COLLEGE DOCTORAL PAYS DE LA LOIRE
SCIENCES DE L'INGENIERIE ET DES SYSTEMES

CENTRALE NANTES

MÉMOIRE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES

ÉCOLE DOCTORALE N° 602
Sciences de l'Ingénierie et des Systèmes
Spécialité : *Signal, Image, Vision*

Par
Oriane THIERY

Multimodal graph learning frameworks for interpretable high-risk diffuse large B-cell lymphoma patient identification

Projet de recherche doctoral présenté et soutenu à l'Ecole Centrale de Nantes, le 2 septembre 2025
Unité de recherche : UMR 6004, Laboratoire des Sciences du Numérique de Nantes (LS2N)

Rapporteurs avant soutenance :

Pierre-Henri CONZE — Maître de conférences, HDR, IMT Brest
Carole LARTIZIEN — Directrice de recherche-CNRS, Institut National des Sciences Appliquées, Villeurbanne

Composition du Jury :

Président :	Harold MOUCHÈRE	Professeur des universités, Nantes Université
Examinateurs :	Pierre-Henri CONZE	Maître de conférences, HDR, IMT Brest
	Mathieu HATT	Directeur de recherche, Université de Bretagne Occidentale, Brest
	Carole LARTIZIEN	Directrice de recherche-CNRS, Institut National des Sciences Appliquées, Villeurbanne
	Dorina THANOU	Senior lecturer, Ecole Polytechnique Fédérale (EPF) Lausanne, Suisse
Directrice de recherches doctorales :	Diana MATEUS	Professeure des universités, Ecole Centrale de Nantes
Co-dir. de recherches doctorales :	Thomas CARLIER	Dr, Physicien médical HDR, CHU de Nantes
Co-enc. de recherches doctorales :	Mira RIZKALLAH	Maître de conférences, Ecole Centrale de Nantes

Acknowledgements

Le document que vous avez devant vous résume le travail de trois années de thèse à Centrale Nantes, avec son lot de hauts et de bas, mais surtout son lot de pauses café, de découverte de nouvelles villes et pays, de belles rencontres, de projets de boulangerie/agence de voyage/hôtel/bar à jeu (vous référer à Maria pour le management et à Younès pour le reste), de repas du monde/chandeleurs/galettes des rois/PPP/n'importe quelle occasion SIMS de manger et de partager un moment ensemble, et de questions existentielles comme les doctorant.e.s savent si bien en poser. Comme vous l'aurez compris, ce travail a été loin d'être un exercice solitaire, et j'ai donc quelques personnes à remercier pour avoir permis d'arriver à ce que vous pourrez lire dans la suite de ce manuscrit.

*The document you have in front of you summarizes three years of PhD work at Centrale Nantes, with its ups and downs, but above all with its coffee breaks, discoveries of new cities and countries, wonderful encounters, projects of a bakery/travel agency/hotel/game bar (refer to Maria for management and Younès for the rest), world cuisine meals/*chandeleur/galettes des rois*/PPP/any SIMS occasion to eat and share a moment together, and the existential questions that doctoral students are so good at asking. As you can see, this work was far from a solitary exercise, and I therefore have a few people to thank for making it possible to produce what you will read in the rest of this manuscript.*

Avant toute chose, je tiens à remercier profondément mes encadrant.e.s, Diana, Mira, et Thomas. J'ai conscience de la chance considérable que j'ai eu d'être si bien entourée pour un tel projet, et ce sur tous les plans : par votre apport technique et clinique, vos explications de concepts que je ne maîtrisais pas, vos idées qui m'ont guidées vers de nouvelles voies, vos questions qui ont considérablement enrichi ce projet ; par votre apport académique, sur ce qui est attendu de nous à chaque étape de la thèse et pour la suite ; par votre présence et votre réactivité, malgré vos emplois du temps de ministres, à chaque fois que j'avais besoin de votre aide et pour suivre ce projet avec un investissement considérable tout au long des trois ans ; par votre soutien constant ; et au-delà de tout ça, par votre humanité, car vous êtes des personnes formidables et

j'ai beaucoup apprécié les moments que nous avons passés ensemble. C'est grâce à vous que cette thèse en est arrivée jusqu'ici, et je vous suis extrêmement reconnaissante pour ces trois années : merci pour tout !

First and foremost, I would like to express my deepest gratitude to my supervisors, Diana, Mira, and Thomas. I am aware of how fortunate I have been to have such an excellent team for this project, in every respect: for your technical and clinical input, your explanations of concepts I did not understand, your ideas that guided me toward new paths, and your questions that greatly enriched this project; through your academic input on what is expected of us at each stage of the thesis and beyond; through your presence and responsiveness, despite your busy schedules, whenever I needed your help, and for following this project with considerable investment throughout the three years; for your constant support; and above all, for your humanity, since you are wonderful people and I greatly appreciated the moments we spent together. It is thanks to you that this thesis has come this far, and I am extremely grateful to you for these three years: thank you for everything!

Je voudrais également remercier les membres du jury de cette thèse (Carole Lartizien et Pierre-Henri Conze, qui l'ont rapportée, ainsi que Mathieu Hatt, Dorina Thanou et Harold Mouchère) pour le temps que vous lui avez consacré et pour vos très intéressantes questions sur ce travail.

I would also like to thank the members of the jury (Carole Lartizien and Pierre-Henri Conze, who reported on this thesis, as well as Mathieu Hatt, Dorina Thanou, and Harold Mouchère) for the time you have devoted to our work and for your very interesting questions.

Merci également à Laurence, Khrystyna, Émily, Sabine, Patrick, Robert, Philippe, Peggy, Edith et toutes les personnes qui s'investissent pour que nos thèses et notre travail se passent au mieux : j'ai (et cette thèse avec moi) une dette considérable envers votre travail, donc merci pour tout votre soutien au quotidien !

Many thanks also to Laurence, Khrystyna, Émily, Sabine, Patrick, Robert, Philippe, Peggy, Edith, and everyone else who has worked so hard to ensure that our PhDs and our work go as smoothly as possible: I (and this thesis with me) owe a considerable debt to your work, so thank you for all your daily support!

Je suis également très reconnaissante à Martin, Hakima, Oli, Mariem, Mahdi et toute l'équipe MEDomics pour leur accueil à Sherbrooke et leur excellente compagnie pendant les deux mois où nous avons travaillé ensemble. Au plaisir de vous revoir si vous passez par la France à l'occasion !

I am also very grateful to Martin, Hakima, Oli, Mariem, Mahdi, and the entire MEDomics team for their warm welcome in Sherbrooke and for their excellent company

during the two months we worked together. I look forward to seeing you again if you ever visit France!

Je souhaiterais également remercier sincèrement les étudiant.e.s avec qui j'ai eu la chance de pouvoir travailler au cours de cette thèse, que ça soit Arthur et Damien, Anthonin et Bachar, ou Maria-Clara et Ilyes. J'ai pris beaucoup de plaisir à échanger avec vous et à vous voir apprendre et progresser sur vos projets respectifs, et même si ces projets n'ont pas mené à des contributions présentées dans cette thèse, sachez qu'ils ont été très bien menés et qu'ils pourront grâce à ça être repris pour aboutir aux résultats espérés. Merci pour votre investissement et votre sympathie. Je vous souhaite tout le meilleur dans la suite de vos projets !

I would also like to sincerely thank the students with whom I had the opportunity to work during this PhD, including Arthur and Damien, Anthonin and Bachar, and Maria-Clara and Ilyes. I really enjoyed discussing with you and seeing you learn and progress on your respective projects, and even if these projects did not lead to contributions presented in this thesis, please know that they were very well conducted and that, thanks to this, they can be taken up again to achieve the desired results. Thank you for your dedication and your enthusiasm. I wish you all the best in your future projects!

Merci également à mes enseignants et enseignantes de Polytech, qui m'ont notamment montré cette voie et m'y ont encouragé. Cette thèse n'aurait pas eu lieu sans votre soutien dévoué à vos pauvres étudiant.e.s qui ne comprennent pas grand chose au système académique, donc un grand merci pour votre aide et, plus globalement, votre investissement et nos sympathiques échanges pendant toutes ces années passées à l'école !

Many thanks also to my teachers at Polytech, who showed me this path and encouraged me to follow it. This thesis would not have been possible without your devoted support for your poor students who don't understand much about the academic system, so a big thank you for your help and, more generally, your commitment and our great discussions during all these years in Polytech!

Un grand merci également à l'équipe du LIP6, Andrea et Isabelle, mais aussi Garance, Leandro, Vic et tous les membres des équipes LFI et SYEL avec qui j'ai réalisé mon premier grand pas vers la recherche, pour la super expérience que ça a été.

A big thank you also to the LIP6 team, Andrea and Isabelle, as well as Garance, Leandro, Vic, and all the members of the LFI and SYEL teams with whom I took my first big step toward research, for the great experience it has been.

Bien évidemment, je ne pourrais manquer de remercier les incroyables ami.e.s qui m'entourent plus ou moins au quotidien. Que ça soit celles avec qui on se suit depuis le primaire, collège ou lycée, Ophélie, Emma, Audrey, Aurélie, Clara, Léa et Typhaine ;

mes ami.e.s de prépa, Clemzer et Lisou, qui m'ont fait découvrir de nouveaux horizons, Guigui, qui poursuit sa propre thèse et en arrivera bientôt à bout, et toute l'équipe ; Manon et Jules et nos soirées jeux ; les collègues qui sont parfois devenus ami.e.s, et les si nombreux souvenirs qu'on a ensemble qu'il faudrait une thèse à part entière pour les raconter ; et plus largement, la formidable équipe SIMS, sur laquelle je suis extrêmement chanceuse d'être tombée. Chacun et chacune d'entre vous a du sens pour moi, et je vous remercie d'avoir été à mes côtés toutes ces années.

Of course, I couldn't fail to thank the incredible friends who surround me more or less every day. Whether it's those I've known since elementary school, middle school, or high school -Ophélie, Emma, Audrey, Aurélie, Clara, Léa, and Typhaine; my prep school friends Clemzer and Lisou, who opened up new horizons for me; Guigui, who is working on her own PhD and will soon finish it, and the whole team; Manon and Jules and our game nights; my colleagues who have sometimes become friends, and the countless memories we have together that would require a thesis in itself to recount; and more broadly, the wonderful SIMS team, which I am extremely lucky to have come across. Each and every one of you means a lot to me, and I thank you for being by my side all these years.

Mes derniers mots ici seront pour ma famille : mes parents et mes soeurs exceptionnelles bien évidemment, mais aussi mes cousins et cousines, oncles et tantes, grands-parents et autres liens familiaux qui ne reposent pas forcément sur le sang (comme dirait Ninouille, « En vrai, tu sais »). Je voudrais profiter de cette occasion pour vous remercier pour votre amour incommensurable, votre présence constante et les personnes formidables que vous êtes. Vous êtes la raison pour laquelle j'ai grandi ainsi, et je suis infiniment reconnaissante de la chance que j'ai d'avoir une famille aussi exceptionnelle. Je vous aime.

*My final words here will be for my family: my parents and my exceptional sisters, of course, but also my cousins, uncles, aunts, grandparents, and other family members who are not necessarily related by blood (as Ninouille would say, "*En vrai, tu sais*"). I would like to take this opportunity to thank you for your immeasurable love, your constant presence, and the wonderful people you are. You are the reason I grew up the way I did, and I am infinitely grateful for the privilege of having such an exceptional family. I love you.*

Ces remerciements n'ont plus grand chose à voir avec la thèse que vous allez (ou non) lire, mais si vous êtes arrivés jusqu'ici c'est que ça vous intéresse malgré tout : donc je vais finir de profiter de cette occasion de remercier tous ceux qui comptent pour moi pour avoir un mot plus global. Dans l'ensemble, je suis extrêmement reconnaissante d'être si bien entourée, que ça soit par les personnes citées ici ou par toutes les autres

qui ont croisé mon chemin, que j'ai appris à connaître, qui m'ont soutenue ou avec qui j'ai partagé un moment (voire tout ça à la fois pour certaines). Merci à chaque personne qui a laissé une trace dans ma vie, quelle qu'en soit la durée. Je vous souhaite à tous et toutes le meilleur dans vos vies à venir, et au plaisir de se revoir à l'occasion !

These acknowledgments have now little to do with the thesis you are about to read (or not), but if you have made it this far, it means you are interested nonetheless: so I will take this opportunity to thank those who matter to me with a few more general words. Overall, I am extremely grateful to be surrounded by such wonderful people, whether they are the ones mentioned here or all the others who have crossed my path, whom I have gotten to know, who have supported me, or with whom I have shared a moment (or all of the above). Thank you to every person who has left a trace on my life, no matter how long it lasted. I wish you all the best in your lives, and I look forward to seeing you again someday!

Contents

Résumé en français x

List of Figures xiv

List of Tables xvi

Abbreviations xvii

Notations xix

I Introduction and context 1

1 Introduction 2

2 Clinical context 7

3 Theoretical background on survival analysis 11

3.1 Basics: censoring, survival function and hazard rate 11

3.2 Survival analysis methods 13

3.3 Survival prediction evaluation metrics 19

4 State of the art 22

4.1 DLBCL analysis 22

4.2 Multimodal fusion 23

4.3 Graphs for medical applications 24

II Lesion graph for a better use of DLBCL PET images 31

5.1 Introduction 33

5.2 Method 33

5.3 Experimental validation 37

5.4 Discussion 40

III Interpretable multimodal fusion of clinical and imaging information 43

6.1 Introduction . . . 45
6.2 Method . . . 46
6.3 Experimental validation . . . 51
6.4 Discussion . . . 65

IV Population-graph for post-hoc correction of survival predictions for improved risk stratification 67

7.1 Introduction . . . 69
7.2 Method . . . 70
7.3 Experimental validation . . . 81
7.4 Discussion . . . 87

V Conclusion and perspectives 89

Supplementary material 95

A PET imaging for DLBCL 95

B GAINED features 97

B.1 Clinical features . . . 97
B.2 Radiomics . . . 98
B.3 GAINED radiomic features . . . 98

C Hyperparameter sets searched 100

D Multimodal model: additional information 101

D.1 Model structure when the second multimodal block is used . . . 101
D.2 Computation of the number of parameters of a given configuration of the model . . . 102

Bibliography 104

Résumé en français

Contexte, motivations et objectifs

Ce document présente le travail réalisé dans le cadre de ma thèse portant sur l'apprentissage basé sur les graphes, multimodal et interprétable pour l'identification des patients atteints de lymphome B diffus à grandes cellules (DLBCL) à haut risque.

Plus précisément, ce travail se concentre sur un type spécifique de lymphome, le DLBCL. Certains patients atteints de cette pathologie ne répondent pas à la première ligne de traitement. Identifier dès le moment du diagnostic ces patients à haut risque peut permettre aux médecins de personnaliser leur traitement, et ainsi de potentiellement améliorer leurs chances de survie tout en diminuant à l'inverse la toxicité des traitements donnés aux patients qui ne sont pas dans cette situation.

Actuellement, le diagnostic et le suivi des patients atteints de DLBCL utilisent des biomarqueurs cliniques et des images ^{18}F-Fluorodeoxyglucose (FDG)-TEP/TDM (TEP : Tomographie par Émission de Positons ; TDM : Tomodensitométrie) [Tabaa et al. 2021]. S'il a été prouvé que les biomarqueurs cliniques sont pronostiques pour l'identification des patients à haut risque, les images TEP sont plus complexes à utiliser dans ce but. Des indicateurs semi-quantitatifs à extraire de ces images ont montré un très fort potentiel pour cette identification [Schöder et al. 2016], même s'ils ne sont pas encore utilisés en routine clinique dans ce but. Cependant, les méthodes visant à les utiliser pour de l'identification de patients à haut risque ne considèrent pas comment les combiner avec d'autres types de modalités, et se limitent à un nombre restreint d'informations extraites des riches images TEP. À l'inverse, des méthodes automatiques ont été développées pour apprendre à tirer des informations plus complexes de ces images ; mais celles-ci ne prennent pas en compte explicitement la multiplicité des lésions des patients sur l'image, ni leur dispersion (ou uniquement via des mesures simplifiées comme la plus grande distance entre deux lésions). De plus, il n'existe que peu de méthodes intégrant à la fois de l'information clinique et image comme le font les médecins ; et celles qui le font n'utilisent que des processus de fusion basiques, peu adaptés à la grande différence (que ce soit de dimensionnalité ou de force pronostique) entre ces deux modalités. Ces observations soulèvent les deux premières questions traitées dans cette thèse :

I- Comment utiliser au mieux les images TEP pour l'identification des patients atteints de DLBCL à haut risque ? Et plus précisément, quelles informations devraient être extraites de l'image, et comment prendre en compte le nombre variable de lésions des patients ainsi que leur distribution spatiale ?

II- Comment fusionner efficacement les images TEP avec les biomarqueurs cliniques pour améliorer les performances d'identification des patients atteints de DLBCL à haut risque ?

De plus, ces méthodes n'utilisent encore que l'information de chaque patient individuellement pour faire une prédiction, sans tirer profit de l'information venant d'autres individus ayant un profil semblable. Cependant, les approches exploitant explicitement la connaissance des similarités et des différences entre les patients se sont montrées capables d'améliorer la performance de prédicteurs individuels. En conséquence, nous traitons dans ce travail une dernière question :

III- Comment tirer profit de la similarité inter-patients pour améliorer l'identification des patients à haut risque ?

Résumé des contributions

Cette thèse est structurée en cinq parties.

La première partie présente une vision globale du contexte de ce travail. Ceci comprend des explications cliniques sur le DLBCL et la façon dont les médecins l'abordent aujourd'hui, ainsi que des explications plutôt techniques sur l'analyse de survie, la fusion multimodale et les approches basées sur les graphes.

Les trois parties suivantes présentent les travaux réalisés dans le cadre de cette thèse et les contributions associées. Plus précisément, ces chapitres sont structurés de la façon suivante :

- **Chapitre II :** Ce chapitre traite de la question ***I*** de ce travail, et explore donc la meilleure façon de tirer profit des images TEP pour l'identification de patients atteints de DLBCL à haut risque. Pour ceci, nous comparons cinq ensembles de ca-ractéristiques extraites de ces images. Ces ensembles de caractéristiques reposent sur différentes méthodes de segmentation des lésions ainsi que sur des ensembles variables de radiomiques extraites. Les cinq groupes de caractéristiques ainsi calculés sont utilisés en entrée de plusieurs modèles, qui ont également des façons distinctes de gérer le nombre variable de lésions ainsi que la prise en compte (ou non) de leur dispersion. Nous montrons ainsi qu'aucun ensemble de caractéristiques n'est clairement meilleur que les autres pour prédire la survie sans progression à deux ans, prouvant la complexité de la tâche. Cependant, nous observons également que pour un modèle simple n'utilisant que l'information d'une

seule lésion il est mieux de ne conserver qu'un nombre restreint de caractéristiques ; tandis que des modèles utilisant l'information de toutes les lésions sont capables de tirer profit d'un plus grand nombre de caractéristiques. Par rapport à la gestion du nombre variable de lésions, nous montrons qu'utiliser un graphe de lésions pour représenter l'image comme dans [Aswathi et al. 2023] se révèle être parmi les méthodes les plus performantes indépendamment de l'ensemble de caractéristiques choisi.

- **Chapitre III :** Ce chapitre vise à répondre à la deuxième question de ce travail portant sur la fusion des informations de l'image TEP et des données cliniques. Nous y proposons un nouveau modèle réalisant pour la première fois la fusion entre les données images et cliniques au niveau lésionnel. Dans ce but, nous concevons un module d'attention croisée et présentons une étude poussée de sa structure interne afin qu'il s'adapte au mieux à nos données multimodales complexes. Afin d'agréger l'information provenant d'un nombre variable de lésions, nous présentons une opération d'agrégation par maximum qui sélectionne automatiquement un nombre restreint de lésions pertinentes associées à chaque prédiction, renforçant l'interprétabilité de notre modèle. Finalement, nous discutons en profondeur des résultats de notre meilleur modèle, soulignant son potentiel pour identifier les patients atteints de DLBCL à haut risque tout en offrant de l'information au niveau lésionnel sur les caractéristiques (tant images que cliniques) qui sont les plus importantes pour la décision. En ce sens, tant représenter l'information des lésions en un graphe que la conception du module d'attention croisée pour la fusion multimodale ont le potentiel d'apporter de nouvelles informations aux cliniciens et donc d'améliorer la compréhension de la pathologie.

- **Chapitre IV :** Ce chapitre traite de la dernière question abordée dans cette thèse, qui est l'utilisation de la similarité entre les patients pour améliorer leur stratification (donc leur séparation en groupes de risque). Nous y modifions un modèle propageant l'erreur de prédiction de patients semblables sur un graphe de population [Raymond et al. 2024] afin qu'il soit capable de corriger des prédictions de courbes de survie et de gérer de la censure au lieu de travailler sur de simples étiquettes de régression. Pour ceci, nous proposons d'introduire explicitement le temps d'événement des patients dans la construction du graphe de population. De plus, dans ce but, nous tirons profit d'un modèle basé sur un *transformer* capable d'utiliser les étiquettes connues des patients d'entraînement de façon semi-supervisée [Shi et al. 2021]. Nous proposons également une étape de post-traitement qui permet d'assurer que les courbes de survie corrigées par le

modèle conservent leurs propriétés, tout en gardant des performances similaires. Finalement, nous appliquons tout ce processus de correction post-hoc basé sur un graphe de population à un problème clinique crucial, la stratification de patients, ce qui n'avait à notre connaissance jamais été fait ; et nous montrons que l'application de cette correction permet d'améliorer nettement les performances du modèle corrigé par rapport au modèle initial.

Enfin, nous concluons ce manuscrit en discutant des contributions et des perspectives futures de ce travail.

Cette thèse a donné lieu à 3 publications : le contenu de la partie II a été publié dans le journal *Computerized medical imaging and graphics* [Thiery, Rizkallah, Bailly, et al. 2025], avec le contenu de la partie III qui a également été publié dans l'atelier AI4TREAT de MICCAI 2023 [Thiery, Rizkallah, Bailly, et al. 2023]. Finalement, le contenu de la partie IV, né d'une collaboration avec l'équipe MEDomics de l'Université de Sherbrooke, a été publié dans l'atelier GRAIL de MICCAI 2025 [Thiery, Rizkallah, Laribi, et al. 2025].

Le travail réalisé dans le cadre de cette thèse a été soutenu par le projet AIby4 (Centrale Nantes-Project ANR-20-THIA-0011), INCa-DGOS-INSERM-ITMO Cancer 18011 (SIRIC ILIAD) avec le support de la région Pays de la Loire (GCS IRECAN 220729), le Fonds Européen de Développement Régional (FEDER), la région Pays de la Loire via le programme *Connect Talent MILCOM* et Nantes Métropole (Conv. 2017-10470).

List of Figures

1.1 Summary of the contributions of this thesis. 4

2.1 PET scan and corresponding semi-quantitative indicators computation. 8
2.2 GAINED study protocol (from [Le Gouill et al. 2021]). 9
2.3 PET scans of two patients from the GAINED dataset. 9

3.1 Example of survival analysis curves. 12
3.2 Example of Kaplan-Meier survival estimations for two different groups of patients. In this case, the green curve describes a population with a higher risk than the red curve (as the corresponding patients have lower survival times). 15
3.3 Example of a (non-survival) Random Forest. 17
3.4 Architecture of DeepHit with two competing events. 19

4.1 Example of graph. 24
5.2 Cross-validation process. 38
5.3 Creation of a balanced validation/test set. 38
6.4 Fusion framework overview. 46
6.5 Proposed design choices and configurations for the cross-attention module specially conceived to be combined with a lesion graph. 50
6.6 Explanation of the max pooling role for interpretability. 60
6.7 AU-ROC performance of the best model in patients with progressively removed lesions. 61
6.8 Example of interpretation of the lesion self-attention matrix. 62
6.9 Sum of the clinical Fully Connected layer weights, for each clinical characteristic. 63
6.10 Imaging features for which the distribution of the retained lesions l_+ (in orange), l_- (in green) and all the lesions of the patients (in blue) are statistically different. 64
7.11 Population graph construction. 70
7.12 Population-graph method overview. 74

7.13 UniMP module overview. 76
7.14 Estimated Event Time (EET) and Area Under the Survival Curve (AUSC) computation for high-risk patient identification. 81
7.15 Distribution of censoring and events in the METABRIC dataset. 82
7.16 For a given training/validation/test split, definition of the $\mathcal{P}_{L^\bullet}$, $\mathcal{P}_{L^\circ}$ and $\mathcal{P}_U$ sets. Training and validation only concern the UniMP (Sec. 7.2.4)). 83
7.17 Qualitative evaluation of the population-graph method results. 85

A.1 PET imaging: explanation and example. 96

D.1 Overview of the proposed fusion framework with two multimodal blocks. 101

List of Tables

5.1 AU-ROC results for the DLBCL multi-lesion 2-year PFS prediction with PET images. 39
6.2 AU-ROC results for the data fusion study. 54
6.3 Best structure obtained for each cross-attention configuration and corresponding performance. 56
6.4 Ablation study over the components of our best multimodal model. . . 57
6.5 Comparison with baseline fusion models on the larger Featv5 set of features. 58
7.6 Stratification quality and survival performances of the base model, and of the base model corrected by our survEPN method applied on either random graphs; a feature graph; and our temporal graph. 85
7.7 Stratification quality and survival performances of the base model, the base model corrected by our survEPN, and the base model corrected by our survEPN method with sigmoid fitting to get proper survival curves. 87

B.1 Clinical features considered in the GAINED dataset. 97
B.2 Lesions 'HeadNeck' imaging features. 98
B.3 Lesions 'FirstOrder' imaging features. 99

C.1 Grid search study (HC = hidden channels ; LR = learning rate). 100

Abbreviations

AU-ROC	Area Under the Receiver Operating Characteristic curve
AUSC	Area Under the Survival Curve
BS	Brier Score
cAMAE	Censoring-aware Average Mean Absolute Error
CLIP	Contrastive Language-Image Pre-training
CNN	Convolutional Neural Network
CT	Computed Tomography
DLBCL	Diffuse Large B-Cell Lymphoma
ECOG	Eastern Cooperative Oncology Group
EET	Estimated Event Time
EPN	Error Passing Network
FC	Fully Connected layer
FDG	FluoroDeoxyGlucose
GATv2/GAT	Graph ATtention network (version 2)
GAINED	GA In NEwly Diagnosed diffuse large B-cell lymphoma
GNN	Graph Neural Network
HC	Hidden Channel
IBS	Integrated Brier Score
IBSI	Image Biomarker Standardisation Initiative
IPI	International Prognostic Index
K-NN	K-Nearest Neighbours
N-MTLR	Neural Multi-Task Logistic Regression
LDH	Lactate DeHydrogenase
LR	Learning Rate
METABRIC	Molecular TAxonomy of Breast Cancer International Consortium

MIL	Multiple Instance Learning
MIP	Maximum Intensity Projection
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
MSE	Mean Squared Error
MTLR	Multi-Task Logistic Regression
MTV	Metabolic Tumor Volume
PET	Positron Emission Tomography
PFS	Progression Free Survival
ReLU	Rectified Linear Unit
RSF	Random Survival Forest
survEPN	Survival Error Passing Network
SUV	Standardized Uptake Value
SVM	Support Vector Machine
TMTV	Total Metabolic Tumor Volume
TLG	Total Lesion Glycolysis
UniMP	Unified Message Passing

Notations

General

$\mathbf{x}^{(n)}$	Feature vector of patient n
$\delta^{(n)}$	Event indicator of patient n: 0 if the patient is censored and 1 if his/her event occurred in the study
$E^{(n)}$	True event time of patient n
$C^{(n)}$	Censoring time of patient n
$t_c^{(n)}$	Continuous survival label of a patient n: equal to $E^{(n)}$ if $E^{(n)} \leqslant C^{(n)}$ ($\delta^{(n)} = 1$) else equal to $C^{(n)}$
$S^{(n)}(t)$	Survival function of patient n
$h^{(n)}(t)$	Hazard rate of patient n
N_{KM}	Number of patients used for Kaplan-Meyer estimation
Υ_i	Number of individuals at risk just before time t_i
d_i	Number of events at time t_i
$\hat{p}_{KM,i}$	Estimated probability of having an event in the interval $]t_{i-1}; t_i]$ while being alive at time t_{i-1}
$\hat{S}^{(n)}(t)$	Estimated survival function of patient n
$\hat{h}(t)$	Estimated hazard rate
N_C	Number of covariates of a patient
η	Parameter vector of a Cox model
$h_0(t)$	Baseline hazard of a Cox model
$\phi_{DeepSurv}$	DeepSurv deep learning model
$\phi_{CoxTime}$	CoxTime deep learning model
$\tilde{\mathcal{R}}^{(i)}$	Subset of the patients at risk at time $t_c^{(i)}$
N_T	Number of discretization intervals to compute the MTLR estimator
θ_i	Logistic regression weight matrix in the MTLR estimator

b_i	Logistic regression bias in the MTLR estimator
N_{BS}	Number of observations for the Brier score computation
$y^{(n)}$	Binary label of patient n
$\hat{p}^{(n)}$	Estimated probability of having $y^{(n)} = 1$
$\hat{G}(t)$	Kaplan-Meier estimate of the censoring survival function $\mathrm{P}(C > t)$
$\mathcal{V}$	Set of nodes of a graph
$\mathcal{E}$	Set of edges of a graph
v	Node of a graph
e	Edge of a graph
w_{ij}	Edge weight between node v_i and node v_j
$\mathbf{z}_i$	Feature vector of node v_i
$\mathcal{N}(i)$	Neighbours of node i
$\mathbf{\Theta}_1$	Projection matrix for the source vector in the GraphConv process
$\mathbf{\Theta}_2$	Projection matrix for the aggregated neighbour vector in the GraphConv process
$\mathbf{\Theta}_s$	Projection matrix for the source vector in the GATv2 process
$\mathbf{\Theta}_t$	Projection matrix for the target vector in the GATv2 process
$\mathbf{\Theta}_e$	Projection matrix for the edge weight in the GATv2 process
$\mathbf{a}$	Learned weighting vector in the GATv2 process
$\rho_{i,j}$	Attention weights between node i and node j in a GATv2 layer
$\mathbf{z}_{i_{GraphConv}}$	Feature vector of node v_i updated by a GraphConv layer
$\mathbf{z}_{i_{GAT}}$	Feature vector of node v_i updated by a GATv2 layer
$\mathbf{Q}$	‘Query‘ matrix of the cross-attention process
$\mathbf{K}$	‘Key‘ matrix of the cross-attention process
$\mathbf{V}$	‘Value‘ matrix of the cross-attention process
$\bar{\mathbf{Q}}$	Projected ‘Query‘ matrix of the cross-attention process
$\bar{\mathbf{K}}$	Projected ‘Key‘ matrix of the cross-attention process
$\bar{\mathbf{V}}$	Projected ‘Value‘ matrix of the cross-attention process
$\mathcal{H}_Q$	Height of the ‘Query‘ matrix of the cross-attention process
$\mathcal{W}_Q$	Width of the ‘Query‘ matrix of the cross-attention process
$\mathcal{H}_K$	Height of the ‘Key‘ matrix of the cross-attention process
$\mathcal{W}_K$	Width of the ‘Key‘ matrix of the cross-attention process

$\mathcal{H}_V$	Height of the ‘Value‘ matrix of the cross-attention process
$\mathcal{W}_V$	Width of the ‘Value‘ matrix of the cross-attention process
$\mathcal{W}_{\bar{Q}}$	Width of the projected ‘Query‘ matrix of the cross-attention process
$\mathcal{W}_{\bar{K}}$	Width of the projected ‘Key‘ matrix of the cross-attention process
$\mathcal{W}_{\bar{V}}$	Width of the projected ‘Value‘ matrix of the cross-attention process
$\mathbf{W}_{CA,Q}$	Projection matrix for the ‘Query‘ matrix of the cross-attention process
$\mathbf{W}_{CA,K}$	Projection matrix for the ‘Key‘ matrix of the cross-attention process
$\mathbf{W}_{CA,V}$	Projection matrix for the ‘Value‘ matrix of the cross-attention process
$\mathbf{Z}^{(n)}_{CrossAtt}$	Feature matrix of patient n after the cross-attention module
β_1	β_1 parameter of the Adam optimizer
β_2	β_2 parameter of the Adam optimizer
ϵ	ϵ parameter of the Adam optimizer

Related to the multimodal lesion graph approach

$\mathcal{I}^{(n)}$	Whole-body PET image of GAINED patient n
$L^{(n)}$	Number of lesions of GAINED patient n
$\mathcal{I}^{(n)}_{seg}$	Segmented whole-body PET image of GAINED patient n
$\mathbf{c}^{(n)}$	Clinical data vector of GAINED patient n
$\bar{\mathbf{c}}^{(n)}$	Projected clinical data vector of GAINED patient n
$\bar{\mathbf{c}}^{(n)}_{concat}$	Matrix resulting from the column-wise concatenation of the vector $\bar{\mathbf{c}}^{(n)}$, $L^{(n)}$ times
$\mathbf{z}^{(n)}_i$	Image feature vector of lesion i of GAINED patient n
$\mathbf{Z}^{(n)}$	Matrix concatenating the image feature vector of all the lesions of GAINED patient n
D_{clin}	Dimension of the GAINED clinical vector
D_{img}	Dimension of the GAINED image feature vector extracted from the PET segmentations of the lesions
D_{proj}	Dimension of the projected GAINED clinical vector
D_{GAT}	Dimension of the output of the GNN operation in the multimodal fusion method
D_{concat}	Dimension of the concatenation of the output of the GNN operation, the initial image feature vector, and the clinical data of a patient: $D_{concat} = (D_{GAT} + D_{img}) + D_{proj}$

$\mathcal{G}_l^{(n)}$	Lesion graph of patient n
$\mathcal{V}_l^{(n)}$	Set of nodes of the lesion graph of patient n
$\mathcal{E}_l^{(n)}$	Set of edges of the lesion graph of patient n
$v_i^{(n)}$	Node of the lesion graph of GAINED patient n corresponding to its i^{th} lesion
$e_{ij}^{(n)}$	Edge between node i and node j in the lesion graph of GAINED patient n
$w_{ij}^{(n)}$	Edge weight between node i and node j in the lesion graph of GAINED patient n
$\mathbf{p}_i^{(n)}$	Centroid of lesion i of GAINED patient n
σ_1	Population-level standard deviation of the GAINED lesion's centroids
σ_2	Population-level standard deviation of the lesion feature distances
γ	Hyperparameter responsible for tuning the lesion graph edge weight distribution
$\mathbf{z}_{i_{GAT}}^{(n)}$	Updated feature vector of a node i of patient n after a GATv2 layer
$\mathbf{z}_{i_{GraphConv}}^{(n)}$	Updated feature vector of a node i of patient n after a GraphConv layer
$\mathbf{Z}_{\text{SelfAtt}}^{(n)}$	Updated image feature matrix of patient n after a self-attention layer
$\mathbf{Z}_{GAT}^{(n)}$	Updated image feature matrix of patient n after a GATv2 layer, obtained by concatenating the $\mathbf{z}_{i_{GAT}}^{(n)}$ updated feature vectors of all nodes of the patient
$\tilde{\mathbf{Z}}^{(n)}$	Image feature matrix of patient n after the skip connection concatenating the output of the GATv2 layer $\mathbf{Z}_{GAT}^{(n)}$ and the original feature matrix $\mathbf{Z}^{(n)}$ extracted from patient n's PET scan
$\mathbf{Z}_{CrossAtt}^{(n)}$	Feature matrix of patient n after the cross-attention module
$\mathbf{W}_{SA,Q}$	Projection matrix for the 'Query' matrix of the self-attention process
$\mathbf{W}_{SA,K}$	Projection matrix for the 'Key' matrix of the self-attention process
$\mathbf{W}_{SA,V}$	Projection matrix for the 'Value' matrix of the self-attention process
d_q	Dimension of the 'Query' projection in the self-attention module of the multimodal process
d_k	Dimension of the 'Key' projection in the self-attention module of the multimodal process
d_v	Dimension of the 'Value' projection in the self-attention module of the multimodal process
μ_i	Weight i of the Mult2Clin $W_{CA,V}$ vector

$\mathbf{a}_i$	Cross-attention row corresponding to lesion i
l_+	Lesion selected by the multimodal model and positively correlated to the predicted PFS score
l_-	Lesion selected by the multimodal model and negatively correlated to the predicted PFS score

Related to the population-graph approach

$\mathbf{X}$	Concatenated features of all the patients in the population-graph framework
D_{feat}	Dimension of the feature vector $\mathbf{x}^{(n)}$ of patient n in the population-graph framework
$\mathcal{P}$	Set including the whole population considered in the population-graph framework
N_{pop}	Number of patients in this population
N_L	Number of labeled patients considered
N_M	Number of masked patients considered for the weighted MSE loss computation
$\mathcal{P}_L$	Set of labeled patients
$\mathcal{P}_U$	Set of unlabeled patients
$\mathcal{P}_{low}$	Set of low-risk patients
$\mathcal{P}_{high}$	Set of high-risk patients
$\mathcal{P}'_{low}$	Set of low-risk patients after survEPN correction
$\mathcal{P}'_{high}$	Set of high-risk patients after survEPN correction
$f_\Theta(\cdot)$	Survival base model to improve by the population-graph framework
Θ	Parameters of the survival base model to improve by the population-graph framework
$\hat{S}^{(n)}_{BM}$	Survival curve predicted by the base model f_Θ for a patient n
$\tilde{S}^{(n)}$	Survival curve of patient n corrected by the survEPN framework
$\tilde{S}^{(n)}_{sig}$	Survival curve of patient n corrected by the survEPN framework and after a sigmoid fitting post-processing step
$\hat{E}^{(n)}$	Predicted event time of patient n
$\mathcal{G}_{\text{pop}}$	Population graph built in our survEPN framework
$u^{(n)}$	Continuous uncensored label of patient n for a regression task

$\hat{u}^{(n)}$	Predicted continuous uncensored label of patient n for a regression task
$\tilde{u}^{(n)}$	Continuous uncensored label of patient n for a regression task corrected by the EPN framework
$\epsilon^{(n)}$	Residual (=prediction error) of patient n
$\mathcal{G}_{EPN}$	Population graph of the EPN framework
$\mathcal{V}_{EPN}$	Set of nodes of the EPN population graph
$\mathcal{E}_{EPN}$	Set of edges of the EPN population graph
$\mathcal{N}_{EPN}(n)$	Neighbours of node n in the EPN population graph
$e_{EPN}^{(i,j)}$	Edge between patient i to patient j in the EPN population graph
$\alpha_{EPN}^{(i,j)}$	Edge weight between patient i to patient j in the EPN population graph
$s_{EPN}^{(i,j)}$	Similarity between patient i to patient j in the EPN process
$\mathbf{W}_{EPN,Q}$	Projection matrix for the source vector in the EPN process
$\mathbf{W}_{EPN,K}$	Projection matrix for the target vector in the EPN process
c_{Φ}	Module predicting pseudo-labels in the population-graph framework
g_{Ψ}	Module computing and propagating the prediction error in the population-graph framework
L_{MSE}	MSE loss
L_{wMSE}	Weighted MSE loss for training the UniMP module
$\mu_k^{(n)}$	Weight of the k^{th} class of patient n in the weighted MSE loss computation
N_{cl}	Number of classes for the discretization of $t_c^{(n)}$
$t_c^{(n)}$	Continuous survival label of a patient n: equal to $E^{(n)}$ if $E^{(n)} \leqslant C^{(n)}$ ($\delta^{(n)} = 1$) else equal to $C^{(n)}$
$\hat{t}_c^{(n)}$	Predicted estimation (pseudo-label) of the continuous survival label $t_c^{(n)}$ of a patient n
$t_d^{(n)}$	Discretized survival label of a patient n
$\hat{t}_d^{(n)}$	Class predicted for a patient n
$\mathbf{t}_{oh}^{(n)}$	Vector of the one-hot discrete label of a patient n
$\mathbf{t}_{oh,k}^{(n)}$	k^{th} element of the one-hot discrete label of a patient n
$\hat{\mathbf{t}}_{oh}^{(n)}$	Vector of the prediction of the discrete label of a patient n
$\bar{\mathbf{t}}_d^{(n)}$	Vector containing the projection of the discrete label of patient n if uncensored and labeled, and null vector otherwise

$\bar{\mathbf{T}}_d$	Matrix containing the projection of the discrete labels of all uncensored labeled patients, and null vectors for the other patients
$\mathbf{H}_l$	Feature matrix of all nodes/patients at layer l of the UniMP module
$\mathbf{h}_l^{(n)}$	Node/patient n's feature vector after layer l of the UniMP module
$\tilde{\mathbf{h}}_l^{(n)}$	Node/patient n's intermediary feature vector at layer l of the UniMP module
L	Number of transformer layers in the UniMP module
M	Number of heads of the transformers of the UniMP module
d_h	Dimension of each head of the transformers of the UniMP module
$\mathbf{W}_T$	Projection matrix for the discrete labels
$\mathbf{W}_{l,m,q}$	Projection matrix for the 'Query' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{W}_{l,m,k}$	Projection matrix for the 'Key' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{W}_{l,m,v}$	Projection matrix for the 'Value' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{W}_{l,r}$	Projection matrix for the residual of the UniMP module at layer l
$\mathbf{W}_{l,g}$	Projection matrix for the gated residual connection of the UniMP module at layer l
$\mathbf{b}_{l,m,q}$	Bias for the projection of the 'Query' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{b}_{l,m,k}$	Bias for the projection of the 'Key' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{b}_{l,m,v}$	Bias for the projection of the 'Value' vector of the UniMP module for the m^{th} attention head of layer l
$\mathbf{b}_{l,r}$	Bias for the projection of the residual of the UniMP module at layer l
$\mathbf{q}_{l,m}^{(n)}$	'Query' vector of patient n of the UniMP module for the m^{th} attention head of layer l
$\mathbf{k}_{l,m}^{(n)}$	'Key' vector of patient n of the UniMP module for the m^{th} attention head of layer l
$\mathbf{v}_{l,m}^{(n)}$	'Value' vector of patient n of the UniMP module for the m^{th} attention head of layer l
$\mathbf{a}_{l,m}^{(ij)}$	Attention between patient i and patient j in the UniMP module for the m^{th} attention head of layer l

$r_l^{(n)}$	Residual at layer l for patient n in the UniMP module
$\beta_l^{(n)}$	Term balancing the weight of the residual connection in the UniMP module, at layer l and for patient n
$\mathcal{G}_{\text{pop}}$	Population graph built in our survEPN framework
$\mathcal{V}_{pop}$	Set of nodes of the population graph built in our survEPN framework
$\mathcal{E}_{pop}$	Set of edges of the population graph built in our survEPN framework
$\mathbf{W}_{pop}$	Matrix of directed edge weights of the population graph built in our survEPN framework
$\mathcal{N}_p(n)$	Neighbours of node n in the survEPN population graph
$s^{(i,j)}$	Similarity between patient i to patient j in the survEPN process
ω	Small value to avoid a division by 0 inside the similarity computation in the survEPN process
$e_s^{(i,j)}$	Edge between patient i to patient j in the survEPN population graph
$\alpha^{(i,j)}$	Edge weight between patient i to patient j in the survEPN population graph
τ	Temperature term controlling the distribution of the attentions $\alpha^{(i,j)}$
A_{sig}	Parameter of the sigmoid fitting step controlling the amplitude of the sigmoid
t_{sig}	Parameter of the sigmoid fitting step controlling the time at which the curve decreases
s_{sig}	Parameter of the sigmoid fitting step controlling the slope of the sigmoid

Part I

Introduction and context

Chapter 1

Introduction

THIS PhD has been done thanks to a collaboration between the CRCI^{2}NA (Centre de recherche en cancérologie et immunologie Nantes-Angers)/Inserm, the LS2N (Laboratoire des Sciences du Numérique de Nantes)/École Centrale de Nantes, and the nuclear medicine team of the Nantes University Hospital, and was integrated into the MILCOM (Multi-modal Imaging and Learning for Computational-based Medicine) project. It has been funded by the AIby4 project (Centrale Nantes-Project ANR-20-THIA-0011), INCa-DGOS-INSERM-ITMO Cancer 18011 (SIRIC ILIAD) with support from the Pays de la Loire region (GCS IRECAN 220729), the European Regional Development Fund (FEDER), the Pays de la Loire region on the Connect Talent MILCOM programme and Nantes Metropole (Conv. 2017-10470).

This work is focused on one pathology, the Diffuse Large B-cell Lymphoma (DLBCL). Some patients affected by this disease do not respond to the standard frontline therapy. Early identification of these patients, who are considered 'high-risk', at the time of diagnosis could enable clinicians to personalize treatment strategies, potentially improving survival outcomes. Conversely, recognizing low-risk patients may help reduce treatment-related toxicity by allowing for therapeutic de-escalation.

Currently, the diagnosis and follow-up of DLBCL patients rely on clinical biomarkers and on the interpretation of ^{18}F-Fluorodeoxyglucose (FDG)-PET/CT images (PET: Positron Emission Tomography; CT: Computed Tomography) [Tabaa et al. 2021]. If the clinical indicators have been proven prognostic for the high-risk patients identification task [Le Gouill et al. 2021], the PET images are less straightforward to use for this end. Their prognostic ability mostly relies on the extraction of semi-quantitative indicators, which show consequent potential for the identification of high-risk patients [Schöder et al. 2016], even if they are still not used in the clinical routine to this end [Boellaard et al. 2024]. However, methods relying on these indicators for high-risk patient identification only consider one or a few imaging measurements from the high-dimensional PET scans, and they do not take into account the information from other modalities. Recently, automatic methods have been developed to extract richer information from PET images,

but very few of these approaches explicitly consider the presence of multiple lesions and their spatial distribution within the patient. When this aspect is addressed, it is done through simplified metrics, such as measuring the largest distance between lesions [Cottereau et al. 2019], which fail to capture the semantic information contained within each lesion and the structural relationship between all of them. Furthermore, only a few methods incorporate both imaging and tabular clinical information, and they do so through basic fusion schemes [Carlier et al. 2024; Eertink et al. 2022; Jiang et al. 2022]. These observations led to the first two questions addressed in this thesis:

(I) How to best use PET images for high-risk DLBCL patient identification?

Especially, which features should be extracted from these images, and how to take into account the variable number of lesions of each patient and their spatial distribution?

(II) How to efficiently fuse PET image information with clinical data to improve the performance of high-risk DLBCL patient identification?

Furthermore, these methods rely solely on the individual information of each patient to make a prediction, without directly taking advantage of the information from other individuals with similar profiles. Yet, approaches that explicitly account for population similarities and differences have been shown to improve the performance of individual predictors. Motivated by this, our work addresses one final question:

(III) How to take advantage of inter-patients similarity to improve high-risk patients identification?

To answer these questions, we rely on the recent development of graph approaches in the medical field. Such graphs have been used, among other applications, to represent individual patients, for example by modeling their blood vessels [Kassam et al. 2022] or, an option of significant interest for our work, by representing the lesions of a patients as a lesion graph to take into account their spatial distribution in addition to their individual information [Aswathi et al. 2023; Kazmierski et al. 2021; Lv et al. 2023]. Another way of taking advantage of graph structures is to use them to represent populations, *i.e.* each node corresponding to one patient, to leverage the similarities and differences between patients to improve individual predictions [Bintsi et al. 2023; Zhang et al. 2023; Zheng et al. 2022]. Both approaches will be explored in this work.

To sum up, in this thesis, we aim to assist in the identification of high-risk DLBCL patients by taking advantage of all the available information at the time of diagnosis: individual PET scans and clinical data, along with other patients' data for whom we already know the treatment response. The document is organized into an initial part dedicated to the introduction and state of the art, followed by three main sections, each addressing one of the central research questions, as illustrated in Fig. 1.1.

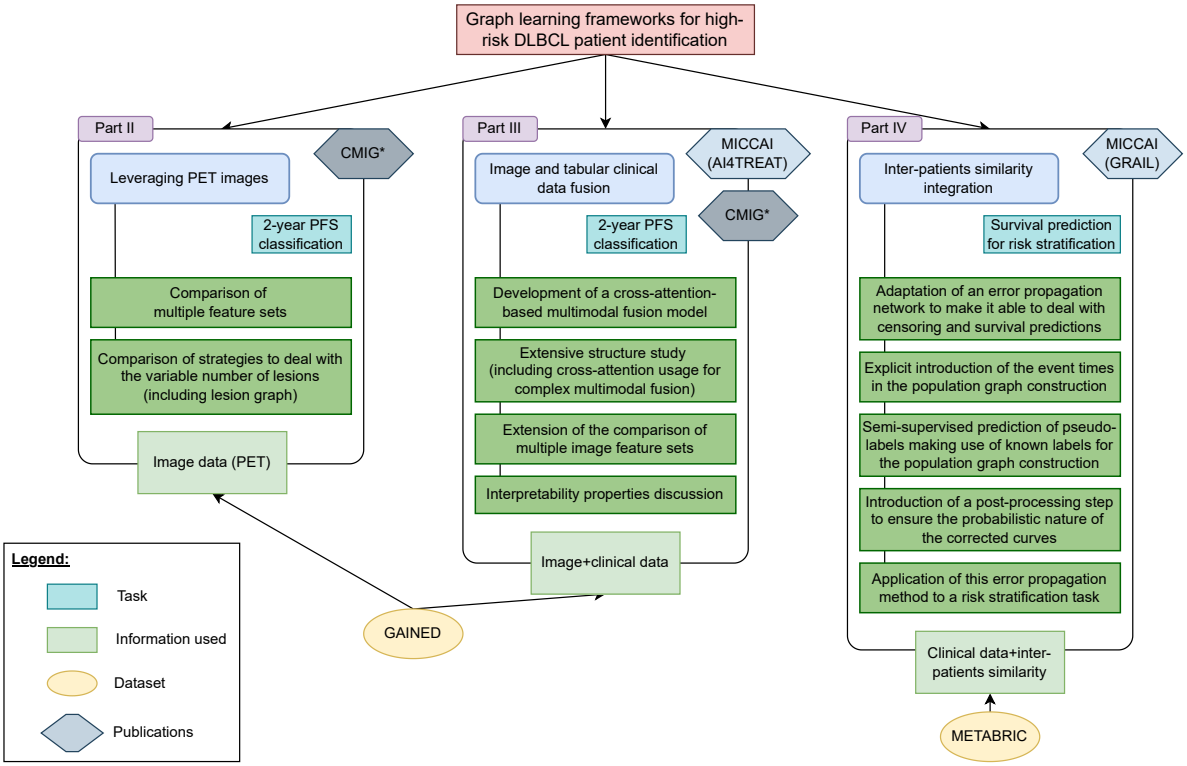


Figure 1.1 – Summary of the contributions of this thesis. PFS means Progression-Free Survival, and $*$ means that the two blocs presenting this publication correspond to a single article (in the journal Computerized Medical Imaging and Graphics).

In Part I, chapter 2 lays the groundwork for the thesis by introducing the clinical context and describing the clinical protocol followed for DLBCL treatment in the clinical study we rely on. Chapter 3 provides a comprehensive overview of survival analysis, a core framework underpinning notably our final contribution. In this chapter are defined key concepts, taxonomies, and notations used throughout the thesis. Then, in chapter 4, we give a broad background of the state of the art concerning the different methodologies tackled in this work: DLBCL analysis, multimodal fusion and graphs.

Next are presented our different contributions related to the questions introduced above. In a nutshell, the rest of the thesis is divided into three parts corresponding to our three major contributions: leveraging PET images, multimodal multi-lesion fusion, and exploiting inter-patient similarity for improved patient stratification.
In Part II, we seek to **leverage the PET imaging data** (Question (I)). More precisely:

- We analyze what imaging features are most relevant for a 2-year PFS prediction task in DLBCL. To this end, we compare five sets of PET radiomic features, based on different lesion segmentation strategies and sets of radiomics extracted. We first show that none clearly outperforms the others. However, we observe that for a model taking only information from one lesion it is better to keep a low number

of features; while richer models using multiple lesions can benefit from using a higher number of features.

- We explore different strategies to deal with the variable number of lesions of each patient, and prove the value of a lesion graph approach to keep all relevant image information while offering a compact trade-off between the characterization of full images and single lesions, performing well regardless of the set of imaging features considered.

This work has been published in the journal *Computerized Medical Imaging and Graphics* [Thiery, Rizkallah, Bailly, et al. 2025].

Part III deals with the **clinical and image multimodal fusion** (Question (II)). More precisely:

- We conceive a new multi-modal fusion model to take advantage of the available image and clinical tabular data, proposing for the first time a fusion at the lesion level. To this end, we design a cross-attention module and present an extensive study of its inner structure such that it best adapts to our complex multimodal data with different and variable dimensionalities.
- We introduce a max pooling operation to aggregate the fused information from a variable number of lesions, which automatically selects a limited number of relevant lesions associated with each prediction, supporting the interpretability of our model.
- We extend our investigation to find the optimal imaging features to extract from PET scans for our task. We demonstrate that our previous findings hold when integrating these features with clinical data: while simple fusion models struggle with high-dimensional imaging inputs, our multimodal fusion model achieves the best performance when using the largest feature set.
- Finally, we discuss in depth the results of our best model, highlighting its potential to identify high-risk DLBCL patients, while directly providing per-lesion information on the features (both imaging and clinical) most relevant to the decision. In this sense, both representing the lesions' information as a graph and designing a cross-attention module for the multimodal fusion bring new information to the clinicians and thus have the potential to improve the understanding of the disease.

This work about multimodal fusion has also been published in the previously mentioned article in *Computerized Medical Imaging and Graphics* [Thiery, Rizkallah, Bailly, et al. 2025], as well as in the AI4TREAT workshop at MICCAI 2023 [Thiery, Rizkallah, Bailly, et al. 2023].

In Part IV, we tackle the third question (III) and leverage the **inter-patients similarity** to improve patients stratification:

- We adapt an error propagation network [Raymond et al. 2024] to deal with both survival predictions and censoring instead of simple regression labels.
- We propose to explicitly introduce the patients' event times in the population graph construction of the framework.
- To predict pseudo-labels for test patients, we make use of a transformer-based model [Shi et al. 2021] taking advantage of the training patients' known labels in a semi-supervised way.
- We introduce a post-processing step ensuring that the corrected survival curves keep their probabilistic nature (*e.g.* to keep them bounded and monotonically decreasing) with similar performances.
- And we apply for the first time such a propagation-based post-hoc method for the clinically crucial problem of risk stratification, and show a consequent improvement of the performance of the model after the post-hoc correction compared to the initial model.

This work was produced in collaboration with the MEDomics laboratory of Université de Sherbrooke following a two-month visit carried out as part of this PhD. An article covering it has been published in the GRAIL workshop at MICCAI 2025 [Thiery, Rizkallah, Laribi, et al. 2025].

Chapter 2

Clinical context

Diffuse Large B-cell Lymphoma (DLBCL) is a lymphatic cancer whose incidence is steadily increasing and which accounts for 30-40% of the 77240 new cases of Non-Hodgkin Lymphoma cases in the US in 2020 [Susanibar-Adaniya et al. 2021]. Its diagnosis and follow-up rely on the analysis of clinical biomarkers and the interpretation of ^{18}F-Fluorodeoxyglucose (FDG)-PET/CT images. This imaging modality, which is detailed in Appendix A, is now considered the reference imaging modality for DLBCL. It is mostly used to evaluate the therapeutic response of a patient over time by comparing visually and/or in a semi-quantitative way the patient's baseline PET scan to further scans realized along the treatment [Tabaa et al. 2021].

However, approximately 30-40% of DLBCL patients still do not respond to standard frontline therapy [Sehn et al. 2021]. Therefore, there is a need for the discovery and the use of new biomarkers to identify patients at high risk of relapse. In this respect, ^{18}F-FDG PET is currently being studied as a means to extract image biomarkers useful for the early identification of high-risk patients before treatment is started.

Currently, in clinical routine, the analysis of PET images is based on a visual analysis of medullary invasion and/or on the semi-quantitative estimation of the Standardized Uptake Value (SUV), methods which are only used to evaluate therapeutic response. Other indicators (presented in Fig. 2.1) such as the dispersion of the lesions [Cottereau et al. 2019] or the Total Metabolic Tumor Volume (TMTV) [Schöder et al. 2016] have shown consequent potential for the identification of high-risk patients, but they are still not used in clinical routine to this end [Boellaard et al. 2024]. Furthermore, their effectiveness in identifying high-risk patients could be further improved, as current methods relying on these semi-quantitative image measurements do not integrate other available information, *e.g.* known clinical biomarkers.

In this context, we propose to develop a computer-assisted learning-based method to assist in the identification of high-risk DLBCL patients at diagnosis (*i.e.* those more likely to experience disease progression), by efficiently fusing all available information, including PET imaging and spatial distribution of multiple lesions, tabular clinical data,

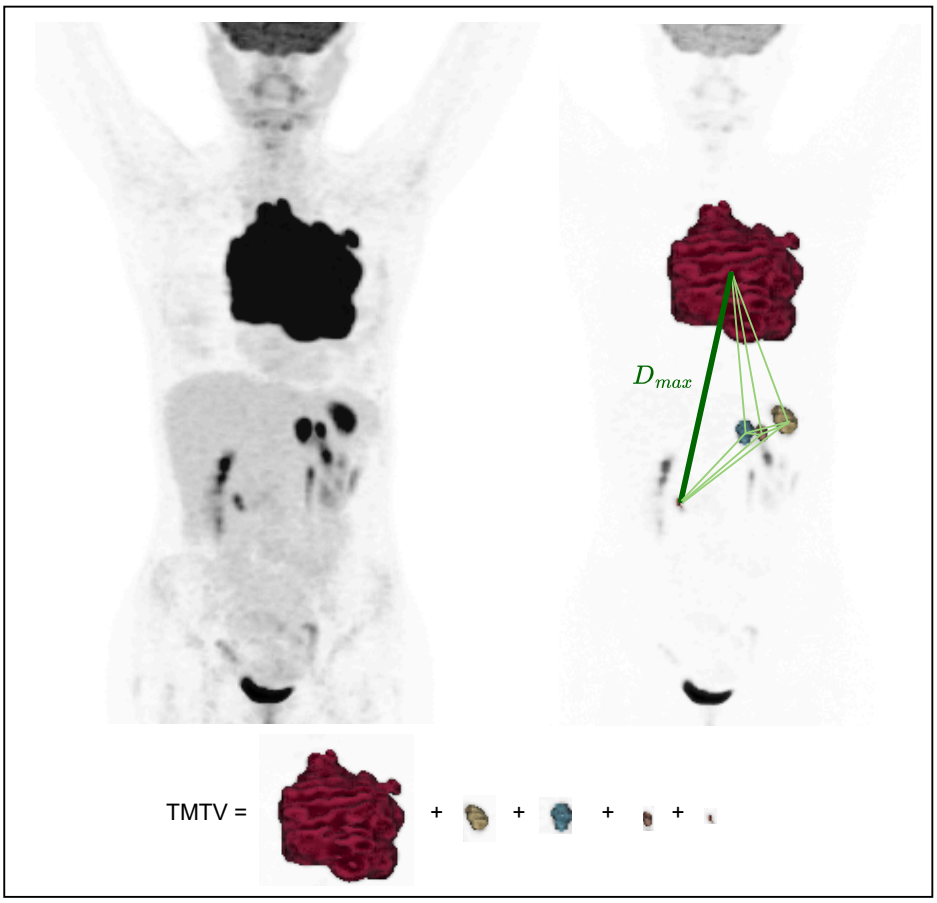


Figure 2.1 – PET scan and corresponding semi-quantitative indicators computation: D_{max} measuring the dispersion of the lesions corresponds to the highest distance between the centroid of two lesions (in dark green in the image), and the TMTV is computed as the sum of the metabolic volume of all the lesions.

and even knowledge of the treatment response of patients sharing similar characteristics.

For the development of our contributions and learning frameworks, we rely on a unique dataset of 670 newly diagnosed and untreated DLBCL patients from 99 centers (1 to 33 patients per center, 32 different camera models) based on the prospective GAINED study (NCT 01659099) [Le Gouill et al. 2021]. In this study, the consolidation treatment given to a patient is determined by their response to chemotherapy, evaluated by the evolution of the disease on three successive PET scans: one at baseline, one after two chemotherapy cycles, and one after four cycles. In the corresponding dataset, each patient is associated with a PET scan at baseline and 10 clinical indicators such as age, Ann Arbor stage or number of extranodal sites (full list in Appendix B.1); as well as an intermediary and a final PET scan. In this thesis, we do not consider the last two scans, as we want to rely only on the baseline information. The PET scans of the GAINED patients have significant variability in terms of number, size and dispersion

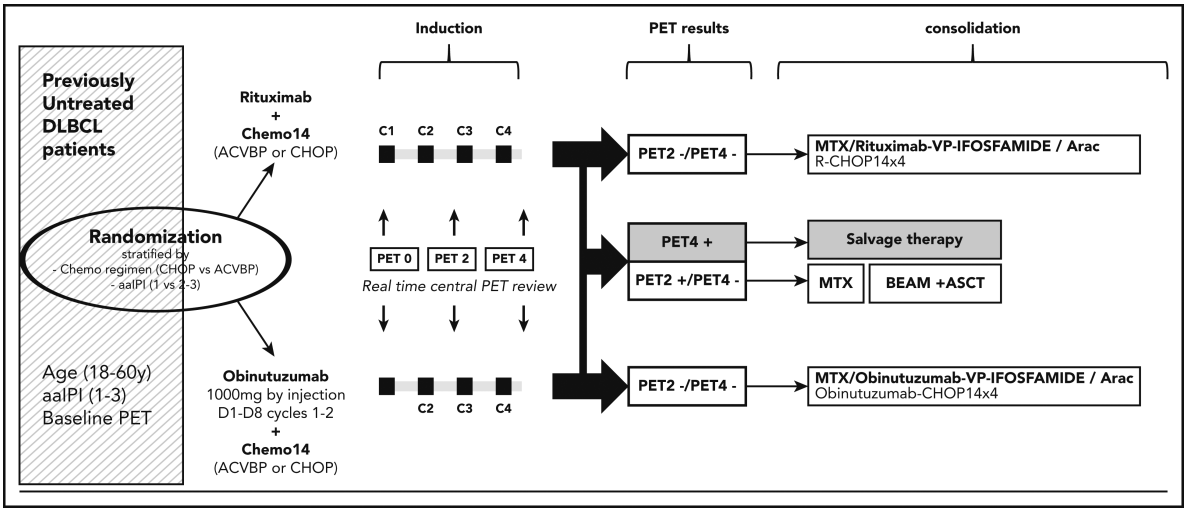


Figure 2.2 – GAINED study protocol (from [Le Gouill et al. 2021]).

of the lesions, as can be seen in Fig. 2.1 and Fig. 2.3.

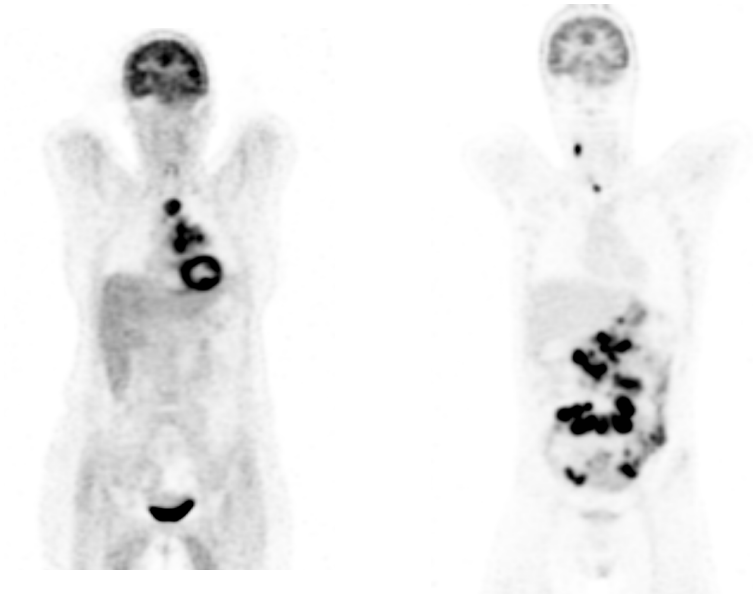


Figure 2.3 – PET scans of two patients from the GAINED dataset.

As a first approximation, we target a progression-free survival (PFS) 2-year classification, intended to predict whether the disease of a patient will progress within two years after the beginning of the treatment, by relying on clinical indicators and full-body PET images at baseline. This can be considered a reasonable surrogate for high-risk patient identification, with patients more likely to progress being considered high-risk. To obtain a prediction of 2-year PFS, we exclude patients whose data are censored before this time and those whose PET images do not meet the protocol quality criteria.

This leaves 545 patients for our experiments. Among these patients, 113 are deemed as positive for the PFS, either because of a progression or a relapse of the disease within two years or because of death within this time.

In the rest of this document, this GAINED dataset is used for the training and evaluation of the methods predicting the 2-year PFS value in Parts II and III. In Part IV, the method we propose will be applied to a survival analysis task, and thus rely on another dataset with a lower censoring rate. This concept of censoring, among other basic survival analysis knowledge, will be presented in the following section.

Chapter 3

Theoretical background on survival analysis

While for the first contributions of this work we simplify the problem of identifying high-risk patients to a 2-year progression-free classification task, the last contribution in Part IV relies instead on a survival analysis formulation. For this reason and for the sake of clarity, we introduce here the main concepts and definitions of survival analysis related to this thesis.

3.1 Basics: censoring, survival function and hazard rate

In the medical field but also beyond, survival analysis methods have been developed with the aim of estimating the time until a specific event. In medical applications, this event can be for example the death of a patient or a relapse of the disease. This survival estimation can be used, among others, to study if patients with different characteristics or biomarkers are differently affected by a pathology, or, in our case, to identify high-risk patients. In the following, we give a brief overview of some basic concepts in survival analysis. For a more in-depth introduction to the field, see for example [Klein et al. 2003].

The major specificity of survival analysis methods is that they can deal with censoring, which corresponds to the lack of information about some time-to-event durations. In this sense, an individual is said to be ***uncensored*** if we know both the beginning time and the event time (for example, both the time of diagnosis for a patient and his time of death are known). Missing data can occur in two main ways: an individual is considered ***left-censored*** when the start time is unknown, and ***right-censored*** when the event time is unknown, for instance, if the patient leaves the study before the event occurs or dies from a cause unrelated to the studied disease. An individual can also be both left- and right-censored.

In the studies considered in this thesis, we only encounter right-censored individuals. To model the censoring in our data, we use an indicator function $\delta^{(n)}$ which is equal to 1 if the event of the n^{th} patient happened and we know his/her true event time $E^{(n)}$ (uncensored), and $\delta^{(n)} = 0$ if the true event time is censored and we only know the censoring time $C^{(n)} < E^{(n)}$.

Formally, survival analysis methods focus on estimating one or various closely related functions. The first is the ***survival function*** $S^{(n)}(t)$ for individual n, which corresponds to the probability that his/her event has not yet occurred at time t:

$$S^{(n)}(t) = \mathrm{P}(E^{(n)} > t), \tag{3.1}$$

with $t > 0$. Another important survival concept is the ***hazard rate*** $h(t)$, a measure of the event rate at time t knowing that the event did not occur before time t:

$$h^{(n)}(t) = \lim_{\Delta t \to 0} \frac{\mathrm{P}(t \leqslant E^{(n)} < t + \Delta t | E^{(n)} \geqslant t)}{\Delta t}. \tag{3.2}$$

A real example of these two types of curves can be found in Fig. 3.1. Other quantities can also be derived from these functions (see [Klein et al. 2003]), but since they are not essential for the understanding of this work, they will not be introduced here.

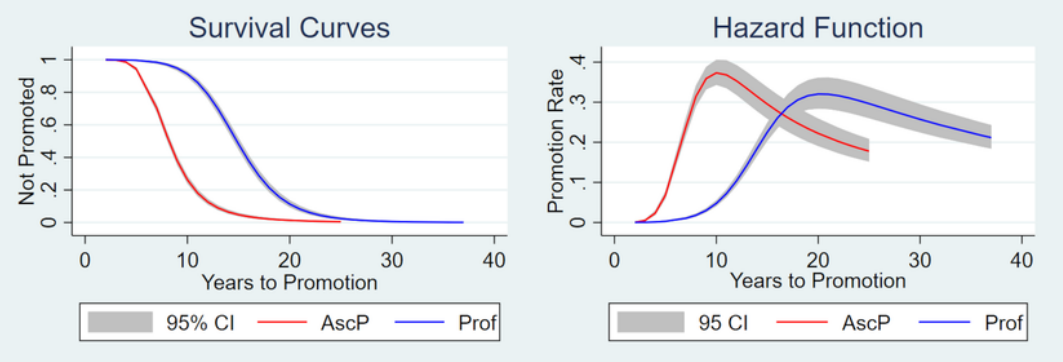


Figure 3.1 – An example of survival analysis curves from Fig. 1 in [Pickett et al. 2022], showing (A) the survival curve of the proportion of professors not promoted vs. years to promotion, and (B) the hazard function showing rate of promotion vs. years to promotion. In both graphs red indicates associate professor (AscP) promotion, and blue full professor (Prof) promotion. 95% CI is the 95% confidence level.

Since all survival analysis functions are closely related to each other, knowing one of them allows for the estimation of the other ones. In our case, we focus on the initial survival function S, which is the most commonly used by clinicians. As we will explain in Part IV, our goal is to improve a model that predicts such survival curves by leveraging the curves of similar patients whose event times are known.

3.2 Survival analysis methods

Survival analysis can be done either by statistical estimation or by learning methods. Statistical estimations are the most commonly used among clinicians. Such methods include:

— Non-parametric approaches such as the Kaplan-Meier or Nelson-Aalen estimator, which are used when no assumption can be made on the distribution of event times;

— Parametric approaches such as the Weibull model, which take into account a vector of covariates characterizing a patient to predict individual survival curves or risks, but rely on strong assumptions on the distribution of the event times;

— And semi-parametric approaches like the Cox model, which assumes an unknown baseline survival curve, and focuses on modeling the effects of different covariates on this baseline, avoiding the need for strong assumptions on the distribution of event times or risks.

As for learning methods, they can be divided into continuous-time approaches, discrete-time approaches, and methods predicting a single event time value, with a variety of methods in each.

3.2.1 Statistical analysis methods

3.2.1.1 Kaplan-Meier estimator

The most used survival estimator is the **Kaplan-Meier estimator** [Kaplan et al. 1958], as it avoids making assumptions on the distribution of the event times. Since it is non-parametric, it does not estimate a survival curve for each individual but instead relies on the list of event times of a full group of patients (and their censoring if any) to make a survival estimation for this population. Thus, it can be used for describing the survival of the considered population, for estimating its median survival and the survival rate at a given time, or for comparing the survival rate of different populations, often by using the log-rank test [Fleming et al. 1981].

In the following, and for simplicity, we consider the event of interest to be the death of a patient. In this setting, surviving after a time t means being alive just before time t and not dying at time t. Generally, we can express the probability that a patient n survives until time t_i by computing the product of the survival probabilities at every time $t_{k \leqslant i}$ knowing that the event $E^{(n)}$ did not happen at time t_{k-1}, with $t_0 = 0$ and $k \in [1, ..., i]$:

$$\mathrm{P}(E^{(n)} > t_i) = \prod_{k=1}^{i} \mathrm{P}(E^{(n)} > t_k | E^{(n)} > t_{k-1}). \tag{3.3}$$

Let's consider the list of events of the considered population, ordered according to their survival times (event or censoring depending on the first to come) $\{t_i\}_1^{N_{KM}}$, N_{KM} corresponding to the number of patients in this population. Additionally, we denote Υ_i the number of individuals at risk just before t_i, *i.e.* patients whose events did not occur and who were not censored before that time, and d_i the number of events at t_i. The probability $\hat{p}_{KM,i}$ that an event occurs in the interval $]t_{i-1}; t_i]$ knowing that a patient was alive at time t_{i-1} (*i.e.* $\hat{p}_{KM,i} = \mathrm{P}(E^{(n)} \leqslant t_i | E^{(n)} > t_{i-1})$) can be computed as:

$$\hat{p}_{KM,i} = \frac{d_i}{\Upsilon_i}. \tag{3.4}$$

If the survival times t_i (event or censoring times) are supposed distinct, $d_i = 0$ when $\delta^{(i)} = 0$ (censored individual), and $d_i = 1$ when $\delta^{(i)} = 1$. Thus, the Kaplan-Meier estimator corresponds to:

$$\hat{S}(t) = \prod_{i|t_i \leqslant t} \Big(1 - \frac{\delta^{(i)}}{\Upsilon_i}\Big). \tag{3.5}$$

The Kaplan-Meier survival curve obtained for a population is represented by a succession of steps of unequal height, this height being proportional to the number of events that occurred at the time of the step in this population. This representation clearly displays the difference between the survival behavior of two groups, as can be seen from the example in Fig. 3.2, and will be used for this property in Part IV.

3.2.1.2 Cox model

The semi-parametric **Cox model** is also widely used in the literature. It is employed when the objective is to assess the effect of a given variable (a *covariate*) on the survival times. It helps explain the occurrence of an event over time through one or more explanatory variables, which can be qualitative or quantitative.

Let's set $\mathbf{x}^{(n)}$ as the N_C-dimensional covariates vector of a patient n. The hazard rate estimated by the Cox model for this patient is given by:

$$\hat{h}(t|\mathbf{x}^{(n)}) = h_0(t) e^{\eta_1 x_1^{(n)} + \ldots + \eta_{N_C} x_{N_C}^{(n)}}, \tag{3.6}$$

with $h_0(t)$ the baseline hazard, which does not change between individuals, and η the vector of the model's parameters, constant over time and representing the covariates' effect on the hazard rate. From this equation, we can observe that for any two patients

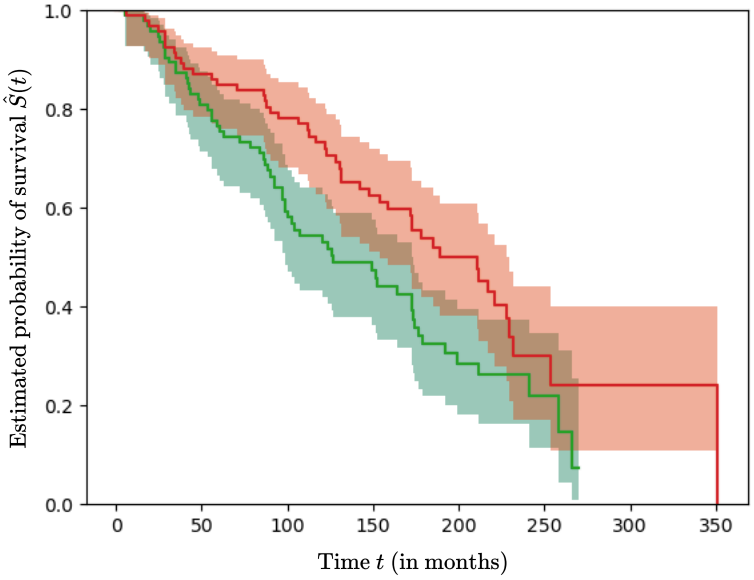


Figure 3.2 – Example of Kaplan-Meier survival estimations for two different groups of patients. In this case, the green curve describes a population with a higher risk than the red curve (as the corresponding patients have lower survival times).

n_1 and n_2,

$$\frac{\hat{h}(t|\mathbf{x}^{(n_1)})}{\hat{h}(t|\mathbf{x}^{(n_2)})} = \frac{e^{\eta^\top \cdot \mathbf{x}^{(n_1)}}}{e^{\eta^\top \cdot \mathbf{x}^{(n_2)}}}, \tag{3.7}$$

meaning that the ratio of the hazard rate of two individuals does not depend on time. This assumption made by the Cox model is called the proportional hazard assumption. The Cox model is fitted in two steps. First, the η parameters are estimated by maximizing the Cox partial likelihood, which does not depend on the baseline hazard; and then the non-parametric baseline hazard $h_0(t)$ is estimated based on the parametric results.

This model is specifically adapted for censored data: if the follow-up period is fixed and there is no censored data, a logistic regression model would be as relevant as the Cox model to estimate the hazard rate. However, even in the context of censored data, the Cox model is limited by two weaknesses: first, it assumes that the influence of the covariates on the hazard rate does not depend on time (proportional hazard assumption), and it considers only linear interaction between the covariates. In clinical practice and research, these methods (and especially the Kaplan-Meier estimator) are significantly used, for example to study whether two groups of patients have different survival tendencies.

3.2.2 Learning-based survival analysis methods

Beyond traditional statistical methods, a growing research trend has focused on the development of machine learning survival analysis approaches aiming to capture complex non-linear relationships between covariates and patient outcomes. The methods currently used to learn to predict an individual's survival fall into three categories: *continuous-time* models which provide a survival value for any time t, *discrete-time* models, where the survival prediction is given only for a limited number of time intervals, or the *event time prediction* models which predict a single event time value. In the following, we will consider $t_c^{(n)}$ as the event time $E^{(n)}$ of patient n if not censored and to its censoring time $C^{(n)}$ otherwise.

3.2.2.1 Continuous-time methods

One of the first and most common survival models is called **DeepSurv** [Katzman et al. 2018]. This continuous-time method simply uses a deep learning model $\phi_{DeepSurv}$ to predict the log-risk component of a Cox model as follows:

$$\hat{h}(t|\mathbf{x}^{(n)}) = h_0(t)e^{\phi_{DeepSurv}(\mathbf{x}^{(n)})}. \tag{3.8}$$

The $\phi_{DeepSurv}$ parameters are learned with a negative partial log-likelihood loss function [Katzman et al. 2018]. Contrary to the original Cox model, DeepSurv can learn non-linear relationships between the covariates and the hazard rate. However, it is still constrained by the proportional hazard assumption and does not consider a time-dependent influence of the covariates.

[Kvamme, Borgan, and Scheel 2019] proposed **CoxTime**, an extension of DeepSurv that addresses these two problems. They do so by including a time variable along with the covariates in the prediction of the log-risk component:

$$\hat{h}(t|\mathbf{x}^{(n)}) = h_0(t)e^{\phi_{CoxTime}(\mathbf{x}^{(n)},t)}. \tag{3.9}$$

This method is thus able to predict a survival evolution of the patients that depends on the covariates in a way that is both non-linear and time-dependent. To learn the parameters of the model $\phi_{CoxTime}$, this approach relies on a loss function evaluating how well the patients are ordered. To do so, it compares the predicted survival of uncensored patients n at their event time $t_c^{(n)}$ to the predicted survival at the same time $t_c^{(n)}$ of a subset of patients at risk at $t_c^{(n)}$. A good ordering means that the survivals of the patients at risk at $t_c^{(n)}$ are higher than the survival of the patient having the event at this time. Formally, this gives the following loss function:

$$\text{loss}_{CoxTime} = \frac{1}{\#\{n|\delta^{(n)}=1\}} \overbrace{\sum_{n|\delta^{(n)}=1} \log\Big(\sum_{i\in\tilde{\mathcal{R}}^{(n)}} e^{\phi_{CoxTime}(\mathbf{x}^{(i)},t_c^{(n)})-\phi_{CoxTime}(\mathbf{x}^{(n)},t_c^{(n)})}\Big)}^{\text{Comparison of pairs of survival predictions at each time } t_c^{(n)} \text{ with an event}} + \lambda \underbrace{\sum_{n|\delta^{(n)}=1}\sum_{i\in\tilde{\mathcal{R}}^{(n)}} |\phi_{CoxTime}(\mathbf{x}^{(i)})|}_{\text{Regularization}} \tag{3.10}$$

with # the cardinal function and $\tilde{\mathcal{R}}^{(n)}$ a subset of the patients at risk at time $t_c^{(n)}$. This use of a subset of the patients at risk instead of considering all of them is the main difference between this loss function and the negative partial log-likelihood used to train DeepSurv, and reduces the computation cost of the non-proportional extension proposed in CoxTime. In our experiments making use of this loss function, we will consider only one sample in this risk set as in [Kvamme, Borgan, and Scheel 2019].

Another commonly used continuous survival model is the **Random Survival Forest (RSF)** [Ishwaran et al. 2008], which is an ensemble of tree-based models making survival predictions (see an example of non-survival Random Forest in Fig. 3.3). A tree is a prediction model built as a succession of decision nodes that partition the feature space into regions, enabling a hierarchical representation of relationships between input variables and output labels. In random survival forests, one tree is built by randomly selecting a few candidate variables for each node, and splitting it by using the candidate variable that maximizes the survival difference between daughter nodes. In the end, global predictions are formed by aggregating the survival predictions of the individual trees in the ensemble, thus improving the accuracy of the prediction and controlling overfitting. This RSF model differs from the previously presented continuous-time methods, but it is also commonly used for its good performance, and as it also deals, by construction, with non-linear relationships between the covariates and non-proportional hazards.

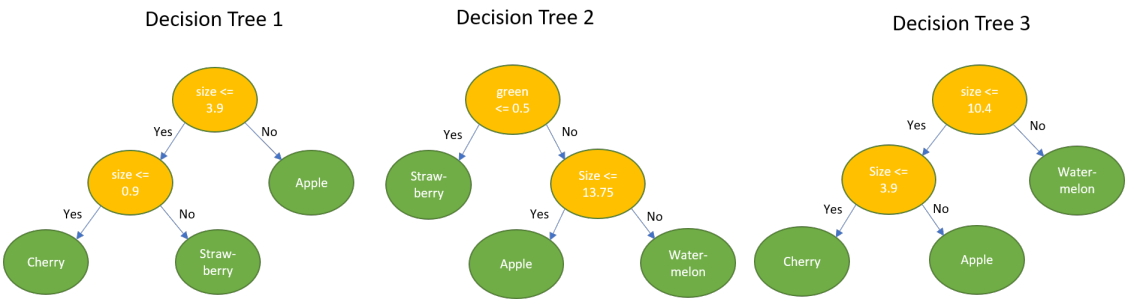


Figure 3.3 – Example of a (non-survival) Random Forest (from https://developer.nvidia.com/blog/accelerating-random-forests-up-to-45x-using-cuml/).

3.2.2.2 Discrete-time methods

Among the discrete-time methods, one of the first proposed approaches was the **Multi-Task Logistic Regression (MTLR)** model [Yu et al. 2011]. This model, trained with a regularized log-likelihood loss function, predicts the probability that an individual's event happens during or after a time interval, by using a distinct logistic regression model for each of the N_T time intervals:

$$\mathrm{P}_{\theta_i}(E^{(n)} \geqslant t_i | \mathbf{x}^{(n)}) = \frac{1}{e^{\theta_i^\top \mathbf{x}^{(n)} + b_i}}, i \in [\![1; N_T]\!]. \tag{3.11}$$

Thus, it is able to learn time-varying relationships between covariates and risks. The **Neural Multi-Task Logistic Regression (N-MTLR)** [Fotso 2018] extends this model by using Multi-Layer Perceptrons (MLPs) instead of logistic regressors for the different intervals, introducing non-linearity in the covariates' interaction.

A second widely-used discrete-time method is **DeepHit** [Lee et al. 2018], which has the specificity of dealing with multiple concurring events (**DeepHitSingle** being the counterpart for unique event settings). This approach predicts for each competing event and for each time interval a probability that the patient will experience the given event *during* this time interval with the model presented in Fig. 3.4. Thus, it can learn non-proportional and potentially non-linear relationships between covariates and risks. The loss function used to train this model is composed of two elements: first, the log-likelihood of the joint distribution of the first hitting time and event; and a combination of cause-specific ranking loss functions. By this, it constrains the model to order the patients correctly, while the MTLR and N-MTLR only focus on the precision of the prediction.

3.2.2.3 Event time value prediction

Instead of predicting a complete survival analysis curve, **RankDeepSurv** [Jing et al. 2019] predicts a single event time value based on a deep learning model. The training of this model relies on both a regression Mean Squared Error (MSE) loss, adapted to deal with censoring, and a ranking loss evaluating how well the patients' predictions are ordered. Other methods have been developed for single value survival prediction [Morvan et al. 2023], but overall these approaches are not adapted to our problem as we want to predict complete survival analysis curves (and especially the survival function S).

Among all these models, we choose in this work to use a continuous model to avoid having to arbitrarily decide upon a discretization scheme, *i.e.* a number of intervals and their distribution (equidistant or by quantiles): indeed, Kvamme and Borgan 2021

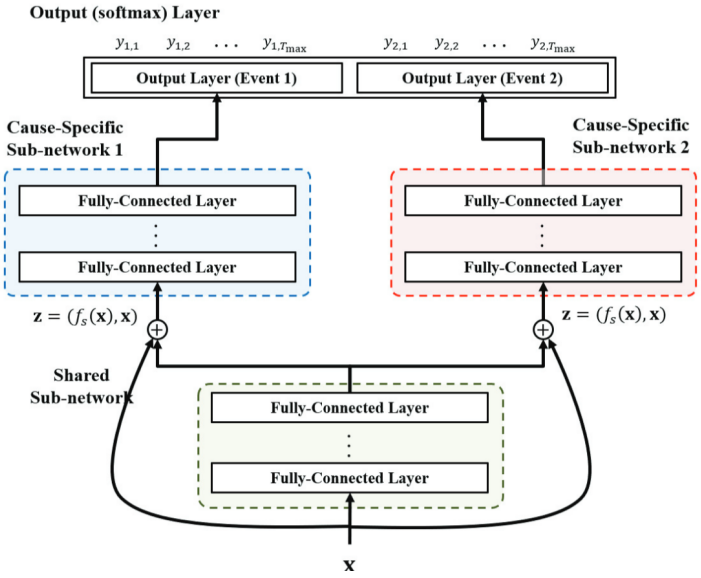


Figure 3.4 – Architecture of DeepHit with two competing events (from [Lee et al. 2018]).

showed that such discretization choices had a large impact on the performance of the methods. Furthermore, [Kvamme, Borgan, and Scheel 2019] conducted a comparative study including DeepSurv, DeepHit, CoxTime, and an RSF model for survival analysis prediction on multiple datasets. Their results highlighted the good performance of CoxTime in most cases, especially ranking first or second in all metrics on the METABRIC dataset which we will be using in Part IV. In a later study including these methods and others [Kvamme and Borgan 2021], CoxTime is still very competitive on the METABRIC dataset, further reinforcing its reliability. Given these strong results, and as the CoxTime model is not limited by a proportional hazard assumption and considers time-dependent and non-linear relationships between the covariates and the predicted output, we choose to rely on this method for survival training in Part IV.

3.3 Survival prediction evaluation metrics

To evaluate the quality of a survival prediction, we rely in this work on three common metrics. First, the c-index measures how well the model orders the patients. Then, the Brier score (and the corresponding integrated Brier score) takes into account calibration to evaluate the quality of each individual prediction. Finally, we compare the Kaplan-Meier curves of predicted risk groups to estimate the stratification ability of a given model.

3.3.1 C-index

The c-index, or concordance index [Harrell et al. 1982], is one of the most commonly used metrics to assess the quality of a survival analysis model. Essentially, it measures the probability that, for any two randomly selected individuals, the model's predicted survival times will be ranked the same as their true survival times (for an in-depth explanation, refer to [Ishwaran et al. 2008]). The original metric from [Harrell et al. 1982] compares the individuals' hazard rates without considering their time component, which is only adapted for proportional hazards models. In our analysis, we rely instead on a new metric derived from the *time-dependent c-index* introduced by [Antolini et al. 2005] which is able to deal with non-proportional hazards models. This time-dependent c-index corresponds to the probability that the survival times of two *comparable* individuals i and j are correctly ordered by the model as follows:

$$\text{c-index} = \mathrm{P}\{\hat{S}(t_c^{(i)}|\mathbf{x}^{(i)}) < \hat{S}(t_c^{(i)}|\mathbf{x}^{(j)})|t_c^{(i)} < t_c^{(j)}, \delta^{(i)} = 1\}. \tag{3.12}$$

Thus, a value of c-index closer to 1 indicates a better model.

To handle tied event times and survival estimates, we incorporate the adjustments specified by Ishwaran et al. 2008 (Section 5.1, step 3). These modifications ensure that if the predictions are independent of the covariates $\mathbf{x}$, i.e. $\hat{S}(t|\mathbf{x}) = \hat{S}(t)$, the c-index will be equal to 0.5 for unique event times. It is worth noting that for proportional hazards models, this adjusted metric is equivalent to the traditional c-index.

3.3.2 Brier score and integrated Brier score

The Brier Score (BS) is a metric closely related to binary classification. It evaluates both the discrimination and calibration of a model's estimates. In short, for N_{BS} binary labels $y^{(n)} \in \{0, 1\}$, the BS is the Mean Squared Error (MSE) of the probability estimates $\hat{p}^{(n)}$ that an individual n has a label $y^{(n)} = 1$, *i.e.* $\text{BS} = \frac{1}{N_{BS}} \sum_{n=1}^{N_{BS}} (y^{(n)} - \hat{p}^{(n)})^2$. To apply this metric for survival analysis predictions, we first need to convert time-to-event data into binary outcomes: for this, we select a specific time point t and label each individual's data based on whether their event time occurs before or after this threshold. Additionally, to deal with censoring, Graf et al. 1999 proposed an extension of this metric weighting the scores by the inverse censoring distribution. This approach increases the weight of patients with late labels, as fewer events can be observed at these times because of censoring, and therefore ensures a more balanced contribution to the MSE across time. These two modifications give the following expression for the BS computation:

$$\mathrm{BS}(t) = \frac{1}{N_{BS}} \sum_{n=1}^{N_{BS}} \Bigg[\overbrace{\frac{\hat{S}(t|\mathbf{x}^{(n)})^2 \mathbb{1}\{t_c^{(n)} \leqslant t, \delta^{(n)} = 1\}}{\hat{G}(t_c^{(n)})}}^{\text{Individuals whose event occurred before time t}} + \overbrace{\frac{(1 - \hat{S}(t|\mathbf{x}^{(n)}))^2 \mathbb{1}\{t_c^{(n)} > t\}}{\hat{G}(t)}}^{\text{Individuals at risk at time t}} \Bigg], \tag{3.13}$$

with N_{BS} being the number of observations, $\mathbb{1}$ the indicator function and $\hat{G}(t)$ the Kaplan-Meier estimate of the censoring survival function $\mathrm{P}(C > t)$. We assume that the censoring and survival times are independent.

The Brier Score can be extended from a single time t to a time interval $[t_1; t_2]$ by computing the Integrated Brier Score (IBS):

$$\mathrm{IBS} = \frac{1}{t_2 - t_1} \int_{t_1}^{t_2} \mathrm{BS}(t) dt. \tag{3.14}$$

In practice, we approximate this integral using numerical integration, setting the time span to match the duration of the test set. The number of grid points is fixed to 100 as in [Kvamme, Borgan, and Scheel 2019]. For both the BS and the IBS, the lower the score, the better the model.

3.3.3 P-values between the Kaplan-Meier curves of two groups of patients

In order to evaluate the quality of the patients' stratification into low- and high-risk groups, a common method is to first estimate the Kaplan-Meier curve of each group, then apply a log-rank test [Fleming et al. 1981] with the hypothesis that these curves are identical. The lower the obtained p-value is, the more strongly this hypothesis is rejected, and the better the model is for patient stratification.

The three metrics explained above will be used in Part IV of this thesis to compare the survival curves of a baseline survival model with those adjusted by our post-hoc correction framework.

Chapter 4

State of the art

This chapter reviews the state of the art on which the contributions of the following parts rely. First, we explain recent advances related to the automatic interpretation of DLBCL patients' information. Then, we review the state-of-the-art methods existing for multimodal data fusion. Finally, a complementary section lays the theoretical foundation for graph-based approaches and their applications in the medical field.

4.1 DLBCL analysis

Several automatic methods have been applied in the literature to support clinicians in the analysis of the information from DLBCL patients, relying on tabular or imaging data. Tabular clinical data, including information such as blood test results or disease staging, is typically used to support the DLBCL diagnosis and subsequent follow-up in clinical studies, using classical but interpretable methods [Le Gouill et al. 2021].

In the field of PET image analysis, trends are to automatically extract quantitative information (radiomic features) to train predictive machine learning methods [Carlier et al. 2024; Eertink et al. 2022; Jiang et al. 2022; Yousefirizi, Gowdy, et al. 2024] or to directly use deep learning methods on the image [Ferrández et al. 2023; P. Liu et al. 2022; Yuan et al. 2022]. Despite knowing that the dispersion of the lesions is prognostic for the pathology [Cottereau et al. 2019], the above approaches do not explicitly account for multiple lesions nor their spatial distribution, or only through simplified measures such as the largest distance between lesions. Furthermore, only a few methods incorporate both imaging and tabular clinical information, and they do so through basic fusion schemes [Carlier et al. 2024; Eertink et al. 2022; Jiang et al. 2022].

In this context, we instead develop a learning-based method to automatically identify high-risk DLBCL patients at diagnosis by efficiently fusing all available information, including the tabular clinical data, the imaging information from all the lesions of the patients and their spatial distribution.

4.2 Multimodal fusion

Beyond the scope of DLBCL diagnosis and follow-up, there is a wide variety of methods to fuse the information coming from different data modalities, ranging from classical early, late, and hybrid fusion approaches to kernel-based methods, graphical models and neural networks [Baltrušaitis et al. 2019].

Deep learning is now widely used in medical image analysis for both decision-making and feature extraction, with a majority of methods dedicated to image-to-image fusion problems [T. Zhou et al. 2023]. Recently, there has been considerable interest in the fusion of medical images with text, *e.g.* from clinical reports. For instance, [Z. Wang, Z. Wu, et al. 2022] achieved the contrastive pre-training of a large text-image CLIP (Contrastive Language-Image Pre-training) model from [Radford et al. 2021] exploiting unpaired medical data.

Less work has focused on the deep learning fusion of tabular and medical image data. For instance, [Pölsterl et al. 2021] addresses the problem by learning affine transformations that dynamically modify convolutional neural network (CNN)-based image features according to tabular data. Also, [Hager et al. 2023] propose a contrastive learning framework with a CLIP loss to align image and tabular embeddings in the context of cardiac classification tasks on the large UK Biobank dataset [Sudlow et al. 2015]. After pretraining, the prediction tasks rely on a concatenation of the latent vectors of unimodal autoencoders. Similarly, [W. Huang 2023] creates a joint embedding of labels and images, followed by a contrastive alignment of tabular data with the image embedding considered as the main modality. During inference, the embeddings of different modalities are concatenated and directly compared with the label encoding for zero-shot predictions.

Despite these recent advances, there is still no widely accepted method for dealing with image-tabular data fusion for medical data. In particular, contrastive multi-modal fusion frameworks are difficult to implement for our targeted problem, even as pre-training, due to the presence of censored data and the lack of large publicly available PET-tabular datasets. Moreover, the relevant information in PET images of DLBCL patients is concentrated in a varying number of multiple small and sparsely distributed lesions. It is therefore inefficient to process the whole image directly as input, and the approaches mentioned above fail to process the varying number of lesions considered independently, as would be the case with concatenation operations. Instead, we aim at a unique dynamic encoding that can produce an adaptive multi-modal embedding of each lesion, and at the same time is able to integrate the information across the variable number of lesions.

Unlike classical contrastive methods that rely on static encoders (for all [Hager et al.

2023] or at least one modality [W. Huang 2023]), we opt for a dynamic adaptation of the multimodal embedding, similar in idea to [Pölsterl et al. 2021] but relying on a cross-and-self-attention mechanism instead of an affine transformation. Such transformer-inspired cross-attention modules (presented in Sec. 4.3.5) have been investigated for multi-modal fusion for non-medical [Nagrani et al. 2021], as well as medical data, as in [Gao et al. 2023; G. Wang et al. 2024] and reviewed in [Andrade-Miranda et al. 2023]. As with contrastive methods, most existing approaches fuse modalities with similar dimensions, with some exceptions [Gao et al. 2023; Golovanevsky et al. 2022]. In our case, we are faced with the fusion of modalities with heterogeneous dimensions. More importantly, none of the methods reviewed considers multiple lesions of a patient individually.

For the reasons presented above, we propose to combine our cross-modal fusion presented in Part III with a lesion graph representation introduced in Part II, which is able to explicitly model the variable number of lesions of each patient individually, as explained in the following section.

4.3 Graphs for medical applications

4.3.1 Basics on graphs

Graphs are a kind of data representation used to describe and analyze entities with relationships or interactions between one another. They are usually defined by a set of ***nodes*** $\mathcal{V}$, where any two nodes $v_i, v_j \in \mathcal{V}$ can be connected to one another by an ***edge*** $e_{ij} \in \mathcal{E}$. A node can also be connected to itself: such an edge is called a ***self-loop***. To enrich this representation, it is possible to assign a ***feature vector*** $\mathbf{z_i}$ to each node i, and/or to define ***edge weights***, *i.e.* a value w_{ij} on each edge indicating how strong the connection is between the two corresponding nodes v_i and v_j. For a visual example of these concepts, see Fig. 4.1.

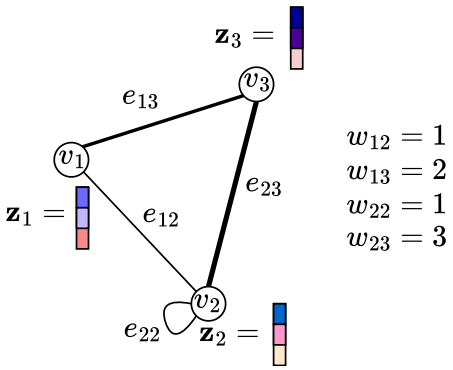


Figure 4.1 – Example of graph.

4.3.2 Patient graphs

In the biomedical field, multiple ways of representing a single patient as a graph have been proposed, depending on the task at hand (for example, to represent the brain [Sarkis et al. 2025] or the blood vessels of a patient [Kassam et al. 2022]). For this work, one of our motivations to use a graph is to explicitly represent each lesion while being able to learn predictors across patients with a variable number of lesions. This calls for what is named a lesion graph, where each lesion of a patient is represented as a node: thereby, the image features of each lesion as well as their structural relationships are jointly encoded.

Such lesion graphs have been explored in [Kazmierski et al. 2021] and [Lv et al. 2023] in combination with a Graph Neural Network (GNN) to predict the probability of distant metastasis over time and the PFS of head and neck cancer, from CT scans and both CT and PET scans, respectively. [Prabhakar et al. 2023] also rely on a lesion graph and a GNN to predict inflammatory activity in multiple sclerosis, while [Rist et al. 2023] present a comparison of multiple lesion graphs and learning processes to estimate the degree of metastatic progression in liver cancer, and [Rochman et al. 2024] propose a longitudinal lesion graph to classify lesion changes over time on CT scans. In our case, we rely in Parts II and III on the work from [Aswathi et al. 2023] and their lesion graph based on PET images of DLBCL patients used for 2-year PFS prediction. In this framework, the feature vector of a node corresponds to imaging features extracted from the corresponding lesion on the PET scan, all pairs of nodes are connected by edges including self-loops, and edge weights are defined in a way that favors stronger connections between closer and more similar lesions.

Despite the rich literature on the subject, the integration of multimodal data with such lesion graphs has not been fully explored. While [Aswathi et al. 2023; Prabhakar et al. 2023; Rist et al. 2023; Rochman et al. 2024] do not address multi-modal fusion, [Kazmierski et al. 2021; Lv et al. 2023] only use a simple concatenation of a clinical data vector with the output of the GNN before the prediction layer (naive late fusion). In general, few graph-based methods have addressed the fusion of tabular and imaging data. One exception is the work of [S. Liu et al. 2023], which relies on a cross-attention module to fuse a patient graph representing the brain of a post-stroke patient with a set of tabular clinical data. However, their approach is based on a fixed graph structure and is therefore incompatible with a variable number of lesions.

To address the aforementioned limitations and challenges, we propose in Part III to combine the lesion graph from [Aswathi et al. 2023] with a cross-attention module, allowing for an efficient fusion with clinical data and accommodating a variable number of lesions.

4.3.3 Population graphs

Shifting from the individual level to the population level, another type of graph, known as a population graph, can be constructed by representing each patient as a node. By leveraging similarities between patient profiles, representing the data as population graphs has already demonstrated effectiveness in tasks such as brain disorder classification and brain age estimation [Bintsi et al. 2023; Zhang et al. 2023].

However, how to connect the nodes in such a graph is still an open question in the field. A lot of frameworks rely on static graphs where the connectivity is fixed once and for all. Those are built without any data-driven learning, either based on prior knowledge of what 'similar' patients are [Parisot et al. 2018], or by computing some sort of similarity between the feature vectors of the patients [Kim 2023; Larose, Touma, et al. 2022]. Overall, many approaches have been proposed, based on either static graphs or graphs dynamically learned and optimized for the task at hand, as reviewed in [Mueller et al. 2024].

Population graphs, whether static or dynamically learned, can be exploited in multiple ways for learning tasks. The most common approach is to propagate the *features* of the patients on the graph with a Graph Neural Network (GNN) (more details on these are given in the next section), and then make a prediction based on the updated feature vector of each patient [Bintsi et al. 2023; Zhang et al. 2023; Zheng et al. 2022]. Another class of methods focuses on label propagation, where, as the name suggests, *labels* themselves are propagated in a semi-supervised manner [Ghorbani et al. 2021; D. Zhou et al. 2003]. For this method to be effective, patients with similar labels should be more strongly connected in the underlying population graph. This property is called *homophily*. A third category of approaches leverages population graphs to mitigate label noise, for example by using the graph to compute uncertainty on the labels to keep only the robust ones [L. Chen et al. 2020]. Finally, a more recent line of work, termed *residual propagation* or *prediction correction*, assumes a trained model that generates predictions from individual feature vectors. The idea behind this method is to leverage similarities between subjects to identify and correct biases in this model's predictions. To this end, these approaches propagate the residuals, *i.e.* the prediction error, across samples in order to correct the predictions of the initial model, leading to an increase in its performance [Q. Huang et al. 2020; Raymond et al. 2024; Yang et al. 2023].

In this PhD, we explored multiple strategies including both static and dynamically-learned graphs, and label propagation and prediction correction methods. In the following, we present only the most conclusive one, which is a prediction correction approach developed in Part IV. We focus on this method with the aim of applying it to our multimodal model presented in Part III, to fully integrate all the information we have

for the high-risk DLBCL patient identification task in one global framework.

For this work, we switched from the 2-year PFS classification task to a survival analysis task, as the second is more informative for the identification of high-risk patients. However, *residual propagation* methods have been applied to regression [Raymond et al. 2024], classification [Q. Huang et al. 2020; Yang et al. 2023] or language modeling tasks [Yang et al. 2023], but to the best of our knowledge, the application of post-hoc correction methods on survival predictions using population graphs has never been done. The use of population graphs in this context improves the identification of high-risk patients in a way that distinguishes them from traditional calibration techniques [Silva Filho et al. 2023]. Indeed, while some calibration approaches also apply post-hoc corrections to predictions, their goal is fundamentally different: calibration adjusts the overall distribution of predicted probabilities without altering their relative order. In contrast, our objective is to modify this order to better separate low- and high-risk patients.

Based on the aforementioned insights, we develop a novel framework to take a survival analysis model used for risk stratification, and to improve a posteriori its ability to divide patients into low- and high-risk groups by taking advantage of the residuals of similar patients. Our proposition, presented in Part IV, is based on the Error Passing Network (EPN) from [Raymond et al. 2024]. We chose to rely on this specific model as it is designed for regression tasks, which is the closest we found to our objective. Details on the construction of the population graph and the error propagation scheme are provided in Part IV.

4.3.4 Graph neural networks

Generally, in the same way one can learn to extract relevant information from tabular or imaging data relying on neural networks, support vector machine (SVM) or random forests for example, some models have been designed to learn a task by taking full advantage of all the information of a graph. Indeed, standard multi-layer perceptron (MLP) and convolutional neural networks (CNN) models for images, for example, rely on the assumption that the data lies on a regular grid, and are therefore not adapted to handle irregular structures such as graphs. Thus, there is a need to design operations on the graph that consider both the feature and structure information. Models doing so are named ***Graph Neural Networks*** (GNNs): see [Z. Wu et al. 2021] for a review of some of them. In this work, we focus on the two following models built on convolution schemes.

The first considered model is the common **GraphConv** [Morris et al. 2019]. For each node, the **GraphConv** operation aggregates the information from the neighbours of the node and itself, by weighting the representations of the neighbouring nodes with

the corresponding edge weight before summing and projecting them. This aggregated representation of the neighbourhood of a node is then added to the projection of its own feature vector. Formally, this module updates the feature vector $\mathbf{z}_i$ of a node v_i as follows:

$$\mathbf{z}_{i_{\text{GraphConv}}} = \mathbf{\Theta}_1 \mathbf{z}_i + \mathbf{\Theta}_2 \left(\sum_{j|v_j \in \mathcal{N}(i)} w_{ij} \mathbf{z}_j \right), \tag{4.1}$$

with $\mathbf{\Theta}_1$ and $\mathbf{\Theta}_2$ being learned projection matrices and $\mathcal{N}(i)$ the neighboring nodes of v_i.

The second GNN used in this work is the **GATv2** layer [Brody et al. 2022], whose strength lies in its ability to update each node based only on the information propagated from the most relevant neighbouring nodes. To identify the latter, this module learns to compute attention weights $\rho_{i,j}$ between each pair of connected nodes i and j as follows [1]:

$$\rho_{i,j} = \frac{\exp(\mathbf{a}^T \,\text{LeakyReLU}(\mathbf{\Theta}_s \mathbf{z}_i + \mathbf{\Theta}_t \mathbf{z}_j + \mathbf{\Theta}_e w_{ij}))}{\sum_{k|v_k \in \mathcal{N}(i) \cup \{i\}} \exp(\mathbf{a}^T \,\text{LeakyReLU}(\mathbf{\Theta}_s \mathbf{z}_i + \mathbf{\Theta}_t \mathbf{z}_k + \mathbf{\Theta}_e w_{ik}))}, \tag{4.2}$$

with $\mathbf{a}$, $\mathbf{\Theta}_s$, $\mathbf{\Theta}_t$ and $\mathbf{\Theta}_e$ learned parameters. Specifically, $\mathbf{\Theta}_s$ learns to project the representation of the *source* node, $\mathbf{\Theta}_t$ the representation of the *target* node (or the neighbour for two different nodes), and $\mathbf{\Theta}_e$ to project the representation of the *edge weight*. The updated feature vector of a node is then computed as:

$$\mathbf{z}_{i_{\text{GAT}}} = \rho_{i,i} \mathbf{\Theta}_s \mathbf{z}_i + \sum_{j|v_j \in \mathcal{N}(i)} \rho_{i,j} \mathbf{\Theta}_t \mathbf{z}_j. \tag{4.3}$$

In the following, we refer to this layer simply as a 'GAT' layer.

The GNN models can be applied to different tasks: from node prediction, where we try to predict a class or value for each of the nodes; to graph prediction, to classify multiple graphs one by one; or even edge prediction to learn which nodes should be connected. In our case, the two GNNs presented will be applied in Parts II and III to a graph prediction task, as we want to classify the patients with each of them being associated with one lesion graph.

4.3.5 Transformer attention process

Finally, beyond graphs and GNNs, another way of propagating information between samples is to use transformers [Vaswani et al. 2017], particularly through their attention module. This module takes as input a query $\mathbf{Q} \in \mathbb{R}^{\mathcal{H}_Q \times \mathcal{W}_Q}$ and a key/value pair $\mathbf{K} \in \mathbb{R}^{\mathcal{H}_K \times \mathcal{W}_K}$ and $\mathbf{V} \in \mathbb{R}^{\mathcal{H}_V \times \mathcal{W}_V}$ (with $\mathcal{H}_K = \mathcal{H}_V$); and it returns a weighted sum of the

1. This implementation corresponds to the `torch_geometric` version of the GATv2 operator.

projected values $\mathbf{V}$. The weight assigned to each value is computed from a compatibility function, corresponding to a scalar product, of a projection of the query with the corresponding projected key (all this normalized by the key dimension). Formally,

$$\begin{aligned}\mathbf{Z}^{(n)}_{\text{CrossAtt}} &= \text{CrossAtt}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) \\ &= \text{softmax}\left(\frac{\bar{\mathbf{Q}}\bar{\mathbf{K}}^\top}{\sqrt{\mathcal{W}_{\bar{K}}}}\right)\bar{\mathbf{V}}, \end{aligned} \tag{4.4}$$

where $\bar{\mathbf{Q}} = \mathbf{Q}\mathbf{W}_{CA,Q}$, $\bar{\mathbf{K}} = \mathbf{K}\mathbf{W}_{CA,K}$ and $\bar{\mathbf{V}} = \mathbf{V}\mathbf{W}_{CA,V}$ are the learned latent representations via three learnable weight matrices $\mathbf{W}_{CA,Q} \in \mathbb{R}^{\mathcal{W}_Q \times \mathcal{W}_{\bar{Q}}}$, $\mathbf{W}_{CA,K} \in \mathbb{R}^{\mathcal{W}_K \times \mathcal{W}_{\bar{K}}}$ and $\mathbf{W}_{CA,V} \in \mathbb{R}^{\mathcal{W}_V \times \mathcal{W}_{\bar{V}}}$. The output of the cross-attention operation is a matrix $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ of size $\mathcal{H}_Q \times \mathcal{W}_{\bar{V}}$. Intuitively, the matrices $\mathbf{W}_{CA,Q}$ and $\mathbf{W}_{CA,K}$ project the query and key information into a common space, before computing their compatibility. The softmax operation then provides the attention scores that should be given to each element of the value matrix. Finally, these attention scores are multiplied by the value matrix which is also projected into a relevant space by $\mathbf{W}_{CA,V}$.

When the query, key and value matrices correspond to the features of the different samples, the attention computation corresponds to creating a fully connected graph structure between these samples, with self-loops and with edge weights corresponding to the attention weights computed. It is possible to enforce sparsity in this graph by applying a mask on these attention weights. Multiplying the attention/graph connectivity matrix by the value matrix then corresponds to propagating the projected samples' features on the graph.

We will explore in Part III if learning the graph structure with such an attention process applied to all the lesions performs better than using the pre-defined graph structure combined with the GNNs presented above. This attention process is also used for the multimodal fusion in the same part, by considering as inputs the different data modalities we have.

Part II

Lesion graph for a better use of DLBCL PET images

Abstract

In this first contribution part of this thesis, we focus on leveraging PET imaging data to improve the identification of high-risk DLBCL patients. Central to this work is the introduction of a lesion graph-based representation, specifically designed to account for the presence of multiple lesions per patient and their spatial relationships. To assess the value of this approach, and as there is currently no consensus on the features to extract from PET DLBCL images for high-risk patients identification, we construct and compare five imaging feature sets derived from three distinct radiomic configurations applied to different lesion segmentation strategies. We also compare various modeling strategies that differ in how they handle the variable number of lesions across patients: using only a single representative lesion, incorporating all lesions without spatial context, or additionally integrating spatial distribution through the lesion graph. These approaches are evaluated based on their ability to predict 2-year progression-free survival (PFS) using the GAINED dataset.

5.1 Introduction

As explained before, leveraging PET images for an automatic DLBCL risk stratification task is not trivial, as the information in the PET volumes of DLBCL patients is spread over multiple, typically small lesions, making feature extraction difficult. Current methods do not explicitly account for these multiple lesions, and overlook the prognostic information contained in their spatial distribution (or only consider it through simple measurements). Furthermore, in these images both the number of lesions and image resolution can vary significantly between patients, making generalization difficult.

To extract features from PET scans, a first option is to feed the whole body images directly into a CNN [Ferrández et al. 2023; P. Liu et al. 2022; Yuan et al. 2022]; but this option comes with a risk of introducing irrelevant information as lesions occupy only a small part of the image, and such CNN features are difficult to interpret. We therefore pursue a second strategy based on the extraction of radiomic features (see Appendix B.2 for a simple introduction to radiomics) from the segmented lesions to preserve only relevant information while improving interpretability [Carlier et al. 2024; Eertink et al. 2022; Jiang et al. 2022; Yousefirizi, Gowdy, et al. 2024]. However, there is currently no consensus in the community as to which ensemble of radiomic features is the most predictive for the risk stratification task. Furthermore, the extracted features' relevance also depends on the lesion segmentation strategy. To overcome those challenges, we explore in the following the performance of three common types of segmentation strategies as well as three sets of radiomic features on the DLBCL risk stratification task.

To broaden the conclusions of this evaluation, we compare the proposed feature sets and segmentation strategies using a range of baseline classification models. This diversity of models enables us to explore and compare various methods for incorporating imaging features from the multiple lesions present in each patient. Specifically, we examine approaches that either average the lesion features, rely solely on the largest lesion, apply a Multiple Instance Learning (MIL) framework, or utilize the lesion graph strategy proposed in [Aswathi et al. 2023]. Notably, the lesion graph is the only method that explicitly captures the spatial distribution of lesions.

5.2 Method

Let's consider a DLBCL clinical examination of a patient n at baseline before treatment, consisting of a set of tabular clinical indicators $\mathbf{c}^{(n)} \in \mathbb{R}^{D_{\text{clin}}}$ and a whole-body PET scan $\mathcal{I}^{(n)}$. After manual segmentation of this scan by clinicians, we get a full-body

3D PET image with 3D lesion segmentation $\mathcal{I}_{seg}^{(n)}$. To understand how to best make use of this PET scan for high-risk DLBCL patient identification, we target a progression-free survival (PFS) 2-year classification, intended to predict whether the disease of a patient will progress within two years after the beginning of the treatment. Five distinct sets of features extracted from $\mathcal{I}_{seg}^{(n)}$, are used as input to five different models. We present in the following these feature sets explored as well as the models we compare in this study.

5.2.1 Choice of PET imaging features

This experiment first requires extracting image features from each of the $L^{(n)}$ lesions appearing on the PET scan of the n^{th} patient. This imaging information is composed of both conventional PET features (see Appendix B.3 for more details) and radiomic features. The difference between the feature sets considered lies in the choice of radiomic features concatenated to the conventional ones, which are included in all sets. For the selection of those radiomics, it is necessary to take into account the lesion segmentation strategy considered, as it heavily influences the value of the radiomic features extracted. Thus, we present in this section the three common types of segmentation strategy as well as the three sets of features we compare in this study.

5.2.1.1 Segmentation strategies

The segmentation strategies considered include raw bounding boxes, majority voting, and threshold-based segmentation. In more detail, the respective binary segmentation masks are obtained from:

— the raw polygonal bounding boxes drawn by clinicians ('Raw' in Table 5.1);

— a majority vote ('Majority vote' in Table 5.1) between three automatic segmentations computed on the raw bounding boxes combining i) a K-means clustering with K = 2, ii) a thresholding to retain only voxels with intensity values greater than 40% of the maximum intensity, and iii) a second thresholding to retain voxels whose normalized SUV (Standardized Uptake Value) is greater than 2.5;

— a normalized SUV thresholding on the bounding boxes with a threshold set to 4 ('SUV4' in Table 5.1) [Malmon et al. 2025].

5.2.1.2 Sets of radiomic features

Depending on the segmentation strategy, we investigate three sets of radiomic features extracted from each segmented lesion. These features, after extraction, are included in the vector $\mathbf{z}_i^{(n)} \in \mathbb{R}^{D_{\text{img}}}$ assigned to the lesion i of a patient n, where D_{img}

is the dimension of the vector including conventional and radiomic features. First, we consider a set of $D_{img} = 15$ features inspired by the work of Kazmierski et al. 2021 on head and neck cancer ('HeadNeck' in Table 5.1). The second option investigated is the 'FirstOrder' set of features, for which we remove any second-order feature as they are known to be unreliable when computed on too small lesions [Zwanenburg et al. 2020]. In this set, we additionally change the first-order features considered and increase the quantity of shape features, to create a set of $D_{img} = 34$ features. The last feature set considered ('All' in Table 5.1) is a full set of $D_{img} = 650$ image features[2]. These radiomics can only be computed on raw bounding boxes to avoid computing texture and statistics on regions that are too small, which, as explained before, is known to be unreliable. For an explicit description of the features considered in the three sets, as well as the pre-processing steps applied, see Appendix B.3.

5.2.1.3 Complete feature sets explored

Five combinations of these segmentation strategies and feature sets were investigated. First, the Featv1 feature set corresponds to 'HeadNeck' features extracted from lesions segmented by majority vote. The Featv2, Featv3 and Featv4 feature sets are all composed of 'FirstOrder' features but are extracted from raw bounding box, majority vote and thresholding segmentations respectively. Finally, the Featv5 set consists of the full 650 imaging features computed directly on the bounding boxes of the lesions.

5.2.2 Models considered to study the usage of PET image data

Once the selected features have been extracted from the PET images, we measure the influence of this feature set selection on the 2-year PFS prediction task. To this end, we compare several baseline models that deal differently with the variable number of lesions of the patients. We present in the following the lesion graph used for the graph-based approaches as well as the full list of models we investigated.

5.2.2.1 Lesion graph creation

In order to efficiently represent the imaging information extracted from multiple lesions, we propose to rely on a lesion graph defined in [Aswathi et al. 2023]. More precisely, we build a fully connected graph $\mathcal{G}_l^{(n)} = \{\mathcal{V}_l^{(n)}, \mathcal{E}_l^{(n)}\}$. Each of its $L^{(n)}$ nodes $v_i^{(n)} \in \mathcal{V}_l^{(n)}$ corresponds to a single lesion i, from which image features are extracted

2. For the sake of clarity, the exhaustive list of these features is not developed in this document, but it can be found in Supplementary material 2 of our article [Thiery, Rizkallah, Bailly, et al. 2025] at the following link: https://www.sciencedirect.com/science/article/pii/S0895611124001587?via%3Dihub

and represented in the feature vector $\mathbf{z}_i^{(n)} \in \mathbb{R}^{D_{\text{img}}}$. From now on, we will denote by $\mathbf{Z}^{(n)} \in \mathbb{R}^{L^{(n)} \times D_{\text{img}}}$ the matrix concatenating row-wise all the node features $\mathbf{z}_i^{(n)}$.

In terms of underlying graph connectivity, we draw edges $e_{ij}^{(n)} \in \mathcal{E}_l^{(n)}$ between any pair of nodes $v_i^{(n)}$ and $v_j^{(n)}$ and include self-loops. To favor message passing between closer and more similar lesions, weights $w_{ij}^{(n)}$ are assigned to each edge, their values being defined based on the distances between both the lesion centroids ($\mathbf{p}_i^{(n)}$, $\mathbf{p}_j^{(n)}$) and their feature vectors ($\mathbf{z}_i^{(n)}$, $\mathbf{z}_j^{(n)}$):

$$w_{ij}^{(n)} = e^{\left(-\frac{||\mathbf{p}_i^{(n)} - \mathbf{p}_j^{(n)}||_2}{\gamma \sigma_1^2}\right)} \cdot e^{\left(-\frac{||\mathbf{z}_i^{(n)} - \mathbf{z}_j^{(n)}||_2}{\gamma \sigma_2^2}\right)}, \tag{5.5}$$

where $||.||_2$ stands for the L2 norm; γ is a hyperparameter tuned to find the best edge weight distribution for our task; and σ_1, σ_2 denote the population-level standard deviations of the centroid and feature distances, respectively.

5.2.2.2 Baseline classification models

To compare the feature sets selected on the 2-year PFS prediction task, we use them as input for multiple models dealing differently with the variable number of lesions of the patients. We investigate models which either consider only one lesion among all, average the features of all lesions, or consider them individually with or without taking into account their spatial distribution.

They are defined as follows:

— **MLP**: This model consists of two linear layers with ReLU (Rectified Linear Unit) activations and a 1-dimensional linear output layer with sigmoid activation. The two intermediate layers have the same dimension, chosen by a grid search. This MLP model allows us to see how predictive the image modality is for the given task with a very simple approach. However, as it requires fixed-size inputs, the lesion image information must be reduced to a fixed-size vector, which we do either by averaging the features from all lesions or by considering only the data from a single lesion.

— **MIL**: A Multiple Instance Learning approach that takes as input the image features from a patient's $L^{(n)}$ lesions, applies a one-layer MLP followed by a ReLU to each lesion's feature vector, aggregates the results by a maximum operation, and projects them linearly (with a sigmoid activation) to obtain the prediction. This approach allows us to use the image information from each lesion individually, but without using the topological information from the graph.

— **GraphConv**: A GNN approach that takes as input the lesion graph presented in Sec. 5.2.2.1. It consists of two GraphConv [Morris et al. 2019] convolutional

layers presented in Sec. 4.3.4, the first with an output dimension determined by grid search and followed by ReLU activation, and the second with an output size of 1 and followed by max pooling and sigmoid activation to predict the PFS. Both this model and the following GAT aim to explore the interest of graph structure over just considering lesion image features, as the MIL does.

— **GAT**: A model with the same structure as the GraphConv model, but with GAT layers [Brody et al. 2022] as presented in Sec. 4.3.4 instead of GraphConv layers.

5.3 Experimental validation

5.3.1 Dataset and experimental setup

For this study, we rely on the GAINED dataset as described in Part I, chapter 2.

With this dataset, the exploitation of the PET image of a patient n begins with the lesion detection, performed manually by a clinician to get $\mathcal{I}_{seg}^{(n)}$. The lesions are then segmented within the initial bounding box using the different approaches presented in Sec. 5.2.1.1. Before computing the radiomic imaging features, the segmented lesions are resampled to the same voxel size ($2\times2\times2$ mm) using a bicubic spline interpolation. After optional pre-processing (see Appendix B.3), the imaging features are extracted and standardized by removing the mean and scaling their variance to unity. Normalization of lesion position during the graph creation is performed in the same way, but individually for each patient to account for differences in patients' size.

For the experiments in the following section, we rely on a cross-validation scheme and a balanced sampling strategy, with a fixed random seed over all experiments. To account for random uncertainty in the data, we report the average of the results over five random train-validation/test splits. For each of these five sets of test patients, we perform ten loops of training and validation, with random splits of the patients in the training and validation sets (from those not in the test set) (see Fig. 5.2). The results obtained with this process are thus robust to data variations. The proportions of training, validation and test patients in each loop are 80%, 10% and 10% respectively, and we force the proportion of positive patients to be the same in all sets. Furthermore, to ensure that the scores are computed on balanced sets, we repeat the validation and test phases five times as presented in Fig. 5.3: for each run, we build a balanced validation set (resp. test set) with the available positive data and a 1/5 randomly sampled subset of the negative data in the validation set (resp. test set), where this ratio corresponds to the ratio of positive to negative samples in the dataset. We then average the metrics of these five balanced runs to produce the final validation or test results. The metrics we report are the 2-year PFS classification performance of the

evaluated methods in terms of AU-ROC, based on the presented computation process. Thus, the results presented in this paper correspond to the test performance, averaged over the five train-validation/test splits and the ten balancing loops, of the model with the best validation AU-ROC of the loop.

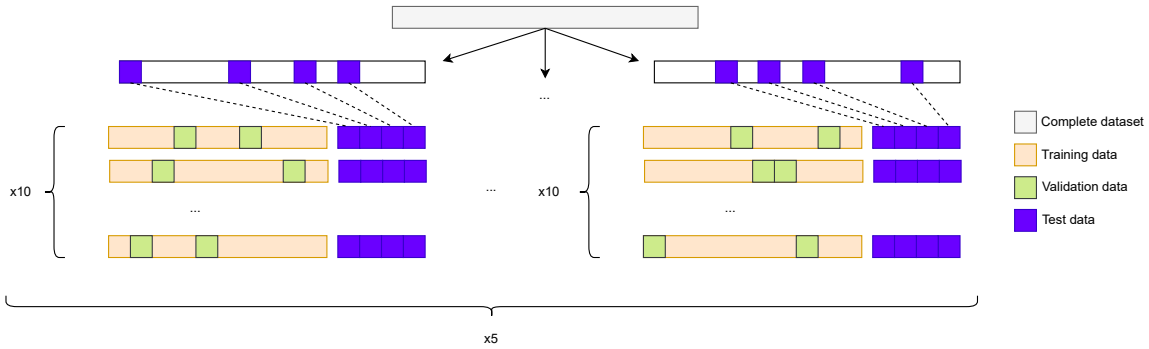


Figure 5.2 – Cross-validation process. A total of five random train-validation/test splits are created. For each of the five sets of test patients, ten loops of training and validation are performed.

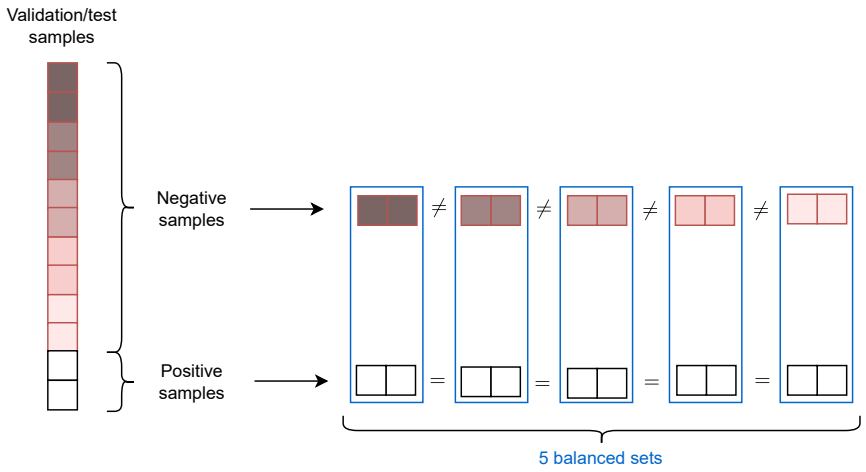


Figure 5.3 – Creation of a balanced validation/test set. We divide the negative samples into five sets, each of which is combined with all the positive samples to create balanced validation/test sets for testing.

Learning is controlled by a weighted binary cross-entropy loss function, with weights compensating for class imbalance (ratio of positive to negative samples $\sim 1:5$). The models are optimized via Adam [Kingma et al. 2014] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$; and this optimization includes a learning rate scheduler which multiplies the learning rate by 0.5 if the training loss did not decrease for 5 epochs, without

going lower than 10^{-6}. A grid search is performed on all evaluated models by selecting the configuration with the best validation AU-ROC. These grid searches explore the space of hyperparameters, including the learning rate, the output dimension of the projection layers (Fully Connected (FC), GraphConv or GAT layers) or, for the GNN methods, the parameter γ used in the construction of the lesion graphs. The full set of hyperparameters explored is presented in Appendix C.

The whole framework was coded in Python using the `PyTorch` [Paszke et al. 2019] and `pytorch-geometric` [Fey et al. 2019] modules, and the radiomics were extracted from the lesions using `pyradiomics` [Griethuysen et al. 2017].

5.3.2 Results

To the best of our knowledge, we are the first to report performance of DLBCL multi-lesion 2-year PFS prediction: thus, we only compare our results between the different models and image features selected. The results of this comparison are presented in Table 5.1.

	r	Featv1	Featv2	Featv3	Featv4	Featv5
Segmentation		Majority vote	Raw	Majority vote	SUV4	Raw
Image features		HeadNeck	FirstOrder	FirstOrder	FirstOrder	All
MLP	1	0.53 ± 0.08	0.52 ± 0.12	$\mathbf{0.56 \pm 0.09}$	$\mathbf{0.56 \pm 0.11}$	0.50 ± 0.09
MLP	2	$\mathbf{0.61 \pm 0.11}$	0.52 ± 0.08	0.53 ± 0.10	0.51 ± 0.09	0.56 ± 0.08
MIL	3	0.52 ± 0.10	0.46 ± 0.09	0.51 ± 0.10	0.50 ± 0.08	0.55 ± 0.12
GraphConv	3	0.54 ± 0.11	$\mathbf{0.56 \pm 0.11}$	$\underline{0.54 \pm 0.09}$	$\underline{0.54 \pm 0.09}$	$\mathbf{0.61 \pm 0.07}$
GAT	3	$\underline{0.55 \pm 0.10}$	$\underline{0.54 \pm 0.10}$	0.53 ± 0.10	0.51 ± 0.11	$\underline{0.59 \pm 0.10}$

Table 5.1 – AU-ROC results for the DLBCL multi-lesion 2-year PFS prediction with PET images. $\boldsymbol{r}$ indicate which radiomic data is used for the model tested: (1) means that we consider only the data from the largest lesion; (2) that we average the imaging data from all lesions; and (3) that we consider the data from each lesion individually. For a set of features, the best model is in bold and the second is underlined.

First, we observe from the low performances reached that predicting the 2-year PFS relying only on PET imaging features is a complex task, whether we use only one lesion or all of them.

Furthermore, the impact of the choice of features and segmentation strategies on the performance varies depending on the prediction model. Nevertheless, some trends can be observed. Regarding the choice of feature set, the HeadNeck features (Featv1) perform best when averaged over all lesions segmented by majority voting (MLP with $\boldsymbol{r}$=2), while other image-based methods perform less well on this feature set (note that both these observations and the following should be taken with a certain amount of caution considering the standard deviations reported). On the other hand, removing second-order features (Featv2-Featv3-Featv4) does not usually improve the results, except when

relying only on the features of the largest lesion (MLP with $\boldsymbol{r}$=1). Moreover, among the image-based methods dealing with the first-order feature sets (Featv2-Featv3-Featv4), the majority vote segmentation (Featv3) ranks higher in all cases except for the graph-based methods (GraphConv and GAT), which perform better without additional segmentation on the raw bounding boxes (Featv2). It is also useful to note that adding more features (Featv5) degrades the MLP performance compared to the HeadNeck set (Featv1), but improves the scores for methods that process data from each lesion individually (MIL and graph-based methods). Overall, the type of features extracted from the lesions has an impact on the performance which depends on the classification model considered. As no feature set particularly outperforms the others, we consider again those five sets in the next part (Part III) to see how they perform when combined with clinical data.

Looking at the performances depending on the strategy to deal with the variable number of lesions of each patient, we observe that the MIL approach is almost always the worst method. This shows that considering the multiple lesions individually is not relevant if not done appropriately. However, the graph-based approaches (GraphConv and GAT), using both all the lesions of the patients and their spatial distribution, rank always best or second best regardless of the feature set, and perform especially well when no additional segmentation is applied on the raw bounding boxes defined by clinicians (Featv2 and Featv5). The MLP using only the features from one lesion does not perform particularly well, except when using the first-order feature sets with additional segmentation of the lesions (Featv3 and Featv4). These observations support our choice to use a lesion graph to make use of the information from all the lesions of each patient and from their spatial distribution.

5.4 Discussion

In a nutshell, in this part we investigated how to best make use of the PET image data of DLBCL patients for high-risk patient identification. To this end, we compared three segmentation strategies and three sets of radiomics extracted, combined in five imaging feature sets. We explored using these imaging features as input to either single- or multiple-lesions approaches, taking into account or not the spatial distribution of the lesions. Those strategies were compared on a 2-year PFS classification task on the GAINED dataset. The methodology and results presented in this part have been published in the journal *Computerized Medical Imaging and Graphics* [Thiery, Rizkallah, Bailly, et al. 2025].

From those experiments, we deduced multiple conclusions:

— Our first observation is that predicting the 2-year PFS from only PET imaging features is a complex task, whether we use only one *representative* lesion or all of them.

— The experiments also revealed complex interactions between the feature sets and the multi-lesion strategies, with no set of radiomic features or segmentation strategy being clearly better than the others. This motivates exploring again all those sets for the integration of this imaging data with the clinical data in the next part of this thesis.

— Lastly, as the lesion graph methods rank among the best models regardless of the feature set, we choose to keep this approach in the following to fuse the information from the multiple lesions of the patients.

Some perspectives for this work include testing the framework from [Loutfi et al. 2024], who proposed an extensive process to automatically select the most relevant and least complex radiomics to extract from an image for a given task: we would like to apply this method on our DLBCL 2-year PFS prediction task to see both which radiomic features would be selected and if it could improve our performances. It would also be interesting to conduct an extended study of the graph structure considered for the lesion graph, which could help to take into account the dispersion of the lesions more efficiently. Finally, we would like to see how the lesion graph representation would perform with more powerful GNNs, as the ones considered (GraphConv and GAT) are widely used in the literature but lack expressive power compared to some more recent modules tailored to graph classification.

Part III

Interpretable multimodal fusion of clinical and imaging information

Abstract

We now focus on realizing the fusion of the imaging information studied in the previous part with clinical data. To this end, we propose a solution based on a graph neural network model, capable of fusing information from both data modalities and over lesions while modeling the variable number of lesions across patients. In this solution, we opt for an attributed lesion graph to represent the PET image of each patient. Moreover, we design a cross-attention module to fuse the image attributes with clinical data, which is particularly challenging given the large difference in dimensionality and prognostic strength of each modality. Thus, we propose several cross-attention configurations, discuss the implications of each design, and experimentally compare their performances. The last module fuses the updated attributes across lesions and makes a probabilistic prediction of the patient's 2-year PFS. We carry out the experimental validation of our proposed framework and show its interpretable properties on the GAINED dataset.

6.1 Introduction

Current methods for DLBCL risk stratification do not take into account both modalities of data available, *i.e.* the PET scan and the clinical information of each patient, or only through basic fusion schemes [Carlier et al. 2024; Eertink et al. 2022; Jiang et al. 2022]. In this part, we instead propose a model able to fuse efficiently the image information studied in the previous section with clinical data in order to identify high-risk DLBCL patients.

While developing a risk classification approach from real-world, prognostic, multicentric, unbalanced, and multimodal data is already a complex task, we face additional challenges related to the disease under consideration and the available data. First, the efficient integration of multiple data modalities is still an open question in the field [Baltrušaitis et al. 2019], and it faces specific additional challenges in our case. Indeed, our fusion implies combining an already identified set of highly predictive clinical features with image information for which there is no consensus (with few exceptions) on their informative value for our task; and there is also a significant gap between the dimensionality of the two modalities, with very few clinical features and many imaging features, increasing the complexity of the task. A second challenge is that the size of the imaging features depends on the number of lesions of each patient: our fusion framework must be able to deal with this varying dimensionality. Finally, while our work relies on a relatively large DLBCL database for the field, it is not comparable to the huge databases required to train complex fusion models in computer vision [Z. Wang, Z. Wu, et al. 2022].

The method we propose to address these challenges first extracts image information at all relevant locations, then relies on a lesion graph and a Graph Neural Network (GNN) to handle the varying number of lesions in each patient, and a cross-attention fusion module to integrate the two heterogeneous modalities.

Beyond the sole development of this model, we performed multiple analyses to improve its fusion ability, prove the relevance of the structural choices made and show its interpretability properties. First, to improve the quality of multimodal fusion and the performance of high-risk patient identification, we conducted an extensive study and explored several variants of the inputs to the cross-attention module, a key design element that we focus on. Indeed, as will be shown later, simple fusion approaches are less powerful and fail to extract complementary information from the different modalities: thus, the design of this fusion block is crucial to get the most out of both modalities. We also perform an ablation study on the structural configuration of the best model found, and validate its performance on a complex set of imaging features. Moreover, using the best-performing model, we analyze the attention patterns to highlight which

lesions are most relevant to the task and provide an interpretation for each patient. Finally, we take advantage of the interpretability of our novel multi-lesion multimodal cross-attention design to rank the most salient clinical features.

6.2 Method

As in Part II, we consider a DLBCL clinical examination of a patient n at baseline before treatment, consisting of a set of tabular clinical indicators $\mathbf{c}^{(n)} \in \mathbb{R}^{D_{\text{clin}}}$ and a whole-body PET image of the patient $\mathcal{I}^{(n)}$. The proposed framework, presented in Fig. 6.4, takes the whole-body 3D PET image with 3D lesion segmentation $\mathcal{I}_{seg}^{(n)}$ and the corresponding clinical tabular data as input, and is trained to predict the probability of 2-year Progression-Free Survival (PFS).

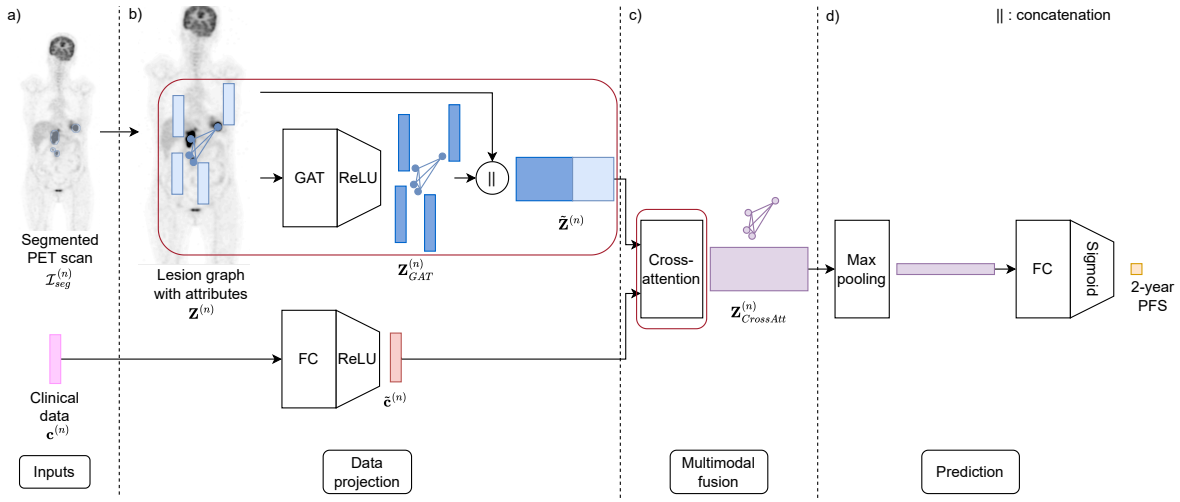


Figure 6.4 – Overview of the proposed framework: a) For a patient n, this framework takes as input the whole-body PET image with 3D lesion segmentation $\mathcal{I}_{seg}^{(n)}$ and the corresponding clinical tabular data $\mathbf{c}^{(n)}$; b) each data modality is projected into a more informative embedding space: a linear projection is applied to the clinical data with a fully connected (FC) layer, and the imaging information is first modeled as a lesion graph, and the features of all the lesions are merged by a graph attention module (GAT) and concatenated with their initial features; c) the two projected data modalities are fused thanks to a cross-attention module; d) a max pooling operation allows us to obtain a single vector, which is fed into an FC layer leading to the 2-year PFS prediction. In red are our structural contributions: the graph-based representation for dealing with the variable number of lesions of the patients (cf. Sec. 5.2.2.1), and the conception of a cross-attention module adapted to the multi-lesion multimodal fusion problem (cf. Sec. 6.2.3).

Our proposed end-to-end learning framework consists of several blocks. First, we create a lesion graph to simultaneously represent the image features of individual lesions and their spatial distribution. Then, a Graph Neural Network (GNN) coupled with a

cross-attention module learns to predict 2-year PFS from this lesion graph and the clinical data. To do this, the information from all lesions is first integrated thanks to a GAT model [Brody et al. 2022], which has the ability to adaptively learn the attentions between lesions for each patient. In parallel, the clinical data is projected with a Fully Connected (FC) layer. The outputs of these operations are then fused by the cross-attention module, which can bring together multiple modalities of data of different dimensions with interpretable behavior. These steps are followed by a pooling operation to reduce the output of the cross-attention layer to a single vector of fixed size, which is sent to an FC layer for prediction. This framework and the corresponding modules are described in more detail in the following.

6.2.1 Lesion graph creation

The first step of this framework is to create a fully connected graph, as presented in Sec. 5.2.2.1, to group the information of the $L^{(n)}$ lesions appearing on the PET scan $\mathcal{I}_{seg}^{(n)}$ of the n^{th} patient. Each of its nodes $v_i^{(n)} \in \mathcal{V}_l^{(n)}$ corresponds to a single lesion i, from which the image data (including both conventional and radiomic features) are extracted and represented in the feature vector $\mathbf{z}_i^{(n)} \in \mathbb{R}^{D_{\text{img}}}$. We denote by $\mathbf{Z}^{(n)} \in \mathbb{R}^{L^{(n)} \times D_{\text{img}}}$ the matrix concatenating all the node features $\mathbf{z}_i^{(n)}$, and $e_{ij}^{(n)} \in \mathcal{E}_l^{(n)}$ the edge between a pair of nodes $v_i^{(n)}$ and $v_j^{(n)}$ of weight $w_{ij}^{(n)}$.

6.2.2 Multi-lesion and clinical data representation learning

6.2.2.1 Multi-lesion representation

In order to incorporate the information from all the patient's lesions while taking into account their relationships, we design a GNN to be applied over our lesion graph. We define a GAT layer with two heads [Brody et al. 2022] to update each node based only on the information propagated from the most relevant neighboring lesions. This operation, described in Part I (Sec. 4.3), returns an updated feature vector $\mathbf{z}_{i_{\text{GAT}}}^{(n)}$ for every node i.

Following this GAT operation, a ReLU activation layer is applied on the updated lesion representations. The resulting feature matrix of dimension $(L^{(n)} \times D_{\text{GAT}})$ is a concatenation of the lesion feature vectors: $\mathbf{Z}_{\text{GAT}}^{(n)} = \text{ReLU}([\mathbf{z}_{1_{\text{GAT}}}^{(n)} || \dots || \mathbf{z}_{L^{(n)}{}_{\text{GAT}}}^{(n)}]^\top)$. Finally, a skip connection is added to preserve the original information of the lesions with a simple concatenation operation resulting in $\tilde{\mathbf{Z}}^{(n)} \in \mathbb{R}^{L^{(n)} \times (D_{GAT} + D_{img})}$.

To fuse the lesion information, we explore in total two graph-based modules and a learned-attention process. Hence, a node embedding $\mathbf{z}_i^{(n)}$ can be updated either (i) by a GAT layer (as explained before), or (ii) by a GraphConv layer [Morris et al. 2019] as

described in Part I, Sec. 4.3 to get an updated feature vector $\mathbf{z}^{(n)}_{i_{\text{GraphConv}}}$.

In the last option (iii), unlike the previous two message passing schemes where the structure of the graph and the edge weights are imposed, the graph (*i.e.* the strength of the connections between lesions) is dynamically learned by a self-attention module [Vaswani et al. 2017]. In this case, the features are updated as:

$$\mathbf{Z}^{(n)}_{\text{SelfAtt}} = \text{softmax}\left(\frac{\mathbf{Z}^{(n)}\mathbf{W}_{SA,Q}(\mathbf{Z}^{(n)}\mathbf{W}_{SA,K})^\top}{\sqrt{d_k}}\right)\mathbf{Z}^{(n)}\mathbf{W}_{SA,V}, \tag{6.6}$$

with three learnable matrices $\mathbf{W}_{SA,Q} \in \mathbb{R}^{D_{\text{img}} \times d_q}$, $\mathbf{W}_{SA,K} \in \mathbb{R}^{D_{\text{img}} \times d_k}$ and $\mathbf{W}_{SA,V} \in \mathbb{R}^{D_{\text{img}} \times d_v}$, having the role of projecting the image features to improve the quality of their combination. This last option based on the self-attention module directly returns the projected feature matrix for all lesions, $\mathbf{Z}^{(n)}_{\text{SelfAtt}}$, of the same dimension as $\mathbf{Z}^{(n)}$.

6.2.2.2 Clinical data representation

In parallel with the multi-lesion representation learning step, a fully connected layer followed by a ReLU activation projects the patient's clinical data $\mathbf{c}^{(n)}$ into a more informative space. We will refer to the output of this step as $\bar{\mathbf{c}}^{(n)}$, of dimension D_{proj}.

6.2.3 Multimodal Multi-lesion Cross-Attention

After the construction of a lesion graph and the design of the GNN to efficiently model the relevant image information, a second core element of our model is a cross-attention module for multi-modal fusion. This module, inspired by self-attention blocks in natural language processing tasks [Vaswani et al. 2017], fuses the information from multiple modalities by computing the similarity between a query $\mathbf{Q}$ and a key $\mathbf{K}$ to get attention scores, which are used to weight a value matrix $\mathbf{V}$ as presented in Part I, Sec. 4.3.5. Using this cross-attention for multimodal fusion is done by considering each modality, in our case, the lesion graph with attributes $\tilde{\mathbf{Z}}^{(n)}$ and the projected clinical data $\bar{\mathbf{c}}^{(n)}$, as either the query, the key/value pair or both of them.

More precisely, we consider different instantiations of this cross-attention module to investigate how such implementations affect the fusion performance and address the specific multi-modal and multi-lesion challenges we have encountered. Indeed, not only do we have to deal with different dimensions and predictive capabilities between the two modalities, but we also have to deal with the varying dimensionality of image data between patients due to the inconsistent number of their lesions. Therefore, we take advantage of the adaptability of the cross-attention module to either let each modality contribute asymmetrically (one modality influencing the weighting of the other), or to use it in a more classical way with a self-attention process (Mult2Mult implementation).

Overall, five configurations using different modalities as query $\mathbf{Q}$, key $\mathbf{K}$ and value $\mathbf{V}$ matrices are explored. These configurations, presented in Fig. 6.5, are the following:

— **Img2Clin**: First, we consider a model where $\mathbf{Q} = \tilde{\mathbf{Z}}^{(n)}$, and $\mathbf{K} = \mathbf{V} = \bar{\mathbf{c}}^{(n)}$. As the clinical features are already known to be predictive for the 2-year PFS [Carlier et al. 2024], using it as the value of the module should improve this prediction.

— **Mult2Clin**: A configuration similar to Img2Clin, but including self-attention to the clinical data in addition to the cross-attention between the two modalities. If the clinical data supports the imaging information to decide which clinical elements are most relevant for prediction, this could allow us to make the most of its predictive capabilities. So, we still have $\mathbf{K} = \mathbf{V} = \bar{\mathbf{c}}^{(n)}$, but $\mathbf{Q}$ is equal to $\tilde{\mathbf{Z}}^{(n)}||\bar{\mathbf{c}}^{(n)\top}_{concat}$, where $\bar{\mathbf{c}}^{(n)}_{concat}$ is the matrix resulting from the column-wise concatenation of the vector $\bar{\mathbf{c}}^{(n)}$, $L^{(n)}$ times. Thereby, the shape of $\mathbf{Q}$ is $(L^{(n)} \times D_{concat})$ with $D_{concat} = (D_{GAT} + D_{img}) + D_{proj}$.

— **Mult2Mult**: A configuration that uses both self- and cross-attention on each type of data, so $\mathbf{Q} = \mathbf{K} = \mathbf{V} = \tilde{\mathbf{Z}}^{(n)}||\bar{\mathbf{c}}^{(n)\top}_{concat}$. This corresponds to a common way of fusing multimodal data with a cross-attention module, where the input key, query and value are all a concatenation of both modalities.

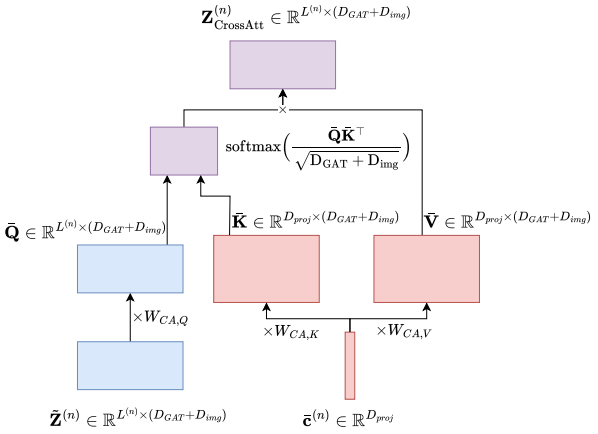
— **Clin2Img**: This model is the opposite of Img2Clin, as we invert $\mathbf{Q}$ and $\mathbf{K} = \mathbf{V}$ to get $\mathbf{Q} = \bar{\mathbf{c}}^{(n)\top}$, and $\mathbf{K} = \mathbf{V} = \tilde{\mathbf{Z}}^{(n)}$. With this configuration, we investigate whether our method can extract information from the lesion data that is relevant to predicting 2-year PFS, based on the clinical data.

— **Mult2Img**: Our final model is an extension of Clin2Img where, in addition to simple cross-attention between the two modalities, we include self-attention to the image data. We expect this configuration, as in our GNN module, to use the relationship between the lesions in addition to the clinical data to select which lesion's updated information is most predictive for the learning task. Thus, we choose $\mathbf{Q} = \tilde{\mathbf{Z}}^{(n)}||\bar{\mathbf{c}}^{(n)\top}_{concat}$ and $\mathbf{K} = \mathbf{V} = \tilde{\mathbf{Z}}^{(n)}$.

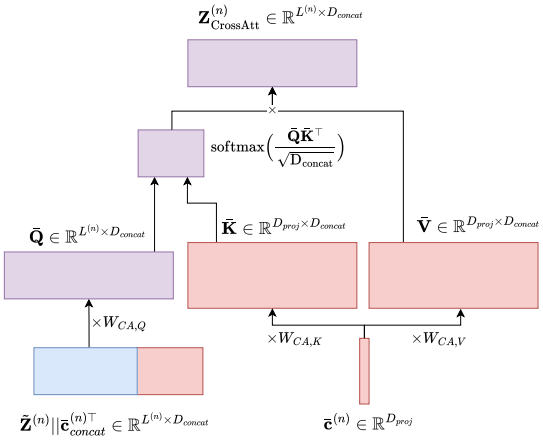
After this fusion process, we obtain a matrix $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$, whose shape depends on the cross-attention configuration considered.

6.2.4 Prediction module

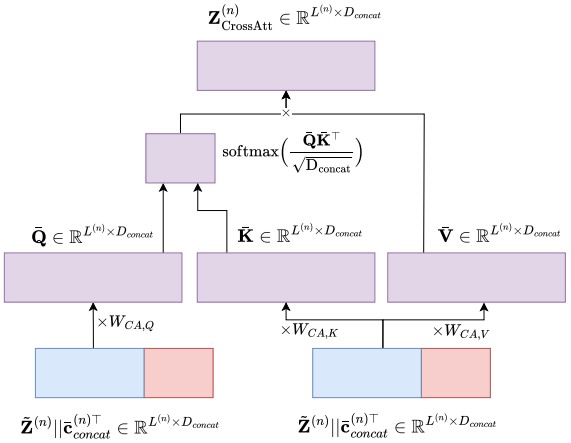
Finally, a max pooling operation, taking the channel-wise maximum across the node dimension, reduces the dimensionality of $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ into a single $\mathcal{H}_Q$-dimensional vector (depending on the shape of $\mathbf{Q}$), and a fully connected layer with sigmoid activation predicts the probability of the patient's 2-year PFS from this vector. Learning is, as in Part II, controlled by a weighted binary cross-entropy loss function, with weights compensating for class imbalance.



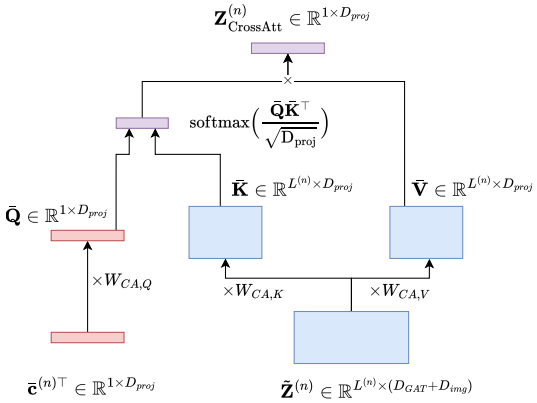
(a) Img2Clin



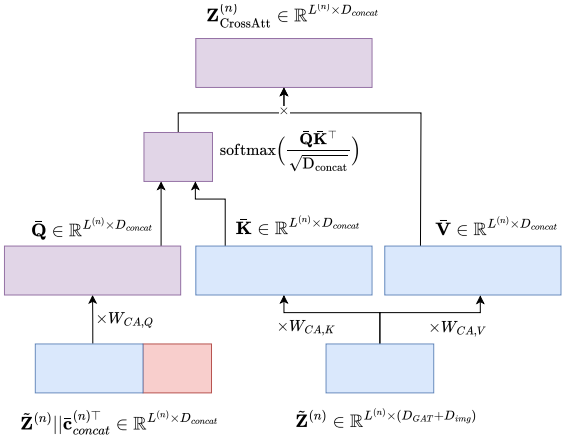
(b) Mult2Clin



(c) Mult2Mult



(d) Clin2Img



(e) Mult2Img

Figure 6.5 – Proposed design choices and configurations for the cross-attention module specially conceived to be combined with a lesion graph.

The max pooling operation, for the Img2clin and Mult2Clin cross-attention configurations, forces the model to concentrate all information from the patient on two lesions that positively and negatively influence its 2-year PFS prediction, respectively. This behavior, which is very useful in terms of interpretability, is explained in Sec. 6.3.4. In general, the whole fusion framework presented is similar to a MIL (Multiple Instance Learning) or late fusion process in the sense that the prediction is made immediately after the cross-attention module responsible for the fusion of the two modalities. However, it has two important differences from these types of models:

— For the image information, we do not consider the features of each lesion separately, as a MIL would do, but rely on all of them and on the underlying topology to update the representation of each of these lesions.

— In the cross-attention module, we do not simply sum or concatenate the features of the two modalities to make the prediction, but we dynamically and adaptively discover from one or both modalities which features of the other are most relevant for a given patient to predict its 2-year PFS.

These differences, as shown in Sec. 6.3, allow our proposed framework to outperform conventional and state-of-the-art models by taking advantage of the multiple modalities available.

6.3 Experimental validation

6.3.1 Dataset and experimental setup

To validate our model, we still rely on the GAINED dataset as described in Part I, chapter 2. We consider the same experimental setup described in Part II, Sec. 5.3.1. As previously mentioned, a grid search is performed on all evaluated models, including the baseline models we compare to and the different configurations studied, by selecting the configuration with the best validation AU-ROC. These grid searches explore the space of hyperparameters, including the learning rate, the output dimension of the projection layers (FC, GraphConv or GAT layers) and, for the GNN methods, the parameter γ used in the construction of the lesion graphs. The full set of hyperparameters explored is presented in Appendix C. In addition, the clinical data are standardized as the image data by removing the mean and scaling their variance to unity, and we use a McNemar test [Mcnemar 1947] on our balanced test predictions to validate the statistical significance of our results.

6.3.2 Choice of clinical and image input features and comparison to baseline methods

6.3.2.1 Experiment

Our first experiment aims to extend the image features selection study from Part II to explore how these feature sets perform when fused with clinical data, either in basic ways or by our proposed framework. To this end, we compare again on the 2-year PFS prediction task the feature sets presented in Sec. 5.2.1, with several baseline models using either clinical data, imaging data, or both. These models deal differently with the variable number of lesions of the patients and rely on different fusion approaches when both modalities are used. The methods investigated are:

— **MLP**: This model consists of two linear layers with ReLU activations and a 1-dimensional linear output layer with sigmoid activation. The two intermediate layers have the same dimension, chosen by a grid search. The MLP model allows us to see how predictive each modality is for the given task with a very simple approach, and to test a simple early fusion scheme. However, as it requires fixed-size inputs, when the lesion image information is used it must be reduced to a fixed-size vector, either by averaging the features from all lesions or by considering only the data from a single lesion.

— **XGBoost**: This state-of-the-art tree-based model [T. Chen and Guestrin 2016] is a representative of traditional machine learning methods and is tested, like the MLPs, on clinical data concatenated either with the mean of the features of the lesions of each patient or with the features of the largest lesion. Several parameters of this model are chosen by grid search, including the number of estimators and the maximum depth of the trees (the range of values considered is described in Appendix C).

— **MIL**: A Multiple Instance Learning approach that takes as input the image features from a patient's $L^{(n)}$ lesions, applies a one-layer MLP followed by a ReLU to each lesion's feature vector, aggregates the results by a maximum operation, and projects them linearly (with a sigmoid activation) to obtain the prediction. This approach allows us to use the image information from each lesion individually, but without using the topological information from the graph. This model is only used on imaging features alone.

— **GraphConv**: A GNN approach that takes as input the lesion graph presented in Sec. 5.2.2.1. It consists of two GraphConv [Morris et al. 2019] convolutional layers, the first with an output dimension determined by grid search and followed by ReLU activation, and the second with an output size of 1 and followed by max

pooling and sigmoid activation to predict PFS. Both this model and the following GAT aim to explore the interest of graph structure over just considering lesion image features, as the MIL does, and are also only used on imaging features alone.

— **GAT**: A model with the same structure as the GraphConv model, but with GAT layers [Brody et al. 2022] as presented in Sec. 5.2.2.1 instead of GraphConv layers.

— **Cross-attention:** The intermediate fusion model presented in Sec. 6.2, with its original configuration from our preliminary work [Thiery, Rizkallah, Bailly, et al. 2023] based on two Img2Clin cross-attention modules, to show that it is possible to take advantage of both modalities while considering all lesions.

— **Cross-attention 2:** An updated Cross-attention model based on two Img2Clin cross-attention modules, with the addition of two successive fully connected layers followed by ReLU activations to project the clinical data (one before each cross-attention module).

The models detailed above take as input either the image information $\mathbf{Z}^{(n)} \in \mathbb{R}^{L^{(n)} \times D_{\text{img}}}$, the clinical data vector $\mathbf{c}^{(n)} \in \mathbb{R}^{D_{\text{clin}}}$, or both. For the models based on imaging features, we explore both single- and multiple-lesion approaches. For example, for the MLP models that require a fixed-size vector as input, we consider either averaging the imaging data from all lesions or using only the data from the largest lesion. For multiple lesion approaches, we compare a simple MIL strategy with graph-based fusion approaches. For the MLPs and XGBoost models where we use both clinical and image data, the corresponding vectors are concatenated before being given as input to the model (early fusion).

6.3.2.2 Results

Regarding the comparison of **clinical-only vs image-only methods**, the results in Table 6.2 show a first observation, which is the performance gap between the two modalities. This is an expected result, as the clinical data considered have shown their efficacy for the task at hand [Carlier et al. 2024]. On the other hand, although there is some evidence that radiomics supports 2-year PFS classification for DLBCL with simpler prediction models [Carlier et al. 2024], their combination with more complex models has not yet been shown to be more effective than clinical data alone [Carlier et al. 2024].

With respect to **image-only models**, as in Part II, Sec. 5.3, we observe complex interactions between the feature sets and multi-lesion strategies. Also, the lesion graph approaches consistently rank among the best models no matter the feature set considered.

	$\boldsymbol{c}$	$\boldsymbol{r}$	Featv1	Featv2	Featv3	Featv4	Featv5
Segmentation			Majority vote	Raw	Majority vote	SUV4	Raw
Image features			HeadNeck	FirstOrder	FirstOrder	FirstOrder	All
MLP	1		$\underline{0.65 \pm 0.09}$*	$\underline{0.65 \pm 0.09}$*	0.65 ± 0.09*	0.65 ± 0.09*	$\mathbf{0.65 \pm 0.09}$
MLP		1	0.53 ± 0.08*	0.52 ± 0.12*	0.56 ± 0.09*	0.56 ± 0.11*	0.50 ± 0.09*
MLP		2	0.61 ± 0.11*	0.52 ± 0.08*	0.53 ± 0.10*	0.51 ± 0.09*	0.56 ± 0.08*
MIL		3	0.52 ± 0.10*	0.46 ± 0.09*	0.51 ± 0.10*	0.50 ± 0.08*	0.55 ± 0.12*
GraphConv		3	0.54 ± 0.11*	0.56 ± 0.11*	0.54 ± 0.09*	0.54 ± 0.09*	0.61 ± 0.07*
GAT		3	0.55 ± 0.10*	0.54 ± 0.10*	0.53 ± 0.10*	0.51 ± 0.11*	0.59 ± 0.10*
MLP	1	1	0.61 ± 0.10*	0.62 ± 0.12*	0.61 ± 0.11*	$\underline{0.66 \pm 0.09}$	0.58 ± 0.09*
MLP	1	2	0.63 ± 0.05*	0.60 ± 0.09*	0.62 ± 0.07*	0.60 ± 0.08*	0.59 ± 0.09*
XGboost	1	1	0.61 ± 0.08*	0.63 ± 0.07*	$\underline{0.66 \pm 0.07}$	0.63 ± 0.08	0.63 ± 0.06
XGboost	1	2	0.63 ± 0.08	0.61 ± 0.08	0.64 ± 0.07	0.63 ± 0.09*	0.61 ± 0.09*
Cross-att.	1	3	0.59 ± 0.11*	0.63 ± 0.08	0.64 ± 0.08	0.63 ± 0.08*	0.64 ± 0.08
Cross-att. 2	1	3	$\mathbf{0.67 \pm 0.09}$	$\mathbf{0.66 \pm 0.08}$	$\mathbf{0.67 \pm 0.08}$	$\mathbf{0.67 \pm 0.07}$	$\underline{0.64 \pm 0.10}$

Table 6.2 – AU-ROC results for the data fusion study: $\boldsymbol{c}$ and $\boldsymbol{r}$ indicate whether clinical ($\boldsymbol{c}$) and/or radiomic ($\boldsymbol{r}$) data are used for the model tested. For $\boldsymbol{r}$, a value of (1) means that we consider only the data from the largest lesion; (2) that we average the imaging data from all lesions; and (3) that we consider the data from each lesion individually. For a set of features, the first is bold and the second is underlined; and we indicate with a star * that the performance of a model on a set of features is significantly different from the performance of the Cross-attention 2 model on the same set (p-value<0.05).

Regarding the performance of the **multimodal models**, the results in Table 6.2 show that the cross-attention model with projection of clinical data is statistically significantly better (p-value < 0.05) than most models (including unimodal ones) using the same type of features, except for the set with the highest number of imaging features (Featv5) where the model using only clinical data is the best. This highlights the importance of the cross-attention design to really take advantage of both types of data, compared to simple fusion methods, which mostly degrade the results compared to the reference MLP relying only on clinical data. Regarding the Featv5 set, which includes the 650 imaging features, we observe that its combination with clinical data leads to a significant drop in performance for simple fusion models (less important for our cross-attention models) compared to other feature sets; and this despite the predictive power of the clinical features, as shown by the MLP performance on clinical data only. We attribute this drop in performance to the difference in dimensionality between the two modalities, which may make this feature set more difficult to learn. Finally, we observe an important difference between the performance of the cross-attention model with and without clinical data projection: this motivates our experiment in Sec. 6.3.3 to explore a wide range of cross-attention model instantiations to find the most efficient one for this complex task.

Overall, the type of features extracted from the lesions still has an impact on the performance that varies depending on the model considered: therefore, in the following

experiments, we will focus on the Featv3 set because of its good performance with the two Cross-attention models.

6.3.3 Cross-modal attention design

Next, we evaluate the influence of the configuration of the cross-attention module on multimodal fusion. For each configuration presented in Sec. 6.2.3, we study the impact of several structural elements of our model, including:

— the learning mechanism for multi-lesion fusion between a GAT, a GraphConv or a self-attention module;

— the presence of a skip connection to keep the raw image data accessible (or not) after the multi-lesion fusion stage;

— the projection of the clinical data with a linear layer vs the implementation of cross-attention directly on the raw clinical data;

— the use or not of a ReLU activation after the projection of the clinical data;

— and the number of cross-modal fusion blocks (1 or 2), taking into account that one block consists of a multi-lesion fusion module and a multi-modal cross-attention layer. In the case of two blocks, the output of the first cross-attention module is sent to another GAT with ReLU activation, then concatenated with the initial image information $\mathbf{Z}^{(n)}$ and fused again with the projected clinical data $\bar{\mathbf{c}}^{(n)}$ by a second cross-attention module before the max pooling operation and the prediction (cf. Appendix D.1).

We consider and evaluate several combinations of these elements for each instantiation of the cross-attention module presented in Sec. 6.2.3, and detail the results in Table 6.3, including, for each cross-attention design, the performance of the best model in terms of test AU-ROC and the retained structure.

Regarding the design of the **cross-attention**, our results show that for the problem studied it is crucial to use the computed attention scores to modulate the clinical data, *i.e.* to include the clinical data as part of the 'Value' matrix as it is the case for the Img2Clin, Mult2Clin and Mult2Mult modules. Instead, the poorer performance of the Clin2Img and Mult2Img models shows that we fail to extract meaningful information from using image data as the 'Value' matrix of the cross-attention module, even when using the clinical data to modulate this extraction. We attribute this behavior gap to the higher predictive power of the clinical data, making it easier to extract relevant information from this modality. This conclusion suggests that in a cross-attention module used for fusion, it is better to use the more predictive information as the key/value, regardless of the data used as the query. Other choices, such as using self-attention in

Cross-att. module	Best AU-ROC	GAT	GraphConv	Self-attention	Skip connection	No skip connection	1 multimodal block	2 multimodal blocks	Clinical FC layer	No clinical FC layer	ReLU	No ReLU	N. of parameters
Img2Clin	$\underline{0.68 \pm 0.08}$	✓				✓		✓	✓		✓		110,305
Mult2Clin	$\mathbf{0.68 \pm 0.08}$			✓	✓		✓		✓		✓		19,637
Mult2Mult	0.67 ± 0.09			✓		✓		✓	✓		✓		43,091
Clin2Img	0.61 ± 0.07*	✓			✓		✓	-	✓			✓	85,633
Mult2Img	0.62 ± 0.08*		✓			✓	✓		✓		✓		105,089

Table 6.3 – Best structure obtained for each cross-attention configuration and corresponding performance (- means impossible). We indicate with a star * the methods whose performance is significantly lower (p-value<0.05) than our best Mult2Clin model. The configuration of each layer is as follows: the GAT layer has two heads and is followed by a ReLU activation. ReLU activations are also applied to the GraphConv and self-attention layers if selected. The GAT, self-attention and cross-attention layers have a dropout rate chosen by grid search (cf. Appendix C). If we have an FC layer projecting the clinical data, its activation may be ReLU or no activation depending on the structure selected. We refer the readers to Appendix D.1 for the detailed model structure in case the second multimodal block is used, and to Appendix D.2 for an equation to compute the number of parameters of the model depending on its configuration.

addition to cross-attention within this module (Img2Clin vs Mult2Clin and Clin2Img vs Mult2Img), have less impact on the results.

Furthermore, we observe that all models require the projection of clinical data to perform best. However, when this is excluded, the best configuration is very different depending on the use of cross-attention, highlighting the complexity of constructing an efficient model when working with real-world, multicentric, prognostic and multimodal data.

Finally, we observe no direct correlation between performance and the number of learned parameters of the configuration retained by the grid search for the different cross-attention instantiations. This behavior can be explained in light of the bias-variance tradeoff: while a higher number of parameters usually translates to a better training performance (bias reduction), there is also a risk of higher variance in inference and, thus, of overfitting. Our results show that the model's structural design is critical in preventing such overfitting.

Overall, the Mult2Clin cross-attention module outperforms the others in its ability to predict 2-year PFS: we therefore retain the Mult2Clin design with its best configuration for the next experiments.

Ablation study To show the influence of each element of the structure of our best model, we perform an ablation study, presented in Table 6.4. This study highlights the relevance of each structural choice, with the model obtained from our previous experiments performing better than any other on the Featv3 feature set. From this experiment, we can also identify the elements in our model structure that are the most important among those evaluated. The most important in this case is the skip connection, followed by the choice of the multi-lesion fusion block and the clinical projection. However, relying on one or two blocks for multimodal fusion gives similar results.

Best AU-ROC	GAT	GraphConv	Self-attention	Skip connection	No skip connection	1 multimodal block	2 multimodal blocks	Clinical FC layer	No clinical FC layer	ReLU	No ReLU	N. of parameters
0.64 ± 0.09*			✓		ⓒ	✓		✓		✓		25,359
0.65 ± 0.09*		ⓒ		✓		✓		✓		✓		109,559
0.65 ± 0.08*			✓	✓		✓		✓			ⓒ	41,237
0.66 ± 0.09*	ⓒ			✓		✓		✓		✓		113,783
0.67 ± 0.09			✓	✓		✓			ⓒ	-	-	17,475
$\underline{0.68 \pm 0.09}$*			✓	✓			ⓒ	✓		✓		84,963
0.68 ± 0.08			✓	✓		✓		✓		✓		19,637

Table 6.4 – Ablation study over the components of our best model (with Mult2Clin cross-attention module). Circled tickmarks highlight the modified component with respect to the best reference configuration (bottom). We indicate with a star * the models whose performance is significantly lower (p-value<0.05) than our best model without ablation (bottom row). Please refer to Table 6.3 or Sec. 6.2 for a more precise description of the layer settings.

Translation to a larger feature set To test the robustness of our best Mult2Clin model, whose structure was optimized on the Featv3 set of imaging features (first-order features on a majority vote segmentation), we also train it on the Featv5 set (full set of image radiomics on raw segmentation), which was the most difficult to learn for models based on both clinical and imaging data (cf. Table 6.2). With an AU-ROC of 0.69 ± 0.06 (and 5,165,701 parameters), there is a slight increase in performance compared to all baseline models in Table 6.2 (with a maximum performance of 0.67 ± 0.08 for 20,225 parameters). This is especially true compared to the previous best baseline using the Featv5 set of image features (0.64 ± 0.10*) (566,977 parameters), and even compared to the same best proposed model learning on the Featv3 set of features (0.68 ± 0.08*)

(19,637 parameters). The difference in the number of parameters is directly linked to the increase in imaging features and to the use of a skip connection. In this paragraph, we have indicated with a star (*) the models whose performance is significantly lower (p-value<0.05) than our best model Mult2Clin on the Featv5 set of features. Despite the moderate gain, this experiment proves the ability of our model to take advantage of both modalities even when there is an important difference in dimensionality between the two. This is especially helpful when we do not know which features of a modality are the most important and thus cannot manually restrict the set to rely on. For the sake of clarity, we will in the following restrict the analysis to the best-performing Mult2Clin model (with the optimized structure) trained on the full set of lesion features (Featv5).

Computational complexity of multimodal models We compare in Table 6.5 the characteristics of the baseline fusion models against our best model. The better accuracy we get compared to simple machine and deep learning models comes at the expense of a higher number of parameters and computational complexity. However, note that the runtime values presented here are only indicative: indeed, they depend on both the programming language (Python here), the hardware architecture (no GPU used), and the implementation of the model. We believe that an optimized implementation run on appropriate hardware could greatly reduce the execution time of our methods.

Model	***clin***	***rad***	AU-ROC	N. of parameters	Training time	Inference time
MLP	1	1	0.58 ± 0.09*	10,865	6min 20s	3.00×10^{-4}s
MLP	1	2	0.59 ± 0.09*	10,865	9min 59s	4.04×10^{-4}s
XGboost	1	1	$\underline{0.63 \pm 0.06}$	100	18.7s	4.53×10^{-3}s
XGboost	1	2	0.61 ± 0.09	100	23s	4.89×10^{-3}s
Mult2Clin	1	3	$\mathbf{0.69 \pm 0.06}$	5,165,701	28min 19s	2.69×10^{-3}s

Table 6.5 – Overall comparison with baseline fusion models on the larger Featv5 set of features: $\boldsymbol{c}$ and $\boldsymbol{r}$ indicate whether clinical ($\boldsymbol{c}$) and/or radiomic ($\boldsymbol{r}$) data are used for the model tested. For $\boldsymbol{r}$, a value of (1) means that we consider only the data from the largest lesion; (2) that we average the imaging data from all lesions; and (3) that we consider the data from each lesion individually. The 'training' time includes both training, validation and test time for all 50 train/validation/test sets (the validation/test time is negligible compared to training time). We indicate by * that the performance of a model is significantly lower (p-value<0.05) than the performance of ours, *i.e.* the Mult2Clin one. Runtime values computed on 16 Intel(R) Xeon(R) W-11955M CPU @ 2.60 GHz.

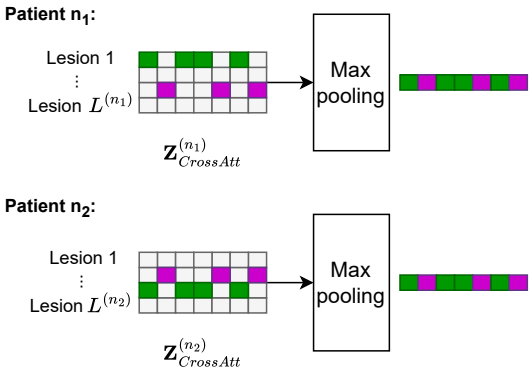
6.3.4 Model interpretability

In this section, we show how our model is designed to enhance interpretability, potentially helping clinicians to understand its predictions. To this end, we will analyze

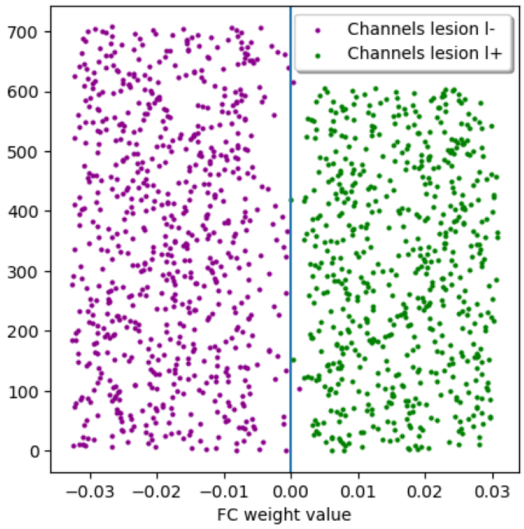
the learned attention patterns and illustrate how these provide valuable insights at both the lesion and feature level.

The first element supporting the interpretability of our model comes from the combination of the multi-lesion graph with the proposed cross-attention module. With this combination and the Mult2Clin cross-attention configuration, the computed cross-modal attentions can be interpreted as weights that modulate the clinical data according to the (graph-updated) image features of each lesion. A second design element important for interpretability, which we will focus on below, is the max pooling operation applied to the cross-attention output. While the primary role of this operation is to recover a fixed-sized embedding vector from a variable number of lesions as input, we observe two interesting mechanisms: (1) the identification of exactly two meaningful lesions for each patient; (2) one of these lesions is positively associated with 2-year PFS, while the other is negatively associated with it.

Two lesions retained by the Max Pooling In our design, the max pooling operation is applied to the columns of the cross-attention output $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ (see Fig. 6.4). In practice, the pooling operation only retains features from two rows/lesions of $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ (see Fig. 6.6a). This selection results from our cross-attention design and, more specifically, from the projection matrix $\mathbf{W}_{CA,V}$ applied to the clinical data (cf. Fig. 6.5). Indeed, given that the clinical vector $\bar{\mathbf{c}}^{(n)}$ has a single dimension, the effect of the $\mathbf{W}_{CA,V}$ projection is to replicate $\bar{\mathbf{c}}^{(n)}$ on the columns of the projected 'Value' matrix $\bar{\mathbf{V}}$, each time weighted by a different learned scalar, *i.e.* $\bar{\mathbf{V}} = \bar{\mathbf{c}}^{(n)}\mathbf{W}_{CA,V} = [\mu_1\bar{\mathbf{c}}^{(n)}, \mu_2\bar{\mathbf{c}}^{(n)}, \ldots, \mu_{D_{concat}}\bar{\mathbf{c}}^{(n)}]$. Since the $\mu_\cdot$ weights can be either positive or negative, the highest values in each $\bar{\mathbf{V}}$ column correspond to either the highest or lowest values in $\bar{\mathbf{c}}^{(n)}$. Also, since the elements of the attention matrix are all positive because they pass through a softmax function, they do not change the sign of the $\bar{\mathbf{V}}$'s columns when multiplied. Therefore, when the max-pooling operation is reached, the elements of $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ still carry the signs of $\mu_\cdot$. Let $\mathbf{a}_i$ denote the cross-attention row corresponding to lesion i, then the elements of $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ are all scaled versions of $\mathbf{a}_i\bar{\mathbf{c}}^{(n)}$. Since the elements of a $\mathbf{Z}^{(n)}_{\text{CrossAtt}}$ column share the $\mu_\cdot$'s scale and sign, the only remaining difference that matters in the max-pooling operation in that column is the attention $\mathbf{a}_i$ that each lesion gives to the clinical data $\bar{\mathbf{c}}^{(n)}$. Therefore, the two lesions that "agree" most with the positive or negative versions of $\bar{\mathbf{c}}^{(n)}$ get higher values and are retained by the max pooling. Furthermore, since the $\mu_\cdot$ are constant across patients, the two sets of columns where the value of the first or second lesion is selected form two distinct groups that are strictly the same for all patients (cf. Fig. 6.6a).



(a) Behavior of the max pooling operation



(b) Distribution of final fully connected layer weights as a function of the lesion from which the max pooling operation selected the corresponding feature (the y-axis: weight identifier).

Figure 6.6 – Explanation of the max pooling role for interpretability.

Association between the selected lesions and the 2-year PFS Our learning architecture allows us to understand not only which lesions the model relied on, but also how these lesions influenced the prediction. More specifically, by analyzing the behavior of the fully connected layer that follows the max pooling operation and outputs the 2-year PFS prediction, we observe that one of the two selected lesions, called l_+ from here on, is positively associated with the predicted 2-year PFS, while the other l_- is negatively associated with it. Indeed, this fully connected layer learns weights that are consistent with the lesions retained by max pooling: the weights related to the channels retrieved from the feature vector of a given lesion (green or purple in Fig. 6.6b) are almost exclusively of the same sign (except for 2 weights over 1316), indicating a very strong and complementary influence of each of these lesions on the 2-year PFS prediction.

Relevance of the selected lesions To experimentally demonstrate that our model selects lesions that are relevant for predicting 2-year PFS, we investigate how it performs when we prevent it from using these lesions. To do this, we iteratively remove pairs of lesions from the lesion graphs (from 0 to 16 lesions). At each iteration, the removed pairs are the two lesions selected by the model, plus those already removed in previous iterations. For each number of lesions removed, we examine the performance of the model using AU-ROC only on those patients for whom we still retain at least

80% of their lesions. For the sake of statistical significance, especially when we remove many lesions and thus have fewer patients to test on, we perform this test on our full dataset. For a given number of lesions removed, we compare the performance of the model on the same subset of patients (1) when we remove the pairs of lesions selected with the max pooling (blue in Fig. 6.7), (2) when we remove randomly selected pairs of lesions (orange in Fig. 6.7) to see if the lesions retained by max pooling are particularly significant, and (3) without any lesions removed (green in Fig. 6.7) to take into account the effect of changing the subset of patients we are testing on. We observe that for almost any number of lesions removed, the performance of the model does not drop much when we randomly remove lesions compared to keeping all lesions. However, it drops much more when we remove the lesions selected by the max pooling operation. This highlights the prognostic importance of the lesions that the model has learned to include in its prediction.

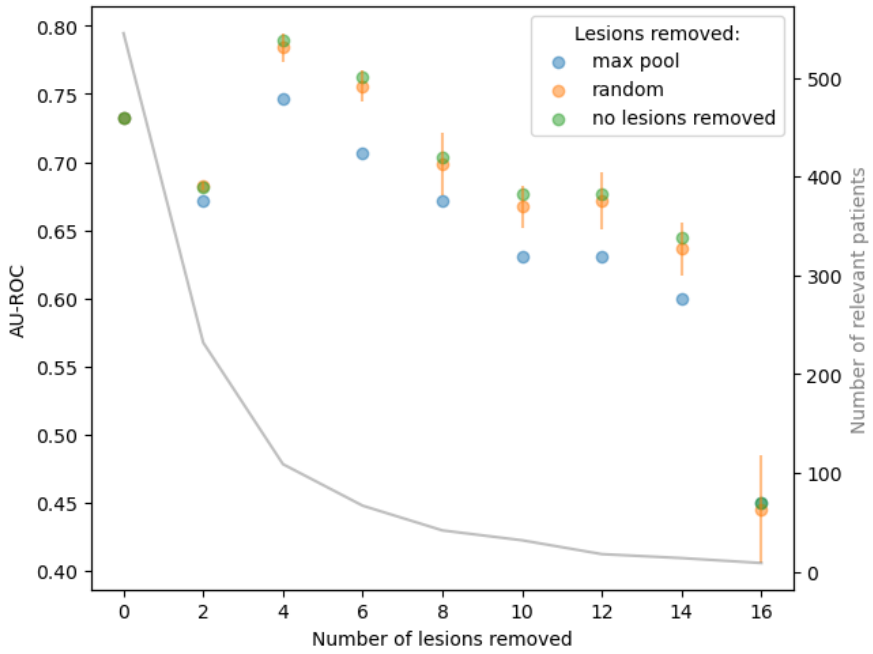
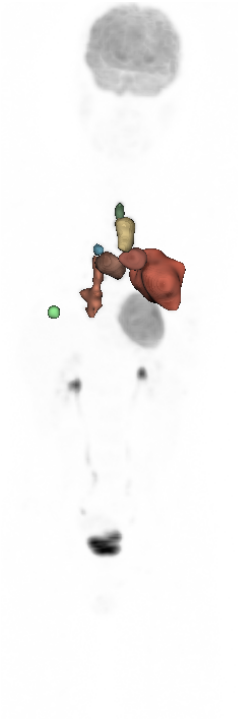
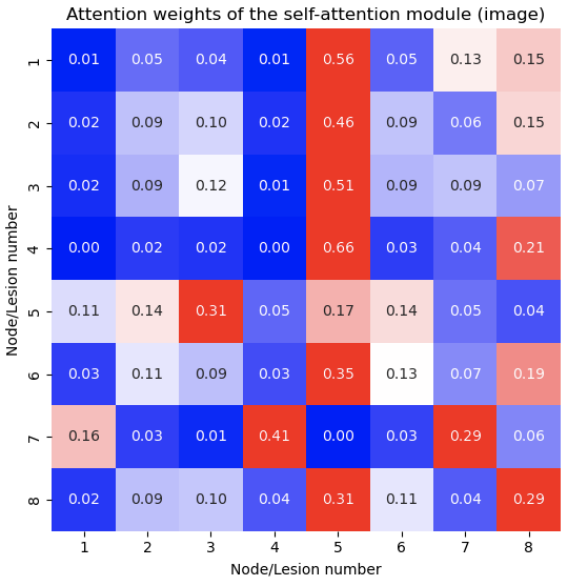


Figure 6.7 – AU-ROC performance of the best model in patients with progressively removed lesions, depending on how these removed lesions are selected (randomly or lesions retained by the max pooling operation).

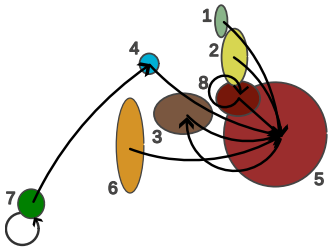
Lesion to lesion Attentions While the two previous mechanisms make it possible to identify the two complementary salient lesions that are mainly used for prediction, the self-attention matrix indicates less straightforwardly secondary lesions that are also taken into account for the prediction. They correspond to those to which l_+ and l_- pay attention in the self-attention module (which replaces the GAT in our model following



(a) Maximum Intensity Projection (MIP) of a patient and his segmented lesions. A total of 8 lesions were segmented.



(b) Lesion self-attention matrix for this patient, where l_- is lesion 3, l_+ is lesion 7, and both the label and the prediction of this patient's 2-year PFS are 1 (meaning there was a progression before 2 years after starting treatment).



(c) Schematic representation of the patient's lesions (color and size as in 6.8a) and the attention they pay to each other (only the most important attentions are represented).

Figure 6.8 – Example of interpretation of the lesion self-attention matrix.

the structural study) (see Fig. 6.8b). Indeed, the activated output of the self-attention module is concatenated with the initial feature vector of the lesion before the cross-attention module. Thus, the lesions l_+ and l_- used for the prediction contain image data from both their own feature vector and the feature vectors of other lesions after the self-attention module. Therefore, these secondary lesions are also associated with the prediction of 2-year PFS with the same type of influence on it (positive or negative) as the lesion (l_+ or l_-) to which it pays attention in the self-attention module. This ability to identify lesions associated with either a positive or negative 2-year PFS could, following appropriate external validation, provide clinicians with valuable diagnostic insight and address the current lack of information on how individual lesions may influence treatment response.

Interpretability at the feature level Another advantage of our model is its ability to extract information about the most important features for predicting 2-year PFS. First, there is a positive bias towards some clinical features that are more strongly

represented in the vector of projected clinical data $\bar{\mathbf{c}}^{(n)}$, *i.e.* the sum of the weights associated with these values in the fully connected layer projecting the clinical data is higher than the sum of the weights associated with other features (see Fig. 6.9). More specifically, we observe three salient features: the Ann Arbor stage of the disease, the number of extranodal sites and the binary value indicating whether the patient underwent salvage therapy.

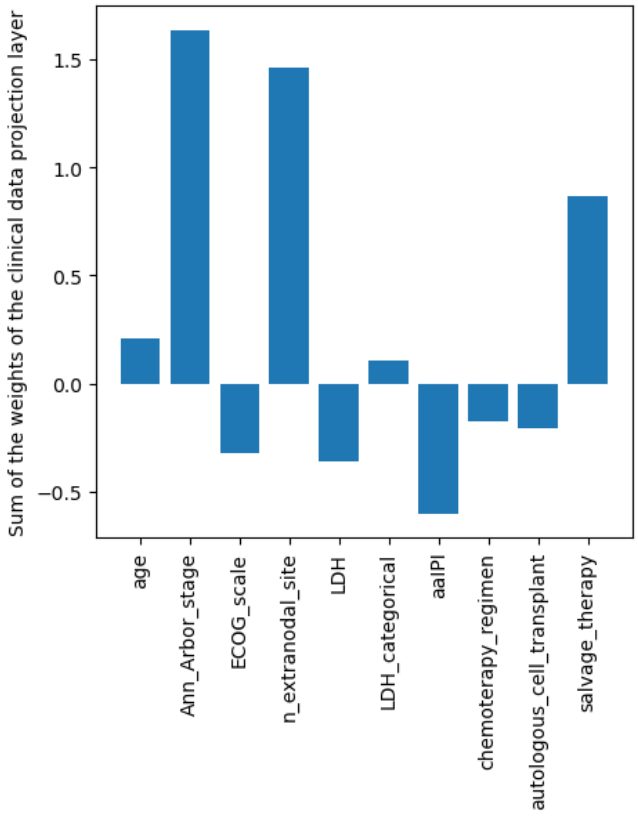
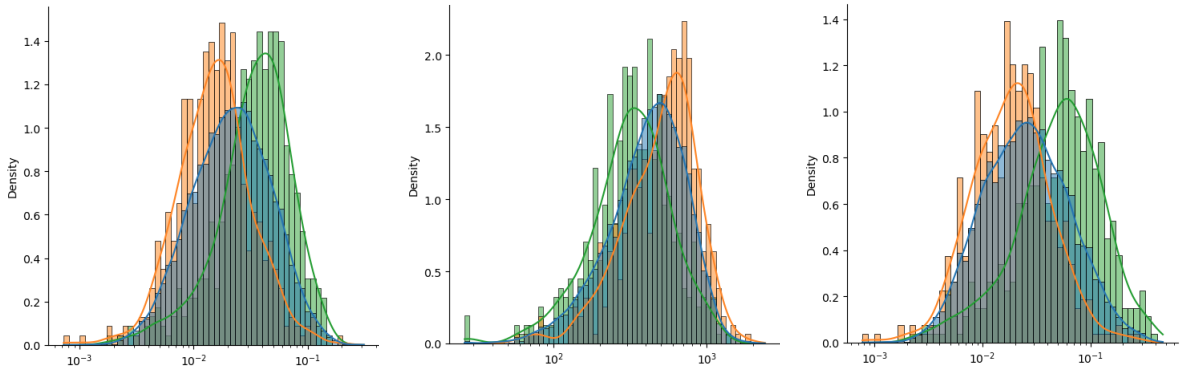


Figure 6.9 – Sum of the clinical Fully Connected layer weights, for each clinical characteristic.

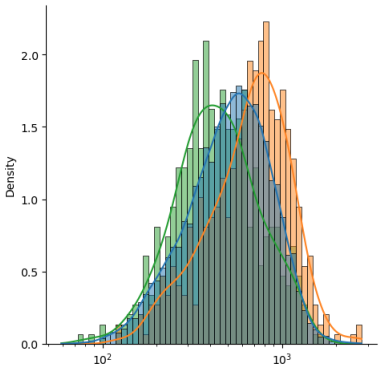
Regarding the image-based information, we observe that some features of the lesions l_+ and l_- (among the image features of the Featv5 set) have statistically different distributions, either one lesion against the other or each of them against the distribution of the features of all lesions. This could mean that these features are important for identifying the lesions that are essential for predicting 2-year PFS. This behavior is observed in particular for a subset of 9 textural features presented in Fig. 6.10.



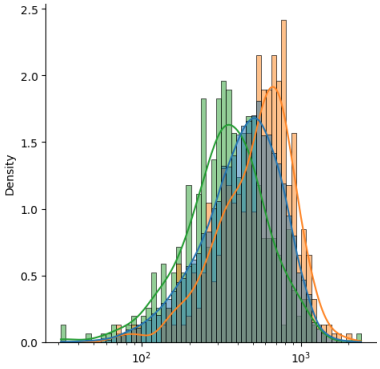
(a) original glrlm ShortRunLowGrayLevelEmphasis BC

(b) original glrlm ShortRunHighGrayLevelEmphasis BC

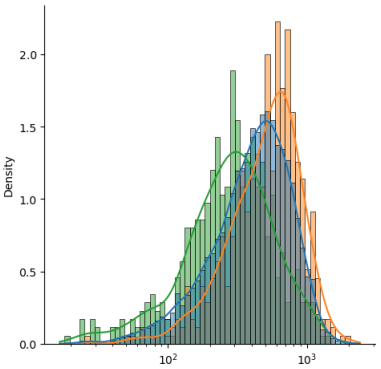
(c) original glrlm LowGrayLevelRunEmphasis BC



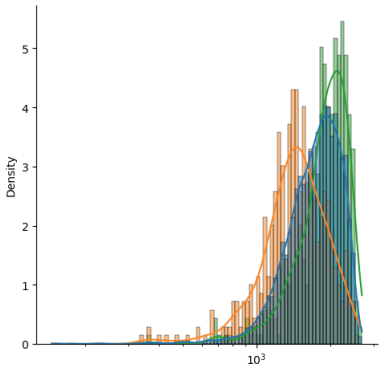
(d) original glrlm LongRunHighGrayLevelEmphasis BC



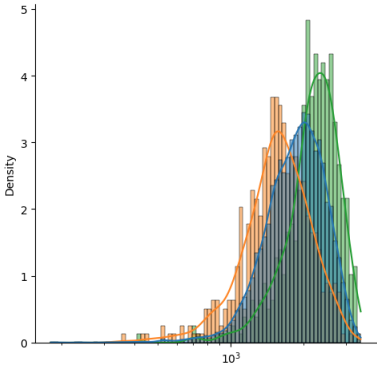
(e) original glrlm HighGrayLevelRunEmphasis BC



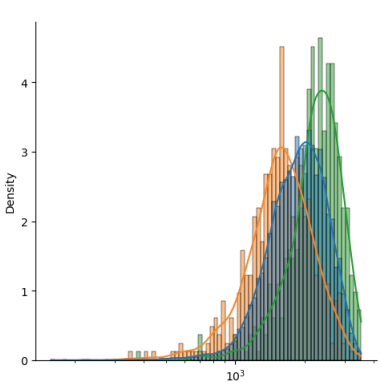
(f) original gldm HighGrayLevelEmphasis BC



(g) log sigma=6mm 3D glrlm ShortRunHighGrayLevelEmphasis BC



(h) log sigma=6mm 3D glrlm HighGrayLevelRunEmphasis BC



(i) log sigma=6mm 3D gldm HighGrayLevelEmphasis BC

Figure 6.10 – Imaging features for which the distribution of the retained lesions l_+ (in orange), l_- (in green) and all the lesions of the patients (in blue) are statistically different.

6.4 Discussion

We addressed in this part the problem of identifying high-risk DLBCL patients by efficiently fusing their imaging and clinical data to automatically predict their 2-year PFS. The methodology and experimental results have been published in an article at the AI4TREAT workshop of MICCAI 2023 [Thiery, Rizkallah, Bailly, et al. 2023], and in the journal *Computerized Medical Imaging and Graphics* [Thiery, Rizkallah, Bailly, et al. 2025]. In summary, we have proposed multiple contributions:

— First, we designed a model that is able to exploit imaging information from a variable number of lesions thanks to the creation of a lesion graph. This information is then fused with tabular clinical data through a cross-attention module, which is able to deal with the significant difference in dimensionality and prognostic strength of the two modalities, as well as the variable number of lesions across patients. Several studies of this graph-based cross-attention model, exploring both the input features to be used and the structure of the model, have highlighted the complexity of constructing a model that performs well on complex, real-world, multicentric, unbalanced, and multimodal data.

— This model outperforms baselines based on different input features (either based on clinical data, imaging data or both) and fusing the imaging data from multiple lesions in variable ways. Our results have been validated on multiple splits of train/validation/test patients (cf. Fig. 5.2), proving the robustness of our model to the data.

— Our experiments have shown the interest of using a large number of image features in the case of multimodal fusion, but only if special attention is paid to defining the structure of the model so that this dimensionality does not prevent the exploitation of the features of other modalities.

— Regarding the choice of cross-attention module inputs, we observed that using the modality with the higher predictive power as the 'Value' matrix outperforms the other strategies for multimodal fusion.

— Finally, the strength of our model in terms of interpretability was highlighted: we showed how its structure allows clinicians to better understand the prediction and gain clinical insight into which clinical features and lesions are most relevant for predicting 2-year PFS in DLBCL.

However, a limitation of our work is that the prediction performance is not yet high enough for clinical deployment, and therefore, the results regarding the relevance of the features considered should be interpreted with caution.

Overall, beyond just seeking performance improvement, this work could be extended

in multiple ways. For example, an interesting development would be to consider cost functions specifically adapted to survival tasks (cf. Part I, Sec. 3.2). This would allow us to take advantage of patients censored before two years, and to obtain a more accurate representation of patients' risk in the short and long term by predicting full survival curves instead of a binary classification estimation. We would also like to validate the robustness of our best model and its ability to generalize to other pathologies presenting similar dispersion of the lesions, such as multiple myeloma [Kraeber-Bodéré et al. 2022].

Part IV

Population-graph for post-hoc correction of survival predictions for improved risk stratification

Abstract

The following part addresses how to make use of the last piece of information at our disposal to improve the stratification of DLBCL patients, beyond PET imaging and clinical data: *i.e.*, we aim at leveraging the predictions and treatment responses of other patients in the cohort. To this end, we propose a model that corrects the initial predictions of a baseline survival model relying on a time-aware population graph. This graph is designed to connect patients whose baseline predictions exhibit similar errors, while also considering their event time. The structure of the graph allows us to anticipate the errors made by the baseline method and to propagate corrections accordingly. These corrected predictions are then used to stratify patients into two risk groups. The proposed framework is validated using the METABRIC dataset for survival analysis in the context of breast cancer, which has a lower censoring rate than the GAINED dataset.

7.1 Introduction

As in the previous parts, we aim to identify high-risk DLBCL patients from the information at our disposal at the time of diagnosis in order to better personalize their treatment. While we previously simplified this risk stratification task to a 2-year PFS classification, it is usually more interesting to predict survival curves that bring richer information on the patient's survival and then use these curves to split the population into risk groups. As described in Part I, Sec. 3.2, several methods exist to predict such survival or risk probabilities from individual patient vector descriptors.

Beyond using graphs to model imaging data and deal with the variable number of lesions of the patients, another way of taking advantage of graphs to improve our DLBCL risk stratification are population graphs. Such graphs take advantage of the similarity between patients to improve the performance of individual predictors [Parisot et al. 2018; Zhang et al. 2023; Zheng et al. 2022]. In this part, we are specifically interested in population graphs in the contexts of feature and label propagation and error correction. As a reminder, as their names suggest, the two first methods respectively imply propagating the node features to learn from neighbours' representations, [Bintsi et al. 2023; Zhang et al. 2023; Zheng et al. 2022] or directly propagating the nodes labels in a semi-supervised way [Ghorbani et al. 2021; D. Zhou et al. 2003] over the graph structure. In our proposed framework, we choose to rely on the work of [Shi et al. 2021], as it combines these two processes, propagating both features and known labels between samples. As for graph-based error correction methods, applied on the predictions of a trained baseline model, they compute the prediction error of known samples and propagate these errors (usually called *residuals*) on a population graph, in order to correct the predictions of new samples [Q. Huang et al. 2020; Raymond et al. 2024; Yang et al. 2023]. Such methods have been applied to regression [Raymond et al. 2024], classification [Q. Huang et al. 2020; Yang et al. 2023] or language modeling tasks [Yang et al. 2023], but to the best of our knowledge, the application of graph-based post-hoc correction methods on survival predictions has not been explored yet.

In this context, we introduce a novel framework that takes a survival analysis model used for risk stratification, and improves a posteriori its ability to divide patients into low- and high-risk groups by correcting the error of newly predicted individuals by leveraging the residuals of similar patients. Our proposition takes inspiration from the Error Passing Network (EPN) [Raymond et al. 2024], designed to correct regression errors (which task was the closest to our objective). In this section we propose to extend this approach to correct survival predictions. Such extension poses several challenges, detailed in Sec. 7.2.2.3: handling censorship, taking into account the times-to-event to build the graph, adapting the scalar corrections to survival curves, and keeping the

probabilistic nature of the corrected survival curves.

To cope with the above challenges, we propose a survEPN method that first predicts time-to-event pseudo-labels which enables building time-aware connections between patients. These connections are then used for refining the survival predictions by propagating the prediction error of labeled patients. The first stage is inspired by [Taherkhani et al. 2019], who predict pseudo-labels for the unlabeled data with a matrix completion approach and connect data points if their true/pseudo-labels are of the same class. In a similar way, we take advantage of the training event times and predict pseudo-labels for the test patients in order to build a time-to-event-aware graph. However, matrix completion is sensitive to outliers and noisy observations, and makes restrictive assumptions on the rank of the matrix [Ahn et al. 2021]. Therefore, we instead opt for a transformer [Shi et al. 2021], which propagates the known labels along with the patient's features towards the unlabeled patients of the cohort to predict pseudo-labels for them. We then rely on the labels and pseudo-labels similarities to build the time-aware population graph, connecting labeled and unlabeled patients. Finally, we deploy on this graph our correction module, specifically adapted to survival analysis, to refine the base model predictions and separate the risk groups on the basis of the corrected survival curves.

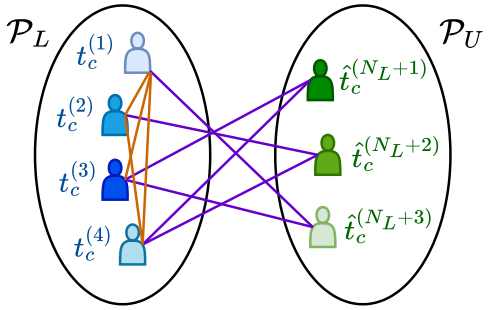


Figure 7.11 – Population graph construction: the bipartite graph with violet edges is built by connecting the labeled patients (from $\mathcal{P}_L$) to unlabeled patients (from $\mathcal{P}_U$) based on the similarity between $\mathcal{P}_L$ patients' labels $t_c^{(j)}$ and $\mathcal{P}_U$ patients' pseudo-labels $\hat{t}_c^{(i)}$; in addition brown edges connect $\mathcal{P}_L$ patients together based on the similarity between their known true labels. N_L corresponds to the number of labeled patients.

7.2 Method

7.2.1 Problem statement

Assume we have a population $\mathcal{P}$ with N_{pop} patients. Let $\mathbf{x}^{(n)} \in \mathbb{R}^{D_{\text{feat}}}$ be the D_{feat} dimensional feature vector associated with a patient $n \in \{1, \ldots, N_{\text{pop}}\}$, $\delta^{(n)}$ its

censorship indicator equal to 0 if the patient is censored and 1 otherwise, and $t_c^{(n)}$ the continuous label corresponding to the patient's event time $E^{(n)}$ when uncensored ($\delta^{(n)} = 1$) and to its censoring time $C^{(n)}$ otherwise ($\delta^{(n)} = 0$). We also consider a base model $f_\Theta(\cdot)$ parameterized by Θ, which takes as input the feature vector $\mathbf{x}^{(n)}$ of a patient n and predicts a continuous survival curve $\hat{S}_{BM}^{(n)}(t) = f_\Theta(\mathbf{x}^{(n)}, t)$. As a reminder (cf. Part I, Sec. 3.1), this survival curve models the probability of an event taking place after time t given the observed features $\mathbf{x}^{(n)}$, *i.e.* $S^{(n)}(t, \mathbf{x}^{(n)}) = \mathrm{P}(E^{(n)} > t|\mathbf{x}^{(n)})$ for $t > 0$. A prediction for the time-to-event $\hat{E}^{(n)}$ can be derived from the predicted survival curve, by taking the time at which the survival curve reaches a probability of 0.5.

We seek to stratify population $\mathcal{P}$ into a low $\mathcal{P}_{low}$ and a high $\mathcal{P}_{high}$ risk groups, such that $\mathcal{P} = \{\mathcal{P}_{low} \cup \mathcal{P}_{high}\}$ and that $\mathcal{P}_{low}$ and $\mathcal{P}_{high}$ are well separated, as evaluated by a log-rank test on their Kaplan-Meyer curves (cf. Part I, Sec. 3.3.3). Specifically, our objective is to improve the risk stratification ability of the base model f_Θ by taking advantage of a population graph built in a semi-supervised way and an error correction strategy.

To this end, our proposed framework consists of three stages. The first step computes time-aware similarities between patients to create an attributed population graph $\mathcal{G}_{\text{pop}}$. This population graph connects labeled (train) patients to unlabeled (test) ones according to their time-to-event predictions, as well as labeled patients together according to their labels (cf. Fig. 7.11). Then, the correction stage computes the errors of the predicted survival curves of the labeled population of N_L patients, and propagates the obtained residuals on the $\mathcal{G}_{\text{pop}}$ graph from the labeled towards all patients, to obtain a refined set of predictions $\{\tilde{S}^{(n)}\}_{n=1}^{N_{\text{pop}}}$. Finally, these corrected survival curves are used to split the patients into improved risk groups (compared to those derived from the initial predictions).

In the next sections, we introduce the original Error Passing Network (EPN) [Raymond et al. 2024] defined for a regression task, discuss the challenges facing its extension to the survival analysis set-up and describe in detail our proposed solution.

7.2.2 Original EPN

Formally, the original EPN addresses prediction correction in the context of a regression problem. For regression, the target value of the base model $f_\Theta(\cdot)$ is a scalar u and the corresponding predictions $\hat{u}^{(n)} = f_\Theta(\mathbf{x}^{(n)})$ are also scalar. A first part of the population is assumed to have access to both features and true target values, *i.e.* $\mathcal{P}_L = \{\mathbf{x}^{(j)}, u^{(j)}\}_{j=1}^{N_L}$, while the second has only access to features $\mathcal{P}_U = \{\mathbf{x}^{(i)}\}_{i=N_L+1}^{N_{\text{pop}}}$, and $\mathcal{P} = \{\mathcal{P}_L \cup \mathcal{P}_U\}$. The EPN seeks to improve the quality of the initial predictions $\hat{u}^{(n)}$ by

computing the residuals $\{\epsilon^{(j)}\}_{j=1}^{N_L}$ for the labeled patients in $\mathcal{P}_L$ and propagating them towards the whole population $\mathcal{P}$ to obtain improved predictions $\{\tilde{u}^{(n)}\}_{n=1}^{N_{\text{pop}}}$. The propagation operation is done over an attributed population graph $\mathcal{G}_{\text{EPN}} = \{\mathcal{V}_{EPN}, \mathcal{E}_{EPN}\}$, where each of the N_{pop} nodes in $\mathcal{V}_{EPN}$ represents a patient and where edges $e_{EPN}^{(n,j)} \in \mathcal{E}_{EPN}$ with weights $\alpha_{EPN}^{(n,j)}$ connect a pair of patients (n,j).

7.2.2.1 Building the population graph

The edges $e_{EPN}^{(n,j)}$ of the EPN's population graph connect any patient $n \in \mathcal{P}$ to all labeled individuals $j \in \mathcal{P}_L$ (with no self-loops), defining a neighbourhood $\mathcal{N}_{EPN}(n) = \{j | e_{EPN}^{(n,j)} \in \mathcal{E}_{EPN}\}$. Raymond et al. 2024 added learned edge weights $\alpha_{EPN}^{(nj)}$ to these connections, computed as:

$$\alpha_{EPN}^{(n,j)} = \text{softmax}(\{s_{EPN}^{(n,j)}\}_{j|j\in\mathcal{N}_{EPN}(n)}), \tag{7.7}$$

with:

$$s_{EPN}^{(n,j)} = \frac{(\mathbf{W}_{EPN,Q}\mathbf{x}^{(n)})^\top \mathbf{W}_{EPN,K}\mathbf{x}^{(j)}}{\sqrt{D_{\text{feat}}}}, \tag{7.8}$$

and where D_{feat} is the number of covariates in the feature vector $\mathbf{x}^{(n)}$, and $\mathbf{W}_{EPN,Q}$ and $\mathbf{W}_{EPN,K}$ are two projection matrices learned with a MSE loss. This edge weight computation is similar to the attention process of a transformer as presented in Part I, Sec. 4.3.5, but with no self-attention.

7.2.2.2 Propagating the errors and correcting predictions

The EPN method then computes the residuals for the labeled patients $j \in \mathcal{P}_L$, that is $\epsilon^{(j)} = u^{(j)} - f_\Theta(\mathbf{x}^{(j)})$, and propagates the corresponding corrections towards the nodes $n \in \mathcal{P}$ following the population graph as follows:

$$\tilde{u}^{(n)} = \left(\sum_{j|j\in\mathcal{N}_{EPN}(n)} \alpha_{EPN}^{(n,j)} (u^{(j)} - f_\Theta(\mathbf{x}^{(j)})) \right) + f_\Theta(\mathbf{x}^{(n)}). \tag{7.9}$$

7.2.2.3 Challenges of extending the EPN method to survival analysis tasks

Extending such a graph-based error propagation method to survival predictions raises multiple challenges. Contrary to a regression or classification problem, a survival model must compare a scalar label (event time) and a curve prediction: thus, error computations must be adapted to compare these two forms of survival representations. A second challenge is to maintain after correction the specific properties of survival curves (*e.g.* bounded and monotonically decreasing) to guarantee their correct interpretation. The third challenge is linked to censoring, *i.e.* having patients for whom we do not know

the true event time but just the time at which they left the study. Therefore, the error computation of these patients must also be adapted. Finally, the definition of the population graph on which the residuals are propagated is not straightforward. Indeed, our assumption is that patients with similar event times should have their survival curves corrected in a similar way, *i.e.* increased before and decreased after their (close) event time. Thus, typical edge weights defined based on the similarity of patient features are not suited to capture the required temporal information in the relationships, especially when the population size is limited.

7.2.3 Overview of the proposed survEPN framework

As mentioned earlier, our objective is to improve the risk stratification ability of a base model f_Θ, such that the curves of the two risk groups are better separated after the correction, *e.g.* as measured by a log-rank test.

To this end, we propose a population graph method, similar to the EPN but adapted to survival analysis, hereafter called survEPN. Our method takes as input the concatenated features $\mathbf{X} \in \mathbb{R}^{N_{\text{pop}} \times D_{\text{feat}}}$ of all the patients, their labels $\{t_c^{(n)}\}$ and censorship values $\{\delta^{(n)}\}$ when available (for the labeled patients from $\mathcal{P}_L$), and all the survival curves $\{\hat{S}_{BM}^{(n)}\}$ predicted by f_Θ. The model is designed to return an improved risk group separation $\mathcal{P} = \{\mathcal{P}'_{low} \cup \mathcal{P}'_{high}\}$, where $\mathcal{P}'_{low}$ and $\mathcal{P}'_{high}$ stand for the low and high risk groups respectively obtained after the survEPN correction step. We can also retrieve, is necessary, the corrected survival curves $\{\tilde{S}^{(n)}\}$ from which the patients are stratified into risk groups.

The proposed survEPN method is decomposed into three stages: the first two steps, as for the EPN, are the graph construction and then the computation and propagation of the corrections. The third and final step relies on the corrected curves to split the patients into risk groups. These stages and their components are illustrated in Fig. 7.12.

For the first semi-supervised stage, finding similarities between patients to create the population graph could be done in principle by comparing the feature description for every pair of patients. However, this could miss the temporal relationship needed for a good correction process, since, as explained before, patients must be connected to neighbours having similar event times for their survival curve to be properly corrected. Therefore, we opt for building a graph relying instead on the similarity of the times-to-event. The first step of survEPN is thus to create these time-aware connections between patients.

Formally, a part of the population is assumed to have access to both features as well as true censorship and time-to-event information, *i.e.* $\mathcal{P}_L = \{\mathbf{x}^{(n)}, \delta^{(n)}, t_c^{(n)}\}_{n=1}^{N_L}$, while the second has only access to features $\mathcal{P}_U = \{\mathbf{x}_n\}_{n=N_L+1}^{N_{\text{pop}}}$, and $\mathcal{P} = \{\mathcal{P}_L \cup \mathcal{P}_U\}$. The

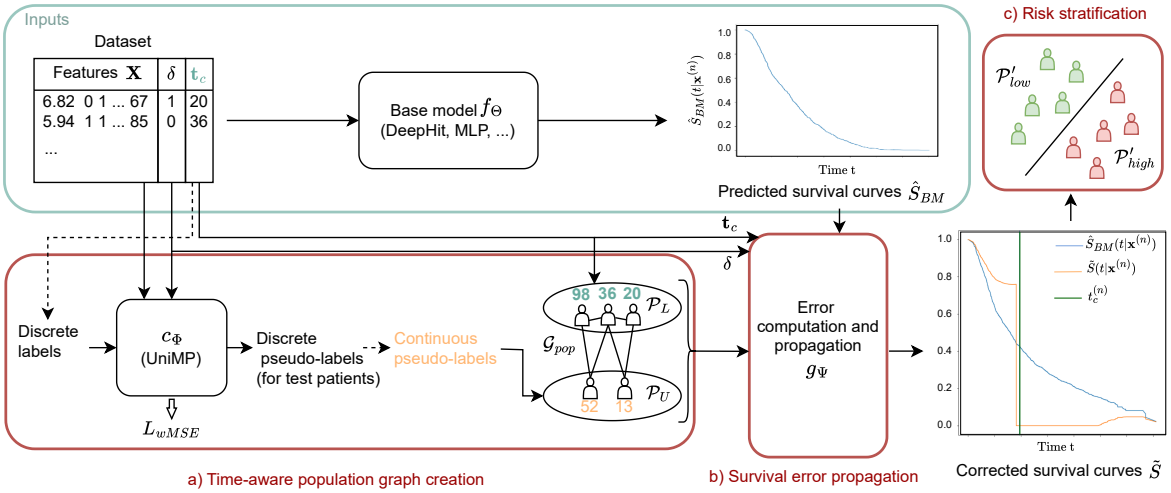


Figure 7.12 – Method overview: our survEPN framework takes as input the feature vectors describing all patients $\mathbf{X}$, their continuous labels $\mathbf{t}_c$, their event indicator $\boldsymbol{\delta}$ and their survival curves predicted by the base model f_Θ, which we want to improve, $\hat{S}_{BM}$. It then (a) creates a time-aware population graph $\mathcal{G}_{\text{pop}}$ by learning to predict pseudo-labels for test patients with a model c_Φ, and connecting these $\mathcal{P}_U$ test patients to the $\mathcal{P}_L$ training patients whose ground truth label is the closest (as well as connecting labeled patients based on their labels). In our case, c_Φ is a Unified Message Passing (UniMP) model [Shi et al. 2021]. After that, (b) the correction module g_Ψ propagates over this population graph the prediction error of the $\hat{S}_{BM}$ curves. Finally, (c) the patients are divided into low- $\mathcal{P}'_{low}$ and high-risk $\mathcal{P}'_{high}$ groups based on their corrected survival curve $\tilde{S}$. In this framework, only the c_Φ module is trained, relying on the loss function L_{wMSE}, without taking into account the subsequent error propagation process.

objectives of this step are to make pseudo-label predictions $\hat{t}_c^{(n)}$ for all patients in the unlabeled dataset $\mathcal{P}_U$ and to find similarities between patients based on their predicted and true times to event. To this end, we design and train a model c_Φ to predict times-to-event for unlabeled patients from their own features but also from features of other unlabeled patients and features and labels of the labeled population. We then connect unlabeled patients i to labeled patients j, based on the similarities between pseudo-labels and true times to event, *i.e.* $\text{sim}(\hat{t}_c^{(i)}, t_c^{(j)})$; and labeled patients together based on the similarities between their known true labels. Finally, the computed similarities guide the construction of the time-to-event-aware graph $\mathcal{G}_{\text{pop}}$ (cf. Fig. 7.11) by determining the edge connections and weights.

In the error correction stage, we start by making predictions for all the subjects, unlabeled or not, using the baseline model to get $\{\hat{S}_{BM}^{(n)}(t)\}_{n=1}^{N_{\text{pop}}}$. We then design a module $g_\Psi(\cdot)$ to compute the prediction error of each patient from $\mathcal{P}_L$, similar to the original EPN but with two main modifications to deal with survival predictions (cf. Sec. 7.2.6). Then, for each patient in $\mathcal{P}$ the correction module adds the weighted sum of its neighbours' survival residuals to obtain corrected curves $\tilde{S}^{(n)}$.

Finally, these corrected curves are used to divide the unlabeled patients into $\mathcal{P}'_{low}$ and $\mathcal{P}'_{high}$, based either on the Area Under their Survival Curve (AUSC) or on their Estimated Event Time (EET).

The three steps mentioned above are described more formally in the following sections.

7.2.4 Pseudo-labels prediction

For the first semi-supervised stage, and in order to build the population graph that explicitly relies on the event time of the patients, we learn to predict pseudo-labels for the $\mathcal{P}_U$ patients with the module $c_\Phi(\cdot)$. To this end, we take advantage of the patients $\mathcal{P}_L$'s labels by considering $c_\Phi(\cdot)$ as a UniMP (Unified Message Passing) model [Shi et al. 2021]. The UniMP, based on transformer blocks, has the advantage of propagating both the features and true labels of $\mathcal{P}_L$ along with the features of $\mathcal{P}_U$ towards the unlabeled patients in $\mathcal{P}_U$, returning a class prediction for every subject. This enables the incorporation of additional knowledge beyond patients' features to enhance the estimation of their time-to-event.

For the training of c_Φ, we partition $\mathcal{P}_L$ into two distinct subsets: $\mathcal{P}_{L^\bullet}$ with access to the true labels, and $\mathcal{P}_{L^\circ}$, whose labels are deliberately masked (not used as inputs) during training and validation. The number of patients considered in each of the two sets is denoted $N_{L^\bullet}$ and N_{L°, respectively. Formally, during training c_Φ predicts a pseudolabel $\hat{t}_c^{(j^\circ)}$ for every masked patient $j^\circ \in \mathcal{P}_{L^\circ}$, taking into account all the features as well as the unmasked labels; that is, $\hat{t}_c^{(j^\circ)} = c_\Phi\big(\{\mathbf{x}^{(n)}\}_{n=1}^{N_{\text{pop}}}, \{t_c^{(j^\bullet)}\}_{j^\bullet=1}^{N_{L^\bullet}}, \{\delta^{(j^\bullet)}\}_{j^\bullet=1}^{N_{L^\bullet}}\big)$.

The reason why the original UniMP model is based on a classification task is that it necessitates the partially-observed labels to be one-hot encoded for their projection. As a result, we first discretize the continuous labels $\{t_c^{(n)}\}_{n=1}^{N_L}$ into N_{cl} classes to get discrete labels $\{t_d^{(n)}\}_{n=1}^{N_L}$, where $t_d^{(n)} \in [1, \dots, N_{cl}]$. Instead of a regular discretization of the time, we rely on time intervals defined by quantiles, *i.e.* ensuring the same number of training patients in each class. This class definition allows us to deal with the uneven distribution of the event times (cf. Fig. 7.15), artificially creating a balanced classification problem.

Next, we describe the details of the UniMP model, which first learns to project the information from labels to combine them with the features, and then relies on several multi-head transformer blocks to compute the attention between patients and propagate this information, to finally predict a class per patient (as presented in Fig. 7.13).

First, we create the initial feature matrix $\mathbf{H}_0$ by adding to the original patients' features $\mathbf{X}$ the uncensored and unmasked discrete labels available, converted into one-hot encoding $\mathbf{t}_{oh}^{(j^\bullet)} \in \mathbb{R}^{N_{cl}}$ and projected by a learned matrix $\mathbf{W}_T \in \mathbb{R}^{N_{cl} \times D_{\text{feat}}}$ in the

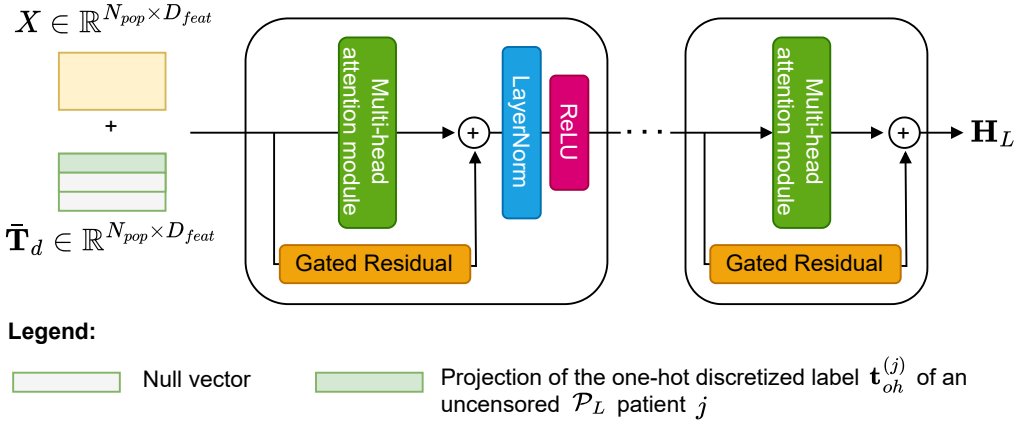


Figure 7.13 – UniMP module overview: the patients' features $\mathbf{X}$ are summed to the projection of the partially observed uncensored labels $\bar{\mathbf{T}}_d$, then updated via transformer layers, which finally returns a discrete time-to-event prediction for all patients $\mathbf{H}_L \in \mathbb{R}^{N_{\text{pop}} \times N_{cl}}$.

same space as the features:

$$\mathbf{H}_0 = \mathbf{X} + \bar{\mathbf{T}}_d, \tag{7.10}$$

with both $\mathbf{H}_0, \bar{\mathbf{T}}_d \in \mathbb{R}^{N_{\text{pop}} \times D_{\text{feat}}}$, and $\bar{\mathbf{t}}_d^{(n)}$ the n-th row of $\bar{\mathbf{T}}_d$:

$$\bar{\mathbf{t}}_d^{(n)} = \begin{cases} \mathbf{t}_{oh}^{(n)} \mathbf{W}_T & \text{if } n \in \mathcal{P}_{L\bullet} \text{ and } \delta^{(n)} = 1 \\ 0_{D_{\text{feat}}} & \text{otherwise.} \end{cases} \tag{7.11}$$

Two types of multi-head transformer blocks are then used: one for intermediate layers and one for the output, the difference lying in the aggregation and activation steps. In both cases, a self-attention module per head is applied on $\mathbf{H}_l$, transforming the features of the l^{th} layer according to learned projection matrices and biases. For the transformer at layer l of the UniMP having M attention heads, the self-attention provides an intermediate representation $\tilde{\mathbf{h}}_{l+1}^{(n)}$ for each patient n:

$$\tilde{\mathbf{h}}_{l+1}^{(n)} = \text{self-attention}(\mathbf{H}_l, \{\mathbf{W}_{l,m,q}, \mathbf{b}_{l,m,q}, \mathbf{W}_{l,m,k}, \mathbf{b}_{l,m,k}, \mathbf{W}_{l,m,v}, \mathbf{b}_{l,m,v}\}_{m=1}^{M}). \tag{7.12}$$

This representation $\tilde{\mathbf{h}}_{l+1}^{(n)}$ is obtained by computing a common query-key-value operation per node $n \in \mathcal{P}$, based on the representation $\mathbf{h}_l^{(n)}$ of all patients, and relying on trainable query ($\mathbf{W}_{l,m,q}$, $\mathbf{b}_{l,m,q}$), key ($\mathbf{W}_{l,m,k}$, $\mathbf{b}_{l,m,k}$) and value ($\mathbf{W}_{l,m,v}$, $\mathbf{b}_{l,m,v}$) parameters per attention head m [3].

3. For each pair of patient, the m^{th} self-attention head first transforms the source node feature $\mathbf{h}_l^{(s)}$ and distant node feature $\mathbf{h}_l^{(d)}$ into query vector $\mathbf{q}_{l,m}^{(s)} \in \mathbb{R}^{d_h}$ and key vector $\mathbf{k}_{l,m}^{(d)} \in \mathbb{R}^{d_h}$ respectively, with d_h the hidden size of each head:

$$\mathbf{q}_{l,m}^{(s)} = \mathbf{W}_{l,m,q} \mathbf{h}_l^{(s)} + \mathbf{b}_{l,m,q}, \tag{7.13}$$

A gated residual connection is introduced between layers to prevent the model from oversmoothing, computing an adaptive parameter $\beta_l^{(n)}$ for each patient and layer from the learned parameters $\mathbf{W}_{l,r}$, $\mathbf{b}_{l,r}$ and $\mathbf{W}_{l,g}$, as follows:

$$\mathbf{r}_l^{(n)} = \mathbf{W}_{l,r}\mathbf{h}_l^{(n)} + \mathbf{b}_{l,r}, \tag{7.18}$$

$$\beta_l^{(n)} = \text{sigmoid}(\mathbf{W}_{l,g}[\,\tilde{\mathbf{h}}_{l+1}^{(n)}||\mathbf{r}_l^{(n)}||\tilde{\mathbf{h}}_{l+1}^{(n)} - \mathbf{r}_l^{(n)}]\,). \tag{7.19}$$

This parameter is tailored to balance the weight of the previous and newly learned representations of the node. As such, the final output of the intermediate transformer block after normalization and activation is computed as:

$$\mathbf{h}_{l+1}^{(n)} = \text{ReLU}(\text{LayerNorm}((1-\beta_l^{(n)})\tilde{\mathbf{h}}_{l+1}^{(n)} + \beta_l^{(n)}\mathbf{r}_l^{(n)})), \tag{7.20}$$

where ReLU is a non linear activation function [Fukushima et al. 1982] and LayerNorm denotes the layer normalization as detailed in [Ba et al. 2016].

For the last output layer (layer L), the transformer block instead employs averaging over the multi-head outputs and removes the non-linear transformation as follows:

$$\tilde{\mathbf{h}}_L^{(n)} = \frac{1}{M}\sum_{m=1}^{M}\Big[\sum_{d=1}^{N_{\text{pop}}} a_{L-1,m}^{(nd)}\mathbf{v}_{L-1,m}^{(d)}\Big], \tag{7.21}$$

$$\mathbf{h}_L^{(n)} = (1-\beta_{L-1}^{(n)})\tilde{\mathbf{h}}_L^{(n)} + \beta_{L-1}^{(n)}\mathbf{r}_{L-1}^{(n)}. \tag{7.22}$$

This $\mathbf{h}_L^{(n)}$, corresponding to $\hat{\mathbf{t}}_{oh}^{(n)}$ in the following, is of dimension N_{cl}. To get a continuous pseudo-label for each unlabeled or masked patient i, we first take the class corresponding to the argmax of $\hat{\mathbf{t}}_{oh}^{(i)}$, named hereafter $\hat{t}_d^{(i)}$. We then convert it back to continuous time by taking the middle value of this class's time interval. This corresponds to the final continuous pseudo-label $\hat{t}_c^{(i)}$ for patient i.

$$\mathbf{k}_{l,m}^{(d)} = \mathbf{W}_{l,m,k}\mathbf{h}_l^{(d)} + \mathbf{b}_{l,m,k}. \tag{7.14}$$

The attention between this pair of patients is then computed for each head:

$$a_{l,m}^{(sd)} = \text{softmax}\Big(\frac{\mathbf{q}_{l,m}^{(s)\top}\mathbf{k}_{l,m}^{(d)}}{\sqrt{d_h}}\Big). \tag{7.15}$$

Then, for each of the M attention heads we do the message aggregation from the N_{pop} distant patients d to the source s, and aggregate them by a concatenation operation:

$$\mathbf{v}_{l,m}^{(d)} = \mathbf{W}_{l,m,v}\mathbf{h}_l^{(d)} + \mathbf{b}_{l,m,v}, \tag{7.16}$$

$$\tilde{\mathbf{h}}_{l+1}^{(s)} = ||_{m=1}^{M}\Big[\sum_{d=1}^{N_{\text{pop}}} a_{l,m}^{(sd)}\mathbf{v}_{l,m}^{(d)}\Big]. \tag{7.17}$$

To train the UniMP, we use a multi-class weighted MSE loss penalizing more the predictions that are far from the true class in an ordinal way (*i.e.* as in ordinal classification methods). For the loss computation on censored patients, we avoid considering the predictions of classes after their censoring time. This gives the following formula:

$$L_{wMSE} = \frac{1}{N_{L^\circ} N_{cl}} \sum_{j^\circ=1}^{N_{L^\circ}} \sum_{k=1}^{N_{cl}} \Big[\mu_k^{(j^\circ)} \cdot \overbrace{\big(\mathbf{t}_{oh,k}^{(j^\circ)} - \text{softmax}(\hat{\mathbf{t}}_{oh}^{(j^\circ)})_k\big)^2}^{\text{MSE over the prediction of each class k}} \Big], \tag{7.23}$$

$$\text{with } \mu_k^{(j^\circ)} = \underbrace{\big(\delta^{(j^\circ)} |t_d^{(j^\circ)} - k| + (1 - \delta^{(j^\circ)}) \max((t_d^{(j^\circ)} - k), 0)\big)}_{\text{Censoring-dependent ordinal penalization}}$$

and $\mathbf{t}_{oh,k}^{(i)}$ being the k^{th} element (so the element of the k^{th} class) of the one-hot vector encoding the discrete label of patient i; and $\text{softmax}(\hat{\mathbf{t}}_{oh}^{(j^\circ)})_k$ the k^{th} element (so the element of the k^{th} class) of the output vector of the UniMP after a softmax operation [4]. Thus, our method avoids enforcing predictions after censoring time, and penalizes more strongly wrong predictions from classes far from the true class in an ordinal way.

During test time, c_Φ is applied to the unseen population $\mathcal{P}_U$ to predict pseudolabels $\hat{t}_c^{(i)}$.

7.2.5 Time-aware population graph creation

With these predicted pseudo-labels $\{\hat{t}_c^{(i)}\}$, we can build our population graph based on the event times. To this end, we first define bipartite edge weights based on the similarity $s^{(ij)}$ between each unlabeled patient i and labeled patient j, computed as:

$$s^{(i,j)} = \text{sim}(\hat{t}_c^{(i)}, t_c^{(j)}) = \frac{1}{|\hat{t}_c^{(i)} - t_c^{(j)}| + \omega}, \tag{7.28}$$

4. For an example of computation of this loss and its comparison against a basic MSE, let's consider a situation with $N_{cl} = 3$ classes and two patients p and q with labels $t_d^{(p)} = 2$, $\delta^{(p)} = 1$, $t_d^{(q)} = 1$ and $\delta^{(q)} = 0$. For a UniMP model whose output predictions are $\text{softmax}(\hat{\mathbf{t}}_{oh}^{(p)}) = \{0.3, 0.5, 0.2\}$ and $\text{softmax}(\hat{\mathbf{t}}_{oh}^{(q)}) = \{0.25, 0.4, 0.35\}$, we get:

$$L_{wMSE}^{(p)} = (0 - 0.3)^2 \cdot 2 + (0 - 0.5)^2 \cdot 1 + (1 - 0.2)^2 \cdot 0, \tag{7.24}$$

$$L_{wMSE}^{(q)} = (0 - 0.25)^2 \cdot 0 + (1 - 0.4)^2 \cdot 0 + (0 - 0.35)^2 \cdot 0, \tag{7.25}$$

while for a normal MSE loss:

$$L_{MSE}^{(p)} = (0 - 0.3)^2 + (0 - 0.5)^2 + (1 - 0.2)^2, \tag{7.26}$$

$$L_{MSE}^{(q)} = (0 - 0.25)^2 + (1 - 0.4)^2 + (0 - 0.35)^2. \tag{7.27}$$

with $|\cdot|$ meaning the absolute value and ω being a small value to avoid dividing by 0. In the same way, we compute the similarity between all pairs of distinct labeled patients $p, q \in \mathcal{P}_L$ based on their known true labels:

$$s^{(p,q)} = \text{sim}(t_c^{(p)}, t_c^{(q)}) = \frac{1}{|t_c^{(p)} - t_c^{(q)}| + \omega}. \tag{7.29}$$

Based on the above similarities, the graph connectivity $\mathcal{E}_{pop}$ is defined using a K-Nearest Neighbors (K-NN) approach, where only the K most similar neighbors for each patient n are retained. The similarities for the unlabeled patients are computed relying only on Eq 7.28, and for labeled patients only on Eq. 7.29. This makes a *directed* graph, meaning that one patient can have an edge towards another without the opposite being true.

Finally, edge weights $\alpha^{(n,d)}, \forall n, d \in \mathcal{P}$, are computed on $\mathcal{E}_{pop}$ edges based on the softmax of the similarities $s^{(n,d)}$. This way, the contribution of the corrections computed in the neighbourhood $\mathcal{N}_p(n)$ of patient n is weighted as:

$$\alpha^{(n,d)} = \text{softmax}(\{\tau s^{(n,d)}\}_{d|d\in\mathcal{N}_p(n)}), \tag{7.30}$$

τ being a temperature factor to control the distribution of the attentions in the population graph (controlled as a hyperparameter). As $\alpha^{(n,d)}$ does not depend on time, we are dealing with a static graph: the only elements evolving depending on time are the initial survival curve predictions and their corresponding errors.

The result of this stage is the population graph $\mathcal{G}_{\text{pop}} = \{\mathcal{V}_{pop}, \mathcal{E}_{pop}, \mathbf{W}_{pop}\}$ with $\mathbf{W}_{pop}$ the matrix containing the directed edge weights $\alpha^{(n,d)}$.

7.2.6 Computation and propagation of the survival error

In the second stage of our framework, we take advantage of the previous graph construction step to analyze the errors made by the baseline model f_Θ on the labeled subjects $\mathcal{P}_L$ and compute corrections for the unlabeled set $\mathcal{P}_U$. Corrections are also computed for $\mathcal{P}_L$ patients, but only for guiding the final stratification step of the $\mathcal{P}_U$ patients, as explained in Sec. 7.2.7. Inspired by [Raymond et al. 2024], we refine the predictions with an error propagation module, denoted hereafter g_Ψ. The model g_Ψ takes as input the event and censorship times of the labeled subjects $\{t_c^{(j)}, \delta^{(j)}\}_{j\in\mathcal{P}_L}$, the predicted survival curves of the baseline model on the entire population $\{\hat{S}_{BM}^{(n)}\}_{n\in\mathcal{P}}$, and the time-aware population graph from the prior step $\mathcal{G}_{\text{pop}}$, and predicts a corrected version $\tilde{S}^{(n)}$ of each survival curve:

$$\{\tilde{S}^{(n)}\}_{n\in\mathcal{P}} = g_\Psi(\{t_c^{(j)}, \delta^{(j)}\}_{j\in\mathcal{P}_L}, \{\hat{S}_{BM}^{(n)}\}_{n\in\mathcal{P}}, \mathcal{G}_{\text{pop}}) \tag{7.31}$$

The idea is to correct the survival curves predicted by the base model according to the error made on patients with similar event times.

Apart from the graph construction described in the previous section, we make three other adaptations to the original EPN to deal with survival analysis tasks: firstly, the single-value event-time labels are converted to approximate the true survival function $\{S^{(j)}\}$ of patient j, so that they can be compared to the predicted survival curves; secondly, the error of the censored patients after their censoring time is constrained to be equal to 0; and optionally, we propose a post-processing step to ensure the corrected curves stay bounded and monotonically decreasing as they should.

To first estimate the error of the survival curves predicted by the baseline model on the labeled population $\{\hat{S}^{(j)}_{BM}\}_{j\in\mathcal{P}_L}$, we approximate the true survival function $\{S^{(j)}\}$ of patient j with as a step-function dropping at the event time $t_c^{(j)}$, as it is often done to compute the Brier score (cf. Part I, Sec. 3.3.2). However, to avoid propagating uncertain errors because of censoring, we keep $\{S^{(j)}\}$ equal to the predicted curve on the censored intervals. Formally:

$$S^{(j)}(t) = \begin{cases} 1 & \text{if } t \leqslant t_c^{(j)} \\ 0 & \text{if } t > t_c^{(j)} \text{ and } \delta^{(j)} = 1 \\ \hat{S}^{(n)}_{BM}(t) & \text{otherwise.} \end{cases} \tag{7.32}$$

In this way, the error $S^{(j)}(t) - \hat{S}^{(j)}_{BM}(t)$ equals 0 for censored times.

Next, we correct the survival curves of all patients by propagating these errors following the edges of the previously defined graph. The corrected survival curve for patient n becomes:

$$\tilde{S}^{(n)}(t) = \Big(\sum_{j|j\in\mathcal{N}_p(n)} \alpha^{(nj)} \left(S^{(j)}(t) - \hat{S}^{(j)}_{BM}(t) \right) \Big) + \hat{S}^{(n)}_{BM}(t), \tag{7.33}$$

The resultant survival curves $\tilde{S}^{(n)}(t)$ are clamped between 0 and 1 if they exceed these values. It is important to note that in this step we do not learn any additional parameter, *i.e.* the correction is solely based on the population graph previously built using pseudo-label predictions.

Finally, the corrected curves may not respect the expected properties of survival curves $S^{(n)}(t) = \mathrm{P}(E^{(n)} > t)$, *e.g.* monotonically decreasing and bounded between 0 and 1. To address this issue, we propose an optional extension of our framework. When beyond the stratification of the population, proper survival curves are required

per patient, we fit a sigmoid to the corrected curves as follows:

$$\tilde{S}^{(n)}_{sig}(t) = 1 - \frac{A_{sig}}{1 + \exp(\frac{t - t_{sig}}{s_{sig}})} \tag{7.34}$$

with parameters A_{sig} controlling the amplitude of the sigmoid and bounded between 0 and 1; t_{sig} linked to the time at which the curve decreases, enforced to be higher than 0; and s_{sig} corresponding to the slope of the function, constrained to be negative. These constraints, beyond monotony and boundaries, enforce the fitted curve to be equal to 1 at time $t = 0$, which is also a necessary property of a survival curve.

7.2.7 Risk stratification

Finally, we use the corrected curves $\{\tilde{S}^{(n)}\}_{n\in\mathcal{P}}$ (or $\{\tilde{S}^{(n)}_{sig}\}_{n\in\mathcal{P}}$) to divide the patients into low- and high-risk groups by using the Area Under their Survival Curve (AUSC) or their Estimated Event Time (EET), the latter computed as the time at which the survival curve of a patient equals 0.5 as described in Fig. 7.14. In the two cases, we take the median AUSC or EET value of the training patients $\mathcal{P}_L$'s corrected curves, and use it as a threshold to split the $\mathcal{P}_U$ test patients into $\mathcal{P}'_{low}$ and $\mathcal{P}'_{high}$ from their corrected curves $\{\tilde{S}^{(i)}\}_{i\in\mathcal{P}_U}$.

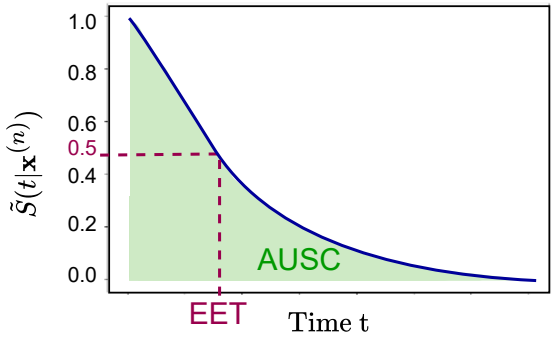


Figure 7.14 – Estimated Event Time (EET) and Area Under the Survival Curve (AUSC) computation for high-risk patient identification.

7.3 Experimental validation

7.3.1 Dataset

The experiments from this section have exceptionally been done on the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) dataset [Curtis et al. 2012], which includes a slightly higher number of patients and, more importantly, a lower censoring rate than the GAINED dataset (42% vs 80%). It consists of gene

and protein expression profiles (with $D_{feat} = 9$ features as described in [Katzman et al. 2018]) used to determine new breast cancer subgroups in order to help clinicians provide better treatment recommendations. It is composed of 1,904 patients, and the events corresponding to the patients' deaths are distributed as shown in Fig. 7.15.

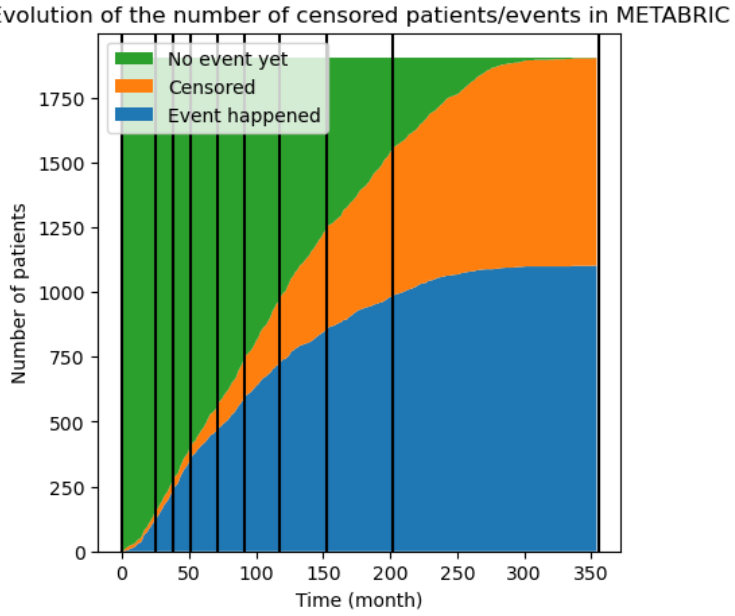


Figure 7.15 – Distribution of censoring and events over time in the METABRIC dataset. The black lines show the quantile classes defined.

7.3.2 Experimental setup

For the following experiments, to account for the random uncertainty in the data, we report the results over five random train-validation/test splits. For each of these five sets of test patients, we perform ten loops of training and validation, with random splits of the rest of the patients in training and validation sets. The proportions of training, validation and test patients in each loop are 80%, 10% and 10% respectively, and we force the proportion of censored patients to be the same in all sets. For fair comparison, these sets are the same as those used to train the base model whose predictions we want to improve. Thus, the results presented hereafter correspond to the test performance of the model with the UniMP having the best validation cAMAE (for censoring-aware Average Mean Absolute Error) of the loop, computed as follows ($\#\mathcal{P}$ being the cardinal of set $\mathcal{P}$):

$$cAMAE = \frac{1}{N_{cl}} \sum_{k=1}^{N_{cl}} \sum_{j^\circ | t_d^{(j^\circ)} = k} \frac{\delta^{(j^\circ)} |t_d^{(j^\circ)} - \hat{t}_d^{(j^\circ)}| + (1 - \delta^{(j^\circ)}) \max(t_d^{(j^\circ)} - \hat{t}_d^{(j^\circ)}, 0)}{\#\{j^\circ | t_d^{(j^\circ)} = k\}} \tag{7.35}$$

The metrics we report are the c-index, IBS and stratification p-values derived from the survival curves as presented in Part I, Sec. 3.3. This test performance is averaged over the five train-validation/test splits and the ten training loops for the IBS and c-index values, while for the p-values we report the median, first and third quartiles of the 50 obtained values (as a mean operation is not relevant considering the variability in the orders of magnitude of the p-values).

In our framework, the training and validation steps correspond to the training of the UniMP module, with the validation patients used for early stopping, *i.e.*, to select the best model among those from different epochs for one fold. The test step corresponds to using the predictions of the trained UniMP to build the population graph and to realize the error propagation and risk stratification processes. In this context and as explained previously, during the training and validation of the UniMP module, we mask the labels of some of the patients in $\mathcal{P}_L$ to compute the L_{wMSE} loss and the cAMAE metric. We illustrate the masking process for training, validation, and test in Fig. 7.16. During training, only the features of the masked patients are used as input, excluding the projection of their one-hot discrete labels in Eq. 7.11. The masked labels instead serve as supervision for computing the weighted Mean Squared Error (MSE) loss (as shown in Eq. 7.23), guiding the update of the learnable parameters of the UniMP to make good pseudo-label predictions for the masked patients. To make the predictions for the validation step, it is the validation patients that are modeled as masked, but unlike in the training stage, the masked labels do not contribute to the loss computation or the parameter update. Instead, masked labels during validation serve to compute the cAMAE, which is our criterion for selecting the best model.

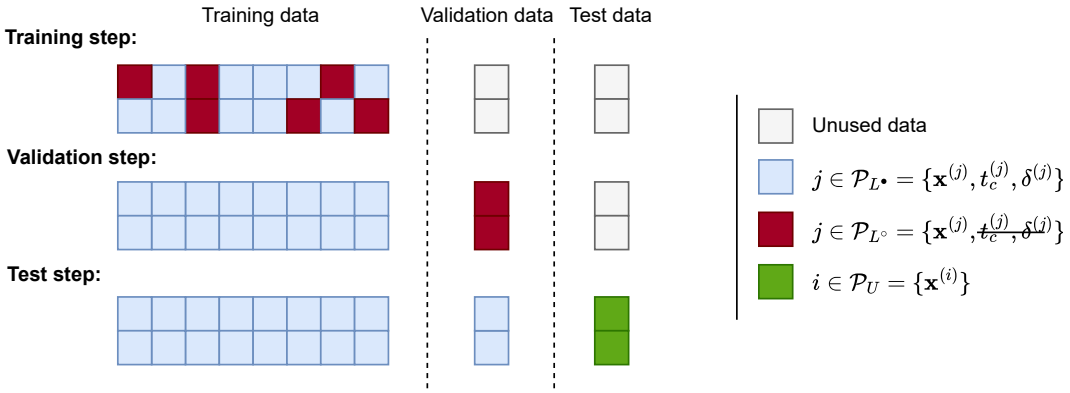


Figure 7.16 – For a given training/validation/test split, definition of the $\mathcal{P}_{L^\bullet}$, $\mathcal{P}_{L^\circ}$ and $\mathcal{P}_U$ sets. Training and validation only concern the UniMP (Sec. 7.2.4)).

The numerical covariates are standardized by removing their mean and scaling their variance to unity to obtain $\mathbf{X}$, while the binary ones are left unchanged. The base model f_Θ considered in the following experiments is a simple 2-layer MLP with ReLU

intermediary activations, trained with the CoxTime survival loss function as presented in Part I, Sec. 3.2, using a batch size of 256 patients on 100 epochs (with early stopping if the loss value does not improve for 10 epochs). We selected its configuration by a grid search on its hidden size (16 from $\{16, 32, 64, 128\}$) and learning rate (0.001 from $\{0.1, 0.01, 0.001, 0.0001\}$). As for the UniMP module, it is trained on 400 epochs (and stopped early if its cAMAE does not improve for 200 epochs) with a dropout rate of 0.3. The difference in the number of epochs between the base model and the UniMP module comes from the difference in their speed of convergence, as the MLP does not improve when trained on more than 100 epochs. An exploration of the hyperparameters of the UniMP led to considering $L = 2$ layers, $M = 2$ attention heads, a hidden size d_h of 32, using a learning rate of 0.1, and discretizing the continuous labels into $N_{cl} = 9$ classes.

As for the graph construction, after performing experiments to find the optimal settings, the temporal graph is defined as a K-NN with K = 400 and $\tau = 0.1$. During the training step of UniMP, we consider as in [Shi et al. 2021] a label rate of 0.625, corresponding to the proportion of unmasked patients $\mathcal{P}_{L\bullet}$ randomly selected from $\mathcal{P}_L$ for training. All models are optimized via Adam [Kingma et al. 2014] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$, including a weight decay of 0.0005 for the UniMP. The whole framework was coded in Python using the `PyTorch` [Paszke et al. 2019], `PyCox` [Geck 2012], `scikit-survival` [Pölsterl 2020], and `pytorch-geometric` [Fey et al. 2019] modules.

7.3.3 Experiments and results

First, we compare the stratification performances of the base model before and after applying our correction framework as explained in Sec. 7.2. The evaluation of this stratification performance relies on the quantiles of the p-values (computed on our 50 folds) between the Kaplan-Meier curves of the two risk groups, in the first and last result columns in Table 7.6. These values show an important improvement in the separation between low- and high-risk patients, proving the strength of our method in improving the stratification ability of a survival model. Fig. 7.17b additionally presents a qualitative and quantitative example of the difference between the Kaplan-Meier curves of the two risk groups before and after the correction.

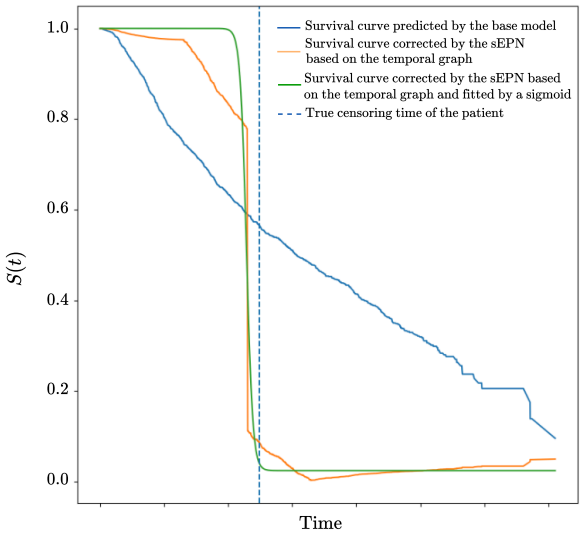
As for the IBS and c-index, their evolutions will be discussed further in this section.

Second, we perform an ablation study to prove both the interest of the survEPN model itself, and of its specific combination with the temporal graph built from the UniMP module. Hereby, we compare the performance of the base model and of our complete framework to two other methods relying on the survEPN module but applied

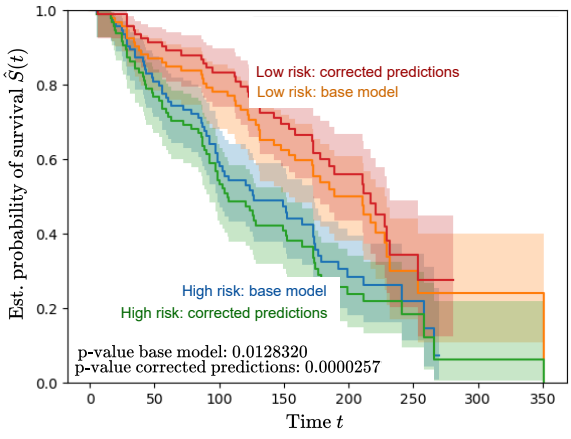
Table 7.6 – Stratification quality and survival performances of the base model, and of the base model corrected by our survEPN method applied on either random graphs; a feature graph; and our temporal graph presented in Sec. 7.2.5. The p-values are computed on the Kaplan-Meier curves of risk groups separated either based on their Estimated Event Time (EET) or the Area Under their Survival Curve (AUSC).

	Quantile	Base model	Random graph survEPN	Feature graph survEPN	Temporal graph survEPN
p-value EET (↓)	25%	1.86×10^{-4}	4.17×10^{-5}	3.53×10^{-6}	1.15×10^{-6}
	50%	2.78×10^{-3}	1.35×10^{-3}	6.86×10^{-5}	1.96×10^{-5}
	75%	1.03×10^{-2}	9.47×10^{-3}	6.44×10^{-4}	1.69×10^{-4}
p-value AUSC (↓)	25%	5.14×10^{-5}	1.54×10^{-5}	1.69×10^{-5}	9.45×10^{-8}
	50%	8.40×10^{-4}	6.26×10^{-4}	1.48×10^{-4}	1.70×10^{-5}
	75%	4.10×10^{-3}	5.18×10^{-3}	2.79×10^{-3}	5.02×10^{-4}
IBS (↓)		0.163 ± 0.004	0.168 ± 0.005	0.165 ± 0.004	0.190 ± 0.007
c-index (↑)		0.66 ± 0.03	0.63 ± 0.03	0.65 ± 0.03	0.64 ± 0.03

on different graphs. To begin with, we apply this module to random graphs [5] (100 different graphs per training/validation/test loop for robustness). We also apply our survEPN module on a graph created on the cosine similarity between the patients' features, also built in a K-NN manner with K = 50 and $\tau = 0.1$. For both graphs the K and τ values have been selected experimentally. The quantitative results are presented in Table 7.6 and some qualitative results are displayed in Fig. 7.17.



(a) Example of survival curve correction.



(b) Kaplan-Meier curves of the two Estimated Event Time (EET) risk groups of the base model's predictions and of the corrected predictions, for one train/validation/test loop.

Figure 7.17 – Qualitative evaluation of the results.

We observe that applying the survEPN on the random graphs to correct the base

5. To build these graphs, we generate a random normal value for each edge, select the 50 highest in a K-NN way, and apply a temperature factor of $\tau = 0.1$ then a softmax on these 50 edges.

model predictions gives equivalent or slightly better stratification results than the base model. However, as soon as the graph becomes informative (as with the feature graph), applying the survEPN on the base model significantly improves its stratification ability, especially when splitting the risk groups based on the estimated event times of the patients. Finally, introducing explicitly the time information in the graph construction further significantly improves the quality of the risk stratification compared to the stratification ability of the base model, even when the split is done based on the area under the survival curves of the patients.

As for the metrics evaluating the quality of the survival curves themselves, we observe that the c-index does not change much, while the IBS is quite degraded from the survEPN based on the temporal graph. Even if our aim is not to improve those metrics, we discuss next why they behave differently from the p-values.

Indeed, while the p-values evaluate how well the patients are split into two groups, the c-index cares about their specific ordering within the groups. Likewise, the degradation of the IBS is primarily due to a few patients experiencing a late event or censoring time but predicted earlier by the UniMP module. As the late classes are broad (cf. Fig. 7.15), an error of one class implies graph connections to patients with significantly earlier events, inducing corrections with low survival probabilities at late times. Such errors are heavily penalized by the IBS. Instead, the p-values evolution depends mainly on the patients who are changing risk groups thanks to the survEPN rather than just the late patients. Thus, even if some survival curves are inaccurate, if a majority of patients are appropriately changing risk groups, the p-values improve despite a decline in the Integrated Brier Score (IBS). It is important to note that a smaller degradation of the IBS would be expected with more balanced datasets, *i.e.* that include a greater number of patients with a later label and thus with thinner late classes.

From a qualitative point of view, we observe from the orange curve in Fig. 7.17a that the curves obtained after our correction process may not respect the expected properties of true survival curves $S^{(n)}(t) = \mathrm{P}(E^{(n)} > t)$, *i.e.* decreasing, equal to 1 at time 0 and always between 0 and 1. We observe qualitatively in Fig. 7.17a the effectiveness of the optional refinement step based on sigmoid fitting (cf. Eq. 7.34) proposed to solve this.

Finally, we evaluate quantitatively the stratification after the optional sigmoid fitting step. The results in the last column of Table 7.7 confirm that the stratification improvement is maintained after the curve refinement, with very close performances to the process without post-processing. This optional step is thus recommended when using our framework to improve risk stratification while retaining the properties of the survival curves, for example for their clinical interpretation.

Table 7.7 – Stratification quality and survival performances of the base model, the base model corrected by our survEPN, and the base model corrected by our survEPN method with sigmoid fitting to get proper survival curves. In this experiment, survEPN relies on a temporal graph. The p-values are computed on the Kaplan-Meier curves of risk groups separated either based on their Estimated Event Time (EET) or the Area Under their Survival Curve (AUSC).

	Quantile	Base model	Temporal graph survEPN	Temporal graph survEPN and sigmoid fitting
p-value EET (↓)	25%	1.86×10^{-4}	1.15×10^{-6}	4.65×10^{-7}
	50%	2.78×10^{-3}	1.96×10^{-5}	1.91×10^{-5}
	75%	1.03×10^{-2}	1.69×10^{-4}	2.98×10^{-4}
p-value AUSC (↓)	25%	5.14×10^{-5}	9.45×10^{-8}	1.04×10^{-7}
	50%	8.40×10^{-4}	1.70×10^{-5}	1.70×10^{-5}
	75%	4.10×10^{-3}	5.02×10^{-4}	4.87×10^{-4}
IBS (↓)		0.163 ± 0.004	0.190 ± 0.007	0.190 ± 0.007
c-index (↑)		0.66 ± 0.03	0.64 ± 0.03	0.63 ± 0.04

7.4 Discussion

We addressed in this part the problem of taking advantage of inter-patient information to improve the stratification ability of a survival prediction model. In a nutshell, our contributions to this task are the following:

— We designed a survival extension to the Error Passing Network, named survEPN, that:
(i) builds and leverages a semi-supervised population graph to model errors of a baseline model and propagate corrections from labeled to unlabeled patients;
(ii) relies on time-aware edges that explicitly account for the event time of the patients when defining graph connections;
(iii) adapts the propagation and correction steps to the survival analysis task, enabling the prediction and correction of survival curves, handling censored data, and splitting patients into risk groups.

— The proposed framework proved its ability to significantly improve the quality of the risk stratification compared to the base model, as long as the underlying population graph is informative enough.

— Our results have been validated on multiple splits of train/validation/test patients, proving the robustness of our method to the data distribution.

— Finally, we proposed a specific way of ensuring that the corrected curves keep their expected properties if needed, with very close performances compared to the original framework.

Regarding the comparison to other approaches evaluated on the public METABRIC

dataset, we found most of them aim at identifying new prognostic biomarkers for the pathology, which differs from our goal. Learning-based methods mainly focus on classification tasks [Khan et al. 2024; Mucaki et al. 2017; Shnain et al. 2024], and previous research work on survival analysis only reports c-index and/or IBS [Katzman et al. 2018; Kvamme and Borgan 2021; Kvamme, Borgan, and Scheel 2019], neither of them reporting stratification performances. The closest works from the literature are [Kvamme and Borgan 2021; Kvamme, Borgan, and Scheel 2019], [Katzman et al. 2018]. The reported C-index and IBS performances in [Kvamme and Borgan 2021; Kvamme, Borgan, and Scheel 2019] are of the same order of magnitude (*i.e.* respectively 0.675 and 0.172 for their best model) as those in Table 7.6. Likewise, the best c-index reported in [Katzman et al. 2018] (equal to 0.65), is of a similar order to us. However, the scores are not directly comparable as they follow a different experimental setup. Unfortunately, two papers proposing a learning-based method for survival analysis and reporting stratification results [Malighetti et al. 2025; Z. Wang, Katsaros, et al. 2023] seem to use another version of the METABRIC dataset, as the number of patients and/or the features declared differ from ours, and the code was not made available for reproduction. Thus, despite our efforts, and given the orthogonal focus of our work, we have oriented our experimental validation towards comparisons against the baseline and ablated versions of our model.

To extend this work, we would like to validate this method on the GAINED dataset to prove its relevance for the main task of the thesis, and to ensure its performance on highly censored datasets. Some other perspectives include exploring alternative ways of integrating the censored labels into the UniMP module, as for the time being we only sum uncensored labels to the features to avoid propagating noise into the model. We would also like to improve the discretization process of this UniMP module, which causes the IBS of the curves to be degraded by the correction. To this end, we envision using a finer discretization to reduce the impact of a classification error on the pseudo-label estimation, and exploring other ways of converting the predicted discrete pseudo-labels to continuous values, for example by sampling a normal distribution instead of using the median value of the class. Additionally, we would be interested in making the survEPN process learnable in an end-to-end way, with a joint formulation of the graph creation and error propagation processes. Finally, we would like to study whether our framework can improve the stratification ability of other state-of-the-art models for survival analysis.

The content of this part has been published in the GRAIL workshop of MICCAI 2025 [Thiery, Rizkallah, Laribi, et al. 2025].

Part V

Conclusion and perspectives

The main goal of this PhD was to assist in the identification of high-risk DLBCL patients by taking advantage of all the available information at the time of diagnosis: PET scans, clinical data and other patients for whom we know the treatment response. To get closer to this goal, we proposed three complementary contributions.

First, we explored in Part II how to best leverage the information from the patients' PET scans for 2-year PFS prediction. To identify the right features for this task, we compared five sets of features composed of distinct radiomic sets extracted from lesions segmented with different strategies. We additionally compared multiple approaches to deal with the varying number of sparsely distributed lesions, and showed the relevance of a lesion graph strategy to get a compact but complete representation of this information.

In a second contribution, we proposed in Part III a new model able to fuse efficiently imaging features from multiple lesions to tabular clinical data. To this end, we rely on a cross-attention mechanism, on which we realized an in-depth study such that it best adapts to our complex multimodal data with different prognostic strength and dimensionality. Our model realizes the fusion for the first time at the lesion level, and a max pooling operation automatically selects a limited number of relevant lesions associated with each prediction, supporting its interpretability. This interpretability is also strengthened by attention processes, and we thus showed the potential of our method to bring new information to the clinicians on the features (both imaging and clinical) most relevant to the decision, and therefore to potentially improve the understanding of the disease. Lastly, we extended our study on the most relevant imaging features for the 2-year PFS prediction task, and notably showed that for simple fusion models it is best to keep a low number of imaging features to not overwhelm the very predictive clinical data, while more elaborate models benefit from a higher number of features.

Finally, in Part IV we extended an error propagation population graph-based method to survival analysis and high-risk patient identification tasks. This included adjusting the error computation process to deal with censoring and survival predictions, adding a post-processing step to enforce the corrected survival curves to keep their probabilistic meaning, and explicitly introducing event times in the population graph creation. For this last step, we proposed to make use of a transformer-based model taking advantage of the training patients' labels in a semi-supervised way. We applied for the first time, to the best of our knowledge, this propagation-based post-hoc correction method to a risk stratification task, showing a strong improvement in the stratification ability of a survival base model.

There would be many ways to extend this work further. We present in the following a few of them, related either to the data and features considered; to the formalisation of the considered problems; to the modeling of the graph structures proposed; or to the multimodal fusion of tabular and image data.

Data and features

First of all, we explored how to extract richer information from PET scans by learning to extract features with a CNN that are enforced to be non-redundant with the radiomic features we have already chosen to use, as in [Vétil et al. 2023]. This work, produced with our two students Maria-Clara VALLE and Ilyes LAMSSALAK, is still unfinished but seems promising to improve the automatic exploitation of PET images, as we tried to do in Part II.

Another way of potentially improving the use of these images beyond what we did in Part II would be to apply the framework of [Loutfi et al. 2024] to automatically select the most relevant and least complex radiomics to extract from the PET scans for our task: it would be interesting to see both which radiomic features would be selected and if it could improve our performances.

Additionally, our frameworks in Parts II and III currently rely on PET scans on which the lesions have been segmented manually by clinicians. However, this segmentation being highly time-consuming and labor-intensive, we explored how to automate this process. To this end, we considered both foundational segmentation models such as UniverSeg [Butoi et al. 2023] and Segment Anything [Kirillov et al. 2023] with Damien GAUTIER and Arthur ASTIER, and models specifically designed for this task as the TMTV-net [Yousefirizi, Klyuzhin, et al. 2024] with Anthonin FALK and Bachar HAMIE, which all showed limited performance on our dataset. Thus, we would like to push this study further, as beyond the gain in labor and time this approach could provide more consistent segmentations (with no intra- and inter-operator variability), which could help the subsequent feature extraction process extract more robust features.

Furthermore, we would like to extend the validation of all our work by applying the proposed methods to external datasets. As such, we plan to evaluate the population graph method from Part IV on the GAINED dataset for the task we are interested in; which is not trivial because of its high censoring rate. It would also be interesting to apply our multimodal graph-based model from Part III to another pathology with consequent dispersion of the lesions, as multiple myeloma for example, to ensure its robustness.

Problem formalisation

Beyond the data and features to use for DLBCL patient stratification, other questions can be raised on the formalisation of the proposed methods and of the problem itself.

For example, another interesting perspective would be to adapt our multimodal model presented in Part III to target a survival analysis task instead of a simple classification, in order to get richer information about the patients in the output. Further-

more, this would allow us to combine this multimodal model with our survEPN module presented in Part IV to fully take advantage of all the data we have for DLBCL stratification at once.

We would also be interested in making the survEPN process presented in Part IV learnable in an end-to-end way, with a joint formulation of the graph creation and error propagation processes.

Additionally, in recent years, there has been growing interest in calibration techniques for deep learning models [Galdran et al. 2023], aimed at aligning predicted probabilities with true confidence levels and thereby enhancing model robustness in a real-world application. It would be interesting to see if there are links between our survEPN correction approach presented in Part IV and calibration methods dedicated to survival [Austin et al. 2020].

Graph based representations

We would also like to refine our graph creation processes, as they are the backbone of our proposed methods and as such have a huge impact on the final quality of our risk stratification. To this end, we identified three major elements to focus on:

- First, it could be interesting to explore other lesion graph structures beyond the one from [Aswathi et al. 2023] presented in Part II, for example by considering a K-NN structure instead of a fully connected one, by exploring different edge weights, etc. This could improve the lesion features fusion and the integration of their dispersion into the model.
- Additionally, in our survEPN model from Part IV, we would like to improve the time-aware population graph creation by exploring various ways of integrating the censored labels in the UniMP propagation process: currently, we do not use them to avoid propagating noise in the module, but they still carry important information that we would like to make use of.
- Lastly, we thought about extending the initial EPN (cf. Part IV) by introducing *negative* connections instead of only positive ones to take advantage not only of the similarity between patients but also of their difference. We did not pursue this study as it was not compatible with our way of correcting the survival curves, but we still believe it to be an interesting perspective as such negative attentions have already been proven relevant in the literature [Bo et al. 2021].

Multimodality

The last type of perspective we propose is related to the fusion of tabular and image data. Indeed, in parallel with our work, some methods were proposed to fuse clinical data with PET scans or other images in rich ways.

- For example, [Larose, Archambault, et al. 2024] rely on a multi-task framework to predict high-grade prostate cancer prognosis, using for each task separately either one or both modalities between clinical and image, and extracting different imaging features for each task. The prediction of multiple dependent clinical endpoints (for example, both the 2-year PFS and a survival estimation of the time-to-event in our case) could help to be more accurate in our estimation of each of them. Considering specific imaging features for each task would also allow us to extract even richer information from the PET scans.
- Very close to our work, [Stym-Popper et al. 2025] fuse time-series embedding of ultrasound images and tabular data for cardiac hypertension diagnosis. To do so, they decouple their modalities by separating the modality-shared and modality-specific information of the tabular data, and use a combination of self-attention and cross-attention processes to asymmetrically fuse all this information. The process of identifying the modality-shared and modality-specific information has been explored multiple times in the literature [T. Chen, Hong, et al. 2022; X. Wang et al. 2020] and has been proven effective to improve the quality of multi-modal fusion, sparking our interest in this work that additionally relies on similar modalities as us.
- Finally, [C. Wu et al. 2024] proposed a method for diagnosing lymph node metastasis in esophageal cancer based on a combination of tabular (clinical and hematological indicators) and imaging data (both radiomic and learning-based features extracted from CT scans). Their method combines the multimodal data by a graph approach (each node corresponding to a modality) followed by a conditional feature-guided diffusion process for feature redundancy elimination, and is able to fuse imaging information from a variable number of lesions with a concatenation followed by an attention process. As for now we only extract radiomic features from the PET images, we are especially interested in their ability to fuse the clinical data with both radiomic and learning-based imaging features extracted from a variable number of lesions, while explicitly removing feature redundancy between them with their conditional feature-guided diffusion approach.

All those works seem very promising and we would be very interested in applying them to our GAINED dataset to compare their performance with ours.

Supplementary material

Appendix A

PET imaging for DLBCL

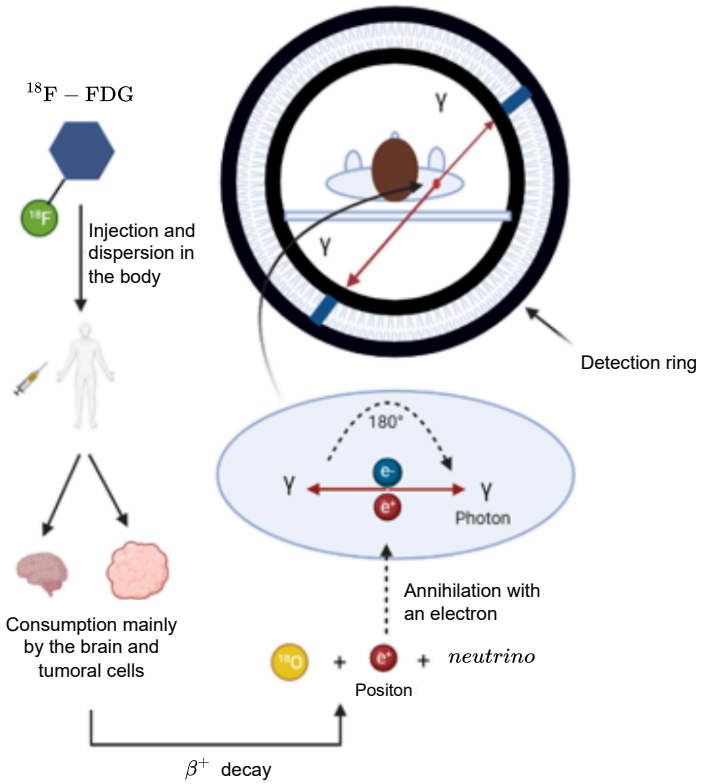
For the DLBCL pathology, lesion detection and diagnosis can be made using ^{18}F-FDG PET images. Unlike Magnetic Resonance Imaging (MRI), FDG-PET has the advantage of being quantitative. Indeed, FDG uptake is characterized by an index, the SUV (Standardized Uptake Value), which takes into account injected activity and patient mass.

The ^{18}F-FDG PET method summarized in Fig. A.1a uses a combination of a carrier, glucose, and an emitter, fluorine-18. Glucose is abundantly consumed by cancerous lesions, but also by the brain, and is eliminated in the bladder. The vector/emitter pair (or radiopharmaceutical drug) will therefore be directed to these areas. In 97% of cases, fluorine-18 disintegrates into oxygen-18 by β^+ decay, forming positrons and neutrinos ν:

$$^{18}_{9}F \rightarrow ^{18}_{8}O + ^{0}_{1}e + \nu \tag{A.1}$$

Once the energy of the positrons has been lost (they travel an average of 0.6 mm), they will annihilate with electrons, producing two γ photons emitted at 180°. These γ photons will be detected by opposite detectors and thus give a response line. By then using adapted tomographic reconstruction algorithms, we obtain the ^{18}F-FDG fixation position. The two photons are validated if their energy is close to 511 keV and if they are detected within a very short time interval (time window of around 4.1 ns). By acquiring data at different angles, a tomographic plane can be reconstructed, and from these slices, the distribution of the radiopharmaceutical drug in the patient's body can be obtained in a three-dimensional space.

Fluorine-18 has the advantage of a reasonably long half-life (110 minutes). Furthermore, its positrons have a relatively short maximum path (2.6 mm in water, compared with 4.1 mm for C-11 and 8.2 mm for oxygen-15, for example), giving a radioactive fixation map closer to reality. A ^{18}F-FDG PET image of a DLBCL patient is shown in Fig. A.1b. These PET images can be used for disease diagnosis and monitoring, as well as for research purposes.



(a) Explanation of the PET process (from [Morvan 2021]).



(b) PET image example: in green are the brain and bladder; in orange are lesions.

Figure A.1 – PET imaging: explanation and example.

Appendix B

GAINED features

This section presents the patients' features in the GAINED dataset: Table B.1 explains the clinical features used for each patient, and is followed by a small introduction on radiomics and the selection of imaging features we considered.

B.1 Clinical features

Table B.1 – Clinical features considered in the GAINED dataset.

Name	Explanation
Age	The higher it is, the higher the risk for the patient.
Lactate dehydrogenase (LDH)	An enzyme present in cells used as a biomarker for DLBCL diagnosis and prognosis.
Categorical LDH	Whether the LDH level is higher than it should be, the expected value depending on the measurement process.
Ann Arbor stage	Ranges from 1 to 4 depending on where the malignant tissues are located.
Number of extranodal sites	Having more than one extranodal site of involvement is usually an indication of a poor prognosis.
ECOG scale	Ranges from 0 to 5 and describes a patient's ability to take care of themselves and perform daily activities.
Age-adjusted IPI (aaIPI)	Classifies the DLBCL patients into four risk groups (0 to 3), based on 3 factors: (a) performance status - a score determining the patient's ability to perform certain day-to-day activities, (b) stage and (c) LDH level.
Treatment type	Course of treatment administered to the patient: chemotherapy regimen, autologous cell transplant or salvage therapy (which are considered as three different binary variables).

B.2 Radiomics

In the medical imaging field, radiomics is defined as the high-throughput extraction of quantitative features from images to enable characterization of tissues [Gillies et al. 2016]. These features can be classified into different categories, each potentially related to different aspects of a tissue phenotype. These categories include morphological features, which describe the shape and size of analyzed regions; intensity features, which capture characteristics related to pixel intensity distributions; and texture features, which quantify spatial intensity patterns. These features can be extracted from the original image intensities, but also from image intensities previously modified by filters that highlight various structures and patterns [Whybra et al. 2024]. Different image pre-processing steps such as interpolation and intensity range definition may also be performed prior to feature extraction. Recently, the Image Biomarker Standardization Initiative (IBSI) defined a standardized workflow for radiomic feature extraction [Zwanenburg et al. 2020], but the selection of a relevant set of radiomic features is task-dependant and highly complex [Loutfi et al. 2024]: with multiple options for image pre-processing steps, adjustable parameters, and if filtering is considered prior to feature extraction, the number of possible features could rise to ∼10,000 and more.

B.3 GAINED radiomic features

We present here the three sets 'HeadNeck', 'FirstOrder' and 'All' of imaging features considered in our experiments. The first two are presented in Tables B.2 and B.3:

Table B.2 – Lesions 'HeadNeck' imaging features.

Conventional PET features	Radiomics	
• Standard uptake value of the maximum intensity voxel inside the lesion (SUVmax) • Metabolic Tumor Volume (MTV) of the lesion • Total lesion glycolysis (TLG) from each of the segmented lesions	Shape First order Second order	• Sphericity • Mean intensity • Standard deviation of intensity • Entropy of intensity • Inverse difference normalized • Joint energy • Contrast • Correlation

The 'HeadNeck' second-order radiomic features are computed on images pre-processed as described in Part II, Sec. 5.3.1 and then normalized using two approaches: a linear equalization with 64 bins and a fixed SUV bin-width of 0.3. As such, they are computed two times, once for each normalization strategy. As it would not make sense to estimate

the other features' values after normalization, they are computed only once before this step to create the 'HeadNeck' set of features of size $D_{img} = 15$.

Table B.3 – Lesions 'FirstOrder' imaging features.

Conventional PET features	Radiomics	
• Standard uptake value of the maximum intensity voxel inside the lesion (SUVmax) • Metabolic Tumor Volume (MTV) of the lesion • Total lesion glycolysis (TLG) from each of the segmented lesions	Shape	• Sphericity • Volume • Surface area • Surface volume ratio • Flatness • Elongation • Maximum 3D diameter • Major axis length • Minor axis length
	First order	• Maximum intensity • Energy

In the same way, the first-order features from the 'FirstOrder' set are computed both on pre-processed images as described in section 5.3.1, and on such images with additional pre-processing. The latter is performed to improve the image characteristics, and includes a wavelet transform using coiflet-1 filters to decompose the original image into 8 frequencies and an edge enhancement Laplacian of Gaussian filter (with 2- and 6-mm σ). These features are thus computed 11 times (8 times for the different frequencies of the wavelet transform, twice for the 2- and 6-mm σ Gaussian filter, and once with no additional pre-processing). Added to the conventional and shape features, this makes a set of $D_{img} = 34$ features [1].

As for the set of 'All' features, for the sake of clarity, its exhaustive list of features is not developed in this document, but it can be found in Supplementary material 2 of our article [Thiery, Rizkallah, Bailly, et al. 2025] at the following link: `https://www.sciencedirect.com/science/article/pii/S0895611124001587?via%3Dihub`.

1. We realized late a mistake in our feature extraction process which led to these features being extracted and the pre-processing steps being applied on first-order features, despite it making no sense. As such, the experiments relying on the 'FirstOrder' set of features could perform even better after correcting these two problems.

Appendix C

Hyperparameter sets searched

Table C.1 – Grid search study (HC = hidden channels ; LR = learning rate).

Model	Search space
MLPs	HC: {16, 32, 64, 128} LR: {0.1, 0.01, 0.001, 0.0001}
XGboost	number of estimators: {5, 10, 15, 20, 25, 30, 40, 50} max depth: {1, 2, 3} subsampling: {0.7, 0.85, 1} LR: {1, 0.1, 0.01}
MIL	HC: {16, 32, 64, 128} LR: {0.1, 0.01, 0.001, 0.0001}
GraphConv	HC: {16, 32, 64, 128} LR: {0.01, 0.001, 0.0001, 1e-05} γ: {0.1, 1, 10, 50}
GAT	HC: {16, 32, 64, 128} LR: {0.01, 0.001, 0.0001, 1e-05} γ: {0.1, 1, 10, 50}
Cross-attention	D_{GAT}: {32, 64, 128} Dropout GAT: {0, 0.3, 0.6} Dropout cross-attention module: {0, 0.3, 0.6} LR: {0.01, 0.001, 0.0001} γ: {0.1, 1, 10, 100}
Cross-attention 2	D_{GAT}: {32, 64, 128} Dropout GAT: {0, 0.3, 0.6} Dropout cross-attention module: {0, 0.3, 0.6} LR: {0.01, 0.001, 0.0001} γ: {0.1, 1, 10, 100}
Our multimodal model (structural study)	D_{GAT}: {32, 64} D_{proj}: {16, 32, 64, 128} Dropout GAT: {0, 0.3, 0.6} Dropout cross-attention module: {0, 0.3, 0.6} LR: {0.01, 0.001, 0.0001} γ: {0.1, 1, 10}

Appendix D

Multimodal model: additional information

In this section is presented additional information about our multimodal model from Part III: first, an explanation of its two-layer version in our structure analysis, and second, equations allowing for the computation of its number of parameters depending on the multiple parameters explored.

D.1 Model structure when the second multimodal block is used

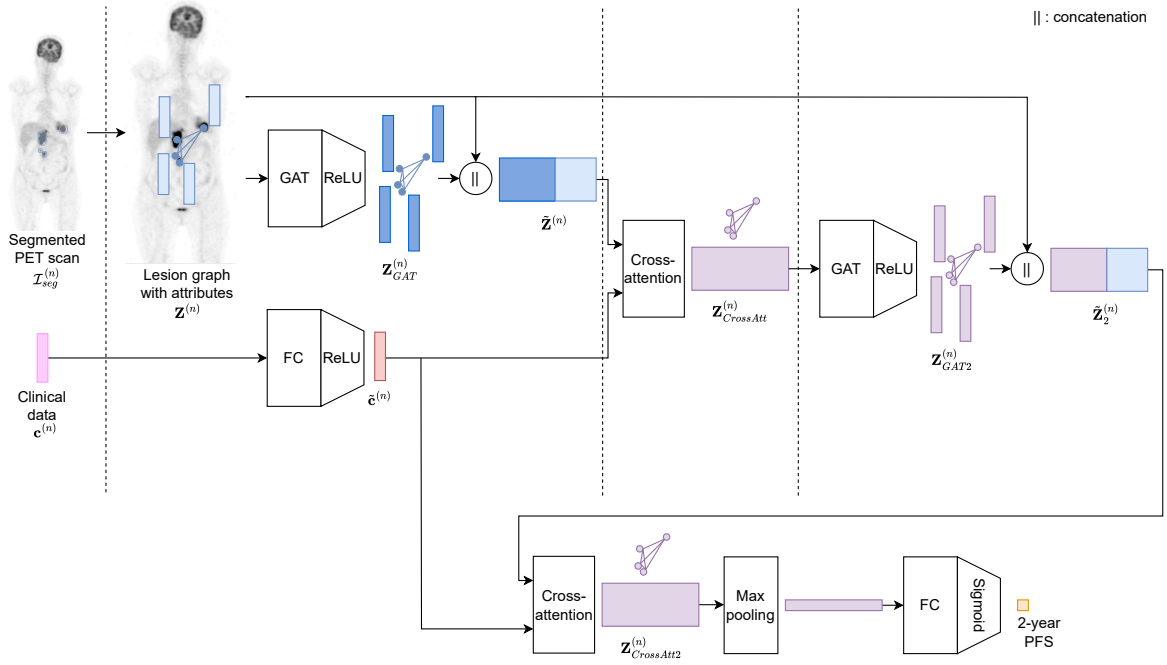


Figure D.1 – Overview of the proposed fusion framework with two multimodal blocks.

D.2 Computation of the number of parameters of a given configuration of the model

Our multimodal model's number of parameters depends on many structural choices made, including: whether the clinical data is projected before the fusion, the choice of GNN to propagate the lesions' information, the cross-attention configuration, the use of a skip connection on the imaging data or not, and the number of layers. The following equations show the complexity of computing its number of parameters in such a case.

$$n_{FC_clin} = \begin{cases} 0 & \text{if no clinical FC layer} \\ D_{clin} \times D_{proj} + D_{proj} & \text{else.} \end{cases} \tag{D.1}$$

$$D_{out_clin} = \begin{cases} D_{clin} & \text{if no clinical FC layer} \\ D_{proj} & \text{else.} \end{cases} \tag{D.2}$$

$$n_{GAT} = \begin{cases} 2 \times (5 \times D_{GAT} + 2 \times (D_{GAT} \times D_{img})) & \text{if GAT} \\ 2 \times (D_{GAT} \times D_{img}) + D_{GAT} & \text{if GraphConv} \\ 4 \times D_{img}^2 + 4 \times D_{img} & \text{if Self-attention} \end{cases} \tag{D.3}$$

$$D_{out_GAT} = \begin{cases} 2 \times D_{GAT} & \text{if GAT} \\ D_{GAT} & \text{if GraphConv} \\ D_{img} & \text{if self-attention} \end{cases} \tag{D.4}$$

$$D_{out_GAT-skip_co} = \begin{cases} D_{out_GAT} & \text{if no skip connection} \\ D_{out_GAT} + D_{img} & \text{else.} \end{cases} \tag{D.5}$$

$$D_{concat} = D_{out_GAT-skip_co} + D_{out_clin} \tag{D.6}$$

$$n_{cross-att} = \begin{cases} 6 \times D_{out_GAT-skip_co} + 2 \times D_{out_GAT-skip_co}^2 & \text{if Img2Clin} \\ 6 \times D_{concat} + 2 \times D_{concat}^2 & \text{if Mult2Clin} \\ 4 \times D_{concat} + 4 \times D_{concat}^2 & \text{if Mult2Mult} \\ 4 \times D_{out_clin} + 2 \times D_{out_clin} \times D_{out_GAT-skip_co} + 2 \times D_{out_clin}^2 & \text{if Clin2Img} \\ 4 \times D_{concat} + 2 \times D_{concat} \times D_{out_GAT-skip_co} + 2 \times D_{concat}^2 & \text{if Mult2Img} \end{cases} \tag{D.7}$$

$$D_{out_catt} = \begin{cases} D_{out_GAT-skip_co} & \text{if Img2Clin} \\ D_{concat} & \text{if Mult2Clin} \\ D_{concat} & \text{if Mult2Mult} \\ D_{out_clin} & \text{if Clin2Img} \\ D_{concat} & \text{if Mult2Img} \end{cases} \tag{D.8}$$

$$n_{GAT_2} = \begin{cases} 2 \times (5 \times D_{GAT} + 2 \times (D_{GAT} \times D_{out_catt})) & \text{if GAT} \\ 2 \times (D_{GAT} \times D_{out_catt}) + D_{GAT} & \text{if GraphConv} \\ 4 \times D_{out_catt}^2 + 4 \times D_{out_catt} & \text{if Self-attention} \end{cases} \tag{D.9}$$

$$D_{out_GAT_2} = \begin{cases} 2 \times D_{GAT} & \text{if GAT} \\ D_{GAT} & \text{if GraphConv} \\ D_{out_catt} & \text{if self-attention} \end{cases} \tag{D.10}$$

$$D_{out_GAT-skip_co_2} = \begin{cases} D_{out_GAT_2} & \text{if no skip connection} \\ D_{out_GAT_2} + D_{img} & \text{else.} \end{cases} \tag{D.11}$$

$$D_{concat_2} = D_{out_GAT-skip_co_2} + D_{out_clin} \tag{D.12}$$

$$n_{cross-att_2} = \begin{cases} 6 \times D_{out_GAT-skip_co_2} + 2 \times D_{out_GAT-skip_co_2}^2 & \text{if Img2Clin} \\ 6 \times D_{concat_2} + 2 \times D_{concat_2}^2 & \text{if Mult2Clin} \\ 4 \times D_{concat_2} + 4 \times D_{concat_2}^2 & \text{if Mult2Mult} \\ 4 \times D_{out_clin} + 2 \times D_{out_clin} \times D_{out_GAT-skip_co_2} + 2 \times D_{out_clin}^2 & \text{if Clin2Img} \\ 4 \times D_{concat_2} + 2 \times D_{concat_2} \times D_{out_GAT-skip_co_2} + 2 \times D_{concat_2}^2 & \text{if Mult2Img} \end{cases} \tag{D.13}$$

$$D_{out_catt_2} = \begin{cases} D_{out_GAT-skip_co_2} & \text{if Img2Clin} \\ D_{concat_2} & \text{if Mult2Clin} \\ D_{concat_2} & \text{if Mult2Mult} \\ D_{out_clin} & \text{if Clin2Img} \\ D_{concat_2} & \text{if Mult2Img} \end{cases} \tag{D.14}$$

$$n_{FC_prediction} = \begin{cases} D_{out_catt} + 1 & \text{if 1 layer} \\ D_{out_catt_2} + 1 & \text{if 2 layers} \end{cases} \tag{D.15}$$

$$n_{total} = \begin{cases} n_{FC_clin} + n_{GAT} + n_{cross-att} + n_{FC_prediction} & \text{if 1 layer} \\ n_{FC_clin} + n_{GAT} + n_{cross-att} + n_{GAT_2} + n_{cross-att_2} + n_{FC_prediction} & \text{if 2 layers} \end{cases} \tag{D.16}$$

Bibliography

Ahn, J., A. M. Elmahdy, S. Mohajer, and C. Suh (2021), « On the Fundamental Limits of Matrix Completion: Leveraging Hierarchical Similarity Graphs », *in*: *IEEE Transactions on Information Theory* 70, pp. 2039–2075.

Andrade-Miranda, G., V. Jaouen, O. Tankyevych, C. Cheze Le Rest, D. Visvikis, and P.-H. Conze (Dec. 2023), « Multi-modal medical Transformers: A meta-analysis for medical image segmentation in oncology », *in*: *Computerized Medical Imaging and Graphics* 110, p. 102308, ISSN: 0895-6111, DOI: 10.1016/j.compmedimag.2023.102308.

Antolini, L., P. Boracchi, and E. Biganzoli (2005), « A time-dependent discrimination index for survival data », *in*: *Statistics in Medicine* 24.*24*, pp. 3927–3944, ISSN: 1097-0258, DOI: 10.1002/sim.2427.

Aswathi, A., M. Rizkallah, G. Frecon, C. Bailly, C. M. Bodet-Milin, O. Casasnovas, S. Le Gouill, F. Kraeber-Bodéré, T. Carlier, and D. Mateus (Apr. 2023), « Lesion graph neural networks for 2-year progression free survival classification of Diffuse Large B-Cell Lymphoma patients », *in*: *International Symposium on Biomedical Imaging*.

Austin, P. C., F. E. Harrell, and D. van Klaveren (2020), « Graphical calibration curves and the integrated calibration index (ICI) for survival models », *in*: *Statistics in Medicine* 39, pp. 2714–2742.

Ba, J. L., J. R. Kiros, and G. E. Hinton (2016), *Layer Normalization*, arXiv: 1607.06450 [stat.ML].

Baltrušaitis, T., C. Ahuja, and L.-P. Morency (Feb. 2019), « Multimodal Machine Learning: A Survey and Taxonomy », *in*: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41.*2*, pp. 423–443, ISSN: 1939-3539.

Bintsi, K.-M., V. Baltatzis, R. A. Potamias, A. Hammers, and D. Rueckert (2023), « Multimodal Brain Age Estimation Using Interpretable Adaptive Population-Graph Learning », *in*: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, pp. 195–204, ISBN: 978-3-031-43993-3, DOI: 10.1007/978-3-031-43993-3_19.

Bo, D., X. Wang, C. Shi, and H. Shen (May 2021), « Beyond Low-frequency Information in Graph Convolutional Networks », *in*: *Proceedings of the AAAI Conference on*

Artificial Intelligence, vol. 35, 5, journalAbbreviation: AAAI, pp. 3950–3957, DOI: 10.1609/aaai.v35i5.16514.

Boellaard, R., I. Buvat, C. Nioche, L. Ceriani, A.-S. Cottereau, L. Guerra, R. J. Hicks, S. Kanoun, C. Kobe, A. Loft, H. Schöder, A. Versari, C.-A. Voltin, G. J. C. Zwezerijnen, J. M. Zijlstra, N. G. Mikhaeel, A. Gallamini, T. C. El-Galaly, C. Hanoun, S. Chauvie, R. Ricci, E. Zucca, M. Meignan, and S. F. Barrington (2024), « International Benchmark for Total Metabolic Tumor Volume Measurement in Baseline 18F-FDG PET/CT of Lymphoma Patients: A Milestone Toward Clinical Implementation », *in*: *Journal of Nuclear Medicine* 65, pp. 1343–1348.

Brody, S., U. Alon, and E. Yahav (2022), « How Attentive are Graph Attention Networks? », *in*: *International Conference on Learning Representations*.

Butoi, V. I., J. J. G. Ortiz, T. Ma, M. R. Sabuncu, J. Guttag, and A. V. Dalca (2023), « UniverSeg: Universal Medical Image Segmentation », *in*: *International Conference on Computer Vision*.

Carlier, T., G. Frécon, D. Mateus, M. Rizkallah, F. Kraeber-Bodéré, S. Kanoun, P. Blanc-Durand, E. Itti, S. Le Gouill, R.-O. Casasnovas, C. Bodet-Milin, and C. Bailly (Jan. 2024), « Prognostic Value of 18 F-FDG PET Radiomics Features at Baseline in PET-Guided Consolidation Strategy in Diffuse Large B-Cell Lymphoma: A Machine-Learning Analysis from the GAINED Study », *in*: *Journal of Nuclear Medicine* 65.*1*, pp. 156–162, ISSN: 0161-5505, 2159-662X, DOI: 10.2967/jnumed.123.265872.

Chen, L., Y. Huang, B. Liao, K. Nie, S. Dong, and J. Hu (Dec. 2020), « Graph Learning Approaches for Graph with Noise: Application to Disease Prediction in Population Graph », *in*: *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 2724–2729, DOI: 10.1109/BIBM49941.2020.9313348.

Chen, T., R. Hong, Y. Guo, S. Hao, and B. Hu (2022), « MS^2-GNN: Exploring GNN-Based Multimodal Fusion Network for Depression Detection », *in*: *IEEE Transactions on Cybernetics*, pp. 1–11, ISSN: 2168-2275, DOI: 10.1109/TCYB.2022.3197127.

Chen, T. and C. Guestrin (2016), « XGBoost: A Scalable Tree Boosting System », *in*: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Cottereau, A.-S., C. Nioche, A.-S. Dirand, J. Clerc, F. Morschhauser, O. Casasnovas, M. Meignan, and I. Buvat (2019), « 18F-FDG PET Dissemination Features in Diffuse Large B-Cell Lymphoma Are Predictive of Outcome », *in*: *The Journal of Nuclear Medicine* 61, pp. 40–45.

Curtis, C., S. P. Shah, S.-F. Chin, G. Turashvili, O. M. Rueda, M. J. Dunning, D. Speed, A. G. Lynch, S. A. Samarajiwa, Y. Yuan, S. Gräf, G. Ha, G. Haffari, A. Bashashati, R. Russell, S. McKinney, A. Langerød, A. R. Green, E. Provenzano,

G. C. Wishart, S. E. Pinder, P. H. Watson, F. Markowetz, L. C. Murphy, I. O. Ellis, A. Purushotham, A.-L. Børresen-Dale, J. D. Brenton, S. Tavaré, C. Caldas, and S. Aparicio (2012), « The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups », *in*: *Nature* 486, pp. 346–352.

Eertink, J. J., G. J. C. Zwezerijnen, S. E. Wiegers, S. Pieplenbosch, M. E. D. Chamuleau, P. J. Lugtenburg, D. de Jong, B. Ylstra, M. Mendeville, U. Dührsen, C. Hanoun, A. Hüttmann, J. Richter, W. Klapper, Y. W. S. Jauw, O. S. Hoekstra, H. C. W. de Vet, R. Boellaard, and J. M. Zijlstra (2022), « Baseline radiomics features and MYC rearrangement status predict progression in aggressive B-cell lymphoma », *in*: *Blood Advances* 7, pp. 214–223.

Ferrández, M. C., S. S. V. Golla, J. J. Eertink, B. M. de Vries, P. J. Lugtenburg, S. E. Wiegers, G. J. C. Zwezerijnen, S. Pieplenbosch, L. Kurch, A. Hüttmann, C. Hanoun, U. Dührsen, H. C. W. de Vet, J. M. Zijlstra, and R. Boellaard (Aug. 2023), « An artificial intelligence method using FDG PET to predict treatment outcome in diffuse large B cell lymphoma patients », *in*: *Scientific Reports* 13.*1*, p. 13111, ISSN: 2045-2322, DOI: 10.1038/s41598-023-40218-1.

Fey, M. and J. E. Lenssen (2019), « Fast Graph Representation Learning with PyTorch Geometric », *in*: *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

Fleming, T. R. and D. P. H. and (1981), « A class of hypothesis tests for one and two sample censored survival data », *in*: *Communications in Statistics - Theory and Methods* 10.*8*, pp. 763–794, DOI: 10.1080/03610928108828073.

Fotso, S. (Jan. 2018), « Deep Neural Networks for Survival Analysis Based on a Multi-Task Framework », *in*: *arXiv:1801.05512*.

Fukushima, K. and S. Miyake (1982), « Neocognitron: a self-organizing neural network model for a mechanism of visual pattern recognition », *in*: *Competition and cooperation in neural nets*, Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 267–285, ISBN: 978-3-642-46466-9.

Galdran, A., J. W. Verjans, G. Carneiro, and M. A. González Ballester (2023), « Multi-head multi-loss model calibration », *in*: *Medical image computing and computer assisted intervention – MICCAI 2023*, Cham: Springer Nature Switzerland, pp. 108–117, ISBN: 978-3-031-43898-1.

Gao, X., F. Shi, D. Shen, and M. Liu (2023), « Multimodal transformer network for incomplete image generation and diagnosis of Alzheimer's disease », *in*: *Computerized Medical Imaging and Graphics* 110, p. 102303, ISSN: 0895-6111, DOI: https://doi.org/10.1016/j.compmedimag.2023.102303.

Geck, M. (Aug. 2012), « PyCox: computing with (finite) Coxeter groups and Iwahori–Hecke algebras », *in*: *LMS Journal of Computation and Mathematics* 15, pp. 231–256, ISSN: 1461-1570, DOI: 10.1112/s1461157012001064.

Ghorbani, M., M. Bahrami, A. Kazi, M. Soleymani Baghshah, H. R. Rabiee, and N. Navab (2021), « GKD: Semi-supervised Graph Knowledge Distillation for Graph-Independent Inference », *in*: ed. by M. De Bruijne, P. C. Cattin, S. Cotin, N. Padoy, S. Speidel, Y. Zheng, and C. Essert, vol. 12905, Lecture Notes in Computer Science, Book Title: Medical Image Computing and Computer Assisted Intervention – MICCAI 2021, Cham: Springer International Publishing, pp. 709–718, ISBN: 978-3-030-87239-7, DOI: 10.1007/978-3-030-87240-3_68.

Gillies, R. J., P. E. Kinahan, and H. Hricak (Feb. 2016), « Radiomics: Images Are More than Pictures, They Are Data », *in*: *Radiology* 278.*2*, pp. 563–577, ISSN: 0033-8419, DOI: 10.1148/radiol.2015151169.

Golovanevsky, M., C. Eickhoff, and R. Singh (2022), « Multimodal Attention-based Deep Learning for Alzheimer's Disease Diagnosis », *in*: *Journal of the American Medical Informatics Association*.

Graf, E., C. Schmoor, W. Sauerbrei, and M. Schumacher (Sept. 1999), « Assessment and comparison of prognostic classification schemes for survival data », *in*: *Statistics in Medicine* 18.*17–18*, pp. 2529–2545, ISSN: 0277-6715, DOI: 10.1002/(sici)1097-0258(19990915/30)18:17/18<2529::aid-sim274>3.0.co;2-5.

Griethuysen, J. J. M. van, A. Fedorov, C. Parmar, A. Hosny, N. Aucoin, V. Narayan, R. Beets-Tan, J.-C. Fillion-Robin, S. D. Pieper, and H. J. Aerts (2017), « Computational Radiomics System to Decode the Radiographic Phenotype. », *in*: *Cancer research* 77 (21), e104–e107.

Hager, P., M. J. Menten, and D. Rueckert (2023), « Best of Both Worlds: Multimodal Contrastive Learning with Tabular and Imaging Data », *in*: *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23924–23935.

Harrell, F. E., R. M. Califf, D. B. Pryor, K. L. Lee, and R. A. Rosati (May 1982), « Evaluating the yield of medical tests », *in*: *Journal of the American Medical Association* 247.*18*, pp. 2543–2546, ISSN: 0098-7484.

Huang, Q., H. He, A. Singh, S.-N. Lim, and A. R. Benson (Oct. 2020), « Combining Label Propagation and Simple Models Out-performs Graph Neural Networks », *in*: *International Conference on Learning Representations*.

Huang, W. (2023), « Multimodal Contrastive Learning and Tabular Attention for Automated Alzheimer's Disease Prediction », *in*: *IEEE/CVF International Conference on Computer Vision Workshops*, pp. 2465–2474, DOI: 10.1109/ICCVW60793.2023.00261.

Ishwaran, H., U. Kogalur, E. Blackstone, and M. Lauer (Dec. 2008), « Random Survival Forests », *in*: *The Annals of Applied Statistics* 2, DOI: 10.1214/08-AOAS169.

Jiang, C., A. Li, Y. Teng, X. Huang, C. Ding, J. Chen, J. Xu, and Z. Zhou (July 2022), « Optimal PET-based radiomic signature construction based on the cross-combination method for predicting the survival of patients with diffuse large B-cell lymphoma », *in*: *European Journal of Nuclear Medicine and Molecular Imaging* 49.*8*, pp. 2902–2916.

Jing, B., T. Zhang, Z. Wang, Y. Jin, K. Liu, W. Qiu, L. Ke, Y. Sun, C. He, D. Hou, L. Tang, X. Lv, and C. Li (July 2019), « A deep survival analysis method based on ranking », *in*: *Artificial Intelligence in Medicine* 98, pp. 1–9, ISSN: 0933-3657, DOI: 10.1016/j.artmed.2019.06.001.

Kaplan, E. L. and P. Meier (1958), « Nonparametric Estimation from Incomplete Observations », *in*: *Journal of the American Statistical Association* 53.*282*, pp. 457–481, ISSN: 01621459, 1537274X.

Kassam, J., F. Thamm, L. Rist, O. Taubmann, and A. K. Maier (2022), « Detecting Large Vessel Occlusions using Graph Deep Learning », *in*: *GEOmedia*.

Katzman, J. L., U. Shaham, A. Cloninger, J. Bates, T. Jiang, and Y. Kluger (Feb. 2018), « DeepSurv: personalized treatment recommender system using a Cox proportional hazards deep neural network », *in*: *BMC Medical Research Methodology* 18.*1*, p. 24, ISSN: 1471-2288, DOI: 10.1186/s12874-018-0482-1.

Kazmierski, M. and B. Haibe-Kains (June 2021), « Lymph Node Graph Neural Networks for Cancer Metastasis Prediction », *in*: *arXiv:2106.01711*.

Khan, K. S., A. Kanamarlapudi, and A. R. Singh (2024), « Diagnosis and Classification of Breast Cancer Using Data Visualization and Deep Learning Techniques », *in*: *2024 5th International Conference for Emerging Technology (INCET)*, pp. 1–6.

Kim, S. Y. (June 2023), « Personalized Explanations for Early Diagnosis of Alzheimer's Disease Using Explainable Graph Neural Networks with Population Graphs », *in*: *Bioengineering* 10.*6*, p. 701, ISSN: 2306-5354, DOI: 10.3390/bioengineering10060701.

Kingma, D. P. and J. Ba (2014), « Adam: A Method for Stochastic Optimization », *in*: *CoRR* abs/1412.6980.

Kirillov, A., E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. B. Girshick (2023), « Segment Anything », *in*: *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3992–4003.

Klein, J. P. and M. L. Moeschberger (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, Springer New York, NY, DOI: 10.1007/b97377.

Kraeber-Bodéré, F., S. Zweegman, A. Perrot, C. Hulin, D. Caillot, T. Facon, X. Leleu, K. Belhadj, E. Itti, L. Karlin, C. Bailly, M.-D. Levin, M. C. Minnema, B. Jamet,

C. Bodet-Milin, B. de Keizer, M. C. Béné, H. Avet-Loiseau, P. Sonneveld, L. Pei, F. Rigat, C. J. de Boer, J. Vermeulen, T. Kampfenkel, J. Lambert, and P. Moreau (2022), « Prognostic value of positron emission tomography/computed tomography in transplant-eligible newly diagnosed multiple myeloma patients from CASSIOPEIA: the CASSIOPET study », *in*: *Haematologica* 108, pp. 621–626.

Kvamme, H. and Ø. Borgan (Oct. 2021), « Continuous and discrete-time survival prediction with neural networks », *in*: *Lifetime Data Analysis* 27.*4*, pp. 710–736, ISSN: 1572-9249, DOI: 10.1007/s10985-021-09532-6.

Kvamme, H., Ø. Borgan, and I. Scheel (2019), « Time-to-Event Prediction with Neural Networks and Cox Regression », *in*: *Journal of Machine Learning Research* 20.*129*, pp. 1–30, ISSN: 1533-7928.

Larose, M., L. Archambault, N. Touma, R. Brodeur, F. Desroches, N. Raymond, D. Bédard-Tremblay, D. LeBlanc, F. Rasekh, H. Hovington, B. Neveu, M. Vallières, and F. Pouliot (Nov. 2024), « Multi-task Bayesian model combining FDG-PET/CT imaging and clinical data for interpretable high-grade prostate cancer prognosis », *in*: *Scientific Reports* 14.*1*, p. 26928, ISSN: 2045-2322, DOI: 10.1038/s41598-024-77498-0.

Larose, M., N. Touma, N. Raymond, D. LeBlanc, F. Rasekh, B. Neveu, H. Hovington, M. Vallières, F. Pouliot, and L. Archambault (2022), « Graph Attention Network for Prostate Cancer Lymph Node Invasion Prediction », *in*: *Medical Imaging with Deep Learning*.

Le Gouill, S., H. Ghesquières, L. Oberic, F. Morschhauser, H. Tilly, V. Ribrag, T. Lamy, C. Thieblemont, H. Maisonneuve, R. Gressin, K. Bouhabdallah, C. Haioun, G. Damaj, L. Fornecker, R. Bouhabdallah, P. Feugier, D. Sibon, G. Cartron, C. Bonnet, M. Andre, L. Chartier, P. Ruminy, F. Kraeber-Bodéré, C. Bodet-Milin, A. Berriolo-Riedinger, J. Brière, J. P. Jais, T. J. Molina, E. Itti, and R. O. Casasnovas (2021), « Obinutuzumab vs rituximab for advanced DLBCL: a PET-guided and randomized phase 3 study by LYSA », *in*: *Blood* 137.*17*, pp. 2307–2320.

Lee, C., W. Zame, J. Yoon, and M. Van der Schaar (Apr. 2018), « DeepHit: A Deep Learning Approach to Survival Analysis With Competing Risks », *in*: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.*1*, ISSN: 2374-3468, 2159-5399, DOI: 10.1609/aaai.v32i1.11842.

Liu, P., M. Zhang, X. Gao, B. Li, and G. Zheng (2022), « Joint Lymphoma Lesion Segmentation and Prognosis Prediction From Baseline FDG-PET Images via Multitask Convolutional Neural Networks », *in*: *IEEE Access* 10, pp. 81612–81623, ISSN: 2169-3536, DOI: 10.1109/ACCESS.2022.3195906.

Liu, S., B. Zhang, R. Fang, D. Rueckert, and V. A. Zimmer (2023), « Dynamic Graph Neural Representation Based Multi-modal Fusion Model for Cognitive Outcome Pre-

diction in Stroke Cases », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer*, Lecture Notes in Computer Science, pp. 338–347, ISBN: 978-3-031-43993-3, DOI: 10.1007/978-3-031-43993-3_33.

Loutfi, M. A. L., T. B. Podasca, A. Zwanenburg, T. Upadhaya, J. Barrios, D. R. Raleigh, W. C. Chen, D. P. I. Capaldi, H. Zheng, O. Gevaert, J. Wu, A. C. Silva, P. J. Zhang, H. X. Bai, J. Seuntjens, S. Löck, P. O. Richard, O. Morin, C. Reinhold, M. Lepage, and M. Vallières (2024), « Unraveling Radiomics Complexity: Strategies for Optimal Simplicity in Predictive Modeling », *in*: *ArXiv* abs/2407.04888.

Lv, W., Z. Zhou, J. Peng, L. Peng, G. Lin, H. Wu, H. Xu, and L. Lu (Mar. 2023), « Functional-structural sub-region graph convolutional network (FSGCN): Application to the prognosis of head and neck cancer with PET/CT imaging », *in*: *Computer Methods and Programs in Biomedicine* 230, p. 107341, ISSN: 0169-2607.

Malighetti, F., M. Villa, A. M. Villa, S. Pelucchi, A. Aroldi, D. L. Cortinovis, S. Canova, S. Capici, M. E. Cazzaniga, L. Mologni, D. Ramazzotti, and N. Cordani (2025), « Prognostic Biomarkers in Breast Cancer via Multi-Omics Clustering Analysis », *in*: *International Journal of Molecular Sciences* 26.

Malmon, S., M.-H. Elsensohn, C. Thieblemont, F. Morschhauser, O. Casasnovas, M. André, S. Le Gouill, Y. A. Tabaa, P. B. Durand, C. Bailly, V. Edeline, L. Vija, L. Vercellino, R. Ricci, S. Kanoun, and A.-S. Cottereau (2025), « Prognostic impact of metabolic tumor volume using the SUV4.0 segmentation threshold in 1,960 lymphoma patients from prospective LYSA trials. », *in*: *European journal of nuclear medicine and molecular imaging*.

Mcnemar, Q. (1947), « Note on the sampling error of the difference between correlated proportions or percentages », *in*: *Psychometrika* 12, pp. 153–157.

Morris, C., M. Ritzert, M. Fey, W. Hamilton, J. Lenssen, G. Rattan, and M. Grohe (July 2019), « Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks », *in*: *AAAI Conference on Artificial Intelligence* 33, pp. 4602–4609.

Morvan, L. (Dec. 2021), « Prédiction de la progression du myélome multiple par imagerie TEP : Adaptation des forêts de survie aléatoires et de réseaux de neurones convolutionnels », fr, PhD thesis, École centrale de Nantes, URL: https://theses.hal.science/tel-03678933.

Morvan, L., C. Nanni, A.-V. Michaud, B. Jamet, C. Bailly, C. Bodet-Milin, S. Chauvie, C. Touzeau, P. Moreau, E. Zamagni, F. Kraeber-Bodéré, T. Carlier, and D. Mateus (2023), « Multiple Myeloma Prognosis From PET Images: Deep Survival Losses and Contrastive Pretraining », *in*: *IEEE Transactions on Radiation and Plasma Medical Sciences* 7.7, pp. 732–741, DOI: 10.1109/TRPMS.2023.3283562.

Mucaki, E. J., K. Baranova, H. Q. Pham, I. Rezaeian, D. Angelov, A. Ngom, L. Rueda, and P. K. Rogan (2017), « Predicting Outcomes of Hormone and Chemotherapy in

the Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) Study by Biochemically-inspired Machine Learning », *in*: *F1000Research* 5.

Mueller, T. T., S. Starck, A. Dima, S. Wunderlich, K.-M. Bintsi, K. Zaripova, R. F. Braren, D. Rueckert, A. Kazi, and G. Kaissis (2024), « A Survey on Graph Construction for Geometric Deep Learning in Medicine: Methods and Recommendations », *in*: *Trans. Mach. Learn. Res.* 2024.

Nagrani, A., S. Yang, A. Arnab, A. Jansen, C. Schmid, and C. Sun (2021), « Attention Bottlenecks for Multimodal Fusion », *in*: *Neural Information Processing Systems*.

Parisot, S., S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker, and D. Rueckert (June 2018), « Disease prediction using graph convolutional networks: application to Autism Spectrum Disorder and Alzheimer's disease », *in*: *Medical Image Analysis* 48, pp. 117–130, ISSN: 1361-8415, DOI: 10.1016/j.media.2018.06.001.

Paszke, A., S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala (2019), « PyTorch: An Imperative Style, High-Performance Deep Learning Library », *in*: *Advances in Neural Information Processing Systems 32*, Curran Associates, Inc., pp. 8024–8035.

Pickett, J. and P. Savala (Oct. 2022), « Rate of publication hastens, but number of publications slows academic promotion », *in*: *PLOS ONE* 17, e0276616, DOI: 10.1371/journal.pone.0276616.

Pölsterl, S. (2020), « scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn », *in*: *Journal of Machine Learning Research* 21.*212*, pp. 1–6.

Pölsterl, S., T. N. Wolf, and C. Wachinger (2021), « Combining 3D Image and Tabular Data via the Dynamic Affine Feature Map Transform », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 688–698.

Prabhakar, C., H. B. Li, J. C. Paetzold, T. Loehr, C. Niu, M. Mühlau, D. Rueckert, B. Wiestler, and B. Menze (2023), « Self-pruning Graph Neural Network for Predicting Inflammatory Disease Activity in Multiple Sclerosis from Brain MR Images », *in*: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer, pp. 226–236.

Radford, A., J. W. Kim, C. Hallacy, A. Ramesh, G. Goh, S. Agarwal, G. Sastry, A. Askell, P. Mishkin, J. Clark, G. Krueger, and I. Sutskever (2021), « Learning Transferable Visual Models From Natural Language Supervision », *in*: *International Conference on Machine Learning*.

Raymond, N., H. Laribi, M. Caru, M. Mitiche, V. Marcil, M. Krajinovic, D. Curnier, D. Sinnett, and M. Vallières (July 2024), « Development of Error Passing Network for

Optimizing the Prediction of VO_2 peak in Childhood Acute Leukemia Survivors », *in*: *Proceedings of the fifth Conference on Health, Inference, and Learning*, PMLR, pp. 506–521.

Rist, L., O. Taubmann, A. Mühlberg, F. Denzinger, F. Thamm, M. Sühling, D. Nörenberg, J. W. Holch, S. Maurus, L. Gebauer, T. Huber, and A. Maier (2023), « Spatial Lesion Graphs: Analyzing Liver Metastases with Geometric Deep Learning for Cancer Survival Regression », *in*: *IEEE International Symposium on Biomedical Imaging*, pp. 1–5, DOI: 10.1109/ISBI53787.2023.10230367.

Rochman, S., A. Szeskin, R. Lederman, J. Sosna, and L. Joskowicz (Feb. 2024), « Graph-based automatic detection and classification of lesion changes in pairs of CT studies for oncology follow-up », *in*: *International Journal of Computer Assisted Radiology and Surgery* 19.*2*, pp. 241–251, ISSN: 1861-6429, DOI: 10.1007/s11548-023-03000-2.

Sarkis, M., M. Rizkallah, and S. Moussaoui (May 2025), « Graph Force Learning from EEG Signals for Mental Workload Classification », *in*: *HAL*, hal-05058873.

Schöder, H. and C. H. Moskowitz (2016), « Metabolic Tumor Volume in Lymphoma: Hype or Hope? », *in*: *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 34 30, pp. 3591–3594.

Sehn, L. H. and G. Salles (2021), « Diffuse Large B-Cell Lymphoma. », *in*: *The New England journal of medicine* 384 9, pp. 842–858.

Shi, Y., Z. Huang, S. Feng, H. Zhong, W. Wang, and Y. Sun (Aug. 2021), « Masked label prediction: Unified message passing model for semi-supervised classification », *in*: *Proceedings of the thirtieth international joint conference on artificial intelligence, IJCAI-21*, pp. 1548–1554, DOI: 10.24963/ijcai.2021/214.

Shnain, A. H., G. Pushpalakshmi, th K. Hemalatha, R. Veer, S. Sihman, and B. Rupavath (2024), « A Deep Learning Based Metastasis Prediction System Using Multi Omics Data and Recursive Neural Network », *in*: *2024 4th International Conference on Mobile Networks and Wireless Communications (ICMNWC)*, pp. 1–5.

Silva Filho, T., H. Song, M. Perello-Nieto, R. Santos-Rodriguez, M. Kull, and P. Flach (Sept. 2023), « Classifier calibration: a survey on how to assess and improve predicted class probabilities », *in*: *Machine Learning* 112.*9*, pp. 3211–3260, ISSN: 1573-0565, DOI: 10.1007/s10994-023-06336-7.

Stym-Popper, J., N. Painchaud, C. Rambour, P.-Y. Courand, N. Thome, and O. Bernard (2025), « DAFTED: Decoupled Asymmetric Fusion of Tabular and Echocardiographic Data for Cardiac Hypertension Diagnosis », *in*: *MIDL 2025*.

Sudlow, C. L. M., J. E. Gallacher, N. E. Allen, V. Beral, P. Burton, J. Danesh, P. Downey, P. Elliott, J. Green, M. J. Landray, B. Liu, P. M. Matthews, G. J. Ong, J. P. Pell, A. J. Silman, A. Young, T. Sprosen, T. C. Peakman, and R. Collins

(2015), « UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age », *in*: *PLoS Medicine* 12.

Susanibar-Adaniya, S. and S. K. Barta (May 2021), « 2021 Update on Diffuse large B cell lymphoma: A review of current data and potential applications on risk stratification and management », *in*: *American journal of hematology* 96.*5*, pp. 617–629, ISSN: 0361-8609.

Tabaa, Y. A., C. Bailly, and S. Kanoun (2021), « FDG-PET/CT in Lymphoma: Where Do We Go Now? », *in*: *Cancers* 13.

Taherkhani, F., H. Kazemi, and N. M. Nasrabadi (Jan. 2019), « Matrix completion for graph-based deep semi-supervised learning », *in*: *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence and Thirty-First Innovative Applications of Artificial Intelligence Conference and Ninth AAAI Symposium on Educational Advances in Artificial Intelligence*, AAAI'19/IAAI'19/EAAI'19, AAAI Press, pp. 5058–5065, ISBN: 978-1-57735-809-1, DOI: 10.1609/aaai.v33i01.33015058.

Thiery, O., M. Rizkallah, C. Bailly, C. Bodet-Milin, E. Itti, R.-O. Casasnovas, S. Le Gouill, T. Carlier, and D. Mateus (2023), « Graph-based multimodal multi-lesion DLBCL treatment response prediction from PET images », *in*: *AI4Treat@MICCAI*.

— (Mar. 2025), « PET-based lesion graphs meet clinical data: An interpretable cross-attention framework for DLBCL treatment response prediction », *in*: *Computerized Medical Imaging and Graphics* 120, p. 102481, ISSN: 0895-6111, DOI: 10.1016/j.compmedimag.2024.102481.

Thiery, O., M. Rizkallah, H. Laribi, M. Vallières, T. Carlier, and D. Mateus (2025), « Population-graph post-hoc corrections of survival predictions for improved risk stratification », *in*: *GRAIL@MICCAI*.

Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin (Dec. 2017), « Attention is All you Need », *in*: *NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems*, pp. 6000–6010.

Vétil, R., C. Abi-Nader, A. Bône, M.-P. Vullierme, M.-M. Rohé, P. Gori, and I. Bloch (2023), « Non-redundant combination of hand-crafted and deep learning radiomics: Application to the early detection of pancreatic cancer », *in*: *Cancer prevention through early detection, CaPTion2023@MICCAI2023*, Cham: Springer Nature Switzerland, pp. 68–82, ISBN: 978-3-031-45350-2.

Wang, G., F. Fan, S. Shi, S. An, X. Cao, W. Ge, F. Yu, Q. Wang, X. Han, S. Tan, Y. Tan, and Z. Wang (2024), « Multi modality fusion transformer with spatio-temporal feature aggregation module for psychiatric disorder diagnosis », *in*: *Computerized Medical Imaging and Graphics* 114, p. 102368, ISSN: 0895-6111, DOI: https://doi.org/10.1016/j.compmedimag.2024.102368.

Wang, X., M. Zhu, D. Bo, P. Cui, C. Shi, and J. Pei (Aug. 2020), « AM-GCN: Adaptive Multi-channel Graph Convolutional Networks », *in*: *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1243–1253, DOI: 10.1145/3394486.3403177.

Wang, Z., D. Katsaros, J. Wang, N. Biglio, B. Y. Hernandez, P. Fei, L. Lu, H. A. Risch, and H. Yu (2023), « Machine learning-based cluster analysis of immune cell subtypes and breast cancer survival », *in*: *Scientific Reports* 13.

Wang, Z., Z. Wu, D. Agarwal, and J. Sun (2022), « MedCLIP: Contrastive Learning from Unpaired Medical Images and Text », *in*: *Conference on Empirical Methods in Natural Language Processing*.

Whybra, P., A. Zwanenburg, V. Andrearczyk, R. Schaer, A. Apte, A. Ayotte, B. Baheti, S. Bakas, A. Bettinelli, R. Boellaard, L. Boldrini, I. Buvat, G. J. R. Cook, F. Dietsche, N. Dinapoli, H. S. Gabryś, V. Goh, M. Guckenberger, M. Hatt, M. Hosseinzadeh, A. Iyer, J. Lenkowicz, M. A. L. Loutfi, S. Löck, F. Marturano, O. Morin, C. Nioche, F. Orlhac, S. Pati, A. Rahmim, S. M. Rezaeijo, C. G. Rookyard, M. R. Salmanpour, A. Schindele, I. Shiri, E. Spezi, S. Tanadini-Lang, F. Tixier, T. Upadhaya, V. Valentini, J. J. M. van Griethuysen, F. Yousefirizi, H. Zaidi, H. Müller, M. Vallières, and A. Depeursinge (2024), « The Image Biomarker Standardization Initiative: Standardized Convolutional Filters for Reproducible Radiomics and Enhanced Clinical Insights. », *in*: *Radiology* 310 2, e231319.

Wu, C., C. Wang, H. Zhou, Y. Zhang, Q. Wang, Y. Wang, and S. Wang (2024), « MM-Fusion: Multi-modality Diffusion Model for Lymph Node Metastasis Diagnosis in Esophageal Cancer », *in*: *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, vol. 15005, Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, pp. 469–479, ISBN: 978-3-031-72085-7, DOI: 10.1007/978-3-031-72086-4_44.

Wu, Z., S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu (Jan. 2021), « A Comprehensive Survey on Graph Neural Networks », *in*: *IEEE Transactions on Neural Networks and Learning Systems* 32.*1*, pp. 4–24, ISSN: 2162-2388, DOI: 10.1109/TNNLS.2020.2978386.

Yang, Z., M. Lukasik, V. Nagarajan, Z. Li, A. Rawat, M. Zaheer, A. K. Menon, and S. Kumar (Dec. 2023), « ResMem: Learn what you can and memorize the rest », *in*: *Advances in Neural Information Processing Systems* 36, pp. 60768–60790.

Yousefirizi, F., C. Gowdy, I. S. Klyuzhin, M. Sabouri, P. Tonseth, A. R. Hayden, D. Wilson, L. H. Sehn, D. W. Scott, C. Steidl, K. J. Savage, C. F. Uribe, and A. Rahmim (2024), « Evaluating Outcome Prediction via Baseline, End-of-Treatment, and Delta Radiomics on PET-CT Images of Primary Mediastinal Large B-Cell Lymphoma », *in*: *Cancers* 16.

Yousefirizi, F., I. S. Klyuzhin, J. H. O, S. Harsini, X. Tie, I. Shiri, M. Shin, C. Lee, S. Y. Cho, T. J. Bradshaw, H. Zaidi, F. Bénard, L. H. Sehn, K. J. Savage, C. Steidl, C. F. Uribe, and A. Rahmim (2024), « TMTV-Net: fully automated total metabolic tumor volume segmentation in lymphoma PET/CT images - a multi-center generalizability analysis. », *in*: *European journal of nuclear medicine and molecular imaging*.

Yu, C.-N., R. Greiner, H.-C. Lin, and V. Baracos (Dec. 2011), « Learning Patient-Specific Cancer Survival Distributions as a Sequence of Dependent Regressors », *in*: *Neural Information Processing Systems*.

Yuan, C., Q. Shi, X. Huang, L. Wang, Y. He, B. Li, W. L. Zhao, and D. Qian (2022), « Multimodal deep learning model on interim 18F-FDG PET/CT for predicting primary treatment failure in diffuse large B-cell lymphoma », *in*: *European Radiology* 33, pp. 77–88.

Zhang, H., R. Song, L. Wang, L. Zhang, D. Wang, C. Wang, and W. Zhang (Feb. 2023), « Classification of Brain Disorders in rs-fMRI via Local-to-Global Graph Neural Networks », *in*: *IEEE Transactions on Medical Imaging* 42.*2*, pp. 444–455, ISSN: 1558-254X, DOI: 10.1109/TMI.2022.3219260.

Zheng, S., Z. Zhu, Z. Liu, Z. Guo, Y. Liu, Y. Yang, and Y. Zhao (Sept. 2022), « Multi-Modal Graph Learning for Disease Prediction », *in*: *IEEE Transactions on Medical Imaging* 41.*9*, pp. 2207–2216, ISSN: 1558-254X, DOI: 10.1109/TMI.2022.3159264.

Zhou, D., O. Bousquet, T. N. Lal, J. Weston, and B. Scholkopf (2003), « Learning with Local and Global Consistency », *in*: *Neural Information Processing Systems*.

Zhou, T., Q. Cheng, H. Lu, Q. Li, X. Zhang, and S. Qiu (June 2023), « Deep learning methods for medical image fusion: A review », *in*: *Computers in Biology and Medicine* 160, p. 106959, ISSN: 0010-4825, DOI: 10.1016/j.compbiomed.2023.106959.

Zwanenburg, A., M. Vallières, M. A. Abdalah, H. J. Aerts, V. Andrearczyk, A. Apte, S. Ashrafinia, S. Bakas, R. J. Beukinga, R. Boellaard, M. Bogowicz, L. Boldrini, I. Buvat, G. J. R. Cook, C. Davatzikos, A. Depeursinge, M.-C. Desseroit, N. Dinapoli, C. V. Dinh, S. Echegaray, I. M. E. Naqa, A. Fedorov, R. Gatta, R. J. Gillies, V. J. Goh, M. Götz, M. Guckenberger, S. M. Ha, M. Hatt, F. Isensee, P. Lambin, S. Leger, R. T. H. Leijenaar, J. Lenkowicz, F. Lippert, A. Losnegård, K. Maier-Hein, O. Morin, H. Müller, S. Napel, C. Nioche, F. Orlhac, S. Pati, E. A. G. Pfaehler, A. Rahmim, A. Rao, J. Scherer, M. M. Siddique, N. M. Sijtsema, J. S. Fernandez, E. Spezi, R. J. H. M. Steenbakkers, S. Tanadini-Lang, D. Thorwarth, E. G. C. Troost, T. Upadhaya, V. Valentini, L. V. van Dijk, J. J. M. van Griethuysen, F. H. P. van Velden, P. Whybra, C. Richter, and S. Löck (2020), « The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. », *in*: *Radiology*, p. 191145.

COLLEGE SCIENCES
DOCTORAL DE L'INGENIERIE
PAYS DE LA LOIRE ET DES SYSTEMES

CENTRALE NANTES

Titre : Apprentissage graphique multimodal et interprétable pour l'identification des patients atteints de lymphome B diffus à grandes cellules à haut risque.

Mot clés : Apprentissage sur graphe, Graphe de lésions, Graphe de population, Réponse au traitement, Analyse de survie, DLBCL, TEP, Fusion multimodale, Attention croisée, Interprétabilité

Résumé : Le lymphome B diffus à grandes cellules (DLBCL) est un cancer lymphatique en croissance constante. Son diagnostic et suivi reposent sur l'analyse de biomarqueurs cliniques et d'images 18F-FDG TEP/TDM. Cette thèse vise à identifier précocement les patients atteints de DLBCL à haut risque en utilisant ces images et données cliniques tabulaires. Trois contributions sont apportées : premièrement, l'étude des meilleures caractéristiques des images pour prédire la survie sans progression à deux ans, montrant la pertinence d'une représentation par graphe des lésions pour modéliser l'information contenue dans les multiples lésions de l'image, qui sont distribuées et de nombre variable chez chaque patient. Deuxièmement, le développement d'un modèle multimodal utilisant un mécanisme d'attention croisée pour fusionner efficacement des données multimodales très hétérogènes, avec une étude poussée du module d'attention croisée et de la structure du modèle pour assurer son interprétabilité. Enfin, une approche de graphe de population est proposée pour propager l'erreur de prédiction d'un modèle de base et améliorer la qualité de ses prédictions de façon post-hoc, approche adaptée pour l'analyse de survie afin de gérer la censure et la correction de courbes complètes, et appliquée à l'identification de patients à haut risque.

Title: Multimodal graph learning frameworks for interpretable high-risk diffuse large B-cell lymphoma patient identification.

Keywords: Graph learning, Lesion graph, Population graph, Treatment response, Survival analysis, DLBCL, PET, Multimodal data fusion, Cross-attention, Interpretability

Abstract: Diffuse Large B-Cell Lymphoma (DLBCL) is a steadily growing type of lymphatic cancer. Diagnosis and follow-up rely on analyzing clinical biomarkers and 18F-FDG PET/CT images. The aim of this PhD is to identify high-risk DLBCL patients at an early stage relying on both images and tabular clinical data. Three major contributions are made: firstly, the exploration of the imaging features that are the most relevant for predicting two-year progression-free survival (PFS), while proving the relevance of a lesion graph representation to model the information contained in the multiple lesions on the image, which are sparsely distributed and variable in number for each patient. Secondly, the development of a multimodal model using a cross-attention mechanism to efficiently merge highly heterogeneous multimodal data, with an in-depth study of the cross-attention module and of the internal structure of the model to ensure its interpretability. Finally, a population graph approach is proposed to propagate the prediction error of a baseline model and improve the quality of its predictions in a post-hoc manner. This approach has been adapted for survival analysis to manage censoring and correction of complete curves instead of single values, and applied to the identification of high-risk patients.