahdi S. Hosseini, Adyn Miles, Konstantinos N. Plataniotis, and Zhou Wang. Focuslitenn: High efficiency focus quality assessment for digital pathology, 2020.

[235] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.

[236] Zhou Wang and Alan C. Bovik. Modern image quality assessment. *Synthesis Lectures on Image, Video, and Multimedia Processing*, 2005.

[237] Benjamin T Whitfield and Jason T. Huse. Classification of adult-type diffuse gliomas: Impact of the world health organization 2021 update. *Brain Pathology*, 32, 2022.

[238] Gio Wiederhold and John McCarthy. Arthur samuel: Pioneer in machine learning. *IBM Journal of Research and Development*, 36:329 – 331, 06 1992.

[239] Wasswa William, Andrew Ware, Annabella Habinka Basaza-Ejiri, and Johnes Obungoloch. Cervical cancer classification from pap-smears using an enhanced fuzzy c-means algorithm. *Informatics in Medicine Unlocked*, 14:23 – 33, 2019.

[240] Marcin Woźniak, Jakub Siłka, and Michał Wieczorek. Deep neural network correlation learning mechanism for ct brain tumor detection. *Neural Computing and Applications*, 03 2021.

[241] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[242] Haitao Xiong, Hongfu Liu, Bineng Zhong, and Yun Fu. Structured and sparse annotations for image emotion distribution learning. *Proceedings of the AAAI Conference on Artificial Intelligence*, 33(01):363–370, Jul. 2019.

[243] Tao Xu, Han Zhang, Xiaolei Huang, Shaoting Zhang, and Dimitris N. Metaxas. Multimodal deep learning for cervical dysplasia diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016.

[244] Tao Xu, Han Zhang, Xiaolei Huang, Shaoting Zhang, and Dimitris N. Metaxas. Multimodal deep learning for cervical dysplasia diagnosis. In *MICCAI 2016*, pages 115–123, Cham, 2016. Springer International Publishing.

[245] Tao Xu, Han Zhang, Cheng Xin, Edward Kim, L. Long, Zhiyun Xue, Sameer Antani, and Xiaolei Huang. Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognition*, 63, 09 2016.

[246] Tao Xu, Han Zhang, Cheng Xin, Edward Kim, L. Long, Zhiyun Xue, Sameer Antani, and Xiaolei Huang. Multi-feature based benchmark for cervical dysplasia classification evaluation. *Pattern Recognition*, 63, 09 2016.

[247] Peng Xue, Man Ng, and Youlin Qiao. The challenges of colposcopy for cervical cancer screening in lmics and solutions by artificial intelligence. *BMC Medicine*, 18, 12 2020.

[248] Yunzhe Xue, Yanan Yang, Fadi G. Farhat, Frank Y. Shih, Olga Boukrina, A. M. Barrett, Jeffrey R. Binder, William W. Graves, and Usman W. Roshan. Brain tumor classification with tumor segmentations and a dual path residual convolutional neural network from mri and pathology images. In Alessandro Crimi and Spyridon Bakas, editors, *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, pages 360–367, Cham, 2020. Springer International Publishing.

[249] Jing Yan, Shenghai Zhang, Qiuchang Sun, Weiwei Wang, Wenchao Duan, Li Wang, Tianqing Ding, Pei Dongling, Chen Sun, Wenqing Wang, Zhen Liu, Xuanke Hong, Xiangxiang Wang, Yu Guo, Wencai Li, Jingliang Cheng, Xianzhi Liu, Zhi-Cheng Li, and Zhenyu Zhang. Predicting 1p/19q co-deletion status from magnetic resonance imaging using deep learning in adult-type diffuse lower-grade gliomas: a discovery and validation study. *Laboratory Investigation*, 102:154–159, 03 2022.

[250] Bin Yang and Shutao Li. Multifocus image fusion and restoration with sparse representation. *IEEE T. Instrumentation and Measurement*, 59:884–892, 04 2010.

[251] Samuel J Yang, Marc Berndl, D Michael Ando, Mariya Barch, Arunachalam Narayanaswamy, Eric Christiansen, Stephan Hoyer, Chris Roat, Jane Hung, Curtis T Rueden, Asim Shankar, Steven Finkbeiner, and Philip Nelson. Assessing microscope image focus quality with deep learning. *BMC Bioinformatics*, 19(1):77, 2018.

[252] Xiefeng Yang, Yu Lin, Zhen Xing, Dejun She, Yanqing Su, and Dairong Cao. Predicting 1p/19q codeletion status using diffusion-, susceptibility-, perfusion-weighted, and conventional mri in idh-mutant lower-grade gliomas. *Acta Radiologica*, 62:1657 – 1665, 2020.

[253] TTE Yeo, SH Ong, Jayasooriah, and R Sinniah. Autofocusing for tissue microscopy. *Image and Vision Computing*, 11(10):629–639, 1993.

[254] Nicole S. L. Yeo-Teh, Yoshiaki Ito, and Sudhakar Jha. High-risk human papillomaviral oncogenes e6 and e7 target key cellular pathways to achieve oncogenesis. *International Journal of Molecular Sciences*, 19(6), 2018.

[255] Yao Yu, Jie Ma, Weidong Zhao, Zhenmin Li, and Shuai Ding. Msci: A multistate dataset for colposcopy image classification of cervical cancer screening. *International Journal of Medical Informatics*, 146:104352, 2021.

[256] Sergey Zagoruyko and Nikos Komodakis. Wide residual networks. *arXiv preprint arXiv:1605.07146*, 2016.

[257] Ling Zhang, Le Lu, Isabella Nogues, Ronald Summers, Shaoxiong Liu, and Jianhua Yao. Deeppap: Deep convolutional networks for cervical cell classification. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 05 2017.

[258] Qiang Zhang and Bao long Guo. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing*, 89(7):1334–1346, 2009.

[259] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, and Jian Sun. ShuffleNet: An extremely efficient convolutional neural network for mobile devices. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6848–6856, 2018.

[260] Yongxin Zhang, Li Chen, Zhihua Zhao, and Jian Jia. Multi-focus image fusion with sparse feature based pulse coupled neural network. *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, 12:357, 06 2014.

[261] Hao Zhou, Ken Chang, Harrison Bai, Bo Xiao, Chang Su, Wenya Linda Bi, Paul Zhang, Joeky Senders, Martin Vallières, Vasileios Kavouridis, Alessandro Boaro, Omar Arnaout, li Yang, and Raymond Huang. Machine learning reveals multimodal mri patterns predictive of isocitrate dehydrogenase and 1p/19q status in diffuse low- and high-grade gliomas. *Journal of Neuro-Oncology*, 142, 04 2019.

[262] Ying Zhou, Hui Xue, and Xin Geng. Emotion distribution recognition from facial expressions. In *Proceedings of the 23rd ACM International Conference on Multimedia*, MM '15, page 1247–1250, New York, NY, USA, 2015. Association for Computing Machinery.

[263] Haiping Zhu, Yuheng Zhang, Guohao Li, Junping Zhang, and Hongming Shan. Ordinal distribution regression for gait-based age estimation, 05 2019.

Appendix A

Image Enhancement using Deep Based Extended Depth of Field Methods

Table A.1: Results Fraunhofer dataset no chromatic align, only elastic.

		No Chromatic alignment + Elastic Alignment			
EDoF-CNN	**stack**	**MSE**	**RMSE**	**PSNR (dB)**	**SSIM (%)**
Max [148]	3	0.00310 ± 0.00069	0.07427 ± 0.00846	27.962 ± 0.915	94.869 ± 0.533
	5	0.00310 ± 0.00068	0.07408 ± 0.00843	28.010 ± 0.909	95.070 ± 0.759
3D [148]	3	0.00311 ± 0.00069	0.07439 ± 0.00851	27.940 ± 0.930	95.241 ± 0.744
	5	0.00310 ± 0.00069	0.07406 ± 0.00850	28.023 ± 0.940	95.591 ± 0.592
Fast (P)	3	0.00300 ± 0.00069	0.07193 ± 0.00902	28.526 ± 1.132	$\mathbf{95.955 \pm 0.798}$
	5	$\mathbf{0.00300 \pm 0.00069}$	$\mathbf{0.07182 \pm 0.00891}$	$\mathbf{28.538 \pm 1.097}$	95.678 ± 0.686
Pairwise (P)	3	0.00303 ± 0.00069	0.07246 ± 0.00886	28.393 ± 1.079	95.910 ± 0.620
	5	0.00301 ± 0.00069	0.07209 ± 0.00898	28.482 ± 1.120	95.893 ± 0.785

Appendix B

Non-parametric Unimodal Losses (CO, CO2, and HO2)

Table B.1: Results in terms of Kendall's τ for 7 class problem, averaged for 10 folds (higher is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	75.9±4.1	76.9±4.8	74.5±5.8	*77.9±4.8*	57.3±9.3	**79.9±4.7**	77.5±5.5
GoogLeNet	*80.9±2.7*	*81.9±4.1*	79.3±4.4	*81.9±4.3*	76.9±4.0	*81.7±4.2*	**82.8±3.7**
MobileNet_v2	83.2±2.9	84.5±1.9	**85.6±2.5**	*84.4±2.1*	75.6±11.5	83.7±3.2	83.8±3.0
ResNet18	*83.1±5.1*	**84.7±3.1**	*83.4±3.0*	*83.5±4.9*	76.9±5.1	*83.8±4.0*	*83.4±5.1*
ResNeXt50_32x4d	83.2±4.5	*85.2±2.8*	*84.7±1.4*	83.2±3.6	81.5±3.4	*84.0±3.6*	**85.4±3.1**
ShuffleNet_v2_x1_0	78.8±4.0	78.9±3.6	*80.9±2.4*	**81.7±3.1**	77.8±4.5	80.1±2.6	*81.6±3.4*
SqueezeNet1_0	**81.0±3.9**	*80.4±2.6*	77.0±4.0	*79.6±5.1*	54.1±14.6	*79.4±4.1*	78.1±4.8
VGG16	*81.6±5.1*	*81.3±2.2*	78.3±6.4	*81.9±3.5*	68.8±6.9	*82.6±3.2*	**82.8±4.6**
Wide_ResNet50_2	83.4±3.1	83.8±2.8	84.9±3.4	**85.9±2.7**	79.8±6.1	84.1±3.6	82.7±4.3
Avg	81.2	82.0	80.9	82.2	72.1	82.2	82.0
Winners	1	1	1	2	0	1	3

Table B.2: Results in terms of Kendall's τ for 4 class problem, averaged for 10 folds (higher is better).

	CE	BU	PU	OE	CO	CO2	HO2
AlexNet	73.9±6.3	75.4±4.0	**77.4±3.8**	*76.0±5.7*	58.7±19.3	*76.5±5.7*	75.0±5.9
GoogLeNet	80.3±4.2	*81.3±3.0*	*80.4±4.9*	*81.2±4.4*	70.1±14.2	**82.5±4.2**	80.2±5.1
MobileNet_v2	81.3±5.1	**83.9±2.0**	81.3±3.7	*81.8±6.2*	*82.3±3.7*	80.4±4.3	82.3±4.4
ResNet18	81.0±2.7	80.3±4.4	*81.2±4.4*	**82.8±3.5**	77.5±7.7	*82.5±3.2*	79.5±5.1
ResNeXt50_32x4d	*83.2±3.1*	*83.2±2.9*	81.8±2.8	**84.3±3.4**	78.1±5.1	*83.9±2.4*	82.3±3.9
ShuffleNet_v2_x1_0	77.1±4.7	76.1±3.7	**78.8±4.1**	74.8±6.2	71.1±9.1	*77.3±5.6*	*78.0±4.5*
SqueezeNet1_0	*76.2±5.2*	*77.5±4.5*	*75.8±5.6*	**78.2±5.9**	48.5±11.0	*76.6±3.8*	74.8±3.8
VGG16	77.9±5.4	*79.2±4.5*	*80.0±4.4*	**81.0±2.4**	63.2±10.3	79.2±4.2	78.3±5.4
Wide_ResNet50_2	81.7±4.6	*83.4±2.8*	**84.4±4.6**	*82.2±5.1*	67.8±18.6	*83.2±4.3*	*82.6±2.8*
Avg	79.2	80.0	80.1	80.2	68.6	80.3	79.2
Winners	0	1	3	4	0	1	0

bold: best model, *italic:* statistically similar to best (paired t-test).

Table B.3: Aggregate results for 7 class problem, averaged for 10 folds.

	CE	BU	PU	OE	CO	CO2	HO2
				Mode			
UOC	*38.8±7.5*	44.2±7.2	41.1±7.7	**38.6±7.5**	53.5±14.7	*39.0±7.3*	39.5±8.1
MAE	0.37±0.08	0.41±0.09	0.40±0.10	**0.36±0.08**	0.60±0.26	*0.37±0.07*	*0.37±0.09*
Accuracy	**73.4±4.8**	66.2±5.5	71.5±4.9	72.8±5.1	58.6±13.4	72.6±4.5	72.0±5.6
Kendall's τ	81.2±4.7	*82.0±4.2*	80.0±5.8	**82.2±4.5**	72.1±12.4	*82.2±4.2*	*82.0±4.9*
ROC AUC	**95.9±1.4**	93.0±2.2	95.5±1.4	95.5±1.5	82.3±9.9	94.5±1.8	93.9±1.9
Gini	85.1±0.2	64.0±1.6	84.8±0.5	84.8±0.4	28.2±33.3	50.1±6.3	45.0±4.2
				Mean			
UOC	**39.2±7.4**	42.1±7.5	41.7±8.2	*39.7±7.4*	75.9±20.6	79.8±4.9	83.7±1.7
MAE	*0.37±0.08*	0.39±0.08	0.39±0.09	**0.37±0.08**	1.17±0.50	0.91±0.13	1.03±0.04
Accuracy	**72.1±5.1**	67.7±5.8	70.1±5.6	*71.6±5.1*	33.6±19.9	28.3±5.4	25.0±3.1
Kendall's τ	*82.3±4.3*	*82.5±4.1*	80.9±5.5	**82.6±4.3**	*nan±nan*	78.3±3.6	76.0±3.2
ROC AUC	**95.9±1.4**	93.0±2.2	95.5±1.4	95.5±1.5	82.3±9.9	94.5±1.8	93.9±1.9
Gini	85.1±0.2	64.0±1.6	84.8±0.5	84.8±0.4	28.2±33.3	50.1±6.3	45.0±4.2

Table B.4: Aggregate results for 4 class problem, averaged for 10 folds.

	CE	BU	PU	OE	CO	CO2	HO2
				Mode			
UOC	*33.1 ± 6.3*	33.9 ± 5.9	33.6 ± 5.7	**32.7 ± 7.0**	45.8 ± 14.9	*33.1 ± 6.3*	34.2 ± 6.4
MAE	0.25 ± 0.06	0.27 ± 0.06	*0.25 ± 0.06*	**0.24 ± 0.06**	0.40 ± 0.20	*0.25 ± 0.06*	0.26 ± 0.06
Accuracy	*79.2 ± 4.1*	76.5 ± 4.5	*78.7 ± 4.0*	**79.2 ± 4.8**	67.4 ± 13.7	78.5 ± 4.2	77.8 ± 4.3
Kendall's τ	79.2 ± 5.5	*80.0 ± 4.7*	*79.7 ± 5.4*	*80.2 ± 5.8*	68.6 ± 15.7	**80.3 ± 5.1**	79.2 ± 5.4
ROC AUC	*94.5 ± 1.7*	92.7 ± 1.9	**94.6 ± 1.5**	*94.5 ± 1.7*	83.5 ± 10.5	93.2 ± 2.1	92.9 ± 2.0
Gini	74.3 ± 0.3	58.4 ± 2.0	74.2 ± 0.4	74.2 ± 0.3	31.5 ± 31.4	44.5 ± 16.3	36.6 ± 11.2
				Mean			
UOC	*33.4 ± 6.3*	34.2 ± 5.8	*34.0 ± 5.9*	**33.2 ± 7.1**	57.0 ± 17.3	52.2 ± 13.5	58.4 ± 8.5
MAE	*0.25 ± 0.06*	0.26 ± 0.05	*0.25 ± 0.06*	**0.25 ± 0.07**	0.50 ± 0.20	0.42 ± 0.14	0.47 ± 0.09
Accuracy	**78.5 ± 4.4**	77.0 ± 4.2	77.8 ± 4.4	*78.4 ± 5.1*	54.7 ± 17.8	60.1 ± 13.4	54.6 ± 9.3
Kendall's τ	79.7 ± 5.1	*80.4 ± 4.6*	*80.1 ± 4.9*	**80.5 ± 5.8**	63.8 ± 16.6	73.4 ± 7.3	70.5 ± 5.6
ROC AUC	*94.5 ± 1.7*	92.7 ± 1.9	**94.6 ± 1.5**	*94.5 ± 1.7*	83.5 ± 10.5	93.2 ± 2.1	92.9 ± 2.0
Gini	74.3 ± 0.3	58.4 ± 2.0	74.2 ± 0.4	74.2 ± 0.3	31.5 ± 31.4	44.5 ± 16.3	36.6 ± 11.2

bold: best model, *italic:* statistically similar to best (paired t-test).

Appendix C

Multimodal Data Integration

C.1 Multimodal Context-Aware Detection of Glioma Biomarkers using MRI and WSI

Foreword on Author Contributions
The results of this work have been disseminated in the form of two papers at international conferences.

> [9] - Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, Peter Schüffler, "Multimodal Identification of Brain Tumor Biomarkers using a Deep Fusion Strategy from MRI and Digital Pathology exams". 19th European Congress on Digital Pathology (ECDP), Budapest, Hungary, 2023.
>
> [4] - Tomé Albuquerque, Mei Ling Fang, Benedikt Wiestler, Claire Delbridge, Maria João M. Vasconcelos, Jaime S. Cardoso, Peter Schüffler, "Multimodal Context-Aware Detection of Glioma Biomarkers using MRI and WSI", 4th International Workshop on Multiscale Multimodal Medical Imaging (MMMI) - MICCAI, Vancouver, Canada, 8-12 October, 2023

The most malignant tumors of the central nervous system are adult-type diffuse gliomas. Historically, glioma subtype classification has been based on morphological features. However, since 2016, WHO recognizes that molecular evaluation is critical for subtyping. Among molecular markers, the mutation status of IDH1 and the codeletion of 1p/19q are crucial for the precise diagnosis of these malignancies. In pathology laboratories, however, manual screening for those markers is time-consuming and susceptible to error. To overcome these limitations, we propose a novel multimodal biomarker classification method that integrates image features derived from brain magnetic resonance imaging and histopathological exams. The proposed model consists of two branches, the first branch takes as input a multi-scale Hematoxylin and Eosin whole slide image, and the second branch uses the pre-segmented region of interest from the magnetic resonance imaging. Both branches are based on convolutional neural networks. After passing the exams by the two embedding branches, the output feature vectors are concatenated, and a multilayer perceptron is used to classify the glioma biomarkers as a multi-class problem. In this work,

several fusion strategies were studied, including a cascade model with mid-fusion; a mid-fusion model, a late fusion model, and a mid-context fusion model. The models were tested using a publicly available data set from The Cancer Genome Atlas. Our cross-validated classification models achieved an area under the curve of 0.874, 0.863, and 0.815 for the proposed multimodal, magnetic resonance imaging, and Hematoxylin and Eosin stain slide images respectively, indicating our multimodal model outperforms its unimodal counterparts and the state-of-the-art glioma biomarker classification methods.

C.1.1 Motivation and Context

Adult-type diffuse gliomas are the most common malignant tumors that grow in the human brain [171]. Diffuse tumors spread into healthy tissues, making it difficult to determine precisely a clear boundary between healthy and cancerous tissues. In addition, gliomas are highly infiltrative and resistant to immunotherapy, rendering them largely incurable. Despite advances in our understanding of gliomas, survival greatly depends on the subtype of tumors present. Therefore, classification and grading of gliomas have significant implications on prognostic and treatment planning [237]. In the past, the classification of diffuse gliomas has been based on morphological features, and, more recently, molecular markers have become increasingly important in the diagnosis process. Since 2016, the World Health Organization (WHO) expanded its definition of grading adult-type diffuse gliomas by solidifying the centrality of molecular biomarkers and reducing the dependency on phenotypic features. The subtyping under the update varies greatly on isocitrate dehydrogenase 1 (IDH1) mutation status and the 1p/19q codeletion status [145].

The current diagnostic workflow to detect adult-type diffuse gliomas includes Magnetic Resonance Imaging (MRI) and tissue acquisition (histology) [82] (Figure C.1). Usually, several complimentary 3D modalities, such as native T1, T1 with contrast agent (T1ce), T2, and Fluid Attenuation Inversion Recovery (FLAIR), are acquired through brain MRI to enable an initial phenotypic inspection. Then a biopsy, which remains the gold standard for diagnosis, is taken and routinely assessed by molecular analysis in order to differentiate diffuse gliomas [94]. Nevertheless, the genotypic analysis of molecular markers to confirm the histological tumor typing and grading is time-consuming, subjective, and expensive. With the advance of deep learning techniques and the rising computing capability, patterns that are indifferent to human eyes can be recognized by deep neural networks in a relatively short period. Considering the digital data derived from MRI and pathological slides in the form of whole slide images (WSI), it is now possible to apply deep learning to combine both modalities to predict the molecular markers and further subtype gliomas [16].

Since the update of the WHO protocol, intensive research has been conducted on biomarker classification and glioma grading. In recent years, several researchers tried to detect IDH1 and 1p/19q biomarkers from MRI or WSI [56, 135, 190]. However, most of the works only make a binary classification between IDH1-mutant and IDH1-wildtype or 1p/19q codeleted and non-codeleted. Figure C.2 shows the most recent and relevant methods found in the literature for glioma biomarker detection.

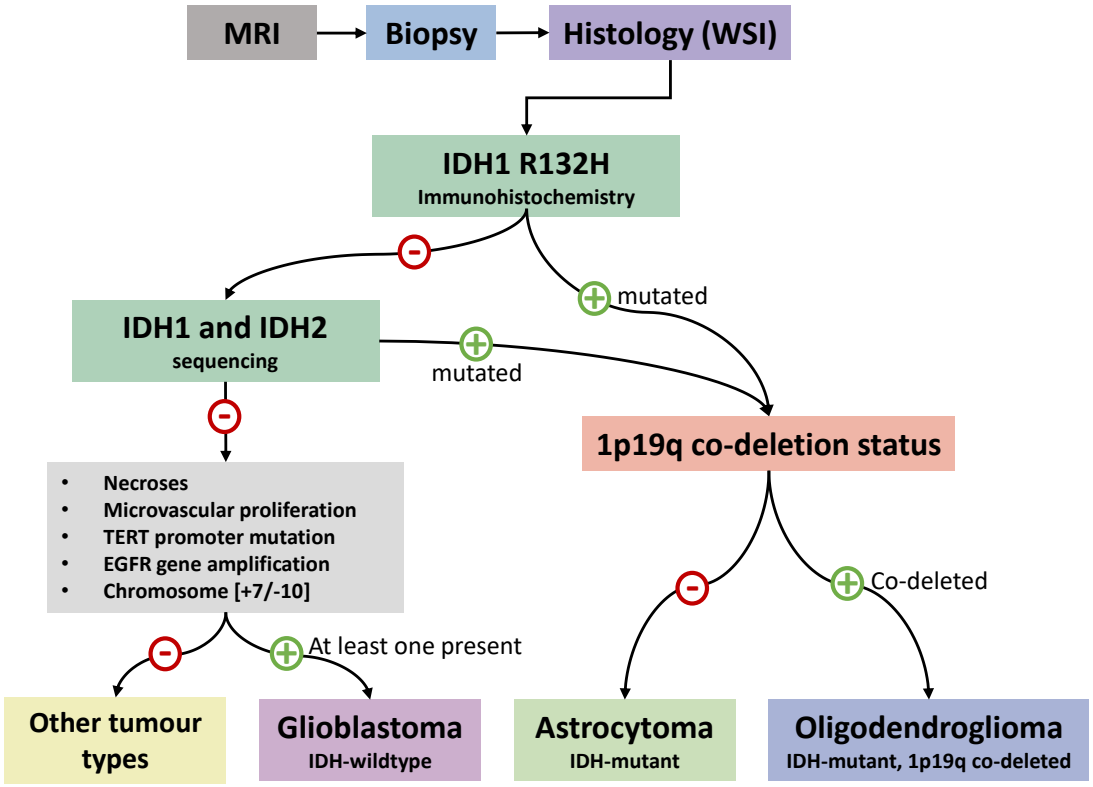


Figure C.1: Simplified overview of the classification pathway for the majority of adult-type diffuse gliomas based on WHO (2021) classification of CNS tumors.

In [248], the authors proposed a multi-modal tumor segmentation network that uses four MRI sequences, then combined with pathology images for tumor subtype classification based on biomarkers. Unlike [248], our proposed model is not focused on tumour subtype classification but on detecting the biomarkers. Also, we use public data and do cross-validation in contrast to [248]. Histopathology and MRI both assess the tumor phenotype, albeit on a vastly different scale (MRI: millimeter, microscopy: micrometer). Also, MRI captures the phenotype of the entire tumor. Thus, these methods promise to provide complimentary information on the tumor phenotype, and consequently the tumor genotype. However, given challenges in fusing these modalities (or their latent representations), exploring this synergy remains understudied to-date.

In this work, we propose a joint analysis of unimodal models using MRI and WSI against multiple multimodal models by combining both exams (MRI and WSI) for brain tumor biomarkers classification (IDH1 and 1p/19q) following the 2021 WHO criteria. This problem is tackled as a multiclass problem and it can be divided into three different targets: (class 0) IDH1-wildtype (glioblastoma and diffuse astrocytic glioma with molecular features of glioblastoma), (class 1) IDH1-mutant and 1p/19q codeletion (oligodendroglioma), and (class 2) IDH1-mutant (lower grade astrocytoma) (Table C.1). Specifically, the main contributions of this work are: (i) the proposal of new fusion models including a context-aware model and (ii) the presentation of the important role of combing MRI with WSI for detecting IDH mutation and 1p/19q codeletion status for better glioma stratification. The code to reproduce this work is publicly available[1].

[1] https://github.com/tomealbuquerque/multimodal-glioma-biomarkers-detection

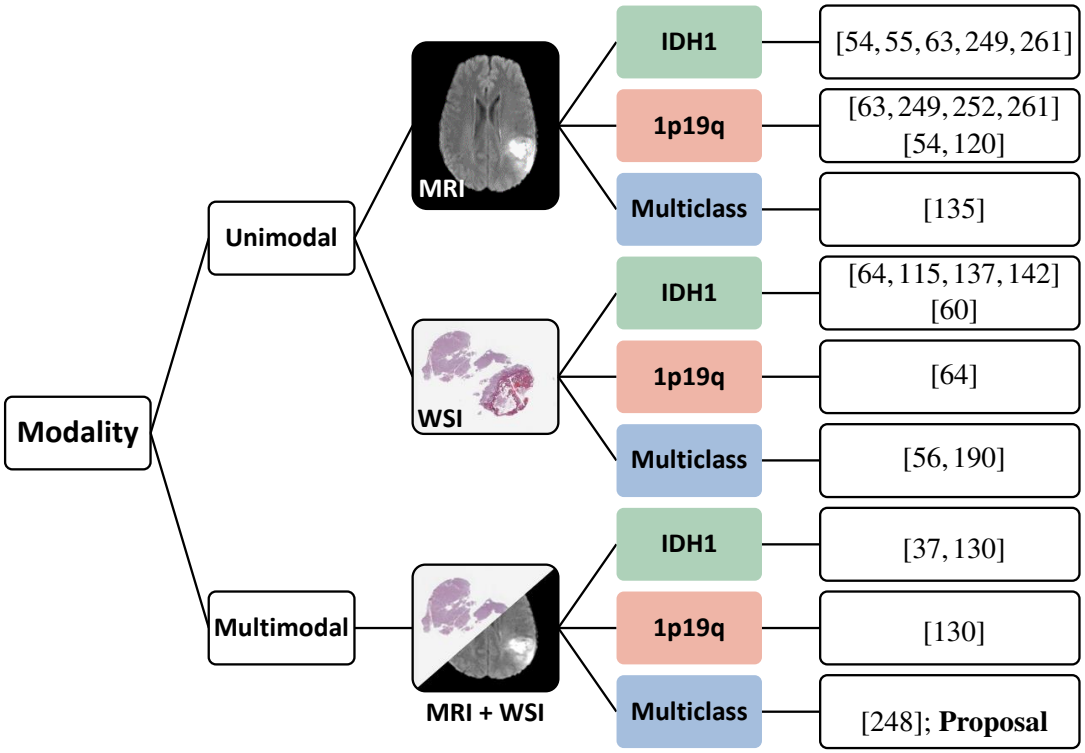


Figure C.2: Overview of the most relevant methods present in the literature for glioma biomarkers detection.

C.1.2 Proposed Methodology

This study is divided into two different approaches: unimodal and multimodal. In the unimodal part, which serves as the basis for comparison with the multimodal part, different models were tested for each modality (MRI and WSI). Regarding MRI, two different approaches were investigated, with the original ($255 \times 255 \times 255$ pixels (px)) MRI exams and also the segmented ROI blocks ($96 \times 96 \times 96$ px). Both methods were tested with all four MRI exams (T1, T1ce, T2, and FLAIR) with and without (standalone) combinations to determine the best method, using DenseNet121 as the backbone for the classification of the biomarkers. For WSI, we incorporated a weakly supervised deep learning model based on multiple instance learning (MIL), using only the reported diagnosis as a label per slide. Using ResNet34, in a binary classification problem for the training process of the embedder/encoder, the top 10 tiles per slide were selected based on the output score. Then, a multilayer perceptron model (MLP) was trained to aggregate the 10 tiles for the multiclass classification problem. Three different methods for relevant tiles selection were tested: top tiles containing the biomarkers (Top tiles), top tiles per class (Mix tiles), and the expected value ($E[X]$) per tile, to select the most relevant ones (Expected tiles), defined as:

$$E[X] = \sum_{k=1}^{K} x_k \times p(x_k). \tag{C.1}$$

where $x_1, x_2, ..., x_K$ are the possible outcomes of the tile X, $p(x_1), p(x_2), ..., p(x_K)$ are their corresponding probabilities, and K is the number of classes. For the multimodal part, four different models combining MRI with WSI were tested and proposed as represented in Figure C.3. The different multimodal methods reutilize the same backbone as previously described in the unimodal models for MRI and WSI respectively. However, the final model layers are modified to integrate the feature vectors from both modalities. The proposed multimodal fusion models are the following:

(a) **Cascade with fusion model** - This model is a two-stage classifier where the decision for 1p19q codeletion arises from the MRI branch. Moreover, a feature vector (penultimate linear layer) from the MRI branch is concatenated in the WSI branch through a middle fusion approach to regularize the training process for the IDH1 mutation classifier (Figure C.3 (a)).

(b) **Mid fusion model** - A mid-fusion approach is tested in this model using the same backbone from MRI and WSI unimodal branches. The last layer for classification was discarded and the two penultimate linear layers were saved as feature vectors from both modalities. The embedder weights were frozen and just the new MLP was trained (Figure C.3 (b)).

(c) **Late fusion model** - This model follows the same approach as the Mid fusion model, however, instead of concatenating the MRI feature vector in the middle layers of the MLP, they are concatenated in the late layers. Followed by a dilatation layer to learn more complex representations and finalize with a linear layer for the multi-class classification (Figure C.3 (c)).

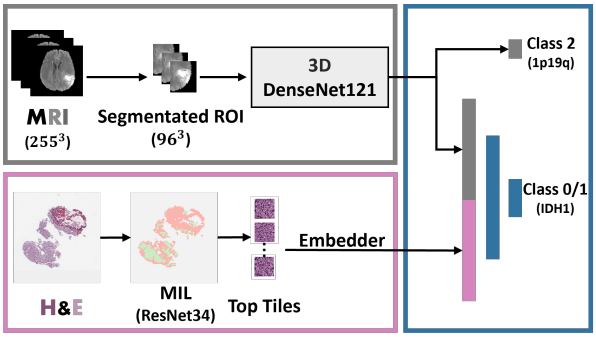
(d) **Mid context fusion model** - Based on the Mid context fusion model, an extra branch is added with larger tiles centered on the original tiles to give more context to the model concerning the surrounding areas. This model is based on the base routine of a pathologist who looks at a higher level at the microscope to have more information concerning the ROI. Thus, an extra embedder is trained using 2048×2048 tiles (Figure C.3 (d)).

C.1.3 Experimental Details

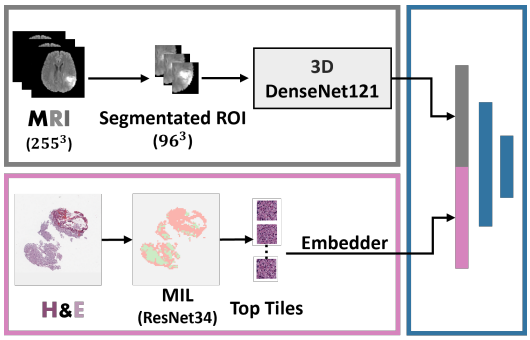
C.1.3.1 Dataset

The dataset used in this work was obtained from The Cancer Genome Atlas (TCGA), a publicly available database [2]. It includes 3D MRI images ($255 \times 255 \times 255$ px) from four modalities: T1, T1ce, T2, and FLAIR with segmentation blocks ($96 \times 96 \times 96$ px), digital pathology slides with hematoxylin and eosin staining (H&E), and clinical data (age and gender) from 187 patients. Table C.1 summarises the target encoding and frequency of samples per class.

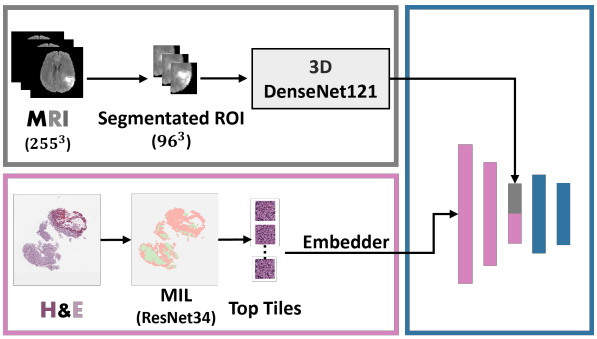
[2] https://www.cancerimagingarchive.net/



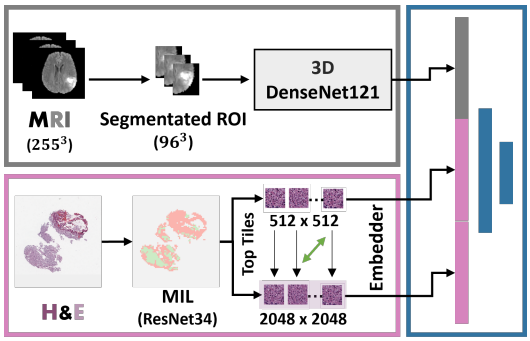
(a) Cascade w/ Fusion.



(b) Mid Fusion.



(c) Late Fusion.



(d) Mid context Fusion.

Figure C.3: Schematic representation of the models proposed in this study: (a) Cascade w/ Fusion (b) Mid Fusion (c) Late Fusion (d) Mid context Fusion.

Table C.1: Summary of TCGA dataset with targets encoding and frequency of samples per class.

	Biomarker		**Frequency**
class	IDH1 mutation	1p/19q codeletion	
0	-	-	102
1	●	-	58
2	●	●	27

C.1.3.2 Data Pre-processing

To preserve the class ratios, the dataset was partitioned into five distinct folds using stratified cross-validation. The results are the average and standard deviation of these 5 folds. For the Pathology deep learning embedder it was necessary to extract tiles with 512×512 px and 2048×2048 px dimensions from all WSI. Only tiles with more than 60% of tissue are used. Normalization was also performed to scale the pixel values to 0-1, which speeds up the training process and also helps the network in regularization. Data augmentation was used to cope with the limited amount of data. Each pathology image is subjected to a series of random transformations during training, including HEDJitter (random perturbation in the HED color space of the image), random elastic transformations, random color affine, random Gaussian blur, and also random image rotations. Since multi-channel augmentations for MRI images take more time and slow down the training process, only random image rotations were employed for MRI images.

C.1.3.3 Training Details

For the Pathology deep learning embedder, a grid search was first performed to find the best combination of hyperparameters, varying the image size (224×224, 552×552), batch size (128,256,384,512) and training epochs (25,50,100). For the MIL models, the best parameters found for the fine-tuned model were an image size of 552×552, a batch size of 384, and 50 training epochs. These hyperparameters were chosen with respect to the test subset of the first fold. ADAM was used in both modalities as the optimizer and started with a learning rate of 10^{-4} for all models. Cross-entropy (CE) was used as loss function. All the experiments are implemented using the PyTorch framework and the training process was conducted on a single Nvidia Tesla A100 (80GB) GPU.

C.1.3.4 Evaluation Metrics

Four different metrics were used to assess the performance of the models: accuracy (ACC), mean average error (MAE), F-score (F1), and the area under the curve (AUC). Accuracy is the most popular metric in classification problems [7] and can be defined as the proportion of cases correctly classified. MAE is calculated by dividing the sum of absolute errors by the sample size. Therefore, the further the predicted class is from the actual class, the more significant the error becomes. F1 is the weighted harmonic mean of the precision and recall by giving equal weight to precision and recall. An overall measure of performance across all potential classification thresholds is provided

by AUC, which can be understood as the probability that a random positive example will be ranked higher than a random negative example in the model.

C.1.4 Results and Discussion

The average performance for the 5-folds of the three different modalities unimodal-WSI, unimodal-MRI, and Multimodal (WSI+MRI) are presented in Table C.2. The unimodal-WSI MIL results include three different aggregation methods of the 10 relevant tiles (top tiles for biomarkers, mix tiles, and also expected tiles). For unimodal-MRI all four MRI exams (T1, T1ce, T2, and FLAIR) were tested with combinations and without combinations (standalone), and the best models are presented in Table C.2 for T1ce and also T1ce with FLAIR. The best models are shown in bold. Looking across the different unimodal models, regarding MRI, the best model was obtained using the pre-segmented blocks instead of the original MRI exam. The T1ce and also T1ce with FLAIR showed to be the best combination of MRI modalities when compared with T1 and T2. Concerning WSI, the best model was obtained using 10 expected tiles. This aggregation method will therefore be used for the multimodal models. Comparing the results of unimodal vs multimodal models, it is possible to infer that multimodal models surpass the unimodal counterparts in all metrics. Mid context Fusion model using T1ce and FLAIR in the MRI branch achieved the best results in two metrics with an ACC of 83.4% and an MAE of 0.224. Mid fusion model using also T1ce and FLAIR remains very competitive in several metrics and surpassed mid context fusion model in AUC, with 0.874. Comparing the ACC of the best-proposed model with the state-of-the-art [248], it is clear that the proposed model is above the state-of-the-art results with an ACC of 88.1% against 84.9% for the best fold. FigureC.4 shows the receiver operating characteristic (ROC) plot for the 3 best models based on the AUC metric, by comparing unimodal vs multimodal approaches. The previously mentioned ROC plot, emphasizes the use of multimodal approaches, showing a higher curve for the multimodal mid fusion model. The confusion matrices for all models are in the Supplementary Tables S1-S5.

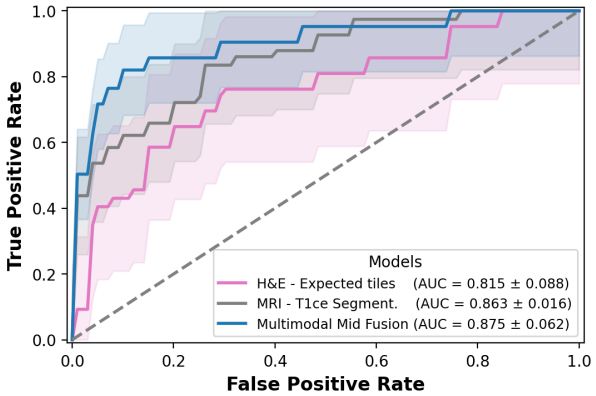
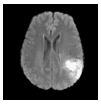


Figure C.4: ROC plot of the 5-folds cross-validation results for the best model per modality.

Table C.2: Results for all modalities (5-fold cross validation, CE loss) - TCGA dataset.

Modality		Model	Evaluation Metrics			
			Acc (%)	MAE	F1 (%)	AUC
	orig.	**T1ce**	65.818 ± 5.430	0.449 ± 0.112	60.260 ± 8.158	0.662 ± 0.048
		T1ce + FLAIR	68.478 ± 4.542	0.385 ± 0.066	63.770 ± 7.090	0.713 ± 0.038
	Seg.	**T1ce**	80.214 ± 3.609	0.283 ± 0.048	75.570 ± 6.545	0.863 ± 0.015
		T1ce + FLAIR	73.244 ± 2.717	0.385 ± 0.081	64.290 ± 9.381	0.840 ± 0.037
	MIL	**Top tiles**	73.243 ± 5.963	0.321 ± 0.084	70.649 ± 6.822	0.808 ± 0.101
		Mix tiles	73.243 ± 5.963	0.321 ± 0.084	70.649 ± 6.822	0.808 ± 0.101
		Expected tiles	74.851 ± 3.754	0.310 ± 0.054	71.518 ± 3.462	0.815 ± 0.088
	T1ce	**Cascade w/ Fusion**	77.568 ± 6.269	0.288 ± 0.088	77.970 ± 6.221	0.833 ± 0.068
		Mid Fusion	80.199 ± 2.239	0.262 ± 0.039	78.129 ± 4.765	0.839 ± 0.059
		Late Fusion	79.147 ± 1.942	0.251 ± 0.034	76.337 ± 3.660	0.842 ± 0.058
		Mid Context Fusion	80.740 ± 5.482	0.246 ± 0.074	80.034 ± 5.524	0.866 ± 0.056
	T1ce+F	**Cascade w/ Fusion**	81.707 ± 10.190	0.237 ± 0.144	**82.304 ± 9.814**	0.860 ± 0.068
		Mid Fusion	82.319 ± 3.791	0.246 ± 0.068	81.893 ± 3.627	**0.874 ± 0.061**
		Late Fusion	80.725 ± 3.216	0.251 ± 0.043	78.221 ± 5.039	0.849 ± 0.056
		Mid context Fusion	**83.414 ± 4.658**	**0.224 ± 0.071**	82.094 ± 5.083	0.858 ± 0.041

C.1.5 Summary and Conclusions

This work introduced and proposed novel multimodal deep learning methods for glioma biomarkers detection. The multimodal networks achieved the best results in several metrics, highlighting the favorable performance gains of using multimodal deep learning methods to tackle this problem. The most effective MRI modalities for the identification of both biomarkers were T1ce and FLAIR. Our proposed multimodal mid context fusion model proved to be very competitive against the other models, however, more experiments varying the size of context tiles should be performed for a better understanding of the extra context information impact. This work is a step toward more efficient use of MRI and histopathological data together and is particularly valuable in encouraging future research to lower the hurdles for point-of-care facilities without rapid access to a molecular genetics laboratory. For further work, we aim to validate the models with external data, however, for complete reproducibility, we used publicly available data and also turn public our code. By combining WSI and MRI into a unified latent space, multimodal solutions may improve the diagnosis with the addition of more training data.