

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département électronique

End-of-study project dissertation
for obtaining the State Engineer's degree in Electronics

Machine Learning and Deep Learning methods for cancer prediction and responses to its treatment

Wissal ACHOUR & Ferial BOUDJATIT

Under the supervision of Mrs. Nour El Houda BENALIA MCA. ENP, Alger

Presented and publicly defended on 02/07/2024 in front of the members of the jury :

President	M. Hicham BOUSBIA-SALAH	Prof.	ENP, Alger
Examiner	Mrs. Rachida TOUHAMI	Prof.	ENP, Alger
Supervisor	Mrs. Nour El Houda BENALIA	MCA.	ENP, Alger

ENP 2024

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Alger, Algérie.
www.enp.edu.dz

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département électronique

End-of-study project dissertation
for obtaining the State Engineer's degree in Electronics

Machine Learning and Deep Learning methods for cancer prediction and responses to its treatment

Wissal ACHOUR & Feriel BOUDJATIT

Under the supervision of Mrs. Nour El Houda BENALIA MCA. ENP, Alger

Presented and publicly defended on 02/07/2024 in front of the members of the jury :

President	M. Hicham BOUSBIA-SALAH	Prof.	ENP, Alger
Examiner	Mrs. Rachida TOUHAMI	Prof.	ENP, Alger
Supervisor	Mrs. Nour El Houda BENALIA	MCA.	ENP, Alger

ENP 2024

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Alger, Algérie.
www.enp.edu.dz

République Algérienne Démocratique et Populaire
الجمهورية الجزائرية الديمقراطية الشعبية
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
وزارة التعليم العالي و البحث العلمي
École Nationale Polytechnique



المدرسة الوطنية المتعددة التقنيات
Ecole Nationale Polytechnique

Département électronique

Mémoire de projet de fin d'études
Pour l'obtention du diplôme d'Ingénieur d'État en Électronique

Machine Learning and Deep Learning methods for cancer prediction and responses to its treatment

Wissal ACHOUR & Ferial BOUDJATIT

Sous la direction de Mme. Nour El Houda BENALIA MCA. ENP, Alger

Présenté et soutenu publiquement le 02/07/2024 auprès des membres du jury :

Président	M. Hicham BOUSBIA-SALAH	Prof. ENP, Alger
Examinatrice	Mme. Rachida TOUHAMI	Prof. ENP, Alger
Promotrice	Mme. Nour El Houda BENALIA	MCA. ENP, Alger

ENP 2024

10, Avenue des Frères Oudek, Hassen Badi, BP. 182, 16200 El Harrach, Alger, Algérie.
www.enp.edu.dz

ملخص

لا يزال السرطان تحديًا صحيًا عالميًا كبيرًا، يؤثر على الأفراد من جميع الأعمار. يعد الكشف المبكر والعلاج الشخصي أمرين حاسمين، حيث يحسنان بشكل كبير من توقعات سير المرض ونتائج العلاج. أظهرت التطورات الحديثة في أساليب التعلم الآلي (ML) والتعلم العميق (DL) وعودًا كبيرة في تعزيز الكشف عن السرطان من خلال تحليل الصور الطبية وتوقع استجابات الأدوية المحددة لكل مريض. تركز هذه الدراسة على تصنيف الأورام الدبقية، وهي نوع من أورام الدماغ، إلى أورام دبقية منخفضة الدرجة (LGG) وأورام دبقية عالية الدرجة (HGG) من خلال اقتراح نموذج تصنيف شامل للأورام يعمل على شرائح التصوير بالرنين المغناطيسي (MRI). بالإضافة إلى ذلك، تستكشف الدراسة تطوير نموذج تنبؤي لاستجابة الأدوية لعلاج السرطان من خلال الاستفادة من بيانات الأدوية الجزيئية ومعلومات خطوط الخلايا السريرية.

كلمات مفتاحية : السرطان، ورم الدماغ، استجابة أدوية السرطان، التصوير بالرنين المغناطيسي، التعلم الآلي، التعلم العميق، التصنيف.

Résumé

Le cancer reste un problème de santé mondial important, qui touche des personnes de tout âge. La détection précoce et le traitement personnalisé sont essentiels car ils améliorent considérablement le pronostic et les résultats du traitement. Les travaux récents dans les méthodes d'apprentissage automatique (ML) et d'apprentissage profond (DL) se sont révélés très prometteurs pour améliorer la détection du cancer par l'analyse d'images médicales et pour prédire les réponses aux médicaments spécifiques aux patients. Cette étude se concentre sur la classification des gliomes, un type de tumeur cérébrale, en gliomes de bas grade (LGG) et gliomes de haut grade (HGG) en proposant un modèle de classification des tumeurs de bout en bout qui fonctionne sur des coupes d'IRM. En outre, elle explore le développement d'un modèle prédictif de la réponse aux médicaments anticancéreux en exploitant les données moléculaires des médicaments et les informations sur les lignées cellulaires cliniques.

Mots clés : Cancer, Tumeur cérébrale, Réponse aux médicaments anticancéreux, IRM, Apprentissage automatique, Apprentissage profond, Classification.

Abstract

Cancer remains a significant global health challenge, affecting individuals of all ages. Early detection and personalized treatment are crucial as they significantly improve prognosis and treatment outcomes. Recent advancements in machine learning (ML) and deep learning (DL) methods, have shown considerable promise in enhancing cancer detection through medical image analysis and predicting patient-specific drug responses. This study focuses on the classification of Gliomas, a type of brain tumor, into Low-Grade Gliomas (LGG) and High-Grade Gliomas (HGG) by proposing an end-to-end tumor grading model that performs on MRI slices. Additionally, it explores the development of a predictive model for cancer drug response by leveraging drug molecular data and clinical cell line information.

Keywords : Cancer, Brain tumor, Cancer Drug Response, MRI, Machine learning, Deep learning, Classification.

Dedication

It is with profound gratitude that I dedicate this final project to all the people who contributed to its realization. The path to this moment has been challenging, but with your support, love, and encouragement, I have navigated each step with determination and perseverance.

I extend my heartfelt gratitude to my parents, whose unwavering support and encouragement have been my greatest source of strength throughout my academic journey. Your boundless love, patience, and belief in my abilities have empowered me to pursue my dreams and overcome every obstacle. From the countless hours of guidance to the sacrifices you have made, your dedication has laid the foundation for my success. I am deeply thankful for your constant presence in my life and for the values you have instilled in me. This achievement would not have been possible without you.

To my brothers Zaki and Youness, to my dear aunt whom I consider my second mother and to my uncles, thank you for your words of encouragement, kindness, and presence. And to my dear cousins Doudou, Khawla, Sheyma, and Kakou, whom I consider my brothers and sisters, thank you for enriching my life in such a wonderful way.

I want to express my deep gratitude to the friends I have met along this journey. You have profoundly impacted my understanding of trust. Thank you for being a part of my life—Rana, Taqwa, Ferial, Bouchra, Sabrina, and Zahra—your presence has brought both laughter and tears of joy into my life.

I would also like to extend my deepest gratitude to my friends and my wonderful cohort, "ELN complot". You have made these past years an unforgettable period. To the new people I got to know along my journey at ENP, from mem-

bers of VIC, IEEE SB to Vniverse, you created a healthy environment to unleash creativity. Thank you for being there with us through happy moments and challenging times. Your presence has made difficult times easier to bear. Thank you for everything you have done.

To all those dear to me, to each of you.”

- Wissal

Dedication

This work is dedicated, above all, to the memory of my beloved father, whose presence will forever live in our hearts.

And to my mother, the light of my heart and the source of my happiness. She has been both parents in one, a strong woman who has imparted her strength to me. Thank you for your unwavering gentleness that is uniquely yours. Thank you for your constant support, sacrifices and encouragement. If I am anything today, it's because you believed in me every step of the way.

To my lovely sisters, Amira and Dounia, who have been two important pillars in my life, and to my brother Souhil, for his generosity, kindness, and the long talks we've shared along the way. I might add a special dedication to my niece Randa and nephew Zayd, as they are rays of sunshine that make my life better. Thank you all for bringing warmth into my life.

After my family, I would like to thank the friends I have made along my journey at ENP. Ranging from the members of IEEE, especially my committee members, and VIC to my promotion, a special thanks to Twin Tower and Mehdi. Befriending each one of you have considerably contributed to my growth and made my three years at this school unforgettable.

I would like to make a special dedication to my friends of heart, Mika, Mehdi, Wissal, Sabrina and Bouchra. Thank you for the love and support, for your patience, and for both the silly and profound conversations we've had. You all have immensely impacted my life, and I am more than grateful to have you.

Finally, I am honored to dedicate this work to all cancer patients and their families who face the challenges of this disease with courage and resilience every day. I hope that the scientific progress to which we dedicate our time bring some ease and equality regarding the daily privileges we often take for granted.

- Ferial

Acknowledgements

Above all, we thank God, the Almighty, for having given us the courage, patience, willpower and strength to face all the difficulties and obstacles that have stood in our way throughout this work.

First and foremost, we would like to express our deepest gratitude and thanks to our project supervisor, Mrs. BENALIA Nour El Houda, for her constant guidance and support. Her unwavering kindness, gentleness, and presence have been invaluable, and we are deeply appreciative of the knowledge she has always shared with us.

We extend our sincere thanks to the President of the jury, Mr. BOUSBIA Hicham, Professor at the Ecole Nationale Polytechnique, and the Examiner, Mrs. TOUHAMI Rachida, also a Professor at the Ecole Nationale Polytechnique, for agreeing to be members of the reading committee. Their constructive analysis of this work, as well as their time, teaching, and support over the past three years, are greatly appreciated.

We are profoundly grateful to all our professors at the Ecole Nationale Polytechnique, who have guided us throughout our five years of study.

Finally, we thank everyone who has helped us, directly or indirectly, throughout our academic journey.

Wissal and Ferial.

Contents

List of Tables

List of Figures

List of Abbreviations

General Introduction **17**

Part 1 Early detection of cancerous tumors **19**

1	Background knowledge and related research	21
1.1	Introduction	22
1.2	State of the art	22
1.2.1	Search and selection criteria of the literature	22
1.2.2	Literature survey	23
1.2.2.1	Brain tumor detection survey	23
1.2.2.2	Lung cancer detection survey	25
1.2.2.3	Breast cancer detection survey	28
1.2.2.4	Skin cancer detection survey	31
1.2.2.5	Leukemia cancer detection survey	34
1.2.2.6	Other cancer types	37
1.2.3	Overall key findings	40
1.3	Brain tumors: Impact and treatment considerations	40
1.3.1	Clinical background	44
1.3.2	Classification	45
1.3.3	Imaging modalities	46
1.3.3.1	Magnetic Resonance Imaging (MRI)	47
1.4	Fundamental tools	49
1.4.1	Dataset	49
1.4.2	Preprocessing tools	51
1.4.2.1	Enhancing the tumor region	52
1.4.2.2	Background removal	53
1.4.2.3	Resizing	54
1.4.3	Feature engineering	54
1.4.3.1	Feature extraction	55
1.4.3.1.1	Radiomics	55
1.4.3.1.2	Grey-Level Co-occurrence Matrix (GLCM) features	57

1.4.3.2	Feature selection	58
1.4.3.3	Feature reduction	58
1.4.4	Classifiers	59
1.4.5	Machine Learning models	59
1.4.5.1	Support Vector Machine (SVM)	59
1.4.5.2	Random Forest (RF)	61
1.4.6	Deep Learning models	62
1.4.6.1	3D Convolutional Neural Networks (3D-CNNs)	62
1.4.6.2	Autoencoders	64
1.4.7	Performance Evaluation Metrics	65
1.5	Conclusion	68
2	Contributions	69
2.1	Introduction	70
2.2	Assessment of different models for cancer tumor grading	70
2.2.1	Radiomics features with Machine Learning classifier	70
2.2.2	AutoEncoder feature extractor with Machine Learning classifier	71
2.2.3	Proposed 3D Multi-Scale Convolutional Neural Networks	72
2.3	Experiments and results	77
2.3.1	Radiomics features with Machine Learning classifier	77
2.3.2	AutoEncoder feature extractor with Machine Learning classifier	79
2.3.3	Our proposed 3D Multi-Scale Convolutional Neural Networks	80
2.3.4	Performance comparison of our proposed model and the Machine Learning methods	81
2.4	Conclusion	82
Part 2	Cancer drug response prediction	84
1	Background knowledge and related research	86
1.1	Introduction	87
1.2	State of the art	87
1.2.1	Search and selection criteria of the literature	87
1.2.2	Literature survey	88
1.3	Clinical background	96
1.3.1	Simplified Molecular-Input Line-Entry System (SMILES)	97
1.3.2	Multi-omics data	98
1.3.2.1	Genomics (Genomic mutations)	99
1.3.2.2	Transcriptomics (Gene expressions)	100
1.3.2.3	Epigenomics (DNA methylation)	101
1.4	Fundamental tools	102
1.4.1	Dataset	103
1.4.2	Deep Learning models	104
1.4.2.1	Transformers	104
1.4.2.2	Dense Neural Networks	108
1.4.2.3	2D Convolutional Neural Networks	108
1.4.3	Performance Evaluation Metrics	109

1.4.3.1	Root Mean Square Error (RMSE)	109
1.4.3.2	Pearson Correlation Coefficient (PCC)	110
1.4.3.3	Spearman's Correlation Coefficient (SCC)	110
1.5	Conclusion	110
2	Contributions	111
2.1	Introduction	112
2.2	Proposed end-to-end Deep Learning model	112
2.2.1	Data preparation	113
2.2.2	Framework	114
2.2.2.1	Transformer's encoder for drug data processing	115
2.2.2.2	Hybrid Neural Network for Multi-omics data processing	116
2.2.2.3	Predictor for combining features and making predictions	117
2.3	Experiments and results	118
2.3.1	Experimental conditions	119
2.3.2	Regression performance evaluation	119
2.3.3	Case prediction across different cancer types	120
2.3.4	Case prediction across different drugs	122
2.4	Conclusion	125
	General Conclusion	127
	Bibliography	129
	Webography	141

confidentiel