Deep brain unsupervised anomaly detection model based on multimodality medical imaging

Daria Zotova

N°d'ordre NNT : 2024ISAL0042

THESE de DOCTORAT DE L'UNIVERSITE DE LYON
opérée au sein de
(Institut National des Sciences Appliquées, INSA - Lyon)

Ecole Doctorale N° 160
(ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE)

Spécialité/ discipline de doctorat :
Traitement du Signal et de l'Image

Soutenue publiquement le 30/05/2024, par :
Daria Zotova

Deep brain unsupervised anomaly detection model based on multimodality medical imaging

---

Devant le jury composé de :

Oliver, Arnau	Professeur des Universités	University of Girona	Rapporteur
Burgos, Ninon	Chargé de recherche HDR	CNRS	Rapporteur
Hammers, Alexander	PU-PH	King's College London	Examinateur
Jung, Julien	Professeur associé	Hospices Civils de Lyon	Examinateur
Lartizien, Carole	Directeur de recherche	INSA Lyon (CNRS)	Directrice de thèse

Département FEDORA – INSA Lyon - Ecoles Doctorales

SIGLE	ECOLE DOCTORALE	NOM ET COORDONNEES DU RESPONSABLE
ED 206 CHIMIE	**CHIMIE DE LYON** https://www.edchimie-lyon.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage secretariat@edchimie-lyon.fr	**M. Stéphane DANIELE** C2P2-CPE LYON-UMR 5265 Bâtiment F308, BP 2077 43 Boulevard du 11 novembre 1918 69616 Villeurbanne directeur@edchimie-lyon.fr
ED 341 E2M2	**ÉVOLUTION, ÉCOSYSTÈME, MICROBIOLOGIE, MODÉLISATION** http://e2m2.universite-lyon.fr Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.e2m2@univ-lyon1.fr	**Mme Sandrine CHARLES** Université Claude Bernard Lyon 1 UFR Biosciences Bâtiment Mendel 43, boulevard du 11 Novembre 1918 69622 Villeurbanne CEDEX e2m2.codir@listes.univ-lyon1.fr
ED 205 EDISS	**INTERDISCIPLINAIRE SCIENCES-SANTÉ** http://ediss.universite-lyon.fr Sec. : Bénédicte LANZA Bât. Atrium, UCB Lyon 1 Tél : 04.72.44.83.62 secretariat.ediss@univ-lyon1.fr	**Mme Sylvie RICARD-BLUM** Laboratoire ICBMS - UMR 5246 CNRS - Université Lyon 1 Bâtiment Raulin - 2ème étage Nord 43 Boulevard du 11 novembre 1918 69622 Villeurbanne Cedex Tél : +33(0)4 72 44 82 32 sylvie.ricard-blum@univ-lyon1.fr
ED 34 EDML	**MATÉRIAUX DE LYON** http://ed34.universite-lyon.fr Sec. : Yann DE ORDENANA Tél : 04.72.18.62.44 yann.de-ordenana@ec-lyon.fr	**M. Stéphane BENAYOUN** Ecole Centrale de Lyon Laboratoire LTDS 36 avenue Guy de Collongue 69134 Ecully CEDEX Tél : 04.72.18.64.37 stephane.benayoun@ec-lyon.fr
ED 160 EEA	**ÉLECTRONIQUE, ÉLECTROTECHNIQUE, AUTOMATIQUE** https://edeea.universite-lyon.fr Sec. : Philomène TRECOURT Bâtiment Direction INSA Lyon Tél : 04.72.43.71.70 secretariat.edeea@insa-lyon.fr	**M. Philippe DELACHARTRE** INSA LYON Laboratoire CREATIS Bâtiment Blaise Pascal, 7 avenue Jean Capelle 69621 Villeurbanne CEDEX Tél : 04.72.43.88.63 philippe.delachartre@insa-lyon.fr
ED 512 INFOMATHS	**INFORMATIQUE ET MATHÉMATIQUES** http://edinfomaths.universite-lyon.fr Sec. : Renée EL MELHEM Bât. Blaise PASCAL, 3e étage Tél : 04.72.43.80.46 infomaths@univ-lyon1.fr	**M. Hamamache KHEDDOUCI** Université Claude Bernard Lyon 1 Bât. Nautibus 43, Boulevard du 11 novembre 1918 69 622 Villeurbanne Cedex France Tél : 04.72.44.83.69 direction.infomaths@listes.univ-lyon1.fr
ED 162 MEGA	**MÉCANIQUE, ÉNERGÉTIQUE, GÉNIE CIVIL, ACOUSTIQUE** http://edmega.universite-lyon.fr Sec. : Philomène TRECOURT Tél : 04.72.43.71.70 Bâtiment Direction INSA Lyon mega@insa-lyon.fr	**M. Etienne PARIZET** INSA Lyon Laboratoire LVA Bâtiment St. Exupéry 25 bis av. Jean Capelle 69621 Villeurbanne CEDEX etienne.parizet@insa-lyon.fr
ED 483 ScSo	**ScSo**[1] https://edsciencessociales.universite-lyon.fr Sec. : Mélina FAVETON Tél : 04.78.69.77.79 melina.faveton@univ-lyon2.fr	**M. Bruno MILLY** (INSA : J.Y. TOUSSAINT) Univ. Lyon 2 Campus Berges du Rhône 18, quai Claude Bernard 69365 LYON CEDEX 07 Bureau BEL 319 bruno.milly@univ-lyon2.fr

[1] ScSo : Histoire, Géographie, Aménagement, Urbanisme, Archéologie, Science politique, Sociologie, Anthropologie

Abstract

According to the World Health Organization (WHO), nearly 65 million people worldwide are affected by epilepsy. Epilepsy is a chronic neurological disorder characterized predominantly by recurrent and unpredictable interruptions of normal brain function, called epileptic seizures. For a third of diagnosed patients seizures cannot be controlled by pharmacotherapy. For such patients, a treatment would be to perform surgical resection of the epileptogenic zone. The success of such surgeries largely depends on the accuracy of the epileptogenic zone localization. Neuroimaging, including magnetic resonance imaging (MRI) and positron emission tomography (PET), has been increasingly considered in the pre-surgical examination routine.

This work attempts to enhance a computer-aided diagnosis (CAD) system for epileptogenic lesion detection, leveraging multimodal neuroimaging data, building upon the foundation laid in previous work by Alaverdyan et al., 2020. The proposed system employs unsupervised deep siamese networks for learning normal brain representations from non-pathological scans, followed by a series of one-class SVM models at the voxel level to generate an anomaly score map. The model, initially tested on T1-weighted and FLAIR MRI exams, demonstrated a sensitivity of 61%. The inclusion of PET imaging data is explored to assess its additional value. However, due to the ethical considerations surrounding radiation exposure in healthy individuals, the development and integration of synthetic PET data, simulating healthy brain scans, emerges as a potential solution.

This thesis makes contributions by: (1) developing strategies for generating synthetic PET images from T1 MRI scans, demonstrating their utility in augmenting original PET data to enhance the detection capabilities of the CAD system; (2) investigating the use of synthetic PET images as substitutes for real PET data in scenarios of incomplete training sets, showing that synthetic PET significantly addresses the missing modality challenge; and (3) identifying the optimal strategy for integrating multiple imaging modalities within the detection model, with late fusion of T1, FLAIR, and PET (both real and synthetic) showing to be the most effective in terms of detection ability of epileptogenic zones.

Structured into two main parts, the thesis first reviews recent deep learning advancements in medical imaging, particularly in brain image analysis, existing fusion strategies in medical imaging, methods to tackle missing data challenges including synthetic image generation, and CAD systems for brain pathology and epilepsy detection. The second part focuses on experiments for PET synthesis from T1 MRI images, the application of out-of-distribution (OOD) techniques to validate the similarity of synthetic to real images, and the detailed experimental results showcasing the improved performance of the epilepsy CAD model with the integration of synthetic PET data.

This work not only advances the field of medical imaging in epilepsy research but also offers a roadmap for future studies to further refine and enhance the detection and localization of epileptogenic lesions, potentially leading to better surgical outcomes and patient care.

Résumé étendu

Selon l'Organisation mondiale de la santé (OMS), près de 65 millions de personnes dans le monde sont touchées par l'épilepsie. L'épilepsie est un trouble neurologique chronique caractérisé principalement par des interruptions récurrentes et imprévisibles du fonctionnement normal du cerveau, appelées crises d'épilepsie. Pour un tiers des patients diagnostiqués, les crises ne peuvent être contrôlées par la pharmacothérapie. Pour ces patients, le traitement consisterait à effectuer une résection chirurgicale de la zone épileptogène. Le succès de ces opérations dépend en grande partie de la précision de la localisation de la zone épileptogène. La neuro-imagerie, y compris l'imagerie par résonance magnétique (IRM) et la tomographie par émission de positons (TEP), est de plus en plus souvent prise en compte dans les examens préchirurgicaux de routine.

Cette étude a pour but de voir comment améliorer un système d'aide au diagnostic (CAD) pour la détection des lésions épileptogènes, en s'appuyant sur des données de neuro-imagerie multimodales, ainsi que sur les bases posées dans des travaux antérieurs par Alaverdyan et al., 2020. Le système proposé utilise des réseaux siamois profonds non supervisés pour apprendre les représentations normales du cerveau à partir de scanners non pathologiques, suivis d'une série de modèles SVM à une classe au niveau du voxel pour générer une carte de score d'anomalie. Le modèle, initialement testé sur des examens IRM pondérés en T1 et FLAIR, a démontré une sensibilité de 61 %. L'inclusion de données d'imagerie TEP est envisagée pour évaluer sa valeur supplémentaire. Cependant, en raison des considérations éthiques entourant l'exposition aux radiations chez les individus sains, le développement et l'intégration de données PET synthétiques, simulant des scanners cérébraux sains, apparaît comme une solution potentielle.

Cette thèse apporte des contributions (1) en développant des stratégies pour générer des images TEP synthétiques à partir d'IRM T1, en démontrant leur utilité pour augmenter les données TEP originales, afin d'améliorer les capacités de détection du système CAD ; (2) en étudiant l'utilisation d'images TEP synthétiques comme substituts aux données TEP réelles dans des scénarios d'ensembles d'entraînement incomplets, en montrant que les images TEP synthétiques répondent de manière significative au défi de la modalité manquante ; et (3) en identifiant la stratégie optimale pour l'intégration des données TEP synthétiques dans le système CAD ; et (3) l'identification de la stratégie optimale d'intégration de modalités d'imagerie multiples dans le modèle de détection, la fusion tardive des images T1, FLAIR et TEP (réelles et synthétiques) s'avérant la plus efficace en termes de capacité de détection des zones épileptogènes.

Structurée en deux parties principales, cette thèse passe d'abord en revue les avancées récentes de l'apprentissage profond en imagerie médicale, en particulier dans l'analyse d'images cérébrales, les stratégies de fusion existantes en imagerie médicale, les méthodes pour relever les défis des données manquantes, y compris la génération d'images synthétiques, et les systèmes de CAD pour la pathologie cérébrale et la détection de l'épilepsie. La deuxième partie se concentre sur des expériences de synthèse de TEP à partir d'images IRM T1, sur l'application de techniques

hors distribution (OOD) pour valider la similarité des images synthétiques avec les images réelles, et sur les résultats expérimentaux détaillés montrant l'amélioration des performances du modèle de CAD pour l'épilepsie avec l'intégration de données synthétiques de TEP.

Ce travail ne fait pas seulement progresser le domaine de l'imagerie médicale dans la recherche sur l'épilepsie, mais offre également une feuille de route pour les études futures afin d'affiner et d'améliorer la détection et la localisation des lésions épileptogènes, ce qui pourrait conduire à de meilleurs résultats chirurgicaux et à de meilleurs soins pour les patients.

Synthèse par chapitre

Chapitre 1 : Apprentissage profond pour l'imagerie médicale

L'apprentissage profond est un ensemble de méthodes d'apprentissage visant à modéliser les données à l'aide d'architectures complexes combinant différentes transformations non linéaires. Les réseaux neuronaux d'inspiration biologique permettent aux ordinateurs d'apprendre à partir de données d'observation et sont les éléments clés de l'apprentissage profond. Ils sont composés de plusieurs couches pour apprendre des représentations cachées des données avec plusieurs niveaux d'abstraction. Chacune de ces couches effectue des transformations non linéaires des données d'entrée avant de les transmettre à une autre couche, et c'est ainsi que des fonctions très complexes sont apprises. Nous pouvons donc formuler l'objectif d'un réseau neuronal comme étant l'approximation d'une fonction f pour mettre en correspondance une entrée x avec une catégorie $y : y = f(x;\theta)$, où θ sont les paramètres pouvant être appris qui résultent en la meilleure approximation de la fonction. Généralement, le réseau est représenté par la composition de plusieurs fonctions. L'estimation des paramètres est obtenue en minimisant une fonction de perte sur certaines données d'apprentissage à l'aide d'un algorithme de descente de gradient.

Les premières formes de réseaux neuronaux, comme le perceptron, ont évolué vers des structures plus complexes telles que les perceptrons multicouches (MLP) et les réseaux neuronaux convolutifs (CNN), ces derniers ont révolutionné le traitement des images en capturant les relations spatiales par le biais de filtres, en réduisant les paramètres grâce à des poids partagés et en regroupant les couches.

Dans le domaine médical, l'apprentissage profond a connu un succès notable, en particulier dans l'imagerie médicale, en tirant parti des big data et des ressources informatiques avancées. Il identifie des modèles de manière autonome, ce qui facilite son utilisation par des non-spécialistes. Les applications d'apprentissage profond couvrent l'enregistrement, la segmentation, la reconstruction, la classification, la génération et la détection d'images médicales. Il contribue ainsi de manière significative aux processus de diagnostic et de traitement. Par exemple, les algorithmes d'apprentissage profond ont montré leur potentiel dans la détection des lésions dans les mammographies, l'aide au diagnostic du cancer de la prostate et l'estimation des risques de malignité des nodules pulmonaires, avec des performances comparables à celles des experts humains. Cependant, la dépendance des méthodes d'apprentissage profond à l'égard de grands ensembles de données annotées pose des problèmes, en particulier dans le domaine médical, en raison de la confidentialité, des préoccupations éthiques et de la variabilité des données. Cette situation a suscité l'intérêt pour les méthodes d'apprentissage faiblement ou non supervisées et la génération de données synthétiques pour améliorer la formation avec des ensembles de données incomplets. L'apprentissage faiblement supervisé utilise des annotations imparfaites, l'apprentissage semi-supervisé exploite les données étiquetées et

non étiquetées, et l'apprentissage non supervisé découvre des modèles dans des ensembles de données non étiquetés. La génération de données synthétiques répond aux problèmes de rareté et de confidentialité des ensembles de données médicales, bien que son efficacité et son acceptation restent des domaines de recherche active. L'intégration de multiples modalités d'imagerie permet d'obtenir des informations complètes sur les conditions médicales, ce qui est essentiel pour des diagnostics précis tels que la détection de l'épilepsie. La combinaison de données provenant de différentes sources, telles que l'IRM et la TEP, nécessite des techniques de fusion sophistiquées afin d'exploiter les informations complémentaires et d'améliorer ainsi la précision du diagnostic imagerie médicale.

Chapitre 2 : Utilisation des images multimodales

La fusion d'images médicales provenant de différentes sources améliore la précision du diagnostic et est de plus en plus utilisée pour des tâches telles que la segmentation et le développement de modèles de diagnostic ou de pronostic. Les différents types d'images fournissent des informations uniques : L'IRM offre des vues détaillées des organes et des tissus mous sans radiation, la tomodensitométrie excelle dans la représentation des os, tandis que la TEP et la TEMP fournissent des données métaboliques et fonctionnelles. L'imagerie multimodale améliore l'apprentissage des réseaux neuronaux en combinant des caractéristiques provenant de différentes sources pour une meilleure représentation des données et une meilleure performance du modèle. Les stratégies de fusion comprennent la fusion précoce (au niveau de l'entrée), intermédiaire (au niveau de la couche) et tardive (au niveau de la décision), chacune avec des mécanismes distincts pour l'intégration des données d'image.

Fusion précoce

La première stratégie, la fusion précoce, combine les modalités d'imagerie dès le départ, en utilisant soit des données brutes, soit des caractéristiques spécifiques à la modalité. Cette méthode est simple et implique généralement la concaténation d'entrées provenant de différentes sources. Elle fonctionne le mieux avec des données multimodales homogènes, le terme "homogène" signifiant que les données ont une résolution spatiale, des contrastes et des informations anatomiques ou pathologiques similaires. L'homogénéité au sein d'une modalité est cruciale pour une analyse fiable, bien que les différences entre les modalités soient attendues et fournissent des informations complémentaires.

Parmi les premiers avantages de la fusion, citons la réduction de la complexité informatique grâce à l'apprentissage d'un modèle unique, adapté aux données homogènes mais potentiellement limité par des entrées hétérogènes. Elle a été appliquée à diverses tâches médicales, telles que la segmentation des lésions de la sclérose en plaques et la classification multi-classes des patients atteints de la maladie d'Alzheimer, démontrant une amélioration des performances et de la précision de la classification [Brosch et al., 2016, Thung et al., 2017].

Outre les données d'imagerie, la fusion précoce peut également intégrer du texte, des caractéristiques élaborées à la main ou des signaux 1D [Engemann et al., 2020, Vaghari et al., 2022]. Les exemples incluent la combinaison de l'IRM avec des informations sur le patient, des images radiographiques avec des données sur le patient, et l'utilisation d'approches multimodales pour améliorer la précision de la classification dans les diagnostics médicaux, illustrant la polyvalence et le potentiel de la

fusion précoce dans les applications médicales.

Fusion intermédiaire

La fusion intermédiaire consiste à former des réseaux neuronaux distincts pour chaque modalité d'imagerie, puis à combiner les caractéristiques de leurs couches intermédiaires pour former une nouvelle entrée pour le modèle final. Cette approche facilite l'intégration d'entrées hétérogènes en tirant parti d'un seul modèle d'apprentissage final. Des études ont montré une amélioration des performances avec la fusion intermédiaire par rapport à l'utilisation des seules données d'imagerie en intégrant des caractéristiques d'image et cliniques, comme on le voit dans la classification de la maladie d'Alzheimer et la prédiction du risque de cancer du sein [Spasov et al., 2018, Yala et al., 2019]. Par exemple, le diagnostic de la maladie d'Alzheimer à l'aide de données IRM, TEP et génétiques utilise un réseau neuronal profond à trois niveaux pour traiter efficacement les différents types de données. Une autre application dans la segmentation des disques intervertébraux démontre que la fusion intermédiaire est plus performante que la fusion précoce. La combinaison de stratégies de fusion précoce et intermédiaire, comme dans la segmentation des tumeurs cérébrales [L. Chen et al., 2018], améliore les performances d'extraction des caractéristiques et de segmentation, certaines méthodes obtenant des scores Dice élevés dans des défis tels que le défi BraTS 2020 [W. Zhang et al., 2021].

Fusion tardive

Dans la fusion tardive, des réseaux distincts traitent chaque modalité d'imagerie et leurs résultats sont combinés pour prendre la décision finale, souvent à l'aide d'une moyenne, d'un vote majoritaire ou de méta-classifieur. Cette technique utilise les informations distinctes de chaque modalité pour une analyse plus complète.

La fusion tardive s'est avérée prometteuse dans diverses applications, comme l'amélioration de la segmentation des tissus cérébraux des nourrissons en surpassant les modèles de fusion unimodaux et multimodaux précoces [Nie et al., 2016]. En outre, pour la segmentation des tumeurs cérébrales, une combinaison de techniques de fusion précoce et tardive a été employée sur l'ensemble de données BraTS 2017, ce qui a permis d'améliorer la précision en se concentrant de manière séquentielle sur différents composants tumoraux [G. Wang et al., 2017].

En outre, Qiu et al., 2018 a appliqué la fusion tardive pour classer les troubles cognitifs légers par rapport à la cognition normale, démontrant des améliorations par rapport aux modèles reposant sur une seule modalité. Cela illustre l'efficacité de la fusion tardive dans l'exploitation d'informations complémentaires provenant de sources multiples pour améliorer les résultats diagnostiques et analytiques en imagerie médicale.

Fusion basée sur le mécanisme de l'attention

Les progrès récents en matière de fusion de caractéristiques pour l'imagerie médicale ont mis l'accent sur l'utilité des mécanismes d'attention, qui peuvent être classés en attention spatiale, en attention de canal et en un hybride des deux. Ces mécanismes visent à mettre en évidence les caractéristiques les plus significatives pour des tâches médicales spécifiques.

[Oktay et al., 2018] a présenté un U-Net d'attention pour la segmentation abdominale par tomodensitométrie, employant des portes d'attention pour améliorer la

sensibilité aux régions de premier plan, améliorant ainsi la qualité de la segmentation. Le mécanisme d'attention met sélectivement l'accent sur les régions saillantes de l'image et supprime les zones moins pertinentes.

Dans une approche innovante de la segmentation des tumeurs cérébrales, [T. Zhou et al., 2020] a utilisé un réseau à trois niveaux incorporant un mécanisme d'attention pour la fusion des caractéristiques. Cette méthode commence par des segmentations grossières à partir d'un réseau 3D, suivies d'un modèle de fusion à plusieurs encodeurs qui affine ces segmentations. Le mécanisme d'attention permet d'identifier les caractéristiques les plus informatives pour une segmentation précise.

Un autre travail remarquable de [T. Zhou et al., 2021] a présenté un modèle d'apprentissage des corrélations latentes multi-sources pour la segmentation des tumeurs cérébrales à partir de données d'IRM. Il comprend un nouveau modèle de corrélation qui combine les représentations modales individuelles en un ensemble de caractéristiques unifiées et informatives pour améliorer les résultats de la segmentation.

Ces études soulignent l'efficacité des mécanismes d'attention pour améliorer la fusion des caractéristiques et la précision de la segmentation dans les tâches d'imagerie médicale.

Conclusions

Le choix d'une stratégie de fusion efficace pour l'apprentissage profond reste une question importante. D'un point de vue méthodologique, chacune d'entre elles a ses avantages et ses inconvénients, et il est rare que les trois méthodes aient été étudiées pour la même tâche sur le même ensemble de données. La première stratégie de fusion, qui consiste à concaténer les modalités pour former l'espace d'entrée, est la plus répandue, mais elle n'exploite pas les relations entre les différentes modalités. En revanche, dans le cas de la fusion intermédiaire, la connexion entre les différentes couches permet de saisir les relations complexes entre les modalités. La stratégie de fusion tardive permet généralement d'obtenir de meilleures performances que la fusion précoce, en particulier pour les tâches de segmentation [T. Zhou, Ruan et al., 2019], un seul réseau apprenant une représentation indépendante des caractéristiques des différentes modalités, mais au prix d'une perte de mémoire et de temps de calcul. Dans la pratique médicale, il n'est pas rare d'être confronté à des données manquantes ou incomplètes, lorsque certains patients ne disposent que de données cliniques ou ne disposent pas d'une modalité d'imagerie particulière. Dans ce cas, la fusion tardive conserve la capacité de faire des prédictions, puisque les fonctions d'agrégation (vote majoritaire ou calcul de la moyenne) peuvent être appliquées même en cas de modalité manquante. La fusion tardive est favorable dans ce scénario, car elle prend en compte chaque modalité séparément. Les mécanismes basés sur l'attention améliorent les prédictions du modèle en éliminant les régions non significatives tout en supprimant les caractéristiques extraites lorsqu'ils sont utilisés aux niveaux intermédiaires ou aux couches de décodage pour moduler la focalisation spatiale.

Chapitre 3 : Apprendre avec des données manquantes

Méthodes pour gérer les données manquantes

Dans la pratique médicale, les médecins ont souvent recours à différents types d'imagerie pour diagnostiquer et traiter les patients. Parfois, certaines modalités

d'imagerie peuvent être manquantes pour diverses raisons telles que des différences dans les protocoles cliniques, l'incapacité du patient à passer certains examens ou l'absence de données fournies par certaines institutions. Le traitement de ces données manquantes est crucial pour une analyse d'image non biaisée et complète. Si l'on se contente d'éliminer les sujets dont les données sont manquantes, on perd beaucoup d'informations précieuses et on réduit le nombre d'échantillons disponibles pour l'entraînement des modèles d'apprentissage automatique.

Une méthode courante pour traiter les données manquantes est l'imputation ou l'attribution, où les valeurs manquantes sont complétées sur la base de suppositions faites à partir du reste des données. Cette méthode s'est avérée plus efficace que celle qui consiste à ignorer les sujets dont les données sont incomplètes, bien qu'elle ne soit pas toujours idéale pour les ensembles de données complexes et de grande taille.

Pour les données d'imagerie manquantes, il existe trois stratégies principales :

1. N'utiliser que les images disponibles.
2. Créer des versions artificielles des images manquantes et les utiliser pour compléter l'ensemble de données.
3. Combiner les images disponibles de manière à capturer leurs caractéristiques les plus importantes, même sans les images manquantes.

Récemment, de nouvelles méthodes ont été mises au point, notamment en ce qui concerne la troisième stratégie. Ces méthodes combinent les images disponibles sous une forme plus simple qui représente les principales caractéristiques de toutes les images, ce qui peut être particulièrement utile car cela ne dépend pas de la qualité des images créées artificiellement.

Un exemple notable est l'approche de segmentation d'image hétéro-modale (HeMIS) pour la segmentation des tumeurs cérébrales. Cette méthode traite chaque type d'image disponible séparément avant de les combiner à l'aide de méthodes statistiques pour capturer les caractéristiques générales présentes dans toutes les images. Elle est conçue pour fonctionner même si certains types d'images sont manquants.

Une autre technique, visant la même tâche de segmentation des tumeurs cérébrales, traite chaque type d'image séparément, mais utilise ensuite un module spécial pour améliorer les caractéristiques des images manquantes avant de tout combiner. Cette approche vise également à combler les lacunes laissées par les données manquantes et s'est avérée efficace et robuste.

Dans l'ensemble, ces stratégies et techniques visent à utiliser au mieux les images disponibles, tout en essayant de relever le défi des données manquantes.

Synthèse des données manquantes

Lorsque l'acquisition est impossible ou limitée, il peut être utile de générer des modalités manquantes au lieu d'effectuer une nouvelle acquisition. Différents réseaux et architectures ont été proposés ces dernières années pour cette tâche.

Autoencodeurs variationnels (VAE) : Les VAE sont des réseaux neuronaux qui projettent la distribution des données dans un espace latent, ce qui permet de générer de nouveaux points de données par interpolation/extrapolation. Ils comprennent un encodeur qui fait correspondre les données d'entrée à un espace latent et un décodeur qui reconstruit les données à partir de cet espace. Efficaces en imagerie médicale, les VAE ont été utilisés pour synthétiser des images d'échographie et d'IRM, démontrant leur utilité en tant que technique d'augmentation des données [Pesteie et al., 2019].

U-net : Conçues initialement pour la segmentation d'images biomédicales, les architectures U-net s'appuient sur une structure d'auto-encodeur augmentée de connexions de contournement (skip connections) pour conserver les informations spatiales perdues lors du rééchantillonnage (upsampling). Cette caractéristique facilite la reconstruction et la segmentation précise des images. Les adaptations des U-net pour la synthèse d'images médicales, telles que la tomodensitométrie à partir d'images MR et la TEP à partir de données IRM, ont montré leur capacité à générer des images synthétiques réalistes pour diverses applications médicales [Ronneberger et al., 2015, X. Han, 2017, Chartsias et al., 2017, Sikka et al., 2018, Kalantar et al., 2021].

Réseaux adversaires génératifs (GAN) : Dotés d'un générateur et d'un discriminateur en compétition, les GAN permettent de créer des images photoréalistes. Leur application à la synthèse d'images médicales couvre la conversion d'images IRM en images CT et l'inverse, les innovations CycleGAN incorporant la cohérence des cycles pour assurer la fidélité dans la traduction du domaine. Les GAN ont surpassé d'autres méthodologies dans la production d'images réalistes, en particulier pour la segmentation des tumeurs cérébrales et la synthèse PET à partir de l'IRM, en offrant un niveau de détail et de réalisme supérieur [Goodfellow et al., 2014, J.-Y. Zhu et al., 2017, H. Yang et al., 2018, Armanious et al., 2019, Wei et al., 2019, Yaakub et al., 2019, Y. Wang, Zhou et al., 2018, Flaus et al., 2023].

Transformeurs : Inspirés par leur succès dans le traitement du langage naturel, les transformateurs ont été étudiés pour la synthèse d'images médicales grâce à leur capacité à modéliser les dépendances à longue portée spatiale via des mécanismes d'attention. Les propositions comprennent des transformateurs de vision (ViT) et des modèles hybrides fusionnant des CNN et des transformeurs pour des tâches telles que la reconstruction PET à partir d'images PET et IRM à faible dose, et la synthèse d'IRM T1 à partir de scans T2. Ces modèles soulignent le potentiel des transformeurs pour discerner des modèles complexes dans les images médicales, améliorant ainsi la qualité et l'efficacité de la synthèse [Vaswani et al., 2017, Dosovitskiy et al., 2020, Watanabe et al., 2021, Luo et al., 2021, X. Zhang et al., 2021, Dalmaz et al., 2022, Shin et al., 2020].

Chaque méthode présente une nouvelle voie pour relever le défi des données manquantes dans l'imagerie médicale, de l'apprentissage statistique des VAE à l'apprentissage adversaire des GAN et à la compréhension contextuelle des transformeurs. Leur application à la synthèse de l'imagerie médicale a donné des résultats prometteurs, améliorant la qualité et le réalisme des images synthétiques, ce qui est essentiel pour faire progresser le diagnostic médical et la planification des traitements.

Chapitre 4 : Systèmes CAD pour la pathologie cérébrale détections

La technologie de détection assistée par ordinateur (CAD) vise à réduire les erreurs et les faux positifs dans l'interprétation des images médicales par les médecins. Les systèmes CAD aident les médecins en fournissant une analyse quantitative des images, en calculant les probabilités de diagnostic et en identifiant les anomalies. Un système de CAD typique comprend les étapes suivantes : prétraitement de l'image, extraction des caractéristiques et conception d'un modèle statistique (à l'aide d'un algorithme d'apprentissage automatique ou d'apprentissage profond).

1. **Prétraitement des images :** Cette étape initiale consiste à préparer les images

avant qu'elles ne soient introduites dans le modèle CAD. Les tâches de prétraitement courantes en imagerie médicale comprennent le débruitage, la correction du champ de distorsion, l'enregistrement, la normalisation et la standardisation. Ces étapes sont cruciales pour améliorer la robustesse des modèles.

2. **Extraction des caractéristiques :** Ce processus consiste à réduire la dimensionnalité de la région d'intérêt (ROI) dans les images à des fins d'analyse. Il convertit les données pixellisées de bas niveau en représentations de plus haut niveau, ce qui permet d'extraire des informations utiles tout en réduisant le volume de données. Les caractéristiques peuvent être apprises automatiquement à l'aide d'architectures d'apprentissage profond adaptées à des tâches spécifiques.
3. **Conception de modèles statistiques :** Cette étape consiste à développer des modèles statistiques axés sur les données et accordés sur un ensemble de données d'entraînement. Le processus de réglage ajuste les hyperparamètres des modèles en fonction d'une fonction de perte qui évalue les performances. L'objectif est de créer une fonction de décision qui traite les vecteurs de caractéristiques d'entrée et produit des variables de décision pour les prédictions. Le choix du modèle et la conception de la fonction de perte dépendent de la tâche (par exemple, régression, classification) et de la nature des données (entièrement étiquetées, partiellement étiquetées ou entièrement non étiquetées), ce qui correspond aux paradigmes d'apprentissage supervisé, semi-supervisé ou non supervisé. Cela permet de s'assurer que le modèle est conçu de manière optimale pour le problème spécifique, avec des hyperparamètres réglés pour améliorer les performances et la précision.

Approches supervisées pour la détection des lésions cérébrales

Dans les approches d'apprentissage supervisé pour la détection des lésions cérébrales, les modèles sont formés avec des ensembles de données étiquetées pour effectuer des tâches telles que la segmentation, la classification et la détection des pathologies cérébrales. L'accent mis sur la sclérose en plaques (SEP) révèle la complexité de son diagnostic, les examens IRM étant essentiels pour visualiser et détecter les lésions, dont la taille et la visibilité varient en fonction du stade de la maladie. Les ensembles de données notables pour la recherche sur la sclérose en plaques comprennent l'ensemble de données ISBI, l'ensemble de données MSSEG (Multiple Sclerosis Segmentation) Challenge et MSSEG-2, chacun offrant des données d'IRM avec des annotations d'experts pour le développement et le test d'algorithmes.

Plusieurs études ont proposé des techniques de segmentation automatique pour la SEP, les classant en deux catégories : celles qui se concentrent uniquement sur la segmentation des lésions et celles qui s'intéressent à la fois à la segmentation du cerveau et à celle des lésions de la SEP. Les techniques impliquent l'utilisation d'ensembles de données publics tels que l'ISBI challenge et le MSSEG pour la validation. Les recherches menées par Wei et al [Wei et al., 2019, Wei et al., 2020] mettent en évidence le potentiel de la mesure de la teneur en myéline pour la détection précoce de la SEP à l'aide d'images dérivées de la TEP et de modèles basés sur le GAN.

D'autres pathologies cérébrales telles que la maladie des petits vaisseaux cérébraux (CSVD) [Hsieh et al., 2019, Duan et al., 2020, Shan et al., 2021], la calcification de l'artère carotide intracrânienne (ICAC) [Bortsova et al., 2021, Lai et al., 2022], et la maladie de Parkinson (PD) sont également abordées []. La détection de la MCVS a utilisé des modèles d'apprentissage profond pour identifier les caractéristiques IRM indicatives de la maladie. La détection de l'ICAC se concentre sur l'utilisation de

tomographies assistées par ordinateur et d'images échographiques pour évaluer le risque d'accident vasculaire cérébral ischémique et de démence. La détection de la maladie de Parkinson bénéficie des examens IRM et de modèles innovants pour le diagnostic précoce et le suivi, avec des études employant des réseaux neuronaux convolutionnels 3D et combinant des évaluations numériques des symptômes pour les tâches de classification [Chakraborty et al., 2020, Bhan et al., 2021, S. Zhu, 2022].

Approches non supervisées

Les approches d'apprentissage non supervisé, qui ne reposent pas sur des données étiquetées, offrent une solution en se concentrant sur la déviation des régions pathologiques par rapport aux tissus sains. Les autoencodeurs, en particulier les autoencodeurs variationnels (VAE), jouent un rôle crucial dans les méthodes non supervisées en apprenant à distinguer l'anatomie normale de l'anatomie anormale par la minimisation des erreurs de reconstruction. Ces modèles, entraînés exclusivement sur des données provenant de sujets sains, génèrent des cartes d'anomalies en identifiant les divergences entre les images originales du patient et leurs équivalents reconstruits. Cette approche s'est révélée prometteuse pour la détection des lésions de la sclérose en plaques [Baur, Denner et al., 2021, Baur et al., 2018, Vogelsanger et al., 2021], des tumeurs [Chatterjee et al., 2022, Zimmerer et al., 2019], et d'autres pathologies cérébrales [You et al., 2019] avec des performances comparables aux modèles supervisés comme U-Net [Baur, Wiestler, Muehlau et al., 2021].

En outre, des modèles non supervisés ont été utilisés pour détecter la pathologie de la sclérose en plaques dans des tissus cérébraux d'apparence normale grâce à une représentation unique des caractéristiques [Yoo et al., 2018], et pour détecter la maladie de Parkinson à l'aide d'une étude comparative entre les autocodeurs spatiaux et les autocodeurs siamois [Muñoz-Ramırez et al., 2021].

Les réseaux adverbiaux génératifs (GAN) trouvent également une application dans la détection des anomalies, facilitant l'identification des anomalies cérébrales à différents stades de la maladie dans les IRM et les tomodensitogrammes [C. Han et al., 2021, Simarro Viana et al., 2020]. En outre, l'approche ANT-GAN offre un nouveau moyen de générer des images IRM d'apparence normale à partir de leurs homologues anormaux, en "supprimant" efficacement les lésions sans nécessiter de données d'entraînement appariées [Sun et al., 2020].

Systèmes de CAD pour l'épilepsie

L'épilepsie, telle que définie par l'Organisation mondiale de la santé (OMS), est une maladie neurologique chronique prévalente qui touche environ 65 millions de personnes dans le monde, ce qui en fait l'une des maladies neurologiques les plus courantes. Caractérisée par des crises d'épilepsie récurrentes et ses impacts neurobiologiques, cognitifs, psychologiques et sociaux, environ 70% des personnes touchées peuvent se libérer des crises grâce à un traitement antiépileptique (AED) approprié. Cependant, environ 30% des patients présentent une résistance aux médicaments, ce qui nécessite des traitements alternatifs tels que des interventions neurochirurgicales, qui ont montré leur efficacité dans les cas d'épilepsie médicalement intraitable (EMI) [Thurman et al., 2011, Moshé et al., 2015, Ramey et al., 2013].

L'épilepsie du lobe temporal (ELT), la forme la plus répandue d'EIM, représente environ 80% des chirurgies de l'épilepsie. L'histopathologie la plus courante chez les patients atteints de TLE est la sclérose hippocampique (HS), la Ligue internationale contre l'épilepsie (ILAE) classant la HS en trois types distincts en fonction de la perte

de cellules neuronales et des schémas de gliose [Ramey et al., 2013, Cendes et al., 2014, Malmgren et al., 2012].

Les malformations du développement cortical (MCD) sont des causes importantes d'épilepsie, en particulier chez les enfants, dues à des défauts de développement du cortex cérébral. Les dysplasies corticales focales (DCF) sont les MCD les plus répandues chez les enfants atteints d'épilepsies focales pharmacorésistantes. La classification des DCF en trois types par l'ILAE met en évidence la diversité de ces lésions, qui vont de la dyslamination et de l'architecture tissulaire perturbée à la présence de neurones et de cellules ballonnets dysmorphiques [Raybaud et al., 2011, Guerrini et al., 2015, Kim et al., 2019].

Les techniques d'IRM jouent un rôle crucial dans la caractérisation non invasive des DCF, des caractéristiques IRM spécifiques étant associées aux différents types de DCF, ce qui facilite l'identification de ces lésions [Urbach et al., 2021, Duncan et al., 2016]. En outre, les hétérotopies nodulaires périventriculaires (PNH) et la polymicrogyrie (PMG) représentent d'autres formes de MCD, fréquemment associées à une épilepsie pharmacorésistante [Mirandola et al., 2017, Shain et al., 2013].

La localisation de la zone épileptogène (ZE) est essentielle pour un traitement efficace, en utilisant des méthodes d'imagerie telles que l'IRM et la TEP, ainsi que des techniques d'électroencéphalogramme (EEG). Le rôle de l'IRM est central dans la localisation de la ZE, avec des protocoles d'imagerie spécifiques recommandés pour l'évaluation de l'épilepsie [Duncan et al., 2016, Rüber et al., 2018, Bernasconi et al., 2019, Cendes, 2013, Malmgren et al., 2012].

Les examens TEP, en particulier la TEP au [18F]FDG, sont précieux, en particulier pour les patients négatifs à l'IRM, car ils permettent d'identifier la zone d'apparition de l'épileptogenèse. L'intégration de la TEP-FDG à l'IRM améliore la précision diagnostique pour la détection des ZE dans l'épilepsie focale en combinant les informations anatomiques et fonctionnelles [Willmann et al., 2007, Ding et al., 2018, Wong-Kisiel et al., 2018, Desarnaud et al., 2018, Kikuchi et al., 2021].

Les avancées récentes en matière d'apprentissage automatique et d'apprentissage profond ont considérablement contribué à la recherche sur l'épilepsie, en particulier à la détection automatique des crises, à la planification pré-chirurgicale et à la prédiction des résultats des interventions médicales et chirurgicales. En utilisant les données de l'électroencéphalogramme (EEG), plusieurs études ont visé à identifier les crises d'épilepsie, en obtenant une précision, une sensibilité et une spécificité élevées pour différencier des signaux normaux et épileptiques grâce à l'utilisation de réseaux neuronaux convolutionnels 2D (CNN) et de réseaux neuronaux récurrents à mémoire à long terme (LSTM-RNN) Abbasi et al., 2019 ; Akut, 2019 ; Bouallegue et al., 2020 ; Hussein et al., 2018 ; Najafi et al., 2022 ; San-Segundo et al., 2019 ; Türk et al., 2019.

Dans la détection de l'épilepsie, les systèmes de diagnostic assisté par ordinateur (CAD) s'appuient principalement sur des données de neuro-imagerie, avec deux orientations principales : la discrimination au niveau du patient et la latéralisation la localisation des foyers épileptogènes. Les études de discrimination au niveau du patient ont utilisé des machines à vecteurs de support (SVM) et divers modèles de réseaux neuronaux pour différencier les patients épileptiques des témoins sains, atteignant des taux de précision allant jusqu'à 97,6%. Ces études intègrent souvent la morphométrie basée sur le voxel (VBM) pour analyser les différences anatomiques cérébrales focales, bien que l'application clinique de la différenciation des sujets sains des épileptiques à l'aide de ces modèles reste limitée [Bharath et al., 2019, J. Huang et al., 2020, B. Zhou et al., 2020, S. Chen et al., 2020, Nemoto, 2017, Si et al., 2020, M.-H. Lee et al., 2020, Nguyen et al., 2021].

Pour la latéralisation des foyers d'épilepsie du lobe temporal (ELT) et la localisation des foyers épileptogènes, des techniques d'apprentissage automatique ont été appliquées aux données d'imagerie structurelle et fonctionnelle, montrant un potentiel dans l'identification des patients atteints d'ELT même lorsque les résultats de l'IRM sont négatifs. Ces tâches présentent toutefois des défis, notamment l'obtention d'étiquettes de vérité terrain précises pour les modèles de formation, en particulier pour les cas négatifs en IRM ou lorsque les lésions ne sont pas visibles sur les scanners [Fang et al., 2017, Mahmoudi et al., 2018, Bennett et al., 2019, Beheshti et al., 2020a, Beheshti et al., 2020b].

La plupart des études sur la localisation des lésions épileptiques se concentrent sur les cas positifs à l'IRM en raison de la difficulté d'obtenir des masques de lésion précis pour les patients négatifs à l'IRM. Diverses caractéristiques, y compris celles dérivées de la morphométrie basée sur la surface (SBM) et de l'épaisseur corticale, ont été utilisées. Bien que les caractéristiques artisanales restent courantes, on s'intéresse de plus en plus à l'exploration des caractéristiques pilotées par les données et des données d'imagerie multimodales afin d'améliorer les performances des modèles. Malgré les défis, certaines études ont réussi à détecter des lésions chez des patients négatifs à l'IRM, soulignant le potentiel des approches supervisées et semi-supervisées dans la détection de l'épilepsie [Gill et al., 2017, Gill et al., 2018, Alaverdyan et al., 2020, Adler et al., 2017, El Azami et al., 2016, Ahmed et al., 2015, Ahmed et al., 2016, Jin et al., 2018, Wagstyl et al., 2020].

L'exploration de l'imagerie TEP dans la détection de l'épilepsie, bien que moins courante, présente un potentiel prometteur. Des études exploitant l'imagerie TEP parallèlement aux données IRM ont fait état d'une sensibilité et d'une spécificité accrues dans la détection des lésions épileptiques, ce qui indique l'intérêt de combiner différentes modalités d'imagerie pour une détection plus précise de l'épilepsie [Tan et al., 2018, X. Zhang et al., 2021].

Les conclusions suivantes peuvent être tirées de cette étude :

— Les recherches menées jusqu'à présent se sont largement concentrées sur des types spécifiques d'épilepsie, tels que l'épilepsie du lobe temporal (ELT) et la dysplasie corticale focale (DCF), en adaptant les caractéristiques explorées à ces pathologies particulières. Il est possible d'améliorer les méthodes de détection de l'épilepsie en élargissant la gamme des caractéristiques prises en compte, ce qui pourrait favoriser la généralisation à différents types d'épilepsie.

— La tendance est de plus en plus à l'intégration de plusieurs modalités d'imagerie, les études exploitant de plus en plus les forces combinées des images T1, FLAIR et des scans TEP. Cette approche multimodale gagne en popularité en raison de sa capacité à améliorer les performances des modèles, ce qui indique un changement par rapport à la dépendance à l'égard de types d'imagerie uniques. L'étude suggère d'explorer davantage la manière dont ces modalités peuvent être intégrées, que ce soit par des techniques de fusion précoce, intermédiaire ou tardive, afin de maximiser l'efficacité du diagnostic.

— Une limite importante de la recherche actuelle est l'accent mis sur les cas positifs à l'IRM, où les lésions épileptogènes sont visibles sur les scanners. Le défi de détecter et d'évaluer avec précision l'épilepsie chez les patients négatifs à l'IRM reste un domaine critique pour les progrès futurs.

Chapitre 5 : Vue d'ensemble des données et du modèle

Dans ce chapitre, nous présentons un modèle CAD de base pour la détection de l'épilepsie, ainsi qu'une description détaillée de l'ensemble des données disponibles pour cette étude, contenant à la fois des témoins sains et des patients présentant des lésions épileptiques confirmées.

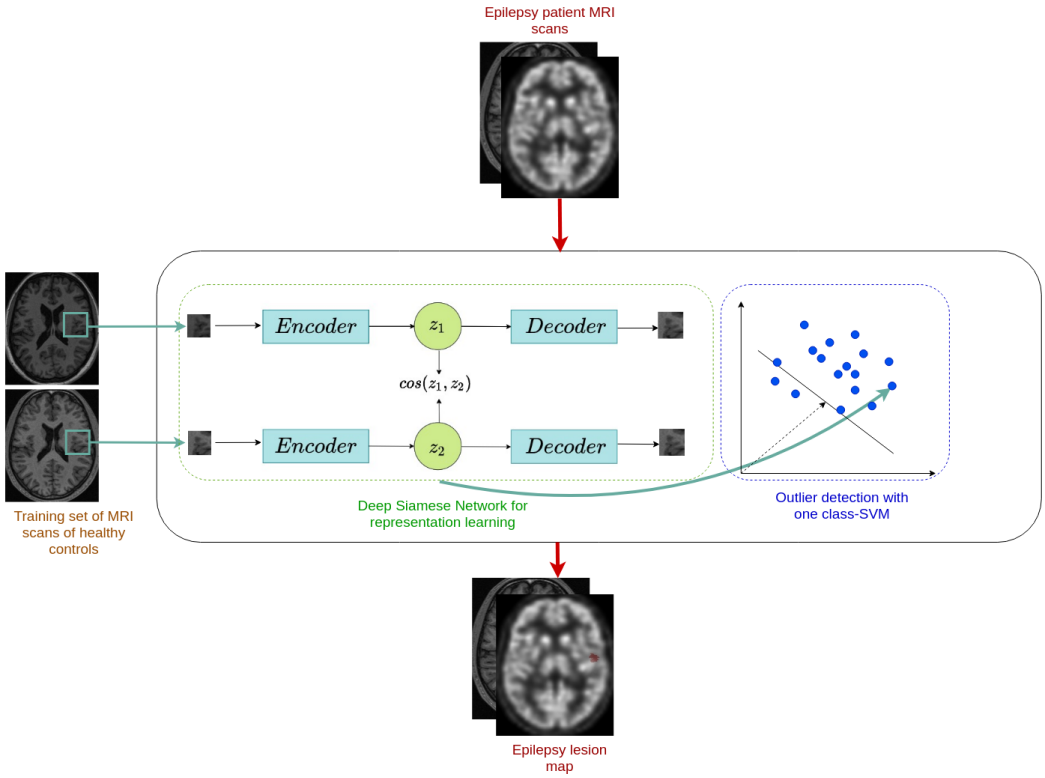

FIGURE 1 – Modèle général de détection de l'épilepsie.

Le modèle global de détection de l'épilepsie est présenté dans la figure 1. Il se compose de deux parties principales :

1. Apprentissage de représentations (cadre pointillé vert)
2. Apprentissage de la détection des valeurs aberrantes (cadre pointillé bleu)

Ces deux composants sont entraînés sur un ensemble de témoins sains. Au début, un réseau neuronal siamois (SNN) est entraîné sur des patchs extraits d'emplacements aléatoires dans le masque cérébral de sujets sains. Une fois l'entraînement terminé, un oc-SVM est construit et entraîné pour chaque voxel en prenant comme entrée les représentations dérivées du SNN. La représentation cachée pour l'oc-SVM correspond aux patchs des sujets sains centrés sur le voxel. Au total, nous avons autant de modèles oc-SVM que le nombre total de voxels dans le masque cérébral. Ensuite, l'image d'un patient peut être traitée par le système. Le système CAD génère une sortie sous la forme d'une carte de scores, qui correspond à la taille de l'image d'entrée. Dans cette carte de scores, chaque voxel est associé à une valeur, représentant la sortie de son modèle oc-SVM correspondant. À la dernière étape, la carte de scores de sortie subit un post-traitement pour générer une carte de grappes mettant en évidence les régions les plus suspectes détectées par le système.

Pour un apprentissage efficace du modèle, les données doivent être prétraitées. La première étape consiste à aligner toutes les acquisitions d'imagerie disponibles sur un modèle commun afin de garantir la correspondance voxel à voxel entre tous les sujets (sains et patients). Pour l'apprentissage la représentations, les images d'entrée sont divisées en patchs 15x15 pour le SNN, soit en tant que modalité unique,

soit dans une architecture multicanal où chaque canal représente une modalité différente. L'exploration de l'efficacité des entrées multimodales est un objectif clé, avec plus d'informations à suivre. Une fois préparées, les données sont transmises au composant suivant du système CAD.

Dans la partie **extraction des caractéristiques**, le modèle utilise un réseau siamois régularisé avec des autoencodeurs convolutifs profonds, tirant parti des avantages des autoencodeurs et des réseaux siamois, pour apprendre des représentations pour les données d'entrée. Cette approche est spécifiquement adaptée à la détection des valeurs aberrantes et consiste en deux sous-réseaux identiques partageant des paramètres et reliés par un module de coût.

Le processus commence par un ensemble de données $X = \{x_i\}_{i=1,\dots,n}, x_i \in \mathcal{R}^d$, visant à projeter les points de données dans un nouvel espace $\mathcal{Z}$ où les éléments similaires sont étroitement regroupés. Les paires de points normaux en entrée (x_1, x_2) sont traitées par des autoencodeurs convolutifs au sein du réseau siamois. Chaque sous-réseau comprend un encodeur E pour la compression des entrées dans un espace caché Z, suivi d'un décodeur D qui reconstruit l'entrée à partir de Z. La fonction de perte, conçue pour améliorer la qualité de la reconstruction et renforcer la similarité des représentations cachées, est définie comme suit :

$$L(x_1, x_2; \theta) = \sum_{t=1}^{2} |x_t - \hat{x}_t|_2^2 - \alpha \cdot cos(z_1, z_2) \tag{1}$$

Ici, la fonction de perte comprend une erreur de reconstruction et un terme pour maximiser la similarité cosinus entre les vecteurs de caractéristiques dans Z, avec α ajustant l'équilibre entre ces composants. La représentation résultante z est ensuite utilisée pour détecter les valeurs aberrantes.

Détection des valeurs aberrantes. La technique SVM à une classe (oc-SVM), une forme d'apprentissage non supervisé initialement introduite par Schölkopf et al. (2001), est examinée pour sa capacité à distinguer les échantillons d'une classe spécifique des autres. Elle fonctionne en recherchant un hyperplan qui sépare de l'origine tous les points de données dans un espace de caractéristiques transformé. Cette transformation utilise une fonction noyau, généralement une RBF (Radial Basis Function), pour cartographier les données dans un espace de dimension supérieure où la séparation est possible, comme le montre la figure 2.

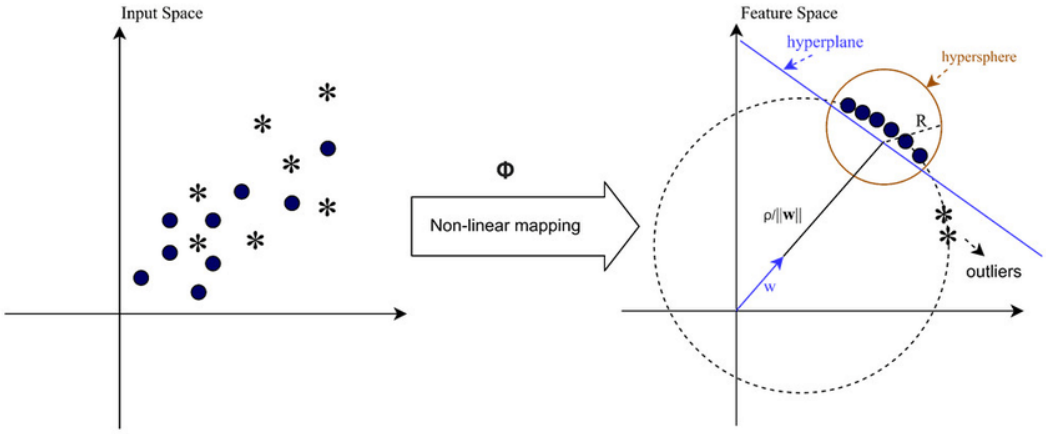

FIGURE 2 – Le concept sous-jacent de la méthode oc-SVM. Les points de l'espace original sont projetés dans un espace de dimension supérieure, où l'on cherche à les séparer du point d'origine en maximisant la marge. L'illustration est tirée de [Yengi et al., 2020].

L'oc-SVM vise à maximiser la marge par rapport à l'origine dans le nouvel espace de caractéristiques en résolvant le problème d'optimisation :

$$\begin{aligned} \min_{\omega,\rho,\xi_i} & \frac{1}{2}\|w\|^2 - \rho + \frac{1}{\nu n}\sum_{i=1}^{n}\xi_i \\ \text{subject to } & w \cdot \phi(x_i) \geq \rho - \xi_i, \\ & \xi_i \geq 0 \end{aligned} \tag{2}$$

Cette formulation incorpore des paramètres tels que la variable d'écart ξ_i, et ν qui régule la fraction des valeurs aberrantes et des vecteurs de soutien. La classification d'un échantillon test dépend de sa position par rapport à l'hyperplan, déterminée par :

$$f(x) = sgn(w^*\phi(x) - \rho^*) \tag{3}$$

Le paramètre ν dicte également l'équilibre entre les erreurs et les vecteurs de support, garantissant qu'une fraction spécifiée des exemples d'apprentissage peut être aberrante.

Pour l'application spécifique de ce pipeline, chaque voxel cérébral v_i est analysé à l'aide d'un classifieur oc-SVM avec le noyau RBF. Le noyau du classifieur est défini par :

$$\frac{|errors|}{n} \leq \nu \leq \frac{|errors| + SVs}{n} \tag{4}$$

où $\gamma = \frac{1}{\sigma^2}$ et ν est fixé à 0,03. Le classifieur oc-SVM de chaque voxel produit un score reflétant sa distance par rapport à l'hyperplan. L'agrégation de ces scores pour tous les voxels génère une carte de distances spécifique au patient D_p, facilitant la détection des valeurs aberrantes (anomalies) au sein de l'ensemble de données.

Description des données

L'efficacité du système CAD a été évaluée sur un groupe de patients épileptiques, en s'appuyant sur un ensemble de données acquises en collaboration avec le Dr Julien Jung des HCL et approuvées par un comité d'examen institutionnel (IRB). Ce projet est un sous-ensemble d'un programme de recherche clinique visant à explorer l'utilité de la neuro-imagerie multimodale dans les évaluations pré-chirurgicales des cas d'épilepsie difficiles.

Aperçu du groupe d'étude. L'étude comprend un ensemble d'entraînement de personnes saines et un ensemble de test de patients épileptiques. L'ensemble de données comprend 75 témoins sains, répartis en deux groupes (DB_{C1} et DB_{C2}), et 31 patients DB_{ep} souffrant d'une épilepsie médicalement réfractaire. Les caractéristiques détaillées des données sont présentées dans un tableau récapitulatif.

Contrôles sains. Ce groupe comprend des individus âgés de 20 à 66 ans, tous soumis à des IRM pondérées en T1 et FLAIR. Seuls 35 témoins ont également subi un examen TEP. *Groupe de patients.* Composé de 31 patients âgés de 17 à 47 ans, ce groupe a subi des évaluations préchirurgicales complètes, y compris des IRM pondérées en T1, FLAIR et TEP, ainsi que des examens EEG intracrâniens pour la localisation de la zone épileptogène. Les patients ont été sélectionnés sur la base de critères stricts afin de garantir la cohérence et la pertinence des données au cours des différentes phases du projet de recherche.

Acquisition de l'IRM et de la TEP. Les témoins et les patients ont été soumis à des protocoles standardisés d'IRM et de TEP utilisant l'équipement Siemens. Les

séquences d'IRM ont capturé des données anatomiques détaillées en 3D, tandis que les scans TEP ont été réalisés pour mesurer le métabolisme cérébral, en utilisant des algorithmes logiciels spécifiques pour la reconstruction d'images afin de garantir une résolution et une précision élevées.

Localisation des lésions cérébrales du patient. L'étude documente méticuleusement l'emplacement des lésions à l'origine de l'épilepsie, sur la base d'évaluations cliniques, d'examens EEG et de résultats chirurgicaux. La plupart des patients ont connu des résultats positifs après l'opération, conformément aux normes de la classification d'Engel. Pour chaque patient, des radiologues experts ont fourni des annotations manuelles de l'emplacement des lésions, qui ont servi de vérité terrain pour évaluer les performances du système CAD.

Prétraitement des données. Le prétraitement des images, effectué à l'aide du logiciel SPM12, a été une étape critique pour améliorer les performances du système CAD. Ce processus visait à améliorer la qualité de l'image, à réduire le bruit et à assurer l'uniformité entre toutes les modalités, facilitant ainsi une analyse et des résultats de détection précis.

Chapitre 6 : Formulation du problème

Ce chapitre présente nos considérations sur le problème de l'amélioration du modèle CAD existant pour la détection de l'épilepsie et donne une idée des choix stratégiques que nous avons dû faire et des solutions que nous avons proposées.

Le chapitre 7 présente les travaux sur la génération d'images synthétiques PET pour l'amélioration des modèles CAD dans la détection de l'épilepsie. En utilisant des réseaux adversoriels génératifs (GAN), la recherche se penche sur les complexités architecturales et la dynamique opérationnelle des modèles GAN, y compris les composants du générateur et du discriminateur, les fonctions de perte et les processus critiques d'ingestion et de prétraitement des données. L'accent est mis sur l'évaluation de la qualité des images synthétiques à l'aide de métriques méticuleusement sélectionnées, préparant le terrain pour une analyse expérimentale complète de l'entraînement des modèles GAN. Ce chapitre présente l'intégration de données synthétiques PET dans des ensembles de données existants, en soulignant les améliorations tangibles des résultats de l'apprentissage des modèles CAD. En comparant les modèles formés sur des ensembles de données enrichis d'images PET synthétiques à ceux utilisant des images PET originales, la recherche révèle une sensibilité accrue dans la détection de l'épilepsie. Ces résultats soulignent non seulement la faisabilité et la fiabilité de l'utilisation de données synthétiques pour l'entraînement des modèles de CAD, mais marquent également une avancée significative dans la résolution des problèmes posés par la rareté des images médicales annotées de haute qualité.

Le chapitre 8 de la thèse aborde le défi complexe de la fusion des données d'imagerie multimodale pour affiner la détection de l'épilepsie par les systèmes de CAD. Ce chapitre explore rigoureusement l'intégration des modalités d'imagerie T1, FLAIR et PET en proposant et en comparant trois stratégies de fusion distinctes : la fusion précoce au niveau du canal, la fusion intermédiaire par concaténation des vecteurs de caractéristiques et la fusion tardive par fusion des cartes de grappes. Grâce à une analyse comparative de ces stratégies, la recherche identifie la fusion tardive comme l'approche la plus efficace pour intégrer les données multimodales, améliorant ainsi la précision de détection du modèle CAD. La configuration et les résultats expérimentaux détaillés permettent de comprendre comment les différentes techniques de

fusion influencent les performances du modèle, en mettant particulièrement l'accent sur la sensibilité et la spécificité supérieures obtenues grâce à la fusion tardive.

Chapitre 7 : Apprendre avec des données synthétiques

Cette section décrit la génération et l'évaluation d'images synthétiques de TEP à l'aide de modèles GAN. Elle détaille l'entraînement de ces modèles sur un ensemble de données composé d'images T1 et TEP appariées, afin de produire des images TEP synthétiques pour des expériences ultérieures. La qualité de ces images synthétiques est évaluée à l'aide de mesures courantes, et une analyse hors distribution est effectuée pour mesurer leur similarité avec les images TEP réelles. La section explore également les implications de l'utilisation de données synthétiques dans l'analyse d'images médicales, en soulignant les avantages pour les modèles de détection d'anomalies non supervisés, tels que la génération rapide de données et la réduction des coûts. En outre, elle reconnaît les limites de l'approche actuelle et suggère des orientations de recherche futures, soulignant le potentiel des données synthétiques pour améliorer la détection des régions épileptiques dans l'imagerie cérébrale.

Modèles GAN pour la synthèse d'images

Les réseaux adversoriels génératifs (GAN) sont puissants dans la création d'images médicales, en particulier pour l'imagerie cérébrale, offrant une solution à la rareté des données TEP en générant des images réalistes. Cette capacité est cruciale pour l'entraînement et le test de modèles d'apprentissage automatique sans encourir de coûts élevés ou d'exposition aux radiations. La structure du GAN est présentée dans la figure 3.

Les GAN se composent d'un générateur (G) qui crée des images ressemblant à des données réelles, et d'un discriminateur (D) qui fait la distinction entre les images réelles et les images générées. L'entraînement implique que G produise des images évaluées par D, le retour d'information étant utilisé pour améliorer le réalisme de l'image.

GAN simple.

Une configuration GAN de base avec un générateur G_B et un discriminateur D_B est utilisée, en se concentrant sur la traduction d'images de la modalité T1 (y_A) en images PET (x_b). La formation optimise un problème min-max en utilisant les pertes du modèle GAN des moindres carrés (LSGAN) pour D_B et G_B :

$$L_{LSGAN}(D_B, A, B) = E_{p(x_b)}[D_B(x_b)^2] + E_{p(y_b)}[(D_B(y_b) - 1)^2] \tag{5}$$

$$L_{LSGAN}(G_B, A, B) = E_{p(x_b)}[(D_B(x_b) - 1)^2] \tag{6}$$

Une perte supplémentaire d'erreur quadratique moyenne (MSE) (L_{mse}) évalue la fidélité des images TEP générées (x_b) par rapport à leurs contreparties réelles (y_b) :

$$L_{mse}(G_B) = E_{p(x_b)}[(x_b - y_b)^2] \tag{7}$$

Cycle-GAN.

Ce modèle comprend deux générateurs et discriminateurs pour la traduction bidirectionnelle d'images entre les domaines A et B. Il incorpore les pertes LSGAN et introduit une perte de cohérence de cycle (L_{cyc}) pour préserver le contenu entre les images traduites :

$$L_{cyc}(G_A, G_B) = E_{p(y_a)}[|y'_a - y_a|1] + Ep(y_b)[|y'_b - y_b|_1] \tag{8}$$

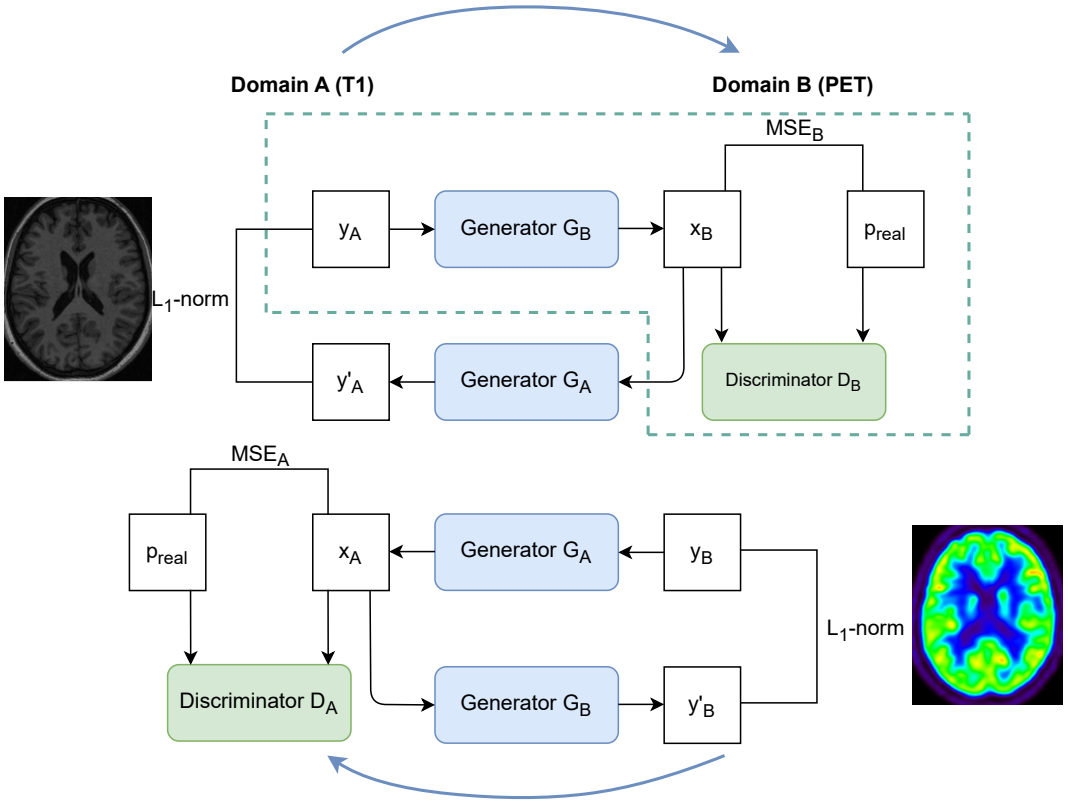

FIGURE 3 – Architecture Cycle-GAN basée sur deux GAN de base traduisant des images du domaine A au domaine B (GAN supérieur) et vice versa (GAN inférieur). Le GAN de base (Sipmle-GAN) est mis en évidence dans l'image par une ligne en pointillés.

Cycle-GAN garantit que l'image générée y'_a (de y_a à G_B puis G_A) reste fidèle à l'original, améliorant ainsi la précision de la traduction.

Vue d'ensemble du générateur.

Différentes architectures sont utilisées pour le générateur des GAN :

— U-Net : Apprend efficacement les correspondances d'image à image, en tirant parti des similitudes structurelles entre les entrées et les sorties. Largement utilisé pour la synthèse d'images médicales.

— ResNet : Surmonte les gradients qui s'évanouissent, en capturant des modèles complexes. Préféré pour son cadre d'apprentissage profond et résiduel, qui permet de générer des images détaillées.

Notre implémentation utilise un générateur basé sur ResNet. Il commence par un émargement pour préserver les dimensions, suivi d'une convolution, d'une normalisation par lots et d'une Leaky ReLU. Le noyau est constitué de blocs résiduels avec convolution, rembourrage, normalisation par lots et sauts de connexion pour faciliter l'apprentissage sans dégradation. En fonction de la taille de l'entrée, 9 blocs sont utilisés pour les entrées semi-3D et 2 pour les entrées 3D-patch. Le décodeur reflète le codeur et aboutit à une activation Tanh.

Vue d'ensemble du discriminateur.

Le discriminateur agit comme un classifieur binaire pour distinguer les images réelles des images générées, en se concentrant sur la qualité de l'échantillon pour l'amélioration du générateur. Contrairement aux GANs traditionnels, notre modèle utilise un discriminateur PatchGAN qui évalue des patchs d'images de taille fixe. Cette

approche améliore la fidélité des détails en se concentrant sur les zones locales. La structure du PatchGAN comprend cinq couches convolutives, une normalisation par lots et un Leaky ReLU, avec un champ réceptif de 70x70. Cette conception segmente efficacement l'image en parcelles pour une évaluation granulaire du vrai par rapport au faux.

Approche du traitement des données

Pour la modélisation des images cérébrales, nous utilisons deux stratégies afin d'équilibrer le contexte spatial et les exigences informatiques : Les modèles semi-3D utilisent trois coupes adjacentes (chacune étant un canal séparé) pour fournir des informations sur la profondeur. Ils évitent ainsi le chargement complet induit par les traitements de données 3D. Les modèles 3D-patch extraient de petites parcelles 3D à partir d'images plus grandes, ce qui permet une analyse locale détaillée avec des exigences de calcul gérables. Pour la semi-3D, les images sont normalisées et redimensionnées à 128x128x136 à partir de leurs dimensions originales avec des voxels isotropes. Les modèles de patchs 3D redimensionnent les images à 160x192x160 pour extraire des patchs de 32x32x32.

Reconstruction d'images

Semi-3D : Reconstruction par empilement de coupes transversales. 3D-patch : Utiliser les parties centrales des patchs pour éviter les prédictions de bord de faible précision, puis empiler pour former le volume. Les deux méthodes utilisent le lissage gaussien pour atténuer les effets de bord et la correspondance d'histogramme standard pour la normalisation de l'intensité.

Métriques d'évaluation de la qualité visuelle.

Les paramètres suivants ont été utilisés pour évaluer objectivement le score quantitatif des images traduites par rapport à l'image réelle de référence :

Mean squared error (MSE) estime l'erreur moyenne basée sur les pixels entre l'image générée (x) et l'image de référence (y) comme suit

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |x(i,j) - y(i,j)|^2 \tag{9}$$

Peak Signal to Noise Ratio (PSNR) calcule le rapport entre la puissance maximale possible (valeur d'intensité des pixels) de l'image générée y' et le MSE caractérisant la puissance du bruit corrupteur qui affecte la fidélité de sa représentation, comme suit

$$PSNR = 20log_{10}\left(\frac{max_x}{\sqrt{MSE}}\right) \tag{10}$$

où max_x est la valeur maximale possible des pixels de l'image x.

La métrique de l'indice de similarité structurelle (SSIM) de Z. Wang et al., 2004 estime la différence perceptive entre deux images en utilisant la moyenne (μ) et l'écart type (σ) sur les valeurs de pixel des images générées () et de l'image réelle (σ). Deux variables, $c_1(= 0,01L)^2$ et $c_2(= 0,03L)^2$, sont incluses pour stabiliser la division avec un faible dénominateur, L étant la plage dynamique des valeurs des pixels, comme suit

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1) \cdot (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) \cdot (\sigma_x^2 + \sigma_y^2 + c_2)} \tag{11}$$

Learned Perceptual Image Patch Similarity (LPIPS) de R. Zhang et al., 2018 est utilisée pour juger de la similarité perceptuelle entre deux images et il a été démontré qu'elle correspond à la perception humaine.

Il calcule la distance entre les cartes d'activation des images générées (x_0) et le patch original (x) pour un réseau prédéfini.

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{hw} \left\| w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw}) \right\|_2^2 \tag{12}$$

Où $\hat{y}^l_{hw}$ et $\hat{y}^l_{0hw}$ sont les piles de caractéristiques extraites de L couches normalisées dans la dimension du canal. Nous avons utilisé AlexNet comme réseau prédéfini.

Expériences avec des modèles GAN et résultats.

Dans la configuration expérimentale, nous avons utilisé deux approches :

— Modèles semi-3D : Quatre variantes de GAN (Simple-GAN et Cycle-GAN, avec et sans terme de perte MSE supplémentaire) ont été testées sur des exemples appariés afin d'évaluer leur performance dans la génération d'images synthétiques.

— Modèles 3D-patch : Compte tenu de l'impact positif observé lors de l'inclusion d'une perte MSE dans l'approche semi-3D, les expériences avec Simple-GAN et Cycle-GAN pour les modèles de patchs 3D ont incorporé la perte MSE dès le départ.

Ces expériences se sont appuyées sur la base de données DB_{C1}, comprenant 35 témoins avec des examens TEP FDG et IRM pondérés T1 appariés, évaluant les performances du modèle à l'aide des mesures de qualité visuelle établies précédemment.

Une validation croisée quadruple a été effectuée, divisant l'ensemble de données en 26 contrôles pour la formation et 9 pour la validation. L'indice de similarité structurelle (SSIM) entre les images réelles et synthétiques a servi de mesure principale pour optimiser les configurations pendant la phase de formation. Lors de la validation, les mesures SSIM ont été adaptées aux spécificités de chaque approche : pour les modèles semi 3D, le SSIM a été calculé après avoir redimensionné les tranches d'images à leurs dimensions d'origine ; pour les modèles de patchs 3D, il a été évalué au niveau du patch, puis une moyenne a été calculée pour tous les contrôles de l'ensemble de validation.

À l'issue des expériences de validation croisée, les deux configurations de modèle les plus efficaces ont été sélectionnées pour la suite de la formation. Ces modèles finaux ont ensuite été appliqués pour générer des images synthétiques de TEP pour la base de données DB_{C2}, qui ne contient pas de scans TEP originaux. Cette étape prépare le terrain pour des expériences ultérieures, impliquant potentiellement des modèles CAD pour l'épilepsie. Une analyse détaillée des procédures expérimentales, des ensembles de données et des résultats est présentée dans l'annexe.

L'étude a permis de générer avec succès des images synthétiques de TEP pour les configurations semi-3D et 3D à l'aide de modèles Cycle-GAN, démontrant ainsi l'efficacité des GAN dans la synthèse d'images médicales. L'application d'un lissage gaussien 3D aux images reconstruites afin d'atténuer les effets de bordure était une exigence notable, une largeur à mi-maximum de 1,5 mm étant considérée comme optimale pour améliorer les valeurs de l'indice de similarité structurelle (SSIM).

Les images synthétiques semblaient initialement plus lumineuses que les images TEP originales, ce qui a incité à utiliser la correspondance des histogrammes pour

aligner les intensités des pixels plus étroitement sur celles des images originales. Ce processus a permis d'adapter les images TEP synthétisées pour refléter la distribution d'intensité d'une image TEP originale sélectionnée au hasard dans l'ensemble de données, garantissant ainsi une ressemblance réaliste en termes de détails structurels et de valeurs d'intensité.

L'évaluation des performances de diverses configurations de GAN a révélé que le modèle Cycle-GAN à patchs 3D avec perte MSE donnait les meilleurs résultats en termes de SSIM, de rapport signal/bruit maximal (PSNR) et de similarité de patchs d'images perceptives apprises (LPIPS). L'inclusion de la perte MSE a amélioré de manière significative les trois mesures, démontrant ainsi son importance dans l'amélioration des performances du modèle.

Les expériences suivantes se sont concentrées sur les modèles les plus performants pour chaque configuration - modèles Cycle-GAN semi-3D et 3D avec perte MSE. Ces modèles ont ensuite été appliqués à la génération d'images TEP manquantes pour un ensemble de données distinct (DB_{C2}), ce qui a permis d'illustrer davantage leur utilité pratique. Le processus de formation a été soigneusement contrôlé afin d'éviter tout surajustement, et des ajustements ont été effectués en fonction de la progression de la perte et des mesures SSIM.

En fin de compte, l'étude souligne le potentiel des GAN, en particulier le Cycle-GAN avec perte MSE, dans la génération d'images PET synthétiques réalistes à partir de scans MRI, offrant des implications précieuses pour l'imagerie médicale et la recherche, en particulier dans les scénarios où les scans PET originaux ne sont pas disponibles.

Détection de rupture de distribution.

L'objectif principal de ces ensembles de données synthétiques est de reproduire fidèlement les images médicales réelles afin de garantir leur utilité dans la recherche et les applications cliniques. L'efficacité de ces données synthétiques est souvent évaluée à l'aide de mesures de qualité standard telles que le rapport signal-bruit maximal (PSNR) et la mesure de l'indice de similarité structurelle (SSIM). Bien que ces mesures permettent de comparer les caractéristiques de bruit et de texture des données de référence et des données synthétiques, elles ne peuvent pas déterminer si les images pseudo-vraies fonctionneront de manière similaire pour la tâche considérée.

Une façon d'évaluer si les données synthétiques suivent la même distribution que les données réelles de référence est de comparer les performances de la tâche clinique en question avec et sans données synthétiques

Pour évaluer si la distribution des données synthétiques de TEP générées correspond à celle des vraies données de TEP, nous dérivons des mesures de détection hors distribution (OOD) adaptées à la tâche spécifique de détection d'anomalies non supervisée en question.

La première mesure est définie comme MSE globale, qui est la moyenne de toutes les erreurs basées sur les pixels entre une image d'entrée u (qu'elle soit synthétique ou réelle) et sa reconstruction $AE(u)$.

La deuxième métrique OOD, inspirée de l'idée développée dans González et al., 2022, est dérivée du calcul de la distance de Mahalanobis d_m entre la représentation latente z_u de toute image d'entrée u dans l'autoencodeur AE et la distribution de cette variable latente dans les données normales de la population d'apprentissage, comme suit

$$d_M = (z_u - \mu)^T \Sigma^{-1} (z_u - \mu) \tag{13}$$

où z_u est la représentation latente de l'image u, μ et Σ sont respectivement la moyenne et la covariance empiriques, calculées sur la représentation latente z_i des N éléments de l'ensemble de données d'apprentissage comme suit :

$$\mu = \frac{1}{N}\sum_{i=1}^{N} z_i, \Sigma = \frac{1}{N}\sum_{i=1}^{N}(z_i - \mu)(z_i - \mu)^T \tag{14}$$

La métrique globale d_M quantifie si les fausses données TEP FDG imitent les vraies données TEP FGD dans la distribution (ID) *dans l'espace latent*, validant ainsi l'utilisation de fausses données TEP pour former les modèles UAD effectuant la tâche de détection dans l'espace de représentation latent.

La comparaison des distributions de MSE et de la distance de Mahalanobis d_M des vraies images y et des fausses images x permettra de détecter tout changement de domaine dans les espaces image et latent, respectivement.

Des taux d'erreur de reconstruction plus faibles ont été observés dans les groupes de contrôle avec des images synthétiques de TEP, l'amélioration la plus significative étant observée dans les images de TEP générées en semi-3D. Cela indique que les images synthétiques générées de cette manière sont très similaires aux images PET réelles.

La variabilité des distances de Mahalanobis entre les images TEP réelles et synthétiques était moindre. Le réglage du patch 3D a permis de différencier clairement les sujets témoins ayant des images TEP synthétiques de ceux ayant des images TEP réelles. En revanche, la frontière était moins nette dans la configuration sem-i3D, ce qui suggère que les images TEP générées en semi-3D ressemblent davantage aux images TEP réelles.

Dans l'ensemble, les images synthétiques de TEP s'avèrent très similaires aux données réelles de TEP, tant au niveau des représentations visuelles que latentes, ce qui justifie leur utilisation pour l'apprentissage des modèles UAD.

Un autre aspect de notre travail de recherche consiste à estimer la capacité des données synthétiques à être utilisées à la place des données réelles en cas de données manquantes. Dans cette partie, l'entrée de l'encodeur siamois est constituée d'images T1 et TEP FGD de sujets normaux combinées en tant que canaux. Nous nous appuyons sur les techniques de détection des OOD pour dériver des métriques évaluant si les distributions des images T1+vraie TEP et T1+TPE synthétiques générées se ressemblent à la fois dans le domaine de l'image et dans le domaine de la représentation latente du modèle UAD considéré.

All groups exhibited similar levels of reconstruction error, indicating that both in-distribution (ID) samples (test controls) and out-of-distribution (OOD) samples (real controls and patients) produce reconstructed images of comparable quality. Based on the global OOD metrics used, patients cannot be distinguished as outliers from the control distribution. This highlights the challenge of detecting epilepsy lesions, which are subtle enough not to significantly affect the global OOD metrics at the patient level.

Mesure de la qualité axée sur les tâches.

Notre hypothèse est que l'ajout d'images synthétiques à l'ensemble de formation peut servir à améliorer les performances des modèles de diagnostic basés sur

l'apprentissage automatique. Précédemment, nous avons décrit le développement et l'implémentation du modèle UAD. Ce modèle est entraîné sur trois bases de données différentes : la série de 35 données TEP de contrôle réelles (DB_{C1}) et deux bases de données hybrides mélangeant ces 35 données TEP de contrôle réelles avec 40 données TEP FDG de contrôle synthétiques générées par les modèles Cycle-GAN semi-3D (DB_{C2}^{semi3D}) et 3D-patch ($DB_{C2}^{3Dpatch}$) avec une perte MSE à partir des données T1 d'origine.

L'analyse révisée a révélé que l'ajout de données synthétiques à la formation, qui revient ici à doubler le nombre d'échantillons de formation, n'a pas permis d'améliorer les performances. La performance la plus élevée n'a montré qu'une sensibilité de 41,2% (7 détections sur 17) avec le modèle entraîné sur 35 scanners TEP réels.

La série d'expériences suivante a été réalisée sur des modèles UAD constitués de données T1+PET (originales ou synthétiques) combinées au début au niveau du canal. Le modèle $UAD_{mripet}^{3Dpatch}$ entraîné sur $DB_{C2}^{3Dpatch}$ a obtenu les meilleures performances de détection, avec une sensibilité de 74% (14 des 19 lésions) et un rang moyen de $2,1$, ce qui signifie que les lésions épileptiques détectées figuraient en moyenne parmi les trois groupes les plus suspects signalés par ce modèle. Le modèle UAD_{mripet}^{semi3D} entraîné sur DB_{C2}^{semi3D} a obtenu une sensibilité de 58% (11 lésions sur 19) et un rang moyen de 2.4, surpassant ainsi le modèle entraîné sur l'IRM T1 et la TEP FDG de DB_{C1}, dont la sensibilité et le rang moyen étaient respectivement de 42% (8 lésions sur 19) et de 3.9. Ces résultats sont comparables à ceux de l'analyse visuelle quantitative et montrent que les images obtenues à partir du modèle CycleGAN 3D-patch semblent être visuellement réalistes et adaptées à l'entraînement du modèle UAD. Les images obtenues à partir du modèle CycleGAN 3D-patch semblent être visuellement réalistes et adaptées à l'entraînement du modèle UAD.

Chapitre 8 : Fusion multimodale

Dans ce chapitre, nous examinons trois stratégies de fusion, à savoir la fusion précoce, la fusion intermédiaire et la fusion tardive, afin d'exploiter tout le potentiel des modalités disponibles pour les patients (T1, FLAIR et PET).

Fusion précoce

Nous utilisons la stratégie de fusion précoce en concaténant différentes modalités d'imagerie au niveau des canaux. Plus précisément, chaque échantillon d'entrée dans notre ensemble de données d'apprentissage $X = \{x_1; x_2; ...; x_N\}$ est un patch multicanal dérivé des modalités choisies. Pour les combinaisons T1+ FLAIR, T1+PET et FLAIR+PET, chaque patch est de taille 15x15x2, ce qui représente des dimensions spatiales de 15x15 et deux canaux correspondant aux modalités d'imagerie respectives. Dans le cas de la combinaison des trois modalités (T1+FLAIR+PET), la taille du patch devient 15x15x3, incorporant un canal supplémentaire pour la troisième modalité. Il est important de souligner que ces dimensions décrivent un seul patch. La représentation latente extraite par notre réseau pour chaque patch conserve une dimensionnalité de 16, quel que soit le nombre de modalités fusionnées dans l'entrée.

Fusion intermédiaire

Dans ce contexte, nous fusionnons deux ou trois modalités au niveau de la représentation cachée extraite. Après avoir obtenu les vecteurs de représentation latente

des réseaux siamois entraînés séparément, chacun sur une seule modalité, nous cherchons à fusionner ces vecteurs pour créer une représentation composite qui englobe les caractéristiques de chaque modalité. Cette fusion est réalisée par concaténation : pour les deux modalités, la fusion $z = [z_1; z_2]$ donne une dimensionnalité de 32 et $z = [z_1; z_2; z_3]$ pour l'intégration des modalités T1, FLAIR et PET avec une dimensionnalité de 48.

Fusion tardive

L'approche de fusion tardive utilisée dans cette étude est conçue pour améliorer la détection de l'épilepsie en intégrant des cartes de grappes générées par des modèles de détection d'anomalie non supervisée (UAD), chacun formé sur une modalité d'imagerie différente (T1, FLAIR et TEP). Ce processus comprend plusieurs étapes clés afin de créer une carte unifiée des grappes qui représente avec précision les anomalies significatives dans toutes les modalités :

— Création d'une carte unifiée des grappes : Les grappes provenant des différentes modalités sont fusionnées pour former une carte complète, capturant à la fois les points communs et les anomalies significatives spécifiques à chaque modalité.

— des grappes et attribution des rangs : Un système de classement évalue l'importance de chaque grappe. Lorsque des grappes de différentes modalités se chevauchent, le rang de la grappe combinée dans la carte unifiée reflète l'importance la plus élevée (rang numérique le plus bas) parmi ces grappes qui se chevauchent. Cela permet de s'assurer que les grappes les plus significatives conservent leur priorité dans la carte finale.

— Hiérarchisation des grappes : Les grappes résultant de la combinaison de différentes modalités sont plus significatives, ce qui souligne leur importance multimodale. Néanmoins, les grappes exceptionnellement significatives (avec des rangs comme 1 ou 2) conservent leur rang élevé, reconnaissant ainsi leurs contributions individuelles essentielles.

— Limitation du nombre de grappes : Pour que l'analyse reste ciblée, le nombre total de grappes dans la carte finale est plafonné à 15, ce qui garantit que seules les grappes les plus pertinentes et les plus significatives sont incluses.

Cette méthodologie de fusion tardive vise à tirer parti des atouts des différentes modalités d'imagerie tout en minimisant les faux positifs, améliorant ainsi la fiabilité de la détection des anomalies dans le cadre du diagnostic de l'épilepsie.

Résultats des expériences sur les stratégies de fusion.

Dans un premier temps, l'étude des données provenant d'une seule modalité a servi de référence. La modalité FLAIR s'est distinguée en identifiant le plus grand nombre de détections, 12 sur 26, bien que seulement 4 de ces détections aient été classées au plus haut niveau en termes de régions épileptiques probables. Les modalités T1 et PET ont permis de détecter l'épilepsie chez 8 patients chacune, mais seules quelques-unes de ces détections étaient de haut niveau. Il est intéressant de noter que les trois modalités n'ont pu identifier correctement les lésions que chez 4 patients, ce qui souligne la nature complémentaire des données d'imagerie, chaque modalité fournissant des informations uniques sur le processus de détection.

En ce qui concerne la stratégie de fusion précoce, on a observé que la combinaison des modalités TEP et FLAIR permettait d'obtenir le plus grand nombre de

détections, 4 patients présentant le rang de grappe le plus élevé. En revanche, la combinaison des modalités T1 et TEP s'est avérée moins efficace. Il est important de noter que la fusion des trois modalités à ce stade n'a pas amélioré les performances du modèle, probablement en raison des limites de l'architecture du réseau siamois dans la gestion de la complexité et de la haute dimensionnalité des modalités combinées.

La stratégie de fusion intermédiaire n'a pas répondu aux attentes. Les combinaisons T1+TEP et FLAIR+TEP ont permis de détecter respectivement 7 et 6 lésions, et seule une poignée de ces détections a reçu un rang élevé. Les défis de cette approche comprennent la capture des interactions non linéaires entre les modalités et la gestion de la dimensionnalité accrue du vecteur de caractéristiques, qui complique considérablement le processus de modélisation.

En revanche, la combinaison des modalités à un stade avancé a surpassé l'efficacité de la fusion précoce, en améliorant la confiance dans le modèle et la précision de la prise de décision. Cette stratégie a permis d'obtenir un total de 17 détections lors de l'intégration des images T1, FLAIR et TEP, avec 11 grappes recevant un classement élevé. La combinaison des images T1 et FLAIR, ainsi que FLAIR et PET, a donné lieu à 16 détections réussies, ce qui indique que la fusion tardive exploite efficacement les points forts de chaque modalité d'imagerie. En fin de compte, l'approche de fusion tardive, qui combine toutes les modalités disponibles, s'est révélée être la stratégie la plus efficace, soulignant l'importance de l'intégration stratégique des modalités pour améliorer la détection de l'épilepsie.

Conclusion et perspectives

Cette thèse a fait progresser les systèmes de CAD pour la détection de l'épilepsie à l'aide de données de neuro-imagerie, en s'appuyant sur le cadre établi par la recherche précédente. Dans un premier temps, nous avons passé en revue les systèmes de CAD existants, en soulignant leurs forces et leurs limites, et nous avons mis l'accent sur le potentiel de la génération de données synthétiques et de la fusion multimodale pour l'amélioration du système.

Notre étude a révélé que les réseaux adversaires génératifs (GAN), en particulier avec un générateur ResNet et une fonction de perte MSE supplémentaire, pouvaient produire des images TEP synthétiques de haute fidélité. Malgré leur qualité, l'application directe d'images synthétiques dans l'apprentissage du modèle a montré des améliorations limitées de la sensibilité de détection. Toutefois, des expériences intégrant des images synthétiques de TEP avec des données réelles d'imagerie par résonance magnétique (IRM) ont suggéré des améliorations modestes de la performance des modèles.

Nous avons ensuite exploré diverses stratégies de fusion pour combiner différentes modalités d'imagerie, concluant que les techniques de fusion tardives, qui capitalisent sur les détections les plus suspectes des modèles formés à une seule modalité, offraient les meilleures perspectives d'amélioration de la précision du diagnostic.

Orientations futures :

— Transformateurs dans les GAN pour la génération d'images : L'intégration de transformateurs dans le cadre des GAN pourrait améliorer de manière significative la génération d'images réalistes, ce qui permettrait de surmonter les défis posés par les ensembles de données limités.

— Raffinement des techniques de fusion : L'exploration de méthodes de fusion avancées, y compris les mécanismes basés sur l'attention, pourrait améliorer l'extraction des caractéristiques entre les modalités, améliorant ainsi les capacités de diagnostic du système de CAD.

— Disponibilité des données : L'augmentation de la diversité et du volume des données de neuro-imagerie peut considérablement améliorer la robustesse et la précision des modèles. Des projets de collaboration tels que MELD pourraient atténuer les difficultés liées à l'accessibilité des données.

— Intégration de sources de données multiples : La combinaison des modalités d'imagerie avec d'autres données diagnostiques, telles que l'électroencéphalographie (EEG), pourrait permettre une compréhension plus complète de l'épilepsie, ce qui pourrait conduire à des avancées significatives dans les stratégies de détection et de traitement.

Nos résultats soulignent le potentiel de l'intégration de l'imagerie multimodale et des techniques sophistiquées de synthèse des données dans l'amélioration des systèmes de CAD, ouvrant la voie à de futures innovations dans le diagnostic de l'épilepsie.

Acknowledgments

This work would not be possible without my great support network. First and foremost, my deep gratitude goes to my supervisor Carole Lartizien. You created a space where we could freely brainstorm the ideas and potential ways of improvements, I felt inspired after our discussions, especially when results were not as good as expected. Thank you for your patience, understanding and support.

Thanks to my colleagues from CREATIS laboratory, with some of you we have truly become friends sharing not only our sincere curiosity towards the world of medical imaging, but also many nights playing board games, sharing conversations while going to new restaurants and supporting each other through difficult times.

Thanks to my dear friends from abroad and from my motherland, my friends whom I met in France, my coco-caline family from Erasmus mundus program, you were one click away online in Zoom, always ready to listen to me and encourage me not to give up.

A big thank you goes to my podcast co-host (and now my colleague!) with whom we recorded hours of content exploring the possibilities of building the scientific career and discussing all the struggles of living abroad. And thank you dear listeners for your feedback.

At last, I want to thank my family, my parents who always tried to do their best to give me all the opportunities to pursue my dreams. Thank you for being with me and by my side.

Contents

List of Figures

List of Tables

Introduction

According to the World Health Organization (WHO), nearly 65 million people worldwide are affected by epilepsy. Epilepsy is a chronic neurological disorder characterized predominantly by recurrent and unpredictable interruptions of normal brain function, called epileptic seizures [Fisher et al., 2005]. Epilepsy varies in severity, in symptoms depending on the type of seizure and can effect people of all ages.

Epilepsy treatment involves consistent intake of antiepileptic drugs, and as a result, patients with epilepsy may lead a normal life. Despite the continuous improvement in drugs and research, approximately 35% of epilepsy patients experience recurrent spontaneous seizures [Duncan et al., 2006]. In such cases of drug resistant epilepsy the remaining option is to perform a surgery where the area in the brain that is no longer properly functioning and is generating seizures is removed. The success of such a surgery heavily depends on how accurate the epileptogenic zone is localized. Thus, as a pre-surgery step, it is recommended that a patient goes through a series of clinical assessments potentially including EEG, Magnetic Resonance Imaging (MRI), Positron Emission Tomography (PET), single-photon emission computed tomography (SPECT), and functional MRI (fMRI) [Yoganathan et al., 2023]. The availability of multimodal imaging techniques allows surgeons to better identify and characterizing areas of abnormal brain activity and structural anomalies associated with drug-resistant epilepsy. The challenge arises from the fact that sometimes epileptogenic lesions are too subtle in images, so their scans are considered normal. As an example, when the lesion is not visible on structural MRI imaging, we refer to such patient as MRI-negative patient. Specialists would gain immensely from an automated system designed to detect such small epileptogenic lesions using multimodal imaging. This system would integrate and analyze data from available imaging through machine learning models, and would aid in the accurate localization of seizure foci, providing surgeons with invaluable insights for planning precise resections while minimizing potential postoperative risks.

This work represents an attempt to improve a computer aided diagnosis system for epileptogenic lesion detection based on neuroimaging data introduced in Alaverdyan et al., 2020. The proposed CAD system consists of unsupervised deep siamese networks to learn normal brain representations using a series of non-pathological brain scans, followed by a set of one-class SVM models at voxel-level that produce the resulting anomaly score map. The model demonstrated 61% in sensitivity when it was trained on pairs of T1-weighted and FLAIR MRI exams. We would like to investigate the added value of incorporating PET modality into the model since it is another useful imaging modality available for patients. However, for healthy individuals, undergoing PET scans, which involves radiation exposure, is not advisable. The development of synthetic PET data in this case, mimicking those of healthy populations, emerges as a promising alternative.

Our first contribution consists in exploring strategies to generate synthetic PET images from T1 MRI. We showed that artificial PET scans can serve as a valuable addition to original PET and improve the detection abilities of the unsupervised

detection model. We further investigated if synthetic PET can be used as a substitution for real PET images in case of incomplete training set: we simulate in the experiments a situation where only T1 images for training set are available and we complement the model with synthetic PET (also referred to as fake PET) images to train the detection system. Our results demonstrate that generated PET images are indistinguishable from real ones, and play an essential role in tackling the missing modality challenge.

The second contribution consists in determining the optimal strategy to combine multiple imaging modalities in the unsupervised detection model originally proposed in Alaverdyan et al., 2020. We explore if integrating images at different stages - early, in the middle or late at the output level - significantly impacts the model's performance in detecting subtle epileptogenic lesions. Our key finding is that late fusion showed the highest number of true detections when we combined all available modalities, namely T1, FLAIR and PET (both real and synthetic images). By optimizing the fusion strategy, this research contributes to more accurate localization of epileptogenic zones, potentially leading to improved patient outcomes and tailored surgical interventions.

This work is divided into two main parts. Part I provides an examination of the recent advancements in deep learning applied to medical imaging, with a specific focus on brain image analysis presented in Chapter 1. The following chapter 2 focuses on existing fusion strategies in medical imaging, highlighting the strengths and weaknesses for every level of fusion methodology. Chapter 3 delves into a methodical overview of approaches for addressing missing data challenges, placing a strong emphasis on the generation of synthetic images. In 4 we describe elements of CAD systems designed for brain pathology detection, offering an exploration of epilepsy and the role of different imaging modalities contributing to the detection and understanding of epilepsy. We also give a review of existing CAD systems for epilepsy detection. We provide a problem formulation and proposed approaches in chapter 5. Finally, we present the clinical datasets that were used throughout this work as well as an overview of the baseline unsupervised detection model in chapter 6.

In Part II, we introduce the main contributions of this work. Chapter 7 describes the principal and main components of generative adversarial networks (GAN) that serve as basis for the synthesis of PET from T1 MRI images. It is followed by description of out-of-distribution (OOD) techniques we applied to ascertain whether synthetic data closely resemble real images. It also contains the details of the experimental part regarding the training of GAN models for the synthesis of PET images from T1 MRI, OOD evaluation and use of these generated synthetic PET images in the epilepsy CAD model. Chapter 8 presents the results obtained with the proposed CAD system on the combination of T1, FLAIR and PET data with 3 fusion strategies, discussion regarding the best performing strategy and the best approach to incorporate information from multiple imaging modalities. The manuscript ends with the key takeaways and offers directions for the potential future research work.

Part I

Medical and scientific context (State-of-the-art)

Chapter 1

Deep learning for medical imaging

Deep learning (DL) is a set of learning methods attempting to model data with complex architectures combining different non-linear transformations. Biologically-inspired neural networks enable computers to learn from observational data and are the key bricks of deep learning. They are composed of multiple layers to learn hidden representations of data with multiple levels of abstraction. Each of these layers performs non-linear transformations of input data before passing it on to another layer, and in this way, very complex functions are learned. We can, thus, formulate the goal of a neural network as to approximate some function f to map an input x to a category y: $y = f(x;\theta)$, where θ are the learnable parameters that result in the best function approximation (Figure 1.1). Typically, the network is represented by composing together many functions. The estimation of the parameters is obtained by minimizing a loss function on some training data with a gradient descent algorithm.

The simplest form of a neural network - perceptron - was designed to perform binary classifications. Its structure comprises a single layer of input nodes directly connected to an output node, which makes it a type of feedforward network. Further evolution of neural network architectures led to the development of the multilayer perceptron (MLP). The MLP introduces one or more hidden layers of nodes between the input and output layers, enabling it to learn nonlinear functions. While MLPs are powerful, they are not well-suited for tasks that require understanding the spatial relationships between data points, such as image recognition. This is where convolutional neural networks (CNNs) come into play. Their appearance has revolutionized the domain of computer vision and image processing. CNNs are able to successfully capture the spatial and temporal dependencies in an image through the application of relevant filters, they can detect features regardless of their spatial location, enhancing the robustness of solutions. Through the use of shared weights and pooling layers, CNNs significantly reduce the number of parameters that need to be trained lowering the computational cost. Compared to traditional machine learning algorithms, deep learning autonomously identifies meaningful representations without requiring the specialized knowledge of domain experts, thus enabling non-experts to effectively utilize deep learning techniques.

In past years, deep learning has been widely used in the medical domain, achieving remarkable performance in many medical imaging applications. Its success has mostly become possible thanks to the availability of big data for a particular task and the advances in computer hardware and software allowing computers to perform an increasing number of tasks that have not been possible before. DL excels at recognizing complex patterns in images and provides a quantitative assessment in an automated fashion. It has found its use in such tasks as:

— medical image registration [Fu et al., 2020]

— segmentation [Tajbakhsh et al., 2020]

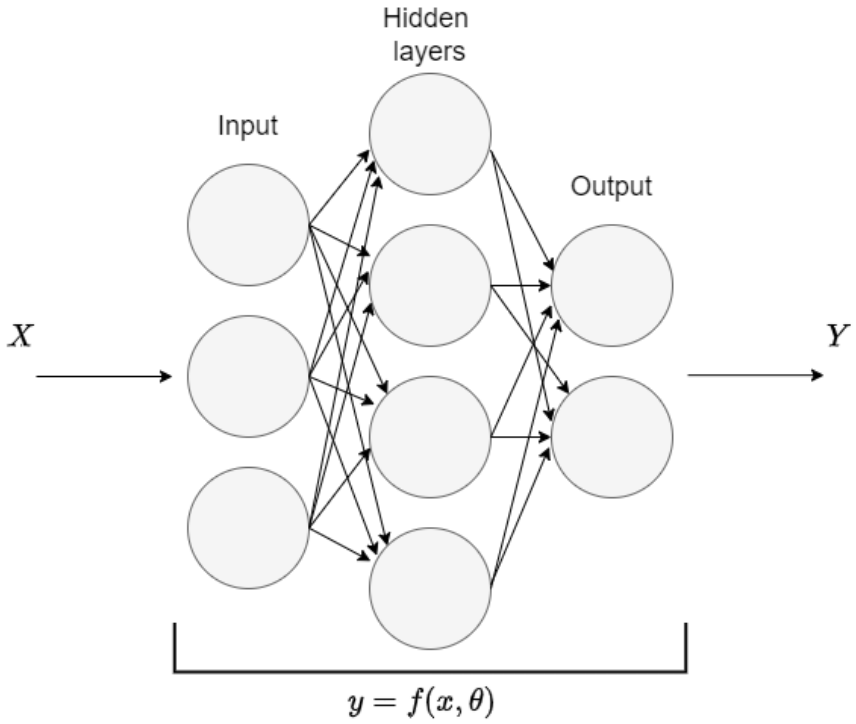

FIGURE 1.1 – General representation of a neural network.

— reconstruction [G. Wang et al., 2021]
— classification [C. Wang et al., 2021]
— generation [Islam et al., 2020]
— detection and decision support tools [Chan et al., 2020]
— and others [S. K. Zhou et al., 2021]

Integrating DL-based systems as an assisting tool for physicians into clinical workflows has started showing more accurate assessments [Alowais et al., 2023]. As an example, deep learning algorithms were shown to perform comparably with certified screening radiologists in detecting lesions in mammograms Kooi et al., 2017. In D. Li et al., 2021 it has been shown that deep learning model trained on multiparametric magnetic resonance imaging (mpMRI) and whole-mount histopathology data improve the diagnosis of prostate cancer in both junior and senior radiologists. A deep-learning algorithm has been applied in Venkadesh et al., 2021 for malignancy risk estimation of pulmonary nodules detected at low-dose screening CT. This DL algorithm demonstrated excellent performance, comparable to thoracic radiologists, with the potential to provide reliable and reproducible malignancy risk scores for clinicians.

Regarding brain disorders, the analysis of neuroimaging data has advanced significantly in the past years, further enhanced with the development of deep learning methods [Avberšek et al., 2022]. There has been extensive development of brain image analysis to devise imaging-based diagnostic and classification systems of strokes [Y. Yu et al., 2020], psychiatric disorders [Quaak et al., 2021], neurodegenerative disorders [Zaharchuk et al., 2018], demyelinating diseases [C. Huang et al., 2022]. Our work is dedicated to the problem of epilepsy detection, there has been a lot of attempts to apply deep learning techniques to develop a model able to accurately identify and predict epileptic seizures from various types of medical data [Shoeibi et al.,

2022]. Many challenges are associated with epileptic seizure detection, such as data availability, extensiveness and differences of seizure patterns. We mostly attempt to tackle two main problems in the context of epilepsy detection, but our findings and contributions may also be applied in other domains of medical imaging.

It is noteworthy that the majority of these solutions have been addressed with supervised learning approaches, meaning that the models were trained on a labeled dataset. Gathering annotated datasets, especially in medical domain poses a significant challenge due to privacy and ethical concerns, limited time of healthcare professionals, diverse data across multiple institutions etc. These challenges have motivated methodological developments:

1. in the domain of weakly or unsupervised deep learning
2. in generation of synthetic data to have better use of incomplete clinical datasets (i.e. with missing modalities)
3. to benefit from multiple imaging modalities.

Non fully-supervised methods

DL methods rely on good annotations, the systems we mentioned above exploit supervised approaches to solve particular tasks. Supervised learning, however, is not always possible to apply not only because of lack of well-annotated data, but in some cases, the training set is not sufficient to account for the complexity of the task. For example, this is a case in detecting brain pathologies when the lesions are subtle and vary largely in their shapes and textures. Thus, weakly supervised, semi-supervised and unsupervised methods come in handy.

Weakly supervised learning is a subset of machine learning techniques that lies between supervised and unsupervised learning. It involves training models on a dataset where the annotations are incomplete, inexact, or uncertain. This approach is particularly valuable in scenarios where obtaining fully labeled data is too costly, time-consuming, or challenging. Weakly supervised learning aims to leverage the available, albeit imperfect, annotations to learn meaningful patterns, structures, or relationships within the data [M. Liu et al., 2019, Ren et al., 2023].

Semi-supervised machine learning approaches are particularly useful when one has a large amount of input data but only a small portion of it is labeled. The key idea behind semi-supervised learning is to leverage the large volume of unlabeled data, in addition to the labeled data, to improve learning accuracy and model performance. This approach is based on the assumption that the distribution of unlabeled data can provide additional information that is beneficial for learning the structure of the dataset or for making more accurate predictions [Chebli et al., 2018, Rieu et al., 2021, Jiao et al., 2023]. In semi-supervised learning, both labeled and unlabeled data are used during the training process. The labeled data are used to guide the learning process with explicit instructions, while the unlabeled data are used to extract hidden structures or patterns that can help in improving learning accuracy. One of the main challenges in semi-supervised learning is ensuring that the unlabeled data actually contributes positively to the learning process. There is a risk that incorrect assumptions about the data distribution or incorrect pseudo-labels generated by the model could lead to decreased performance.

Unsupervised machine learning is a type of machine learning where algorithms are trained on datasets without labels [Raza et al., 2021]. Unsupervised learning focuses on identifying patterns, structures, or features within the data itself. It is often used for exploratory data analysis to find hidden patterns or grouping in data.

Synthetic data generation and usage.

Medical imaging has certain traits that influence the way we propose deep learning solutions. One of the issues is associated with annotations of medical images. The labels we get from physicians are sparse and noisy due to inconsistency, their acquisition is a time-consuming and expensive process. Therefore, the establishment of gold standards for image labeling remains an open question.

Deep learning models thrive on large volumes of data, which they use to learn the complex patterns and nuances inherent in medical images. Developing robust models requires a large amount of high-quality and diverse data. However, there are limited numbers of publicly available datasets, and most neuroimaging datasets have been limited to small-scale collections. Another limitation comes from the fact that patient data is highly sensitive and subject to strict privacy laws and regulations. Giving these challenges, the necessity for synthetic data in medical imaging emerges. Methods to supplement real medical data to overcome the hurdles associated with privacy, standardization, annotation, and access to data received increased interest and popularity. Adding synthetic data may produce good results when training neural networks, but a lack of trust towards these methods prevails in the community. Thus, in our work, we focus not only on generating synthetic images (namely, PET modality for healthy subjects) but also on methods that explore the nature of such data. As it has been show in a work of Skandarani et al., 2023, synthetic data might not always be suitable for training deep models even if they are nearly indistinguishable from real data.

In chapter 3 we provide an overview of latest advanced algorithms in the field of medical imaging generation, and show later in 7 how synthetic images can be used for an epilepsy detection task. We indeed show a case when adding generated images might not lead to improvements of the model, and in what scenario synthetic images demonstrate superior performance compared to a model trained solely on real images.

Usage of multiple imaging modalities.

Different medical imaging modalities have been used in clinical applications strengthening the the clinical precision, as different modalities bring new insights and highlight various aspects of pathological conditions. When it comes to epilepsy detection, combining structural information from MRI with functional insights from PET or EEG can provide a more accurate picture of the epileptogenic zone. The challenge lies in effectively processing and merging heterogeneous data, requiring sophisticated algorithms and models tailored to capture the complementary information each modality provides.

To improve the performance of deep models that can also deal with multiple imaging modalities, it is required to investigate the ways for efficient image fusion. This often leads to robust information processing that can reveal information that is otherwise invisible to the human eye [James et al., 2014].

Chapter 2 covers the latest research in multimodal medical image fusion techniques. We demonstrate later in 8 how we integrated MRI and PET images at different stages of the model, and discuss the best approach.

Chapter 2

Usage of multimodal images

2.1 Fusion of multimodality images (State-of-the-art)

Combining medical images from different modalities has been proven very efficient in improving diagnosis performance of expert clinicians. It has also shown a growing tendency for many tasks of medical image processing and analysis including segmentation, or the development of diagnosis or prognosis models. As exemplified in Figure 2.1, different modalities offer different information. MR images (here T1 and FLAIR images) provide detailed images of the organs and soft tissues within the body without radiation, Computed Tomography (CT) shows bony structures with high resolution, functional images, such as Positron Emission Tomography (PET) and Single-Photon Emission Computed Tomography (SPECT) can provide quantitative metabolic and functional information about diseases [T. Zhou, Ruan, et al., 2019]. Multi-modal images have found their usage in the training of neural networks. Extracting features from different modalities and bringing complementary information indeed leads to better data representation and thus better discriminative power of the neural models.

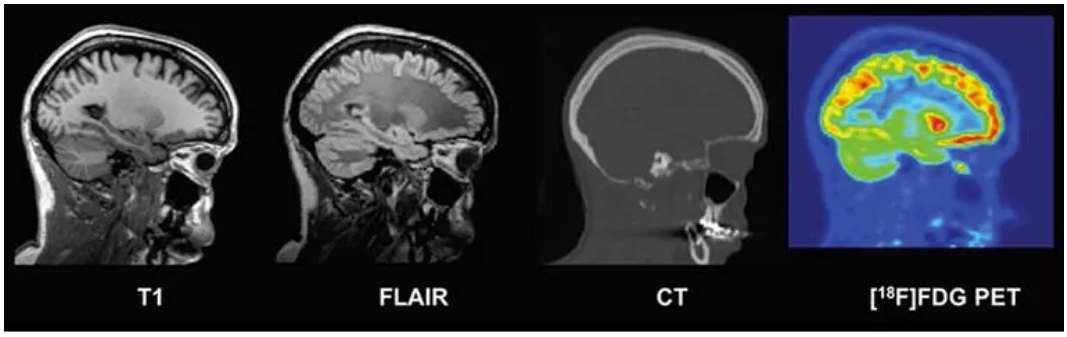

FIGURE 2.1 – Example of several modalities (sagittal plane) for one subject: coregistered T1 MRI, FLAIR MRI, CT and [18F] FDG PET images. Illustration from Mérida et al., 2021 licensed under CC BY-NC-ND 4.0 (https://creativecommons.org/licenses/by-nc-nd/4.0/).

In general, the fusion strategies of different imaging modalities or inputs are categorized into three groups: Early fusion (input-level fusion), intermediate fusion (layer-level fusion), and late fusion (decision-level fusion). Figure 2.2 depicts the mechanism of these different scenarios.

2.1.1 Early fusion

The first strategy consists of merging the modalities at the beginning so that the input to the model consists of their combined representations. The input can be the

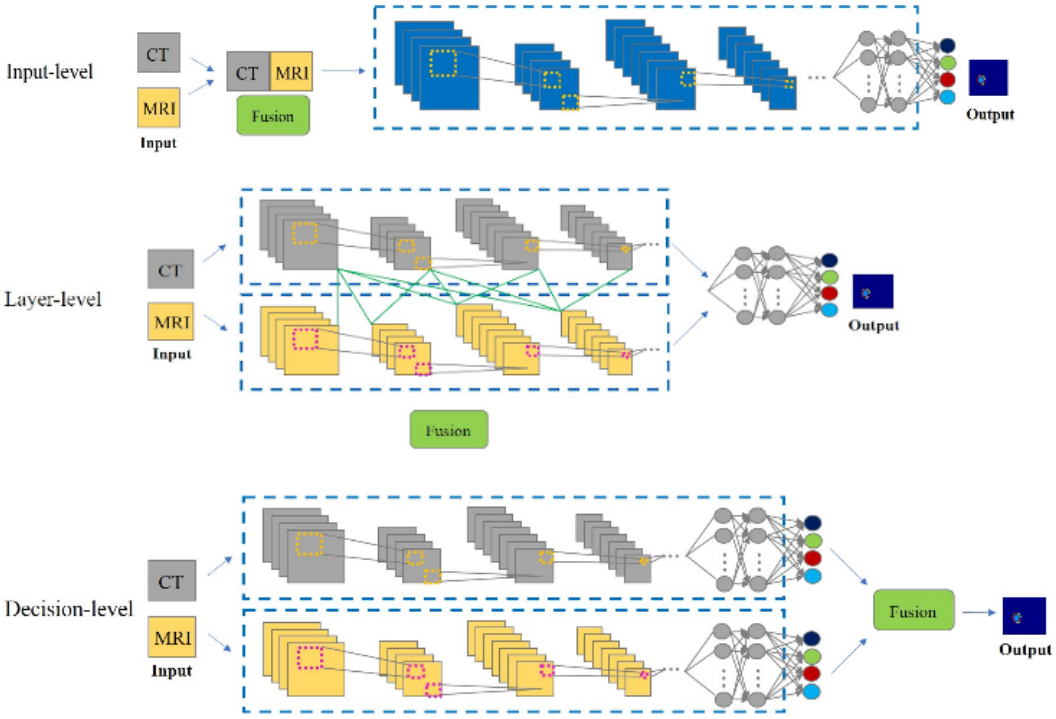

FIGURE 2.2 – The generic network architecture for 3 main fusion strategies. Illustration from T. Zhou, Ruan, et al., 2019 licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

raw imaging data or features relevant to each modality. The straightforward approach would be to concatenate the input from different modalities. The input-level fusion would be beneficial when the multimodal inputs are homogeneous. In this context, 'homogeneous' refers to the degree of compatibility and similarity in the characteristics of imaging data, homogeneity can be assessed in terms of spatial resolution, contrast patterns, and the specific anatomical or pathological information each modality emphasizes. Images are considered homogeneous if they present a unified representation where the features of interest are consistently enhanced across modalities. Non-homogeneous images may lead to misalignment of anatomical features across modalities, the model may struggle to extract relevant features due to the conflicting information presented by the heterogeneous inputs, potentially leading to poorer performance. It's essential to evaluate how similar the image characteristics are and whether any preprocessing steps are needed to harmonize the modalities before fusion. When addressing homogeneity, it's crucial to differentiate between homogeneity within a single modality and across different modalities: the lack of homogeneity between different modalities, such as between T1 and FLAIR images, is inherent and expected due to the different physical principles and acquisition parameters underlying each modality. This difference is not only acceptable but also advantageous, as each modality provides complementary information, however, homogeneity within a modality is essential for reliable analysis.

An additional benefit of early fusion is that only one learning model is trained as opposed to other approaches, reducing computational complexity and resource requirements. However, this approach might face limitations in the case of heterogeneous data.

Early fusion has been used in various medical applications. T1w, T2w, proton density-weighted (PDw), and FLAIR MRIs were joint as channels in Brosch

et al., 2016 for multiple sclerosis lesion segmentation. In a work of Thung et al., 2017, a challenging problem has been solved, namely, multi-class classification of Alzheimer's patients with incomplete data. Here, three prediction tasks associated with different combinations of imaging modalities were learnt jointly to improve the performance of each task. MRI and PET data were used from the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset where PET data have far less number of samples compared with MRI data. There are two sets of neurons in the input layer, each of which receives data from one modality. The output layer consists of three sets of neurons, which correspond to three different classification tasks - classification using only MRI data, classification using only PET data, and classification using MRI+PET data. The model uses categorical crossentropy as the loss function for each classification task and trains the network iteratively and subnet-wise, adjusting the network weights based on the modality data available for each specific task. This study demonstrates that the performance of the multi-task deep learning model is enhanced when combining the tasks over training a single model using limited T1+PET datasets with the improvement of classification accuracy from 62.4% to 65.8%.

Not only imaging data can be combined at the early stages, but other source data as text, handcrafted features, or 1D signals. Kiela et al., 2019 presented a supervised multimodal bitransformer (this type of architecture will be covered in the next sections) that jointly finetunes unimodally pretrained text and image encoders. MRI images were combined with patients' information in Bhagwat et al., 2018, or X-ray images with patient data in LIU, 2018. In Engemann et al., 2020 MRI and fMRI were combined with MEG for prediction of age. In Vaghari et al., 2022 MEG proved to improve the classification of mild cognitive impaired patients from normal controls when combined with MRI data.

2.1.2 Intermediate fusion

By concatenating image modalities at the input of the network, it is assumed that the relation between different modalities is simple, which is not always the case. In the intermediate fusion, we train different neural networks, each trained on one imaging modality, and join features from intermediate layers to create a new feature representation that is inputted to the final model at hand. The advantage of this approach is that only a single learning model is trained while the difficulty of combining heterogeneous inputs is diminished. Spasov et al., 2018 for the classification of Alzheimer's disease (AD) patients and Yala et al., 2019 for breast cancer risk prediction implemented convolutional neural networks to learn image features and fused these feature representations with clinical features before feeding them into a feed-forward neural network. In both works, simple concatenation was used to fuse imaging and clinical features. Reported results showed an improvement in performance using fusion compared to image-only models. A framework of three-stage deep neural network was proposed in T. Zhou, Thung, et al., 2019 for AD diagnosis using MRI, PET, and genetic data. Direct concatenation was not possible here since PET exam was missing for many patients, and also because images are continuous and low-dimensional while genetic data are discrete and high dimensional. First, latent representations (high-level features) for each modality were learnt independently in the first stage. Then, in the second stage, latent feature representations for each pair of modality combination (e.g., MRI and PET, MRI and SNP, PET and SNP, where SNP stands for single nucleotide polymorphism) were learnt and fed to the third stage that was trained to predict the diagnostic labels.

To better account for the complexity of multi-modal data, in Dolz et al., 2018, MRI data were combined in a middle fusion manner to solve a problem of intervertebral disc segmentation. Such an intermediate fusion strategy was shown to outperform the early fusion based on segmentation metric (Dice score).

A combination of early and intermediate fusion is presented in a work of L. Chen et al., 2018 in the context of brain tumor segmentation. Patches from FLAIR and T2 MRI are concatenated as the first input branch of a network, and T1 with T1-CE are combined together in a second branch. The first pathway performs a binary classification that segments the whole tumor from the background, this segmentation map is later combined with the second pathway (features extracted from T1 and post contrast T1-weighted images) to be fed into a 4-class softmax classifier (background, the necrotic and non-enhancing tumor (NCR/NET), peritumoral edema (ED) and GD-enhancing tumor (ET)), the final output from the model, thus, involves the hierarchical segmentation of different brain tumor regions. It was shown that the improved information flow extracts better features compared to traditional networks.

An interesting approach was proposed in W. Zhang et al., 2021 for MRI brain tumor segmentation. They proposed to use an autoencoder with 4 encoders (one encoder for each of 4 MRI modalities) to perform one-to-one feature extraction, and later merge the feature maps of the four modalities into one decoder. Their method demonstrated superior performance in the BraTS 2020 Challenge achieving high Dice scores comparable with state-of-the-art models.

2.1.3 Late fusion

For the last strategy called late fusion, each modality serves as an individual input for separate networks whose outputs are then merged to get the final results. Each network thus is designed to learn the complementary information from each modality independently. Usually, the final output is based on such aggregation functions as averaging, majority voting, weighted voting or a meta-classifier based on the predictions from each model. The choice usually depends on the nature of the task and is made empirically. Another option is to train an additional model on the decisions of the trained models for each modality to output the final decision.

In Nie et al., 2016, late-fusion strategy was applied in the context of tissue segmentation of infant brain. T1, T2, and fractional anisotropy (FA) images were used for training fully convolutional networks (FCN) individually for each modality and then features from the last layer were fused. Results were compared with those gained from a FCN model that combines 3 modalities as input and other commonly used segmentation methods with a single modality image as an input. The proposed multimodal late fusion FCN model was shown to perform better in terms of Dice ratio than the other unimodality models or the multimodality one based on early fusion.

A combination of early and late fusion was done in G. Wang et al., 2017, where MR T1, T1c, T2 and FLAIR image modalities (from BraTS 2017 dataset) were fused at the input level for brain tumor segmentation. Three networks (WNet, TNet, and ENet) were used to hierarchically and sequentially segment substructures of brain tumor, each network dealt with a binary segmentation problem. WNet was responsible for segmenting the whole tumor from the multi-modal MRI volumes. This includes the primary tumor mass along with any surrounding edema. TNet took the output of WNet (the whole tumor region) to focus on the tumor core, which excludes the edema and segments the more solid parts of the tumor. ENet further refined the segmentation by focusing on the enhancing tumor core, which is usually

indicative of the most aggressively growing tumor areas and is delineated using the output from TNet. Additionally, each of the networks was trained in axial, sagittal and coronal views to maximize the accuracy of the segmentation. During the testing time, predictions in these three views were fused to get the final segmentation. It has been shown that fusion helps to reduce the noise of the segmentation and improve segmentation accuracy. Kamnitsas et al., 2017 also applied the late fusion strategy to solve a problem of brain tumor segmentation on BraTS 2017 dataset. Three networks were trained separately to average the confidence of the individual networks. The final segmentation was obtained by assigning each voxel with the highest confidence. For the majority voting strategy, the final label of a voxel depends on the majority of the labels of the individual networks.

Qiu et al., 2018 trained three independent VGG-like networks that took as an input a single MRI slice (each from a specific anatomical location). Max, mean and majority voting were applied to aggregate predictions and perform the binary classification task between mild cognitive impairment and normal cognition. In the end, predictions from the three base models were combined by applying another majority voting to generate a final prediction of the multimodal fusion model. The fusion model showed improvements in performance when compared to models that used only one single modality. In the study by Reda et al., 2018, the methodology employed for diagnosing prostate cancer involved the use of prostate-specific antigen (PSA) screening results, alongside MR-based features, through a two-stage classification framework. Initially, a K-nearest neighbor classifier was utilized to convert PSA screening results into diagnostic probabilities. These probabilities were then combined with probabilities derived from MR-based features, specifically the cumulative distribution functions of apparent diffusion coefficients from diffusion-weighted MRI, as inputs. A stacked nonnegativity constraint sparse autoencoder was trained with these combined probabilities to classify the prostate volume as benign or malignant.

In Z. Guo et al., 2018, all three fusion schemes were investigated with the objective of enhancing the segmentation accuracy for soft tissue sarcoma. The performance of single modality networks based on PET, CT or T1 were way beyond results from the network that used fusion. Feature-level (input-level) fusion and classifier-level (layer-level) fusion both showed comparable performance, however, experiments demonstrated that bad training samples could result in unstable performance and decreased robustness for the early fusion strategy.

2.1.4 Fusion based on Attention Mechanism

Depending on the medical task, not all features extracted from the encoding part may be useful. In recent years, there has been research in finding effective ways to fuse features that are most informative. To this end, attention mechanisms have been introduced. It is mainly divided into three sub-categories: spatial attention model, channel attention model, spatial and channel hybrid attention model. The goal of the first type of attention mechanism is to identify the most significant spatial locations within the data that are crucial for the task at hand. Channel attention models define the importance of each feature channel and then either enhance or suppress it depending on the task. The integration of these two attention mechanisms allows learning the importance of each channel and the importance of each spatial localization in the image respectively.

Oktay et al., 2018 proposed an attention U-net, which uses the channel attention mechanism to fuse the high-level and low-level features for CT abdominal segmentation. Attention gates (AGs) in the decoding part of the U-Net filter the features propagated through the skip connections improving the model's sensitivity to foreground pixels, and the segmentation quality as a result. Mathematically, AG is defined as: for each pixel vector $x_i^l \in \mathbb{R}^{F_l}$ (where F_l is the number of feature maps in layer l), attention coefficients $\alpha_i \in [0, 1]$ identify salient image regions and the output from the AG can be seen as the element-wise multiplication of the input feature map and attention coefficients as $\widehat{x}_i^l = x_i^l \cdot \alpha_i$.

A three-stage segmentation network with an attention mechanism at a fusion level was proposed by T. Zhou et al., 2020. At first, four MRI imaging modalities (Flair, T1, T1c, T2) serve as input to a 3D-Unet model to get rough segmentations of brain tumor. After post-processing, this segmentation map serves as a context constraint for the following multi-encoder based fusion model. In this network, each imaging modality is encoded by a single encoder to obtain the individual latent representations, and the context constrain provides boundary information to refine the segmentation result, then the five encoders (one for the context constraint received after the first step and four encoders for each of the imaging modalities) are fused into the shared representation space with the fusion block. With the assistance of the attention mechanism, the feature representation is separated along channel-wise and space-wise, so that the most informative feature is obtained as the shared latent representation. Finally, the fused latent representation is decoded by the decoder to obtain the final segmentation result.

A correlation model (CM) was introduced in T. Zhou et al., 2021 to learn latent multi-source correlation representation from multiple MRI sources for brain tumor segmentation. Each input modality is encoded by an individual encoder to obtain the individual representation. The proposed correlation model and fusion block (as described above) project the individual representations into a fused representation, which is finally decoded to form the reconstructed images and the segmentation result. The correlation model consists of two parts: Model Parameter Estimation (MPE) and Linear Correlation Expression Module (LCE module). The MPE is a neural network with two fully connected layers and LeakyReLU that maps the modality-specific representation $f_i(X_i|\theta)$, where X_i is input modality and θ denotes the parameters of an encoder, to a set of independent parameters $\alpha_i, \beta_i, \gamma_i, \delta_i$, also individual for every modality. The correlation representation is then obtained with the help of the LCE Module as:

$$F_i(X_i|\theta_i) = \alpha_i \odot f_i(X_i|\theta_i) + \beta_i \odot f_k(X_k|\theta_k) + \gamma_i \odot f_m(X_m|\theta_m) + \delta_i, (i \neq j \neq k \neq m) \tag{2.1}$$

The correlation representations obtained from CM for every modality are then concatenated to form an input for a fusion block with an attention mechanism.

Ranjbarzadeh et al., 2021 proposed to use a novel Distance-Wise Attention (DWA) mechanism used for brain tumor segmentation based on multi-parametric MRI. The DWA module explores distance-wise dependencies in each slice of the four employed modalities for the selection of useful features.

J. Yu et al., 2021 proposed a CT and MRI image fusion method for healthy abdominal organ segmentation. Firstly, the network called VoVet effectively extracts features by utilizing four One-Shot Aggregation (OSA) modules, a residual branch in OSA and a channel attention module added to backbone to establish the interdependence between channels. VoVNet outputs multi-scale features that go to the

feature fusion (FF) module to get an aggregation feature (M) that integrates multi-scale context information. Finally, the mixed domain attention module is used after the FF module to enhance the interdependence of channels and the interdependence of location information and to produce the final segmentation.

Jiang et al., 2020 proposed a Max-Fusion U-Net for pathology segmentation given aligned multi-modal images. Modality-specific features are extracted by dedicated encoders and the attention mechanism. D. Li et al., 2022 presented multi-modality MRI fusion U-Net for cardiac pathology segmentation. Three dedicated encoders were used to extract independent specific modal features, and one fusion encoder with the channel attention to fuse specific modal information from the three independent encoders.

L. Xu et al., 2020 introduced a global spatial attention mechanism in CNNs for medical image classification. This attention mechanism is designed to differentiate between important and unimportant pixels across all images in a dataset, based on their intensities, using a binary classification approach. This enables the network to focus on areas within the images that are more relevant for making accurate predictions, thereby improving the performance of the CNN by enhancing feature selection during the learning process.

2.1.5 Conclusion

Choosing an effective fusion strategy for deep learning is still an important issue. Methodologically, each of them has its advantages and disadvantages, and rarely all three methods have been investigated for the same task on the same dataset. The early fusion strategy concatenating modalities to form the input space is prevailing, but it does not exploit the relationships among the different modalities. In contrast, for intermediate fusion, the connection among different layers can capture complex relationships between modalities. The late fusion strategy usually achieves better performance compared to the early fusion, especially for segmentation tasks [T. Zhou, Ruan, et al., 2019], with one single network learning independent feature representation from different modalities but at a cost of memory and computational time. In medical practice, it is not rare to face missing or incomplete data, when some patients have only clinical data available or lack or particular imaging modality. In such case, late fusion retains the ability to make predictions, since aggregation functions (majority voting or averaging) can be applied even in case of missing modality. Late fusion is favorable in this scenario, as it considers each modality separately. Attention-based mechanisms demonstrate improvements in model's predictions removing insignificant regions while suppressing extracted features when used at the intermediate levels or at the decoding layers to modulate the spatial focus.

Chapter 3

Learning with missing data

3.1 Ways of dealing with missing data

In medical practice, physicians usually use the information of different imaging modalities. However, certain modalities may be missing caused by various reasons from different protocols used at different time or institutions, patient inability to perform some exam or to institutions not providing data. The ability to overcome such issue would lead to more unbiased and statistically valid image analyses. Removing subjects with incomplete data will result in discarding a large amount of the acquired data and may lead to significant reduction of training samples.

The first strategy to deal with missing data is called imputation. Here missing values are replaced by some reconstructed values based on a mathematical assumption of the distribution of the original data. Such approach has been used to deal with missing parametric data. For example, a comparative study of different imputation techniques for filling missing records in the ADNI (Alzheimer's Disease Neuroimaging Initiative) data set on classification task has been conducted in Campos et al., 2015. The results showed that training classifiers with imputed data is better than constructing a predictive model with a reduced number of subjects with complete records. However, traditional imputation techniques are not suitable for large-scale high-dimensional datasets.

Dealing with missing imaging modalities can be grouped into three categories:

— Train a model on available modalities only

— Synthesize missing modalities and use a further complete set of modalities for model training and performing the final task

— Fuse available modalities in a latent space and learn a shared feature representation

Recently, several approaches have been proposed to synthesize missing modalities that we will consider further in this chapter. We start by reviewing some references of the last category of methods that perform fusion of the available modalities and imputation of the missing ones in a lower dimensional latent representation space. Exploiting latent feature representation may be beneficial as, in that case, we do not depend on the quality of synthetic imaging modality.

The state-of-the-art method has been proposed in Havaei et al., 2016 introduced the seminal Hetero-Modal Image Segmentation (HeMIS) architecture for brain tumor segmentation. In this method, each modality is initially processed by its own convolutional pipeline, independently of all others. The architecture for processing each modality is conceptually the same, ensuring that each modality's information is captured before fusion. Then, feature maps from all available modalities are merged by computing mapwise statistics. The mean and variance calculated across modalities provide a statistical summary of the information present across the different

inputs. This approach allows the network to effectively combine information from available modalities, compensating for any missing ones by leveraging the statistical properties (mean and variance) of the modalities that are present. In the end, the mean and variance feature maps are concatenated and fed into a final set of convolutional stages to obtain a segmentation. Any subset of available modalities can be provided as input, so that this method is robust to missing modalities.

Another algorithm for the same task of brain tumor segmentation has been proposed by Y. Zhu et al., 2021. They designed a brain tumor segmentation algorithm that is robust to the absence of any modality. Here each modality is also processed individually by its own encoder. The Cascade Supplement Module (CSM) is employed to augment the features of missing modalities by leveraging both simple operations and a cascading technique to generate diverse shared features. This module specifically addresses the challenge of missing data modalities in brain tumor segmentation by filling in these gaps. Following this, the Modality Fusion Module (MFM) takes over, where the enhanced features from CSM along with the original modalities are combined. The MFM utilizes a sophisticated selection mechanism to selectively integrate these features, thereby optimizing the process. Ultimately, this results in a refined fusion of features, significantly boosting the accuracy and robustness of the brain tumor segmentation outcome.

As stated above, the strategies to handle missing modalities include synthesizing the missing one or learning a modality-invariant feature space.

3.2 Synthesis of missing data

When acquisition is infeasible or limited, generating missing modalities instead of incurring an actual scan, can be beneficial. Various networks and architectures have been proposed in recent years for this task.

3.2.1 Variational Autoencoders

The first group of networks is based on Variational Autoencoders (VAEs). Originally proposed by Kingma et al., 2013, Variational Autoencoder (Figure 3.1 consists of an encoding and a decoding part just like the classical autoencoder, however, the fundamental difference is that the VAE learns the distributions of latent variables (variables that are part of the model, but which we don't observe, and are not part of the dataset) based on their mean values (μ_x) and variances (σ_x), thus providing the generative capability to the entire space.

Such model was used in a work of Pesteie et al., 2019 as an effective data augmentation technique to synthesize medical data. The effectiveness has been demonstrated on ultrasound images of the spine and MRI images of the brain.

3.2.2 U-net

U-Net was proposed in Ronneberger et al., 2015 for biomedical image segmentation. This artificial neural network uses the auto-encoder structure with skip connections. The advantage of such connections is that the spatial information lost during maxpooling operations in the encoding branch can be recovered. The encoder is designed to extract features from the images and the decoder is designed for upsampling and constructing the image segmentation by using those extracted spatial features. The architecture of the network is presented in Fig. 3.2. Most studies employing U-net follow the above architecture with some variants and improvements.

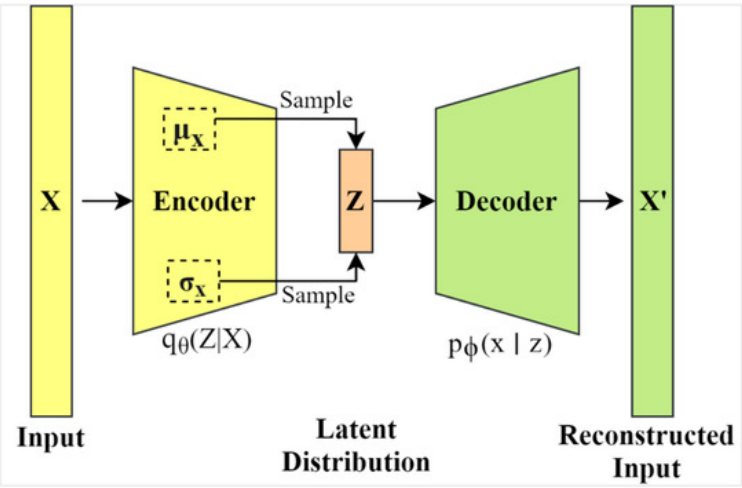

FIGURE 3.1 – The Variational Autoencoder architecture. Here, the approximation function $q_\theta(Z|X)$ is the probabilistic encoder, and the conditional probability $p_\phi(x|Z)$ is a decoder.

One of the first attempts to apply U-net-like architecture for image synthesis was by X. Han, 2017. Synthetic CT scans were derived from MR images, and their quality was evaluated on a set of brain tumor patient images for which real CT scans existed. U-net has become an inspiration for the encoding part of the deep convolutional network for synthesis of MR images in a work of Chartsias et al., 2017.

In Sikka et al., 2018, it was shown that 3D convolutional U-Net managed to produce realistic PET from T1 MRI data, that were used further for multimodal Alzheimer's disease classification (using only MRI gave 70.18% of accuracy while joint classification using synthetic PET and MRI resulted in 74.43%).

In Kalantar et al., 2021, U-net was used as a baseline to synthesize T1-weighted MRI images from pelvic CT scans. The results showed that despite U-net generated relatively realistic predictions for pelvic slices consisting of fixed and bony structures, the synthetic T1 MRI appeared to be blurry and locally unrealistic for deformable pelvic structures. The other type of neural networks, namely generative adversarial networks (GAN) has shown better performance for this task, and we are going to describe it in the next subsection.

3.2.3 Generative adversarial networks

In recent years, Generative Adversarial Networks (GANs) have become increasingly popular due to their powerful ability to produce photorealistic images. The GAN architecture was first described in Goodfellow et al., 2014. Its basic structure 3.3 consists of the generator that we train to generate new examples, and the discriminator that tries to distinguish examples being real or fake. Two models are trained simultaneously and compete against each other.

To train a GAN, the following optimization problem that the discriminator is trying to maximize and the generator is trying to minimize must be solved:

$$\min_G \max_D V(D,G) = \mathop{\mathbb{E}}_{\mathbf{x}\sim p_{\text{data}}} [\log D(\mathbf{x})] + \mathop{\mathbb{E}}_{\mathbf{z}\sim p_z} [\log(1 - D(G(\mathbf{z})))] \tag{3.1}$$

Following the main idea of GAN networks, J.-Y. Zhu et al., 2017 proposed a new approach to translate images from a source domain to a target domain in the absence

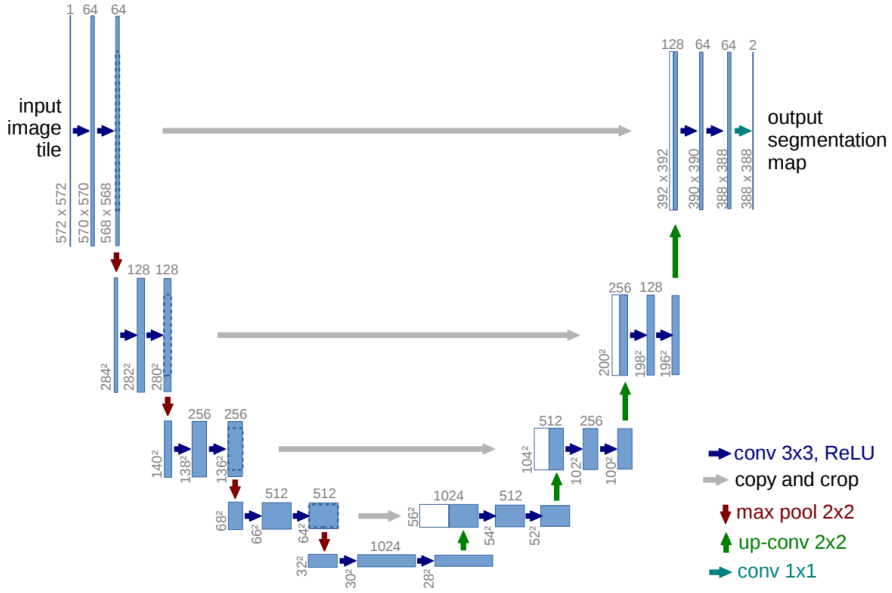

FIGURE 3.2 – The U-Net architecture

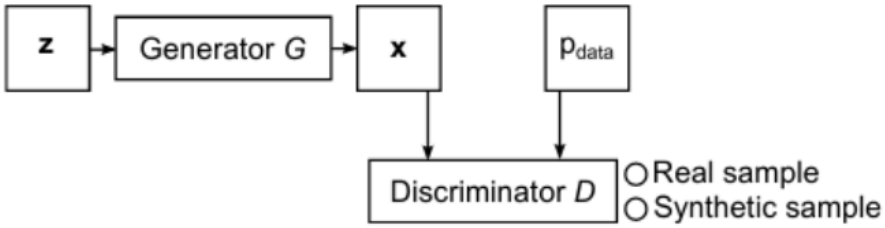

FIGURE 3.3 – The basic GAN scheme: z is a vector samples from a distribution p_z, the generator G takes z as input and transforms this vector into a sample x, the discriminator D tries to distinguish generated samples from samples from the real distribution p_data.

of paired examples, and it is called CycleGAN (figure 3.4) which is based on the combination of adversarial losses 3.1 and cycle-consistency loss 3.2:

$$L_{cyc}(G_A, G_B) = \mathbb{E}y_a \sim p_{real}[\| y'_a - y_a \|_1] + \mathbb{E}y_b \sim p_{real}[\| y'_b - y_b \|_1], \tag{3.2}$$

where $y'_a = G_A(G_B(y_a))$ and $y'_b = G_B(G_A(y_b))$

The key element of a CycleGAN is a cycle consistency that assumes that an image y_A from domain A can be translated to a domain B and back, thus the original y_A image should be similar to a reconstructed image y'_A. Two cycles are trained simultaneously, one from domain A to domain B and back, and one from domain B to domain A and back.

In H. Yang et al., 2018, CycleGAN was used for brain MR-to-CT synthesis using unpaired data. In addition to an adversarial loss and a cycle-consistency loss, a structure-consistency loss has been proposed. The structure-consistency loss is designed to ensure structural consistency between synthetic and input images by mapping these images into a common feature domain using a modality-independent structural feature, specifically the Modality Independent Neighbourhood Descriptor (MIND). This loss enforces the extracted MIND features in the synthetic image to be

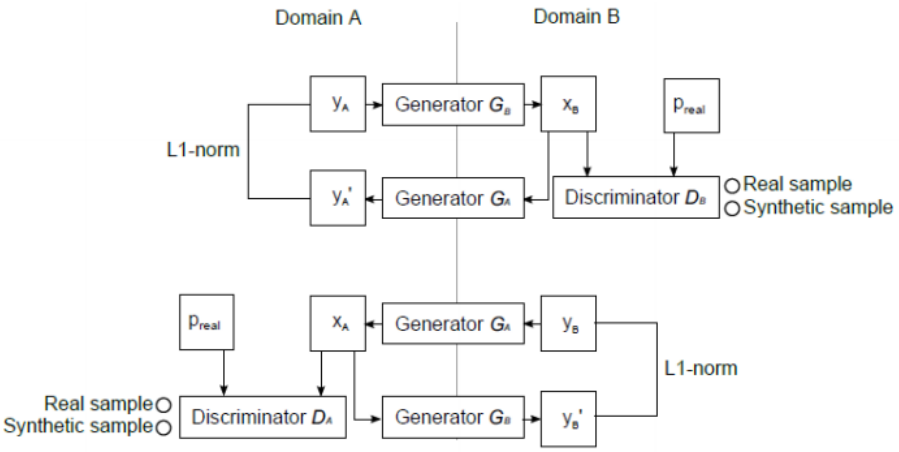

FIGURE 3.4 – CycleGAN consists of two generators (G_a, G_b) and two discriminators (D_a, D_b). The generators translate images from domain A to domain B and vice versa. The discriminators try to distinguish between real and synthetic samples in each domain. An additional L1-losses determine whether samples are consistently recovered after cyclic translation.

voxel-wise close to those extracted in the input image. It is calculated using the L1 norm across all voxels in the non-local region around each voxel, thus measuring the difference between the MIND features of synthetic and input images and ensuring that these features are closely aligned. The model generated better synthetic CT images than the conventional CycleGAN and produced results similar to a CycleGAN trained with paired data.

CycleGAN has been adapted for medical applications in Armanious et al., 2019 with the inclusion of two new non-adversarial losses: the cycle-perceptual loss aims to minimize perceptual discrepancies and enhance the global consistency of the output images by calculating the mean absolute error (MAE) between feature maps extracted from the input and the cycle-reconstructed images using a pre-trained feature extractor network; the cycle-style loss, on the other hand, focuses on matching the texture style and fine details between the input and cycle-reconstructed images. It is computed by calculating the feature map correlations over the depth dimension, represented by the Gram matrices, and then calculating the weighted average of the squared Frobenius norm of these matrices (here, The Gram matrix is used to measure the similarity between all pairs of vectors in a given set of vectors. It represents the inner products between all pairs, capturing the angles and magnitudes of these vectors in relation to each other. The Frobenius norm is a measure used to quantify the size of a matrix, calculated as the square root of the sum of the absolute squares of its elements). Together, these losses guide the generator architectures to produce outputs with minimized perceptual and textural discrepancies, leading to more detailed and higher quality translated images. Cycle-MedGAN framework has been used in a context of PET-CT translation and MR motion correction. For the quantitative comparison, metrics such as Peak Signal to Noise Ratio (PSNR), Structural Similarity Index (SSIM), Learned Perceptual Image Patch Similarity (LPIPS) and Visual Information Fidelity (VIF) were used with outperforming results for Cycle-MedGAN compared to Cycle-GAN (corresponding metrics are 0.9115 vs 0.8963, 24.08 vs 23.35, 0.4275 vs 0.3831, 0.2233 vs 0.2561).

A Sketcher-Refiner GAN model was proposed in Wei et al., 2019 to get PET-derived demyelination from multimodal brain MRI. While the sketcher part generates an image with global anatomical information, the refiner one pays more attention to lesional areas, thus allowing to better learn complex relationship between the two domains. This study also demonstrated better performance when including more MR modalities as inputs.

A 3D approach for PET synthesis from T1 MRI was proposed in Yaakub et al., 2019 with a GAN model where the generator is based on a residual U-Net and the discriminator is a convolutional network similar to what has been used in previous works but in 3D. The model was trained on patches of size 32x32x32. Generated PET images were compared to those produced by U-Net and high-resolution dilated CNN. 3D GAN model was shown to outperform the two other models based on mean absolute error (MAE) and PSNR metrics. The generated PET images served for the identification of hypometabolism in epilepsy patients. Real PET images from patients were subtracted from pseudo-normal PET images produced by the proposed GAN model to detect regions of hypometabolism. This approach was shown to outperform standard statistical parametric mapping (SPM) showing high sensitivity in MRI-positive and MRI-negative patients.

A 3D auto-context-based locality adaptive multi-modality generative adversarial networks model (LA-GANs) was proposed in Y. Wang, Zhou, et al., 2018 to synthesize high-quality FDG PET image from low-dose PET (L-PET) with the accompanying MRI images, namely T1-MRI and diffusion tensor image (DTI), that provide anatomical information. Instead of treating each image modality as an input channel and apply the same kernel to convolve the whole image, the authors proposed a new mechanism to fuse multi-modality information so that the weight of each imaging modality can vary with image locations for better serving the synthesis of full-dose PET (F-PET). The whole pipeline looks as following: the locality-adaptive fusion network takes an L-PET, a T1-MRI, an FA-DTI and an MD-DTI images as inputs, and generates a fused image by learning different convolutional kernels at different image locations. Then the generator is trained to produce a synthetic F-PET from the fused image, while the discriminator subsequently takes a pair of images to distinguish between the real and synthetic pairs. In as second phase, the synthetic F-PET images generated from the LA-GANs for all training samples are used as context information, together with the original images of each modality to train a new auto-context LA-GANs model, which further refines the synthesized F-PET image. The training is done on image patches of size 64x64x64. Compared to a model trained on L-PET only, the authors show that employing L-PET together with T1 allows achieving slighlty better performance, with PSNR improved from 24.29 to 24.58 and the SSIM increased from 0.982 to 0.985, respectively, for normal controls.

In Flaus et al., 2023, the authors trained a U-Net-based adversarial network to generate synthetic pseudo-normal FDG PET images from MR T1 images of control subjects. In a second phase, they used this generative model to generate control-like pseudo-FDG PET images from MR T1 exams of epilepsy patients, which were then subtracted from the patient's true FDG PET images to localize hypometabolic regions. This method was shown to outperform standard statistical parametric mapping (SPM) analysis.

As a conclusion, GAN-based architectures have been extensively used for medical image synthesis with good performance. One potential limit is that GANs are made from CNNs, where convolutions operate in a fixed-sized window with a locally receptive field and thus cannot capture long-range dependencies and second, convolutional operation cannot flexibly adopt for different inputs because of fixed

weights of the convolution filter after training. Transformers have been proposed as an interesting alternative to CNN to remedy to these limitations. The following section relates recent state of the art transformer-based model for medical image synthesis.

3.2.4 Transformers

The models we have seen so far are based on convolutional neural networks that include an encoding and a decoding parts. Another approach for transformers that are based solely on attention mechanism is proposed by Vaswani et al., 2017. They have become dominant in the field of natural language processing (NLP), but their demonstrated exemplary performance led to a great interest for computer vision applications [Khan et al., 2021].

Transformers are based on a self-attention mechanism that learns the relationships between elements of a sequence. Dosovitskiy et al., 2020 kept the key elements and principals of a transformer, but applied it for image classification task. In that case, an input sequence becomes a set of image patches. The structure of such a Vision Transformer (ViT) is given in Figure 3.5.

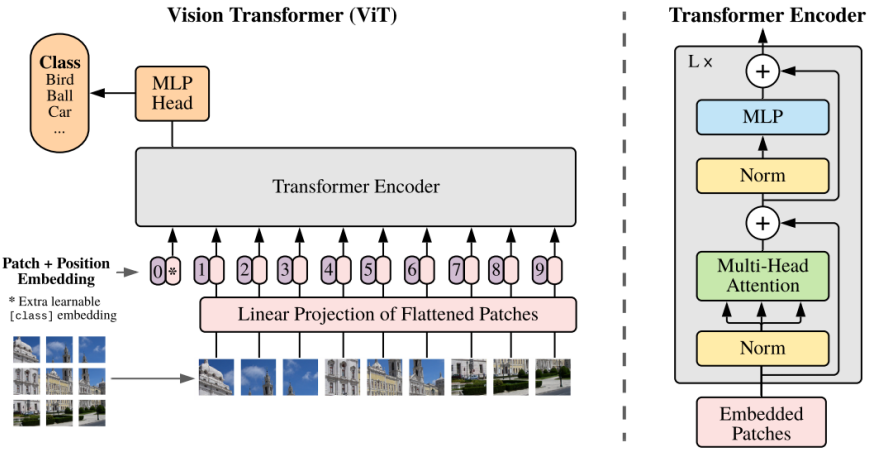

FIGURE 3.5 – Overview of a vision transformer.

The key element of the transformer's encoder is a multi-head attention that consists of several attention mechanisms (called "Scaled Dot-Product Attention" in the original paper) running in parallel. The basic scheme of such a concept is represented in Figure [3.3]. Given a sequence of items (or image patches) self-attention estimates the relevance of one item to all other items through a set of mathematical operations. Let's define the input sequence as $X \in R^{n \times d}$, where n is a total number of items and d is the embedding dimension that represents each item. The input sequence X is first projected onto three vectors, namely, Keys (K), Queries (Q) and Values (V) through corresponding learnable weights matrices W^K, W^Q, W^V. The matrix of outputs is computed as:

$$Attention(Q, K, V) = softmax(\frac{QK^T}{\sqrt{d_k}})V \tag{3.3}$$

The self-attention computes the dot-product of the query with all keys, followed by the normalization using softmax operator to get the attention scores.

Multi-head attention allows the model to attend to information from different representation subspaces at different positions. It is calculated as:

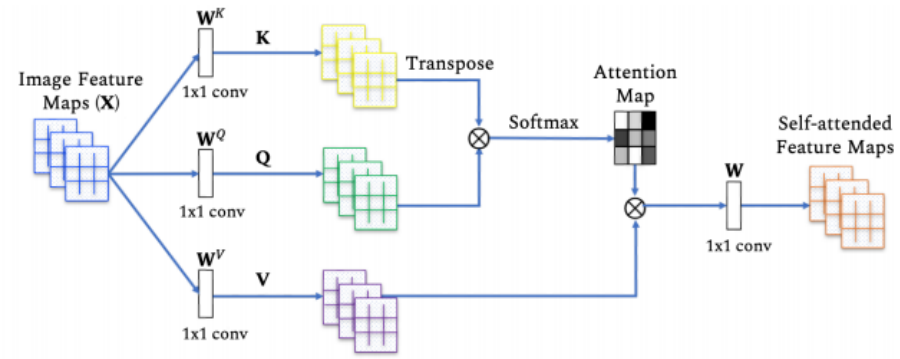

FIGURE 3.6 – Self-attention mechanism in the ViT. Given the input sequence, the vectors of Keys, Queries and Values are calculated followed by attention calculation and applying it to reweight the values. A single head is shown here and an output projection (W) is applied to get output features of the same dimension as the input.

$$MultiHead(Q, K, V) = Concat(head_1, ..., head_h)W, \tag{3.4}$$

where $head_i = Attention(QW_i^Q, KW_i^K, VW_i^V)$

Watanabe et al., 2021 proposed a model called Generative Image Transformer (GIT), focused on generating and transforming SPECT images characteristic of Parkinson's disease. The input consists of 40 superior slices of 3D volume SPECT images, and the output consists of the 51 sequentially generated rest of the inferior slices, including transformed images from healthy to Parkinson's disease-like characteristics. GIT utilizes a transformer model architecture based on transformer decoder blocks. It includes pre-layer normalization, multi-head attention, residual connections, and position-wise feed-forward phases, with a total of 16-layer transformer decoder blocks. Attention blocks are located within the transformer decoder layers, where they are used for capturing dependencies without regard to their distance in the input sequence.

Vision transformers are highly promising for the goal of image synthesis since attention operators learn contextual features that should improve sensitivity for long-range interactions. Recent studies consider hybrid architectures or computation-efficient attention operators to adopt transformers in medical imaging tasks. In Luo et al., 2021, the objective of the research is to reconstruct standard-dose PET (SPET) image from low dose PET (LPET) thus obtaining clinically acceptable PET images while reducing the radiation exposure. The proposed solution called MEaTransGAN exploits advantages of both GANs and Transformer networks: CNN-based GANs can describe the local spatial features, and transformers are good at capturing the long-range semantic information. In this work, the generator is designed as a CNN-based encoder for the compact feature representation extraction, followed by a transformer encoder (TransEncoder) to capture the long-range dependencies, and a CNN-based decoder for restoring the reconstructed SPET image. Quantitative analysis demonstrates the MEaTransGAN model's significant advancement over existing methods, including the Transformer-GAN, in terms of PSNR and NMSE metrics. This highlights the effectiveness of incorporating anatomical information from T1-MRI into PET reconstruction, which is reflected in the improved PSNR from 24.818 to 25.426 for NC subjects and from 25.249 to 26.041 for MCI subjects, respectively

In X. Zhang et al., 2021, a novel MRI synthesis framework was proposed, namely, Pyramid Transformer Net (PTNet) to synthesize T1w scans using good-quality T2w

scans . PTNet's architecture consists of transformer layers, skip-connections, and a multiscale pyramid representation. It mimics the classical U-Net structure and inherit the skip connection, but CNN-based encoder/decoder is replaced by so-called performer-based encoder/decoder. Performer is an attention-based architecture similar to transformer but with a simplified self-attention model, thus requiring less computation than a standard transformer. The proposed model was compared with pix2pix and pix2pixHD (the conditional GAN for high-resolution image synthesis) for the task of generating T1w scans from T2w scans. SSIM and PSNR were calculated on test dataset and showed superiority of the PTNet model.

Dalmaz et al., 2022 proposed a generative adversarial approach called ResViT for multi-modal medical image synthesis. ResViT is an adversarial model with a hybrid CNN-transformer architecture as generator and a conditional PatchGAN as discriminator. The generator in ResViT is based on an encoder-decoder architecture with a central information bottleneck in the middle. The encoder and decoder comprise convolutional layers to maintain local precision and inductive bias in learned structural representations, while the information bottleneck comprises a stack of novel aggregated residual transformer (ART) blocks. ART blocks organised as the cascade of a transformer module learn contextual representations, and synergistically fuse CNN-based local and transformer-based global representations. ResViT was applied for synthesizing missing sequences in multicontrast MRI, and CT images from MRI and was shown to outperform several state-of-the-art convolutional and transformer models in PSNR and SSIM metrics.

Another hybrid architecture is used in Shin et al., 2020. The authors built a generative adversarial network by utilizing the Bidirectional Encoder Representations from Transformers (BERT) architecture, namely GANBERT, to generate PET images from MRI images. A U-Net like architecture first generates PET from T1-MRI input in 3D (the input image has a size of 256x256x256). Next, BERT acts as the GAN discriminator trying to predict if the next sequence (here, images are "summarized" to text-like sequences) is "real" or "generated" PET. Quantitative evaluation demonstrated superiority of a GANBERT model over a pix2pix-GAN [Isola et al., 2017] showing higher values for PSNR, SSIM and Root-Mean-Square Error (RSME) metrics.

Chapter 4

CAD systems for brain pathology detection

Computer aided detection (CAD) is a technology designed to decrease observational oversights, and consequently the false negative rates of physicians interpreting medical images. CAD systems could assist physicians in many ways, from providing quantified image metrics and calculating probabilities of diagnoses to detecting and segmenting abnormalities. This section delves into the fundamental principles underlying CAD systems.

Initially, we explore the general principles of CAD systems, key components, and the role they play in automating the detection and interpretation of brain pathologies. Next, our focus shifts to the heart of CAD system efficacy: the ways of the models are trained. Supervised models, with their ability to learn and predict based on labeled datasets, offer a direct route to identifying known patterns of pathology. Conversely, unsupervised models, which discern structures and relationships in unlabeled data, present opportunities for uncovering novel insights and unknown patterns within complex imaging datasets. We then dive deeper into the latest CAD systems for epilepsy starting with the description of epilepsy as a disease and ending with a detailed literature review of the state of the art CAD systems for epilepsy detection.

4.1 Overview of a CAD system

A typical CAD system includes the following steps: image pre-processing, feature extraction and statistical model design (using machine learning or a deep learning algorithm).

Image pre-processing

Before inputting images into a CAD model, they are usually processed first to make models robust. Common pre-processing steps in medical imaging may include denoising, bias field correction, registration, normalization and standardization, etc. [Masoudi et al., 2021].

Feature extraction

A feature extraction is a process of dimensionality reduction of the region of interest (ROI) for analyzing images. It includes modifying the image from the lower level of pixel data into higher level representations. From these higher level representations, we can gather useful information while effectively reducing the amount of data. The features may be learned automatically by using deep learning architectures for a specific task.

Statistical model design

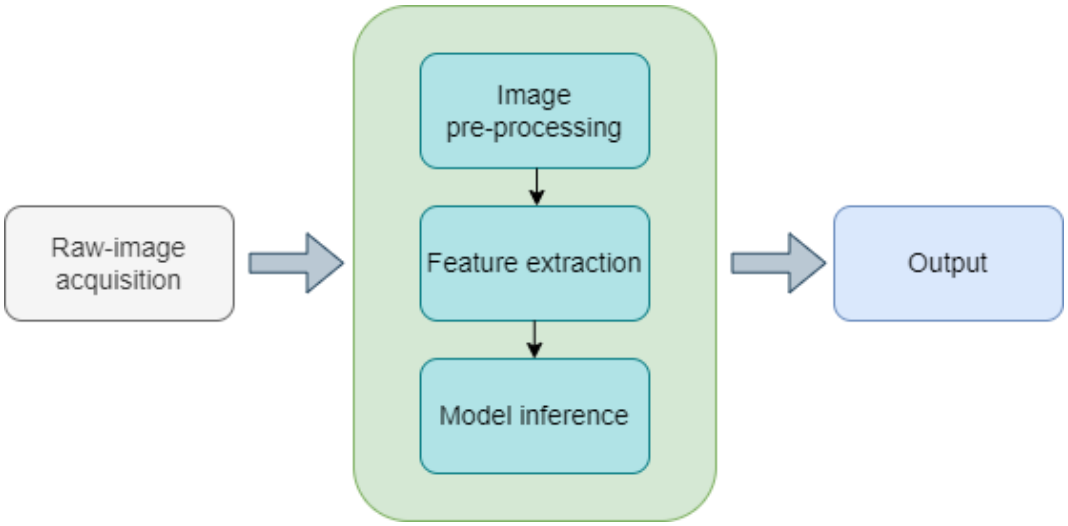

FIGURE 4.1 – Schema of a general CAD-system pipeline.

Statistical model design refers to the development of data-driven models, which are essentially statistical models tuned on a training dataset. This tuning process involves adjusting the models' hyperparameters based on an objective function, a loss function, that evaluates the model's performance. The essence of this approach is to create a decision function that processes an input feature vector and outputs a decision variable, effectively making predictions based on the provided data.

The selection of a specific class of models is inherently task-dependent (regression, classification, etc.). The nature of the available data significantly influences this choice. The data may be fully labeled, partially labeled, or entirely unlabeled, corresponding to supervised, semi-supervised, or unsupervised learning paradigms, respectively. These paradigms dictate the type of loss function to be minimized during the hyperparameter tuning process.

By selecting the model and tailoring the loss function to the specific task and data at hand, it is possible to design a statistical model that optimally addresses the problem, ensuring that the hyperparameters are tuned to enhance model performance and decision-making accuracy.

CAD systems for the analysis of brain neuroimaging data are desirable as a "second opinion" to assist radiologists in interpreting images. In the next sections, we will observe the main trends in developing such CAD systems depending on the type of model training. In neuroimaging, many different medical problems can be addressed with deep learning methods. One of the most common tasks is segmentation of the brain's anatomical structures, for analyzing brain changes, measuring and visualizing, for delineating pathological regions. Brain Tumor Segmentation (BRATS) challenge [Baid et al., 2021] has become a popular competition, increasing the dataset and expanding the scope of proposed problems every year. Among the winning solutions, we can find an encoder-decoder architecture with a dynamic scale attention mechanism [Yuan, 2020], nn-Unet based [Luu et al., 2022], 3D-Unet-like architectures in Demoustier et al., 2022, and with additional attention mechanism in Akbar et al., 2022.

Segmentation tasks take place when precise discrimination of tissue types is needed. In this work, we focus on a detection problem where pathological lesions can be very subtle and may not be easily identified and outlined by experts. In

the next sections, we go through reviews of important works concerning detection problems in neuroimaging.

4.2 Supervised approaches for brain lesion detection

In supervised learning, the model is trained using a labeled dataset. We consider a training set to be composed of pairs $(x_1, y_1), ..., (x_n, y_n)$, where $x_n \in \mathbb{R}^k = X$ are feature vectors and $y_n \in Y$ the corresponding label classes. Then we can introduce the function $f(x, \theta)$ that maps the input to the output, such as $X \rightarrow Y$, with θ being a set of parameters that best approximates the function's output. Typical tasks that are solved with supervised deep learning methods in medical image processing are segmentation, classification, detection. Below we observe several applications of supervised DL for brain pathology which are commonly used for clinical purposes.

Multiple sclerosis (MS) is a chronic autoimmune, inflammatory neurological disease. The diagnosis of MS is rather complex, but today's clinical practice includes brain MRI scans of different modalities (T1, FLAIR, etc.) to visualize and detect lesions. MS lesions can appear with varying characteristics, they vary greatly in size, ranging from a few millimeters to several centimeters in diameter. Early in the disease, lesions may be more subtle and harder to detect, while in later stages, they can be more pronounced and easier to identify. Detection of MS lesions is important for the optimal treatment, and one of the ways is to segment lesions.

Numerous datasets have been made publicly available for researchers and clinicians. The most noticeable ones are: ISBI challenge dataset [Carass et al., 2017] includes MRI data from multiple time points for each subject, accompanied by expert annotations of white matter lesions associated with MS, the dataset comprises training data from five subjects (average of 4.4 time-points per subject) and test data from fourteen subjects (also averaging 4.4 time-points per subject); the Multiple Sclerosis Segmentation (MSSEG) Challenge dataset [Commowick et al., 2018] contains MRI data from 53 MS cases sourced from four different centers, the dataset is divided into a training set of 15 patients, available for algorithm development, and a testing set of 38 patients, reserved for evaluation, each case was annotated manually by seven different experts; MSSEG-2 [Commowick et al., 2021] contains 100 new MS patients (40 patients for training, 60 patients for testing), lesions are manually annotated by four expert neuroradiologists, followed by a consensus formation through senior expert review and majority voting.

A number of automatic segmentation techniques have been proposed for tissue segmentation in MS. The first group of methods focuses only on lesion segmentation [Roy et al., 2018 where they used the public ISBI challenge dataset as well the private one, Salem et al., 2021 with experiments conducted on private dataset, Alijamaat et al., 2021 focused their research on the MSSEG dataset]. However, volumes of gray and white matter and cerebrospinal fluid are also affected by the MS, robust estimation of tissue volumes is necessarily as well. Thus, the second group of methods focuses on both tasks - brain and MS lesion segmentation [Gabr et al., 2020 tested their algorithm on public dataset, McKinley et al., 2021 used MSSEG dataset and the private one for the experimental part].

Measuring myelin content can potentially allow multiple sclerosis to be detected earlier. Wei et al., 2019, Wei et al., 2020 tracked the progression of demyelination in MS patients with the help of [^{11}C]PIB PET-derived images from the Sketcher-Refiner GAN and Conditional flexible self-attention GAN respectively. Another approach

has been implemented with GAN-based model and multimodal images for MS segmentation in C. Zhang et al., 2018: the generator consists of two encoding paths for T1 and FLAIR modalities and one decoding path that led to the MS segmentation mask, and it tries to produce a mask as close to the ground truth as possible fulling the discriminator.

The cerebral small vessel disease (CSVD) is another brain pathology associated with abnormalities related to small blood vessels in the brain. It can cause such consequences as dementia or stroke. On MRI, it can be seen in a form of lacunes, white matter hyperintensities (WMH), small subcortical infarcts, prominent perivascular spaces, cerebral microbleeds, and atrophy [Wardlaw et al., 2013]. Several deep learning models for CSVD detection have been applied [Hsieh et al., 2019, Duan et al., 2020, Shan et al., 2021]. Hsieh et al., 2019 utilized a dataset from Taipei Medical University-Shuang Ho Hospital, comprising MRI images of 50 middle-aged stroke patients, Duan et al., 2020 trained their model on a diverse collection of MRI scans from 1,500 patients, further tested on 30 patients selected at random, encompassing T1-weighted, T2*, DWI, and FLAIR sequences, meanwhile, Shan et al., 2021 analyzed FLAIR imaging data from 1,156 patients diagnosed with CSVD-associated WMH, collected from Beijing Tiantan Hospital over a year.

Intracranial carotid artery calcification (ICAC) is one of the atherosclerotic plaque features, that can be an easily identified on computed tomography scans. ICAC is associated with ischemic stroke [Bos et al., 2014] and cognitive decline, and linked to an increased risk of dementia [Bos et al., 2015], therefore, its detection and assessment is important in clinical practice. Bortsova et al., 2021 used ensemble networks on noncontrast CT scans to produce probability maps representing network confidence in classifying pixels as ICAC. Ultrasound images of the carotid artery were used to train and test the Capsulenet model in Lai et al., 2022 to classify image as normal or abnormal.

The last brain pathology we would like to cover in this section is Parkinson's disease (PD). Early detection and monitoring of the disease leads to improvement in the life of the patients. MRI scans of the brain are widely used for this purpose. Chakraborty et al., 2020 focused on the detection of PD patients as a binary classification problem of MRI scans based on a 3D convolutional neural networks. In Bhan et al., 2021, a 2-class classification task was modeled with the LeNet-5 CNN architecture to distinguish between healthy controls and PD subjects. A hybrid model combining numerical symptoms assessments and MRI images was proposed in S. Zhu, 2022 showing that the model can successfully diagnose and classify Parkinson's disease patients into five categories based on the severity of the disease.

4.3 Unsupervised approaches

The deep architectures described in the previous subsection achieve impressive the state-of-the art results, however, they require large annotated data sets for training. The nature of brain pathologies is highly variable, and well-annotated representative data sets may not be available. Subtle pathological brain regions can be heterogeneous and are more easily described as abnormalities, i.e. defined by their deviation from the characteristics of healthy tissue, than by specific features. To this

end, unsupervised learning approaches come in handy, where input data are not labeled, and no known result is given to the model to train. In this part, we will go through the same brain pathologies and tasks observed in the previous section, but solved with unsupervised deep learning methods.

A variety of works are based on using variations of autoencoders since they are not using labeled data, and their goal is usually to minimize reconstruction error based on a loss function. The key principle is to learn the healthy anatomy through representation learning. If we have a set of normal subjects $X \in \mathcal{R}^{D \times H \times W}$, then the autoencoder learns to project it to and recover it from a lower dimensional space $Z \in \mathcal{R}^{K}$. Such models assume that the autoencoder trained on normal data only will not be able to reconstruct anomalies contained in the patient images. At inference, the anomaly map is obtained by computing the error between the original data and the pseudo-normal data reconstructed by the autoencoder.

The state-of-the-art results of such architectures are currently achieved by variational autoencoders (VAE) [Baur, Denner, et al., 2021]. A comprehensive study by Hassanaly et al., 2023 evaluates the efficacy of seventeen VAE-based approaches in identifying anomalies in 3D Brain FDG PET images, specifically targeting abnormalities associated with Alzheimer's disease and other dementias. VAEs have been used in Baur et al., 2018, Vogelsanger et al., 2021 for MS lesion detection in brain MRI, in Chatterjee et al., 2022, Zimmerer et al., 2019 for tumor detection, in You et al., 2019 for brain tumor and stroke detection. In a work of Baur, Wiestler, Muehlau, et al., 2021 it has been shown that the unsupervised method based on the autoencoder and trained on normal subjects performed similarly to the supervised U-Net.

Yoo et al., 2018 used an unsupervised model for detecting multiple sclerosis pathology on normal-appearing brain tissues using a latent hierarchical myelin-T1w feature representation. It consists of two modality-specific deep belief networks, one for myelin features and the other for T1 features, which are fed into a joint network that learns multimodal features. Gained features are later used to train a random forest that distinguishes each patch being MS patients or a normal subject.

A patch-based unsupervised pipeline was proposed in Muñoz-Ramırez et al., 2021 to detect Parkinson's disease. For this purpose, a comparative study has been conducted to define what works better, a spatial auto-encoder or a patch-fed siamese auto-encoder (SAE).

GANs also found their use in solving anomaly detection tasks.

C. Han et al., 2021 used a medical anomaly detection GAN to reconstruct multiple adjacent brain MRI slices to detect brain anomalies at different stages on multi-sequence structural MRI.

Simarro Viana et al., 2020 used a 3D GAN that detects and localizes traumatic brain injury abnormalities in non-contrast CT images.

An abnormal-to-normal translation generative adversarial network (ANT-GAN) was proposed by Sun et al., 2020. It generates a normal looking MRI brain images based on their abnormal-looking counterpart without the need for paired training data, so in a corresponding image the lesion is "removed".

4.4 CAD systems for epilepsy

4.4.1 Description of epilepsy

According to the World Health Organization (WHO), epilepsy is a chronic disease of the brain that affects people of all ages. Around 65 million people are affected worldwide, which makes epilepsy one of the most common, chronic and serious neurological diseases [Thurman et al., 2011]. Epilepsy is characterized by enduring epileptic seizures and by neurobiological, cognitive, psychological and social consequences of the condition. Overall, about 70% of patients achieve seizure freedom with an appropriate treatment [Moshé et al., 2015], which usually involves taking antiepileptic drugs on a long-term basis. The remaining part of 30% who do not respond to the initial antiepileptic drug is referred as drug-resistant patients. As many as one-third of patients will have a refractory form of disease indicating the need for a neurosurgical evaluation, and resection would be more effective for medically intractable epilepsy (MIE) than anti-epileptic drugs (AED) treatment alone [Ramey et al., 2013].

Temporal Lobe Epilepsy

Temporal Lobe Epilepsy (TLE) is the most common form of MIE and comprises about 80% of epilepsy surgeries [Ramey et al., 2013]. The surgical treatment of TLE has been the most widely practiced and researched. According to Cendes et al., 2014, hippocampal sclerosis (HS) is considered the most frequent histopathology encountered in patients with TLE. The International League Against Epilepsy (ILAE) subdivides HS into three types. The hippocampus is located in the medial temporal lobe and is formed by the interlocking neuronal bands of the dentate gyrus and the hippocampus proper (also known as the "cornu ammonis" = CA)[Malmgren et al., 2012]. The CA areas are all filled with densely packed pyramidal cells having four subfields (CA1–4) (Fig. 4.3). Around 20% of TLE cases do not show significant neuronal cell loss with only reactive gliosis, this group of patients is refereed to as “gliosis only, no-HS”.

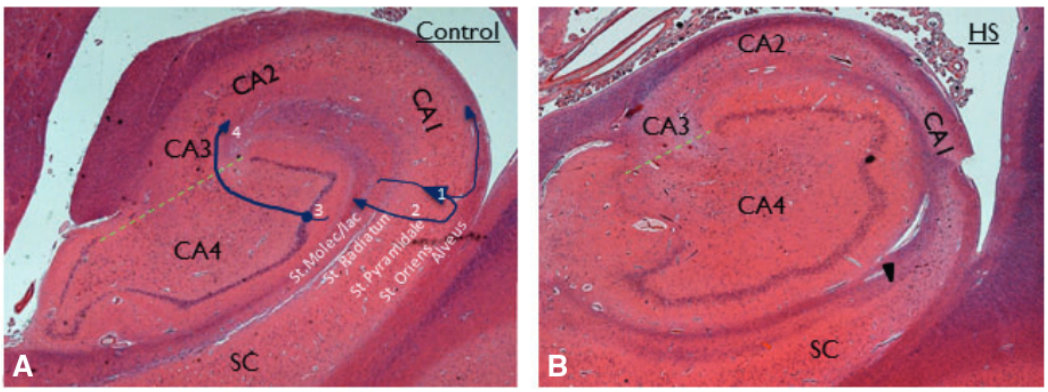

FIGURE 4.2 – Hippocampal subfields. (A) demonstrates a section of the hippocampus from a neurologically normal patient and (B) from a patient with a temporal lobe epilepsy. The subregions of the hippocampus are marked from CA1 to CA4 (CA = cornu ammonis). In the epileptic hippocampus (B), sclerosis is evident - there is a hardening or scarring of tissue, particularly visible as a sharp cutoff between the atrophied CA1 sector and the intact subiculum (SC). Illustration from Malmgren et al., 2012 licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

Type 1 refers to severe neuronal cell loss and gliosis predominantly in CA1 and CA4 regions, while it is usually CA1 predominant neuronal cell loss and gliosis for the type 2, and CA4 predominant neuronal cell loss and gliosis for the type 3.

Malformations of cortical development

The development of the cerebral cortex results from complex and overlapping processes of cellular proliferation, differentiation, and apoptosis, of migration, and development of neuronal connections. The term malformation of cortical development (MCD) describes the structural abnormality resulting from any defect affecting any stage of this development [Raybaud et al., 2011]. MCDs are also an important cause of epilepsy, particularly in children.

Focal cortical dysplasias (FCD) belong to the large spectrum of malformations of cortical development [Guerrini et al., 2015]. They are the most common structural brain lesion in children with drug-resistant focal epilepsies. The ILAE classification system of FCDs has a three-tier system [Kim et al., 2019]. This system is composed of isolated FCDs (FCD type I and II) and variants associated with other (potentially) epileptogenic lesions (FCD type III).

For the type I FCD, the lesions are dyslamination and disrupted organization of tissue architecture, but with morphologically normal neurons and glial cells. Type I FCD is further divided into subtypes Ia (Cortical dyslamination only, with an abnormal radial organization), subtype Ib (Cortical dyslamination, plus giant or immature neurons, with abnormal tangential layering). The combination of both conditions will be classified as FCD type Ic. The FCD type II variants are characterized by the presence of cortical dyslamination and dysmorphic neurons without (FCD type IIa) or with balloon cells (FCD type IIb). The definition of FCD type III is cortical dyslamination abnormalities associated with a principal lesion, usually adjacent to or affecting the same cortical area/lobe. It is also divided into several subtypes: FCD type IIIa (cortical lamination abnormalities in the temporal lobe associated with hippocampal sclerosis), FCD type IIIb (cortical lamination abnormalities adjacent to a glial or glioneuronal tumor), FCD type IIIc (cortical lamination abnormalities adjacent to vascular malformation), and FCD type IIId (cortical lamination abnormalities adjacent to any other lesion acquired during early life, for example, trauma, ischemic injury, encephalitis). Magnetic resonance imaging techniques have provided a non-invasive way for the characterization of some forms of FCDs. According to the findings in Urbach et al., 2021, only FCD type II have distinctive MRI: it may demonstrate an increased cortical thickness, blurring of the gray/white matter junction, abnormal gyral/sulcal pattern. For the FCD type I, the cortex is only little altered so that these lesions are hardly visible with MRI and thus often considered as "non-lesional". Still, they may show blurring of the gray/ white matter junction. FCD type IIIa may show white matter hypoplasia and white matter blurring, but nothing has been discovered for the other subtypes.

MRI abnormalities of FCD are often subtle, often overlooked, and their visibility on MRI depends on the FCD type. Between 15 to 30% of patients with drug-resistant epilepsy are considered as MRI negative, meaning that no structural lesion is identified [Duncan et al., 2016]. However, it comprises patients without a MRI lesion and those, in which a subtle MRI lesion is overlooked.

Periventricular nodular heterotopias (PNHs) are another type of MCD related to neuronal migration disorders, frequently associated with drug-resistant epilepsy [Mirandola et al., 2017]. The term "heterotopia" describes apparently normal cells,

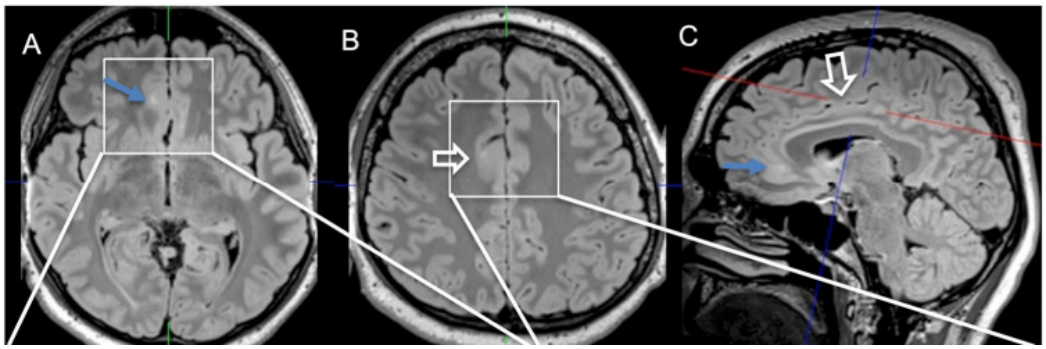

FIGURE 4.3 – FLAIR images of a patient with two FCD type IIa in the right cingulate gyrus. Illustration from Urbach et al., 2021 licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

but in an abnormal position. In case of PNHs, misplaced neurons are located in the white matter along the ventricle walls and are organized in a nodular structure. Heterotopic nodules may vary in their size, number, and localization along the ventricle system. They are sometimes associated with other cerebral malformations, such as FCD.

Polymicrogyria (PMG) is an epileptogenic malformation of cortical development. PMG refers to the excessive gyration or microfolding of the cerebral cortex [Shain et al., 2013]. It may be bilateral or unilateral and may occur in a variety of topographic regions, the most common of which is the perisylvian region. On the MRI, a proportion of PMG malformations include relative hypoplasia of cortical volume in the cortex of affected regions, white matter abnormalities.

4.4.2 Epileptogenic zone localization. Clinical protocol

A "seizure" is a paroxysmal alteration of neurologic function caused by the excessive, hypersynchronous discharge of neurons in the brain [Stafstrom et al., 2015]. The area of cortex that generates seizures is called the epileptogenic zone (EZ). One of the challenges in epilepsy evaluation is the successful identification of the EZ. Various tools and techniques can be applied to that purpose, including imaging methods, mainly MRI and PET, as well as electroencephalogram (EEG), mostly performed with implanted electrodes in intracranial EEG (iEEG) and sometimes associated with video recordings of seizures. A patient can be recommended for surgery in case of a relevant structural lesion detection that is consistent with the results of video EEG telemetry [Duncan et al., 2016]. If no relevant lesion on MRI is found for a patient, FDG PET is a useful next imaging technique that can possibly reveal an area of hypometabolism. Subsequent investigations such as SPECT, electrical source imaging (ESI), magnetic source imaging (MSI), and simultaneous EEG and fMRI (EEG-fMRI) are recommended in case nothing has been detected with the two previous imaging methods. Each method could appear useful in localization of the seizure source.

The role of MRI and PET imaging in the lesion localization

Brain imaging has a crucial role in the evaluation of a person with seizures. Success of a resective surgery depends on the correct localisation of the epileptogenic zone, therefore, preoperative investigations are vital. According to Rüber et al., 2018, MRI constitutes a necessary, albeit not sufficient part of the presurgical

routine for epilepsy patients. MRI is more sensitive than CT and is therefore preferred, especially for the detection of cortical malformation, or hippocampal sclerosis [Stafstrom et al., 2015]. Standard clinical protocols for epilepsy should include T1-weighted, T2-weighted, and fluid-attenuated inversion recovery (FLAIR) images. Three-dimensional (3D) sequences with isotropic voxels (cube-shaped voxels of identical length on each side or image plane) of 1 mm or less are recommended [Bernasconi et al., 2019]. FLAIR images have shown an accuracy of 97% for detecting abnormalities associated with hippocampal sclerosis (HS)[Cendes, 2013]. The MRI features of HS include reduced hippocampal volume, increased signal intensity on T2-weighted imaging, and disturbed internal architecture [Malmgren et al., 2012].

Diffusion-weighted imaging (DWI) is another form of MR imaging that represents diffusion of water molecules. While it is more sensitive to acute changes in stroke and encephalopathy, DWI hyperintensity has also been reported after status epilepticus - a seizure with 5 minutes or more of continuous clinical and/or electrographic seizure activity or recurrent seizure activity without recovery between seizures [Yokoi et al., 2019]. Diffusion tensor imaging (DTI) - a specific type of modeling of the DWI datasets - is another MRI sequence applied for intractable epilepsy detection. It reveals subtle alterations in white matter microstructure proximal and distal to the epileptic focus [Leyden et al., 2015]. In work of Park et al., 2019, grey matter (GM) anatomical features from structural MRI data, and white matter (WM) anatomical features from diffusion MRI were used as markers to discriminate TLE patients from the healthy controls.

The detection of subtle lesions, however, remains challenging, as they can be missed during standard visual inspections of the images. While up to 87% of FCD type I is present on images, this figure is only 33% for the FCD type II [So et al., 2015].

Studies show that PET scans serve as a confirmatory test, especially for MRI-negative patients [Willmann et al., 2007]. Using FDG-PET and high-resolution MRI (HR-MRI) co-registration in patients with MRI-negative refractory extra-temporal lobe epilepsy can improve the identification of the epileptogenic onset zone [Ding et al., 2018]. FCD include changes in gyral size, abnormal gyral shape, decreased cortical T1 intensity, increased T2 signal, and poor gray and white matter differentiation [Wong-Kisiel et al., 2018]. Positron emission tomography with fluorine-18 fluorodeoxy-glucose ([18F]FDG) is an important tool to define the onset zone and to better understand the functional alterations induced by various forms of epilepsy. The epileptogenic focus in the interictal phase usually appears as a hypometabolic area on [18F]FDG-PET. A meta-analysis of the studies evaluating the impact of PET imaging was done in a work of [Willmann et al., 2007]. A sensitivity of 70–85% for [18F]FDG-PET in patients with TLE have been reported [La Fougère et al., 2009]. [18F]F-FDG PET has proved highly sensitive for the detection of FCD type II. It demonstrated the localization of FCD type II in 83% of patients in [Desarnaud et al., 2018] by integration of electroclinical data and coregistered PET and MRI.

Kikuchi et al., 2021 explored the diagnostic accuracy for the epileptogenic zone detection in focal epilepsy and concluded that FDG-PET/MRI scans (Figure 4.4) provide higher visual capabilities than standalone MRI or FDG-PET/CT, since it is a simultaneous acquisition of both anatomical and functional information.

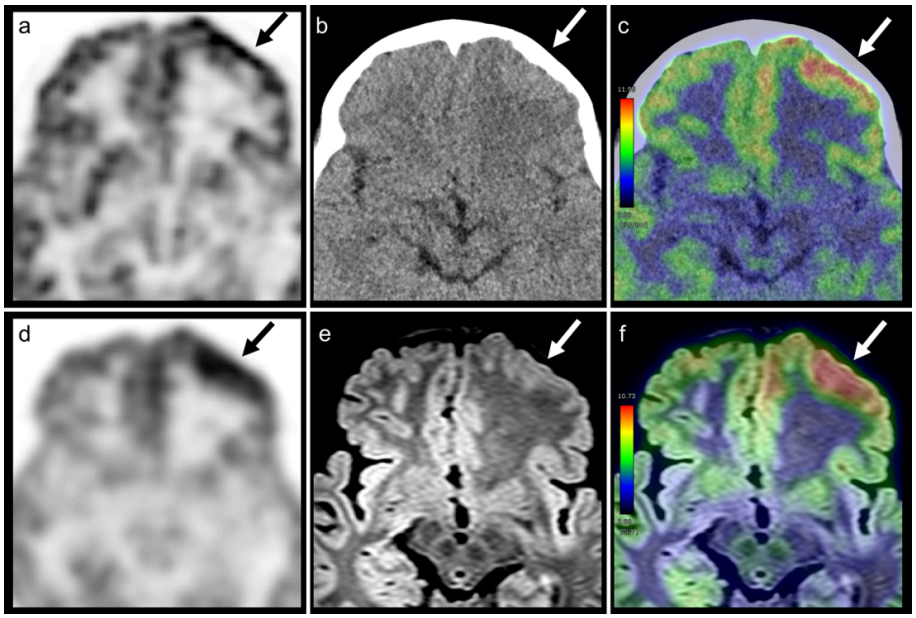

FIGURE 4.4 – Comparison of different examinations of a patient with FCD, type Ib: (a) PET from PET/CT, (b) CT, (c) PET/CT, (d) PET from PET/MRI, (e) FLAIR, (f) PET/MRI. Illustration from Kikuchi et al., 2021 licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

4.4.3 State of the art CAD systems for epilepsy detection

Recent ML/DL applications for epilepsy include automatic seizure detection from clinical data, pre-surgical planning, prediction of medical and surgical outcomes [Abbasi et al., 2019]. Some of the works are based on EEG data with the goal to identify epilepsy seizures on the EEG signals and perform a classification between normal and abnormal signals, linked to epilepsy. The proposed solutions utilize 2D-CNNs [Akut, 2019, San-Segundo et al., 2019, Türk et al., 2019], long short-term memory recurrent neural networks (LSTM-RNN) [Hussein et al., 2018, Bouallegue et al., 2020, Najafi et al., 2022]. The latest work achieved a 96.1% accuracy, a 96.8% sensitivity, and a 97.4% specificity in distinguishing normal subjects from subjects with epilepsy.

In the works utilizing imaging sources for epilepsy detection different kinds of features are extracted.

The majority of CAD systems for epilepsy detection are based on neuroimaging data. The development of these systems can be divided into two main directions:

1. Patient-level discrimination. This task consists of discriminating patients with epilepsy from healthy controls. A number of studies address this question with the help of SVMs [Keihaninejad et al., 2012, Bharath et al., 2019, J. Huang et al., 2020, B. Zhou et al., 2020, S. Chen et al., 2020]. From 84% to 97.6% of accuracy was possible to achieve. In B. Zhou et al., 2020, the authors utilized a combination of features derived from structural MRI (sMRI) and resting-state functional MRI (rs-fMRI) data. From sMRI, three measures were extracted: gray matter (GM), white matter (WM), and cortical thickness. From rs-fMRI, two measures were used: amplitude of low-frequency fluctuation (ALFF) and

regional homogeneity (ReHo). These five measures, encompassing both structural and functional aspects of the brain, were then combined to provide integrated information for the classification model.

The latter work of S. Chen et al., 2020, in fact, uses the combination of SVM and voxel-based morphometry (VBM) that is a neuroimaging technique that investigates focal differences in brain anatomy. The core process of VBM is segmenting the brain into grey matter, white matter, and cerebrospinal fluid [Nemoto, 2017]. The features used for the SVM classification were derived from regions where gray matter volume (GMV) abnormalities were detected through VBM analysis. The volumes of these abnormal regions were calculated for each participant and served as the features in the SVM model.

Neural networks have also been exploited for the classification task [Si et al., 2020, M.-H. Lee et al., 2020, Nguyen et al., 2021]. Though these works show the potential of machine/deep learning methods, the clinical use of the differentiation between healthy subjects and patients with epilepsy is limited. Thus, the next two directions appear to be more clinically important.

2. Lateralization of the TLE foci. Establishing the laterality of the epileptogenic focus with as much certainty as possible is an important task in the preoperative evaluation of patients with TLE. Most of the research works utilize machine learning methods such as SVM applied on morphometric features (usage of the ratio of parahippocampal gyrus volumes to hippocampal volumes, derived from T1-weighted MRI led to 100% of lateralization accuracy in hippocampal sclerosis and 91% of accuracy in TLE [Keihaninejad et al., 2012]), structural connectivity from DTI [Fang et al., 2017], morphological features from T1 [Mahmoudi et al., 2018], morphological features T1, T2 and FLAIR modalities [Bennett et al., 2019], FLAIR and PET signals respectively [Beheshti et al., 2020a, Beheshti et al., 2020b]. The later work stated that FDG-PET and single-photon emission computed tomography can robustly identify TLE patients when the MRI is negative.
3. Localization of the epilepsy foci. Various applications of machine learning have been proposed for lesion identification. The task is more challenging and the majority of works focuses on FCD epilepsy detection. Tables 4.1 - 4.3 summarize the current methods for the identification of epileptogenic foci, with the majority of works focusing on FCD detection.

One of the biggest challenges when developing data driven models that localize the epileptogenic region is getting a ground truth. While ground truth labels are easily mined for discrimination (a healthy subject, or an epilepsy patient) and lateralization tasks, it is time and resource consuming to outline a precise area of an epileptogenic focus. For radiologically positive MRI seizure-onset zones (SOZs) can be proved with the help of iEEG, and usually it is possible to obtain voxel-level annotations. On the contrary, for MRI-negative patients when nothing is visible on images, manual masks for supposedly epilepsy lesion may be only roughly delineated, and it usually requires that the patient undergoes a resection surgery, so further post-surgical justification can be proven including histopathological investigations.

Most of the studies aiming at localizing the epilepsy lesion consider T1w MRI images solely or in combination with FLAIR images [Gill et al., 2017, Gill et al., 2018, Alaverdyan et al., 2020]. In the work of Jin et al., 2018, FLAIR data were not used as a multivariate input, because of their unavailability in the control subjects. Handcrafted features are still widely used. As we can see from tables 4.1 - 4.3, the

popular choice is morphological features (SBM). SBM stands for surface-based morphometry - a group of brain morphometric techniques used to construct and analyze surfaces that represent structural boundaries within the brain. Boundaries between the grey matter and white matter are extracted from brain segmentation, and the corresponding surface is generated by a meshing algorithm that encodes relationships between voxels on the boundary into relationships between polygonal or polyhedral surface elements. Cortical thickness, grey-white matter intensity contrast, curvature, sulcal depth and FLAIR intensity at each vertex of the 3D cortical reconstruction were used as features in [Adler et al., 2017] as they are structural markers of FCD. Grey-white matter junction computed by a convolution of the binarized image of T1-w MRI image quantifies the grey-white matter blurring, and is used to characterize FCD [El Azami et al., 2016]. The *supervised* approaches consider mostly MRI-positive cases due to limitations in obtaining accurate lesion masks for MRI-negative patients. One of the exceptions from this observation is the work of Ahmed et al., 2015. They applied logistic regression, using an iterative-reweighted least squares (IRLS) algorithm on the vertices of the cortical surfaces. The evaluation dataset included 31 patients with confirmed FCD, and the proposed method managed to detect lesions in 6 out of 7 MRI-positive patients, as well as 14 out of 24 FCD lesions in MRI-negative patients. Authors state that the resection zones of MRI-negative patients should not be treated as a gold standard for training models as they include both lesional and nonlesional tissues. Ahmed et al., 2016 extended the previous study by adding new morphological features. A semi-supervised approach (hierarchical conditional random field) reached a 75% detection rate for the MRI-negative patients (compared to a human expert detection rate of 0%). Adler et al., 2017 proposed to use a simple neural network to classify healthy from pathological vertices based on 28 cortical features. The highest performance was reached with using FLAIR intensity (AUC = 0.83). Jin et al., 2018 also used cortical features and a neural network showing sensitivity of 52.9% in the group of patients where the MRI was negative, and sensitivity of 81.8% for MRI-positive patients. An automated epilepsy detection model trained on data-driven features was presented in Gill et al., 2018 with two CNNs trained to classify raw image voxels. The sensitivity and specificity of 91% and 92% respectively have been reached. Here, the training cohort comprised of 40 patients with the location of the seizure focus established using intracranially-implanted electrodes. The testing dataset consisted of patients with histologically-confirmed FCD, as well as of patients with TLE and histologically-verified hippocampal sclerosis. Wagstyl et al., 2020 also used cortical features and a neural network for epilepsy detection. Results were concordant with SEEG seizure onset zone in 62% of focal epilepsies and 86% of histopathologically confirmed FCDs. Consensus clustering (an unsupervised learning technique that identifies stable clusters based on bootstrap-aggregation) was implemented in a work of H. M. Lee et al., 2020 to analyze features of FCD. Alaverdyan et al., 2020 introduced a pipeline that extracts features from a regularised siamese network and fed extracted features into one-class SVM. The possible epilepsy lesions are present as clusters of voxels that were classified as outliers. Trained on the combination of T1 and FLAIR images, the model demonstrated sensitivity of 62% with 21 patients in total out of which there were 18 MRI-negative patients.

The usage of PET images is less common in epilepsy detection, however, the potential is promising. Tan et al., 2018 considered morphology and intensity-based features characterizing FCD lesions from MRI and PET imaging. A 2-step approach based on SVM was proposed to recognize lesional vertices and minimize the level of

false positive detections. The results reported an increase in sensitivity from 82% to 93%, corresponding to the maximum specificity, from 61% to 64%, when PET imaging is considered alongside MRI data compared to quantitative MRI and multimodal visual analysis. The accompanying FP rate in FCD patients, however, increases as well.

Q. Zhang et al., 2021 focused on the detection of epileptic foci in pediatric patients with temporal lobe epilepsy. They hypothesized that epilepsy is strongly correlated to the high-dimensional interhemispheric symmetricity changes in PET images and exploited radiomic features with symmetric information to diagnose TLE. The right and left parts of 18F-FDG PET images were partitioned into pairs of cubes (PoCs), then for each input PET image, a Siamese CNN was applied to all generated PoCs, producing a probability score for each indicating the likelihood of containing an epileptic focus. This method significantly outperformed the physicians blinded or unblinded to clinical information (90% vs. 56% and 68% in detection accuracy) with AUC = 0.93.

We have presented a detailed description of recent methods for automatic epilepsy detection in neuroimaging data. The following conclusions can be drawn out of this review:

— Few works generalize to different types of epilepsy. Most of the proposed methods are focused on a particular type (TLE, FCD). Considered features are therefore relevant to the particular pathology. It would thus might be beneficial to explore a wider range of features.

— Combination of different modalities is becoming more popular. As stated in the review paper of Sone et al., 2021, multimodal imaging is a recent trend in epilepsy research. While early works were mostly using T1 images, other modalities such as FLAIR or PET show added value to the model performance. Early fusion remains the preferred way of concatenating imaging modalities, but intermediate and late fusion schemes should be explored.

— Most studies are evaluated on MRI-positive cases where the lesions are visible on scans. MRI-negative patients remain challenging.

TABLE 4.1 – State-of-the-art methods for the detection of epileptogenic foci. Part I

Study	Data	Imaging Modality	Ground Truth	Features	Classifier	Results
Rudie et al., 2015	169 Epilepsy patients (85 with MTS, 84 without MTS)	T1	N/A (classification task)	Morphological (SBM, VBM)	SVM	ACC = 0.81 for Epilepsy patients with MTS vs. without MTS
Ahmed et al., 2015	31 FCD, 62 HC	T1	Neuropathological examination of the resected tissue	Morphological (SBM)	LR, IRLS	Detection in 6 of 7 MRI-positive patients, 14 of 24 MRI-negative patients (58%)
El Azami et al., 2016	11 FE, 77 HC	T1	manual segmentation, histology, SEEG	GM and WM junction	SVM	Sensitivity: 10/13, Avg. # of FPs: 3.2
Hong et al., 2016	41 FCD, 41 HC	T1	manual segmentation, histology, SEEG	Morphological (SBM)	SVM	ACC = 98% for Type I vs. II, approximately 90% for lateralization, 82–92% to predict seizure outcome
Ahmed et al., 2016	20 FCD, 115 HC	T1	post-surgical MRI of resected zone	Morphological (SBM)	semi-supervised HCRF + LoOP	Detection rate: 52-75%
Adler et al., 2017	22 FCD, 28 HC	T1, FLAIR	manual segmentation, histology	Morph (SBM), FLAIR signal	NN	AUC: 0.51 - 0.83
Gill et al., 2017	38 HC, 41 FCD	T1, FLAIR	manual segmentation	Morphological (SBM)	Decision trees	Sensitivity: 83%, Specificity: 92%

TABLE 4.2 – State-of-the-art methods for the detection of epileptogenic foci. Part II

Study	Data	Imaging Modality	Ground Truth	Features	Classifier	Results
Gill et al., 2018	38 HC, 40 FCD, 67 FCD, 63 TLE/HS	T1, FLAIR	manual segmentation	raw images	2 CNNs	Sensitivity: 91% Specificity: 92%
Y. Wang, Zhou, et al., 2018	12 FCD	DTI, T2	histopathology	FA, MD, VR, T2 signal	GPML, SVM	AUC = 0.76 to automatically detect FCD by GPML
Jin et al., 2018	155 HC, 15 HS, 61 FCD	T1	histopathology, multimodal localization	Morphological (SBM)	NN	AUC = 0.75 to detect FCD
Tan et al., 2018	28 FCD, 23 TLE	T1, FDG-PET	histopathology	Morphological (SBM), GM intensity, PET signal	SVM	Sensitivity: 0.93
Mo et al., 2019	80 HC, 80 TLE-HS (39R, 41L)	T1	histopathology	Visual features, Morphological	SVM, E-net LR	AUC around 0.98–0.99 for TLE-HS vs. HC, 96% detection rate for visually negative HS
H. M. Lee et al., 2020	35 HC, 46 FCD	T1, FLAIR, rs-fMRI	manual segmentation, histopathology	Morph (SBM), FLAIR signal, Gradient, Ratio, fALFF	Consensus clustering (unsupervised)	Four relevant structural profiles (WM, GM, GM and WM, GM-WM interface) were identified
Wagstyl et al., 2020	20 HC, 34 FCD	T1, FLAIR	SEEG	Morph (SBM), FLAIR signal	NN	Sensitivity: 0.74, Specificity: 1.00

Study	Data	Imaging Modality	Ground Truth	Features	Classifier	Results
Alaverdyan et al., 2020	21 FE, 75HC	T1, FLAIR	manual segmentation	T1, FLAIR signals	RSN, SVM	Sensitivity = 0.62 to detect anomaly lesion
Y. Guo et al., 2020	56 FCD, 40 GNTs	T1, T2, FLAIR	VEEG	Visual assessment	RF, SVM, DT, LR, XGBoost, LightGBM, CatBoost	AUC = 0.934 for FCD vs. GNTs by RF-based ML when combined MRI and clinical info
Snyder et al., 2021	15 FCD, 30 HC	T1, T2, FLAIR	histology	Morphological (SBM), signal intensity	Normative model	Sensitivity = 80% Specificity = 70% to detect FCD
Q. Zhang et al., 2021	201 TLE (P), 24 Control (lymphoma)	FDG-PET	manual segmentation	Radiomics	CNN	AUC = 0.93, ACC = 0.90 to detect epileptogenic focus
Kanber et al., 2021	27 FE (MRI-), 42 FE (MRI+), 62 HC	T1, FLAIR	SEEG, manual segmentation	MRI signals	LightGBM	For MRI+: median DSC = 0.59, Sensitivity = 0.55, Specificity = 0.99 For MRI-: Detection rate = 67%

TABLE 4.3 – State-of-the-art methods for the detection of epileptogenic foci. Part III.
ACC, accuracy; AUC, area under the ROC curve; CNN, convolutional neural network DT, decision tree; E-net LR, elastic net logistic regression; FA, fractional anisotropy; fALFF, fractional amplitude of low-frequency fluctuations; FCD, focal cortical dysplasia; FE, focal epilepsy; FP, false positive; GM, gray matter; GNTs, glioneuronal tumors; GPML, Gaussian processes for machine learning; HC, healthy controls; HCRF, hidden-state conditional random fields; HS, hippocampal sclerosis; IRLS, iterative-reweighted least squares; L, left; LoOP, local outlier probabilities; LR, logistic regression; MD, mean diffusivity; MTS, mesial temporal sclerosis; NN, neural network; P, pediatric cases; R, right; RF, random forest; RSN, regularized Siamese neural network; SBM, surface-based morphometry; SEEG, stereotactic EEG; TLE, temporal lobe epilepsy; VBM, voxel-based morphometry; VEEG, Video-EEG; VR, volume ratio; WM, white matter.

Chapter 5

Data and model overview

In chapter 4, we described the key elements of a CAD system for medical image analysis, and went through the latest applications in the domain of neuroimaging, followed by a detailed overview with the state-of-the-art methods for epilepsy detection. Within this project, we aim to address questions of generating missing data and joining multiple modalities to propose a better system that is able to detect subtle anomalous regions in the brain. In this chapter, we present a baseline CAD model for epilepsy detection, as well as a detailed description of the data set available for this study containing both healthy controls and patients with confirmed epileptical lesions.

5.1 CAD model for epilepsy detection

The overall model for the epilepsy detection is shown in Figure 5.1. This model was developed in the group and showed promising performance for T1+FLAIR modalities [Alaverdyan et al., 2020]. It has two main parts:

1. Representation learning (green dotted frame)
2. Outlier detection learning (blue dotted frame)

These two components are both trained on a set of healthy controls. In the beginning, a siamese neural network (SNN) is trained on patches extracted from random locations within the brain mask from healthy subjects. Once the training is finished, a oc-SVM is built and trained for every voxel taking as an input the representations derived from the SNN. The hidden representation for the oc-SVM corresponds to patches of healthy subjects centered at the voxel. In total, we have as many oc-SVMs models as the total number of voxels within the brain mask. After that, a patient's image can be processed by the system. The CAD system generates an output in the form of a score map, which matches the size of the input image. In this score map, each voxel is associated with a value, representing the output of its corresponding oc-SVM model. At the last stage, the output score map undergoes post-processing to generate a cluster map highlighting the most suspicious regions detected by the system.

For the effective model training, the data need to be pre-processed. The first step is to align all available imaging acquisitions to a common template to ensure voxel-to-voxel correspondence between all the subjects (healthy ones and patients). Details for the pre-processing steps will be given further.

For the representation learning, the input image is split into a set of patches of size 15x15 that are fed to the SNN. It can be done in a standalone fashion (one modality) or as multichannel architecture, where each channel corresponds to one modality. The effectiveness of using multiple modalities is one of the goals of this

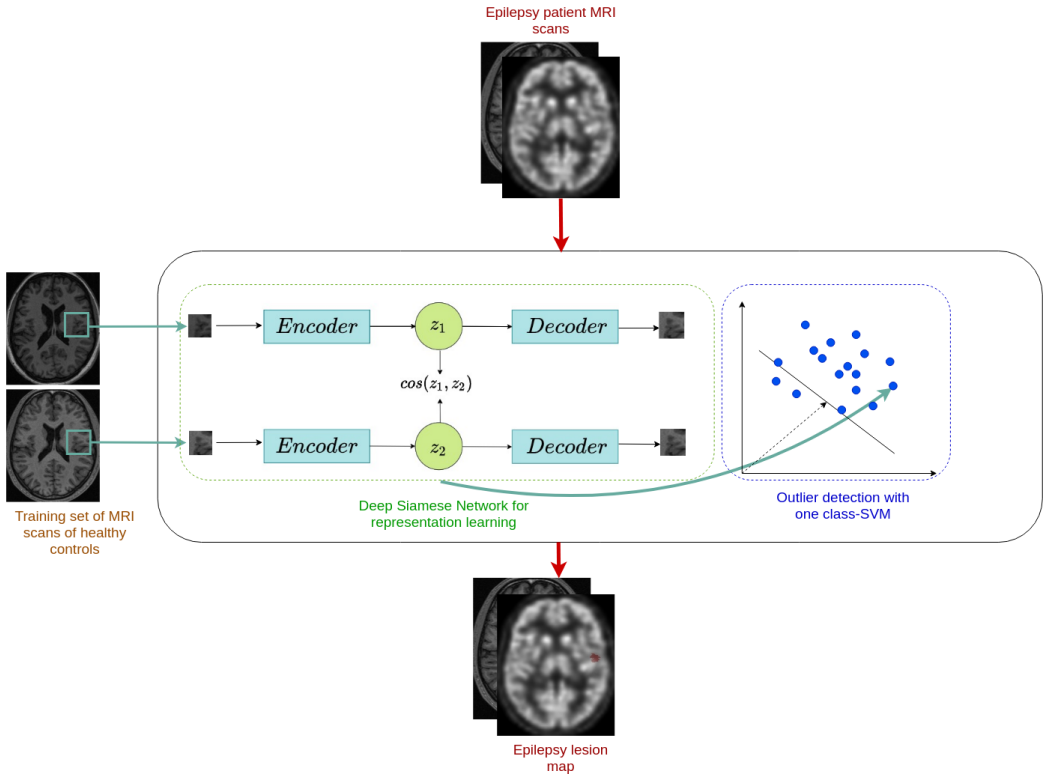

FIGURE 5.1 – General model for the epilepsy detection.

work, and more details will be given further. Once the input is defined, the data are transferred to the next component of the CAD system.

5.1.1 Feature extraction

At this step, the model learns the representation for the provided input. Benefits from both autoencoders and siamese networks are used in the unified framework adapted for the outlier detection problem. It is called a regularized siamese network with deep convolutional autoencoders. A siamese network is composed of two sub-networks, with identical architecture, a shared parameter set, and a cost module. Figure 5.2 shows the whole framework.

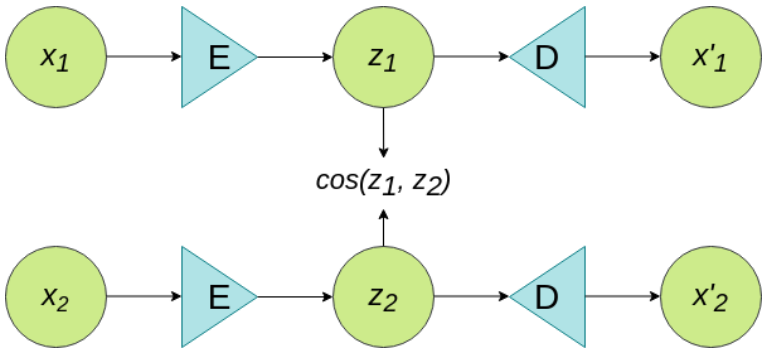

FIGURE 5.2 – Regularized siamese network.

Given a data set $X = \{x_i\}_{i=1,\dots,n}$, $x_i \in \mathcal{R}^d$ composed of n points, we want to find a mapping $D_\theta : \mathcal{X} \to \mathcal{Z}$ to project points from an original space into a representation space where similar examples form a close neighborhood. That goal is reached

with a help of a regularized siamese network. It receives as an input two patches of different controls located at the same location in the brain (x_1, x_2) that propagates further through two subnetworks, namely, convolutional autoencoders. These two subnetworks have identical components - an encoder E and a decoder D. E encodes the input to the hidden space Z with a series of convolutional operations, followed by D performing deconvolutional and upsampling operations. As the output, subnetworks return the reconstructed image x_i' of the given input x_i. The loss function for a single pair is:

$$L(x_1, x_2; \theta) = \sum_{t=1}^{2} \|x_t - \hat{x}_t\|_2^2 - \alpha \cdot cos(z_1, z_2) \tag{5.1}$$

It comprises two parts: the first component minimizes the squared error between the subnetwork input and corresponding output, thus ensuring a better reconstruction quality; the second term imposes a similarity in the hidden representation space by maximizing the cosine similarity of the middle layer feature vectors, where the coefficient α controls the extent of similarity. Further, the extracted representation z is used for the outlier detection task.

5.1.2 Outlier detection

The one-Class SVM (oc-SVM) is an unsupervised learning technique that is able to differentiate the test samples of a particular class from other classes. It is a particular case of the binary SVM. First introduced in Schölkopf et al., 2001, it remains a popular method for detecting anomalies. In oc-SVM the examples of the dataset $X = \{x_i\}_{i=1,\dots,n}$, where $x_i \in \mathcal{R}^d$ are normal (or positive), and a desired hyperplane is sought to separate all the data points from the origin. Since the data are usually not linearly separable in the original space, the first step is to map points into a higher dimensional space through a mapping function $\phi(x)$ with a kernel $K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle$. The kernel function returns the inner product between two points in a suitable feature space. Different kernels are used by SVMs, but RBF remains the most used type of kernel function, showing good empirical performance on various datasets. Figure 5.3 illustrates a mapping of original data samples with RBF kernel in oc-SVM.

Once the data are projected into a new feature space, they have to be separated from the origin with maximum margin. The following problem must be solved:

$$\begin{aligned} \min_{\omega, \rho, \xi_i} \frac{1}{2} \|w\|^2 - \rho + \frac{1}{\nu n} \sum_{i=1}^{n} \xi_i \\ \text{subject to } w \cdot \phi(x_i) \geq \rho - \xi_i, \\ \xi_i \geq 0 \end{aligned} \tag{5.2}$$

where n is the total number of training examples, x_i is the i-th training example from data set X, ξ_i is a slack variable relaxing the inequality constraints, w and ρ define the separating hyperplane, $\nu \in (0, 1)$ is a parameter that characterizes the fractions of support vectors and outliers. When the optimal solution for the hyperplane is found, the decision for a particular example x depends on the side of the hyperplane it falls. This can be expressed as:

$$f(x) = sgn(w^* \phi(x) - \rho^*) \tag{5.3}$$

It will be positive for most examples x contained in the training set with a reasonably small $\|w\|$. the ν coefficient in the formulation of oc-SVM is an upper bound

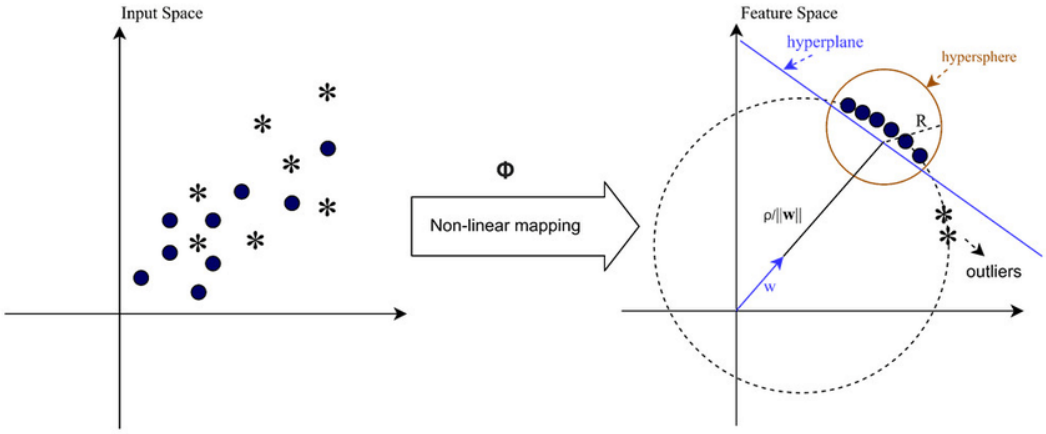

FIGURE 5.3 – The underlying concept of oc-SVM method. The points in the original space are projected into a higher dimensional space, where their separation from the point of origin is sought through maximizing the margin. Illustration from Yengi et al., 2020 licensed under CC BY 4.0 (https://creativecommons.org/licenses/by/4.0/).

on the fraction of permitted errors and a lower bound on the fraction of support vectors. For example, by setting it to be 0.01, it would allow 1% of the training examples to be misclassified as outliers.

$$\frac{|errors|}{n} \leq \nu \leq \frac{|errors| + SVs}{n} \tag{5.4}$$

oc-SVM design

In this pipeline, each voxel v_i is associated with a oc-SVM classifier C_i. The RBF kernel is chosen for all SVMs as the most common function. Thus, for each classifier a kernel is defined as:

$$K_{RBF}(z_{ik}, z_{ij}) = e^{-\gamma \left\| z_{ik} - z_{ij} \right\|^2} \tag{5.5}$$

where z_{ij} is the representation vector corresponding to the patch centered at v_i of subject j, and $\gamma = \frac{1}{\sigma^2}$.

For each voxel v_i, the corresponding oc-SVM classifier outputs a score, i.e. the distance to the found hyperplane:

$$score(v_{i=} w^* \phi(z_i) - \rho^*) \tag{5.6}$$

Eventually, all voxel distance scores combined together yield the distance map D_p for the given patient p.

5.2 Data description

The patient and control database of this project have been collected, curated and annotated as part of long lasting collaboration with Dr Julien Jung from HCL. The study was approved by our institutional review board (IRB) with approval numbers 2012-A00516-37 and 2014-019 B and written consent was obtained from all participants. This research was partly conducted as the part of a research program PHRC (programme hospitalier de recherche clinique) initiated by Pr. F. Maugière and Dr. J. Jung.

	Number of subjects	T1 (1.5T Siemens Sonata)	FLAIR (1.5T Siemens Sonata)	PET (mCT PET-CT Siemens tomograph)
DB_{C1}	35 healthy controls	160 x 192 x 192 1.2mm cubic voxels	176 x 196 x 256 1.2mm cubic voxels	109 x 200 x 200 2.036mm cubic voxels
DB_{C2}	40 healthy controls	160 x 192 x 192 1.2mm cubic voxels	176 x 196 x 256 1.2mm cubic voxels	-
DB_{ep}	31 patients	160 x 192 x 192 1.2mm cubic voxels	176 x 196 x 256 1.2mm cubic voxels	109 x 200 x 200 2.036mm cubic voxels

TABLE 5.1 – Summary of the data.

5.2.1 Study group

This study uses a training set of healthy individuals and a test set of epilepsy patients. The data set is described in table 5.1.

Healthy control group: As part of the PHRC research program, two control databases were acquired consisting of 75 healthy controls (referred to as DB_{C1} and DB_{C2} in Table 5.1) aged between 20 and 66 years. All the subjects have T1-weighted and FLAIR MRI sequences, but only 35 of them from DB_{C1} have PET exams as well.

Patient group: The test group DB_{ep} consists of 31 patients diagnosed with medically intractable epilepsy. The majority of the patients underwent imaging at the Neurological Hospital of Lyon. These patients are aged between 17 and 47 years old, with a median of 29. As a part of pre-surgical evaluation, they all had T1-weighted, FLAIR and PET examinations. Moreover, the patients had to undergo intracranial EEG exam in order to get localizations of the epileptogenic zone (EZ). It is important to note that the availability of patient dataset varied over the course of the study as we wanted to involve as many relevant participants as possible. The detailed lists of patients used in three major experimental phases of this PhD project are provided in Table 5.2 and referred to as DB_{ep1}, DB_{ep2} and DB_{ep3}. DB_{ep1} is the cohort of patients participating in the first experimental phase of chapter 7 dedicated to the generation and usage of synthetic PET images in epilepsy detection. We later included a few more patients to DB_{ep2} but excluded patients from AA to AE for the second major experimental phase of chapter 7, where we combined T1 and either real or synthetic PET images in the training of the detection model, while we only used PET in the first experimental phase. These patients were excluded because they had their PET exam acquired in a different hospital and on an a different scanner than those of the control datasets, thus potentially coming from a different distribution which could potentially negatively impact the performance of a proposed CAD model. For the last part of the study in chapter 8 where we explored the use of multiple modalities for the epilepsy detection, the dataset DB_{ep3} was comprised of patients participating in the 2 previous major experimental phases, with the addition of new patients H, L, W, X, Y resulting in 26 patients in total. In Table 5.2, we show the split of all patients into sub-groups depending on the experiment, we indicate if their PET images were acquired in the Lyon hospital and if patients had unsuccessful outcome after the surgery.

Patient	DBep1	DBep2	DBep3	PET Lyon	bad outcome
Patient A	✓	✓	✓	✓ mCT	
Patient B		✓	✓	✓ mCT	
Patient C	✓	✓	✓	✓ mCT	
Patient D	✓		✓	no	
Patient E	✓	✓	✓	✓ mCT	
Patient F		✓	✓	✓ mCT	
Patient G	✓	✓	✓	✓ mCT	
Patient H			✓	no	
Patient I		✓	✓	✓ mCT	
Patient J	✓	✓	✓	✓ mCT	
Patient K		✓	✓	✓ mCT	
Patient L			✓	no	
Patient M		✓	✓	✓ mCT	
Patient N		✓	✓	✓ mCT	
Patient O	✓	✓	✓	✓ mCT	
Patient P		✓	✓	✓ mCT	
Patient Q	✓	✓	✓	✓ mCT	
Patient R	✓		✓	no	
Patient S	✓	✓	✓	✓ mCT	
Patient T		✓	✓	✓ mCT	
Patient U	✓	✓	✓	✓ mCT	
Patient V	✓		✓	no	
Patient W			✓	no	
Patient X			✓	no	
Patient Y			✓	no	
Patient Z		✓	✓	✓ mCT	
Patient AA	✓			no	✓
Patient AB	✓			✓ mCT	✓
Patient AC	✓			✓ mCT	✓
Patient AD	✓			no	
Patient AE	✓			no	✓

TABLE 5.2 – Patients participation in 3 experimental phases.

5.2.2 MRI and PET acquisition

All the healthy controls and patients had 3D anatomical T1-weighted brain MRI sequences (TR/TE 2400/3.55; 160 slices of 192 x 192 1.2mm cubic voxels) and FLAIR MRI sequences (176 slices of 196 x 256 1.2mm cubic voxels) on a 1.5 T Sonata scanner (Siemens Healthcare, Erlangen, Germany).

PET scans were performed using a Biograph mCT PET-CT machine (Siemens). Head movement was minimized using an airbag, and a camera was used to monitor head position during the scan. Tissue and head support attenuation were measured using a low-dose CT scan taken before the emission data was collected. The emission scan was done dynamically and recorded in list mode for 60 minutes following injection. Static ^{18}F-FDG uptake images were created using a 3D-ordinary Poisson-ordered subset expectation maximization iterative algorithm, incorporating point spread function and time of flight (with a Gaussian filter of 4mm), and corrected for scatter and attenuation. The algorithm was run for 12 iterations with 21 subsets. The reconstructed volumes consisted of 109 contiguous slices, each 2.03mm thick and made up of 200x200 voxels (2.036x2.036mm^2). The actual resolution of the reconstructed images was approximately 2.6mm full width at half maximum in the axial direction and 3.1mm full width at half maximum in the transaxial direction, as measured for a source located 1cm from the field of view Jakoby et al., 2011.

5.2.3 Location of patient's brain lesions

Information on the location of the epilepsy-causing brain lesions in the patients is provided in tables 5.3 and 5.4. These tables include details on the clinical basis for determining the true location of the lesions. All of the patients in the study had an intracranial electroencephalogram (EEG) exam, and many of them had surgery to remove the lesions. Most of these patients had good surgical outcome based on the Engel Classification (Engel I or II), meaning they became seizure-free within six months of the surgery. A few patients instead received thermocoagulation treatment, which was successful in stopping their seizures and corroborated the findings of the EEG exams. A few patients (AA, AB, AC and AE) in table 5.2 from DB_{ep2} were classified with a mitigate outcome (Engel score GT 2). The brain lesions observed in most of the patients either did not fit into the histopathological categories of epilepsy presented in section 4.4.1 or were not analyzed at the time of the study, and therefore their type is mentioned as *Unknown*.

For each patient, there is a manual annotation available carefully drawn by an expert radiologist after reviewing the intracranial EEG results, post-op MR scans and clinical or thermocoagulation reports. In Figures 5.4 and 5.5, the ground truth annotations are superimposed on the corresponding T1-w MR transverse slices. The precise boundary of a true lesion was not obtainable, so that this *ground truth* represents a *reference zone*. If a cluster detected by the model intersects the ground truth area, it is considered to be a true positive (TP) result and false positive (FP) otherwise.

5.2.4 Data pre-processing

Pre-processing of images before feeding them into a CAD system is an important step. It can help to improve the performance of the CAD system by enhancing the quality and clarity of the images, making it easier for the system to perform the task. Additionally, it can help to reduce the amount of noise or other artifacts in the images, which can improve the accuracy of the CAD system. The T1-weighted

Patient	Lesion location	Lesion location confirmation	Lesion type	Age
Patient A	Temporal Lobe L	Intracranial EEG & successful thermoco-agluation	Unknown	17
Patient B	Temporal Pole L Insula L	-	-	-
Patient C	Temporal Lobe R	-	-	-
Patient D	Middle frontal gyrus L	Intracranial EEG & successful thermoco-agluation	FCD type II	43
Patient E	Temporal Pole R Insula L	-	-	-
Patient F	Hippocampus L, parahippocampus L	Intracranial EEG & surgical success	Unknown	41
Patient G	Precentral gyrus R	Intracranial EEG & surgical success	Unknown	19
Patient H	Superior temporal gyrus R	Intracranial EEG & surgical success	Unknown	44
Patient I	Insula L	-	-	-
Patient J	Temporal Pole R	-	-	-
Patient K	Insula	-	-	-
Patient L	Anterior temporal lobe R	Intracranial EEG & surgical success	Unknown	26
Patient M	Temporal Pole R	-	-	-
Patient N	Middle frontal gyrus L	Intracranial EEG & surgical success	Unknown	33
Patient O	Insula	-	-	-

TABLE 5.3 – Patient group description. Part 1

Patient	Lesion location	Lesion location confirmation	Lesion type	Age
Patient P	Hippocampus R	Intracranial EEG & surgical success	Histopathology: FCD type III with HS	41
Patient Q	Lateral remainder of occipital lobe L	Intracranial EEG & surgical success	FCD type II	29
Patient R	Orbital gyrus R	Intracranial EEG & surgical success	Ganglioglioma	47
Patient S	Hippocampus R	Intracranial EEG & surgical success	Histopathology: FCD type IIIa	31
Patient T	Surgical resction R	-	-	-
Patient U	Insula	-	-	-
Patient V	Insula R	Intracranial EEG & successful thermoco-agluation	Unknown	17
Patient W	Hippocampus R	Intracranial EEG & surgical success	Histopathology: FCD type III with HS	41
Patient X	Superior frontal gyrus R	Intracranial EEG & surgical success	FCD type II	21
Patient Y	Inferiolateral remainder of parietal lobe R	Intracranial EEG & surgical success	Unknown	25
Patient Z	Temporal Pole R	-	-	-
Patient AA	Frontal Lobe R	-	-	-
Patient AB	Medial Frontal Lobe R	-	-	-
Patient AC	Temporal Lobe L, Hippocampus L	-	-	-
Patient AD	Hippocampus L, parahippocampus L	Intracranial EEG & surgical success	Unknown	28
Patient AE	Middle frontal gyrus R	Intracranial EEG	Unknown	25

TABLE 5.4 – Patient group description. Part 2

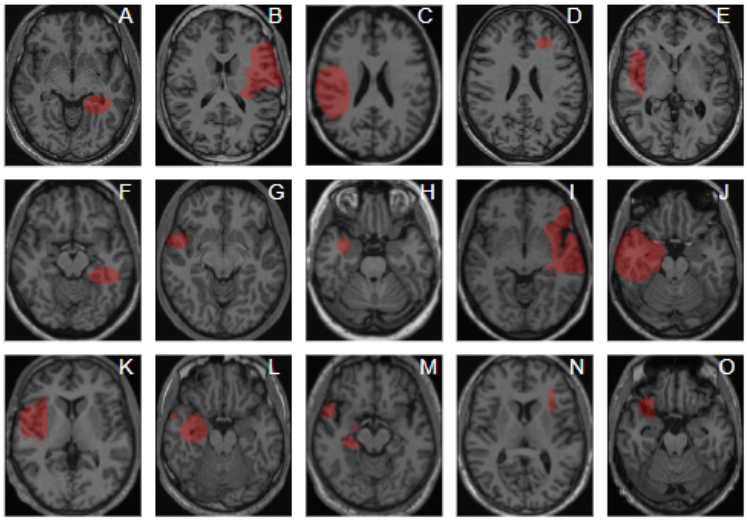

FIGURE 5.4 – The manual ground truth annotations (areas in red) overlaid onto T1-weighted MRI patients' transverse slices. Part 1

MRI volumes of all databases were first processed with the unified segmentation algorithm implemented in SPM12 (https://www.fil.ion.ucl.ac.uk/spm/doc/manual.pdf) using the default parameter values. This algorithm performs tissue segmentation (white/grey matter, cerebrospinal fluid), correction for magnetic field inhomogeneities, and spatial registration to the standard brain template of the Montreal Neurological Institute (MNI) with a voxel size of 1×1×1 mm. FLAIR and PET images were then rigidly aligned to their corresponding individual T1w MR images in the native space. Then, they were co-registered to the MNI space with SPM12 applying the transformation parameters derived from the registration of the T1 images. A masking image in the MNI space derived from the Hammersmith maximum probability atlas described in Hammers et al., 2003 was used at different steps of the pipeline, to focus or exclude specific brain regions. The resulting images size of all modalities after the pre-processing steps is 157x189x136.

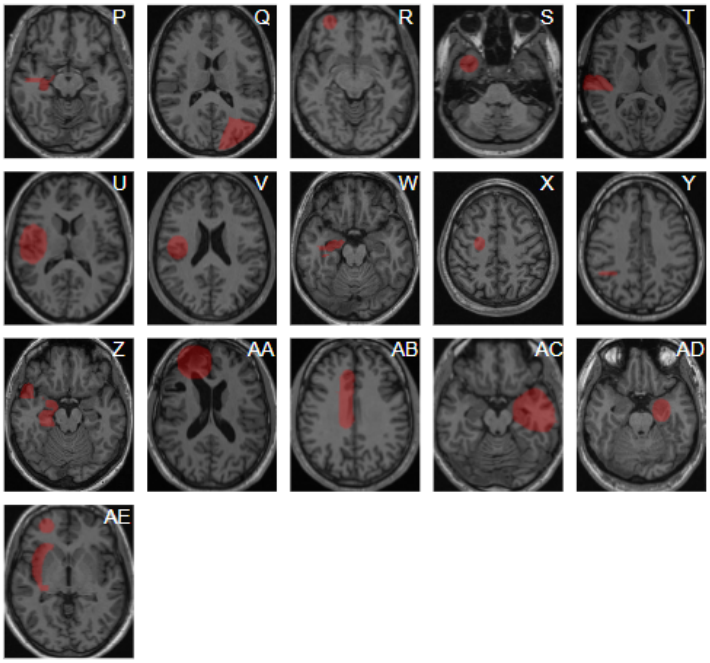

FIGURE 5.5 – The manual ground truth annotations (areas in red) overlaid onto T1-weighted MRI patients' transverse slices. Part 2

Chapter 6

Problem formulation

In the previous chapters we covered different aspects of development of CAD systems for medical imaging. We started by a general introduction into deep learning in the medical imaging in chapter 1. Chapter 2 focused on the utilization of multimodal imaging, highlighting the integration of different imaging modalities to enhance diagnostic accuracy. This chapter delved into three levels of image combination: early fusion, intermediate fusion, and late fusion. In the following chapter 3, we discussed the ways of dealing with missing data and modern approaches on synthetic image generation. In chapter 4, we looked at the main elements of a CAD system, highlighted the difference between supervised and unsupervised methods and we introduced the epilepsy detection problem. We described our backbone CAD model in chapter 5 as well as the available data for model training this model on a control dataset and making inferences on a cohort of epilepsy patients. This chapter provides our considerations for the problem of improving the existing CAD model for epilepsy detection and gives an understanding of the strategic choices we had to make and our proposed solutions.

6.1 Challenges and objectives

The objective of this study is to propose the improved version of the reference CAD model proposed in Alaverdyan et al., 2020 and described in the previous chapter. We continue to adhere to an unsupervised learning approach primarily due to the challenges in acquiring high-quality annotations, particularly those with precisely delineated borders of epileptic lesions. Annotated data remain scarce and often lacks the granularity necessary for supervised learning, thus making unsupervised methods more feasible and appropriate for this research.

As we saw in the data description part 5.2, for all the patients we have 3 available modalities - T1 and FLAIR MRI and PET, while in the healthy controls only DB_{C1} subgroup has the same set of modalities. We'd like the CAD system to be able to process both MRI and PET. Thus, we decided to investigate the possibilities for missing data generation. As described in chapter 3, early works used U-net model and its variations to generate a target modality from a source. Though U-net allows to generate realistically looking images, this network lacks generative capabilities. It was originally proposed for semantic segmentation tasks successfully extracting meaningful and class-relevant features, but it does not inherently possess the generative capabilities required for synthesizing new imaging modalities. The second limitation comes from the fact that preserving fine details and anatomical accuracy is crucial in medical imaging. U-Net's architecture, primarily focused on segmentation, is not optimized to retain the detailed structures in a synthetic image. With these limitations, and based on promising performance reported in chapter 3, we

chose a GAN-based approach for missing modality generation task. GAN models consist of two primary components: the generator and the discriminator. The generator creates images, while the discriminator evaluates them. As a generator U-net is often used, however our choice was made in favor of a ResNet model. ResNets are known for their ability to generalize, the depth and structure of ResNet (especially the skip connections) make it well-suited for learning a wide variety of features from the source data. GANs are difficult to train due to their complex relationships between a generator and a discriminator. Though we can rely on metrics evaluating image quality, the final feasibility of synthesised data can be measured by their added value in task-oriented problem solving. In the chapter 4 we also observed the power of vision transformers (ViT), however, for our specific task of generating synthetic medical images, we have decided not to utilize transformer networks. The primary design of transformers, particularly in their role of predicting the next token, makes them more suitable for such tasks as predicting the missing slices in a volumetric dataset where the transformer can effectively infer the missing part based on the contextual information from the existing slices [J. Xu et al., 2022], or enhancing the image quality [Feng et al., 2021].

The next challenge comes from the need to merge different modalities ensuring that the integrated dataset reflects accurately the underlying pathology and physiological processes of epilepsy. To address this, we plan to explore fusion techniques as outlined in 2.1. Here we present the ideas for every fusion level and justify our choice:

- Early fusion might involve combining data at the input level channel-wise.
- Mid-level fusion could consists of combining features from intermediate layers of a siamese network either by summing or averaging feature maps from corresponding layers, thus equalizing each modality's influence. Alternative method would be to implement a learnable transformation layer in the network or, inspired by the transformer architecture, incorporate attention mechanisms that would allow the network to focus on the most relevant features from each modality. Another approach might consist of exploring strategies to combine extracted from the siamese network features of different modalities before feeding them to the oc-SVM. The most straightforward method is to concatenate the feature vectors end-to-end, but while this method preserves all information, it can lead to high-dimensional data, which increases the complexity of the model. Averaging or weighing feature vectors would be another choice, though leading to potentially losing some distinct features. The concatenated vectors could be fed through additional neural network layers that can learn an optimal combination of features. These methods inherently demand additional time and effort for experimentation to identify the optimal network structure for our specific use case. Considering these factors, we have decided to initially implement the simplest method: feature concatenation. This approach will serve as our starting point, providing a baseline. If time permits, after conducting our main set of experiments, we will explore the feasibility and potential benefits of implementing more advanced fusion techniques to enhance our model's performance further.
- Late fusion focuses on integrating outputs at a stage closer to the final decision-making. One of the choices would be to compute a weighted average of the score maps from each modality, though that would require additional post-processing of score maps since their values are not initially bounded. We

decided to combine modalities at the cluster maps level. For each voxel, we can take the maximum score from the cluster maps of all modalities, prioritizing the detection of any potential anomalies, or averaging the values for equal contribution of each modality. We think it is important to preserve those clusters that were identified as anomalies with higher certainty and give even higher weight to those clusters that were identified as highly anomalistic in one modality and somewhat anomalistic in other modalities. We will present our strategy in the corresponding part of this work.

6.2 Contributions

Part II reports the main contributions of this PhD project. Chapter 7 starts with the detailed overview of GAN models that we utilised for synthetic PET images generation. This part encompasses the foundational structure of GAN models, including the architecture of individual components, loss functions, data ingestion and preprocessing steps. We then identify the metrics to assess the quality of the generated images and move to the experimental part with details for the GAN models training. In the following chapter we formulate the out-of-distribution problem to identify if synthetic images coming from the same distribution as original images and can be reliably used for data augmentation. We describe the concrete implementation of a CAD model for epilepsy detection first introduced in chapter 5 with the detailed architectures of individual components. Chapter 7 includes the results of our experiments where synthetic PET data were added to the original dataset for the CAD model training. The outcomes from these experiments report quantitative performance metrics and qualitative assessments of the detections. The next set of experiments include the setting where generated PET images completely substitute original PET images coupled with original T1 images. We show an improved sensitivity of the model trained on T1 and fake PET images and discuss our findings.

Chapter 8 addresses the challenge of fusing multiple imaging modalities at different levels. We proposed to compare three strategies: channel-level fusion as an early fusion, concatenating feature vectors derived from each modality as an intermediate fusion, and, finally, cluster maps merging as a late fusion. We compare all three levels of available modalities integration (T1, FLAIR, PET) and report the models performance. For the best approach that turned out to be the late fusion for three modalities we highlight the detection outcomes for patients, providing a detailed analysis and interpretation of the results yielded by our study.

Ultimately, the manuscript delivers a conclusive overview and defines our suggestions for future investigation.

Part II

Contributions

Chapter 7

Learning with synthetic data

In the previous chapters, we presented the general unsupervised detection anomaly model that we selected for for epilepsy detection in brain images, motivated the need to fuse PET and MRI imaging to improve performance and presented state-of-the-art approaches to improve the current model by integrating synthetic PET normative data and exploring the ways of efficient multiple modalities merging. We discussed the difficulties and challenges associated with proposed directions and justified our choices. In this chapter, we outline our experimental methodology for fake PET images synthesis, we investigate the characteristics of these generated images through the out-of-distribution problem definition, and examine the efficacy of utilizing synthetic PET data in identifying epileptogenic zones in brain images.

We start by describing the process of generating synthetic PET images. Deep generative models are used to get PET images that are similar in appearance to real PET images. We trained the model on a subset that we referred to as DB_{C1} in 5.2, the dataset consists of paired images of T1 and PET modalities and is used to generate synthetic PET images for our further experiments. We evaluate the quality of PET images qualitatively using most popular metrics that will be explained later on.

Next, we evaluate the quality of synthetic PET images approaching it as an out-of-distribution (OOD) problem to determine whether they are similar enough to real PET images. We discuss the metrics, the experimental pipeline and the results of the evaluation and their implications for the use of synthetic data in our project.

Finally, we discuss the implications of our findings for the use of synthetic data in medical image analysis. We highlight the potential benefits of using synthetic data for the training of our unsupervised anomaly detection model, including the ability to generate large amounts of data quickly and the potential to reduce costs associated with collecting and storing real data. We also discuss the limitations of our experiments and potential future directions for research in this area.

Overall, this chapter provides valuable insights into the use of synthetic data in medical image analysis and highlights the potential benefits and challenges of using synthetic PET data to detect epileptic regions in brain images.

7.1 PET image synthesis with GANs

As outlined in chapter 3, we reviewed various architectures, highlighting their strengths and limitations in the context of generating medical images. Building upon this foundational review, chapter 6 presented a justification for selecting GAN as our model of choice for the task of synthesizing missing PET modalities. The decision was predicated on GANs' superior generative capabilities and their demonstrated efficacy in producing highly realistic images.

GANs have shown great potential in generating medical images, and brain images of various modalities in particular. One of the main benefits of GANs is that they can generate realistic images that closely resemble real images, which can be used for training and testing of machine learning models. Generating large volumes of high-quality images can help overcome the limited availability of real PET data in our case, reducing the costs and eliminating the need to expose subjects to radiation.

In next subsections we provide detailed overview of the GAN models that were used in our experiments, their components and training procedures.

7.1.1 GAN models

The classical GAN model consists of two parts: a generator (G) and a discriminator (D). The generator is responsible for creating synthetic data that resembles the real data, while the discriminator is trained to distinguish between real and synthetic data. In our case, during training, the generator produces synthetic images, which is fed into the discriminator along with real images. The discriminator then evaluates the synthetic input and assigns a probability score indicating whether the input is real or fake. Based on the discriminator's evaluation, the generator receives feedback on how to adjust its output to create more realistic images.

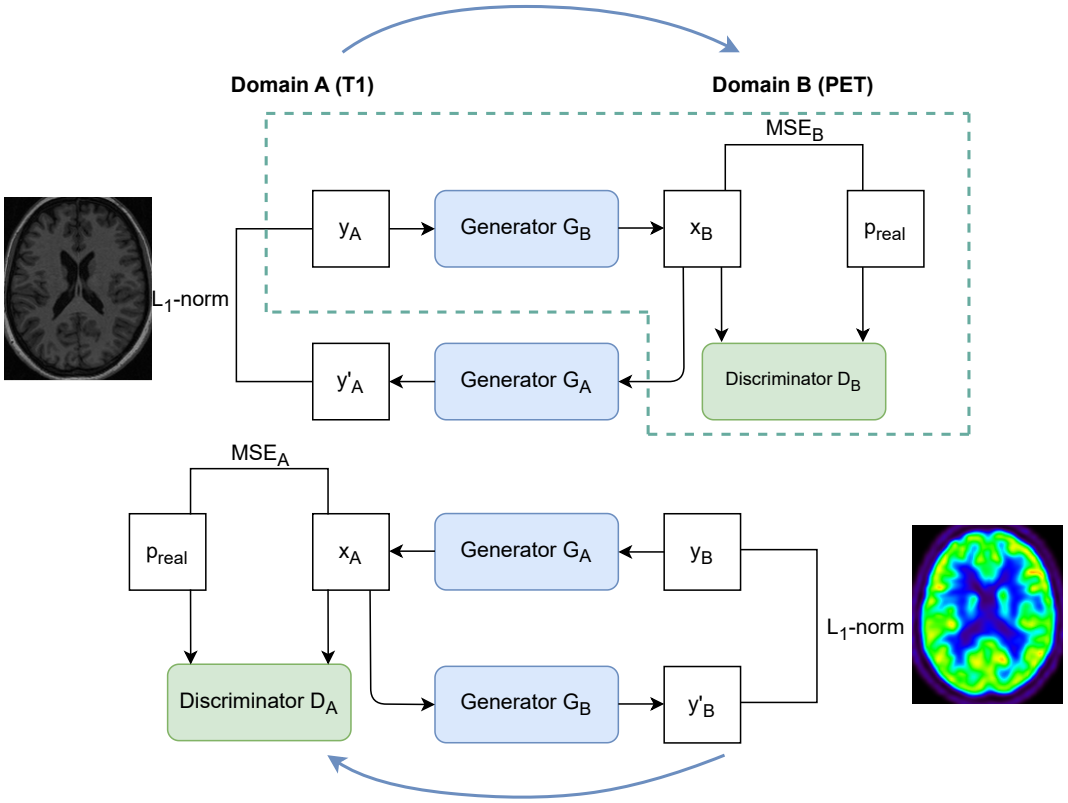

FIGURE 7.1 – Cycle-GAN architecture based on two baseline GANs translating images from domain A to domain B (upper GAN) and vice versa (lower GAN). The baseline GAN (Simple-GAN) is highlighted in the image with a dashed line

Simple-GAN. We first propose to use a standard GAN architecture with one generator G_B and one discriminator D_B as depicted in the upper part of Figure 7.1. Generator G_B attempts to improve the quality of the translated output x_b of domain B (PET modality) from the original input y_A from the original domain A (T1 modality),

thus deceiving the discriminator D_B. The training procedure is formulated as a min-max optimization problem of an objective function that the discriminator is trying to maximise and the generator is trying to minimize. In this study, we implement the least squares GAN (LSGAN) model Mao et al., 2017 that aims to minimize the following discriminator $L_{LSGAN}(D_B, A, B)$ and generator $L_{LSGAN}(G_B, A, B)$ losses :

$$\begin{aligned} L_{LSGAN}(D_B, A, B) &= E_{p(x_b)}[D_B(x_b)^2] + E_{p(y_b)}[(D_B(y_b) - 1)^2] \\ L_{LSGAN}(G_B, A, B) &= E_{p(x_b)}[(D_B(x_b) - 1)^2] \end{aligned} \tag{7.1}$$

where y_a and y_b are true images of domain A and B, respectively, and $x_b = G_B(y_a)$ is the fake image of domain B generated from y_a.

In the context of supervised image translation, where the model can be trained on paired images in both domains at the pixel level (e. g. corresponding images of the same patient), we propose to add a mean squared error (MSE) loss term L_{mse} (see eq. 7.2) between the fake image x_b generated from a true image y_a of domain A and its paired true image y_b in domain B.

$$L_{mse}(G_B) = E_{p(x_b)}[(x_b - y_b)^2] \tag{7.2}$$

Cycle-GAN. Cycle-GAN consists of two generator networks G_A and G_B and two discriminator networks D_A and D_B. The baseline Cycle-GAN model is shown in Figure 7.1. The generators translate images from domain A to domain B and vice versa. Each of the generator networks is trained adversarially using a corresponding discriminator D_A and D_B. In addition to the adversarial loss term of the simple GAN network in eq. 7.1, the key element in training Cycle-GAN network is a cycle-consistency loss function L_{cyc}:

$$L_{cyc}(G_A, G_B) = E_{p(y_a)}[\| y'_a - y_a \|_1] + E_{p(y_b)}[\| y'_b - y_b \|_1] \tag{7.3}$$

where y'_a is the fake image of domain A generated by generator G_A from the fake x_b, that is $y'_a = G_A(x_b)$ with $x_b = G_B(y_a)$. As for the simple GAN formulation, in a paired mode, we add a MSE loss term between real and synthetic images of both domains A and B.

Additionally, an identity loss can be used to ensure that the generative model preserves the content between the input and the output images when the input is already from the target domain, thus ensuring that the translation process does not introduce unnecessary changes (eq. 7.4):

$$L_{identity}(G_A, G_B) = E_{p(y_b)}[\|G_B(y_b) - y_b\|_1] + E_{p(y_a)}[\|G_A(y_a) - y_a\|_1] \tag{7.4}$$

7.1.2 Design of our GAN model

In 3.2.3, we provided an overview of the fundamental GAN architectures and their underlying principles. We began by introducing the basic GAN architecture, consisting of a single generator and discriminator, which forms the foundation of adversarial training. This architecture establishes a competitive learning process between the generator and discriminator networks, enabling the generation of synthetic data samples that closely resemble real data. Continuing our exploration, we delved deeper into the CycleGAN architecture, a significant advancement that introduced the concept of cycle consistency. Now, we describe the details if the architecture of the GAN models utilized in our experiments. By investigating these architectures in greater detail, we aim to gain deeper insights into their design choices,

performance characteristics, and their potential impact on the medical image generation.

7.1.2.1 Architectures of GAN components

Generator

Various architectures are employed for the generator component. In DCGAN [Ghassemi et al., 2020, Islam et al., 2020, Fernandez-Quilez et al., 2022] the generator typically starts with a dense layer that maps the input noise vector to a low-dimensional feature map, followed by a series of transposed convolutions. Each transposed convolutional layer increases the spatial resolution of the feature maps while reducing the number of channels.

U-net that we have described in 3.2.2 is another popular choice as a generator in GANs. It allows a more efficient and effective learning of the image-to-image mapping by exploiting the structural similarities between the input and output images, and has been used for various medical image synthesis applications [Sohail et al., 2019, Armanious et al., 2020, Q. Yang et al., 2020, Kalantar et al., 2021].

ResNet is another choice for a generator part. It is known for its ability to handle tasks without suffering from the vanishing gradient problem, and it is suitable for capturing complex and hierarchical patterns in the generated data. ResNet was chosen as a generator in [Thirumagal et al., 2020, X. Liu et al., 2021].

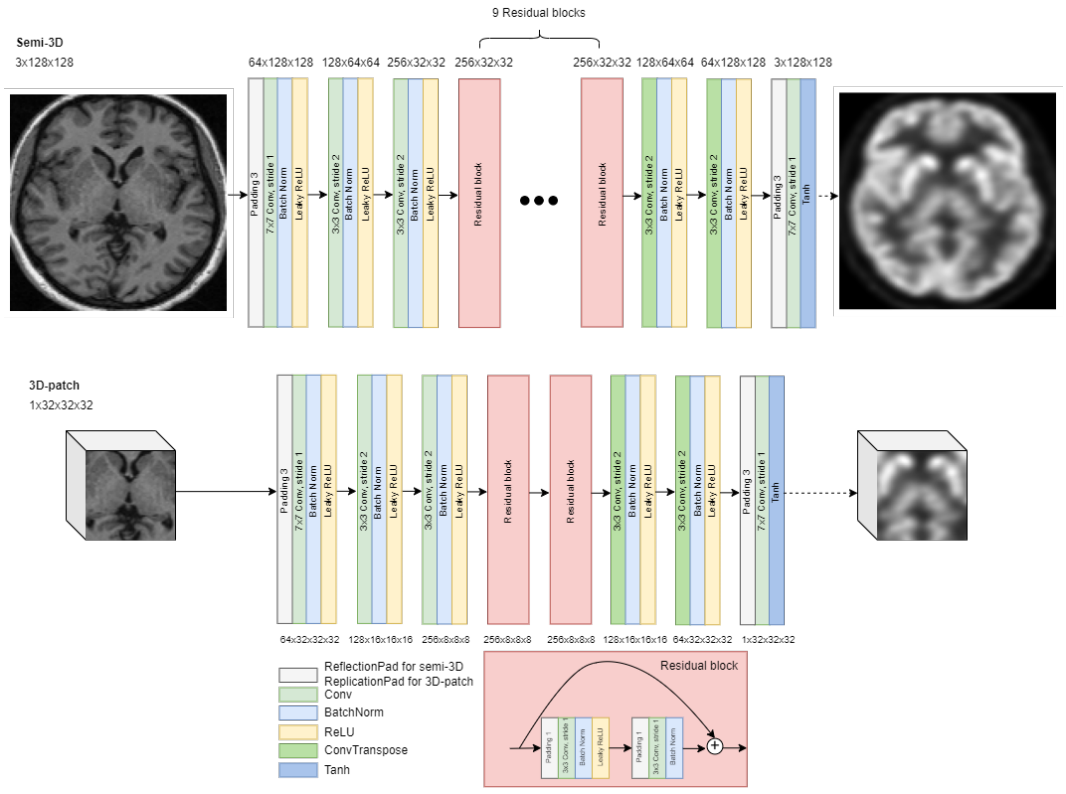

FIGURE 7.2 – ResNet architecture for a generator in either semi-3D or 3D-patch setting.

In our work we utilised the generator based on ResNet, its architecture is given in Figure 7.2. It starts with a set of a padding layer with a padding size of 3 to add extra pixels around the borders of an image to preserve the spatial dimensions of

the input and avoid the reduction in size, followed by a convolutional layer with a 7x7 kernel and a stride of 1, batch normalization and a leaky ReLU; the output next goes to the similar set of layers without a padding operation and with convolutional layers with smaller kernels and different stride. The output of the third set of layers enters the key component of the network that is a residual block. A residual block is composed of a stack of layers, and the output of the previous layer is added to another deeper layer in the block. This connection known as the shortcut or the skip-connection allows for information to bypass intermediate layers within the block. Every residual block encompasses 2 parts: first stack of layers consists of a padding layer with a padding size of 1, a convolutional layer with a kernel size 3x3 and a stride of 1, followed by a batch normalisation layer and a leaky ReLU, the second part is identical to the first one except it does not contain a leaky ReLU. The output of such a block is added to the output of the next block and so on. Depending on the size of the input image (we explain the details further in this section) we use 9 residual blocks for the semi-3d input, and only 2 blocks for the 3d-patch input. The decoding part of the ResNet is identical to the encoding part with a transposed convolutional layer instead of a convolutional one. In the final layer of the network, a Tanh activation function is applied to the output.

Discriminator

The discriminator is a neural network that acts as a binary classifier, distinguishing between real and generated data. Its main purpose is to assess the quality of the generated samples and provide feedback to the generator to improve its performance.

In traditional GAN models, the discriminator provides a single scalar output that represents the likelihood of the entire input being real or fake, however, in the context of image-to-image translation it is beneficial to generate outputs based on assessing local details rather than focusing on global image-level.

We followed the idea proposed in Isola et al., 2017 where they introduced a PatchGAN. The key feature of the PatchGAN is that instead of receiving the entire image as input, it takes as input patches of fixed size. By operating at the patch level, the PatchGAN discriminator captures local information and assess the realism of image more effectively.

Figure 7.3 outlines the PatchGAN architecture we used for our task of synthetic PET generation. It consists of five convolutional layers with a kernel of 4x4 and stride of 2 (apart from the last two layers), batch normalization layers and a Leaky ReLU. The receptive field of the discriminator is 70x70 and this is similar to manually dividing the entire image into overlapping patches of size 70x70. Each element of the discriminator output signifies whether a corresponding patch is real or fake. For the 3D-patch approach each feature in the final output feature map includes all of the input pixels.

Both generator and discriminator models were written by using PyTorch version 1.3.1 and we took python code provided by J.-Y. Zhu et al., 2017 as a baseline.

7.1.2.2 Data ingestion and preprocessing

Two-dimensional slice-based models taking one single slice as input allows capturing the global spatial context but inherently fail to leverage context from adjacent slices, unlike 3D models which can lead to improved performance but comes with a heavy computational and data cost.

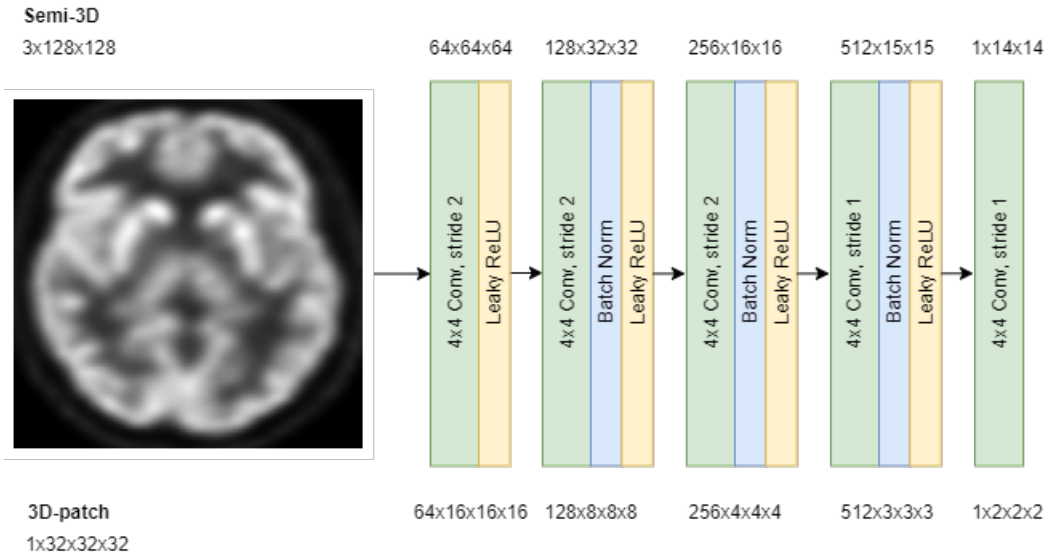

FIGURE 7.3 – PatchGAN architecture for a discriminator in either semi-3D or 3D-patch setting.

As a compromise, we consider two configurations depending on the size and shape of the input data:

— ***semi-3D*** models which receive three adjacent transverse slices as input (each slice corresponding to one channel)

— ***3D-patch*** models where we feed 3D mini patches extracted from the original 3D images into the network

For the semi-3D configuration, the original co-registered 3D images of size 157x189x136 with 1 mm^3 isotropic voxels are normalized and cropped into samples of size 128x128x136. For the 3D-patch configuration, the original 3D images are first resized to 160x192x160 so that we can extract sets of mini-volumes each of size 32x32x32.

Image reconstruction

In the semi-3D configuration, the whole 3D image is reconstructed by stacking the generated transverse slices. In the 3D-patch setting, we crop the generated 3D patches so as to consider only their central part as it has been shown in B. Huang et al., 2018 that predictions for edge pixels have lower accuracy, thus we consider only areas with higher prediction confidence. All patches are then stacked to reconstruct the 3D volume. For both semi-3D and 3D-patch configurations, we finally apply Gaussian smoothing as a post-processing to tackle with "border" effect that may occur when stacking either slices or mini-volumes. As a lightweight normalisation, we also perform standard histogram matching by adjusting intensity distribution of any fake PET to that of a randomly chosen PET image of the database that served to train the GAN synthesis model and considered as the reference image.

7.1.3 Visual quality metrics

The following metrics were used to objectively assess the quantitative score of translated images compared with the true image as reference:

Mean squared error (MSE) estimates the mean pixel-based error between the generated (x) and ground truth (y) images as

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} |x(i,j) - y(i,j)|^2 \tag{7.5}$$

Peak Signal to Noise Ratio (PSNR) computes the ratio between the maximum possible power (pixel intensity value) of the generated image y' and the MSE defined in 7.5 characterizing the power of corrupting noise that affects the fidelity of its representation, as

$$PSNR = 20log_{10}\left(\frac{max_x}{\sqrt{MSE}}\right) \tag{7.6}$$

where max_x is the maximum possible pixel value of image x.

Structural similarity index metric (SSIM) from Z. Wang et al., 2004 estimates the perceptual difference between two images using the mean (μ) and standard deviation (σ) over pixel values of the generated (x) and ground truth (y) images. Two variables, $c_1(= 0.01L)^2$ and $c_2(= 0.03L)^2$, are included to stabilize the division with low denominator, with L being the dynamic range of the pixel values, as

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1) \cdot (2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1) \cdot (\sigma_x^2 + \sigma_y^2 + c_2)} \tag{7.7}$$

Learned Perceptual Image Patch Similarity (LPIPS) from R. Zhang et al., 2018 is used to judge the perceptual similarity between two images and has been shown to match human perception.

It computes the distance between the activation maps of the generated (x_0) and ground truth (x) images for a pre-defined network.

$$d(x, x_0) = \sum_l \frac{1}{H_l W_l} \sum_{hw} \left\| w_l \odot (\hat{y}^l_{hw} - \hat{y}^l_{0hw}) \right\|_2^2 \tag{7.8}$$

Where $\hat{y}^l_{hw}$ and $\hat{y}^l_{0hw}$ are the extracted feature stacks from L layers normalized in the channel dimension. As a pre-defined network, we used AlexNet.

7.1.4 Experiments

For the semi-3D approach, we explore in total 4 variants of GANs for paired examples based on Simple-GAN or Cycle-GAN architectures both with and without the proposed MSE extra loss term. For the 3D-patch approach, we performed experiments with Simple-GAN and Cycle-GAN with MSE loss included as earlier experiments for semi-3D approach demonstrated noticeable improvements in models performance once MSE loss term was added to the total loss.

For the synthetic PET images generation we used the DB_{C1} database introduced in section 5.2 with a total amount of 35 controls with paired series of FDG PET and T1 weighted MRI scans. We compare the performance of each model based on the visual metrics derived in section 7.1.3.

A 4-fold cross-validation performance study is conducted based on 26 controls in the train set and 9 controls in the validation set. During the training, Structural Similarity Index (SSIM), as defined in section 7.1.3, between real and synthetic validation

Experiment 1: GANs comparison for PET synthesis				
Steps description	Model(s)	DB_{C1}	DB_{C2}	Outputs/results
1) Train semi-3D and 3D-patch GAN models in 4-fold cross-validation manner	Simple GAN Cycle-GAN	✓		Table 7.2
2.1) Train the best semi-3D GAN model	Cycle-GAN with MSE loss	✓		semi-3D-GAN model
2.2) Train the best 3D-patch GAN model	Cycle-GAN with MSE loss	✓		3Dpatch-GAN model
3.1) Generate $DB_{C2}^{Fake} = DB_{C2}^{semi3D}$ fake PET images	semi-3D-GAN		✓	DB_{C2}^{semi3D}
3.2) Generate $DB_{C2}^{Fake} = DB_{C2}^{3Dpatch}$ fake PET images	3D-patch-GAN		✓	$DB_{C2}^{3Dpatch}$

TABLE 7.1 – Experimental steps and datasets used to generate synthetic PET images from T1 modality

images serves as a quality metric to define the optimal configuration. During validation phase SSIM is computed on the whole image size for the semi-3D approach (where we first resize all validation image slices from the shape of 128x128 to the original shape of 157x189 and then estimate the SSIM metric) and on a patch level for the 3D-patch, and averaged over all controls in a validation set.

Once the 4-fold cross-validation experiments are complete, we then choose the two best model configurations and train the final versions of models for the generation of synthetic PET images for the DB_{C2} database for which original PET images are missing. Detailed step-by-step breakdown of the experimental procedures, including the specific dataset used at each stage and the corresponding outputs is provided in Table 7.1.

Implementation details of the GANs training for each type of input data is as follows:

semi-3D approach

— 46 triplets of adjacent slices per control training samples are extracted thus resulting in around 1 200 training samples for each model.

— The models are trained for a maximum of 200 epochs with a batch size of 5 and Adam optimizer.

— learning rate of 0.0002 is kept constant up to 100 epochs and linearly decayed to zero over the next 100 epochs.

3D-patch approach

— 6 069 mini-patches of size 32x32x32 are extracted for each control training subject with a stride of 8, thus leading to more than 200 000 training mini-volumes.

— The models are are trained for a maximum of 100 epochs with a batch size of 10 respectively and Adam optimizer.

— The learning rate of 0.0002 is kept constant.

7.1.5 Results

In this section, we delve into the experimental results obtained from training our GAN models with several objectives. Firstly, we aim to examine the loss curves of the GAN models to identify their learning efficacy over the training epochs. Secondly, we evaluate the SSIM metric as a criterion for terminating the training process. By comparing the SSIM scores across different epochs, we seek to establish a robust way of identifying the optimal stopping point that ensures high-quality image generation. Lastly, leveraging the insights from analyzing the loss curves and SSIM scores, we proceed to select the best epoch and use the model's weights of the best epoch to generate DB_{C2}^{Fake}.

7.1.5.1 Comparative Analysis of Loss Functions in CycleGAN Experiments

The success of the deep model training relies heavily on the choice and design of loss functions. In section 7.1.1 we provided mathematical formulation of losses chosen for classical GAN and Cycle-GAN models. In this part, we present graphical visualizations and analysis of losses contributions to the training process.

Since there are two networks trained at the same time that are competing against each other, the problem of convergence in GANs is one of the most challenging problems. The situation where both the generator and the discriminator are stabilised and are producing consistent results is hard to achieve. One reason behind this issue is that as the generator improves over epochs, the discriminator's performance declines due to its difficulty in distinguishing between real and fake instances.

For this reason not only we experimented with hyperparameters, but also introduced SSIM metric tracking to obtain an objective measure of the quality of the PET images. Higher SSIM scores indicate better visual similarity to the real data, indicating that the GAN is producing more realistic results. The SSIM metrics takes into account perceptual factors that might not be explicitly captured by the GAN loss functions alone.

Simple GAN

Figures 7.4-7.6 show the progression of loss functions over epochs. By examining the first two images, we can observe the generator and discriminator losses on both train and validation images, ensuring that the model does not suffer from overfitting. The third image presents the validation loss of both the generator and discriminator, along with the SSIM metric calculated on the validation images. We utilize this metric as a stopping criteria during training - we take the best weights of the model from the epoch where SSIM reached its maximum.

Both the generator and the discriminator demonstrate expected behavior: the training and validation adversarial losses have relatively high values at the beginning, but as the training progresses, we can observe fluctuations indicating the learning process and the model's attempts to generate more realistic images. The train losses G(A) and D(A) decrease over time, and the validation losses follow the same pattern being just slightly higher than the train loss and not overfitting. In Figure 7.6 we show the behaviour of the validation losses for the generator and the discriminator together with the SSIM metric progress. The red line indicates the point at which the SSIM reached the highest value, here, both G(A) and D(A) losses have very small values.

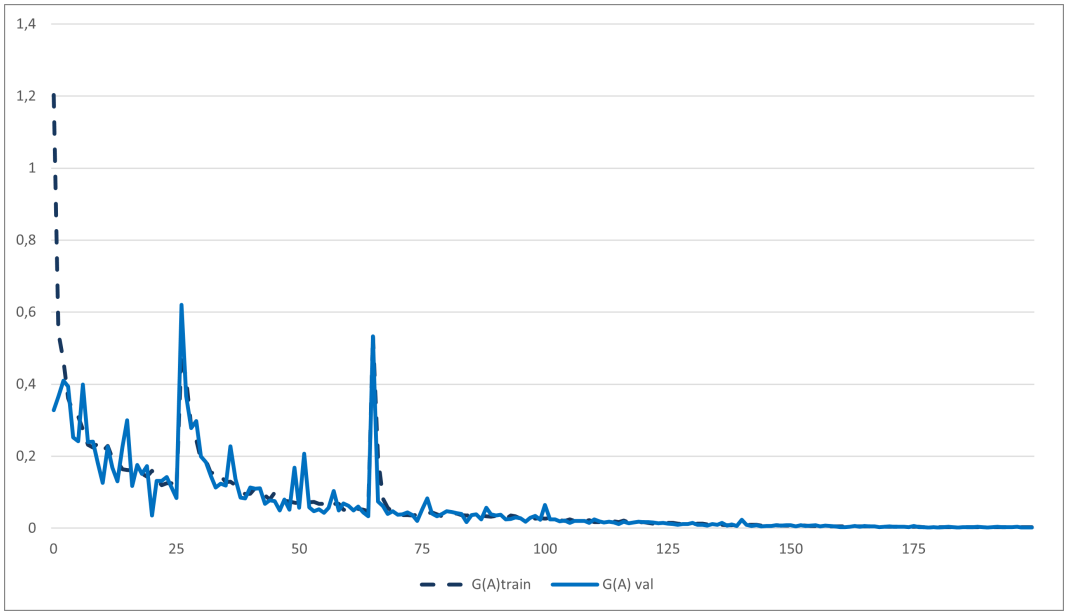

FIGURE 7.4 – Simple GAN Training Progress: Generator Loss G(A) train vs Generator Loss G(A) validation, A stands for the source domain T1 which is further generated into the target domain PET

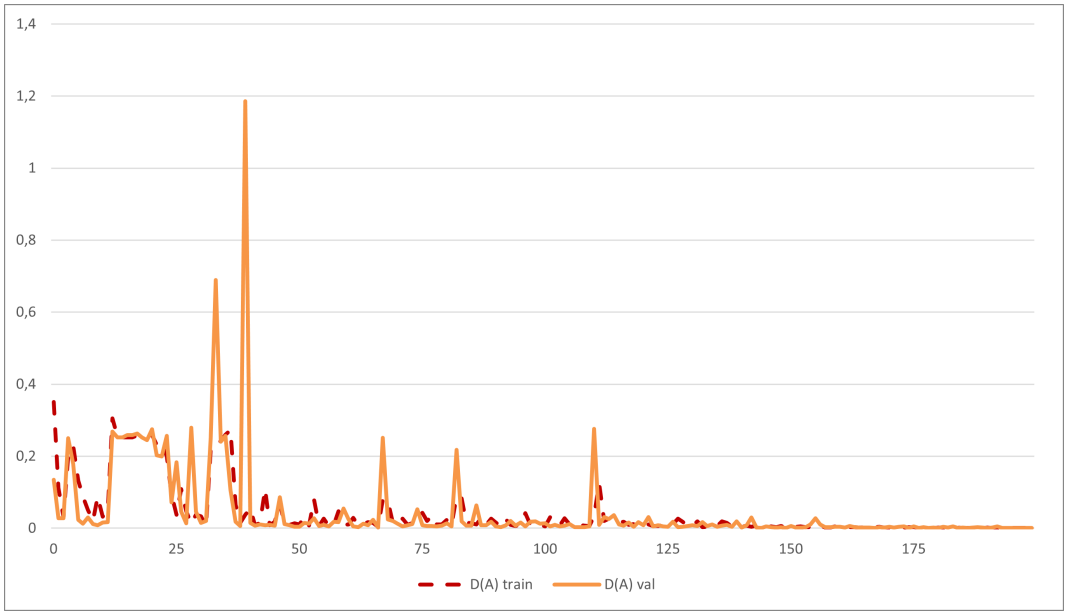

FIGURE 7.5 – Simple GAN Training Progress: Discriminator Loss D(A) train vs Discriminator Loss D(A) validation

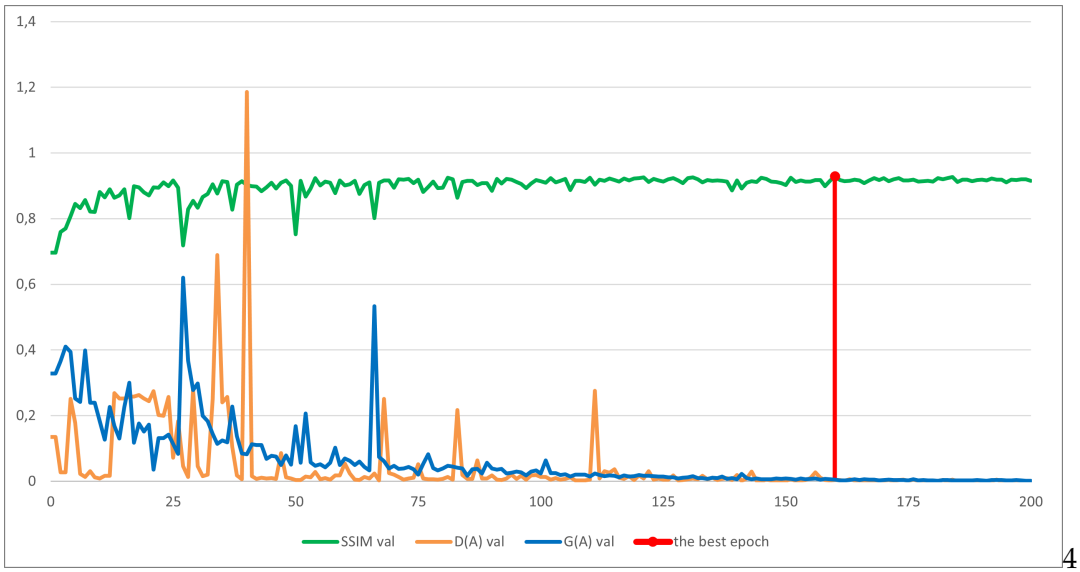

FIGURE 7.6 – Simple GAN Training Progress: SSIM, Generator Loss G(A), Discriminator Loss D(A) over epochs on validation images and the best epoch

Cycle GAN

In addition to the adversarial losses of the generator and the discriminator, Cycle GAN also use the cycle consistency loss, the identity loss. We optionally offer to include the MSE loss taking benefits from the fact of having a paired dataset. Figures 7.7-7.11 depict the changes in all the implemented losses as the model is trained.

The behaviour of loss functions in this case is more chaotic. For training and validation set of images (Figures 7.7 and 7.8), both the generator and the discriminator that translate images from T1 domain to PET domain struggle at the beginning with their losses growing, while the opposite operation of generating fake T1 images from real PET images is more successful as the losses G(B) and D(B) are both decreasing. After the epoch 130, we observe the abrupt fall of the G(A) loss together with the decrease of D(A) loss and a slight increase in losses for G(B) and D(B). This may be a result of the learning rate scheduler, where the learning rate had been decreased by a predefined step which led to the model's parameters better optimization. At the best epoch 189, we observe the highest SSIM value, at this point the generators and discriminators losses demonstrate a stabilized behavior. It is noteworthy to mention that D(A) loss at this point is around the 0.5 which means that the discriminator makes random guesses in defining the real PET images from the fake ones, and that is what we want. MSE, cycle consistency and identity losses demonstrate classical behaviour of training and validation losses both gradually go down with the validation loss reached its plateau after certain amount of epochs.

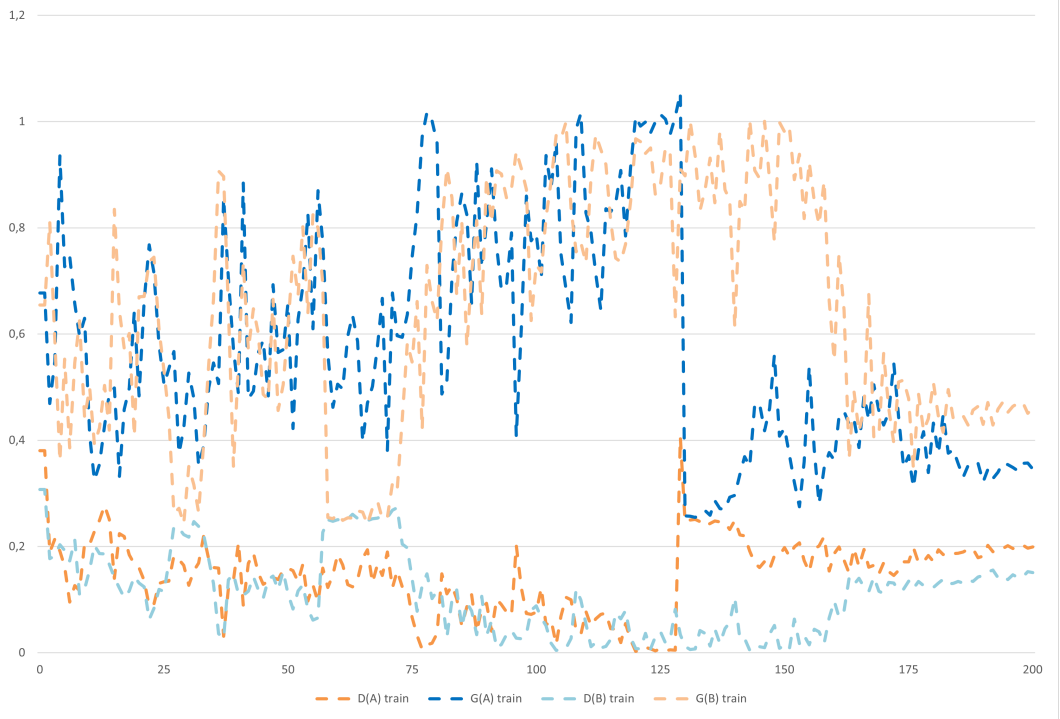

FIGURE 7.7 – Cycle GAN training progress: Discriminator Loss D(A), Generator Loss G(A), Discriminator Loss D(B), Generator Loss G(B) on train images

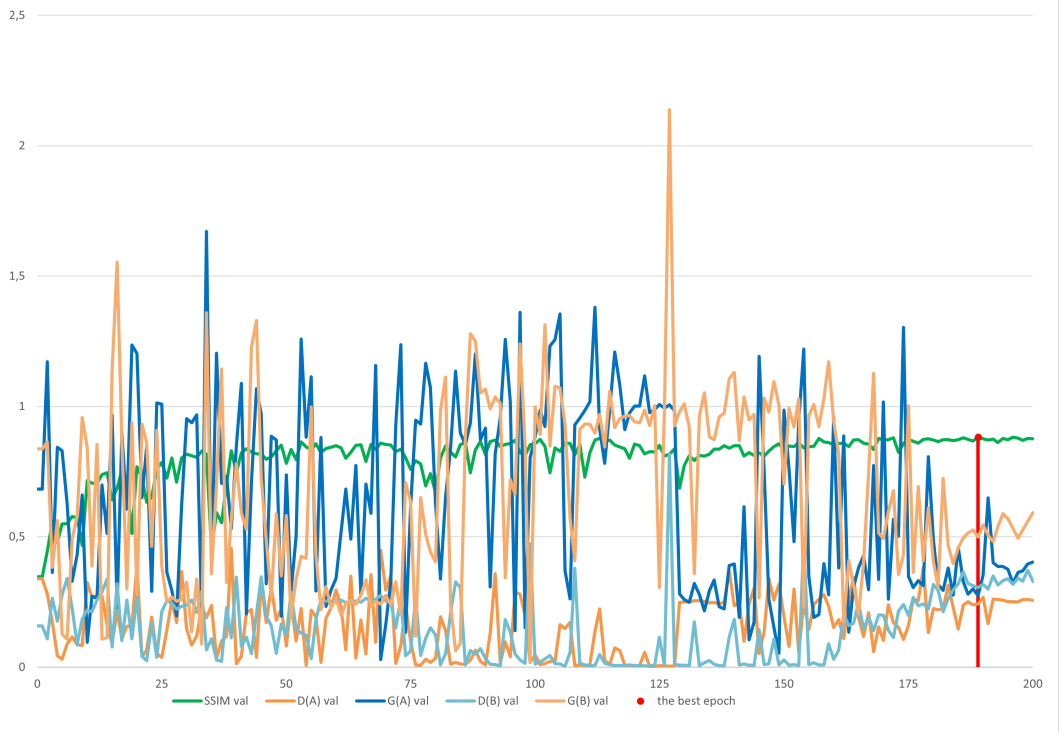

FIGURE 7.8 – Cycle GAN training progress: Discriminator Loss D(A), Generator Loss G(A), Discriminator Loss D(B), Generator Loss G(B), the SSIM metric on validation images, as well as the best epoch indicator

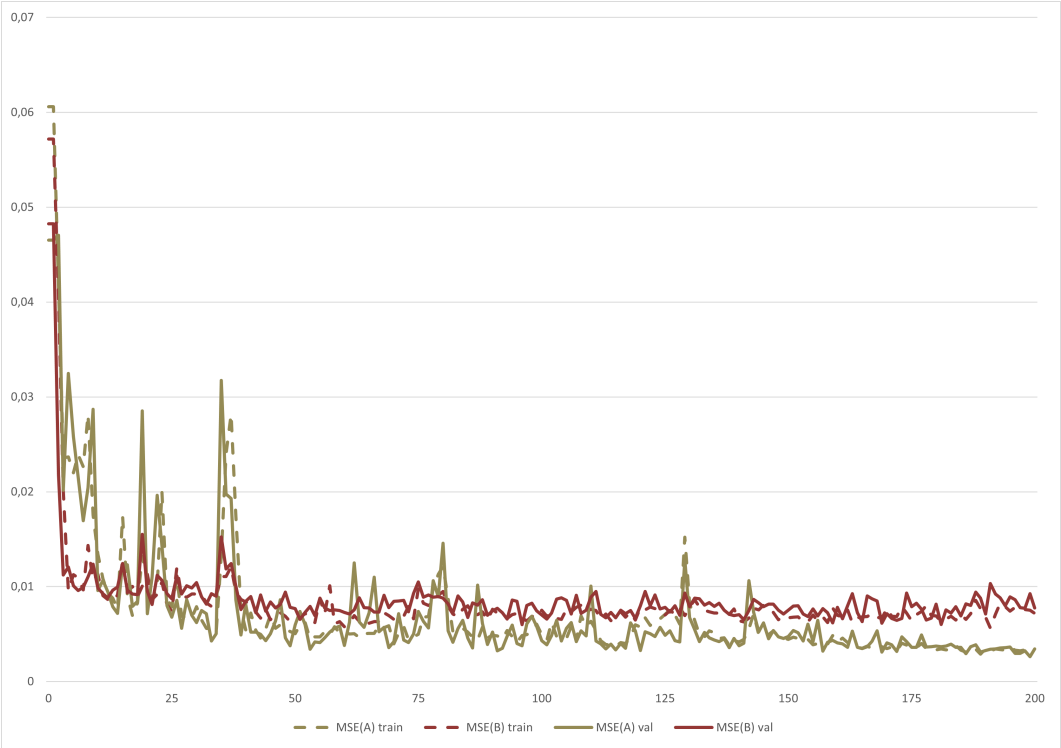

FIGURE 7.9 – Cycle GAN training progress: MSE loss between original PET images and their synthetic versions generated from T1 denoted as MSE(A) and MSE loss between original T1 images and their synthetic versions generated from fake PET denoted as MSE(B) on both train and validation images

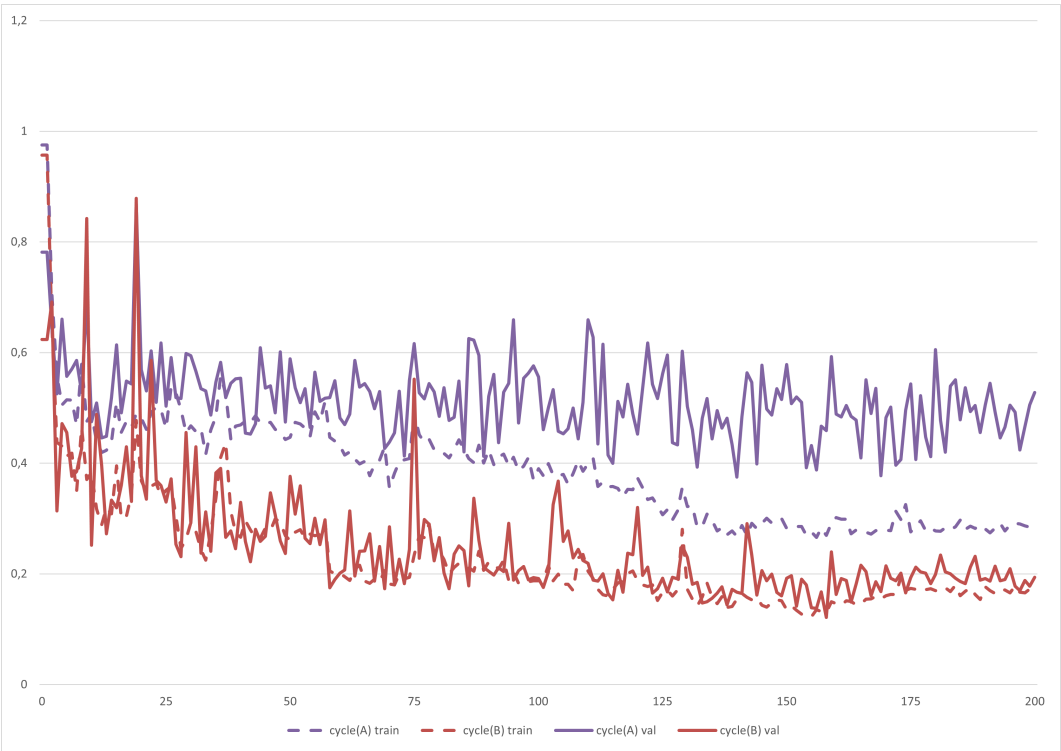

FIGURE 7.10 – Cycle GAN training progress: Cycle consistency loss for the generator A and the generator B for train and validation images

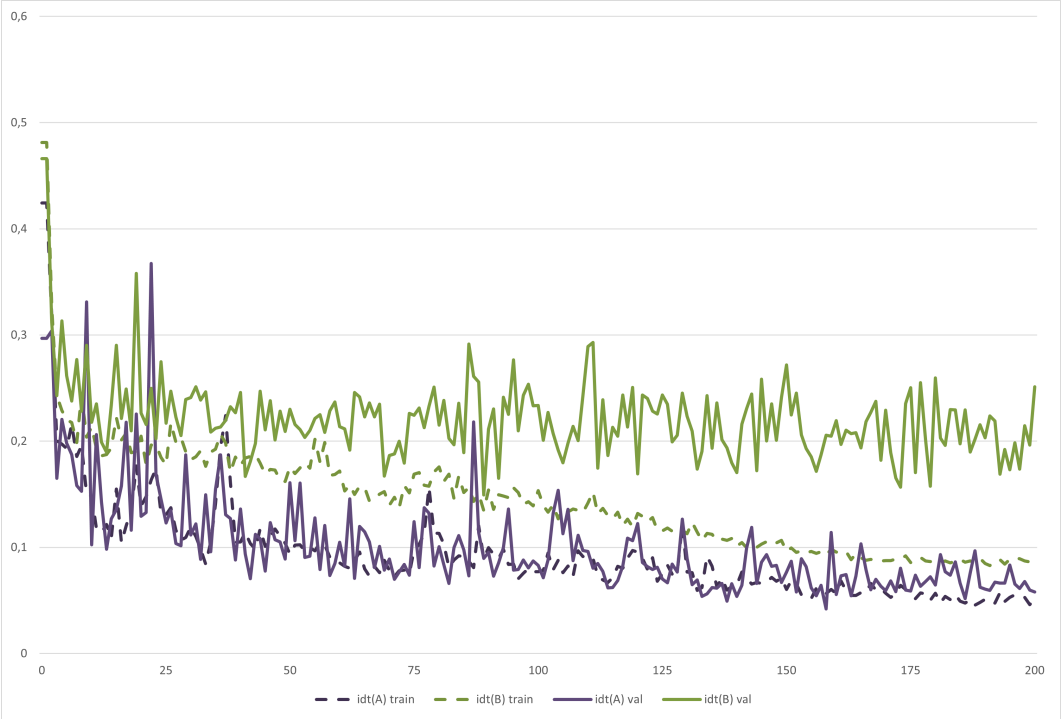

FIGURE 7.11 – Cycle GAN training progress: Identity loss for the generator A and the generator B for train and validation images

7.1.5.2 Evaluation of the visual quality of the synthetic PET images

After the best epochs are defined as described previously we can generate the synthetic images. We did it for all validation images in all cross-folds. Despite the high quality of generated slices for the semi-3D mode or patches in 3D-patch mode, the post-processing is required to soften the transition at the boundaries between slices/patches during the reconstruction of the entire PET image. For this a 3D Gaussian smoothing is applied on both semi-3D and 3D-patch reconstructed PET images to reduce border effects. Among a range of values between 0 and 3 mm FHWM, the value of 1.5 mm is shown to produce the best SSIM values.

We also noticed that after smoothing the resulting images appeared to be slightly brighter compared to original PET images. Figure 7.12 illustrates a comparison between histograms of original and synthesized PET images for a single control in the validation dataset. After the histogram matching the pixel intensities in a synthesized PET image matches closer the target histogram of an original image. In a real case scenario we would not have the original PET images for every T1 image of a patient, but we might have examples of PET images from desired PET scanners, that's why during the histogram matching processing we choose a random original PET image as a reference and not the original PET image of a given control. As a result our final reconstructed PET image exhibits a close resemblance with the original PET in terms of structural details and intensity values (Figure 7.13.)

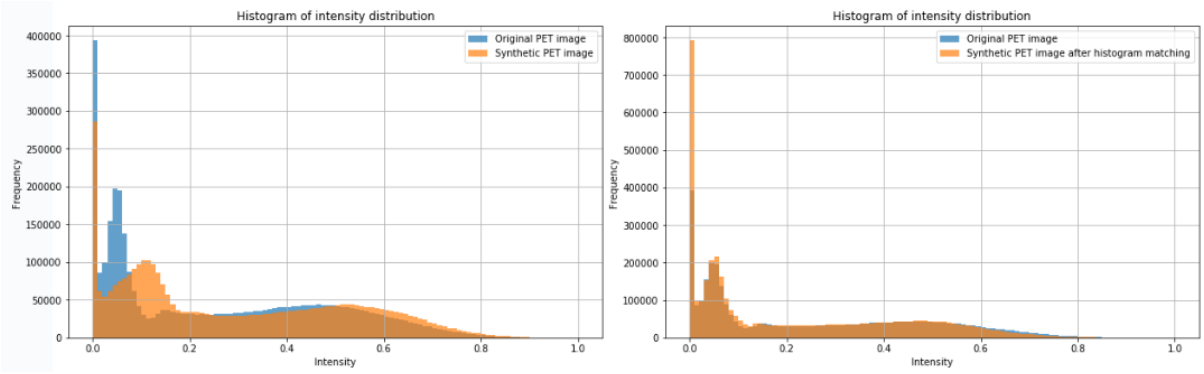

FIGURE 7.12 – A histogram of original and reconstructed PET images for a control in a validation set before and after applying histogram matching.

Table 7.2 reports the mean SSIM, Peak Signal to Noise Ratio (PSNR) and Learned Perceptual Image Patch Similarity (LPIPS) with corresponding standard deviation computed over all validation samples and all 4 folds for each of the six considered models. 3D-patch Cycle-GAN with MSE loss is shown to perform the best among the 6 models considered in this study. Two-tailed Wilcoxon signed rank tests yield no significant differences between the semi-3D and 3D-patch Cycle-GAN models with MSE loss for the PSNR (p-value = 0.79) metric while p-values of 0.0004 and $8.6\text{x}10^{-8}$ are achieved for the LPIPS and SSIM metrics, respectively, in favor of the 3D-patch method. Also note that our proposition to add the MSE loss term to the Cycle-GAN global loss allows a significant improvement of all three metrics.

In the following, we consider the best performing models of each configuration, namely semi-3D and 3D-patch Cycle-GAN models with MSE loss.

TABLE 7.2 – Average visual quality metrics computed on the 35 synthetic PET exams generated from T1 MRI of 35 healthy subjects.

Configuration	Model	SSIM	PSNR	LPIPS
semi-3D	Simple-GAN	0.825 ± 0.02	21.489 ± 0.89	0.033 ± 0.006
	Simple-GAN with MSE loss	0.880 ± 0.02	23.542 ± 1.43	0.022 ± 0.006
	Cycle-GAN	0.884 ± 0.02	23.700 ± 1.43	0.023 ± 0.006
	Cycle-GAN with MSE loss	$\mathbf{0.886 \pm 0.019}$	$\mathbf{23.742 \pm 1.26}$	$\mathbf{0.021 \pm 0.004}$
3D-patch	Simple-GAN with MSE loss	0.883 ± 0.02	23.100 ± 1.38	0.021 ± 0.004
	Cycle-GAN with MSE loss	$\mathbf{0.897 \pm 0.019}$	$\mathbf{23.82 \pm 1.72}$	$\mathbf{0.019 \pm 0.005}$

Further in our work we want to explore the practical aspects of using synthetic data applied on real case of epilepsy detection. For this application, we use DB_{C2} as the test set and generate missing PET images with the two models (semi-3D or 3D-patch) trained on the DB_{C1} database. In order to do that, we have to train each of the two models using as much data as possible. We split our original dataset DB_{C1} into train (29 controls), validation (3 controls) and test (3 controls) sets. This partitioning allows us to have more data for training, and the independent test controls to verify that the weights found based on SSIM on validation data are optimal for synthetic PET data generation for a database without original PET scans. The training parameters remain the same as for the 4-cross-fold validation experiments.

The loss progressions for the semi-3D and 3D-patch models are shown in Figures 7.14 and 7.15 respectively.

For the semi-3D configuration the loss behaviour is different from what we have seen with the cross-validation experiments. Both losses for the generator and discriminator B (with PET images as a source domain and T1 as an output domain) increase after epoch 74, while for the T1 to PET translation, the generator A and discriminator A compete against each other without showing much changes. Thus, we had to stop before the whole model starts to overfit.

The training of the 3D-patch model is less stable, and losses vary a lot. After the epoch 30, generator losses start to increase with the discriminator losses go down and the model produces less realistic patches of PET images. Thus, we stopped at the epoch 24 with all losses being relatively small and SSIM metric reaching its highest value.

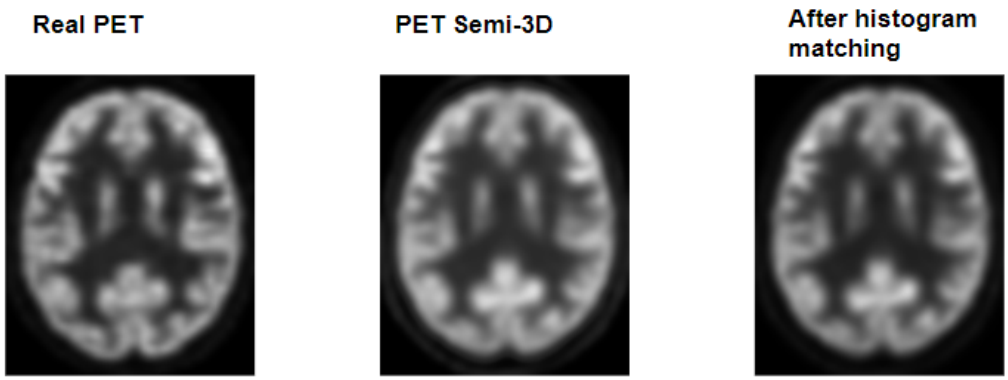

FIGURE 7.13 – Transversal slice for a control in a validation set: original image, before and after applying histogram matching.

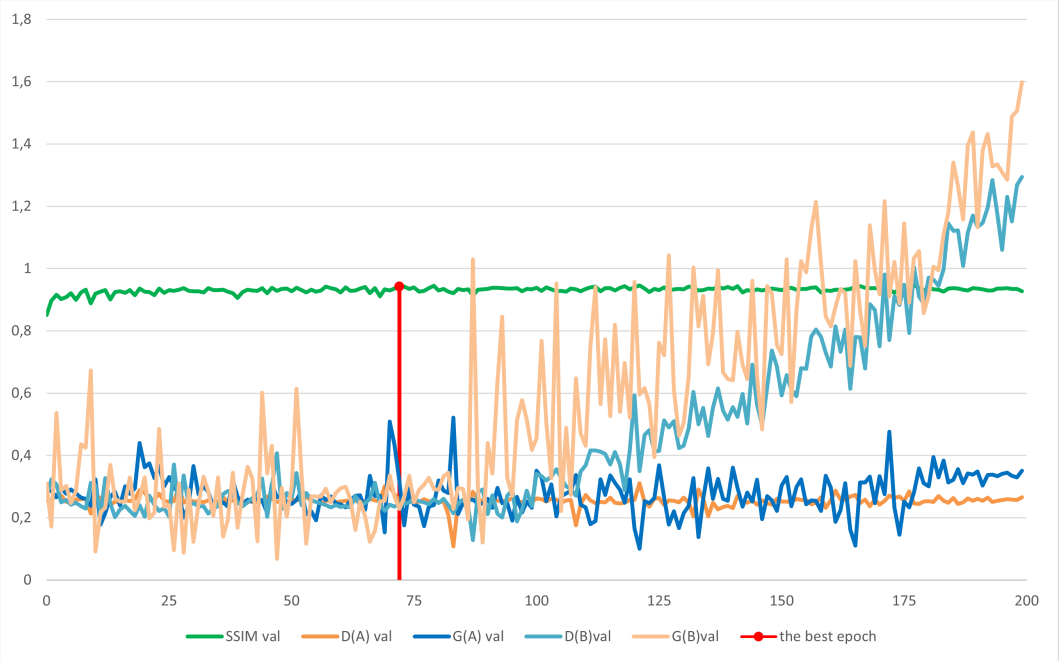

FIGURE 7.14 – Cycle-GAN semi-3D configuration. Generators and discriminators loss functions progress.

Synthetic PET data generated by the semi-3D and the 3D-patch configurations of Cycle-GAN models from the same T1 MRI of a control subject are illustrated in Figure 7.16 and compared with the reference PET image of this subject. Both models allow generating visually realistic FDG PET data.

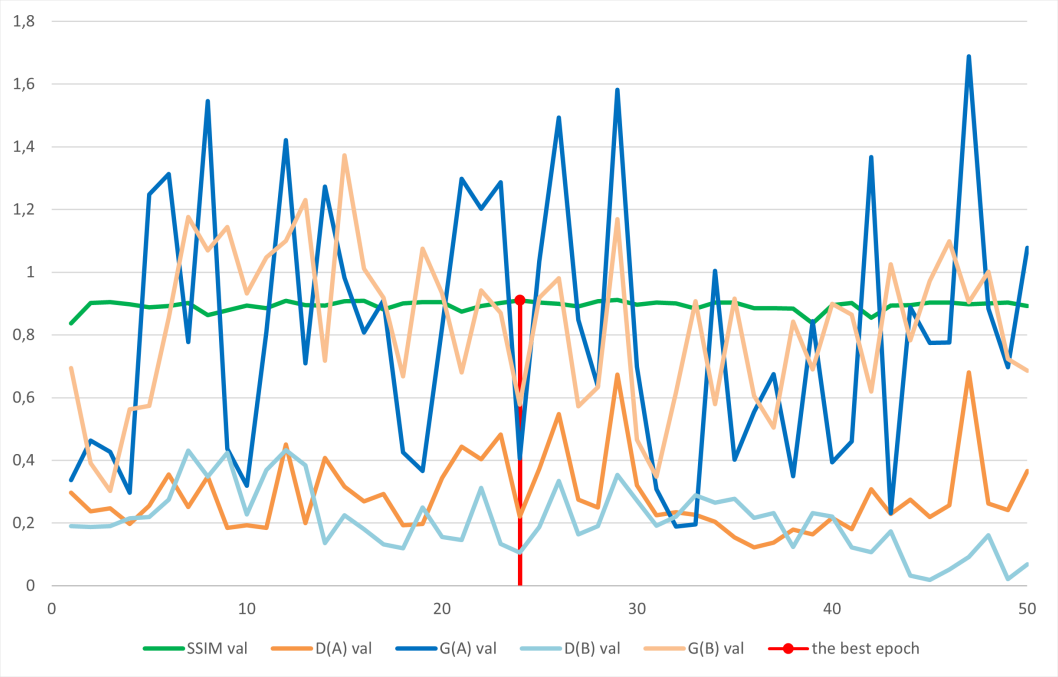

FIGURE 7.15 – Cycle-GAN 3D-patch configuration. Generators and discriminators loss functions progress.

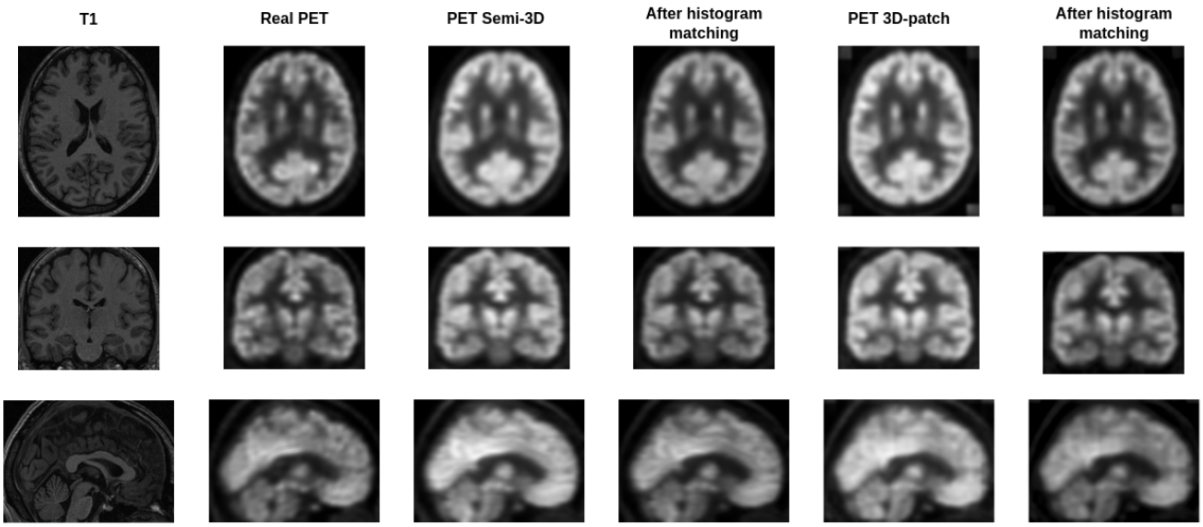

FIGURE 7.16 – Synthetic PET generated from T1 MRI of a test control.

7.2 Evaluation of synthetic PET data based on task-oriented quality metrics

As described in the previous section, synthetic medical data are aimed at augmenting or completing missing multimodality datasets that serve for different tasks of medical image processing, including the training of deep machine learning models. It is thus important that these synthetic data mimic as closely as possible true images. Most of the papers in the domain use standard quality metrics such as PSNR or SSIM to evaluate the performance of the proposed architectures Abu-Srhan et al., 2021; Gao et al., 2022; Qin et al., 2022; Sikka et al., 2018. Although these metrics allow comparing noise and texture characteristics of the reference and synthetic data, they can not determine if the pseudo-true images will perform similarly for the considered task at hand.

Only a few studies have gone beyond the quantitative intensity and noise wise similarity assessment.

Some studies evaluated the added value of synthetic data to perform quantification tasks Dewey et al., 2019; Wei et al., 2019. As an example, Dewey et al., 2019 showed that that the generative DeepHarmony model aiming at harmonizing brain MR T1 data across multiple scanners of different vendors allowed a significant decrease in brain volume measurement error in a longitudinal study. Kläser et al., 2021 showed that pseudo-CT generated from MR image recovered more accurate quantitative HU than a multi-atlas propagation approach and subsequently led to a significant improvement in the PET reconstruction error.

Other studies included synthetic data to the training of segmentation J. Liu et al., 2023; Skandarani et al., 2023 or classification Chadebec et al., 2022; Pan, Chen, et al., 2021; Pan et al., 2018; Pan, Liu, Xia, et al., 2021 networks. Liu et al J. Liu et al., 2023, for instance, generated pseudo T1 Gadolinium images from T1, T2 and FLAIR images of the BraTS dataset and compared tumor segmentation accuracy on true and synthetic T1 Gd images.

Regarding classification, Chadebec et al., 2022 recently trained a variational autoencoder (VAE) to generate T1-weighted (T1w) MR brain images brain that were then used to augment the training set of a CNN whose task is to differentiate Alzheimer's disease (AD) patients from cognitively normal (CN) subjects.

Finally, getting closer to detection tasks targeted in this study, Yaakub et al., 2019 generated synthetic pseudo-normal FDG PET images from MR T1 images of a control subjects. In a second phase, they used this generative model to generate control-like pseudo FDG PET images from MR T1 exams of epilepsy patients, which were then subtracted from the patient true PET images to localize hypometabolic regions.

7.2.1 Out-of-distribution problem formulation

A task-oriented quality evaluation is desirable if the synthetic data are to be used for training machine learning models. Performance of these models indeed rely on the strong assumption that data are identically distributed, meaning that they are sampled from the same distribution. This means that the distributions of test data should be sampled from the same distribution as the data used to train the model, but also that data of each group (e.g training or testing) should be sampled from a unique distribution. A violation of this assumption, referred to as domain shift problem, may result in a performance decrease of the ML model.

One way to evaluate that synthetic data follow the same distribution as the reference true data is to compare performance of the clinical task at hand with and without synthetic data, as done for example in Chadebec et al., 2022.

Another way is to evaluate if the synthetic data are out-of-the distribution (OOD) samples. A few OOD detection methods have been proposed over the past few years, mostly for deep classification networks. Gonzales *et al* González et al., 2022 proposed a lightweight method to detect when deep segmentation models silently fail on out-of-distribution (OOD) data, mainly due to domain shift problem. This method exploits the Mahalanobis distance in the feature space of the deep segmentation model to derive epistemic uncertainty maps which are shown to correctly signal when a model prediction should not be trusted. It demonstrated good performance for different segmentation tasks including the Covid-19 lung CT lesion segmentation challenge gathering multi-centre scans as well as to MRI segmentation tasks of the hippocampus and the prostate.

7.2.2 Out-of-distribution metrics definition

As described in Section 4.3, the main principle of UAD models is to learn a model of normal control population by learning to compress and recover healthy data based on autoencoder-like networks. Once trained, anomalies present in pathological data can be detected as outliers from the modeled, normative distribution, either by computing reconstruction error between the true and reconstructed images *in the image space* or by training uniclass discriminant or generative models *in the latent representation space*. To evaluate if the distribution of the generated synthetic PET data matches that of true PET data, we derive out-of-distribution (OOD) detection metrics fitted to the specific unsupervised anomaly detection task at hand.

The first metric is the global MSE (as defined in eq. 7.5) averaged over all pixel-based error between any input image u (either fake or true) to the autoencoder AE and its reconstruction $AE(u)$. This metric quantifies if the fake FDG PET data mimic in-distribution (ID) true FDG PET data *in the image space*, thus validating the use of fake PET to train UAD models based on the reconstruction error based UAD models.

The second OOD metric, inspired from the idea developed in González et al., 2022 is derived from the computation of the Mahalanobis distance d_m between the latent representation z_u of any input image u to the autoencoder AE and the distribution of this latent variable in the normal training population data, as

$$d_M = (z_u - \mu)^T \Sigma^{-1} (z_u - \mu) \tag{7.9}$$

where z_u is the latent representation for image u, μ and Σ are the empirical mean and covariance, respectively, computed over the latent representation z_i of the N elements of the training data set as:

$$\mu = \frac{1}{N} \sum_{i=1}^{N} z_i, \Sigma = \frac{1}{N} \sum_{i=1}^{N} (z_i - \mu)(z_i - \mu)^T \tag{7.10}$$

Dimension of the data inputted to the AE autoencoder (2D or 3D images or patches) drives the dimension of the d_M distance vector. It is indeed one-dimensional if the UAD model is designed to extract one unique z value per 3D image, while it equals the dimension of the original image if the UAD model is designed to generate

one z value per voxel of the inputted 3D image. In the latter case, following González et al., 2022, we average over all voxels to obtain a global metric, also referred to as d_M to simplify notations, whose value is normalised between the minimum and doubled maximum values for ID train data.

The global d_M metric quantifies if the fake FDG PET data mimic in-distribution (ID) true FDG PET data *in the latent space*, thus validating the use of fake PET to train UAD models performing the detection task in the latent representation space.

Comparing the distributions of MSE and Mahalanobis distance d_M of true y and fake x images will allow detecting any domain shift in the image and latent spaces, respectively.

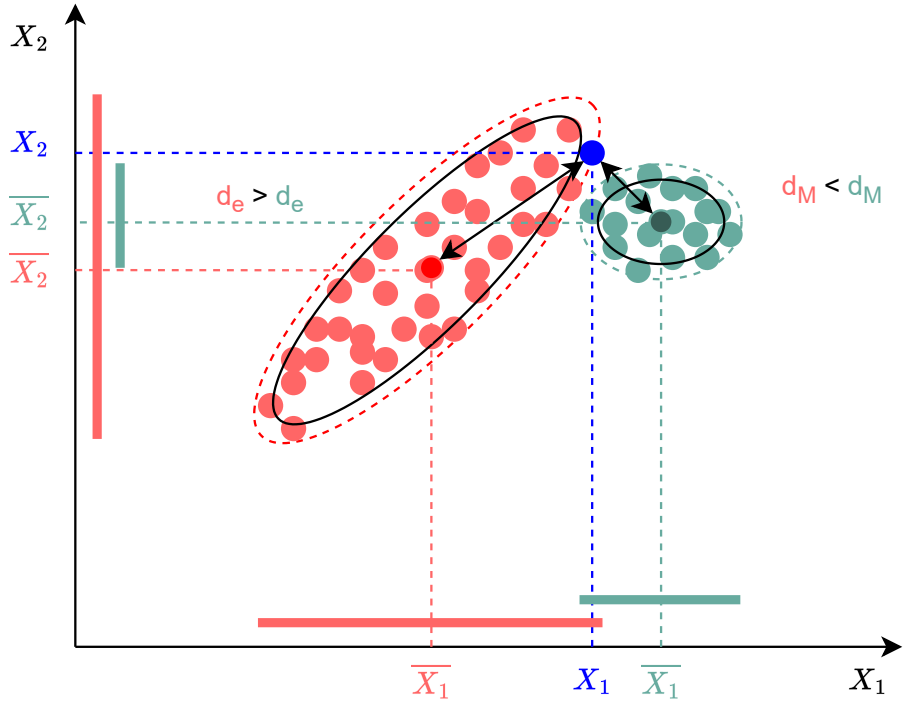

FIGURE 7.17 – Example of the difference between the Euclidean distance (d_e) and Mahalanobis distance (d_M) in 2D space for two clusters and a testing point X. is the vector of the average values for all variables (centroid).

With figure 7.17, we would like to illustrate why the Mahalanobis distance might be a better choice compared to the Euclidean distance. We first remind that the Euclidean distance d_e is defined as

$$d_e = \sqrt{(X - \bar{X})(X - \bar{X})^T} \tag{7.11}$$

In Figure 7.17, d_e from a testing point X (blue) to the centroid of the elongated (red) cluster is higher than to the round (green) cluster. On the contrary, the Mahalanobis distance d_M (7.12, 7.13) that considers the covariance between dimensions is smaller than the elongated cluster. The Mahalanobis distance is a more robust distance metric than the Euclidean distance when dealing with data that may have correlated features.

$$d_M = \sqrt{(X - \bar{X})\Sigma^{-1}(X - \bar{X})^T}, \tag{7.12}$$

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix} \tag{7.13}$$

7.2.3 Experiments

We perform the OOD detection on PET images to explore if synthetic images look alike to the real ones and to ensure that they are representative of the true distribution. The primary hypothesis suggests that the inclusion of synthetic images will serve as an augmentation to the original dataset that, in turn, will lead to the model's improvement in detecting epilepsy.

Another aspect of our research work is to estimate the ability of synthetic data to be used instead of real ones in case of missing data. For this part, the input of the siamese encoder consists of T1 and PET FDG images of normal subjects combined as channels. Our hypothesis is that the model trained on T1 and synthetic PET images can show comparable performance as a model trained on T1 with real PET data. To confirm this hypothesis, we build on OOD detection techniques to derive metrics assessing if the distributions of T1+true PET and T1+generated synthetic PET images look alike both in the image domain and in the latent representation domain of the UAD model under consideration. Then, we compare the performance of this UAD model trained either on normative paired MRI and true PET or on normative paired MRI and synthetic FDG PET for the detection task of subtle epilepsy lesions in T1 MRI and PET patient exams.

To retrieve OOD metrics we designed experiments to detect OOD samples from PET images solely and from a combination of T1+PET images. Our dataset construction and analysis involved distinct phases for PET and T1+PET experiments, utilizing a combination of real and synthetic data samples.

7.2.3.1 Dataset construction

PET experiment: We assembled a dataset comprising 35 real PET data samples from DB_{C1} and 40 synthetic PET data samples generated from DB_{C2} MRI images. These synthetic samples were produced using CycleGAN, augmented with an MSE loss, in two distinct modes: 3D-patch ($DB_{C2}^{3Dpatch}$) and semi-3D (DB_{C2}^{semi3D}). This dataset was divided into 15 folds, each containing 60 images for training and 15 for testing, ensuring a balanced representation of real and synthetic images in the test set. We also include 17 patients from DB_{ep1} (see in Table 5.2) in the test set.

T1+PET experiment: For the T1+PET setup, we created a training dataset including 35 pairs of synthetic PET images (either from DB_{C2}^{semi3D} or $DB_{C2}^{3Dpatch}$) alongside their corresponding T1 MRI images. The test set comprised remaining 5 pairs of T1 and synthetic PET images for evaluating reconstruction error and Mahalanobis distance, plus 35 pairs of T1+real PET from DB_{C1}. Additionally, the test set included paired T1+PET scans of 18 patients DB_{ep2} (see in Table 5.2).

7.2.3.2 Protocol for OOD metrics gathering

1. *Siamese Network Training:* For both experiments, we utilized a Siamese network architecture, trained on the respective datasets (technical details further in this chapter in section 7.3.1.1). For PET experiments, the networks (UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$) were trained on the combined real and synthetic PET images. Similarly, for the T1+PET experiment, the networks (UAD_{mripet}^{semi3D} and $UAD_{mripet}^{3Dpatch}$) were trained on the paired images of T1 and synthetic PET. This training facilitated the calculation of reconstruction error (MSE) and Mahalanobis distance for OOD detection.
2. *Assessment of OOD Metrics:* Utilizing the trained networks, we calculated the MSE between the input and reconstructed images and computed the Mahalanobis distance to the train distribution at the voxel level. For the T1+PET experiment, MSE errors for T1 and PET images were calculated separately and averaged to derive a single metric per subject. The Mahalanobis distances were similarly computed, allowing for an aggregate assessment of OOD metrics. The resulting Mahalanobis distances at voxel level are averaged over all voxel to obtain a single mean value at the subject level.
3. *Visual Representation and Analysis:* For both sets of experiments, we generated single mean MSE errors and mean Mahalanobis distances for each test subject, including controls and patients from respective datasets. These values are used to generate 2D graphics, enabling visual representation and analysis of the experimental results.

Refer to Experiment 2 in Table 7.3 and Experiment 3 in Table 7.4 to track inputs and outputs for either PET or T1+PET experiments.

Experiment 2: OOD detection in PET					
Steps description	Model(s)	DB_{C1}	DB_{C2}^{Fake}	DB_{ep1}	Outputs/results
2) Train UAD model on real PET images + fake PET images (DB_{C2}^{semi3D} and $DB_{C2}^{3Dpatch}$)	UAD model	✓	✓		UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$
3) Compare true and fake PET images, evaluate the synthesis with task-oriented OOD metrics	UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$	✓	✓	✓	Figure 7.18 and Figure 7.19

TABLE 7.3 – Experimental steps and datasets used to perform OOD detection in PET

7.2.4 Results

We showcase the results for OOD detection in PET images in Figures 7.18 and 7.19.

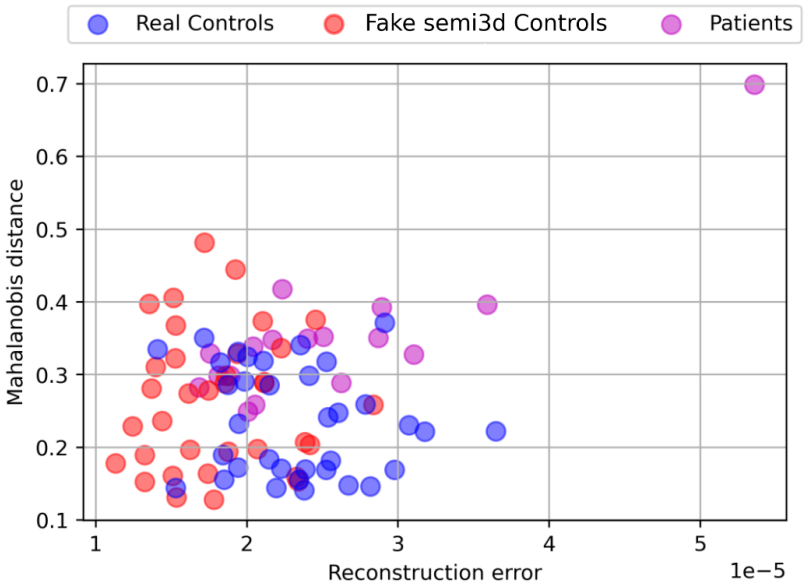

FIGURE 7.18 – OOD estimation for real PET, synthetic PET generated with a semi-3D configuration and PET patients.

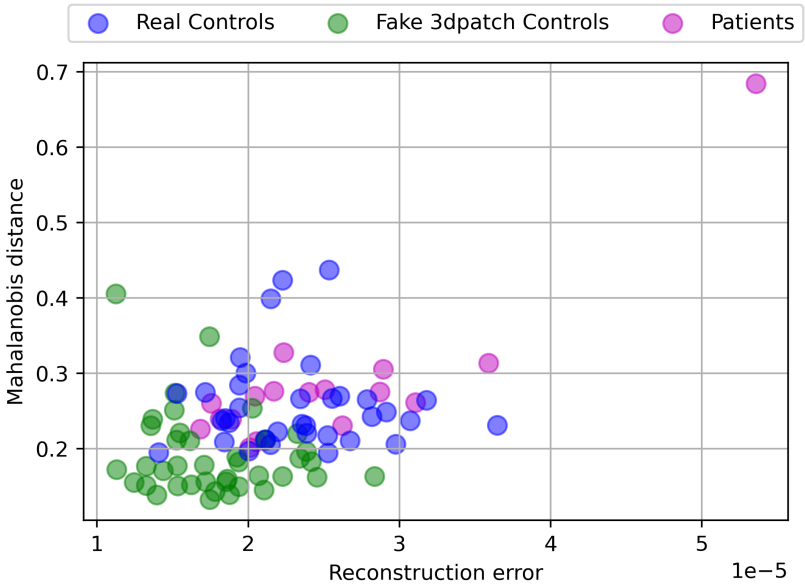

FIGURE 7.19 – OOD estimation for real PET, synthetic PET generated with a 3D-patch configuration and PET patients.

Experiment 3: OOD detection in T1+PET						
Steps description	Model(s)	DB_{C1}	DB_{C2}^{fake}		DB_{ep2}	Outputs/results
			35	5		
1) Tran UAD model on real MRI + real PET images	UAD model	✓				$UAD_{mripet}^{original}$
2) Train UAD model on real MRI + fake PET images (DB_{C2}^{semi3D} and $DB_{C2}^{3Dpatch}$)	UAD model		✓			UAD_{mripet}^{semi3D} and $UAD_{mripet}^{3Dpatch}$
3) Compare true and fake PET images coupled with true MRI, evaluate the synthesis with task-oriented OOD metrics	UAD_{mripet}^{semi3D} and $UAD_{mripet}^{3Dpatch}$	✓		✓	✓	Figure 7.20 and Figure 7.21

TABLE 7.4 – Experimental steps and datasets used to perform OOD detection in T1+PET

We see that the reconstruction error rates are lower in the both control groups with fake PET images with noticeable shift to the left for the DB_{C2}^{semi3D}. Mahalanobis distance show less variability between groups of real and fake PET images. For the 3D-patch setting, we can more clearly distinguish two clusters: control subjects with fake PET images and control subjects and patients with real PETs. In the case of semi3D-generated PET, this boundary is less noticeable, which allows us to conclude that the PET images obtained with UAD_{pet}^{semi3D} are more similar to real PET images.

The Mahalanobis distance is lower between the mixed train set and the test set with images in $DB_{C2}^{3Dpatch}$. This suggests that the fake PET images are more similar to the characteristics learned by the siamese network during training, possibly because the network learned to capture the common features present in the generated images. The higher distances between the train set and the test set with real PET images can be explained by greater variability in the features of real PET scans (due to a range of factors, including patient differences, radiotracers variability, scanner settings, environmental factors, etc.) compared to the training set. Generated fake PET images are more consistent in terms of the features, structures, and patterns that the Siamese network is sensitive to, and this could lead to a more consistent compressed representation and, consequently, more consistent Mahalanobis distances for the fake PET images as well as more consistent reconstruction errors.

In both cases, one patient turned out to be an outlier - Patient R showed high values of Mahalanobis distance and a high reconstruction error. It is worth noting that 11 of the 17 patients underwent PET scanning at the same hospital in Lyon (see Table 5.2, database DB_{ep1}), respectively, their PET images were taken on the same scanner, which allows us to say that these PET images are homogeneous. Patient R whose metrics are markedly different from the control subjects underwent PET scanning at a different hospital, we thus hypothesize that its PET images come from a different distribution. However, the remaining 5 patients (Patients D, V, AA, AD,

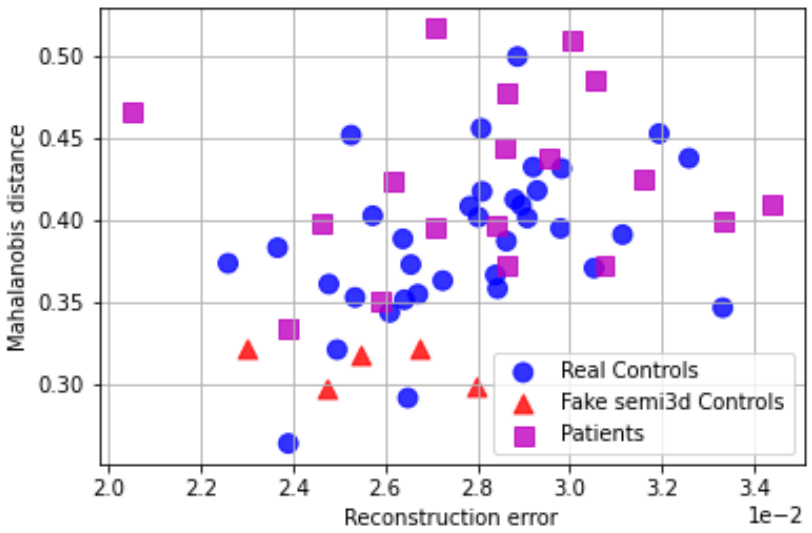

FIGURE 7.20 – Mahalanobis distance d_M and reconstruction error MSE on test subjects inputted to the $UAD^{original}_{mripet}$ model. Blue points correspond to the 35 real controls from DB_{C1}, purple squares to the 18 patients of DB_{ep2}, red triangles correspond to 5 test control samples of DB^{semi3D}_{C2}.

AE), whose data came from a different hospital than Lyon, showed no abnormalities compared to the control subjects. To what extent this affects the ability of the system to detect epilepsy we will examine further.

Results of the OOD detection in T1+PET images are presented in Figures 7.20 and 7.21. As for the previous experiment with real and synthetic PET images, they show the relationship between the mean normalized Mahalanobis distance d_M and the mean reconstruction error MSE estimated on test subjects from DB_{C1}, DB^{semi3D}_{C2}, $DB^{3Dpatch}_{C2}$, and DB_{ep2}. The blue circle symbols and the purple square ones report metrics computed on the real samples of DB_{C1} and 18 patients of DB_{ep2}, respectively, while the red and green triangles correspond to 5 remaining test control samples of DB^{semi3D}_{C2} and $DB^{3Dpatch}_{C2}$, respectively.

For all controls and patients, we observe the same order of reconstruction error denoting that *ID samples* (red and green triangles), *OOD samples* (blue circles for DB_{C1} controls and purple squares for DB_{ep2} patients) produce reconstructed images of similar quality. The difference is more noticeable in the values of the Mahalanobis distance computed in the latent space with average d_M values around 0.4 for real controls (blue circle) and patients (purple square) and 0.32 for the two series of test controls (green and red triangles). Also, note the higher variability for both metrics seen for real controls and patients. These results suggest that the distributions of DB_{C1}, DB^{semi3D}_{C2} and $DB^{3Dpatch}_{C2}$ control samples reasonably overlap in the image and latent spaces thus meaning that fake paired images of DB^{semi3D}_{C2} and $DB^{3Dpatch}_{C2}$ can be considered as inliers of the distribution of true control samples of DB_{C1}.

Also note that the patient and control scatter plots are well-mixed, thus meaning

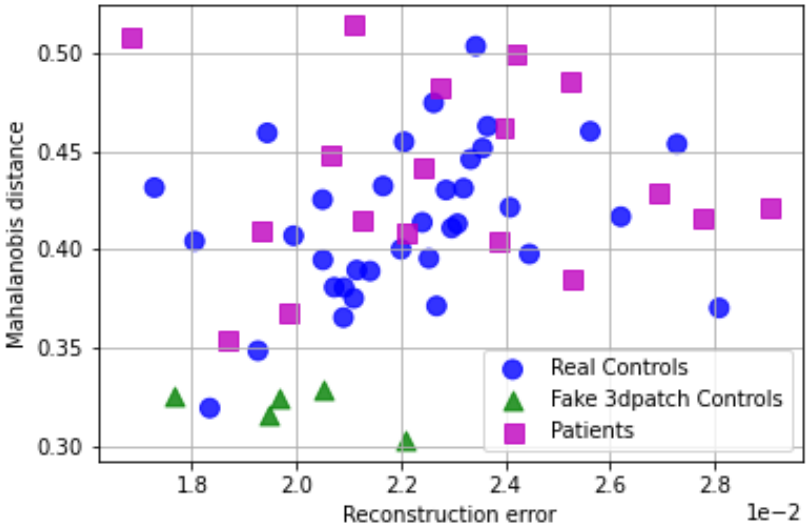

FIGURE 7.21 – Mahalanobis distance d_M and reconstruction error MSE on test subjects inputted to the $UAD^{original}_{mripet}$ model. Blue points correspond to the 35 real controls from DB_{C1}, purple squares to the 18 patients of DB_{ep2}, green triangles correspond to 5 test control samples of DB^{semi3D}_{C2}.

that the patients can not be considered outliers of the distribution of control samples. This underlines the difficulty of the detection task considered in this study, where epilepsy lesions are so subtle that they do not impact the global patient-level OOD metrics considered in this study.

7.3 Application of synthetic PET data to the training of a brain anomaly detection model

Now that we have generated PET images using a CycleGAN and explored the nature of their hidden distribution, we can investigate whether a UAD model trained on a combination of real and generated PET images can outperform a model trained solely on the limited set of available real PET images. We also investigate if synthetic PET data can be used instead of real ones when paired with T1 MRI modality to achieve a similar model's performance in epilepsy detection compared to the model trained on pairs of real T1 and PET images.

7.3.1 Description of the brain UAD model

7.3.1.1 Siamese network for feature extraction

This analysis is based on the UAD model described in section 5.1.
The architecture of the encoder and the decoder used in the regularized Siamese network that was applied to learn the hidden representation of image patches is shown in Figure 7.22. In the case of one modality input, the input consists of pairs of patches of different subjects centered around the same coordinate. Each patch of size 15x15x1 is mapped to a vector $z \in R^{64}$ using the encoder (E), and is subsequently

mapped back to its original space using the decoder (D).

We want to exclude the brain regions (the cerebellum and brain stem) that are not susceptible to epilepsy using a masking image in the MNI space derived from the Hammersmith maximum probability atlas described in Hammers et al., 2003. If a selected patch overlaps with the mask by at least 30% then we keep this patch in a training set, a smaller overlap would lead to a rejection.

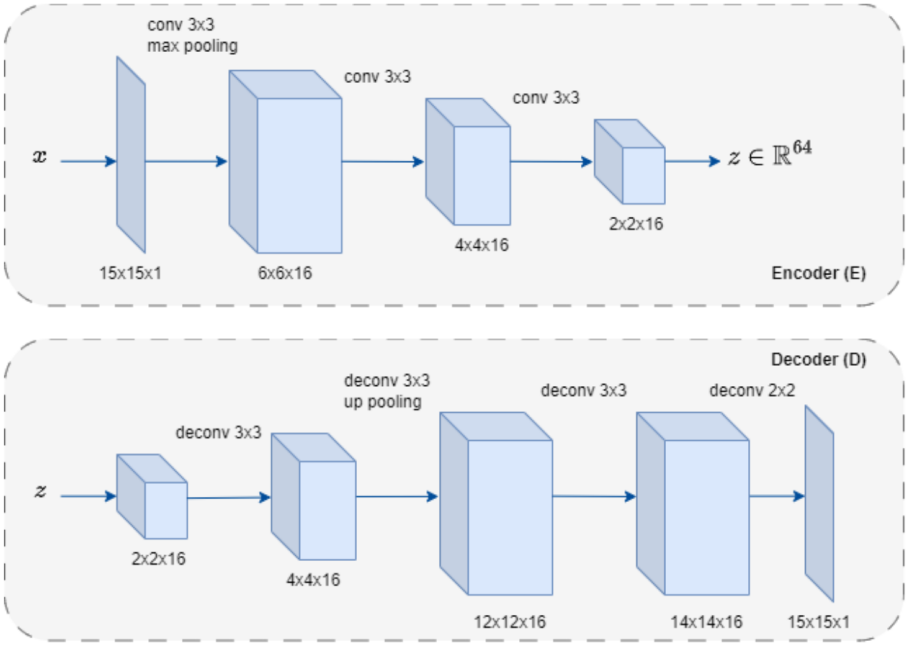

FIGURE 7.22 – Encoder and decoder of the Siamese Network

7.3.1.2 oc-SVM classifier for outlier detection

Every voxel denoted as v_i is associated with a oc-SVM classifier C_i uses the RBF kernel. that is trained on the matrix $M_i = [z_{i1}, ..., z_{in}]$ where every element z_{i1} is the hidden representation of a patch centered at v_i of a subject j out of total number of subjects n. Each classifier C_i uses the RBF kernel as explained earlier in 5.1.2. As an output, for each voxel v_i we get a *score* that is a distance to the found optimal hyperplane that separated the inliers from outliers (or anomalies). In the end, all distance scores combined together yield the *distance map* D_p for the given control/patient.

7.3.1.3 Post-processing

The distance map derived from the previous stage has to be processed to obtain the final detections. We follow the next steps:

1. During oc-SVM training a 10-fold evaluation of the controls is performed. For each fold we get the distance maps based on the oc-SVM models trained on the remaining subjects (For example, if we have 35 control subjects, we train oc-SVM on 28 subjects and get the distance maps for the remaining 7 subjects, and this process repeats 10 times). The distance maps obtained this way allow us to compute the standard deviation of the normal subjects' distance distribution at

voxel-level. We save these values for every voxel in the form of a normalization map N_S computed as:

$$N_S(v_i) \leftarrow std(\{D_s(v_i)\}_{s \in X}) \tag{7.14}$$

X is the total training dataset, and Ds is the distance score map for the healthy control s,

2. For every patient p we compute a normalised distance map D_pn by dividing the output distance map D_p over the N_S voxel-wise.
3. The final distance map D_{pw} (we refer to it as a weighted distance map) is computed as:

$$D_{pw} = \frac{1}{2}(\frac{D_p}{max(abs(D_p))} + \frac{D_{pn}}{max(abs(D_{pn}))}) \tag{7.15}$$

Some areas of the brain have greater variability and are more likely to be seen as anomalies, by weighing them by the standard deviation, the score maps take into account this effect.

4. The next step consists of thresholding the weighted distance map to get *a cluster map* with areas of the highest probability to be an anomaly.
   (a) All scores from D_{pw} are combined into a histogram. Subsequently, this histogram is approximated using a non-parametric distribution through a kernel density estimator.
   (b) With a chosen p-value and a 26-connectivity rule the connected components or *clusters* are identified. P-value is a varying parameter and can be chosen by a clinician to perform the analysis. The minimal size of a cluster is set to be 82.
5. The clusters received at the previous steps are called *detections*. In order the define the cluster of the highest probability to be an anomaly we perform a ranking as:

$$rank(c_i) = \omega * \frac{score(c_i)}{min_j score(c_j)} + (1 - \omega) * \frac{size(c_i)}{max_j size(c_j)} \tag{7.16}$$

here, $score(c_i)$ is the average of the voxel scores in the cluster c_i, $size(c_i)$ indicates the cluster's voxel count, ω is a weight parameter that we set to be 0.5 in our experiments to equally threat the clusters on both their size and average score.

7.3.1.4 Evaluation protocol

To evaluate the CAD system performance with different input modalities, we compare the final cluster maps with the ground truth annotations. A cluster map is compared to the ground truth image, and the overlap between the found clusters and the ground truth cluster is calculated. If there is any overlap for one or more detected clusters, they are referred to as *true positive detections*. The remaining clusters that fall outside the true lesion zone are counted as *false positive detections*. We report the number of correct detections as the total number of patients where there was at least one true positive detection. We are also interested in evaluating the ranks of the true detections found among the top n clusters in all patients.

Additionally, we limit the maximum number of found clusters to 15 thus tolerating no more than 14 false positives. The greater number of clusters would decrease the clinicians ability to estimate the CAD system's results.

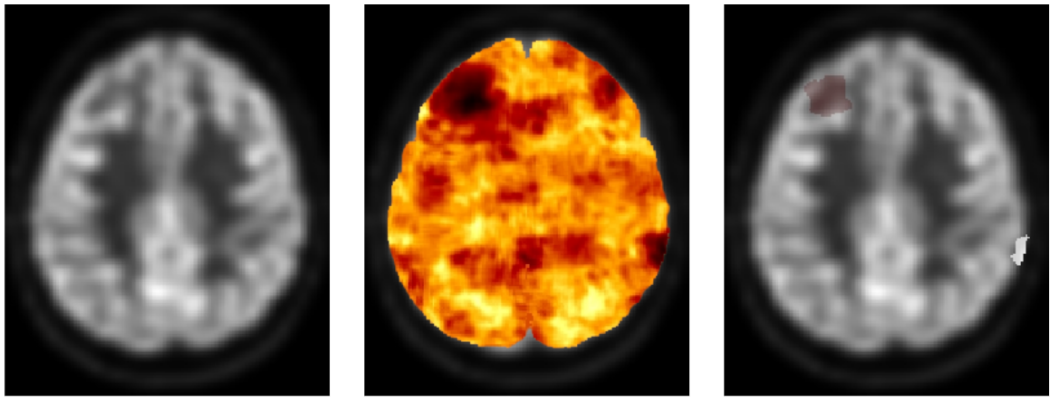

FIGURE 7.23 – An example of post-processing steps on one patient: a) original PET scan transverse slice b) a score map overlayed on top of the original image with dark areas being very negative values corresponding to the probable anomaly zone c) thresholded D_{pw} at a p-value with a maximum of 15 clusters, the top left red cluster has the highest rank and is identified as an anomaly.

7.3.2 Experiments

We conduct two types of experiments: training of UAD models on PET data (solely on real PET images and on the combination of real and synthetic data), and training of UAD models on pairs of T1 MRI and PET data where we use either real PET or synthetic ones thus modeling the situation of missing PET modality in the training dataset. The details for these two types of experiments are given below.

UAD models trained on PET data

These models are trained on three different databases: the series of 35 real control PET dataset (DB_{C1}) described in section 5.2 and two hybrid databases mixing these 35 real control PET with 40 synthetic control FDG PET data generated by the semi-3D (DB_{C2}^{semi3D}) and 3D-patch ($DB_{C2}^{3Dpatch}$) Cycle-GAN models with MSE loss from original T1 data.

For the siamese network training pairs at the control level were randomly selected, and for each control, a fixed number of patches — specifically, 25 000 — was extracted, constituting the optimal quantity for efficient and reliable training of the Siamese network. Thus, with a dataset DB_{C1} consisting of 35 healthy subjects with their real PET images we get 875 000 patches, and we extract 1 875 000 patches for a combined real dataset with synthetic datasets DB_{C2}^{semi3D} and $DB_{C2}^{3Dpatch}$. We trained the network for 30 epochs, The trade-off coefficient controlling the extent of similarity was set to 0 for the first 10 epochs, then linearly increased to the maximum of 0.5 during the next 15 epochs and remained at a plateau for the last 5 epochs. The optimization was performed by using Adam and a learning rate=0.001.

During oc-SVM training, a 10-fold evaluation of the controls is performed. For each fold, we get the distance maps based on the oc-SVM models trained on the

remaining subjects (If we have 35 control subjects, we train oc-SVM on 28 subjects and get the distance maps for the remaining 7 subjects, and this process repeats 10 times). The distance maps obtained this way allow us to compute the standard deviation of the normal subjects' distance distribution at voxel level. After processing distance maps we receive resulting cluster maps with detections as described in section 7.3.1.3.

As the result, we obtain three trained models $UAD_{pet}^{original}$, UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$ and compare their detection performance on DB_{ep1}.

UAD models trained on pairs of T1 MRI and PET data

As for the PET experiments, we use the same architecture for the siamese network with the only difference being that we combine T1 and PET patches at the channel level to form the input. Each training dataset consists of 35 controls with real T1 and either real or synthetic PET data producing 875 000 training patches. The test set DB_{ep2} consists of paired T1+PET scans of 18 patients. The difference from the experiment we made for PET data is that we took only patients whose scans are coming from the Lyon hospital excluding patients AB and AC for which bad outcomes after the surgery were observed and added new patients (see Table 5.2) for the details.

Training of the oc-SVM model is the same as for PET experiments.

Resulting $UAD_{mripet}^{original}$, UAD_{mripet}^{semi3D} and $UAD_{mripet}^{3Dpatch}$) are tested on DB_{ep2} to evaluate the detection performance of each model.

Implementation details

The development and implementation of the UAD model were done in Python, using Tensorflow and Keras libraries for the deep model training and feature extraction. The oc-SVM part was accomplished by partitioning all the voxels into individual subsets, with each subset being allocated to its own thread, and the oc-SVMs were then trained sequentially. Testing was done in the same way. The oc-SVM implementation was provided by the Scikit-learn library.

7.3.3 Results

7.3.3.1 Results of the CAD model trained on real PET and the mix of real and synthetic PET images

Our hypothesis is that adding synthetic images to the train set can serve to improve performance of machine learning based diagnostic models. Previously, we described the development and implementation of the UAD model.

In our paper [Zotova et al., 2021] we reported that the best detection sensitivity of 64.7% (11 out of 17 correct detections) was achieved with the model trained on the hybrid dataset including the 40 PET data generated from the best semi-3D model. In Table 7.5, we however present the updated results after having made a more rigorous and in-depth visual analysis. This review revealed inaccuracies in the initial automatic detection process leading to adjustments in sensitivity rate. A detailed visual analysis revealed that several detections previously classified as true positives did not meet the criteria for TP from a clinical perspective. Our initial approach — comparing a set of 35 true PET images with an augmented set comprising both true

Patient	$UAD_{pet}^{original}$	UAD_{pet}^{semi3D}	$UAD_{pet}^{3Dpatch}$
Patient A	x	x	x
Patient C	✓ (1)	✓ (5)	✓ (1)
Patient D	✓ (1)	✓ (1)	✓ (1)
Patient E	✓ (4)	x	x
Patient G	x	x	x
Patient J	✓ (6)	x	x
Patient O	✓ (1)	x	x
Patient Q	x	x	x
Patient R	x	x	x
Patient S	x	x	x
Patient U	x	x	✓ (9)
Patient V	x	✓ (6)	✓ (4)
Patient AA	x	x	x
Patient AB	x	✓ (5)	✓ (4)
Patient AC	✓ (3)	x	x
Patient AD	x	x	x
Patient AE	✓ (4)	x	x
Overall # of lesion detections	**7**	**4**	**5**
Mean rank	**3.3**	**4.25**	**3.8**

TABLE 7.5 – Performance of the brain anomaly detection model trained on three databases. From left to right: $UAD_{pet}^{original}$: 35 real PET images from DB_{C1}, UAD_{pet}^{semi3d}: 35 real PET images from DB_{C1} and 40 synthetic PET images DB_{C2}^{semi3D}, $UAD_{pet}^{3dpatch}$: 35 real PET images from DB_{C1} and 40 synthetic PET images $DB_{C2}^{3dpatch}$

and synthetic PET images — was reevaluated. This comparison appears to have not been the most appropriate task for assessing the model's performance accurately. The complexity of the task, coupled with the heterogeneous nature of the PET images (acquired with different machines and varying in quality), likely skewed the initial results. Additionally, the inclusion of patients with poor outcomes (patients AA, AB, AC, AE in Table 5.2), who were later excluded in the revised analysis (the next experiments with T1 and PET data), further necessitated this update.

The revised analysis revealed that adding synthetic data to the training, which here amounts to doubling the number of training samples, did not lead to a gain in performance. The highest performance showed only a 41.2% sensitivity (7 out of 17 detections) was achieved with the model trained on 35 real PET scans. Figure 7.24 illustrates anomaly maps derived from the three detection models on three test epilepsy patients.

Interestingly to admit, even though UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$ models demonstrated reduced sensitivity, they were able to detect lesions in patients missed by $UAD_{pet}^{original}$. Factors such as the quality of synthetic images, the degree of realism in replicating clinically relevant features, and the fusion of synthetic with real dataset are critical considerations that could influence the efficacy of deep learning models trained on augmented datasets.

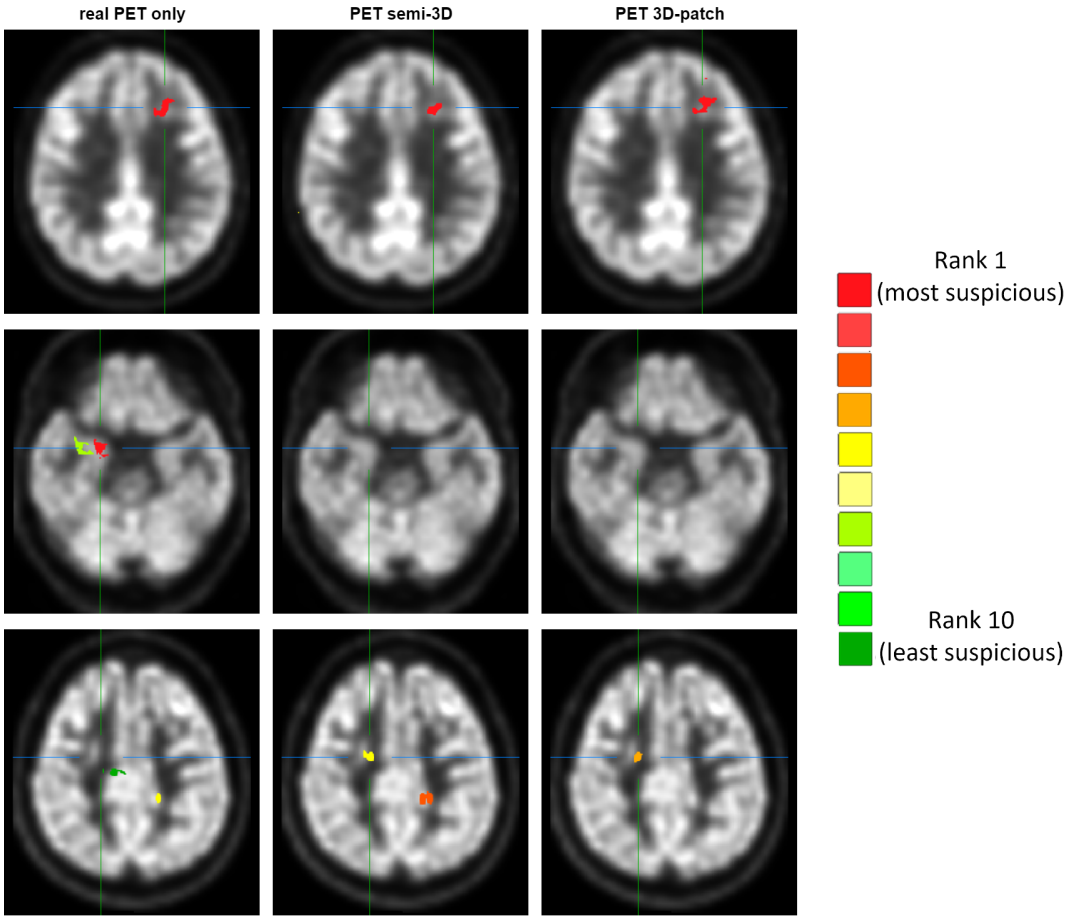

FIGURE 7.24 – Example cluster maps for three patients produced by the detection models, from left to right: 35 real PET scans ($UAD_{pet}^{original}$), 35 real+40 synthetic PET (UAD_{pet}^{semi3D}), 35 real+40 synthetic PET ($UAD_{pet}^{3Dpatch}$). The upper line demonstrates a case for Patient D where all models detected a correct cluster with a high confidence (rank of 1). The middle line shows a result for Patient O, where models trained on the mix of real and synthetic PET failed to detect a cluster, but the correct detection was made by a model trained on real PET images. The bottom line demonstrates a case for Patient AB where both models trained with additional synthetic data managed to detect a lesion with a middle confidence in the right internal frontal lobe, while it is missed by the model trained on real PET data solely. Red clusters indicate a very high probability of anomaly, while green clusters represent the least suspicious areas detected by the model.

7.3.3.2 Results of the CAD model trained on T1+real PET and T1+synthetic PET images

Detection performance results are reported in Table 7.6. Patient *T* (see Table 5.4) was added to evaluate the ability of the UAD models to detect *non-subtle anomalies*, here a surgical resection localization in the right temporal pole. Detection performance achieved for this patient is not included in the overall metrics reported on the last two lines of Table 7.6, thus aggregating performance estimated on the 19 small and subtle epileptogenic lesions of 18 patients included into DB_{ep2} (patient *B* has two lesions in left temporal pole and insula).

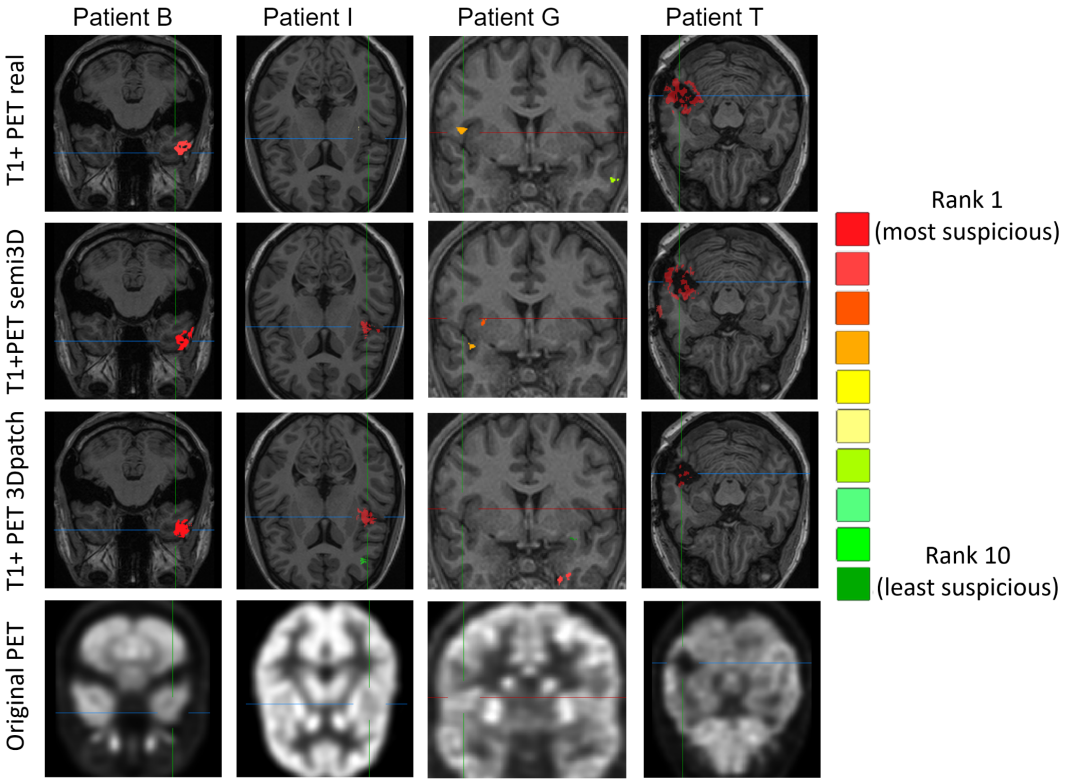

FIGURE 7.25 – Example cluster maps overlaid on T1 MRI by the detection models, from top to bottom: $UAD^{original}_{mripet}$, UAD^{semi3D}_{mripet} and $UAD^{3Dpatch}_{mripet}$, respectively. Selected transverse or coronal slices of patients B, I, and G are centered on confirmed EZ localisations in various areas of the brain, namely the left temporal lobe, left insula, and right precentral gyrus. Illustration of patient T depicts performance for the detection of a large surgical resection area located in the right temporal lobe.The cursor points to suspicious anatomical regions. The color bar displays the most suspicious cluster of rank 1 as bright red and the least suspicious detected cluster of rank 10 as dark green.

The best-reported detection performance was achieved with the $UAD^{3Dpatch}_{mripet}$ model trained on $DB^{3Dpatch}_{C2}$, with 74% sensitivity (14 out of the 19 lesions) and a mean reported rank of 2.1, meaning that the detected epilepsy lesions were, on average, among the top 3 most suspicious clusters reported by this model. The UAD^{semi3D}_{mripet} model trained on DB^{semi3D}_{C2} achieved 58% sensitivity (11 out of the 19 lesions) and a mean reported rank of 2.4, thus outperforming the model trained on the real paired T1 MRI and FDG PET of DB_{C1} whose sensitivity and mean rank were of 42% (8 out of 19 lesions) and 3.9, respectively. These results are on par with those of the quantitative visual analysis and show that images obtained from the 3D-patch CycleGAN model appeared to be visually realistic as well as adapted to the training of the UAD model.

Figure 7.25 illustrates anomaly cluster maps derived from the three detection

Patient	**T1+PET real** $UAD^{original}_{mripet}$	**T1+PET semi-3D** UAD^{semi3D}_{mripet}	**T1+PET 3D-patch** $UAD^{3Dpatch}_{mripet}$
Patient A	✗	✗	✗
Patient B	✓(2)	✓(3)	✓(1) and ✓(5)
Patient C	✓ (3)	✓ (4)	✓ (2)
Patient E	✗	✓ (4)	✓ (5)
Patient F	✓ (5)	✓ (4)	✓ (1)
Patient G	✓ (4)	✓ (3, 4)	✗
Patient I	✗	✓(1,4)	✓(1)
Patient J	✓ (8)	✗	✓ (1)
Patient K	✓(1)	✓(1)	✓(1)
Patient M	✗	✓(2)	✓(1)
Patient N	✓ (1)	✓ (1)	✓ (3)
Patient O	✗	✓ (3)	✓ (5)
Patient P	✗	✗	✗
Patient Q	✗	✗	✗
Patient S	✗	✗	✓ (2)
Patient T	✓ (1)	✓ (1)	✓ (1)
Patient U	✓ (7)	✓ (1)	✓ (1)
Patient Z	✗	✗	✓(1)
Overall # of lesion detections	**8**	**11**	**14**
Mean rank	**3.9**	**2.4**	**2.1**

TABLE 7.6 – Performance of the brain anomaly detection model trained on the three databases: from left to right: $UAD^{original}_{mripet}$: 35 real T1 and PET samples of DB_{C1}, UAD^{semi3D}_{mripet}: 35 paired true T1 MRI and fake PET of DB^{semi3D}_{C2}, $UAD^{3Dpatch}_{mripet}$: 35 paired true T1 MRI and fake PET of $DB^{3Dpatch}_{C2}$. ✓ denotes a true detection followed by its rank inside parentheses. ✗ denotes no true positive detection meaning that the lesion was not detected among the 10 highest-ranked clusters detected by the model. Bottom lines denote the total number of detected lesions over all epileptogenic patients as well as the mean rank score assigned by each model.

models on 4 test epilepsy patients. These visual results confirm the quantitative performance reported in Table 7.6. Lesion of Patient *B* in the left temporal pole was correctly detected by all models but with the highest rank obtained by the $UAD^{3Dpatch}_{mripet}$ model trained on T1 and synthetic PET data generated by the 3D-patch CycleGAN. Lesion of Patient *I* located in the left insula was correctly detected by the $UAD^{3Dpatch}_{mripet}$ and UAD^{semi3D}_{mripet} model with the highest suspicious rank, but it was missed by the $UAD^{original}_{mripet}$ model trained on real T1 and PET data. Lesion of Patient *G* located in the vicinity of the right precentral gyrus and operculum was correctly detected with $UAD^{original}_{mripet}$ and UAD^{semi3D}_{mripet} models with ranks of 4 and 3, respectively, but it was missed by $UAD^{3Dpatch}_{mripet}$ model. At the most right, the surgical resection zone of Patient *T* in the right temporal lobe was correctly detected by all three models with high confidence.

7.4 Discussions and conclusions

In this chapter, we demonstrated that realistic FDG PET exams of healthy subjects can be generated from GAN-based architectures with T1 MRI as input. This is assessed by the quantitative visual metrics, such as PSNR and SSIM reported in Table 7.2 as well as qualitatively, based on example samples reported in Figure 7.16. Quantitative results show that our proposal to train the standard CycleGAN architecture with paired T1 and PET data and additional MSE loss term allowed notable improvements in comparison to standard CycleGAN. We improved the visual quality of synthetic PET data by performing histogram matching of the generated synthetic PET, thus producing synthetic PET images that closely match the original ones.

Visual performance reported in this study is in the range of values obtained in Yaakub et al., 2019 and Flaus et al., 2023 for the same task of generating FDG PET images of control subjects based on their T1 MRI. Both studies implemented GAN-based architectures with resulting PSNR values of 23.2 $\pm$ 2.3 in Yaakub et al., 2019 and 35$\pm$ 3.8 in Flaus et al., 2023.

We also introduce diagnostic task-oriented quality metrics of the synthetic imaging data and strengthen the use case application focusing on the detection of epilepsy lesions in paired FDG PET and T1 MR data. We include an extensive performance evaluation of the proposed UAD model on a homogeneous and extended series of clinical epilepsy patients with confirmed lesion localization and surgery outcome as well as paired T1 MRI and PET exams realized in similar conditions.

We showed that generating realistic PET images can be seen as an augmentation technique leading to the improved performance of the UAD model when added to the set of original PET data. The first examination of this was done by measuring the OOD metrics reported in Figures 7.18 and 7.19. An alignment in the metrics is consistent with the performance of the UAD model trained on the mix of real PET data from DB_{C1} and synthetic PET from DB_{C2}^{semi3D}, showing noticeable improvements in detecting subtle epileptogenic lesions.

We also show that these synthetic PET data could efficiently replace true FDG PET images into multi-modality normative databases to train unsupervised anomaly detection models. As for the previous experiment, this was first assessed based on the OOD metrics reported in figures 7.20 and 7.21 indicating a reasonable overlap of the metrics computed on the paired T1 and real PET and paired T1 and synthetic PET data. This study provides clues of evidence that a UAD detection model trained with normative fake PET would not silently fail when tested on true patient PET data. The detection performance analysis allows confirming conclusions drawn from the OOD metrics.

The slight improvement of detection performance observed with the models trained on synthetic paired data might be explained by the lower mean and standard error values of the Mahalanobis distance D_m estimated on the inliers fake pairs, respectively the 5 test samples of DB_{C2}^{semi3D} (red triangles) in Fig 7.20 and the 5 test samples of and $DB_{C2}^{3Dpatch}$ (green triangles) in Fig. 7.21. This might indicate that the latent distribution of the synthetic representation z is denser than that achieved with the real paired T1 MR and FDG PET samples. The denser the normative distribution in the latent representation space, the more compact the density support estimated by the oc-SVM model, and the more sensitive it is to any deviation from normality

at inference. One possible explanation for the assumed higher density of the synthetic latent distribution is that the synthetic PET data have lower inter-individual pattern variability than the real population. The UAD^{semi3D}_{mripet} and $UAD^{3Dpatch}_{mripet}$ models trained on paired synthetic data might be more sensitive to any subtle deviation from the normative pattern, including those originating from subtle epilepsy lesions. This attempted explanation should be interpreted with caution. Further investigation including reproducibility analysis based on an extended database, is required to confirm this trend.

This study builds on the CycleGAN model that demonstrates impressive performance in synthesizing natural and medical images, as reported in section 3.2.3. We improved the visual quality of the synthetic PET images by training this architecture with paired FDG and T1 exams and adding the MSE loss term. The recent study of Pan et al Pan et al., 2022 underlines that CycleGAN architectures are still competitive models for the task of cross-modality image synthesis. The added value of transformer models was also recently addressed. In Dalmaz et al., 2022, the authors proposed an architecture based on ResVit, which consists of a generator subnetwork that follows an encoder-decoder pathway with aggregated residual transformer (ART) blocks in the bottleneck. This generator is associated with a convolutional discriminator subnetwork. This architecture was designed to enable conditional image synthesis, meaning it can unify various source-target modality configurations into a single model. Performance was evaluated on brain MR images from the IXI dataset to synthesize one missing modality from triplets of T1, T2, or FLAIR MR images. Encouraging visual performance based on SSIM and PSNR was reported, outperforming GAN or U-Net-based architectures for this task. In J. Liu et al., 2023, an encoder-decoder model based on multiscale Swin Transformer blocks was proposed for the same task. As in Dalmaz et al., 2022, this architecture was evaluated on T1, T2, and PD MR images of the IXI database with improved visual metrics compared to standard GAN architectures. The objective of these two studies was to design a model that can take any subset of input contrasts and synthesize those that are missing in a unified and unique framework. This task is thus different from ours so an extensive comparison is not achievable. One comment, however, is that both studies did not report the performance of CycleGAN-based architectures, or diagnostic-based quality metrics. As far as we know, such transformer-based architectures have not been evaluated yet for the synthesis of PET data from T1 MRI.

One perspective to this work would be to assess the reproducibility of our results based on an extended validation study with more control subjects and patients. We also plan to further assess the added value of attention modules, e.g. based on the transformer models that are shown to perform well for the imputation of missing imaging modalities. Finally, as recently investigated in Pan, Chen, et al., 2021 and Pan et al., 2022, one promising way is to design models that jointly tackle the synthesis of the modality and the diagnostic task at hand, by coupling synthesis and diagnostic (e.g. classification) networks.

Chapter 8

Multimodality fusion

8.1 Epilepsy lesion detection on multimodality images

It is common to have multiple sources of information for a given problem, such as different measurements, experiments, or sensors. In medical imaging, multiple images of the same subject can be obtained from different modalities (e.g. MRI, CT and PET) or protocols (e.g. various MRI sequences like FLAIR and DWI). Utilizing multiple modalities can provide a more comprehensive understanding of a patient's condition, as each modality highlights different aspects of the subject. This concept has led to the development of machine learning algorithms based on multimodality imaging that aims at extracting the most discriminant and complementary information from different imaging sources for more accurate diagnosis and treatment.

In chapter 2, we covered the problem of using multiple sources of information for building better models and saw the latest works exploring models' capabilities depending on the level at which the different sources of information are combined. We further focused on the problem of epilepsy detection, and in chapter 5 we formulated it as the problem of detecting subtle lesions in brain imaging as a voxel-wise outlier detection problem. We also presented our strategy where we exploited the idea of combining the siamese networks for hidden representation learning coupled with per voxel oc-SVMs for generating the output maps with regions suspicious to be epileptical. In chapter 7 we demonstrated the power of using two modalities, T1 and PET, and showed that the usage of artificially generated PET images may even outperform the model trained on real images. The best performance of the proposed CAD system trained on real T1 and synthetic PET images showed 74% of sensitivity among the top 3 most suspicious detected clusters.

In this chapter, we examine three fusion strategies, namely, early fusion, intermediate fusion, and late fusion as explained in chapter 2 in order to exploit the full potential of available modalities for patients (T1, FLAIR and PET). We will then assess the effectiveness of each strategy and make a comparison between them.

8.1.1 Data

The patient data set is the database DB_{ep3} described in Table 5.2 of chapter 5. It consists of the DB_{ep2} patients as for our previous experiments with T1 and PET modalities in chapter 7 with some extra patients from DB_{ep1} and some more new patients that were included later. We recall that PET exams of DB_{ep2} patients were acquired on the same scanner as that used to acquire the PET exams in the training dataset DB_{C1} while PET exams for the other patients were acquired on different PET machines. The training dataset consisting of healthy controls is the same as introduced in section 5.2 with additional 40 synthetic PET images from $DB_{C2}^{3Dpatch}$. In

the course of the experiments presented below, we use the T1-weighted, FLAIR MRI and PET sequences for both the 75 healthy controls (including synthetic data) and 31 patients with confirmed epilepsy lesions. All images were normalized to the MNI space with FLAIR and PET images being rigidly co-registered with the corresponding T1-w volumes and further normalized to the MNI space as well. Again, we excluded the brain regions (the cerebellum and brain stem) that are not susceptible to epilepsy using a masking image in the MNI space derived from the Hammersmith maximum probability atlas described in Hammers et al., 2003. After the elimination of the corresponding voxels, the number of remaining voxels adds up to around 1.4 million. In the pre-processing stage, we additionally removed top 1% intensities to avoid a negative impact on the performance and individually scaled the images between 0 and 1.

8.1.2 Experimental setting

For the next set of experiments, we adapted the architecture of the siamese network depicted in Figure 8.1):

— We dropped the max pooling operation in an attempt to preserve the positional information. The dimensionality reduction is now reached by applying different strides in convolutional layers.

— Every convolutional layer is followed by a batch normalization layer, serving as regularization to avoid overfitting and stabilizing the whole training process.

— Different kernel and stride sizes of the convolutional layers lead to the smaller size of z-representation (16 vs the previous size of 64). As we tested the detection model on various combinations of modalities at 3 different levels, we had to reduce the z-dimension so the joint z-representations at the intermediate fusion could still fit the oc-SVM classifier.

— The decoding part is mirrored to the encoder consisting of deconvolutional layers with the same kernel and stride sizes as for the encoder followed by batch normalization.

For both siamese network and oc-SVM trainings, we form a training dataset consisting of 32 controls from DB_{C1} and 37 controls from DB_{C2} and a validation dataset (the remaining 3 controls from DB_{C1} and DB_{C2}).

We extract 250 000 patches instead of 25 000 from each control to train the siamese network. This is motivated by a generalization need and richer feature representations, especially since we reduce the dimension of z-representation of patches. The training details for the siamese network remain the same as for experiments in section 7.3.

When it comes to the oc-SVM part, we preserve the same classifier, but only change the computation of the normalized maps described in section 7.3.1.3. We now compute distance maps of the 6 validation control samples, and since they are not a part of the siamese network's training, it should give us the standard deviation of the distance distribution among normal subjects at each voxel comparable with the previous setting but avoiding the computational costs of performing 10-fold cross-validation.

8.1.3 Detection performance evaluation

We implemented changes in the evaluation protocol after revision of the results obtained in experiments described in sections 7.3.3.1 and 7.3.3.2. In these experiments, a detection was classified as 'true positive' if there was a minimal overlap, as small as a single pixel, between the detected anomaly cluster and the ground truth (GT) annotation. This criterion, however, led to an overestimation of true detections, complicating the assessment of our anomaly detection model's performance. For instance, for some patients in evaluation of $UAD_{pet}^{original}$, UAD_{pet}^{semi3D} and $UAD_{pet}^{3Dpatch}$ models, we noted that, despite the detected cluster had a small overlap with the ground truth on a specific slice, yet the spatial propagation of the cluster was predominantly in a direction divergent from the actual epileptogenic zone. The revised approach below aims to reduce false positives and improve the consistency between detected clusters and GT annotations, thereby providing a more accurate reflection of the model's efficacy in identifying epileptogenic areas in medical imaging.

— **Expanded Ground Truth (GT) Annotations**: we have broadened the scope of our GT files to encompass the entire brain region associated with the anomaly, rather than limiting annotations to localized areas. This will help with the initial step in our evaluation process which involves verifying that a detected cluster is situated within the appropriate brain region. Subsequently, we conduct a visual inspection to assess the extent of overlap between the detected cluster and the precise GT annotations.

— **Single cluster map implementation**: in contrast to our previous approach of generating multiple cluster maps from an output score map at varying probability (p-levels), we now employ a single cluster map in our evaluation. This map contains a maximum of 15 clusters, simplifying the analysis and enhancing the interpretability of results.

— **Intersection threshold**: we require that the model's detected cluster has more than 30% intersection with the marked region of the brain in the GT file. This criterion ensures that the detected cluster encompasses a significant portion of the actual anomaly, thereby enhancing the reliability of the detection.

8.1.4 Early fusion

We employ the early fusion strategy by concatenating different imaging modalities at the channel level. Specifically, each input sample in our training dataset $X = \{x_1; x_2; ...; x_N\}$ is a multi-channel patch derived from the chosen modalities. For the combinations of T1+FLAIR, T1+PET, and FLAIR+PET, each patch is of size 15x15x2, representing spatial dimensions of 15x15 and two channels corresponding to the respective imaging modalities. In the case of combining all three modalities (T1+FLAIR+PET), the patch size becomes 15x15x3, incorporating an additional channel for the third modality. It is important to emphasize that these dimensions describe a single patch. The subsequent latent representation extracted by our network for each patch retains a dimensionality of 16, regardless of the number of modalities fused in the input.

8.1.5 Intermediate fusion

In this setting, we merge two or all three modalities at the level of extracted hidden representation. Upon obtaining the latent representation vectors from the

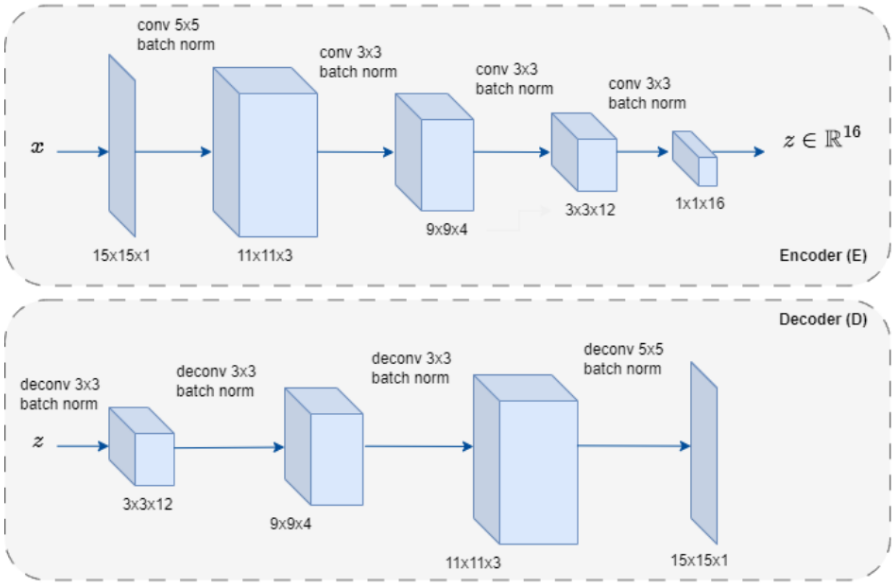

FIGURE 8.1 – Encoder and decoder of the Siamese Network v2

separately trained siamese networks, each on one single modality, we aim to merge these vectors to create a composite representation that encapsulates the features from each modality. This merging is done by concatenation: for the two modalities merging $z = [z_1; z_2]$ resulting in the dimensionality of 32 and $z = [z_1; z_2; z_3]$ for T1, FLAIR and PET modalities integration with the dimensionality of 48.

8.1.6 Late fusion

In the late fusion stage, our methodology centers around the integration of cluster maps, which are derived from individual UAD models. Each of these models is specifically trained on a single modality. The merging process unfolds as follows:

— **Creation of the unified cluster map**. The ultimate cluster map is formed by merging clusters originating from different imaging modalities (T1, FLAIR, and PET). Our aim is to ensure that the final map includes both clusters that are common across different modalities and those that are highly significant in a single modality.

— **Determining cluster significance through rank assignment**. To evaluate the importance of each cluster, we adopt a rank-based system. When there is an overlap of clusters from different modalities, the rank for the combined cluster in the unified map is assigned based on the most significant (i.e., the lowest) rank among these intersecting clusters. For instance, if an overlapping occurs between a cluster ranked 1 in the T1 modality and a cluster ranked 6 in the PET modality, the combined cluster in the unified map is assigned the higher significance with a rank of 1.

— **Prioritizing clusters**. In cases where clusters are formed by the union of different modalities, these clusters are assigned a higher rank to highlight their multimodal importance. Despite this, individual clusters that initially have very high importance, for example, ranks 1 or 2, retain their prominent positions in the final ranking. This approach acknowledges the significant individual contributions of these clusters.

— **Limiting the number of clusters**. To ensure the analysis remains focused and manageable, we limit the total number of clusters in the final joint cluster map to 15, thus concentrating on the most relevant and significant clusters.

This methodology enables the consideration of identified regions of interest across all modalities while preserving high-ranked clusters from individual modalities. Consequently, this should enhance the reliability of detections while reducing the number of false positive detections. Figure 8.2 illustrates the described approach of new cluster formation and their rank assignment.

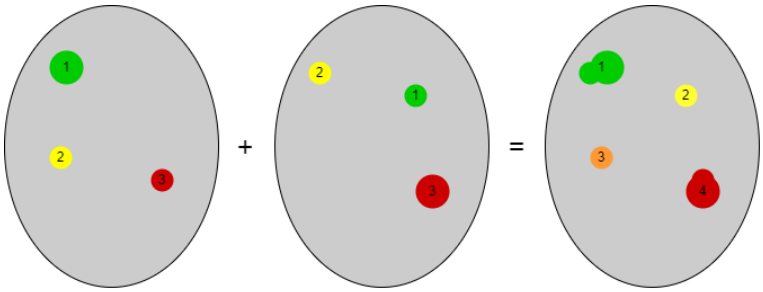

FIGURE 8.2 – Late fusion. Creation of the unified cluster map and rank assignment. Clusters from individual modalities are combined into a unified map, with ranks assigned based on their original positions and overlaps. The methodology aims to retain the significance of high-ranked clusters while integrating single clusters of a high rank from all modalities.

8.2 Results

8.2.1 Comparison of fusion levels

Tables 8.1 - 8.4 present the results of comparison for different fusion strategies including the model performance for single modalities input. As a baseline, we take results of single modalities in Table 8.1 where the FLAIR modality showed the highest number of detections (15 out of 26), however only for 4 patients the detected clusters gained the highest rank of 1 that signifies the most probable cluster to be an epileptic region. T1 and PET modalities managed to detect 11 and 10 patients respectively with a mean rank of 4.2 and 6.6. It is noteworthy that the three monomodal models performed differently in identifying epilepsy across patients. They converged on correctly identifying the true lesions on 4 (patients I, N, R and V) out of the 26 patients only, with different rankings of the lesions. This may suggest that different imaging data indeed contain complementary information which, when combined together, would provide better epilepsy identification in some patients.

For the *early fusion* strategy in Table 8.2, combinations of the PET and FLAIR modalities showed the maximum number of detections with 4 patients (C, G, L and T) showing the highest cluster rank. The least number of detections is observed for the paired T1 and PET modalities. Combining all three imaging modalities at the early stage did not seem to improve the model performance. One possible explanation for that could be that the architecture of the siamese network considered in our study did not allow handling the high dimensionality and complexity of all three modalities combined as input channels. In addition to that, it is likely that T1 and

FLAIR modalities share relevant features so that adding both of them as input channels did not provide significant additional information beyond what one of them already provided.

The *intermediate fusion* fell behind the expectations. Combinations of T1+PET and FLAIR+PET managed to detect 8 and 6 lesions respectively with only 4 and 3 clusters of a high rank (patients G, H, V and Z for the first pair of modalities and patients P, R and Z for the latter). Early fusion allows the model to learn non-linear transformations and interactions between the modalities from the beginning, while capturing these intricate interactions can be more challenging at an intermediate stage. An additional point is that by stacking hidden representations from two modalities, we effectively double the size of the resulting feature vector – expanding it from 16 to 32 dimensions before it is input into the oc-SVM. This increase in dimensionality can be problematic, as SVMs are susceptible to the 'curse of dimensionality' - the phenomenon where the feature space becomes so high-dimensional that the model's performance deteriorates. Performing the experiment for the intermediate fusion with three modalities was not feasible as the computational complexity of this method significantly increased the time required for processing that went beyond the allowed frames of the computational center. Further optimization of the code is required to perform this last experiment. This experience highlights once again the practical constraints often encountered in advanced machine learning research.

Combining modalities at the *late stage* surpassed the efficacy of early fusion, showing improvements in the model's confidence and decision-making accuracy. We observed 17 detections in total when integrating T1, FLAIR and PET images with 11 clusters of a high rank (1-3) and it is worth remembering that the final cluster rank for the late fusion is a combination of individual modalities results, the newly assigned rank 3 could mean that in one individual modality the model predicted a rank of 1, and similarly, detections with a low rank had rather a medium level of confidence in individual modalities rather than low. The combination of T1 and FLAIR, as well as FLAIR and PET, resulted in 16 successful detections, with the majority of these clusters receiving high rankings. With the total number of 17 detections, we conclude that the late fusion combining all three available modalities is the winning approach for merging imaging modalities in our set of experiments.

8.2.2 Qualitative results

In this section, we present a detailed visual analysis of the performance of a CAD model for the detection of epileptogenic regions utilizing T1, FLAIR and PET images merged in the late fusion fashion. Visual results are presented in Figures 8.3 and 8.4. These images are projections of the detected clusters onto a transverse T1-weighted slice with a corresponding slice with ground truth on the left, for all patients in which our CAD system detected clusters. For each patient, we show on the left a transverse T1-weighted slice with the ground truth for epileptic zones delineated in red, and the image with clusters detected by our CAD model on the right.

Patients G, O, P, V, Y and Z demonstrate a high degree of overlap between the ground truth and the model's detections of high ranks, suggesting an effective identification of the epileptogenic region by the CAD model. The model has identified multiple clusters with medium rank for Patients C, H, R, some of which overlap with the extended ground truth region. This may indicate a dispersed epileptogenic zone, but also points to a need for enhanced precision to reduce false positives. The model also identified additional anomalies that do not correspond to the epileptogenic zones. Detected clusters of Patient I have minimal intersections with the

Patient	T1	FLAIR	PET
Patient A	✗	✓(15)	✓(15)
Patient B	✓(15)	✓(8)	✗
Patient C	✗	✓(3)	✗
Patient D	✗	✓(11)	✗
Patient E	✗	✗	✗
Patient F	✗	✗	✓(15)
Patient G	✓(1)	✗	✗
Patient H	✓(2)	✗	✓(1)
Patient I	✓(9)	✓(2)	✓(12)
Patient J	✗	✗	✗
Patient K	✗	✗	✗
Patient L	✗	✓(3)	✓(2)
Patient M	✗	✗	✗
Patient N	✓(9)	✓(7)	✓(11)
Patient O	✗	✗	✓(1)
Patient P	✗	✓(2)	✗
Patient Q	✗	✗	✗
Patient R	✓(2)	✓(4)	✓(7)
Patient S	✗	✓(1)	✗
Patient T	✓(1)	✓(1)	✗
Patient U	✓(3)	✗	✗
Patient V	✓(1)	✓(3)	✓(1)
Patient W	✗	✗	✓(1)
Patient X	✗	✓(4)	✗
Patient Y	✓(2)	✓(1)	✗
Patient Z	✓(1)	✓(1)	✗
Overall # of detections	**11**	**15**	**10**
Mean rank of true detections	**4.2**	**4.4**	**6.6**

TABLE 8.1 – Comparative results of CAD systems for different modalities using single modality inputs at patient level. If the lesion is detected it is marked with either ✓ and a rank of a detected cluster (the lower the rank - the more confident the result) or ✗ in case of no detection.

Patient	T1+FLAIR	T1+PET	FLAIR+PET	T1+FLAIR+PET
Patient A	✗	✗	✗	✓(8)
Patient B	✓(10)	✓(11)	✓(11)	✓(11)
Patient C	✗	✗	✓(1)	✗
Patient D	✓(12)	✗	✓(12)	✗
Patient E	✗	✗	✗	✗
Patient F	✗	✗	✓(15)	✗
Patient G	✗	✗	✓(1)	✗
Patient H	✗	✓(3)	✓(3)	✓(4)
Patient I	✓(6)	✓(11)	✓(11)	✓(3)
Patient J	✗	✗	✗	✗
Patient K	✗	✗	✗	✗
Patient L	✓(3)	✓(1)	✓(1)	✓(3)
Patient M	✗	✗	✗	✗
Patient N	✓(8)	✓(11)	✓(6)	✓(9)
Patient O	✗	✗	✗	✗
Patient P	✓(1)	✗	✓(2)	✓(1)
Patient Q	✗	✗	✗	✗
Patient R	✓(3)	✓(5)	✓(5)	✓(4)
Patient S	✓(3)	✗	✓(3)	✗
Patient T	✓(1)	✓(1)	✓(1)	✓(1)
Patient U	✗	✓(6)	✗	✗
Patient V	✓(5)	✓(3)	✓(4)	✗
Patient W	✗	✗	✗	✗
Patient X	✗	✗	✗	✗
Patient Y	✓(1)	✗	✗	✓(1)
Patient Z	✓(1)	✓(1)	✓(2)	✓(1)
Overall # of detections	**12**	**10**	**15**	**11**
Mean rank of true detections	**4.5**	**5.3**	**5.3**	**4.2**

TABLE 8.2 – Comparative results of CAD systems for different modalities using **early fusion strategy**. If the lesion is detected it is marked with either ✓ and a rank of a detected cluster (the lower the rank - the more confident the result) or ✗ in case of no detection.

Patient	T1+FLAIR	T1+PET	FLAIR+PET
Patient A	✓(14)	✗	✗
Patient B	✓(7)	✓(15)	✓(7)
Patient C	✗	✗	✗
Patient D	✓(10)	✗	✗
Patient E	✗	✗	✗
Patient F	✗	✗	✗
Patient G	✗	✓(1)	✗
Patient H	✗	✓(2)	✗
Patient I	✓(8)	✓(13)	✓(14)
Patient J	✗	✗	✗
Patient K	✗	✗	✗
Patient L	✓(3)	✗	✗
Patient M	✗	✗	✗
Patient N	✓(7)	✓(11)	✓(10)
Patient O	✗	✓(1)	✗
Patient P	✓(1)	✗	✓(1)
Patient Q	✗	✗	✗
Patient R	✓(4)	✗	✓(1)
Patient S	✓(1)	✗	✗
Patient T	✓(1)	✗	✗
Patient U	✗	✗	✗
Patient V	✓(2)	✓(1)	✗
Patient W	✗	✗	✗
Patient X	✗	✗	✗
Patient Y	✓(1)	✗	✗
Patient Z	✓(1)	✓(1)	✓(1)
Overall # of detections	**13**	**8**	**6**
Mean rank of true detections	**4.6**	**5.6**	**5.7**

TABLE 8.3 – Comparative results of CAD systems for different modalities using **intermediate fusion strategy** at patient level. If the lesion is detected it is marked with either ✓ and a rank of a detected cluster (the lower the rank - the more confident the result) or ✗ in case of no detection.

Patient	T1+FLAIR	T1+PET	FLAIR+PET	T1+FLAIR+PET
Patient A	✗	✗	✗	✗
Patient B	✓(14)	✗	✓(15)	✗
Patient C	✓(2)	✗	✓(1)	✓(3)
Patient D	✗	✗	✗	✗
Patient E	✗	✗	✗	✗
Patient F	✗	✗	✗	✗
Patient G	✓(1)	✓(1)	✗	✓(2)
Patient H	✓(2)	✓(2)	✓(2)	✓(3)
Patient I	✓(2)	✗	✓(2)	✓(4)
Patient J	✗	✗	✗	✗
Patient K	✗	✗	✗	✗
Patient L	✓(6)	✓(4)	✓(4)	✓(7)
Patient M	✗	✗	✗	✗
Patient N	✓(9)	✗	✓(13)	✓(15)
Patient O	✗	✓(1)	✓(2)	✓(2)
Patient P	✓(3)	✗	✓(4)	✓(3)
Patient Q	✗	✗	✗	✗
Patient R	✓(3)	✓(3)	✓(8)	✓(4)
Patient S	✓(1)	✗	✓(2)	✓(2)
Patient T	✓(1)	✓(1)	✓(1)	✓(1)
Patient U	✓(6)	✓(5)	✗	✓(9)
Patient V	✓(2)	✓(2)	✓(2)	✓(2)
Patient W	✗	✓(2)	✓(1)	✓(2)
Patient X	✓(8)	✗	✓(7)	✓(10)
Patient Y	✓(1)	✓(3)	✓(1)	✓(1,6)
Patient Z	✓(1)	✓(1)	✓(1)	✓(1)
Overall # of detections	**16**	**11**	**16**	**17**
Mean rank of true detections	**3.9**	**2.3**	**4.1**	**4.2**

TABLE 8.4 – Comparative results of CAD systems for different modalities using **late fusion strategy** at patient level. If the lesion is detected it is marked with either ✓ and a rank of a detected cluster (the lower the rank - the more confident the result) or ✗ in case of no detections.

ground truth, however, the model identified a big area that looks abnormal compared to other scans. Patients K and T with surgical resections were also considered to evaluate the performance of the model in detecting non-epileptogenic regions and large abnormalities. This finding is considered a success within the realm of anomaly detection, as it showcases the model's broader diagnostic capabilities.

A fair comparison with other published works is difficult because of the differences in the patient groups and the nature of their epilepsy. Our model achieved 65.4% sensitivity, and we remind again that this value has to be compared against 0% of sensitivity rate on MRI negative patients when observations were made by experts.

Our model system failed to identify the lesions of 9 patients from DB_{ep3}. For some of them, we could observe the highly anomalous regions in the raw output, but after the post-processing and cluster formation step, those detections did not appear among the top 15 detections. We assume that it might indicate the presence of other more 'anomalous' regions causing greater impact than subtle epileptic zones. For the patients where the model failed to detect the anomalous region, it would require a more thorough analysis of the raw images and the reasons why the model failed in detection.

8.3 Conclusions

In this chapter, we presented the results of experiments for anomaly detection on brain images. Initially trained on single modalities, we further explored the model capability of leveraging complementary information from different modalities combined at different stages of the model, namely, at the input level with channel-wise modalities concatenation (early fusion), at the intermediate level by concatenating feature representations and at the output level (late fusion) where we merged individual cluster maps from monomodal CADs. Our main contributions consist of:

1. Modality fusion for CAD models: enhanced the performance of an existing CAD model by integrating images from multiple modalities (T1, FLAIR, and PET).
2. Evaluation of synthetic PET imaging value: nearly half of the PET data incorporated into the model consisted of synthetic PET images ($DB_{C2}^{3Dpatch}$) generated by the best GAN model, namely, Cycle-GAN with MSE loss.
3. Comprehensive performance analysis: conducted a thorough comparison to determine the optimal level for combining modalities, offering a valuable reference for future work in multimodal image processing and its application in CAD systems, in particular with a focus on brain anomaly detection.

The early fusion strategy in the context of integrating multiple imaging modalities involves combining raw data from T1, FLAIR, and PET images at the input level (channel-wise) before feeding them into a model, the intermediate fusion consists of stacking the extracted hidden z-representation into one vector that goes further to the oc-SVM for training, and for the late fusion we developed a strategy that integrates the most suspicious clusters from individual modalities into one cluster map. The early fusion leveraged the complementary information available across different modalities from the outset and demonstrated a high level of performance, however, it is on par with the results we observed from a model solely trained on FLAIR images. The model trained on T1 and FLAIR images showed a decline in performance which can be explained by potential redundancy between these two modalities. The

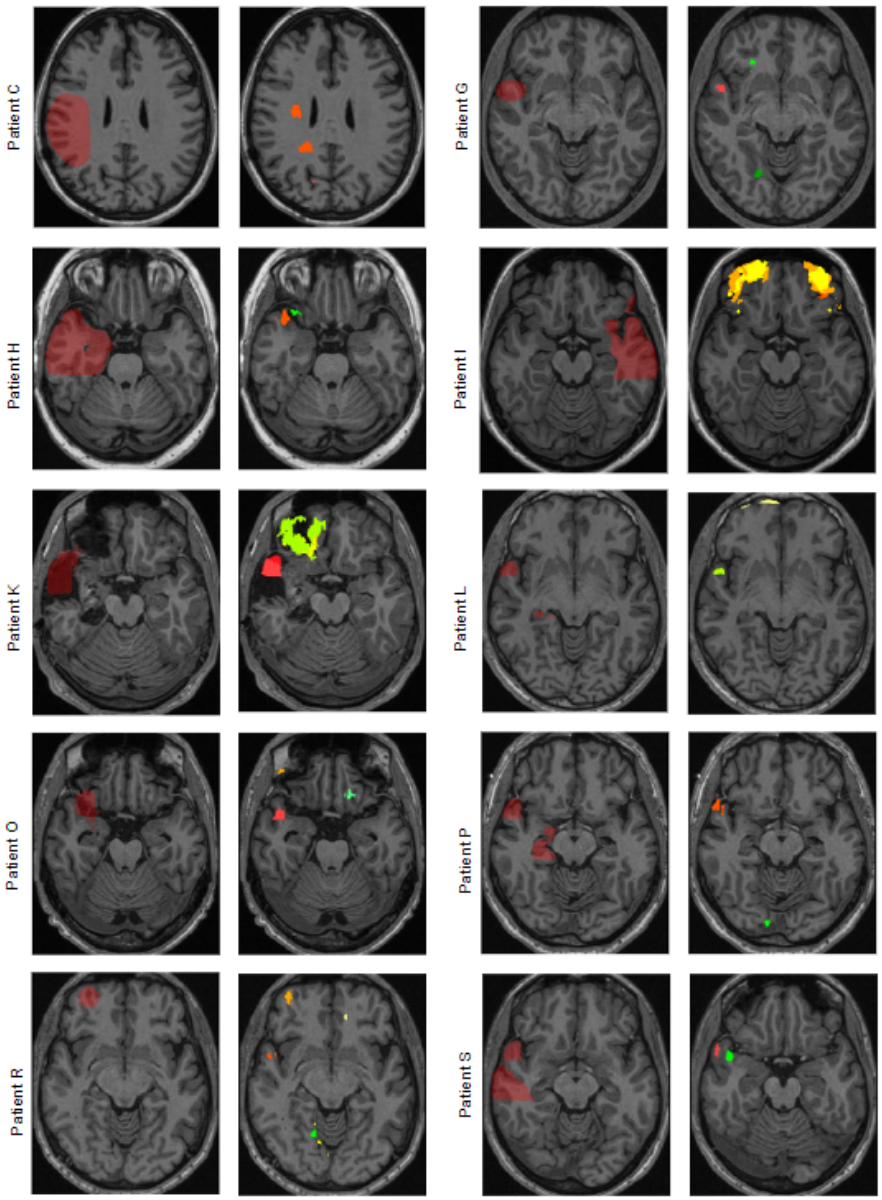

FIGURE 8.3 – Visual results of the detections made by CAD model trained on T1, FLAIR and PET images merged in late fusion fashion. For each patient, the left column displays the ground truth with the epileptogenic region highlighted in red, and the right column shows the model's detection output. Detected clusters are color-coded to indicate the likelihood of each cluster being an epileptogenic zone from bright red (the most suspicious cluster) to bright green (the least suspicious cluster). Part I

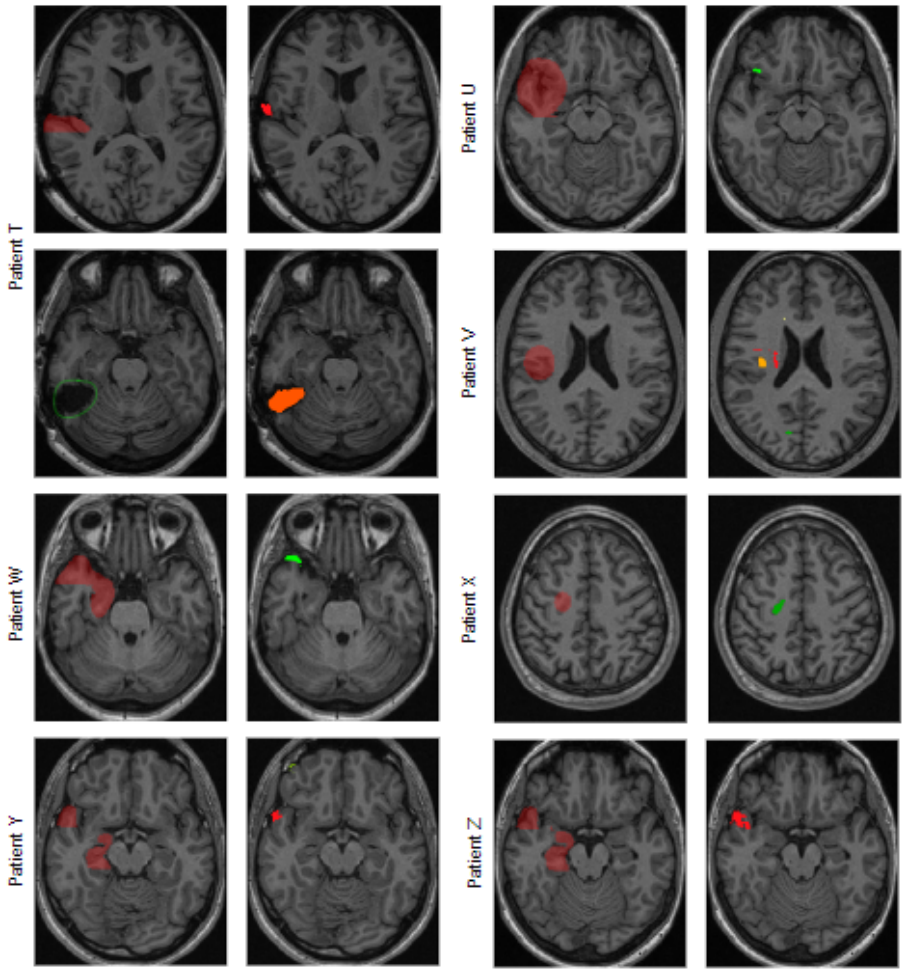

FIGURE 8.4 – Visual results of the detections made by CAD model trained on T1, FLAIR and PET images merged in late fusion fashion. For each patient, the left column displays the ground truth with the epileptogenic region highlighted in red, and the right column shows the model's detection output. Detected clusters are color-coded to indicate the likelihood of each cluster being an epileptogenic zone from bright red (the most suspicious cluster) to bright green (the least suspicious cluster). Part II

intermediate fusion turned out to be the least efficient. Concatenating the hidden representations from images might not effectively capture the complex intermodal interactions that are necessary for distinguishing epileptic regions, it increases the dimensionality of the feature space without ensuring that the combined features are optimally informative for anomaly detection. It is worth exploring post-processing steps for the intermediate fusion: once we receive a concatenated z-vector, we could use dimensionality reduction (with PCA or t-SNE), feature normalization, or more advanced techniques such as attention mechanisms. A solution proposed in W. Zhang et al., 2021 could also lead to a more integrated and efficient feature representation. In our study, late fusion, integrating cluster maps from T1, FLAIR, and both real and synthetic PET images, enhanced model performance and confidence in detecting epilepsy. The inclusion of synthetic PET images helped to train the model with the same number of controls as for T1 and FLAIR models. The results of detection from monomodal CADs showed a variety in patients for which models successfully detected a lesion or failed. Our hypothesis that an efficient strategy to incorporate the best detections from each modality proved to be the best approach in total number of correct detections. However, this approach introduced a longer processing time, suggesting a trade-off between improved diagnostic capability and time efficiency. Future work may focus on optimizing computational strategies to balance these aspects.

Conclusion and perspectives

This dissertation represents an attempt to enhance the efficacy of computer-aided diagnosis system for identifying epileptogenic lesions through neuroimaging data, further developing the foundational work established in Alaverdyan et al., 2020.

In the beginning, we presented a comprehensive overview of the existing CAD systems for epilepsy detection with their achievements and limitations. Dealing with a limited number of data and the lack of accurately annotated lesions we kept using the unsupervised approach, and identified two major ways for improvements as: generating synthetic missing modalities and exploring multimodal fusion.

Recognizing the limitations of existing models like the U-net in generating synthetic images with the necessary detail and accuracy, we pivoted towards a GAN-based methodology. The use of ResNet model as a generator in CycleGAN with additional MSE loss function allowed us to not only generate missing PET data with higher fidelity but also to integrate these synthesized images into our CAD system. The high image quality and close resemblance to the real images does not guarantee that synthetic images will be successfully applied in model training, thus we proposed to perform experiments for identifying out-of-distribution samples. Indeed, fake PET images formed distinct clusters with smaller mean squared error and Mahalanobis distance from the distribution formed as the mix of real and synthetic data. Adding generated PET images did not lead to the detection sensitivity of a CAD model and the best performance was achieved at the level of 41.2% by a model trained on only real PET images.

We then formulated a new problem to identify if synthetic images can replace entirely the missing modality and performed a set of experiments with pairs of T1 and PET images fused as channels for the model input. The OOD detection showed a closer resemblance between real and fake PET images coupled with real T1 modality. We later showed that the usage of fake PET data generated by the semi3d CycleGAN model led to some improvements in the CAD model: 77.7% of sensitivity by $UAD^{3Dpatch}_{mripet}$ versus 44.4% reached by $UAD^{3Dpatch}_{pet}$ model. We improved the quality of the images from $DB^{3Dpatch}_{C2}$ by applying the histogram matching and excluded from the test set those patients whose PET data were coming from other than Lyon hospital, thus we could guarantee a higher level of data homogeneity.

Our next contribution consisted in applying different fusion strategies, we explored three levels of merging T1, FLAIR and PET modalities. From the initial combination of merging data as channels at input level to concatenating feature vectors and cluster maps, each technique was evaluated for its potential to enhance the system's diagnostic capabilities. We concluded that the late fusion approach showed the highest detection sensitivity and reliability as it takes into account the most suspicious detections made by models trained on single modality.

The insights gained from our work suggest a promising horizon for the improvements of CAD systems, where the integration of multimodal imaging and advanced data synthesis techniques can lead to significant advances in the development of diagnostic models.

Future work

We proposed two main ways of improvements for the existing CAD model, and yet there are several more potential directions for future development.

The first aspect concerns missing data generation. In the overview in section 3.2.3 we mentioned the increasingly popular transformers architecture. Originally proposed to solve natural language processing tasks, it has been widely used in other domains including computer vision. Their ability to capture long-range dependencies and understand complex patterns in data offers a novel approach to improving the generative capabilities of GANs. By integrating transformer models into the generator component of GANs, future research could leverage the self-attention mechanism of transformers to enhance the quality and realism of generated images. By generating high-fidelity synthetic images, transformer-enhanced GANs could provide a robust method for data augmentation, helping to address the challenge of limited available datasets. This could improve the performance of CAD systems across diverse datasets.

The second way for enhancement could be the refinement of intermediate fusion techniques across multiple modalities. We implemented and tested only the basic concatenation approach that did not show the high detection ability of the system. An additional layer for learning the best features and their combination could be attached to the model. Alternatively, attention-based mechanism looks promising in extracting features of input images into a single feature map that can be later used as an input to the oc-SVM part of our CAD model. By selectively focusing on the most relevant features across different imaging modalities, attention mechanisms can significantly enhance the interpretability and efficacy of diagnostic models.

The next possible improvement lies in the area of data availability. We worked with a limited dataset consisting of T1, FLAIR and partially PET images. No matter how trivial it may sound, deep models tend to improve their performance with more data in the training set. Access to a rich variety of neuroimaging data enables the development of models that are robust, accurate, and capable of generalizing across different patient populations and imaging technologies. However, the collection and sharing of such data are often hampered by logistical, ethical, and technical challenges, including patient privacy concerns, data standardization issues, and the sheer volume of data generated by modern imaging techniques. The MELD project, with its focus on improving lesion detection in patients with drug-resistant epilepsy, aligns closely with the objectives of advancing CAD systems for epileptogenic lesion detection. By pooling resources, expertise, and data from multiple centers around the world, the MELD project offers a unique opportunity to overcome some of the challenges associated with data accessibility. Such partnerships not only provide access to invaluable datasets but also foster a collaborative ecosystem that can drive innovation and improve diagnostic outcomes for patients with epilepsy and other neurological conditions.

Last, but not least, we would like to highlight the need to incorporate not only imaging modalities but also other sources of data used for epilepsy detection such as EEG. Medical imaging, MRI or PET scans, provides detailed structural and functional insights into the brain, allowing for the identification of physical and metabolic changes associated with epileptogenic lesions. EEG, on the other hand, offers real-time electrical activity mapping, pinpointing the exact moments and locations of epileptic seizures. The integration of these two data sources leverages their complementary strengths, offering a multifaceted view of epilepsy that could lead to breakthroughs in detection, characterization, and treatment planning.

Publications

Journal article:

Zotova Daria, Pinon Nicolas, Trombetta Robin, Bouet Romain, Jung Julien, Lartizien Carole. GAN-based synthetic FDG PET images from T1 brain MRI can serve to improve performance of deep unsupervised anomaly detection models. Submitted to Artificial intelligence In medicine.

Conference papers and abstracts:

Zotova Daria, Jung Julien, Lartizien Carole. GAN-Based Synthetic FDG PET Images from T1 Brain MRI Can Serve to Improve Performance of Deep Unsupervised Anomaly Detection Models. SASHIMI workshop in conjunction with MICCAI conference, Strasbourg (virtual) 2021.

Zotova Daria, Lartizien Carole (2021). Realistic FDG-PET synthesis from T1 MRI based on adversarial deep architectures. Winter school AI4Health 2021, Paris.

Bibliography

Abbasi, B., & Goldenholz, D. M. (2019). Machine learning applications in epilepsy. *Epilepsia*, *60*(10), 2037–2047.

Abu-Srhan, A., Almallahi, I., Abushariah, M. A., Mahafza, W., & Al-Kadi, O. S. (2021). Paired-unpaired unsupervised attention guided gan with transfer learning for bidirectional brain mr-ct synthesis. *Computers in Biology and Medicine*, *136*, 104763.

Adler, S., Wagstyl, K., Gunny, R., Ronan, L., Carmichael, D., Cross, J. H., Fletcher, P. C., & Baldeweg, T. (2017). Novel surface features for automated detection of focal cortical dysplasias in paediatric epilepsy. *NeuroImage: Clinical*, *14*, 18–27.

Ahmed, B., Brodley, C. E., Blackmon, K. E., Kuzniecky, R., Barash, G., Carlson, C., Quinn, B. T., Doyle, W., French, J., Devinsky, O., et al. (2015). Cortical feature analysis and machine learning improves detection of "mri-negative" focal cortical dysplasia. *Epilepsy & Behavior*, *48*, 21–28.

Ahmed, B., Thesen, T., Blackmon, K. E., Kuzniekcy, R., Devinsky, O., & Brodley, C. E. (2016). Decrypting" cryptogenic" epilepsy: semi-supervised hierarchical conditional random fields for detecting cortical lesions in mri-negative patients. *The Journal of Machine Learning Research*, *17*(1), 3885–3914.

Akbar, A. S., Fatichah, C., & Suciati, N. (2022). Unet3d with multiple atrous convolutions attention block for brain tumor segmentation. *International MICCAI Brainlesion Workshop*, 182–193.

Akut, R. (2019). Wavelet based deep learning approach for epilepsy detection. *Health information science and systems*, *7*(1), 1–9.

Alaverdyan, Z., Jung, J., Bouet, R., & Lartizien, C. (2020). Regularized siamese neural network for unsupervised outlier detection on brain multiparametric magnetic resonance imaging: application to epilepsy lesion screening. *Medical image analysis*, *60*, 101618.

Alijamaat, A., NikravanShalmani, A., & Bayat, P. (2021). Multiple sclerosis lesion segmentation from brain mri using u-net based on wavelet pooling. *International journal of computer assisted radiology and surgery*, *16*(9), 1459–1467.

Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., et al. (2023). Revolutionizing healthcare: the role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), 689.

Armanious, K., Hepp, T., Küstner, T., Dittmann, H., Nikolaou, K., La Fougère, C., Yang, B., & Gatidis, S. (2020). Independent attenuation correction of whole body [18f] fdg-pet using a deep learning approach with generative adversarial networks. *EJNMMI research*, *10*(1), 1–9.

Armanious, K., Jiang, C., Abdulatif, S., Küstner, T., Gatidis, S., & Yang, B. (2019). Unsupervised medical image translation using cycle-medgan. *2019 27th European Signal Processing Conference (EUSIPCO)*, 1–5.

Avberšek, L. K., & Repovš, G. (2022). Deep learning in neuroimaging data analysis: applications, challenges, and solutions. *Frontiers in Neuroimaging*, *1*, 981642.

Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Farahani, K., Kalpathy-Cramer, J., Kitamura, F. C., Pati, S., et al. (2021). The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification. *arXiv preprint arXiv:2107.02314*.

Baur, C., Denner, S., Wiestler, B., Navab, N., & Albarqouni, S. (2021). Autoencoders for unsupervised anomaly segmentation in brain mr images: a comparative study. *Medical Image Analysis*, *69*, 101952.

Baur, C., Wiestler, B., Albarqouni, S., & Navab, N. (2018). Deep autoencoding models for unsupervised anomaly segmentation in brain mr images. *International MICCAI brainlesion workshop*, 161–169.

Baur, C., Wiestler, B., Muehlau, M., Zimmer, C., Navab, N., & Albarqouni, S. (2021). Modeling healthy anatomy with artificial intelligence for unsupervised anomaly detection in brain mri. *Radiology: Artificial Intelligence*, *3*(3).

Beheshti, I., Sone, D., Maikusa, N., Kimura, Y., Shigemoto, Y., Sato, N., & Matsuda, H. (2020a). Flair-wise machine-learning classification and lateralization of mri-negative 18f-fdg pet-positive temporal lobe epilepsy. *Frontiers in Neurology*, *11*, 580713.

Beheshti, I., Sone, D., Maikusa, N., Kimura, Y., Shigemoto, Y., Sato, N., & Matsuda, H. (2020b). Pattern analysis of glucose metabolic brain data for lateralization of mri-negative temporal lobe epilepsy. *Epilepsy Research*, *167*, 106474.

Bennett, O. F., Kanber, B., Hoskote, C., Cardoso, M. J., Ourselin, S., Duncan, J. S., & Winston, G. P. (2019). Learning to see the invisible: a data-driven approach to finding the underlying patterns of abnormality in visually normal brain magnetic resonance images in patients with temporal lobe epilepsy. *Epilepsia*, *60*(12), 2499–2507.

Bernasconi, A., Cendes, F., Theodore, W. H., Gill, R. S., Koepp, M. J., Hogan, R. E., Jackson, G. D., Federico, P., Labate, A., Vaudano, A. E., et al. (2019). Recommendations for the use of structural magnetic resonance imaging in the care of patients with epilepsy: a consensus report from the international league against epilepsy neuroimaging task force. *Epilepsia*, *60*(6), 1054–1068.

Bhagwat, N., Viviano, J. D., Voineskos, A. N., Chakravarty, M. M., Initiative, A. D. N., et al. (2018). Modeling and prediction of clinical symptom trajectories in alzheimer's disease using longitudinal data. *PLoS computational biology*, *14*(9), e1006376.

Bhan, A., Kapoor, S., Gulati, M., & Goyal, A. (2021). Early diagnosis of parkinson's disease in brain mri using deep learning algorithm. *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)*, 1467–1470.

Bharath, R. D., Panda, R., Raj, J., Bhardwaj, S., Sinha, S., Chaitanya, G., Raghavendra, K., Mundlamuri, R. C., Arimappamagan, A., Rao, M. B., et al. (2019). Machine learning identifies "rsfmri epilepsy networks" in temporal lobe epilepsy. *European radiology*, *29*(7), 3496–3505.

Bortsova, G., Bos, D., Dubost, F., Vernooij, M. W., Ikram, M. K., van Tulder, G., & de Bruijne, M. (2021). Automated segmentation and volume measurement of intracranial internal carotid artery calcification at noncontrast ct. *Radiology: Artificial Intelligence*, *3*(5).

Bos, D., Portegies, M. L., van der Lugt, A., Bos, M. J., Koudstaal, P. J., Hofman, A., Krestin, G. P., Franco, O. H., Vernooij, M. W., & Ikram, M. A. (2014). Intracranial carotid artery atherosclerosis and the risk of stroke in whites: the rotterdam study. *JAMA neurology*, *71*(4), 405–411.

Bos, D., Vernooij, M. W., de Bruijn, R. F., Koudstaal, P. J., Hofman, A., Franco, O. H., van der Lugt, A., & Ikram, M. A. (2015). Atherosclerotic calcification is related to a higher risk of dementia and cognitive decline. *Alzheimer's & Dementia*, *11*(6), 639–647.

Bouallegue, G., & Djemal, R. (2020). Eeg person identification using facenet, lstm-rnn and svm. *2020 17th International Multi-Conference on Systems, Signals & Devices (SSD)*, 22–28.

Brosch, T., Tang, L. Y., Yoo, Y., Li, D. K., Traboulsee, A., & Tam, R. (2016). Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE transactions on medical imaging*, *35*(5), 1229–1239.

Campos, S., Pizarro, L., Valle, C., Gray, K. R., Rueckert, D., & Allende, H. (2015). Evaluating imputation techniques for missing data in adni: a patient classification study. *Iberoamerican Congress on Pattern Recognition*, 3–10.

Carass, A., Roy, S., Jog, A., Cuzzocreo, J. L., Magrath, E., Gherman, A., Button, J., Nguyen, J., Prados, F., Sudre, C. H., et al. (2017). Longitudinal multiple sclerosis lesion segmentation: resource and challenge. *NeuroImage*, *148*, 77–102.

Cendes, F. (2013). Neuroimaging in investigation of patients with epilepsy. *CONTINUUM: Lifelong Learning in Neurology*, *19*(3), 623–642.

Cendes, F., Sakamoto, A. C., Spreafico, R., Bingaman, W., & Becker, A. J. (2014). Epilepsies associated with hippocampal sclerosis. *Acta Neuropathologica*, *128*(1), 21–37.

Chadebec, C., Thibeau-Sutre, E., Burgos, N., & Allassonnière, S. (2022). Data augmentation in high dimensional low sample size setting using a geometry-based variational autoencoder. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *45*(3), 2879–2896.

Chakraborty, S., Aich, S., & Kim, H.-C. (2020). Detection of parkinson's disease from 3t t1 weighted mri scans using 3d convolutional neural network. *Diagnostics*, *10*(6), 402.

Chan, H.-P., Hadjiiski, L. M., & Samala, R. K. (2020). Computer-aided diagnosis in the era of deep learning. *Medical physics*, *47*(5), e218–e227.

Chartsias, A., Joyce, T., Giuffrida, M. V., & Tsaftaris, S. A. (2017). Multimodal mr synthesis via modality-invariant latent representation. *IEEE transactions on medical imaging*, *37*(3), 803–814.

Chatterjee, S., Sciarra, A., Dünnwald, M., Tummala, P., Agrawal, S. K., Jauhari, A., Kalra, A., Oeltze-Jafra, S., Speck, O., & Nürnberger, A. (2022). Strega: unsupervised anomaly detection in brain mris using a compact context-encoding variational autoencoder. *arXiv preprint arXiv:2201.13271*.

Chebli, A., Djebbar, A., & Marouani, H. F. (2018). Semi-supervised learning for medical application: a survey. *2018 international conference on applied smart systems (ICASS)*, 1–9.

Chen, L., Wu, Y., DSouza, A. M., Abidin, A. Z., Wismüller, A., & Xu, C. (2018). Mri tumor segmentation with densely connected 3d cnn. *Medical Imaging 2018: Image Processing*, *10574*, 357–364.

Chen, S., Zhang, J., Ruan, X., Deng, K., Zhang, J., Zou, D., He, X., Li, F., Bin, G., Zeng, H., et al. (2020). Voxel-based morphometry analysis and machine learning based classification in pediatric mesial temporal lobe epilepsy with hippocampal sclerosis. *Brain imaging and behavior*, *14*(5), 1945–1954.

Commowick, O., Cervenansky, F., Cotton, F., & Dojat, M. (2021). Msseg-2 challenge proceedings: multiple sclerosis new lesions segmentation challenge using a

data management and processing infrastructure. *MICCAI 2021-24th International Conference on Medical Image Computing and Computer Assisted Intervention*, 126.

Commowick, O., Istace, A., Kain, M., Laurent, B., Leray, F., Simon, M., Pop, S. C., Girard, P., Ameli, R., Ferré, J.-C., et al. (2018). Objective evaluation of multiple sclerosis lesion segmentation using a data management and processing infrastructure. *Scientific reports*, *8*(1), 13650.

Dalmaz, O., Yurt, M., & Çukur, T. (2022). Resvit: residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, *41*(10), 2598–2614.

Demoustier, M., Khemir, I., Nguyen, Q. D., Martin-Gaffé, L., & Boutry, N. (2022). Residual 3d u-net with localization for brain tumor segmentation. *International MICCAI Brainlesion Workshop*, 389–399.

Desarnaud, S., Mellerio, C., Semah, F., Laurent, A., Landre, E., Devaux, B., Chiron, C., Lebon, V., & Chassoux, F. (2018). 18f-fdg pet in drug-resistant epilepsy due to focal cortical dysplasia type 2: additional value of electroclinical data and coregistration with mri. *European Journal of Nuclear Medicine and Molecular Imaging*, *45*(8), 1449–1460.

Dewey, B. E., Zhao, C., Reinhold, J. C., Carass, A., Fitzgerald, K. C., Sotirchos, E. S., Saidha, S., Oh, J., Pham, D. L., Calabresi, P. A., et al. (2019). Deepharmony: a deep learning approach to contrast harmonization across scanner changes. *Magnetic resonance imaging*, *64*, 160–170.

Ding, Y., Zhu, Y., Jiang, B., Zhou, Y., Jin, B., Hou, H., Wu, S., Zhu, J., Wang, Z. I., Wong, C. H., et al. (2018). 18f-fdg pet and high-resolution mri co-registration for pre-surgical evaluation of patients with conventional mri-negative refractory extra-temporal lobe epilepsy. *European Journal of Nuclear Medicine and Molecular Imaging*, *45*(9), 1567–1572.

Dolz, J., Desrosiers, C., & Ayed, I. B. (2018). Ivd-net: intervertebral disc localization and segmentation in mri with a multi-modal unet. *International Workshop and Challenge on Computational Methods and Clinical Applications for Spine Imaging*, 130–143.

Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al. (2020). An image is worth 16x16 words: transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Duan, Y., Shan, W., Liu, L., Wang, Q., Wu, Z., Liu, P., Ji, J., Liu, Y., He, K., & Wang, Y. (2020). Primary categorizing and masking cerebral small vessel disease based on "deep learning system". *Frontiers in neuroinformatics*, *14*, 17.

Duncan, J. S., Sander, J. W., Sisodiya, S. M., & Walker, M. C. (2006). Adult epilepsy. *The Lancet*, *367*(9516), 1087–1100.

Duncan, J. S., Winston, G. P., Koepp, M. J., & Ourselin, S. (2016). Brain imaging in the assessment for epilepsy surgery. *The Lancet Neurology*, *15*(4), 420–433.

El Azami, M., Hammers, A., Jung, J., Costes, N., Bouet, R., & Lartizien, C. (2016). Detection of lesions underlying intractable epilepsy on t1-weighted mri as an outlier detection problem. *PloS one*, *11*(9), e0161498.

Engemann, D. A., Kozynets, O., Sabbagh, D., Lemaıtre, G., Varoquaux, G., Liem, F., & Gramfort, A. (2020). Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. *Elife*, *9*, e54055.

Fang, P., An, J., Zeng, L.-L., Shen, H., Qiu, S., & Hu, D. (2017). Mapping the convergent temporal epileptic network in left and right temporal lobe epilepsy. *Neuroscience letters*, *639*, 179–184.

Feng, C.-M., Yan, Y., Fu, H., Chen, L., & Xu, Y. (2021). Task transformer network for joint mri reconstruction and super-resolution. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, 307–317.

Fernandez-Quilez, A., Parvez, O., Eftestøl, T., Kjosavik, S. R., & Oppedal, K. (2022). Improving prostate cancer triage with gan-based synthetically generated prostate adc mri. *Medical Imaging 2022: Computer-Aided Diagnosis*, *12033*, 422–427.

Fisher, R. S., Boas, W. V. E., Blume, W., Elger, C., Genton, P., Lee, P., & Engel Jr, J. (2005). Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ilae) and the international bureau for epilepsy (ibe). *Epilepsia*, *46*(4), 470–472.

Flaus, A., Jung, J., Ostrowky-Coste, K., Rheims, S., Guénot, M., Bouvard, S., Janier, M., Yaakub, S. N., Lartizien, C., Costes, N., et al. (2023). Deep-learning predicted pet can be subtracted from the true clinical fluorodeoxyglucose pet co-registered to mri to identify the epileptogenic zone in focal epilepsy. *Epilepsia Open*, *8*(4), 1440–1451.

Fu, Y., Lei, Y., Wang, T., Curran, W. J., Liu, T., & Yang, X. (2020). Deep learning in medical image registration: a review. *Physics in Medicine & Biology*, *65*(20), 20TR01.

Gabr, R. E., Coronado, I., Robinson, M., Sujit, S. J., Datta, S., Sun, X., Allen, W. J., Lublin, F. D., Wolinsky, J. S., & Narayana, P. A. (2020). Brain and lesion segmentation in multiple sclerosis using fully convolutional neural networks: a large-scale study. *Multiple Sclerosis Journal*, *26*(10), 1217–1226.

Gao, J., Zhao, W., Li, P., Huang, W., & Chen, Z. (2022). Legan: a light and effective generative adversarial network for medical image synthesis. *Computers in Biology and Medicine*, *148*, 105878.

Ghassemi, N., Shoeibi, A., & Rouhani, M. (2020). Deep neural network with generative adversarial networks pre-training for brain tumor classification based on mr images. *Biomedical Signal Processing and Control*, *57*, 101678.

Gill, R. S., Hong, S.-J., Fadaie, F., Caldairou, B., Bernhardt, B., Bernasconi, N., & Bernasconi, A. (2017). Automated detection of epileptogenic cortical malformations using multimodal mri. *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 349–356). Springer.

Gill, R. S., Hong, S.-J., Fadaie, F., Caldairou, B., Bernhardt, B. C., Barba, C., Brandt, A., Coelho, V. C., d'Incerti, L., Lenge, M., et al. (2018). Deep convolutional networks for automated detection of epileptogenic brain malformations. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 490–497.

González, C., Gotkowski, K., Fuchs, M., Bucher, A., Dadras, A., Fischbach, R., Kaltenborn, I. J., & Mukhopadhyay, A. (2022). Distance-based detection of out-of-distribution silent failures for covid-19 lung lesion segmentation. *Medical image analysis*, *82*, 102596.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Guerrini, R., Duchowny, M., Jayakar, P., Krsek, P., Kahane, P., Tassi, L., Melani, F., Polster, T., Andre, V. M., Cepeda, C., et al. (2015). Diagnostic methods and treatment options for focal cortical dysplasia. *Epilepsia*, *56*(11), 1669–1686.

Guo, Y., Liu, Y., Ming, W., Wang, Z., Zhu, J., Chen, Y., Yao, L., Ding, M., & Shen, C. (2020). Distinguishing focal cortical dysplasia from glioneuronal tumors in patients with epilepsy by machine learning. *Frontiers in Neurology*, *11*, 548305.

Guo, Z., Li, X., Huang, H., Guo, N., & Li, Q. (2018). Medical image segmentation based on multi-modal convolutional neural network: study on image fusion schemes. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 903–907.

Hammers, A., Allom, R., Koepp, M. J., Free, S. L., Myers, R., Lemieux, L., Mitchell, T. N., Brooks, D. J., & Duncan, J. S. (2003). Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human brain mapping*, *19*(4), 224–247.

Han, C., Rundo, L., Murao, K., Noguchi, T., Shimahara, Y., Milacski, Z. Á., Koshino, S., Sala, E., Nakayama, H., & Satoh, S. (2021). Madgan: unsupervised medical anomaly detection gan using multiple adjacent brain mri slice reconstruction. *BMC bioinformatics*, *22*(2), 1–20.

Han, X. (2017). Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical physics*, *44*(4), 1408–1419.

Hassanaly, R., Brianceau, C., Colliot, O., & Burgos, N. (2023). Unsupervised anomaly detection in 3d brain fdg pet: a benchmark of 17 vae-based approaches. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 110–120.

Havaei, M., Guizard, N., Chapados, N., & Bengio, Y. (2016). Hemis: hetero-modal image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 469–477.

Hong, S.-J., Bernhardt, B. C., Schrader, D. S., Bernasconi, N., & Bernasconi, A. (2016). Whole-brain mri phenotyping in dysplasia-related frontal lobe epilepsy. *Neurology*, *86*(7), 643–650.

Hsieh, Y.-Z., Luo, Y.-C., Pan, C., Su, M.-C., Chen, C.-J., & Hsieh, K. L.-C. (2019). Cerebral small vessel disease biomarkers detection on mri-sensor-based image and deep learning. *Sensors*, *19*(11), 2573.

Huang, B., Reichman, D., Collins, L. M., Bradbury, K., & Malof, J. M. (2018). Tiling and stitching segmentation output for remote sensing: basic challenges and recommendations. *arXiv preprint arXiv:1805.12219*.

Huang, C., Chen, W., Liu, B., Yu, R., Chen, X., Tang, F., Liu, J., & Lu, W. (2022). Transformer-based deep-learning algorithm for discriminating demyelinating diseases of the central nervous system with neuroimaging. *Frontiers in Immunology*, *13*.

Huang, J., Xu, J., Kang, L., & Zhang, T. (2020). Identifying epilepsy based on deep learning using dki images. *Frontiers in Human Neuroscience*, *14*, 590815.

Hussein, R., Palangi, H., Wang, Z. J., & Ward, R. (2018). Robust detection of epileptic seizures using deep neural networks. *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2546–2550.

Islam, J., & Zhang, Y. (2020). Gan-based synthetic brain pet image generation. *Brain informatics*, *7*, 1–12.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1125–1134.

Jakoby, B., Bercier, Y., Conti, M., Casey, M., Bendriem, B., & Townsend, D. (2011). Physical and clinical performance of the mct time-of-flight pet/ct scanner. *Physics in Medicine & Biology*, *56*(8), 2375.

James, A. P., & Dasarathy, B. V. (2014). Medical image fusion: a survey of the state of the art. *Information fusion*, *19*, 4–19.

Jiang, H., Wang, C., Chartsias, A., & Tsaftaris, S. A. (2020). Max-fusion u-net for multi-modal pathology segmentation with attention and dynamic resampling. *Myocardial pathology segmentation combining multi-sequence CMR challenge*, 68–81.

Jiao, R., Zhang, Y., Ding, L., Xue, B., Zhang, J., Cai, R., & Jin, C. (2023). Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Computers in Biology and Medicine*, 107840.

Jin, B., Krishnan, B., Adler, S., Wagstyl, K., Hu, W., Jones, S., Najm, I., Alexopoulos, A., Zhang, K., Zhang, J., et al. (2018). Automated detection of focal cortical dysplasia type ii with surface-based magnetic resonance imaging postprocessing and machine learning. *Epilepsia*, *59*(5), 982–992.

Kalantar, R., Messiou, C., Winfield, J. M., Renn, A., Latifoltojar, A., Downey, K., Sohaib, A., Lalondrelle, S., Koh, D.-M., & Blackledge, M. D. (2021). Ct-based pelvic t-1-weighted mr image synthesis using unet, unet plus plus and cycle-consistent generative adversarial network (cycle-gan). *FRONTIERS IN ONCOLOGY*, *11*.

Kamnitsas, K., Bai, W., Ferrante, E., McDonagh, S., Sinclair, M., Pawlowski, N., Rajchl, M., Lee, M., Kainz, B., Rueckert, D., et al. (2017). Ensembles of multiple models and architectures for robust brain tumour segmentation. *International MICCAI brainlesion workshop*, 450–462.

Kanber, B., Vos, S. B., de Tisi, J., Wood, T. C., Barker, G. J., Rodionov, R., Chowdhury, F. A., Thom, M., Alexander, D. C., Duncan, J. S., et al. (2021). Detection of covert lesions in focal epilepsy using computational analysis of multimodal magnetic resonance imaging data. *Epilepsia*, *62*(3), 807–816.

Keihaninejad, S., Heckemann, R. A., Gousias, I. S., Hajnal, J. V., Duncan, J. S., Aljabar, P., Rueckert, D., & Hammers, A. (2012). Classification and lateralization of temporal lobe epilepsies with and without hippocampal atrophy based on whole-brain automatic mri segmentation. *PloS one*, *7*(4), e33096.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., & Shah, M. (2021). Transformers in vision: a survey. *ACM Computing Surveys (CSUR)*.

Kiela, D., Bhooshan, S., Firooz, H., Perez, E., & Testuggine, D. (2019). Supervised multimodal bitransformers for classifying images and text. *arXiv preprint arXiv:1909.02950*.

Kikuchi, K., Togao, O., Yamashita, K., Momosaka, D., Nakayama, T., Kitamura, Y., Kikuchi, Y., Baba, S., Sagiyama, K., Ishimatsu, K., et al. (2021). Diagnostic accuracy for the epileptogenic zone detection in focal epilepsy could be higher in fdg-pet/mri than in fdg-pet/ct. *European radiology*, *31*(5), 2915–2922.

Kim, S. H., & Choi, J. (2019). Pathological classification of focal cortical dysplasia (fcd): personal comments for well understanding fcd classification. *Journal of Korean Neurosurgical Society*, *62*(3), 288–295.

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

Kläser, K., Varsavsky, T., Markiewicz, P., Vercauteren, T., Hammers, A., Atkinson, D., Thielemans, K., Hutton, B., Cardoso, M. J., & Ourselin, S. (2021). Imitation learning for improved 3d pet/mr attenuation correction. *Medical image analysis*, *71*, 102079.

Kooi, T., Litjens, G., Van Ginneken, B., Gubern-Mérida, A., Sánchez, C. I., Mann, R., den Heeten, A., & Karssemeijer, N. (2017). Large scale deep learning for computer aided detection of mammographic lesions. *Medical image analysis*, *35*, 303–312.

La Fougère, C., Rominger, A., Förster, S., Geisler, J., & Bartenstein, P. (2009). Pet and spect in epilepsy: a critical review. *Epilepsy & Behavior*, *15*(1), 50–55.

Lai, K. W., Khalil, A., Samiappan, D., et al. (2022). Performance analysis of machine learning and deep learning architectures on early stroke detection using carotid artery ultrasound images. *Frontiers in Aging Neuroscience*, 1013.

Lee, H. M., Gill, R. S., Fadaie, F., Cho, K. H., Guiot, M. C., Hong, S.-J., Bernasconi, N., & Bernasconi, A. (2020). Unsupervised machine learning reveals lesional variability in focal cortical dysplasia at mesoscopic scale. *NeuroImage: Clinical*, *28*, 102438.

Lee, M.-H., O'Hara, N., Sonoda, M., Kuroda, N., Juhasz, C., Asano, E., Dong, M., & Jeong, J.-W. (2020). Novel deep learning network analysis of electrical stimulation mapping-driven diffusion mri tractography to improve preoperative evaluation of pediatric epilepsy. *IEEE Transactions on Biomedical Engineering*, *67*(11), 3151–3162.

Leyden, K. M., Kucukboyaci, N. E., Puckett, O. K., Lee, D., Loi, R. Q., Paul, B., & McDonald, C. R. (2015). What does diffusion tensor imaging (dti) tell us about cognitive networks in temporal lobe epilepsy? *Quantitative imaging in medicine and surgery*, *5*(2), 247.

Li, D., Han, X., Gao, J., Zhang, Q., Yang, H., Liao, S., Guo, H., & Zhang, B. (2021). Deep learning in prostate cancer diagnosis using multiparametric magnetic resonance imaging with whole-mount histopathology referenced delineations. *Frontiers in medicine*, *8*.

Li, D., Peng, Y., Guo, Y., & Sun, J. (2022). Taunet: a triple-attention-based multimodality mri fusion u-net for cardiac pathology segmentation. *Complex & Intelligent Systems*, 1–17.

Liu, J., Pasumarthi, S., Duffy, B., Gong, E., Datta, K., & Zaharchuk, G. (2023). One model to synthesize them all: multi-contrast multi-scale transformer for missing data imputation. *IEEE Transactions on Medical Imaging*.

LIU, M.-Q. (2018). Bone age assessment model based on multi-dimensional feature fusion using deep learning. *Academic journal of second military medical university*, 909–916.

Liu, M., Zhang, J., Lian, C., & Shen, D. (2019). Weakly supervised deep learning for brain disease prognosis using mri and incomplete clinical scores. *IEEE transactions on cybernetics*, *50*(7), 3381–3392.

Liu, X., Emami, H., Nejad-Davarani, S. P., Morris, E., Schultz, L., Dong, M., & K Glide-Hurst, C. (2021). Performance of deep learning synthetic cts for mr-only brain radiation therapy. *Journal of applied clinical medical physics*, *22*(1), 308–317.

Luo, Y., Wang, Y., Zu, C., Zhan, B., Wu, X., Zhou, J., Shen, D., & Zhou, L. (2021). 3d transformer-gan for high-quality pet reconstruction. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 276–285.

Luu, H. M., & Park, S.-H. (2022). Extending nn-unet for brain tumor segmentation. *International MICCAI Brainlesion Workshop*, 173–186.

Mahmoudi, F., Elisevich, K., Bagher-Ebadian, H., Nazem-Zadeh, M.-R., Davoodi-Bojd, E., Schwalb, J. M., Kaur, M., & Soltanian-Zadeh, H. (2018). Data mining mr image features of select structures for lateralization of mesial temporal lobe epilepsy. *Plos one*, *13*(8), e0199137.

Malmgren, K., & Thom, M. (2012). Hippocampal sclerosis—origins and imaging. *Epilepsia*, *53*, 19–33.

Mao, X., Li, Q., Xie, H., Lau, R. Y. K., Wang, Z., & Smolley, S. P. (2017). Least squares generative adversarial networks. *Proceedings of the IEEE conference on computer vision (ICCV)*, 2813–2821.

Masoudi, S., Harmon, S. A., Mehralivand, S., Walker, S. M., Raviprakash, H., Bagci, U., Choyke, P. L., & Turkbey, B. (2021). Quick guide on radiology image preprocessing for deep learning applications in prostate cancer research. *Journal of Medical Imaging*, *8*(1), 010901.

McKinley, R., Wepfer, R., Aschwanden, F., Grunder, L., Muri, R., Rummel, C., Verma, R., Weisstanner, C., Reyes, M., Salmen, A., et al. (2021). Simultaneous lesion and brain segmentation in multiple sclerosis using deep neural networks. *Scientific reports*, *11*(1), 1–11.

Mérida, I., Jung, J., Bouvard, S., Le Bars, D., Lancelot, S., Lavenne, F., Bouillot, C., Redouté, J., Hammers, A., & Costes, N. (2021). Cermep-idb-mrxfdg: a database of 37 normal adult human brain [18 f] fdg pet, t1 and flair mri, and ct images available for research. *EJNMMI research*, *11*, 1–10.

Mirandola, L., Mai, R. F., Francione, S., Pelliccia, V., Gozzo, F., Sartori, I., Nobili, L., Cardinale, F., Cossu, M., Meletti, S., et al. (2017). Stereo-eeg: diagnostic and therapeutic tool for periventricular nodular heterotopia epilepsies. *Epilepsia*, *58*(11), 1962–1971.

Mo, J., Liu, Z., Sun, K., Ma, Y., Hu, W., Zhang, C., Wang, Y., Wang, X., Liu, C., Zhao, B., et al. (2019). Automated detection of hippocampal sclerosis using clinically empirical and radiomics features. *Epilepsia*, *60*(12), 2519–2529.

Moshé, S. L., Perucca, E., Ryvlin, P., & Tomson, T. (2015). Epilepsy: new advances. *The Lancet*, *385*(9971), 884–898.

Muñoz-Ramırez, V., Pinon, N., Forbes, F., Lartizen, C., & Dojat, M. (2021). Patch vs. global image-based unsupervised anomaly detection in mr brain scans of early parkinsonian patients. *International Workshop on Machine Learning in Clinical Neuroimaging*, 34–43.

Najafi, T., Jaafar, R., Remli, R., & Wan Zaidi, W. A. (2022). A classification model of eeg signals based on rnn-lstm for diagnosing focal and generalized epilepsy. *Sensors*, *22*(19), 7269.

Nemoto, K. (2017). Understanding voxel-based morphometry. *Brain and nerve= Shinkei kenkyu no shinpo*, *69*(5), 505–511.

Nguyen, R. D., Kennady, E. H., Smyth, M. D., Zhu, L., Pao, L. P., Swisher, S. K., Rosas, A., Mitra, A., Patel, R. P., Lankford, J., et al. (2021). Convolutional neural networks for pediatric refractory epilepsy classification using resting-state functional magnetic resonance imaging. *World neurosurgery*, *149*, e1112–e1122.

Nie, D., Wang, L., Gao, Y., & Shen, D. (2016). Fully convolutional networks for multi-modality isointense infant brain image segmentation. *2016 IEEE 13Th international symposium on biomedical imaging (ISBI)*, 1342–1345.

Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., et al. (2018). Attention u-net: learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999*.

Pan, Y., Chen, Y., Shen, D., & Xia, Y. (2021). Collaborative image synthesis and disease diagnosis for classification of neurodegenerative disorders with incomplete multi-modal neuroimages. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part V 24*, 480–489.

Pan, Y., Liu, M., Lian, C., Zhou, T., Xia, Y., & Shen, D. (2018). Synthesizing missing pet from mri with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis. *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11*, 455–463.

Pan, Y., Liu, M., Xia, Y., & Shen, D. (2021). Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. *IEEE transactions on pattern analysis and machine intelligence*, *44*(10), 6839–6853.

Pan, Y., Liu, M., Xia, Y., & Shen, D. (2022). Disease-image-specific learning for diagnosis-oriented neuroimage synthesis with incomplete multi-modality data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(10), 6839–6853. https://doi.org/10.1109/TPAMI.2021.3091214

Park, C.-h., & Ohn, S. H. (2019). A challenge of predicting seizure frequency in temporal lobe epilepsy using neuroanatomical features. *Neuroscience Letters*, *692*, 115–121.

Pesteie, M., Abolmaesumi, P., & Rohling, R. N. (2019). Adaptive augmentation of medical data using independently conditional variational auto-encoders. *IEEE transactions on medical imaging*, *38*(12), 2807–2820.

Qin, Z., Liu, Z., Zhu, P., & Ling, W. (2022). Style transfer in conditional gans for cross-modality synthesis of brain magnetic resonance images. *Computers in Biology and Medicine*, *148*, 105928.

Qiu, S., Chang, G. H., Panagia, M., Gopal, D. M., Au, R., & Kolachalama, V. B. (2018). Fusion of deep learning models of mri scans, mini–mental state examination, and logical memory test enhances diagnosis of mild cognitive impairment. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, *10*, 737–749.

Quaak, M., van de Mortel, L., Thomas, R. M., & van Wingen, G. (2021). Deep learning applications for the classification of psychiatric disorders using neuroimaging data: systematic review and meta-analysis. *NeuroImage: Clinical*, *30*, 102584.

Ramey, W. L., Martirosyan, N. L., Lieu, C. M., Hasham, H. A., Lemole Jr, G. M., & Weinand, M. E. (2013). Current management and surgical outcomes of medically intractable epilepsy. *Clinical Neurology and Neurosurgery*, *115*(12), 2411–2418.

Ranjbarzadeh, R., Bagherian Kasgari, A., Jafarzadeh Ghoushchi, S., Anari, S., Naseri, M., & Bendechache, M. (2021). Brain tumor segmentation based on deep learning and an attention mechanism using mri multi-modalities brain images. *Scientific Reports*, *11*(1), 1–17.

Raybaud, C., & Widjaja, E. (2011). Development and dysgenesis of the cerebral cortex: malformations of cortical development. *Neuroimaging Clinics*, *21*(3), 483–543.

Raza, K., & Singh, N. K. (2021). A tour of unsupervised deep learning for medical image analysis. *Current Medical Imaging*, *17*(9), 1059–1077.

Reda, I., Khalil, A., Elmogy, M., Abou El-Fetouh, A., Shalaby, A., Abou El-Ghar, M., Elmaghraby, A., Ghazal, M., & El-Baz, A. (2018). Deep learning role in early diagnosis of prostate cancer. *Technology in cancer research & treatment*, *17*, 1533034618775530.

Ren, Z., Wang, S., & Zhang, Y. (2023). Weakly supervised machine learning. *CAAI Transactions on Intelligence Technology*, *8*(3), 549–580.

Rieu, Z., Kim, J., Kim, R. E., Lee, M., Lee, M. K., Oh, S. W., Wang, S.-M., Kim, N.-Y., Kang, D. W., Lim, H. K., et al. (2021). Semi-supervised learning in medical

mri segmentation: brain tissue with white matter hyperintensity segmentation using flair mri. *Brain Sciences*, *11*(6), 720.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, 234–241.

Roy, S., Butman, J. A., Reich, D. S., Calabresi, P. A., & Pham, D. L. (2018). Multiple sclerosis lesion segmentation from brain mri via fully convolutional neural networks. *arXiv preprint arXiv:1803.09172*.

Rüber, T., David, B., & Elger, C. E. (2018). Mri in epilepsy: clinical standard and evolution. *Current opinion in neurology*, *31*(2), 223–231.

Rudie, J. D., Colby, J. B., & Salamon, N. (2015). Machine learning classification of mesial temporal sclerosis in epilepsy patients. *Epilepsy research*, *117*, 63–69.

Salem, M., Oliver, A., Salvi, J., & Lladó, X. (2021). Msdetector: a fully convolutional neural network for the detection of new t2-w lesion in multiple sclerosis. *MSSEG-2 challenge proceedings: Multiple sclerosis new lesions segmentation challenge using a data management and processing infrastructure*, 115.

San-Segundo, R., Gil-Martın, M., D'Haro-Enrıquez, L. F., & Pardo, J. M. (2019). Classification of epileptic eeg recordings using signal transforms and convolutional neural networks. *Computers in biology and medicine*, *109*, 148–158.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., & Williamson, R. C. (2001). Estimating the support of a high-dimensional distribution. *Neural computation*, *13*(7), 1443–1471.

Shain, C., Ramgopal, S., Fallil, Z., Parulkar, I., Alongi, R., Knowlton, R., Poduri, A., & Investigators, E. (2013). Polymicrogyria-associated epilepsy: a multicenter phenotypic study from the epilepsy phenome/genome project. *Epilepsia*, *54*(8), 1368–1375.

Shan, W., Duan, Y., Zheng, Y., Wu, Z., Chan, S. W., Wang, Q., Gao, P., Liu, Y., He, K., & Wang, Y. (2021). Segmentation of cerebral small vessel diseases-white matter hyperintensities based on a deep learning system. *Frontiers in Medicine*, *8*.

Shin, H.-C., Ihsani, A., Mandava, S., Sreenivas, S. T., Forster, C., Cha, J., & Initiative, A. D. N. (2020). Ganbert: generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis. *arXiv preprint arXiv:2008.04393*.

Shoeibi, A., Moridian, P., Khodatars, M., Ghassemi, N., Jafari, M., Alizadehsani, R., Kong, Y., Gorriz, J. M., Ramırez, J., Khosravi, A., et al. (2022). An overview of deep learning techniques for epileptic seizures detection and prediction based on neuroimaging modalities: methods, challenges, and future works. *Computers in Biology and Medicine*, 106053.

Si, X., Zhang, X., Zhou, Y., Sun, Y., Jin, W., Yin, S., Zhao, X., Li, Q., & Ming, D. (2020). Automated detection of juvenile myoclonic epilepsy using cnn based transfer learning in diffusion mri. *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1679–1682.

Sikka, A., Peri, S. V., & Bathula, D. R. (2018). Mri to fdg-pet: cross-modal synthesis using 3d u-net for multi-modal alzheimer's classification. *International Workshop on Simulation and Synthesis in Medical Imaging*, 80–89.

Simarro Viana, J., de la Rosa, E., Vande Vyvere, T., Robben, D., Sima, D. M., et al. (2020). Unsupervised 3d brain anomaly detection. *International MICCAI Brainlesion Workshop*, 133–142.

Skandarani, Y., Jodoin, P.-M., & Lalande, A. (2023). Gans for medical image synthesis: an empirical study. *Journal of Imaging*, *9*(3), 69.

Snyder, K., Whitehead, E. P., Theodore, W. H., Zaghloul, K. A., Inati, S. J., & Inati, S. K. (2021). Distinguishing type ii focal cortical dysplasias from normal cortex: a novel normative modeling approach. *NeuroImage: Clinical*, *30*, 102565.

So, E. L., & Ryvlin, P. (2015). *Mri-negative epilepsy*. Cambridge University Press.

Sohail, M., Riaz, M. N., Wu, J., Long, C., & Li, S. (2019). Unpaired multi-contrast mr image synthesis using generative adversarial networks. *Simulation and Synthesis in Medical Imaging: 4th International Workshop, SASHIMI 2019, Held in Conjunction with MICCAI 2019, Shenzhen, China, October 13, 2019, Proceedings*, 22–31.

Sone, D., & Beheshti, I. (2021). Clinical application of machine learning models for brain imaging in epilepsy: a review. *Frontiers in Neuroscience*, *15*, 761.

Spasov, S. E., Passamonti, L., Duggento, A., Lio, P., & Toschi, N. (2018). A multimodal convolutional neural network framework for the prediction of alzheimer's disease. *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1271–1274.

Stafstrom, C. E., & Carmant, L. (2015). Seizures and epilepsy: an overview for neuroscientists. *Cold Spring Harbor perspectives in medicine*, *5*(6), a022426.

Sun, L., Wang, J., Huang, Y., Ding, X., Greenspan, H., & Paisley, J. (2020). An adversarial learning approach to medical image synthesis for lesion detection. *IEEE journal of biomedical and health informatics*, *24*(8), 2303–2314.

Tajbakhsh, N., Jeyaseelan, L., Li, Q., Chiang, J. N., Wu, Z., & Ding, X. (2020). Embracing imperfect datasets: a review of deep learning solutions for medical image segmentation. *Medical Image Analysis*, *63*, 101693.

Tan, Y.-L., Kim, H., Lee, S., Tihan, T., Ver Hoef, L., Mueller, S. G., Barkovich, A. J., Xu, D., & Knowlton, R. (2018). Quantitative surface analysis of combined mri and pet enhances detection of focal cortical dysplasias. *Neuroimage*, *166*, 10–18.

Thirumagal, E., & Saruladha, K. (2020). Design of fcse-gan for dissection of brain tumour in mri. *2020 international conference on smart technologies in computing, electrical and electronics (ICSTCEE)*, 1–6.

Thung, K.-H., Yap, P.-T., & Shen, D. (2017). Multi-stage diagnosis of alzheimer's disease with incomplete multimodal data via multi-task deep learning. *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 160–168). Springer.

Thurman, D. J., Beghi, E., Begley, C. E., Berg, A. T., Buchhalter, J. R., Ding, D., Hesdorffer, D. C., Hauser, W. A., Kazis, L., Kobau, R., et al. (2011). Standards for epidemiologic studies and surveillance of epilepsy. *Epilepsia*, *52*, 2–26.

Türk, Ö., & Özerdem, M. S. (2019). Epilepsy detection by using scalogram based convolutional neural network from eeg signals. *Brain sciences*, *9*(5), 115.

Urbach, H., Kellner, E., Kremers, N., Blümcke, I., & Demerath, T. (2021). Mri of focal cortical dysplasia. *Neuroradiology*, 1–10.

Vaghari, D., Kabir, E., & Henson, R. N. (2022). Late combination shows that meg adds to mri in classifying mci versus controls. *NeuroImage*, *252*, 119054.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, *30*.

Venkadesh, K. V., Setio, A. A., Schreuder, A., Scholten, E. T., Chung, K., W. Wille, M. M., Saghir, Z., van Ginneken, B., Prokop, M., & Jacobs, C. (2021). Deep learning for malignancy risk estimation of pulmonary nodules detected at low-dose screening ct. *Radiology*, *300*(2), 438–447.

Vogelsanger, C., & Federau, C. (2021). Latent space analysis of vae and intro-vae applied to 3-dimensional mr brain volumes of multiple sclerosis, leukoencephalopathy, and healthy patients. *arXiv preprint arXiv:2101.06772*.

Wagstyl, K., Adler, S., Pimpel, B., Chari, A., Seunarine, K., Lorio, S., Thornton, R., Baldeweg, T., & Tisdall, M. (2020). Planning stereoelectroencephalography using automated lesion detection: retrospective feasibility study. *Epilepsia*, *61*(7), 1406–1416.

Wang, C., Shao, J., Lv, J., Cao, Y., Zhu, C., Li, J., Shen, W., Shi, L., Liu, D., & Li, W. (2021). Deep learning for predicting subtype classification and survival of lung adenocarcinoma on computed tomography. *Translational oncology*, *14*(8), 101141.

Wang, G., Jacob, M., Mou, X., Shi, Y., & Eldar, Y. C. (2021). Deep tomographic image reconstruction: yesterday, today, and tomorrow—editorial for the 2nd special issue "machine learning for image reconstruction". *IEEE transactions on medical imaging*, *40*(11), 2956–2964.

Wang, G., Li, W., Ourselin, S., & Vercauteren, T. (2017). Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks. *International MICCAI brainlesion workshop*, 178–190.

Wang, Y., Zhou, L., Yu, B., Wang, L., Zu, C., Lalush, D. S., Lin, W., Wu, X., Zhou, J., & Shen, D. (2018). 3d auto-context-based locality adaptive multi-modality gans for pet synthesis. *IEEE transactions on medical imaging*, *38*(6), 1328–1339.

Wang, Y., Zhou, Y., Wang, H., Cui, J., Nguchu, B. A., Zhang, X., Qiu, B., Wang, X., & Zhu, M. (2018). Voxel-based automated detection of focal cortical dysplasia lesions using diffusion tensor imaging and t2-weighted mri data. *Epilepsy & Behavior*, *84*, 127–134.

Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600–612.

Wardlaw, J. M., Smith, E. E., Biessels, G. J., Cordonnier, C., Fazekas, F., Frayne, R., Lindley, R. I., T O'Brien, J., Barkhof, F., Benavente, O. R., et al. (2013). Neuroimaging standards for research into small vessel disease and its contribution to ageing and neurodegeneration. *The Lancet Neurology*, *12*(8), 822–838.

Watanabe, S., Ueno, T., Kimura, Y., Mishina, M., & Sugimoto, N. (2021). Generative image transformer (git): unsupervised continuous image generative and transformable model for [123i] fp-cit spect images. *Annals of nuclear medicine*, *35*(11), 1203–1213.

Wei, W., Poirion, E., Bodini, B., Durrleman, S., Ayache, N., Stankoff, B., & Colliot, O. (2019). Predicting pet-derived demyelination from multimodal mri using sketcher-refiner adversarial training for multiple sclerosis. *Medical image analysis*, *58*, 101546.

Wei, W., Poirion, E., Bodini, B., Tonietto, M., Durrleman, S., Colliot, O., Stankoff, B., & Ayache, N. (2020). Predicting pet-derived myelin content from multisequence mri for individual longitudinal analysis in multiple sclerosis. *NeuroImage*, *223*, 117308.

Willmann, O., Wennberg, R., May, T., Woermann, F., & Pohlmann-Eden, B. (2007). The contribution of 18f-fdg pet in preoperative epilepsy surgery evaluation for patients with temporal lobe epilepsy: a meta-analysis. *Seizure*, *16*(6), 509–520.

Wong-Kisiel, L. C., Quiroga, D. F. T., Kenney-Jung, D. L., Witte, R. J., Santana-Almansa, A., Worrell, G. A., Britton, J., & Brinkmann, B. H. (2018). Morphometric analysis on t1-weighted mri complements visual mri review in focal cortical dysplasia. *Epilepsy Research*, *140*, 184–191.

Xu, J., Moyer, D., Grant, P. E., Golland, P., Iglesias, J. E., & Adalsteinsson, E. (2022). Svort: iterative transformer for slice-to-volume registration in fetal brain mri. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 3–13.

Xu, L., Huang, J., Nitanda, A., Asaoka, R., & Yamanishi, K. (2020). A novel global spatial attention mechanism in convolutional neural network for medical image classification. *arXiv preprint arXiv:2007.15897*.

Yaakub, S. N., McGinnity, C. J., Clough, J. R., Kerfoot, E., Girard, N., Guedj, E., & Hammers, A. (2019). Pseudo-normal pet synthesis with generative adversarial networks for localising hypometabolism in epilepsies. *International Workshop on Simulation and Synthesis in Medical Imaging*, 42–51.

Yala, A., Lehman, C., Schuster, T., Portnoi, T., & Barzilay, R. (2019). A deep learning mammography-based model for improved breast cancer risk prediction. *Radiology*, *292*(1), 60–66.

Yang, H., Sun, J., Carass, A., Zhao, C., Lee, J., Xu, Z., & Prince, J. (2018). Unpaired brain mr-to-ct synthesis using a structure-constrained cyclegan. *Deep learning in medical image analysis and multimodal learning for clinical decision support* (pp. 174–182). Springer.

Yang, Q., Li, N., Zhao, Z., Fan, X., Chang, E. I., Xu, Y., et al. (2020). Mri cross-modality image-to-image translation. *Scientific reports*, *10*(1), 1–18.

Yengi, Y., Kavak, A., & Arslan, H. (2020). Physical layer detection of malicious relays in lte-a network using unsupervised learning. *IEEE Access*, *8*, 154713–154726.

Yoganathan, K., Malek, N., Torzillo, E., Paranathala, M., & Greene, J. (2023). Neurological update: structural and functional imaging in epilepsy surgery. *Journal of Neurology*, *270*(5), 2798–2808.

Yokoi, S., Kidokoro, H., Yamamoto, H., Ohno, A., Nakata, T., Kubota, T., Tsuji, T., Morishita, M., Kawabe, T., Naiki, M., et al. (2019). Hippocampal diffusion abnormality after febrile status epilepticus is related to subsequent epilepsy. *Epilepsia*, *60*(7), 1306–1316.

Yoo, Y., Tang, L. Y., Brosch, T., Li, D. K., Kolind, S., Vavasour, I., Rauscher, A., MacKay, A. L., Traboulsee, A., & Tam, R. C. (2018). Deep learning of joint myelin and t1w mri features in normal-appearing brain tissue to distinguish between multiple sclerosis patients and healthy controls. *NeuroImage: Clinical*, *17*, 169–178.

You, S., Tezcan, K. C., Chen, X., & Konukoglu, E. (2019). Unsupervised lesion detection via image restoration with a normative prior. *International Conference on Medical Imaging with Deep Learning*, 540–556.

Yu, J., Yang, D., & Zhao, H. (2021). Ffanet: feature fusion attention network to medical image segmentation. *Biomedical Signal Processing and Control*, *69*, 102912.

Yu, Y., Xie, Y., Thamm, T., Gong, E., Ouyang, J., Huang, C., Christensen, S., Marks, M. P., Lansberg, M. G., Albers, G. W., et al. (2020). Use of deep learning to predict final ischemic stroke lesions from initial magnetic resonance imaging. *JAMA network open*, *3*(3), e200772–e200772.

Yuan, Y. (2020). Automatic brain tumor segmentation with scale attention network. *International MICCAI Brainlesion Workshop*, 285–294.

Zaharchuk, G., Gong, E., Wintermark, M., Rubin, D., & Langlotz, C. (2018). Deep learning in neuroradiology. *American Journal of Neuroradiology*, *39*(10), 1776–1784.

Zhang, C., Song, Y., Liu, S., Lill, S., Wang, C., Tang, Z., You, Y., Gao, Y., Klistorner, A., Barnett, M., et al. (2018). Ms-gan: gan-based semantic segmentation of multiple sclerosis lesions in brain magnetic resonance imaging. *2018 Digital Image Computing: Techniques and Applications (DICTA)*, 1–8.

Zhang, Q., Liao, Y., Wang, X., Zhang, T., Feng, J., Deng, J., Shi, K., Chen, L., Feng, L., Ma, M., et al. (2021). A deep learning framework for 18f-fdg pet imaging diagnosis in pediatric patients with temporal lobe epilepsy. *European journal of nuclear medicine and molecular imaging*, *48*(8), 2476–2485.

Zhang, R., Isola, P., Efros, A. A., Shechtman, E., & Wang, O. (2018). The unreasonable effectiveness of deep features as a perceptual metric. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 586–595.

Zhang, W., Yang, G., Huang, H., Yang, W., Xu, X., Liu, Y., & Lai, X. (2021). Me-net: multi-encoder net framework for brain tumor segmentation. *International Journal of Imaging Systems and Technology*, *31*(4), 1834–1848.

Zhang, X., He, X., Guo, J., Ettehadi, N., Aw, N., Semanek, D., Posner, J., Laine, A., & Wang, Y. (2021). Ptnet: a high-resolution infant mri synthesizer based on transformer. *arXiv preprint arXiv:2105.13993*.

Zhou, B., An, D., Xiao, F., Niu, R., Li, W., Li, W., Tong, X., Kemp, G. J., Zhou, D., Gong, Q., et al. (2020). Machine learning for detecting mesial temporal lobe epilepsy by structural and functional neuroimaging. *Frontiers of Medicine*, *14*(5), 630–641.

Zhou, S. K., Greenspan, H., Davatzikos, C., Duncan, J. S., Van Ginneken, B., Madabhushi, A., Prince, J. L., Rueckert, D., & Summers, R. M. (2021). A review of deep learning in medical imaging: imaging traits, technology trends, case studies with progress highlights, and future promises. *Proceedings of the IEEE*, *109*(5), 820–838.

Zhou, T., Thung, K.-H., Zhu, X., & Shen, D. (2019). Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis. *Human brain mapping*, *40*(3), 1001–1016.

Zhou, T., Canu, S., & Ruan, S. (2020). Fusion based on attention mechanism and context constraint for multi-modal brain tumor segmentation. *Computerized Medical Imaging and Graphics*, *86*, 101811.

Zhou, T., Canu, S., Vera, P., & Ruan, S. (2021). Latent correlation representation learning for brain tumor segmentation with missing mri modalities. *IEEE Transactions on Image Processing*, *30*, 4263–4274.

Zhou, T., Ruan, S., & Canu, S. (2019). A review: deep learning for medical image segmentation using multi-modality fusion. *Array*, *3*, 100004.

Zhu, J.-Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proceedings of the IEEE international conference on computer vision*, 2223–2232.

Zhu, S. (2022). Early diagnosis of parkinson's disease by analyzing magnetic resonance imaging brain scans and patient characteristic. *2022 10th International Conference on Bioinformatics and Computational Biology (ICBCB)*, 116–123.

Zhu, Y., Wang, S., Lin, R., Hu, Y., & Chen, Q. (2021). Brain tumor segmentation for missing modalities by supplementing missing features. *2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, 652–656.

Zimmerer, D., Petersen, J., & Maier-Hein, K. (2019). High-and low-level image component decomposition using vaes for improved reconstruction and anomaly detection. *arXiv preprint arXiv:1911.12161*.

Zotova, D., Jung, J., & Lartizien, C. (2021). Gan-based synthetic fdg pet images from t1 brain mri can serve to improve performance of deep unsupervised anomaly detection models. *Simulation and Synthesis in Medical Imaging: 6th International Workshop, SASHIMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 6*, 142–152.

FOLIO ADMINISTRATIF

THESE DE L'UNIVERSITE DE LYON OPEREE AU SEIN DE L'INSA LYON

NOM : ZOTOVA DATE de SOUTENANCE : 30 / 05 / 2024

Prénom : Daria

TITRE : Deep brain unsupervised anomaly detection model based on multymodality imaging

NATURE : Doctorat Numéro d'ordre :

Ecole doctorale : Electronique, Electrotechnique, Automatique

Spécialité : Traitement du Signal et de l'Image

RESUME:
Selon l'Organisation mondiale de la santé (OMS), près de 65 millions de personnes dans le monde sont touchées par l'épilepsie. L'épilepsie est un trouble neurologique chronique caractérisé principalement par des interruptions récurrentes et imprévisibles du fonctionnement normal du cerveau, appelées crises d'épilepsie. Pour un tiers des patients diagnostiqués, les crises ne peuvent être contrôlées par la pharmacothérapie. Pour ces patients, le traitement consisterait à effectuer une résection chirurgicale de la zone épileptogène. Le succès de ces opérations dépend en grande partie de la précision de la localisation de la zone épileptogène. La neuro-imagerie, y compris l'imagerie par résonance magnétique (IRM) et la tomographie par émission de positons (TEP), est de plus en plus souvent prise en compte dans les examens préchirurgicaux de routine.

Ce travail tente d'améliorer un système d'aide au diagnostic (CAD) pour la détection des lésions épileptogènes, en s'appuyant sur des données de neuro-imagerie multimodales, en s'appuyant sur les bases posées dans des travaux antérieurs. Le système proposé utilise des réseaux siamois profonds non supervisés pour apprendre les représentations normales du cerveau à partir de scanners non pathologiques, suivis d'une série de modèles SVM à une classe au niveau du voxel pour générer une carte de score d'anomalie. Le modèle, initialement testé sur des examens IRM pondérés en T1 et FLAIR, a démontré une sensibilité de 61%. L'inclusion de données d'imagerie TEP est envisagée pour évaluer sa valeur supplémentaire. Cependant, en raison des considérations éthiques entourant l'exposition aux radiations chez les individus sains, le développement et l'intégration de données PET synthétiques, simulant des scanners cérébraux sains, apparaît comme une solution potentielle.

Ce travail ne fait pas seulement progresser le domaine de l'imagerie médicale dans la recherche sur l'épilepsie, mais offre également une feuille de route pour les études futures afin d'affiner et d'améliorer la détection et la localisation des lésions épileptogènes, ce qui pourrait conduire à de meilleurs résultats chirurgicaux et à de meilleurs soins pour les patients.

MOTS-CLÉS : unsupervised representation learning, multymodality fusion, siamese networks, medical image synthesis, gans, outlier detection, one class SVM, computer aided diagnosis, epilepsy

Laboratoire (s) de recherche : CREATIS, CNRS UMR 5220 – INSERM U1294 – Université Lyon 1 – INSA Lyon – Université Jean Monnet Saint-Etienne.

Directeur de thèse: Carole Lartizien

Président de jury : Hammers Alexander

Composition du jury :

Oliver, Arnau	Professeur des Universités	University of Girona	Rapporteur
Burgos, Ninon	Chargée de recherche	CNRS	Rapporteur
Hammers, Alexander	PU-PH	King's College London	Examinateur
Jung, Julien	Professeur associé	Hospices Civils de Lyon	Examinateur
Lartizien, Carole	Directeur de recherche	INSA Lyon (CNRS)	Directeur de thèse