

Segmentação Automática de Infarto Cerebral Isquêmico em Tomografia Computadorizada Não Contrastada Utilizando Deep Learning para Apoio ao Diagnóstico Clínico de AVC

DANIEL HARDY MELO

Trabalho de Conclusão de Curso
MBA em Inteligência Artificial e Big Data

UNIVERSIDADE DE SÃO PAULO

Instituto de Ciências Matemáticas e de Computação

Segmentação Automática de Infarto Cerebral
Isquêmico em Tomografia Computadorizada Não
Contrastada Utilizando Deep Learning para Apoio ao
Diagnóstico Clínico de AVC

DANIEL HARDY MELO

DANIEL HARDY MELO

Segmentação Automática de Infarto Cerebral Isquêmico em Tomografia Computadorizada Não Contrastada Utilizando Deep Learning para Apoio ao Diagnóstico Clínico de AVC

Trabalho de conclusão de curso apresentado ao Departamento de Ciências de Computação do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo - ICMC/USP, como parte dos requisitos para obtenção do título de Especialista em Inteligência Artificial e Big Data.

Área de concentração: Inteligência Artificial.

Orientador: Prof. Dr. José Fernando Rodrigues Jr.

USP - São Carlos

2025

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

M528s MELO, DANIEL HARDY
Segmentação Automática de Infarto Cerebral
Isquêmico em Tomografia Computadorizada Não
Contrastada Utilizando Deep Learning para Apoio ao
Diagnóstico Clínico de AVC / DANIEL HARDY MELO;
orientador JOSÉ FERNANDO RODRIGUES JUNIOR. -- São
Carlos, 2025.
91 p.

Tese (Doutorado - MBA em Inteligência Artificial
e Big Data) -- Instituto de Ciências Matemáticas e
de Computação, Universidade de São Paulo, 2025.

1. ACIDENTE VASCULAR CEREBRAL. 2. INTELIGÊNCIA
ARTIFICIAL. 3. BIG DATA. 4. TOMADA DE DECISÃO. 5.
SEGMENTAÇÃO. I. RODRIGUES JUNIOR, JOSÉ FERNANDO,
orient. II. Título.

DEDICATÓRIA

À minha esposa Loíse e aos meus filhos Murilo e Massimo, pelo apoio e compreensão durante todos os momentos da execução deste trabalho e nos momentos de ausência, mas sobretudo pelo vosso amor.

Aos meus pais José Luis e Regina, alicerces da minha vida, exemplos de dedicação, amor e carinho. Agradeço a minha formação pessoal e por ensinarem os verdadeiros valores.

AGRADECIMENTOS

Ao meu orientador, Professor Dr. José Fernando Rodrigues Jr, Professor Associado do Departamento de Ciências da Computação do Instituto de Ciências Matemáticas e da Computação da Universidade de São Paulo (ICMC/USP), pelo seu apoio, pela confiança e por me facultar a oportunidade de vencer mais uma etapa em minha jornada.

Ao estimado grupo de professores afiliados ao Departamento de Ciências da Computação do Instituto de Ciências Matemáticas e da Computação da Universidade de São Paulo (ICMC/USP), ilustres docentes envolvidos no programa de pós-graduação MBA em Inteligência Artificial e Big Data liderados pela Professora Dra. Solange Oliveira Rezende e Professora Dra. Roseli Aparecida Francelin Romero, cujos conhecimentos e *insights* coletivos contribuíram muito para minha experiência educacional.

Aos meus colegas da 4ª turma do MBA em IA e Big Data, cujas experiências e interações compartilhadas foram importantes para enriquecer minha jornada educacional, ampliar as perspectivas intelectuais e reforçar o compromisso com a inovação e a excelência acadêmica.

RESUMO

MELO, D. H. **Segmentação Automática de Infarto Cerebral Isquêmico em Tomografia Computadorizada Não Contrastada Utilizando Deep Learning para Apoio ao Diagnóstico Clínico de AVC.** 2025. 90 f. Monografia (MBA em IA e Big Data) – Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

Recentes avanços no desenvolvimento de metodologias de deep learning têm demonstrado ser ferramentas promissoras para otimizar a avaliação de imagens médicas, visando auxiliar os profissionais na tomada de decisão. O presente estudo foi projetado para avaliar modelos de aprendizado profundo adaptados para as tarefas de segmentação e classificação relativas a imagens de tomografia computadorizada sem contraste, com ênfase na identificação e localização de lesões cerebrais isquêmicas causados por acidente vascular cerebral. Para alcançar este objetivo, uma seleção de sete arquiteturas contemporâneas de redes neurais profundas foi feita após uma revisão da literatura e adaptadas às limitações dos recursos computacionais. Estas incluem: U-Net, U-Net++, R2U-Net, V-Net, Attention U-Net, Swin U-Net e ResUnet-a, cada uma com características únicas na realização da tarefa proposta. A implementação de cada um desses modelos foi executada utilizando bibliotecas de código aberto, e eles foram posteriormente treinados em um conjunto de dados rotulado meticulosamente selecionado, que consistia em imagens tomografia computadorizada sem contraste originais que exibiam uma variedade de características morfológicas representativas das lesões isquêmicas encontradas em cenários clínicos habituais. A estrutura metodológica empregada nesta investigação abrangeu várias etapas críticas, incluindo o pré-processamento abrangente das imagens, definição precisa do pipeline analítico, ajuste de hiperparâmetros e a avaliação sistemática do desempenho utilizando um conjunto de métricas de segmentação padronizadas que incluem, mas não estão limitadas ao coeficiente de Dice, interseção sobre união, precisão, sensibilidade, especificidade, F1-Score, área sob a curva (AUC) e precisão média (AP). Os resultados do estudo revelaram que todas as arquiteturas escolhidas demonstraram a capacidade de executar a tarefa de segmentação com um grau satisfatório de eficácia, ao mesmo tempo em que exibem diversos níveis de eficácia de desempenho entre elas. A arquitetura V-Net mostrou-se como a mais eficaz em avaliações clínicas, atingindo uma pontuação de coeficiente Dice de 0,6897 e uma precisão média de 0,7502, com as estruturas ResUnet-a e R2U-Net acompanhando de perto em termos de desempenho, ambas demonstrando eficácia notável. A classificação do ponto de vista da competência na aplicação clínica desses modelos foi formulada, concentrando-se nas métricas mais relevantes que têm importância substancial na prática médica, priorizando assim os modelos que exibiram habilidades aprimoradas na identificação das regiões afetadas e, ao mesmo tempo, reduzindo a incidência de falsos negativos. Em resumo, este estudo sugere que a implantação de redes neurais profundas para a segmentação do AVC isquêmico em imagens de tomografia computadorizada sem contraste é tecnicamente viável e possui o potencial de fornecer assistência substancial nos processos de tomada de decisão clínica, especialmente em ambientes de urgência e emergência. Além disso, o estudo destacou a importância de escolher arquiteturas adequadas e modelos de ajuste fino como componentes fundamentais para alcançar resultados confiáveis dentro da estrutura de ambientes médicos.

Palavras-chave: AVC isquêmico. Aprendizado profundo. Segmentação de imagens. Tomografia computadorizada. Redes neurais convolucionais.

ABSTRACT

MELO, D. H. **Automated Ischemic Stroke Segmentation in Non-Contrast CT using Deep Learning**. 2025. 90 f. Monografia (MBA em IA e Big Data) – Centro de Ciências Matemáticas Aplicadas à Indústria, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2025.

Recent advances in the development of deep learning methodologies have shown to be promising tools to optimize the evaluation of medical images, aiming to assist professionals in decision-making. The present study was designed to evaluate deep learning models adapted for segmentation and classification tasks related to non-contrast computed tomography images, with emphasis on the identification and localization of ischemic brain lesions caused by stroke. To achieve this goal, a selection of seven contemporary architectures of deep neural networks was made after a literature review and adapted to the limitations of computational resources. These include: U-Net, U-Net++, R2U-Net, V-Net, Attention U-Net, Swin U-Net and ResUnet-a, each with unique characteristics in carrying out the proposed task. The implementation of each of these models was performed using open source libraries, and they were subsequently trained on a meticulously selected labeled dataset, which consisted of original uncontrasted computed tomography images that exhibited a variety of morphological characteristics representative of the ischemic lesions found in usual clinical scenarios. The methodological structure employed in this research covered several critical steps, including comprehensive image preprocessing, precise definition of the analytical pipeline, adjustment of hyperparameters and systematic evaluation of performance using a set of standardized segmentation metrics that include, but are not limited to Dice coefficient, Dice, IoU, accuracy, precision, sensitivity, specificity, F1-Score, AUC, AP. The results of the study revealed that all architectures chosen demonstrated the ability to perform the segmentation task with a satisfactory degree of effectiveness, while displaying several levels of performance effectiveness among them. The V-Net architecture proved to be the most effective in clinical evaluations, achieving a Dice coefficient score of 0.6897 and an average accuracy of 0.7502, with the structures ResUnet-a and R2U-Net closely following in terms of performance, both demonstrating remarkable effectiveness. The classification from the point of view of competence in clinical application of these models was formulated, focusing on the most relevant metrics that have substantial importance in medical practice, prioritizing models that exhibit improved skills in identifying the affected regions and, at the same time, reducing the incidence of false negatives. In summary, this study suggests that the implantation of deep neural networks for ischemic stroke segmentation in non-contrast computed tomography images is technically feasible and has the potential to provide substantial assistance in clinical decision-making processes, especially in urgency and emergency environments. In addition, the study highlighted the importance of choosing appropriate architectures and fine-tuning models as key components to achieve reliable results within the framework of medical environments.

Keywords: ischemic stroke. deep learning. image segmentation. computed tomography. convolutional neural networks.

LISTA DE ILUSTRAÇÕES

Figura 1 – Arquitetura U-Net	24
Figura 2 – Arquitetura V-Net	26
Figura 3 – Arquitetura U-Net++	28
Figura 4 – Arquitetura R2U-Net	30
Figura 5 – Arquitetura Attention U-Net	32
Figura 6 – Arquitetura ResUNet-a	34
Figura 7 – Arquitetura Swin U-Net	36
Figura 8a – Visualização da predição em TC de AVC isquêmico U-Net – caso 1	52
Figura 8b – Visualização da predição em TC de AVC isquêmico U-Net – caso 2	52
Figura 9a – Visualização da predição em TC de AVC isquêmico U-Net++ – caso 1	54
Figura 9b – Visualização da predição em TC de AVC isquêmico U-Net++ – caso 2	54
Figura 10a – Visualização da predição em TC de AVC isquêmico R2U-Net – caso 1	57
Figura 10b – Visualização da predição em TC de AVC isquêmico R2U-Net – caso 2	57
Figura 11a – Visualização da predição em TC de AVC isquêmico V-Net – caso 1	59
Figura 11b – Visualização da predição em TC de AVC isquêmico V-Net – caso 2	60
Figura 12a – Visualização da predição em TC de AVC isquêmico Attention U-Net – caso 1	62
Figura 12b – Visualização da predição em TC de AVC isquêmico Attention U-Net – caso 2	62
Figura 13a – Visualização da predição em TC de AVC isquêmico ResUNet – caso 1	64
Figura 13b – Visualização da predição em TC de AVC isquêmico ResUNet – caso 2	65
Figura 14a – Visualização da predição em TC de AVC isquêmico Swin U-Net – caso 1	67
Figura 14b – Visualização da predição em TC de AVC isquêmico Swin U-Net – caso 2	78

LISTA DE TABELAS

Tabela 1 – Computador e sistema utilizado	49
Tabela 2 – Métricas da arquitetura U-Net	51
Tabela 3 – Métricas da arquitetura U-Net++	53
Tabela 4 – Métricas da arquitetura R2U-Net	54
Tabela 5 – Métricas da arquitetura V-Net	58
Tabela 6 – Métricas da arquitetura Attention U-Net	61
Tabela 7 – Métricas da arquitetura ResUNet-a	64
Tabela 8 – Métricas da arquitetura Swin U-Net	66
Tabela 9 – Valores médios das métricas nas diferentes arquiteturas	73

LISTA DE GRÁFICOS

Gráfico 1 – Aprendizado gerados a partir da arquitetura U-Net	51
Gráfico 2 – Aprendizado gerados a partir da arquitetura U-Net++	53
Gráfico 3 – Aprendizado gerados a partir da arquitetura R2U-Net	55
Gráfico 4 – Aprendizado gerados a partir da arquitetura V-Net	58
Gráfico 5 – Aprendizado gerados a partir da arquitetura Attention U-Net	60
Gráfico 6 – Aprendizado gerados a partir da arquitetura ResUNet	63
Gráfico 7 – Aprendizado gerados a partir da arquitetura Swin U-Net	66
Gráfico 8 – Desempenho na detecção e classificação do AVC isquêmico	73
Gráfico 9 – Desempenho na segmentação do AVC isquêmico	74

LISTA DE ABREVIATURAS E SIGLAS

ACM	–	artéria cerebral média
AP	–	average precision
AUC	–	area under the curve
AVC	–	acidente vascular cerebral
AVCi	–	acidente vascular cerebral isquêmico
CUDA	–	Compute Unified Device Architecture
Dice	–	Dice similarity coefficient
FN	–	falso negativo
FP	–	falso positivo
GAN	–	rede adversária generativa
GPU	–	unidade de processamento gráfico
IoU	–	Intersection over Union
OMS	–	Organização Mundial da Saúde
p.p	–	pontos percentuais
PET	–	tomografia por emissão de pósitrons
PSP	–	<i>Pyramid Scene Parsing</i>
RMN	–	ressonância magnética nuclear
RNA	–	rede neural artificial
RNC	–	redes neurais convolucionais
ROC	–	receiver operating characteristic curve
SPECT	–	tomografia por emissão de pósitrons
SUS	–	Sistema Único de Saúde
TC	–	tomografia computadorizada
TCSC	–	tomografia computadorizada sem contraste
USG	–	ultrassonografia

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 Objetivos.....	16
2 REVISÃO BIBLIOGRÁFICA.....	16
2.1 Acidente Vascular Cerebral Isquêmico e Diagnóstico por Imagem	17
2.2 Origem da Aplicação de Inteligência Artificial em Imagens Médicas	19
2.3 Técnicas de Segmentação de Imagens Médicas	22
2.3.1 U-Net: arquitetura original de segmentação biomédica	23
2.3.2 V-Net: segmentação volumétrica 3D	25
2.3.3 U-Net++: conexões aninhadas para reduzir lacuna semântica	27
2.3.4 R2U-Net: U-Net com convoluções recorrentes e residuais	29
2.3.5 Attention U-Net: aprendendo a focar em regiões importantes	31
2.3.6 ResUNet-a: combinação de resíduos, dilatação, pirâmide e multi-tarefas	32
2.3.7 Swin U-Net: <i>transformers</i> puros em estrutura na forma de U	35
2.4 Estudos Recentes sobre Segmentação Automática de AVC	37
2.5 Considerações Finais	42
3 METODOLOGIA	44
3.1 Aquisição e Seleção de Dados	45
3.1.1 Critérios de inclusão	45
3.1.1 Critérios de exclusão	46
3.2 Pré-processamento	46
3.2.1 Divisão das imagens e descarte de regiões irrelevantes	46
3.2.2 Aumento do conjunto de dados	46
3.2.3 Divisão do conjunto de dados	47
3.2.4 Limites de pós-processamento e segmentação	47
3.3 Definições de redes neurais	47
3.4 Métricas de Avaliação e Visualização	49
4 RESULTADOS	50
4.1 U-Net	50
4.2 U-Net ++	52
4.3 R2 U-Net	55
4.4 V-Net	57

4.5 Attetion U-Net	60
4.6 ResUNet-a	62
4.7 Swin U-Net	65
5 DISCUSSÃO	68
6 CONCLUSÕES E PERSPECTIVAS FUTURAS	74
REFERÊNCIAS.....	76
APÊNDICE	76

1 INTRODUÇÃO

O acidente vascular cerebral (AVC) tem sua importância inequívoca por se uma das doenças mais prevalentes na especialidade de Neurologia das principais causas de morte, a maior causa de incapacidade no mundo (MARTINS et al, 2013; VIRANI et al., 2020). De acordo com a Organização Mundial da Saúde (OMS), todos os anos acontecem cerca de 12,2 milhões de novos casos de acidente vascular cerebral (AVC). Destes, aproximadamente 6,5 milhões dos casos evoluirão para o óbito, o que mostra a gravidade do problema em escala mundial (WORLD HEALTH ORGANIZATION, 2023). Quando olhamos apenas para os casos de AVC isquêmico, percebe-se que quase metade dos sobreviventes carrega algum tipo de sequela permanente. Esta perda de autonomia em vários níveis e piora da qualidade de vida, além de exigir consultas médicas frequentes e processos de reabilitação, muitas vezes se tornam parte da rotina do paciente (AMERICAN STROKE ASSOCIATION, 2023).

No Brasil, os dados estão em consonância com essa tendência. Entre 2019 e 2023, foram registrados mais de 174 mil óbitos por acidente vascular cerebral, embora haja uma tendência a redução nos últimos anos (VIEIRA et al., 2020; PINHEIRO et al., 2023). Este comportamento na redução dos casos tem sido atribuído aos avanços nas ações de prevenção, melhorias no tratamento do AVC ou mudanças no perfil etário da população (SANTOS et al., 2025). A taxa de mortalidade varia entre 95 e 130 óbitos por 100 mil habitantes. A maior quantidade dos óbitos ocorre nas regiões Sudeste e Nordeste, porém o impacto desta doença repercute também nos aspectos econômicos e sociais (SANTOS et al., 2025). O Sistema Único de Saúde (SUS) já registrou mais de 200 mil internações anuais, resultando em gastos superiores a R\$ 300 milhões por ano apenas com hospitalizações e reabilitação (PINHEIRO et al., 2024). Além disso, um estudo realizado pela Academia Brasileira de Neurologia demonstrou que menos de 40% dos pacientes com AVC isquêmico recebem atendimento dentro da janela terapêutica ideal, porém nenhum recebeu terapia de reperfusão (JUNIOR et al., 2022).

O AVC pode ser classificado em dois tipos: isquêmico e hemorrágico. O tipo isquêmico representa aproximadamente 80% do total de AVC e é causado pelo suprimento insuficiente de sangue para uma região do cérebro, levando à isquemia e à morte celular (FAN et al., 2023). Já o AVC hemorrágico resulta da ruptura de um vaso sanguíneo, provocando extravasamento de sangue no tecido cerebral. Ambos os tipos de AVC exigem intervenção médica imediata, pois o tempo decorrido desde o início dos sintomas influencia diretamente no prognóstico do paciente (MOHAPATRA et al., 2023).

Atualmente, é estabelecida uma janela terapêutica ideal para o uso da trombólise intravenosa de até 4,5 horas após o início dos sintomas e da trombectomia mecânica, que pode ser realizada em casos selecionados em até 24 horas (NISHI et al., 2023).

Sabe-se que o principal exame para avaliação inicial do AVC isquêmico é a TCSC. No entanto, a identificação precoce da isquemia ainda é um desafio significativo, pois a TCSC apresentar baixa sensibilidade na detecção desses infartos cerebrais, especialmente nas primeiras 6 horas após o evento (SAHOO et al., 2022). O desafio da segmentação visual é agravado pelo baixo contraste das imagens de TCSC e pela urgência em diagnosticar rapidamente as possíveis lesões isquêmicas, que variam em localização e gravidade (SAHOO et al., 2022; MOHAPATRA et al., 2023). Portanto, é comum que estas lesões isquêmicas iniciais não sejam percebidas na avaliação visual e dificultem a tomada de decisão sobre o início do tratamento.

Além das limitações técnicas das TCSC, a interpretação das imagens depende de radiologistas experientes, tornando o processo suscetível a falhas e variações nas avaliações entre os observadores (SAHOO et al., 2022). Estudos indicam que a precisão diagnóstica do AVC isquêmico inicial por TCSC pode variar entre 50% e 70%, dependendo da experiência desse profissional, da qualidade da imagem obtida e do método de visualização do observador (SAHOO, P. K. et al., 2022; MOHAPATRA et al., 2023).

Existe uma crescente necessidade de ferramentas avançadas que ajudem no diagnóstico de AVCs isquêmicos para reduzir o aspecto subjetivo da interpretação médica e aumentar a precisão na identificação de lesões cerebrais isquêmicas em exames de TCSC. Automatizar o processo de segmentação não apenas agilizaria o atendimento inicial ao paciente, mas também auxiliaria no acompanhamento da evolução do tratamento e na previsão da extensão final do infarto (SAHOO et al., 2022; NISHI et al., 2023).

Nas últimas décadas, o desenvolvimento de metodologias de *deep learning* demonstrou ser uma grande promessa para otimizar a segmentação de AVC isquêmico em exames de imagem, para contribuir com o diagnóstico precoce e o desfecho favorável de pacientes com AVC. As RNCs oferecem vantagens significativas na segmentação de imagens de TC em AVC isquêmico, incluindo alta precisão, automação completa do processo diagnóstico, redução do tempo de análise, detecção precoce de lesões sutis e reprodutibilidade superior aos métodos manuais. (SAHOO et al., 2022).

O desenvolvimento de um modelo de segmentação de infarto isquêmico com *deep learning* em TCSC pode, portanto, ser um diferencial para um atendimento mais eficiente e

personalizado aos pacientes com AVC, com possíveis impactos positivos na sua recuperação.

1.1 Objetivos

Portanto, este trabalho visa obter um modelo de *deep learning* capaz de segmentar automaticamente as áreas de infarto isquêmico em imagens de TCSC de forma precisa, com o propósito de auxiliar no diagnóstico precoce e no planejamento do tratamento de pacientes com AVC isquêmico.

Objetivos gerais:

- modelar a segmentação e classificação de imagens para compreender seus fatores determinantes, adaptando o modelo a características específicas;
- estudar e selecionar arquiteturas de *deep learning* adequadas ao problema e viáveis dentro do escopo deste trabalho.
- definir um arcabouço de *deep learning* capaz de processar, segmentar, e classificar imagens de TCSC;
- desenvolver uma atividade prática de alta complexidade envolvendo as técnicas mais atuais de Ciência de Dados e Inteligência Artificial.

Objetivos específicos:

- revisar a literatura e os recursos de software mais atuais para definir um pipeline analítico para a classificação de imagens reais de TCSC;
- elencar um conjunto de imagens com variabilidade e rotulação que possa ser usado para treinar e avaliar um modelo de segmentação e classificação;
- realizar experimentos com os recursos reunidos usando software de acesso aberto que envolvam arquiteturas avançadas de *deep learning*;
- coletar métricas que permitam avaliar o quanto dos objetivos gerais foram alcançados, sobretudo no auxílio à prática clínica de diagnóstico.

2 REVISÃO BIBLIOGRÁFICA

A identificação imediata e o diagnóstico preciso desse incidente cerebrovascular são de extrema relevância para estabelecer intervenções terapêuticas oportunas e reduzir as

repercussões funcionais nos pacientes (FAN et al., 2023). A evolução da inteligência artificial (IA) e das tecnologias de aprendizado profundo facilitou melhorias notáveis na avaliação de imagens médicas, particularmente na detecção imediata de distúrbios neurológicos, como o AVC isquêmico.

Consequentemente, a utilização de metodologias sofisticadas para o delineamento automatizado de lesões isquêmicas no TCSC é uma promessa considerável para melhorar a precisão do diagnóstico e auxiliar nos processos de tomada de decisão clínica. Esta seção elucidará os princípios fundamentais relativos ao AVC isquêmico, processamento de imagens e técnicas de segmentação, juntamente com uma revisão das pesquisas mais contemporâneas neste domínio que sustentam o avanço desta investigação.

2.1 Acidente Vascular Cerebral Isquêmico e Diagnóstico por Imagem

A diminuição crítica do fluxo sanguíneo para uma área de tecido cerebral (ou para a medula espinal ou para o tecido da retina), podendo resultar em hipóxia e morte celular na região afetada, caracteriza a fisiopatologia do AVCi (NISHI et al., 2023).

Os principais fatores que elevam a probabilidade de um AVCi abrangem condições médicas como diabetes mellitus, hipertensão, sedentarismo, uso de tabaco, doenças cardiovasculares, estresse psicológico e distúrbios emocionais, além da hiperlipidemia, obesidade e suscetibilidade hereditária. Além destes, estima-se que aproximadamente 15% das ocorrências de AVCi sejam devido à aterosclerose nas artérias carótidas. A utilização de anticoncepcionais hormonais pode agravar ainda mais esse risco (VIEIRA et al., 2020).

O AVCi é diagnosticado clinicamente e definido como o início agudo de sintomas neurológicos focais como fraqueza ou paralisia, dormência ou formigamento, alterações visuais ou de fala, convulsões ou dificuldades cognitivas. Estes aparecem subitamente e incluem fraqueza ou formigamento de um lado do corpo, paralisia facial, dificuldades na fala, alteração no nível de consciência, crises convulsivas, déficits visuais, vertigem, perda do equilíbrio e dificuldades para caminhar. Além disso, podem ocorrer dores de cabeça intensa, perda de consciência parcial ou total e confusão mental (VIEIRA et al., 2020). As consequências do AVC variam de acordo com a localização e extensão do dano cerebral, além da rapidez para o início do tratamento. As sequelas mais prevalentes são déficits motores, deficiências cognitivas,

dificuldades na comunicação (afasia) e distúrbios emocionais (AMERICAN STROKE ASSOCIATION, 2023).

Na abordagem inicial do paciente com suspeita clínica de AVC isquêmico, a TCSC continua sendo amplamente utilizada devido à sua rapidez e disponibilidade. No entanto, sua sensibilidade limitada na detecção de alguns padrões de lesões, especialmente as precoces, vem canalizando o interesse dos pesquisadores na identificação de sinais radiológicos sutis e precoces do infarto cerebral e da oclusão arterial, além da realização do estudo arterial (angiotomografia de vasos intracranianos e cervicais) e estudo de perfusão (SAHOO et al., 2022; VAN POPPEL et al., 2022). A escala de ASPECTS (*Alberta Stroke Program Early CT Score*) visa quantificar a área isquêmica na fase aguda do AVCi. Uma TC de crânio normal tem uma escala ASPECTS de 10. Esse valor diminui à medida que aumenta a área isquêmica. Devem ser valorizados alguns sinais precoces de isquemia, como a perda de diferenciação entre a substância branca e cinzenta dos núcleos lentiformes e o apagamento de sulcos (PEXMAN et al., 2001).

Na fase inicial do AVCi, há configurações específicas na RMN de crânio que podem ser executadas em menos de 20 minutos como difusão, perfusão, gradiente eco e *fluid attenuated inversion recovery* (Flair), evitando atrasos na condução terapêutica (AOKI et al., 2021). A sequência de difusão permite visualizar a região com isquemia em poucos minutos do início dos sintomas. Tem maior sensibilidade do que a TC de crânio para detectar isquemia em fossa posterior (cerebelo e tronco encefálico) e permite maior acurácia para determinar o tamanho da região isquêmica (JOHANSEN et al., 2024). A sequência Flair só se altera em média 4 horas após o *ictus* do AVC. Dessa forma, o estudo das sequências de difusão e Flair pode favorecer a indicação em pacientes candidatos à trombólise (THOMALLA et al., 2019). Já a sequência de gradiente eco é tão eficaz quanto a TC em detectar AVCs hemorrágicos (JOHANSEN et al., 2024).

Tanto a TC quanto a RMN, caso tenham o software adequado, são capazes de realizar o estudo de perfusão. Ele diferencia uma área central necrótica, denominada núcleo (ou core) isquêmico, de uma região potencialmente recuperável com terapias de reperfusão, chamada de penumbra. A diferença entre essas áreas é o *mismatch*. Se ainda existe uma grande área de penumbra, com *mismatch* importante, significa que uma terapia de reperfusão pode trazer benefícios (TAKÁCS et al., 2025). Selecionando pacientes com perfusão adequada, estudos clínicos conseguiram estender o tempo de trombólise até 9 horas e o tempo de trombectomia mecânica até 24 horas com resultados positivos (AHMED et al., 2024).

2.2 Origem da Aplicação de Inteligência Artificial em Imagens Médicas

Ao longo dos anos, as imagens médicas foram se consolidando como uma das principais fontes de informação para o diagnóstico, terapêutica e nas pesquisas em medicina, biologia e em outros campos relacionados. Através dos aspectos morfológicos das estruturas examinadas e de seus aspectos funcionais (embora em um grau limitado), permitiu-nos inferir sobre sua condição fisiológica normal ou a presença de um estado patológico. A partir daí, nos é possível chegar a conclusões fundamentadas e contribuir no processo de raciocínio diagnóstico e propostas terapêuticas.

As imagens médicas são obtidas através de vários processos físicos. A produção dessas imagens ainda usa técnicas clássicas de raios-X com base no fato de que a quantidade de energia de raios-X absorvida por tecidos específicos difere. A técnica da TC, frequentemente utilizada, envolve medições precisas da quantidade total de raios-X absorvidas ao longo de várias linhas que passam pelo corpo do paciente, em vários pontos e em várias inclinações, fornecendo imagens transversais de cada seção do paciente. Outros processos físicos que envolvem a diferenciação entre os tecidos do corpo, dependendo de sua densidade e de outras propriedades, nos permite combinar o mapeamento da imagem e a morfologia dos órgãos do corpo com as informações que necessitamos. Esses métodos incluem técnicas de imagem que utilizam RMN, PET, SPECT entre outros. Outras técnicas envolvem a USG produzindo imagens de órgãos e tecidos utilizando a absorção e a reflexão das ondas de ultrassom que passam através do corpo e são captadas de receptor especial (transdutor); e ainda, há métodos que consistem na visualização da radiação ionizante introduzida no corpo do paciente, de preferência na forma de um composto químico que participa nos processos de metabolismo, e são ativamente absorvidos ou removidos dos órgãos que precisam ser examinados.

A tecnologia de imagens médicas evoluiu de modalidades baseadas em filme (analógicas) para modalidades de imagens digitais, especialmente a partir da década de 1980, onde foi possível capturar essas imagens com qualidade suficiente para análise em computadores. Como essa tarefa requer o uso de uma grande base de conhecimento para extrair várias características das imagens, bem como interpretá-las, a imagem médica está, portanto, bem posicionada para aplicações de IA (GIGER et al., 1988).

A primeira proposta do uso de computadores como ferramenta de suporte à tomada de decisão clínica surgiu na década de 1950 com o artigo de Ledley e Lusted (1959). Embora os

autores não tenham utilizadas imagens médicas diretamente, eles lançaram as bases matemáticas para o surgimento de sistemas automatizados de diagnóstico, incluindo as imagens digitalizadas, para apoiar o raciocínio clínico (LEDLEY; LUSTED, 1959).

O uso de computadores para avaliar imagens de mamografias remonta o final da década de 1960 (DOI, 2007; TAKAHASHI R, KAJIKAWA Y, 2017) e foi inicialmente concebido para apoiar o radiologista (segunda opinião). Neste período, as aplicações se enquadravam em duas categorias: detecção assistida por computador (CAdE) e diagnóstico auxiliado por computador (CAdx) (GAO et al., 2019).

Embora as primeiras tentativas de análise computadorizada de imagens médicas tenham sido feitas na década de 1960 (LODWICK GS, 1963; MYERS et al., 1964), a investigação sistemática sobre o papel do computador no auxílio ao diagnóstico ganhou impulso na década de 1980 com uma mudança fundamental no conceito de utilização da saída do computador, evoluindo do diagnóstico auxiliado por computador para o diagnóstico automatizado por computador (DOI, 2007).

Na década de 1980, o nível de desempenho da saída do computador passou a ser muito alta em virtude de diagnóstico automatizado por computador ser baseado em algoritmos de aprendizado de máquina. Em um primeiro momento, as tentativas de desenvolver a interpretação automática de imagens médicas basearam-se em modelos decisórios humanos, realizando interpretações de imagens de alto nível (ou seja, simbólicas). Durante esse período, as metodologias de inteligência artificial aplicadas às imagens médicas estavam predominantemente baseadas no que foi chamado de sistemas especializados, que conduziam deduções lógicas com base em regras extrapoladas de metodologias humanas relativas ao processamento de imagens (MATSUYAMA et al., 1988). Em termos práticos, isso englobou operações que são consideradas simplórias pelos benchmarks contemporâneos, como a binarização ou limiar de uma imagem e a identificação de formas geométricas (linhas, círculos, trapézios). Um conjunto de princípios lógicos inferiria posteriormente o conteúdo da imagem por meio da avaliação da existência ou inexistência de tais estruturas (STANSFIELD et al., 1986).

A segunda geração de algoritmos também foi inspirada na cadeia decisória humana, mas, em vez de adotar uma interpretação simbólica das observações, migrou para uma interpretação estatística/probabilística. Nessa categoria, os parâmetros do modelo são calculados a partir de um conjunto de dados de referência rotulados, usando métodos probabilísticos que determinam as soluções mais prováveis (PARK et al., 2003, WARFIELD

et al., 2004). Dessa forma, um modelo estatístico é criado para representar a variabilidade inerente a uma população de referência.

A terceira categoria das tecnologias de inteligência artificial é a abordagem connexionista, baseada em redes neurais profundas (*deep learning*). Embora cada abordagem tenha seus pontos fortes e fracos, a abordagem connexionista vem ganhando bastante atenção recentemente para a resolução de problemas complexos (KIM et al., 2019). Uma rede neural artificial é um algoritmo inspirado nos neurônios e suas conexões, dispostas em camadas com conectadas por nós. Possuem em camadas de entrada e saída, e um conjunto de camadas ocultas. A primeira camada contém valores de entrada e a última camada possui valores rotulados (saída). Durante o treinamento, o valor de cada nó é determinado pela parametrização de pesos por meio de algoritmos de aprendizagem, como a retropropagação. Os pesos de cada nó são otimizados na direção para reduzir perdas e, assim, aumentar a precisão (HE et al., 2016).

Entre as redes neurais artificiais, destacam-se as redes neurais convolucionais. Estas conseguem lidar com um grande volume de dados, quando comparadas a arquiteturas anteriores. Como parte do *deep learning*, uma rede neural convolucional tem se destacado na visão computacional para tarefas de aprendizado supervisionado e não supervisionado (KRIZHEVSKY et al., 2012). Esta arquitetura alcançou recordes históricos de tarefas de visão tradicionais (WANG et al., 2017). A composição estrutural dessa arquitetura compreende camadas convolucionais, camadas de agrupamento e camadas totalmente conectadas. O papel principal da camada convolucional é detectar padrões, bordas e limites, entre outras características. A possibilidade de gerar múltiplas características das imagens permite o sucesso em diversas tarefas de visão computacional, como a segmentação e a classificação.

O sucesso na aplicação de ferramentas atuais de *deep learning* foi possível principalmente devido aos recentes avanços no desenvolvimento de tecnologias de hardware como, por exemplo, as unidades de processamento gráfico (JEON et al., 2021). De fato, em virtude da enorme quantidade de dados gerados, por vezes, bilhões de parâmetros na fase de treinamento para o reconhecimento de padrões e relações complexas, existe a necessidade de um maior poder computacional para analisá-los.

A comunidade envolvida na análise de imagens médicas progrediu lentamente, contando com a extração manual de dados. Antes da AlexNet, várias metodologias para aquisição de recursos eram amplamente aceitas. Bengio (2013) apresentou às CNNs um treinamento extensivo realizado somente após a avaliação, denominado “Treinamento global de modelos profundos”.

As metodologias de aprendizado profundo para análise de imagens médicas se iniciaram em conferências e workshops, fortalecendo-se posteriormente no discurso acadêmico. Houve um aumento significativo na atividade de publicação durante 2015 e 2016. Esse assunto surgiu como um tema predominante nas principais conferências, culminando em uma edição especial do IEEE *Transactions on Medical Imaging* (GREENSPAN; GINNEKEN; SUMMERS, 2016).

Neste contexto, a segmentação de imagens/órgãos – um dos pilares da imagem médica – é um exemplo clássico para demonstrar os conceitos acima. Uma estrutura de interesse em uma nova imagem pode ser descrita pela correspondência de sua aparência com o modelo estatístico previamente obtido. Se o modelo for treinado a partir de imagens de pacientes saudáveis, a presença de lesões em uma nova imagem pode ser interpretada com valores atípicos e anormais em relação à aparência esperada. Essa metodologia é frequentemente empregada em neuroimagem para delinear a arquitetura anatômica do cérebro e determinar lesões resultantes de ferimentos traumáticos, câncer, esclerose múltipla e várias doenças, inclusive AVC (VAN LEEMPUT et al., 2001; ERUS et al., 2014; VIERGEVER et al., 2016). A segmentação automatizada de lesões isquêmicas surge como uma alternativa com o potencial de aprimorar substancialmente o domínio da imagem médica (VIERGEVER et al., 2016).

2.3 Técnicas de Segmentação de Imagens Médicas

As abordagens mencionadas acima vêm se tornando importantes para o desenvolvimento da prática clínica moderna. A implementação destas técnicas pode permitir avanços na medicina diagnóstica e a personalização dos tratamentos através da análise de dados biomédicos, orientando diagnósticos e intervenções médicas no tratamento de doenças complexas.

Os métodos tradicionais de segmentação de imagens utilizam características de pixels (heurísticas), como cor e intensidade, enquanto as abordagens de aprendizado profundo utilizam redes neurais avançadas para reconhecimento de padrões complexos. As saídas produzem máscaras de segmentação que delinham os limites precisos dos pixels e a morfologia de cada classe na imagem. Os RNCs são preferidos para a segmentação médica devido à sua proficiência em aprendizado e extração automática de recursos (MOHAPATRA et al., 2023).

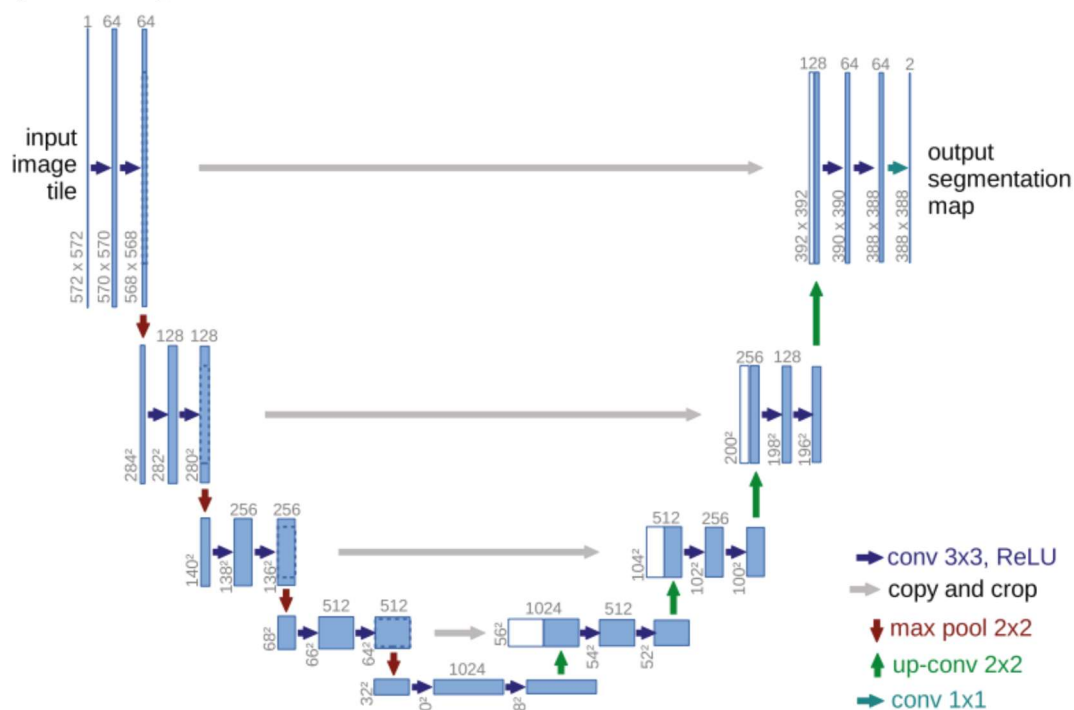
As técnicas de automatização desse processo têm avançado com modelos de deep learning, em especial após o surgimento da U-Net e de suas variantes. Assim, enquanto métodos

tradicionais enfrentam dificuldades para identificar de forma confiável as regiões isquêmicas, as arquiteturas baseadas em U-Net têm sido adotadas e aprimoradas para melhorar a sensibilidade e precisão da segmentação. Nesta revisão, apresentamos as principais arquiteturas derivadas da U-Net trabalhadas nesta investigação: U-Net original, V-Net, U-Net++, R2U-Net, Attention U-Net, ResUNet-a e Swin U-Net – apresentando suas arquiteturas, inovações, características específicas, vantagens e limitações, enfatizando a aplicação em imagens médicas e, em particular, na segmentação de lesões de AVC isquêmico em TC sem contraste. Cada modelo representa um passo evolutivo no esforço de aprimorar a segmentação de estruturas complexas e pouco contrastadas, como as lesões isquêmicas cerebrais.

2.3.1 U-Net: Arquitetura Original de Segmentação Biomédica

A U-Net é uma arquitetura dirigida à segmentação de imagens médicas. Sua forma em “U” combina um caminho de codificação (contração/análise) e outro de decodificação (expansão/síntese), conectados por ligações entre camadas de mesma resolução. O primeiro agrega informações de classificação; o segundo recupera a localização semântica, aumenta a resolução de saída e produz a segmentação. O resultado é um mapa pixel a pixel, com o mesmo tamanho da entrada, obtido por sequências de convoluções e deconvoluções (*transposed convolution*) que primeiro reduzem e depois ampliam a resolução (RONNEBERGER; FISCHER; BROX, 2015) (Figura 1).

Figura 1 – Arquitetura U-Net



Arquitetura U-Net (exemplo para 32x32 pixels na resolução mais baixa). Cada caixa azul corresponde a um mapa de características multicanal, com a contagem de canais especificada acima da caixa. A dimensão x-y é indicada na parte inferior esquerda de cada caixa. As caixas brancas significam mapas de características copiados. As setas indicam as várias operações. (RONNEBERGER; FISCHER; BROX, 2015)

A U-Net introduz melhorias através de conexões de atalho que preservam o contexto global e a resolução espacial. Isso permite uma segmentação precisa mesmo com desequilíbrio de classes e anotações esparsas em imagens médicas. Ele captura habilmente as relações entre imagens e máscaras, mantendo o contexto espacial de estruturas críticas. Além disso, ele demonstra eficiência e alto desempenho com dados limitados. As conexões de atalho aprimoram o decodificador com recursos detalhados do codificador, melhorando o refinamento de bordas e detalhes. Consequentemente, o U-Net segmenta a imagem inteira, mantendo o contexto completo, o que é fundamental para detectar lesões difusas ou sutis, como derrames isquêmicos precoces. Pesquisas indicam que ele tem um desempenho comparável ao dos radiologistas na identificação de alterações isquêmicas precoces na TC quando adequadamente treinado e ajustado (RONNEBERGER; FISCHER; BROX, 2015).

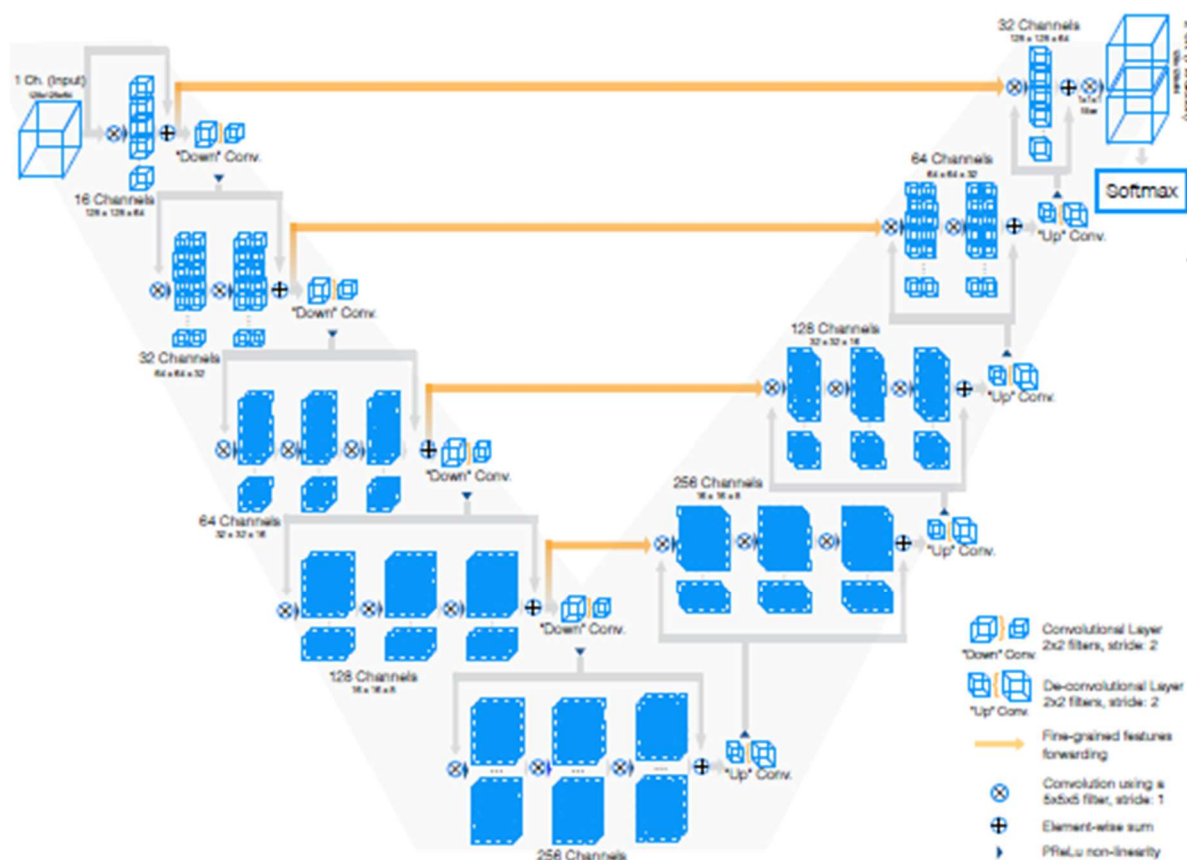
Sua popularidade no domínio médico decorre da adaptabilidade, mas enfrenta restrições. As conexões diretas entre codificador e decodificador correm o risco de mesclar informações semanticamente incongruentes, complicando o aprendizado e prejudicando a segmentação em lesões pequenas ou indistintas. A U-Net convencional utiliza convoluções

locais, que restringem as dependências globais; na tomografia computadorizada do crânio, comparações inter-hemisféricas e padrões distribuídos diferenciados podem permanecer sem solução. Além disso, ele foi projetado inicialmente para aplicações 2D; quando aplicado *slice* a *slice*, ele negligencia a continuidade entre as fatias, resultando em uma perda de informações volumétricas (3D). Por fim, enfrenta desequilíbrio extremo de classes quando a lesão ocupa fração mínima de pixels, a menos que se usem perdas adequadas (p.ex., Dice loss) ou amostragem balanceada. Essas limitações motivaram as variações e aprimoramentos discutidos a seguir (RONNEBERGER; FISCHER; BROX, 2015).

2.3.2 V-Net: Segmentação Volumétrica 3D

O U-Net é basicamente um modelo 2D, enquanto o V-Net adapta essa estrutura para análise volumétrica 3D. Ele utiliza convoluções volumétricas em vez de processamento de fatias, sendo totalmente convolucional. Mantendo a arquitetura de codificador-decodificador U-Net, ele incorpora kernels 3D (por exemplo, $5 \times 5 \times 5$) (MILLETARI et al., 2016). A V-Net apresenta dois avanços importantes: (1) a integração de blocos residuais para aprimorar o treinamento de aprendizado profundo; e (2) a implementação do *pool* de intercâmbio para otimizar a resolução e a expansão do canal. Consequentemente, ele substitui o *pool* tradicional por uma redução de resolução treinável, o que melhora o fluxo do gradiente e pode reduzir os requisitos de memória (MILLETARI et al., 2016) (Figura 2).

Figura 2 – Arquitetura V-Net



Representação esquemática da arquitetura V-Net. Processa dados 3D realizando convoluções volumétricas. Adaptado de Milletari et al., 2016.

O V-Net aprimora a segmentação volumétrica por meio de uma perda baseada no coeficiente de Dice (MILLETARI et al., 2016). Em imagens médicas, a classe de interesse geralmente compreende uma fração mínima de *voxel*; perdas tradicionais, como entropia cruzada, são influenciadas pelo fundo. A perda de dados promove a sobreposição entre previsões e referências, equilibrando efetivamente a importância de pequenas regiões. Isso é vital para lesões isquêmicas: com aproximadamente 1% dos *voxels* representando lesões, a rede pode classificar tudo como plano de fundo e, ao mesmo tempo, atingir alta precisão — a perda de Dice alivia esse viés. Consequentemente, o V-Net aborda o desequilíbrio de classes e delinea com precisão pequenas estruturas em conjuntos de dados 3D (MILLETARI et al., 2016).

A V-Net segmenta com eficiência volumes inteiros. Ela utiliza uma redução profunda da resolução de camadas para impor restrições globais de segmentação. Essa metodologia aprimora a identificação de lesões de baixo contraste em um contexto mais amplo. A V-Net emprega conexões de salto no estilo U-Net para reter recursos detalhados durante o aumento de

escala da amostragem. O estudo descobriu que o PreLU apresenta uma robustez superior ao ReLU. A principal vantagem é a integração de informações contextuais tridimensionais na segmentação, crucial para o delineamento preciso dos volumes de infarto no AVC isquêmico, o que é mais desafiador em avaliações bidimensionais (MILLETARI et al., 2016).

A Dice loss é vantajosa em aplicações médicas. A V-Net demonstrou que essa otimização aprimora a identificação de estruturas minúsculas sem a necessidade de técnicas de reamostragem. Além disso, o uso de blocos residuais facilita uma convergência mais rápida: ao adquirir funções residuais, a rede aprimora a profundidade do treinamento enquanto preserva a qualidade do gradiente, alcançando assim uma convergência mais rápida e superando arquiteturas comparáveis sem resíduos. Em síntese, a V-Net elevou a confiabilidade e a precisão da segmentação volumétrica e segue referência para segmentação 3D de estruturas anatômicas e lesões, com resultados superiores a metodologias anteriores (MILLETARI et al., 2016).

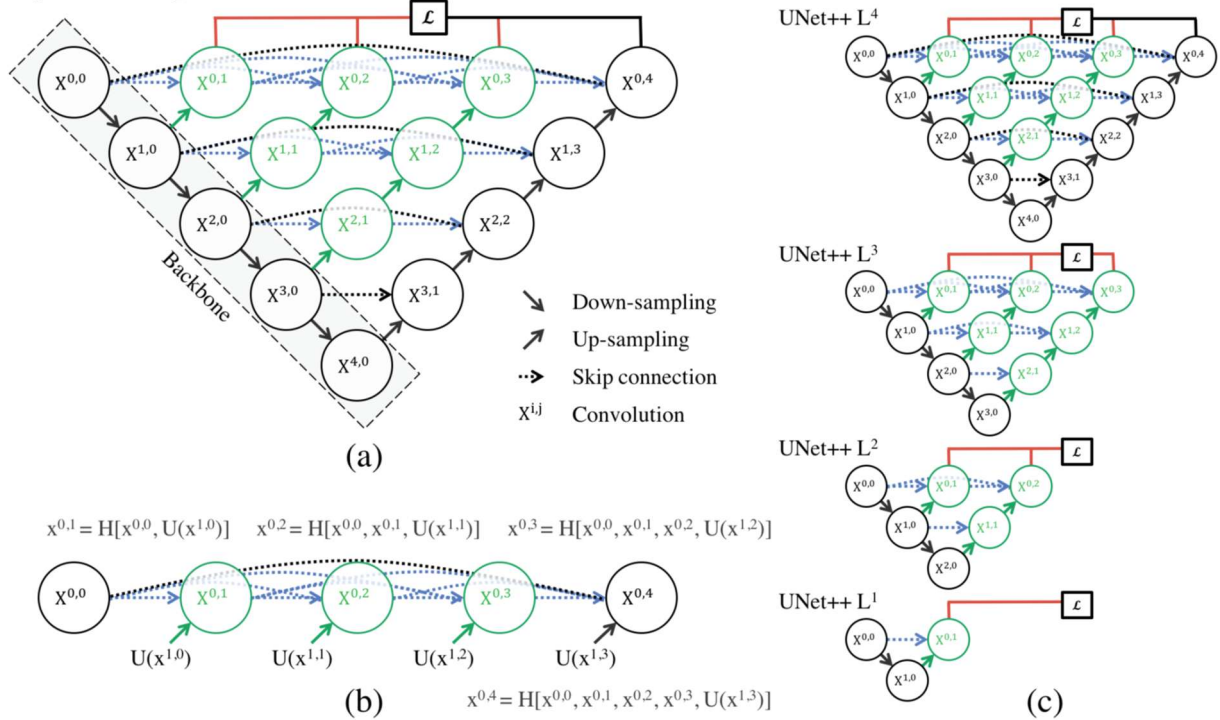
Por outro lado, a V-Net incorre em maiores custos computacionais. O processamento total de volumes em 3D exige mais memória e tempo, o que o torna impraticável para tomografia computadorizada de alta resolução ou vários volumes. A falta de agrupamento limita a eficiência de recursos, mas o modelo permanece consumindo muitos recursos em comparação com as U-Nets 2D e exige GPUs avançadas, limitando a aplicação clínica. A complexidade de anotar dados volumétricos 3D em comparação com imagens bidimensionais torna o estabelecimento de conjuntos de dados de treinamento abrangentes consideravelmente desafiador (MILLETARI et al., 2016).

No contexto do AVC isquêmico, a segmentação abrangente de volumes abrangendo todas as regiões infartadas é pouco frequente; em vez disso, há uma predominância de anotações ou marcadores parciais. Além disso, a V-Net herda certas limitações da U-Net em relação à granularidade: apesar da inclusão do contexto global, ela pode exigir modificações para lesões caracterizadas por heterogeneidade no tamanho (por exemplo, infartos extensos versus lacunares). Por fim, a arquitetura V-Net convencional não integra mecanismos de atenção ou recorrência; sua eficácia depende fundamentalmente de convoluções tridimensionais e pode ser inadequada em casos de contrastes significativamente baixos (MILLETARI et al., 2016).

2.3.3 U-Net++: Conexões Aninhadas para Reduzir Lacuna Semântica

A U-Net++ melhora a U-Net reduzindo as discrepâncias semânticas entre seu codificador e decodificador. A conexão direta do codificador U-Net original a um decodificador mais profundo resulta em inconsistências semânticas. A U-Net++ resolve isso empregando atalhos aninhados e densos que aprimoram iterativamente os recursos do codificador para integração do decodificador. Várias conexões curtas são utilizadas em vez de uma única conexão de salto para alinhar progressivamente as representações. Caminhos densos, que lembram o DenseNet, promovem a reutilização de recursos e o fluxo de gradientes. Além disso, ele incorpora supervisão profunda ao produzir saídas de segmentação em vários estágios do decodificador, facilitando o aprendizado de representações valiosas em camadas intermediárias; máscaras em diversas resoluções podem ser amalgamadas na saída final, aumentando a precisão em determinados contextos (ZHOU et al., 2018) (Figura 3).

Figura 3 – Arquitetura U-Net++



(a) UNet++ - um codificador e um decodificador que são conectados por uma série de blocos convolucionais densos aninhados. No resumo gráfico, o preto indica a U-Net original, o verde e o azul mostram os blocos de convolução densos nas vias de atalho, e o vermelho indica supervisão profunda. Os componentes vermelho, verde e azul distinguem o UNet++ do U-Net. (b) Análise detalhada da primeira via de atalho do UNet++. (c) O UNet++ pode ser reduzido em tempo de inferência, se treinado com supervisão profunda. Adaptado de Zhou et al., 2018

Três pontos definem a U-Net++: (1) convoluções extras nas conexões de atalho (“verdes”) para alinhar semanticamente as features do codificador às do decodificador antes da

fusão; (2) conexões densas (“azuis”) entre nós de atalho para propagar/refinar informações entre níveis, enriquecendo mapas de alto nível com detalhes de baixo nível; e (3) supervisão profunda (“vermelha”), com cabeças de segmentação em cada estágio do decodificador, que força representações úteis em múltiplas escalas. Ao decompor o problema em alvos mais simples, reduz a lacuna semântica e melhora o desempenho; os autores relatam ganho médio de ~ 4 p.p. em IoU sobre a U-Net em vários conjuntos médicos (ZHOU et al., 2018).

A estrutura aninhada torna a rede mais profunda e com mais parâmetros que a U-Net, embora a reutilização densa compense parcialmente. Em troca, produz mapas e saídas multi-escala (p.ex., $1/2$, $1/4$, $1/8$ da resolução), recalculados e refinados, úteis quando o tamanho das estruturas varia muito. Sem supervisão profunda, a U-Net++ já supera a U-Net; com supervisão, os ganhos adicionais aparecem sobretudo em cenários multi-escala (pólipos, lesões hepáticas) e tendem a ser menores quando as estruturas têm dimensões semelhantes — sinal de que a própria arquitetura aninhada resolve boa parte do problema (ZHOU et al., 2018).

A U-Net++ melhora a delimitação de contornos e detalhes finos ao mitigar a divergência semântica entre *encoder* e *decoder*. Conexões densas melhoram as características residuais, promovendo a propagação do gradiente e a utilização do filtro, reduzindo assim o sobreajuste. Em última análise, ele supera a U-Net e suas contrapartes mais simples (por exemplo, Wide U-Net), ao mesmo tempo em que garante despesas computacionais razoáveis. A supervisão múltipla melhora a estabilidade do treinamento e evita “camadas mortas”, resultando em uma convergência mais confiável (ZHOU et al., 2018).

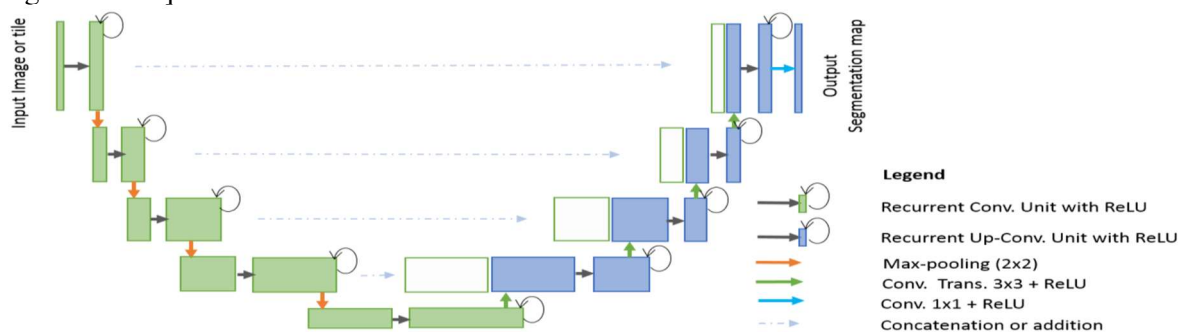
Em síntese, ele demonstra capacidades superiores de segmentação em diversas escalas, mas exige maior complexidade do que o U-Net: exige treinamentos prolongados e durações de inferência, hardware ou otimizações aprimoradas e implementação cuidadosa para alinhar conexões e perdas. Com dados limitados, o potencial de sobreajuste aumenta (parcialmente aliviado por conexões densas). Dada a necessidade de uma interação robusta entre codificador e decodificador, ela pode ser excessiva para tarefas simples. A adoção deve considerar o equilíbrio entre desempenho e despesas computacionais individualmente (ZHOU et al., 2018).

2.3.4 R2U-Net: U-Net com Convoluções Recorrentes e Residuais

A arquitetura R2U-Net, introduzida por Alom et al. (2018), integra conexões residuais internas junto com operações convolucionais recorrentes. Cada bloco de convolução da U-Net

é transformado em um bloco recorrente, refinando progressivamente as características. Esse bloco é envolvido por uma conexão residual que soma a entrada à saída, estabilizando o treino. Assim, combina a acumulação de representações da recorrência com a robustez residual. Existem duas variantes: RU-Net (apenas recorrente) e R2U-Net (recorrente + residual) (ALOM et al., 2018) (Figura 4).

Figura 4 – Arquitetura R2U-Net



Arquitetura R2U-Net com unidades de codificação e decodificação convolucionais usando CCR baseadas na arquitetura U-Net. As unidades residuais são usadas com CCR para a arquitetura R2U-Net. Adaptado de Alom et al., 2018. CCR - camadas convolucionais recorrentes

Em termos práticos, na U-Net uma camada convolucional extrai *features* uma vez; na R2U-Net, a mesma camada é reaplicada sequencialmente (t vezes), sempre recebendo sua própria saída anterior e a entrada original via atalho residual — como “desenrolar” uma micro-RNA convolucional. Passos iniciais capturam padrões grosseiros; os seguintes refinam e corrigem. Os pesos são compartilhados, portanto não há aumento de parâmetros. A conexão residual preserva a identidade quando as iterações pouco ajudam, estabilizando o treino. Assim, a R2U-Net obtém representações mais ricas sem aumentar a largura do modelo (ALOM et al., 2018).

Os benefícios associados ao R2U-Net são: (1) a incorporação de conexões residuais promove a convergência de redes neurais mais profundas, aliviando problemas relacionados à descida do gradiente; (2) o mecanismo de recorrência serve como uma forma implícita de atenção sobre sequências temporais, aprimorando áreas não resolvidas e agregando recursos sem precisar da ampliação dos filtros; (3) desempenho aprimorado com uma contagem de parâmetros equivalente, conforme evidenciado em tarefas de segmentação da retina, dérmica e pulmonar, superando a U-Net e a ResU-Net em termos de Dice e índices de Jaccard (ALOM et al., 2018).

Suas desvantagens compreendem: maior custo de treino e inferência, pois t iterações equivalem a t convoluções empilhadas (tipicamente $t = 2-3$); risco de instabilidade recorrente se mal calibrada (p.ex., gradientes explosivos com t alto); espaço de hiperparâmetros mais complexo (escolha de t , inicialização etc.); maior risco de overfitting em bases pequenas; e menor interpretabilidade do refinamento iterativo em relação a arquiteturas puramente feedforward (ALOM et al., 2018).

2.3.5 Attention U-Net: Aprendendo a Focar em Regiões Importantes

A arquitetura Attention U-Net incorpora habilmente mecanismos sofisticados de atenção à estrutura pré-existente da U-Net, aumentando significativamente a capacidade da rede de discernir e identificar com precisão pontos focais críticos dentro da imagem, o que é essencial para tarefas que exigem detalhes meticulosos e precisão na segmentação. Em vez de simplesmente concatenar as características do codificador com as do decodificador, cada conexão residual atravessa uma “porta de atenção” que, influenciada pelas informações contextuais do decodificador, determina as características salientes enquanto diminui as menos relevantes. Consequentemente, são produzidas máscaras de atenção que, pixel a pixel, avaliam a importância dos mapas de recursos do codificador antes de sua integração. O módulo recebe (1) features do codificador na escala corrente (baixo nível) e (2) features do decodificador (alto nível); combina-as via transformações lineares e ativação sigmoide, produzindo coeficientes entre 0 e 1 para aquela escala (OKTAY et al., 2018).

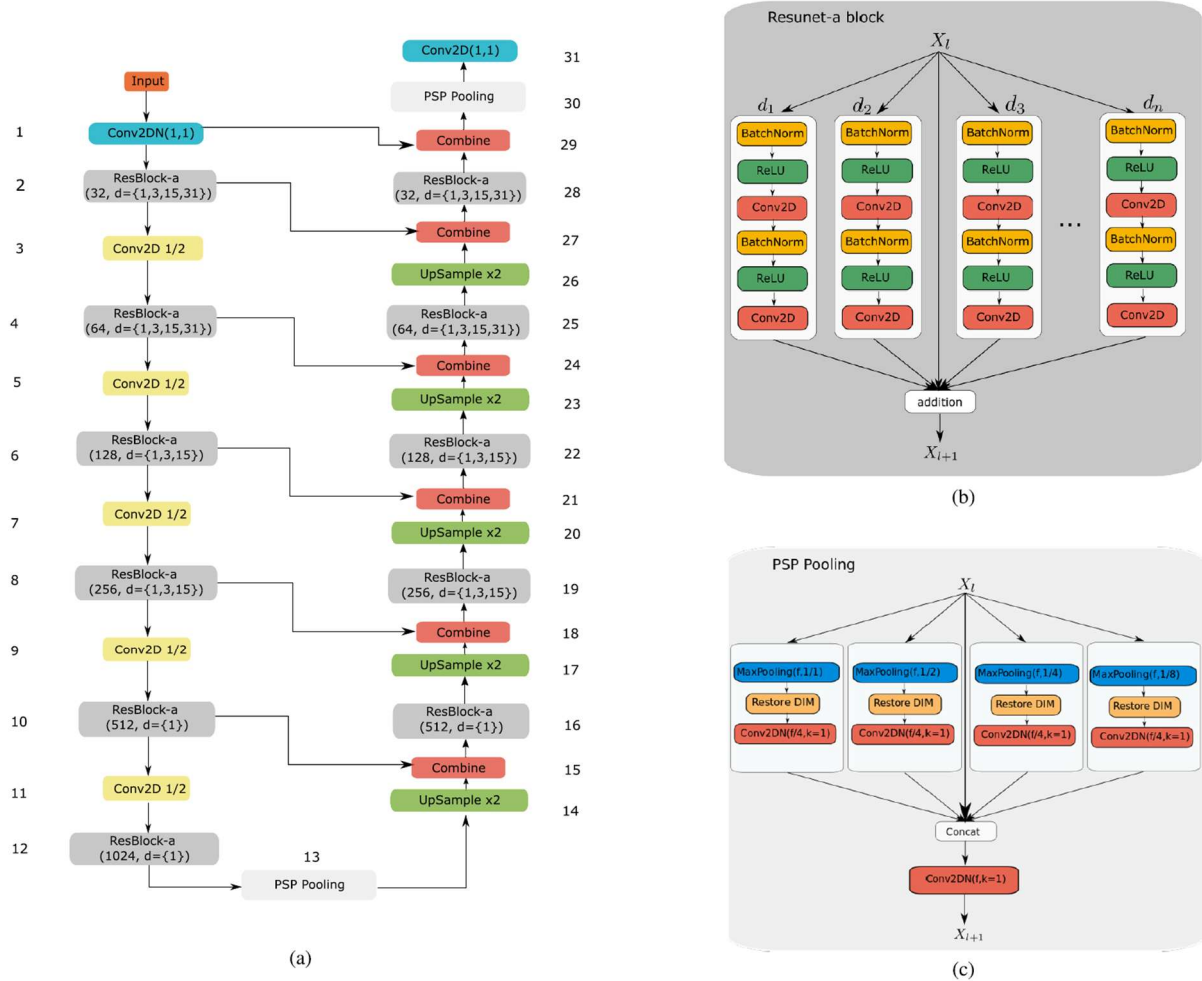
Em seguida, essa máscara multiplica as features do *encoder*, suprimindo regiões irrelevantes e preservando/amplificando as úteis; só então ocorre a concatenação no *decoder*. Repetido em cada conexão, o *decoder* recebe sobretudo informação pertinente, dispensando etapas separadas de localização (p.ex., detecção de órgão seguida de segmentação). A própria Attention U-Net integra localização e segmentação em uma única rede (OKTAY et al., 2018) (Figura 5).

A Resunet-A, que foi projetada para a tarefa de segmentação de imagens no domínio do sensoriamento remoto, incorpora vários avanços essenciais que também apresentam considerável aplicabilidade em contextos de imagens médicas, estabelecendo-se assim como uma iteração significativamente aprimorada da arquitetura U-Net amplamente reconhecida (DIAKOGIANNIS et al., 2020). As principais características do ResUnet-a são:

- A incorporação de blocos residuais nos componentes do codificador e do decodificador (“Res” UNet) aumenta a capacidade de treinar redes de maior profundidade.
- A utilização de convoluções atrosas (dilatadas), que aumentam o campo receptivo sem aumentar os parâmetros de peso (incorporando zeros no kernel) e capturando informações com eficácia em várias escalas.
- A incorporação de um módulo *Pyramid Scene Parsing* (PSP): uma pirâmide de agrupamento de contexto, remanescente da PSPnet, com o objetivo de extrair informações contextuais globais e regionais.
- A execução de inferência multitarefa sequencial: junto com a melhor saída de segmentação, a arquitetura produz mapas de borda, transformações de distância e até reconstruções de cores na mesma rede.

As tarefas estão interligadas (bordas → distâncias → segmentação), em que cada tarefa informa e condiciona os processos subsequentes em todo o pipeline (DIAKOGIANNIS et al., 2020) (Figura 6).

Figura 6 – Arquitetura ResUNet-a



Visão geral da rede ResUNet-a d6. (a) O ramo esquerdo (para baixo) é a parte codificadora da arquitetura. O ramo direito (para cima) é o decodificador. (b) Bloco de construção (residual). (c) Camada de agrupamento de análise sintática de cenas piramidais. Adaptado de Diakogiannis et al., 2020.

O gasto associado à abundância de módulos é a complexidade e o custo computacional. Equilibrar as perdas de diversas saídas (bordas, distância, máscara, reconstrução) representa um desafio para evitar o domínio da tarefa. A variedade de componentes complica a depuração, pois o gargalo pode ocorrer na dilatação, perda ou tarefas auxiliares. Em ambientes com recursos limitados, o modelo apresenta considerável latência e dependência dos recursos da GPU. Além disso, os princípios extrapolados a partir de imagens aéreas, como a reconstrução colorida, possuem aplicabilidade limitada à tomografia computadorizada cerebral, sugerindo que certas simplificações podem ser vantajosas. Em conclusão, embora o modelo possua uma força considerável, sua eficácia pode ser dependente do contexto (DIAKOIANNIS et al., 2020).

Os componentes buscam robustez e precisão em situações de alta variabilidade e baixo contraste, predominantes em satélites e na medicina. Uma perda generalizada de Dice é

proposta para aumentar a convergência e abordar o desequilíbrio de classes de forma eficaz (DIAKOGIANNIS et al., 2020).

A eficácia decorre da sinergia: a incorporação de blocos residuais facilita o aumento da profundidade, melhora a propagação de gradientes e permite a reutilização de recursos; as circunvoluções dilatadas servem para aumentar o campo receptivo sem recorrer à redução excessiva da resolução; o módulo *Pyramid Scene Parsing* (PSP) infunde contexto global (por meio de agrupamento em várias escalas e retroprojeção), que é particularmente benéfico quando o contexto espacial auxilia na diferenciação de estruturas. A estrutura da multitarefa condicionada (de bordas a distâncias e segmentação) funciona como uma forma de regularização, produzindo delineamentos suaves e coerentes. No contexto de conjuntos de dados urbanos desafiadores, essa metodologia atingiu uma *F1-Score* de aproximadamente 92,9%, demonstrando assim sua capacidade de lidar com a variabilidade e o desequilíbrio (DIAKOGIANNIS et al., 2020).

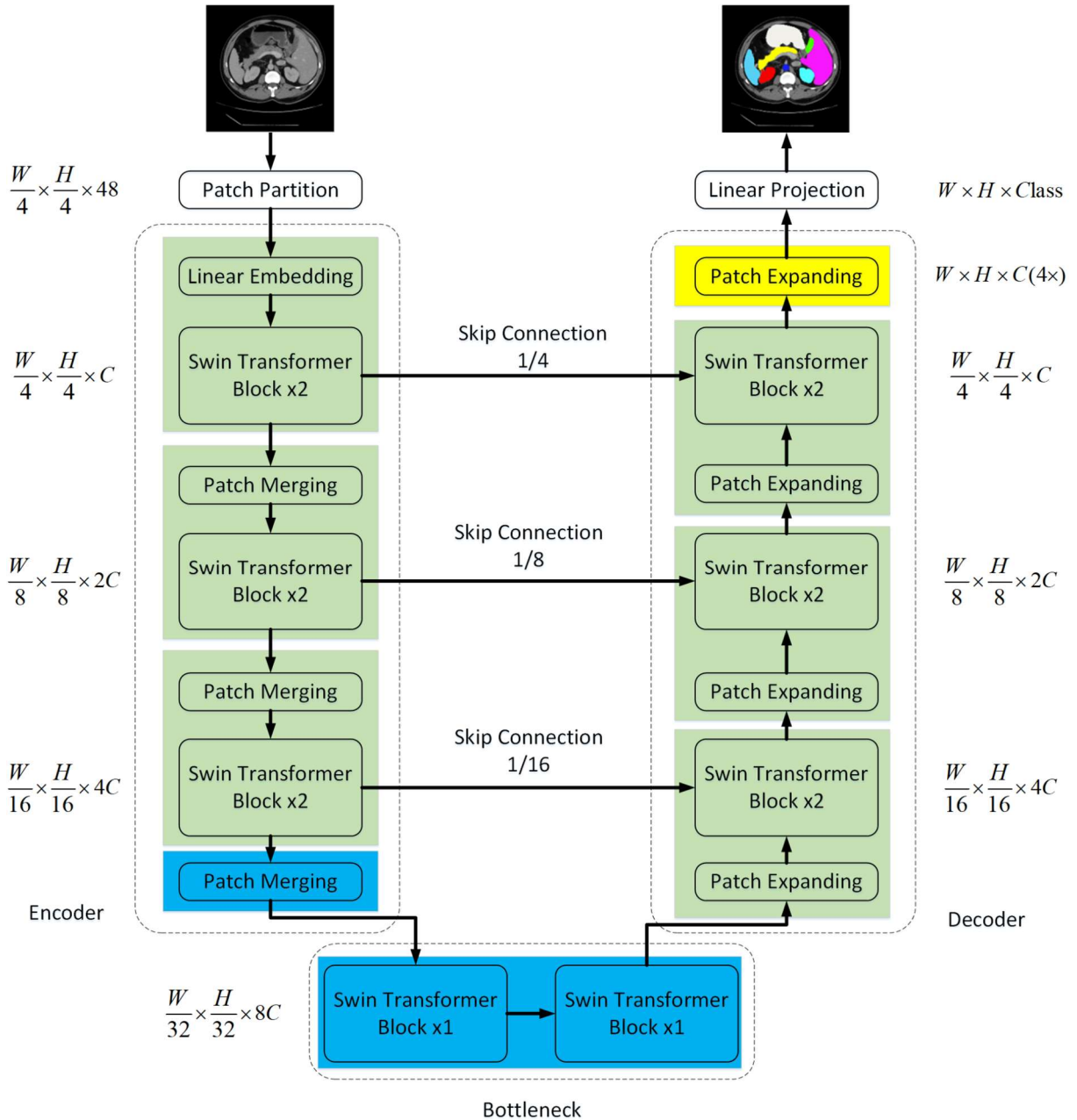
2.3.7 Swin U-Net: *transformers* puros em estrutura na forma de U

O Swin U-Net combina a arquitetura “U” distinta com transformadores de visão, empregando apenas camadas de *Swin Transformers* (CAO et al., 2021). O *Swin Transformer* demonstra uma arquitetura hierárquica e eficácia operacional, executando mecanismos de autoatenção em janelas segmentadas para capturar com proficiência dependências de longo alcance (LIU et al., 2021).

- Codificador: esse componente divide sistematicamente a imagem em patches discretos (*tokens*) e diminui progressivamente a resolução por meio de um processo de mesclagem de patches (quatro patches convergindo em um token). Em cada estágio iterativo, a autoatenção é aplicada às janelas padrão e offset, abrangendo assim todo o domínio espacial e facilitando a extração contextual.
- Decodificador: esse segmento foi projetado para replicar a funcionalidade do codificador, executando a ampliação da amostragem nas camadas do transformador e reconstruindo a máscara; conexões de atalho facilitam a transferência de tokens de alta resolução do codificador para aumentar a fidelidade dos detalhes.
- Gargalo (*bottleneck*): posicionado no nível de resolução mais baixo, um bloco transformador é empregado para capturar uma perspectiva global abrangente.

O grande diferencial é que a arquitetura evita totalmente as convoluções; todos os processos de aprendizagem são conduzidos por meio da autoatenção, que é hábil em modelar relacionamentos remotos dentro da imagem (CAO et al., 2021) (Figura 7).

Figura 7 – Arquitetura Swin U-Net



A arquitetura do Swin-Unet é composta por codificador, gargalo, decodificador e conexões de atalho, todos construídos com base no bloco transformador Swin. Adaptado de Cao et al., 2021

O estímulo para a criação do Swin U-Net surge das limitações inerentes às RNCs com relação à sua capacidade de capturar adequadamente interdependências globais explícitas, atribuíveis às janelas locais restritas dos núcleos convolucionais. No domínio das imagens

médicas complexas, é imperativo estabelecer associações entre *loci* distantes (por exemplo, o exame comparativo de hemisférios cerebrais ou a inter-relação de neoplasias com entidades anatómicas remotas). Os *transformers*, por meio do mecanismo de autoatenção, modelam com eficácia as interações entre diferentes posições (dentro da janela localizada e, posteriormente, em todo o conjunto de dados). O Swin U-Net faz isso de forma hierárquica, o que é congruente com a estrutura multiescala inerente à U-Net (CAO et al., 2021).

- Modelagem de longo alcance: por meio da implementação de janelas de deslocamento, a atenção engloba estrategicamente a totalidade da imagem após várias camadas, estabelecendo correlações entre sinais distribuídos sutis.
- Representação livre de convolução: ao evitar operações convolucionais, ela mitiga os vieses associados às estruturas da rede local e facilita a aquisição de características mais adaptáveis, que se baseiam inteiramente nos mecanismos de atenção.

Na investigação, a Swin U-Net demonstrou um desempenho superior em comparação com arquiteturas exclusivamente convolucionais (como U-Net e 3D U-Net), bem como modelos híbridos (como o TransoNet) nas tarefas de segmentação cardíaca e de múltiplos órgãos. A arquitetura que é fundamentalmente dependente de transformadores demonstrou habilidades aprimoradas de generalização e precisão superior, o que pode ser parcialmente atribuído à preservação de detalhes intrincados (uma diminuição de quatro vezes) e à diminuição da dependência de padrões texturais locais, enfatizando assim as correlações de intensidade e as relações espaciais (CAO et al., 2021).

Os obstáculos e restrições são complexos e variados. Os *transformers* tendem a exigir muitos dados (mais parâmetros e menos viés indutivo); embora o Swin tenha viés local por janela, treinar do zero com poucas anotações é difícil. A autoatenção continua custosa em memória para 3D/alta resolução. A interpretabilidade é difusa (mapas de atenção menos diretos que máscaras explícitas). Posicionamento relativo é crítico: o alinhamento preciso em contextos médicos simétricos pode desafiar a atenção “pura”. Por fim, o ecossistema é menos maduro que o das convoluções/U-Nets, com menos *frameworks* otimizados (CAO et al., 2021; LIU et al., 2021).

2.4 Estudos Recentes sobre Segmentação Automática de AVC

Para segmentar lesões de AVC isquêmico em TC sem contraste, a U-Net costuma servir como baseline e ponto de partida em muitas pesquisas. Alguns trabalhos empregaram a U-Net pura para identificar regiões hipodensas correspondentes ao núcleo isquêmico ou à penumbra em TC, com resultados promissores, porém modestos. As conexões de atalho são úteis para destacar sutis diferenças de tonalidade, e o modelo consegue explorar contexto anatômico global – por exemplo, sabendo que lesões isquêmicas geralmente respeitam limites vasculares (território da artéria cerebral média, etc.) e muitas vezes apresentam assimetria em relação ao outro hemisfério. Entretanto, dada a sutileza de muitas lesões isquêmicas (particularmente nas primeiras horas do AVC), a U-Net pode deixar de detectá-las se essas diferenças forem próximas ao nível de ruído. Para melhorar sua sensibilidade, diversos estudos propuseram pré-processamento (normalização de janelas de intensidade, realce de contraste, remoção do crânio) e ensemble de múltiplas U-Nets treinadas em diferentes condições. Ainda assim, os desafios inerentes à TC sem contraste – baixo contraste e artefatos – incentivaram o desenvolvimento das variantes a seguir, que trazem mecanismos adicionais (atenção, recorrência, conectividade densa, etc.) (RONNEBERGER; FISCHER; BROX, 2015).

A segmentação volumétrica com V-Net apresenta um potencial atraente para detecção de AVCs em 3D. Em contextos clínicos, os radiologistas realizam uma avaliação cognitiva da coleta de cortes de tomografia computadorizada de forma volumétrica — por exemplo, identificando pequenas hipodensidades que abrangem várias seções contíguas. Uma arquitetura V-Net possui o potencial de discernir com mais eficiência essas configurações volumétricas em comparação com uma U-Net 2D. Alguns estudos utilizaram V-Nets ou variantes 3D para segmentar núcleo isquêmico em TC perfusional (onde há mapas 3D de parâmetros como CBV e CBF) e obtiveram boa correspondência com áreas de infarto. No caso de TC sem contraste, menos trabalhos aplicaram V-Net diretamente, mas é plausível supor que integrar informação 3D melhore a robustez – por exemplo, diferenciando hipodensidades verdadeiras de artefatos de uma única fatia, ou usando a simetria 3D do cérebro (V-Net pode comparar o hemisfério esquerdo com o direito ao longo do volume, identificando assimetria sugestiva de isquemia). Um benefício adicional é que a V-Net, ao ver todo o cérebro, pode aprender contextos globais como “a lesão está no território de artéria cerebral média e corresponde a hipodensidade envolvendo os gânglios da base e córtex insular”, padrões difíceis de aprender em 2D puro. Em contrapartida, a aplicação de V-Net requer anotações volumétricas confiáveis como base de treinamento; e se o objetivo é detectar lesões muito precoces (quase invisíveis), mesmo o contexto 3D pode não resolver a falta de contraste intrínseca. Consequentemente, o V-Net é frequentemente integrado a aumentos de pré-processamento (como filtros de extração e

aprimoramento craniano) e, progressivamente, a estruturas suplementares, como mecanismos de atenção, que são empregados para se concentrar em áreas cerebrais que provavelmente são afetadas por lesões (MILLETARI et al., 2016).

As lesões do AVC apresentam uma ampla gama de dimensões, que incluem tanto isquemias arteriais significativas quanto apresentações lacunares sutis. As funcionalidades multiescala do U-Net++ no gerenciamento de configurações dimensionais heterogêneas são particularmente benéficas nesse contexto. Por exemplo, um infarto de ACM (que afeta grande área do hemisfério) e um pequeno infarto requerem do modelo a captura tanto de *big picture* quanto detalhes minúsculos, e a U-Net++ está bem equipada para isso. Além disso, a redução da lacuna semântica ajuda a delinear precisamente os contornos da lesão isquêmica, que podem ser difusos. Supondo que a diferença de densidade entre tecido normal e isquêmico seja muito sutil, cada nuance de nível de cinza conta; a U-Net++ pode garantir que o *decoder* receba features do *encoder* já refinadas para realçar essas nuances (ZHOU et al., 2018). Investigações recentes sobre o delineamento de lesões isquêmicas utilizaram metodologias semelhantes à U-Net++, caracterizadas pela implementação de estruturas de vários níveis ou multirresolução, facilitando assim a identificação aprimorada de lesões diminutas e, ao mesmo tempo, preservando a fidelidade da segmentação para lesões maiores. É notável que uma variante designada como U-Net em várias escalas foi avaliada quanto à sua eficácia na segmentação de regiões hipodensas em imagens de TC de AVC, resultando em maior sensibilidade para a identificação de lesões precoces, ressaltando assim a importância de projetos estruturais aninhados ou em várias escalas. Por outro lado, a necessidade de um conjunto de dados abrangente para o treinamento do U-Net++ é fundamental; em cenários em que a disponibilidade de imagens é limitada, a adoção de técnicas de transferência de aprendizado ou a aplicação sinérgica com a U-Net original pode ser aconselhável para manter a robustez. No entanto, o U-Net++ indiscutivelmente fornece avanços substanciais que podem aumentar a precisão das segmentações do AVC, permitindo assim a detecção de ocorrências isquêmicas que podem ser negligenciadas por abordagens mais simplistas (ZHOU et al., 2018).

A R2U-Net é bastante adequada para tarefas em que detalhes sutis fazem grande diferença, o que é o caso das lesões isquêmicas iniciais. Por exemplo, ao segmentar um pequeno infarto, as primeiras passes recorrentes de uma camada podem identificar uma região aproximadamente mais escura, e passes subsequentes podem aprimorar o contorno ou filtrar falsos positivos, resultando numa delinação mais acurada. Além disso, as recorrências funcionam como “iteradores” de foco, algo parecido com atenção: no contexto de um cérebro inteiro, a rede pode iterativamente concentrar-se em regiões candidatas a lesão, melhorando a

detecção de áreas de baixo contraste. Isso é valioso para não deixar escapar lesões pequenas ou de contraste limítrofe. Em um cérebro onde a diferença entre normal e isquêmico é quase imperceptível, uma única passagem de convolução pode não captar, mas várias iterações podem acumular evidências suficientes. Também, a invariância de parâmetros significa que mesmo sem muitos exemplos de treino, a R2U-Net pode utilizar recorrência para expandir sua profundidade efetiva sem precisar de mais dados para ajustar novos pesos – em problemas de dados escassos, isso é uma grande vantagem. Entretanto, a cautela reside em não superestimar essa vantagem: se o contraste é baixíssimo, a rede pode acabar reforçando ruído também. Alguns estudos de segmentação cerebrais relatam que adicionar recorrência melhora delineamento de estruturas como vasos ou anomalias difusas, sugerindo que R2U-Net ou ideias afins seriam úteis em AVC. Por exemplo, na segmentação de penumbra isquêmica (área de risco ao redor do núcleo do infarto), onde não há um limiar claro de intensidade, uma abordagem recorrente poderia iterativamente expandir a área segmentada até capturar toda zona hipoperfundida. Em resumo, a R2U-Net traz um equilíbrio interessante: mais capacidade de representação sem inchar o modelo – qualidade bem-vinda para detectar AVC isquêmico onde é preciso muita sensibilidade e precisão, mas os dados rotulados são limitados (ALOM et al. 2018).

Na TCSC de crânio, a maior parte da imagem mostra estruturas normais, enquanto a lesão isquêmica é uma região menor e sutil. Um modelo de atenção aumenta a sensibilidade ao permitir o foco em regiões cerebrais alteradas, desconsiderando, assim, as áreas normais. Por exemplo, se a maior diferença causada por um AVC inicial é um leve obscurecimento da junção córtico-subcortical em certo giro, a atenção pode ajudar a realçar esse giro em relação aos demais, baseando-se possivelmente em assimetrias ou padrões detectados pelo *decoder*. Isso é consistente com a prática clínica, pois os neurologistas se concentram nas regiões típicas do infarto, ignorando as áreas normais. A rede de atenção também pode suprimir informações normais e destacar regiões suspeitas. Estudos recentes de detecção de AVC por TC implementaram módulos de cuidado. Um estudo de 2021 introduziu uma U-Net modificada com diversos mecanismos de atenção para segmentação de lesões isquêmicas agudas, demonstrando desempenho superior com variantes de atenção (PNEV et al., 2021). Outra investigação empregou uma estratégia U-Net de duas etapas para a detecção eficaz de infartos indetectáveis na TC. Isso indica que a atenção aumenta a capacidade da rede de priorizar áreas significativas semelhantes às de um especialista, essenciais para reconhecer indicadores isquêmicos sutis (GÓMEZ et al., 2023). No entanto, como a atenção depende de pistas contextuais, os modelos podem exigir informações complementares para uma orientação

precisa em casos de traços mínimos. No entanto, o Attention U-Net apresenta uma estrutura robusta de “foco seletivo” que aprimora as ferramentas de segmentação do AVC isquêmico.

Quando utilizado para a segmentação de lesões isquêmicas, o ResUNet-a possui a capacidade de aumentar a precisão ao delinear de forma mais eficaz as margens da lesão e incorporar informações contextuais (DIAKOIANNIS et al., 2020). Por exemplo, a aplicação de métodos de detecção de borda pode aumentar a capacidade da rede de distinguir entre tecido isquêmico e tecido saudável adjacente, que geralmente aparece sutilmente na TC. A saída da transformada de distância encoraja a rede a entender a morfologia da lesão, possivelmente ajudando-a a não marcar regiões isoladas erráticas, mas sim volumes coesos (pois a distância implica uma única bolha por lesão). Isso poderia reduzir a dispersão de falsos positivos, pois um erro frequente na segmentação do traço é a identificação de pontos fantasmas isolados que não possuem uma estrutura coesa (BORGEFORS, 1986). O PSP *pooling* dentro da arquitetura daria noção global do encéfalo: por exemplo, se a lesão está apenas de um lado, se há simetria alterada, etc. O modelo capta isso com *pooling* global. E, claro, os blocos residuais e dilatações assegurariam que mesmo não ausência de redução excessiva a rede possa captar relações de longo alcance (p. ex., perceber que um edema no sulco central e outro no insular juntos sugerem infarto de ACM) (HE et al., 2014; ZHAO et al., 2017). Até onde sabemos, não há ainda aplicação direta do ResUNet-a em AVC, mas elementos semelhantes já foram usados: *dilated convolutions* são parte do DeepLab (e.g., DeepMedic e outros em segmentação de lesão cerebral já usavam) para capturar contexto sem perder resolução; *pyramid pooling* foi explorado em segmentação de múltiplas estruturas cerebrais; *segmentation with auxiliary tasks* é uma ideia emergente (por exemplo, algumas redes de segmentação de tumor de cérebro incluem predição de edema separado de núcleo do tumor para ajudar) (CHEN et al., 2017a, 2017b). Portanto, o ResUNet-a alinharia com essas tendências. Na prática, talvez se aproveitássemos a essência do ResUNet-a para AVC, poderíamos simplificar algumas coisas (talvez a reconstrução colorida não seja útil, já que TC é em tons de cinza). Mas certamente os conceitos de aproveitar múltiplos critérios (intensidade, borda, forma) para guiar a segmentação podem aumentar a confiabilidade dos resultados em um problema tão difícil quanto destacar uma lesão isquêmica incipiente no meio de um cérebro de aparência quase normal (DIAKOIANNIS et al., 2020).

Para detecção de AVC isquêmico em TC, um modelo como da Swin U-Net tem enorme potencial porque pode comparar regiões distantes e detectar padrões sutis de forma global. Por exemplo, um indício clássico de isquemia precoce é a assimetria sutil entre hemisférios – o lado afetado pode estar um pouco menos denso ou com sulcos mais apagados. Uma RNC pura pode ou não notar isso, mas um *transformer* pode explicitamente atentar o hemisfério esquerdo

contra o direito e destacar diferenças. Além disso, lesões isquêmicas podem apresentar padrões difusos (como leve hipodensidade distribuída em um território vascular); a autoatenção permite que diferentes partes desse território “conversem” na rede, reforçando a evidência coletiva de lesão. Pesquisas investigaram arquiteturas híbridas de *transformers*/RNC no contexto de AVC, revelando que esses modelos excederam o desempenho das U-Net convencionais, particularmente na identificação de eventos isquêmicos menores (DOSOVITSKIY et al., 2021). O Swin U-Net, sendo totalmente *transformer*, poderia levar isso além, embora devamos lembrar que quantidade de dados e qualidade das anotações são determinantes. Um aspecto importante é que, caso o *dataset* de treinamento não seja muito grande, poderíamos usar transfer learning de modelos *transformers* treinados em outras imagens (talvez imagens cerebrais de ressonância, ou mesmo imagens naturais adaptadas via *fine-tuning*). Isso pode ajudar a contornar a limitação de dados. Em termos práticos, um Swin U-Net bem treinado para segmentação de AVC possivelmente marcaria lesões com sensibilidade elevada, mas também com precisão quanto aos contornos, pois a visão global evita que ele confunda ruído isolado com lesão (ele “sabe” se um pixel escuro isolado faz parte de um contexto maior de lesão ou não). A robustez a variações também ajuda – diferentes protocolos de TC, idades de pacientes, etc., poderiam ser gerenciados pela adaptabilidade do *transformer*. Em contrapartida, o custo computacional seria maior: talvez para uso clínico em tempo real ainda seja mais prático uma U-Net otimizada (CAO et al., 2021; LIU et al., 2021). No entanto, à medida que os avanços tecnológicos no hardware progridem, é concebível que as metodologias de *transformers* se tornem rotineiramente aplicáveis. Em resumo, o Swin U-Net estabelece um novo paradigma, aproveitando as capacidades inerentes dos *transformers* para fins de segmentação médica e, especificamente dentro de nosso foco de pesquisa, potencialmente aprimorando a identificação de AVC isquêmico, em que cada indicador distinto distribuído pela região cerebral pode ser de suma importância.

2.5 Considerações Finais

As arquiteturas revisadas – U-Net e suas variações – apresentam um exame evolutivo das estruturas de segmentação no domínio da visão computacional médica. Cada iteração foi desenvolvida para abordar deficiências específicas inerentes à U-Net original e, ao mesmo tempo, aprimorar a segmentação de estruturas complexas. Em resumo:

- A U-Net estabeleceu a estrutura básica com conexões de atalho eficientes que facilitam a integração de informações contextuais e recursos detalhados, permitindo a segmentação com o mínimo de dados e alcançando ampla adoção em vários desafios médicos (RONNEBERGER; FISCHER; BROX, 2015).
- A V-Net modificou o U-Net para aplicações tridimensionais e incorporou a perda de dados para lidar com o desequilíbrio de classes, tornando-o particularmente adequado para dados volumétricos, como tomografias computadorizadas, embora às custas de maior complexidade (MILLETARI et al., 2016).
- A U-Net++ se concentrou em mitigar as discrepâncias semânticas nas ligações codificador-decodificador, empregando caminhos densamente aninhados para capturar várias escalas e obter uma segmentação mais precisa de detalhes intrincados (ZHOU et al., 2018).
- A R2U-Net integrou elementos recorrentes e residuais, facilitando assim a extensão da estrutura U-Net sem um aumento paralelo no número de parâmetros, o que aumenta a capacidade de segmentação de estruturas diminutas e de baixo contraste (ALOM et al., 2018).
- Attention U-Net empregou mecanismos de atenção espacial para concentrar a atenção do modelo nas regiões pertinentes, um aspecto crucial em imagens repletas de informações estranhas, aumentando assim a sensibilidade sem sacrificar a eficiência (OKTAY et al., 2018).
- A ResUnet-a uniu uma série de avanços (abrangendo resíduos, dilatação, pirâmide de agrupamento e tarefas auxiliares) para produzir segmentações altamente precisas e consistentes, ao mesmo tempo em que adquiriu a capacidade de aprender contornos e formas; no entanto, isso acarreta gastos computacionais consideráveis (DIAKOIANNIS et al., 2020).
- A Swin U-Net representa a incorporação de transformadores na arquitetura U-Net, modelando habilmente as dependências globais e locais de uma maneira versátil e alcançando resultados aprimorados em uma variedade de tarefas de segmentação, sugerindo assim uma trajetória promissora para os próximos esforços de pesquisa (CAO et al., 2021).

Contribuições significativas surgem dos avanços na segmentação da lesão isquêmica do AVC na TC sem contraste. A arquitetura fundamental do U-Net estabelece a estrutura essencial e é capaz de detectar inúmeras lesões quando treinada adequadamente. Extensões como U-

Net++ e R2U-Net são mais eficientes em capturar detalhes sutis e aprimorar a detecção de pequenos focos isquêmicos (RONNEBERGER; FISCHER; BROX, 2015; ZHOU et al., 2018; ALOM et al., 2018). Os mecanismos de atenção, demonstrados pela Attention U-Net, aumentam a diferenciação do sinal em exames cranianos, auxiliando na identificação de alterações relacionadas ao AVC (OKTAY et al., 2018). Modelos tridimensionais como o V-Net oferecem um contexto volumétrico crítico para avaliar a continuidade da lesão e as discrepâncias inter-hemisféricas (MILLETARI et al., 2016). ResUNet-a nos lembra da utilidade de impor conhecimento de forma (bordas) e contexto multi-escala, o que em isquemias poderia prevenir segmentações anatomicamente incoerentes (DIAKOGIANNIS et al., 2020). E, por fim, os *transformers* (Swin U-Net) oferecem uma capacidade sem precedentes de correlacionar pistas distantes e generalizar para variações, podendo no futuro próximo se tornar o novo *state-of-the-art* também nesse domínio (CAO et al., 2021).

Em conclusão, a segmentação das lesões isquêmicas do AVC é aprimorada por várias inovações, envolvendo uma combinação de técnicas para enfrentar desafios como baixo contraste e variabilidade (PNEV et al., 2021). O avanço contínuo dessas arquiteturas, potencialmente integrando várias inovações, visa alinhar o desempenho do algoritmo com o de especialistas humanos na detecção de AVC isquêmico por meio de TCSC (CAO et al., 2021). Pesquisas atuais sugerem otimismo de que o aprendizado profundo corresponderá aos métodos clínicos tradicionais na identificação de lesões isquêmicas, em grande parte atribuíveis às melhorias arquitetônicas que otimizam a detecção de sinais sutis, porém críticos (JOHANSEN et al., 2024; TAKÁCS et al., 2025).

3 METODOLOGIA

Esta investigação emprega um desenho experimental e quantitativo, utilizando um banco de dados acessível ao público em conjunto com metodologias de aprendizado profundo para treinar, validar e avaliar um modelo de segmentação para AVC isquêmico em exames de TCNC. A estrutura metodológica incorpora a aquisição e o pré-processamento do conjunto de dados, a formulação dos fluxos de trabalho da arquitetura, a codificação da arquitetura subjacente, a fase do treinamento, a avaliação das métricas de desempenho e a execução da análise preditiva.

3.1. Aquisição e Seleção de Dados

O conjunto de dados de AVC, que compreende cortes axiais de TC da região craniana sem contraste, foi reunido como um elemento integrante da competição de Inteligência Artificial em Saúde que ocorreu em Istambul em 2021, recebendo aprovação do Ministério da Saúde, Direção Geral de Sistemas de Informação em Saúde e operando sob a gestão dos Institutos Turcos de Saúde (TUSEB) - <https://www.kaggle.com/datasets/ozguraslank/brain-stroke-ct-dataset> (KOÇ et al., 2022).

No total, compreende 5.557 imagens que foram coletadas entre os anos de 2019 e 2020 do *e-Pulse and Teleradiology System* afiliado ao Ministério da Saúde da República da Turquia. Todos os dados foram anonimizados e padronizados. Posteriormente, as imagens foram submetidas a avaliação por sete radiologistas, a fim de garantir a consistência diagnóstica e o rigor científico. Para este estudo, foram utilizadas:

- 1.130 imagens de AVC isquêmico, associadas a máscaras binárias que delimitam as regiões afetadas (subpastas ISCHEMIA/PNG e ISCHEMIA/OVERLAY);
- 4.427 imagens normais, oriundas da pasta NORMAL/PNG, utilizadas como grupo controle.

Os critérios adotados foram projetados especificamente com o objetivo de garantir a qualidade técnica das imagens adquiridas e a consistência clínica das anotações anexas, estabelecendo assim uma estrutura robusta para análise.

3.1.1 Critérios de inclusão

- Tomografia computadorizada do crânio que foram obtidas sem meio de contraste nas pastas designadas ISCHEMIA/PNG e NORMAL/PNG;
- As máscaras de segmentação correspondentes que são armazenadas na categoria ISCHEMIA/OVERLAY, que são essenciais para uma análise precisa;

3.1.2 Critérios de exclusão

- Imagens que se enquadram na categoria BLEEDING, que explicitamente não estão incorporadas ao escopo deste estudo específico;
- Exames que exibam artefatos significativos, cortes de imagem incompletos ou possuam uma resolução considerada insuficiente para fins de diagnóstico confiáveis;
- Casos em que há falta de correspondência entre a imagem original e a máscara de anotação associada, comprometendo assim a integridade da análise.

3.2 Pré-processamento

3.2.1 Divisão das imagens e descarte de regiões irrelevantes

Dado que as imagens iniciais exibiam resoluções diversas e não tinham regiões de interesse uniformemente distribuídas, foi determinado que essas imagens deveriam ser particionadas em patches quadrados medindo 256×256 pixels, em monocromático (1 canal), para garantir a compatibilidade com as especificações de entrada exigidas pelas arquiteturas implantadas.

Durante essa segmentação automatizada, a integridade informacional de cada par de imagens e sua máscara correspondente foi meticulosamente avaliada. Quando se descobriu que ambas as entidades continham valores exclusivamente nulos (ou seja, uma ausência completa de patologia e segmentação visíveis), esses pares foram eliminados. Esse procedimento diminuiu significativamente a quantidade de amostras supérfluas e agilizou o processo de treinamento, restando apenas as instâncias clinicamente pertinentes.

3.2.2 Aumento do conjunto de dados

A expansão do conjunto de dados foi executada por meio de duas metodologias complementares:

- Aumento de dados off-line: a mera segmentação de imagens em blocos menores resultou em uma quantidade significativamente aumentada de amostras, sem comprometer as informações anatômicas.
- Aumento de dados on-line: foi executado em tempo real durante todo o processo de treinamento do modelo, empregando metodologias como reflexão estocástica horizontal e vertical.

3.2.3 Divisão do conjunto de dados

O conjunto de dados foi dividido aleatoriamente em três subconjuntos para treinamento e avaliação eficaz do modelo: treinamento (70%), validação (15%), teste (15%).

3.2.4 Limites de pós-processamento e segmentação

Como a função de ativação da camada final dos modelos empregou uma função sigmoide, os resultados da segmentação foram representados como valores contínuos variando de 0 a 1, significando a probabilidade de presença de lesão. Consequentemente, foi possível implementar o pós-processamento por meio de ajuste de limite (limiar) para aumentar a sensibilidade ou especificidade da segmentação, em alinhamento com os objetivos clínicos e métricas priorizadas (como Dice, Intersection over Union (IoU), precisão, sensibilidade, etc.).

3.3 Definições de redes neurais

As arquiteturas foram implementadas utilizando a linguagem de programação Python. Além disso, foram utilizadas as estruturas TensorFlow/Keras, que oferecem interfaces

sofisticadas, uma comunidade de suporte e funcionalidades personalizadas especificamente para GPUs. O modelo de linguagem ChatGPT (OPENAI, 2025) facilitou o desenvolvimento e o refinamento de scripts Python ao propor trechos de código, retificando erros e sugerindo modificações lógicas, com todas as contribuições posteriormente avaliadas e executadas pelo autor. O processo empregou o algoritmo de otimização Adam, juntamente com uma função de perda baseada em *Dice Loss* e *Tversky Loss*, em conjunto com metodologias de regularização, incluindo parada antecipada, para mitigar o risco de *overfitting*. Os pesos das redes neurais foram atualizados na conclusão de cada época, e os modelos que exibiram desempenho superior no conjunto de dados de validação foram mantidos.

O conjunto de dados passou por vários limites de binarização, delineados com base em investigações empíricas conduzidas dentro do conjunto de dados de treinamento. Essa estrutura metodológica facilitou a identificação de valores ideais para métricas de desempenho designadas, como *Dice* e *Intersection over Union* (IoU), que foram posteriormente integradas ao conjunto de testes. A função de ativação que foi empregada no contexto da camada de saída foi especificamente a função sigmoide, selecionada por sua capacidade de produzir efetivamente gradientes de intensidade nas várias regiões segmentadas que foram analisadas durante o processo. Esse recurso permitiu a avaliação de métricas de desempenho, incluindo curvas de aprendizagem (Dice, IoU e Loss) e permitindo assim uma avaliação abrangente do avanço dos modelos ao longo das épocas e da eficácia do critério de parada precoce.

Em todos os protocolos experimentais, vários parâmetros, como tamanho do *batch*, taxa de aprendizado e épocas máximas, foram meticulosamente ajustados para acomodar as especificações de hardware.

O estudo utilizou 7 arquiteturas de U-Net para delinear lesões isquêmicas dentro do contexto do TCSC, garantindo assim que as condições de treinamento fossem uniformes em todos os aspectos. Essa estrutura metodológica permitiu uma avaliação sistemática das configurações de U-Net em um contexto clinicamente pertinente.

A implementação do programa foi realizada por meio da utilização de uma GPU de memória de 8 GB fabricada pela empresa NVIDIA®. A estrutura de desenvolvimento do TensorFlow incorpora a biblioteca CUDA11 em sua versão v11.2, o que facilita a aceleração de aplicativos devido aos seus núcleos de processamento dedicados. É imperativo destacar que os recursos computacionais delineados na Tabela 1 foram empregados para alcançar os resultados derivados das redes propostas neste estudo.

Tabela 1 – Computador e sistema utilizado

Processador	<i>Intel(R) Core(TM) Ultra 9 185H (2.50 GHz)</i>
Memória	<i>LPDDR5X onboard de 32GB (2x16GB)</i>
GPU	<i>NVIDIA® GeForce RTX™ 4070 8 GB GDDR6</i>
Biblioteca (GPU)	<i>CUDA 12.8 / CUDA Toolkit 11.2</i>
Armazenamento	<i>1TB SSD NVMe PCIe 4.0</i>
Sistema Operacional	<i>Windows 11 Home (versão 24H2) 64 bits</i>
Programação	<i>Python 3.8.20 (conda-forge)</i>
Bibliotecas de deep learning	<i>TensorFlow 2.10.1, Keras 2.10.0 e Scikit-learn 1.3.2</i>

Fonte: Elaborado pelo autor

3.4 Métricas de Avaliação e Visualização

A avaliação da segmentação automática de lesões isquêmicas no TCNC utilizou várias métricas estabelecidas na literatura de imagens médicas, fornecendo uma visão multifacetada do desempenho do modelo. Cada métrica contribuiu para uma análise completa da eficácia das arquiteturas implementadas.

Neste estudo, são exibidas as seguintes métricas: acurácia, sensibilidade, precisão, especificidade, coeficiente de similaridade de *Dice*, IoU, *F1-Score* e AUC e precisão média (AP), sendo estas apresentadas em sua forma padrão.

Além das avaliações quantitativas, um protocolo de visualização foi implementado, juxtapondo imagens de TCSC, as máscaras de referência e as predições. Esse método qualitativo auxilia na avaliação anatômica de regiões segmentadas, ressaltando a eficácia e as deficiências do modelo.

As curvas de aprendizado durante o treinamento ilustraram a progressão da função de perda, *Dice* e IoU, permitindo a avaliação da estabilidade do treinamento e os benefícios do *overfitting*. Essas métricas específicas foram de suma importância na avaliação da convergência do modelo, pois desempenharam um papel fundamental na determinação da estabilidade do desempenho do modelo, na garantia da reprodutibilidade e confiabilidade nos resultados.

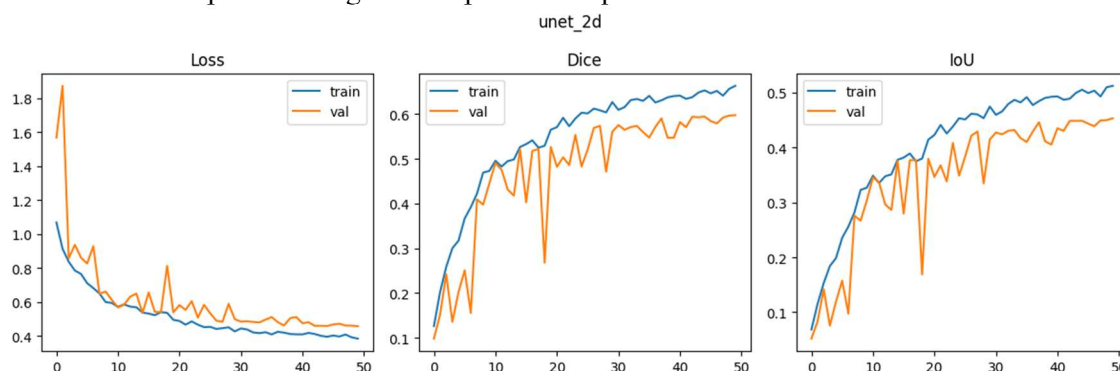
4 RESULTADOS

Ao explorar os diferentes cenários de treinamento, pretendemos avaliar o desempenho e as capacidades de generalização da nossa abordagem de segmentação. Para isso, separamos os resultados em quantitativos e qualitativos para demonstrar melhor os impactos das diferentes arquiteturas. Iremos apresentar estes resultados obtidos a partir da aplicação da arquitetura U-Net e suas variações na sequência de seus treinamentos. Utilizamos as seguintes métricas de desempenho na avaliação dos modelos: *Dice*, IoU, acurácia, precisão, sensibilidade, especificidade, *F1-Score*, AUC e AP. Na avaliação qualitativa, faremos uma interpretação visual e descritiva dos resultados de segmentação produzidos por cada modelo, buscando avaliar aspectos como: nitidez, definição de bordas, extensão da área segmentada e o comportamento do modelo em situações difíceis.

4.1 U-Net

Na análise do Gráfico 1, observamos um declínio consistente do *Loss* ao longo das épocas entre os conjuntos de treinamento e validação, com valores abaixo de 0,5. Tal comportamento indica uma convergência satisfatória do modelo, apesar de pequenas oscilações na validação. A métrica de similaridade (*Dice*), sinaliza para uma melhoria sustentada ultrapassando 0,65 no treinamento e se aproximando de 0,6 na validação. Isto indica uma sobreposição satisfatória entre a máscara real e a predição. A métrica IoU também apresenta um crescimento consistente superior a 0,5 ao final das 50 épocas, também demonstrando um desempenho satisfatório do modelo.

Gráfico 1 – Aprendizado gerados a partir da arquitetura U-Net



Fonte: Elaborado pelo autor

As métricas *Dice* e *IoU* refletem um forte desempenho de segmentação, apesar das complexidades das lesões isquêmicas do AVC (Tabela 2). O modelo apresenta alta especificidade (0,9963) e precisão (0,7527), indicando confiabilidade na minimização de falsos positivos. Por fim, os valores elevados de AUC (0,9808) e AP (0,6938) indicam a eficácia do modelo em vários limites de decisão.

Tabela 2 – Métricas da arquitetura U-Net

Métrica	Valor
Dice	0,6352
IoU	0,4654
Acurácia	0,9873
Precisão	0,7527
Sensibilidade	0,5494
Especificidade	0,9963
F1-Score	0,6352
AUC	0,9808
AP	0,6938

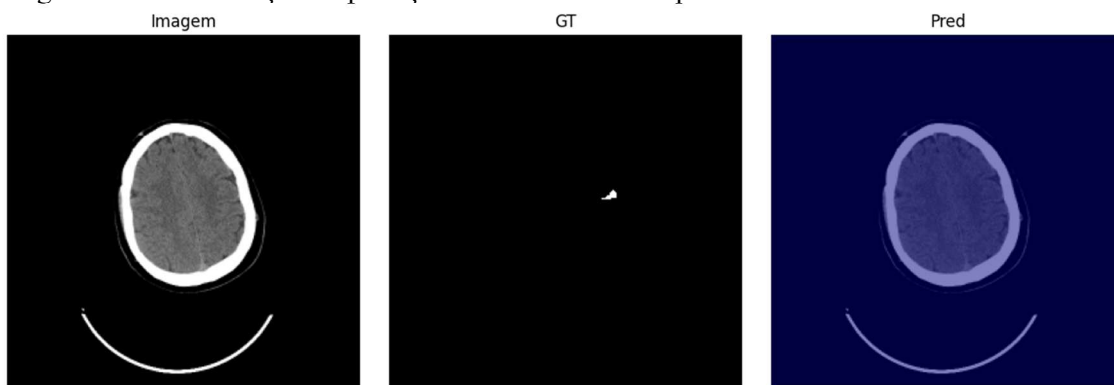
Fonte: Elaborado pelo autor

A avaliação qualitativa das previsões da U-Net é observada nas figuras abaixo. Na Figura 8a, a região segmentada do modelo (à direita) não exibe alinhamento com a máscara (imagem central) e a predição foi subestimada. Isso sugere desafios na detecção de padrões sutis em lesões pequenas, conforme indicado por uma métrica de sensibilidade de 0,5494. No entanto, a especificidade elevada (0,9963) implica na prevenção efetiva de falsos positivos em áreas não lesadas, conforme indicado pela predição conservadora na imagem.

Na Figura 8b, o modelo detectou de forma imprecisa lesões isquêmicas ausentes da máscara. Esse problema pode resultar da dificuldade em distinguir as regiões sadias das áreas isquêmicas. Essa avaliação visual se correlaciona com as métricas acima: o modelo exibe alta

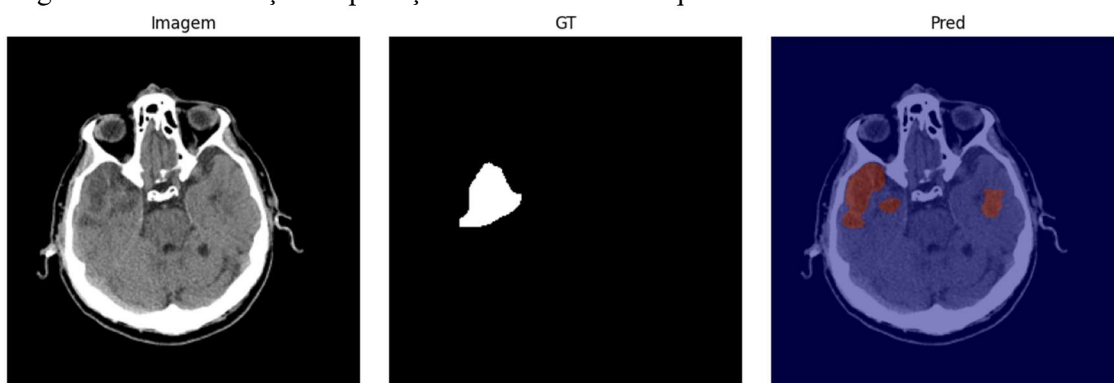
acurácia (0,9873), mas com precisão e *F1-Score* moderados (0,7527 e 0,6352, respectivamente), sugerindo adequação geral, mas notável potencial para aprimoramentos por meio de pós-processamento ou estratégias arquitetônicas sofisticadas.

Figura 8a – Visualização da predição em TC de AVC isquêmico U-Net – caso 1



Fonte: Elaborado pelo autor

Figura 8b – Visualização da predição em TC de AVC isquêmico U-Net – caso 2



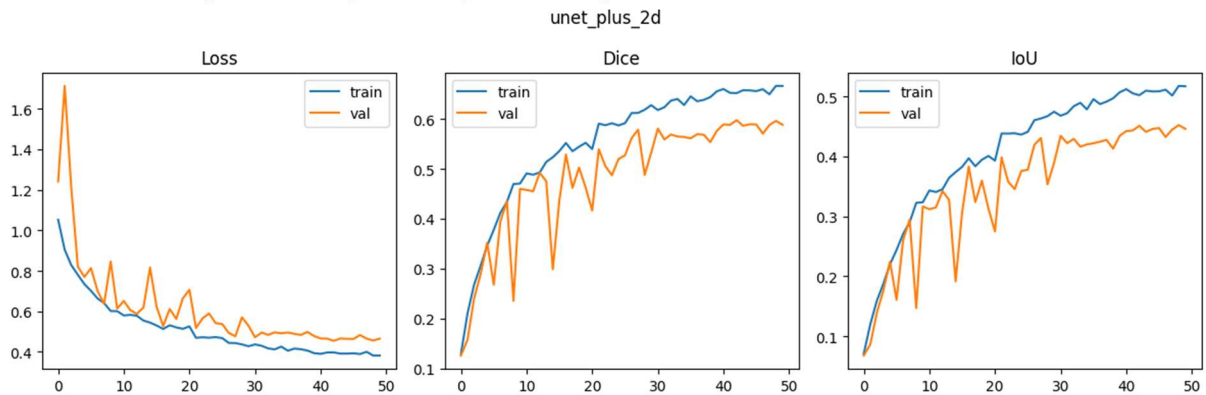
Fonte: Elaborado pelo autor

4.2 U-Net ++

Na análise do Gráfico 2, uma diminuição constante na perda foi observada ao longo das temporadas nos conjuntos de treinamento e validação, refletindo o aprendizado progressivo e a estabilização do modelo até a 30ª época. Isso implica um *overfitting* mínimo. A métrica de similaridade de dados revela um aprimoramento consistente, superando 0,7 no treinamento e quase 0,65 na validação após a 20ª época. Isso indica a capacidade do modelo de identificar padrões relevantes. A métrica IoU ilustra um rápido avanço nos estágios iniciais, com a

validação flutuando entre 0,45 e 0,5, indicando a competência do modelo em identificar com precisão a área da lesão.

Gráfico 2 – Aprendizado gerados a partir da arquitetura U-Net++



Fonte: Elaborado pelo autor

As métricas Dice (0,6316) e IoU (0,4615) indicam um nível moderado de precisão do modelo em sua capacidade preditiva. A *F1-Score* de 0,6316 indica uma relação equilibrada entre precisão e sensibilidade na detecção de lesões. Apesar da alta acurácia (0,9873), é necessário cuidado devido ao possível desequilíbrio de dados que afeta a detecção de pequenos infartos. A precisão (0,7542) significa uma forte taxa de verdadeiros positivos na classificação de lesões, minimizando efetivamente os alarmes falsos. A sensibilidade de 0,5432 significa que cerca de 54% das áreas isquêmicas foram detectadas com precisão, destacando a necessidade de aprimoramentos no modelo para reduzir os falsos negativos. A especificidade (0,9964) destaca a proficiência do modelo em identificar com precisão tecidos saudáveis, o que é vital para a precisão clínica. A AUC (0,9807) demonstra a notável capacidade do modelo de distinguir entre classes, significando generalização efetiva, apesar dos desafios na detecção de lesões. O AP (0,6987) oferece precisão média aceitável em vários limites, relevante para estruturas de avaliação mais detalhadas (Tabela 3).

Tabela 3 – Métricas da arquitetura U-Net++

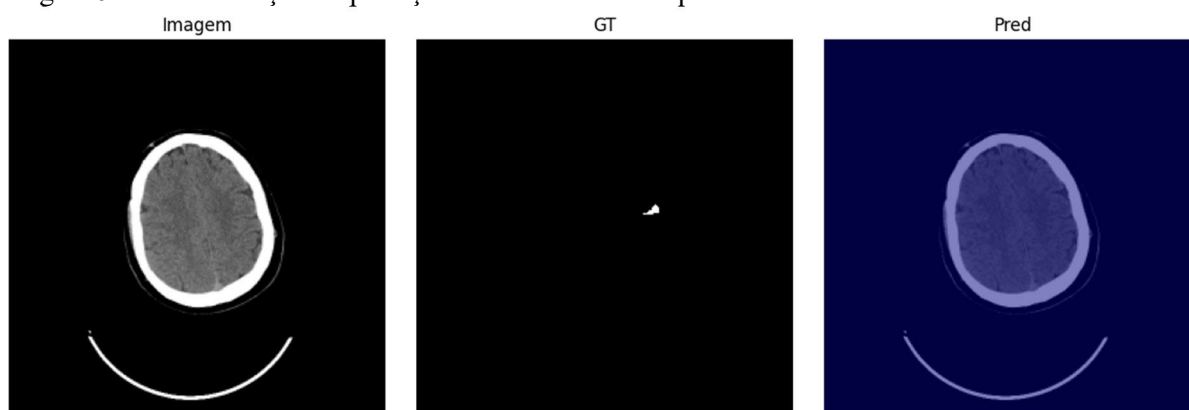
Métrica	Valor
Dice	0,6316
IoU	0,4615
Acurácia	0,9873
Precisão	0,7542
Sensibilidade	0,5432
Especificidade	0,9964
F1-Score	0,6316
AUC	0,9807
AP	0,6987

Fonte: Elaborado pelo autor

Na Figura 9a, a região segmentada do modelo (à direita) não exibe alinhamento com a máscara (imagem central) e a predição foi subestimada, como observado na arquitetura U-Net. Isso sugere falsos positivos e refletindo baixa especificidade local e sensibilidade reduzida para pequenas lesões em meio ao ruído de fundo.

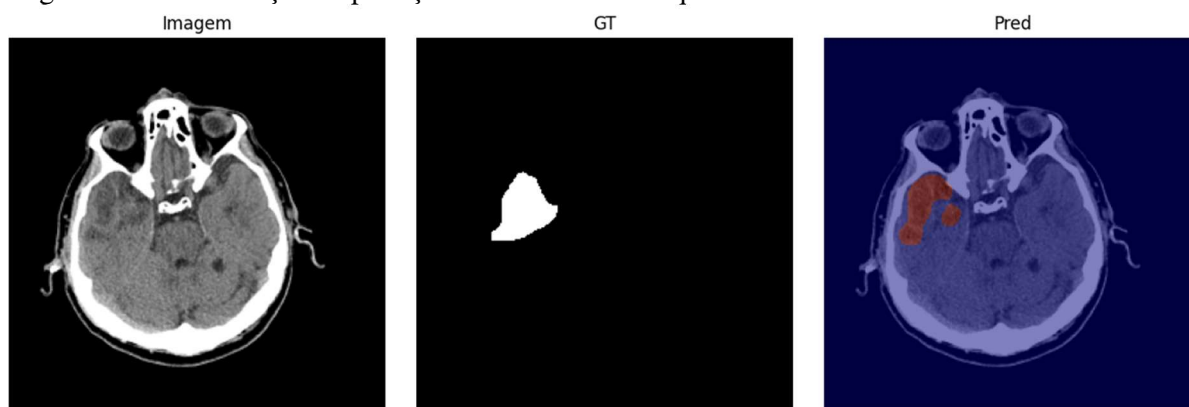
Na segunda imagem, a lesão isquêmica é marcadamente pronunciada na máscara, abrangendo uma área mais extensa e densa. A U-Net++ identifica efetivamente a vizinhança geral da lesão, que apresenta congruência espacial com as características reais do terreno. No entanto, a predição revela regiões maiores do que as indicadas pela máscara e incorpora áreas errôneas adjacentes, o que significa uma superestimação da extensão da lesão. Esse fenômeno pode ser elucidado pelas restrições do modelo em delinear limites precisos com precisão, pois ele tende a adotar uma abordagem mais conservadora ao designar regiões que exibem intensidades de sinal comparáveis (Figura 9b).

Figura 9a – Visualização da predição em TC de AVC isquêmico U-Net++ – caso 1



Fonte: Elaborado pelo autor

Figura 9b – Visualização da predição em TC de AVC isquêmico U-Net++ – caso 2

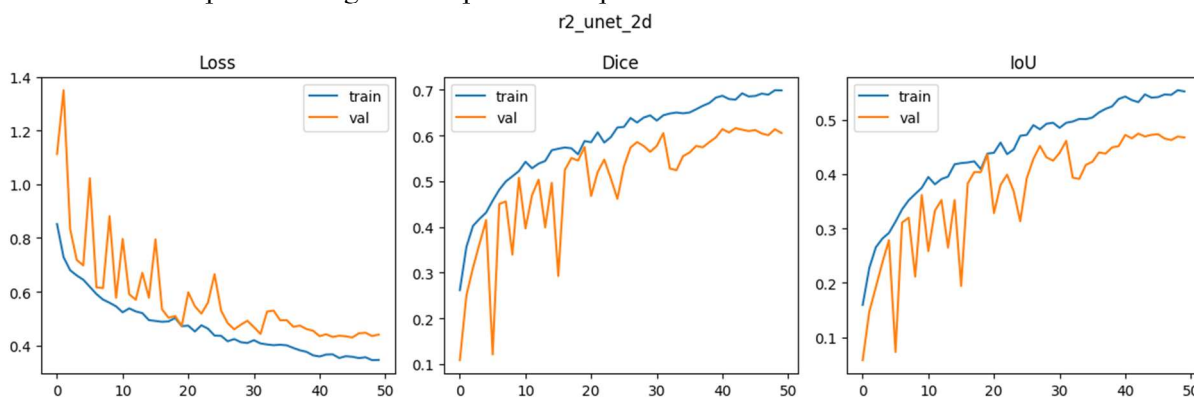


Fonte: Elaborado pelo autor

4.3 R2 U-Net

Na análise do Gráfico 3, uma redução perceptível e consistente na Loss foi observada em várias épocas no conjunto de dados de treinamento. No conjunto de dados de validação, após as flutuações iniciais, a curva atinge um estado de estabilidade após a 20ª época. Esse fenômeno sugere uma convergência adequada do modelo, apesar de sua suscetibilidade a flutuações nos dados de entrada. A métrica de similaridade (*Dice*) demonstra um aumento consistente no conjunto de dados de treinamento, atingindo um valor aproximado de 0,70 após 50 épocas, enquanto no conjunto da validação exibe estabilização entre 0,63 e 0,65. Essa observação significa um grau notável de sobreposição entre a máscara real e a saída preditiva. Além disso, a métrica IoU também demonstra uma tendência ascendente consistente, culminando em 0,55 com a conclusão da fase de treinamento e, apesar de certas instabilidades, a validação atinge a estabilização na faixa de 0,48 a 0,50.

Gráfico 3 – Aprendizado gerados a partir da arquitetura R2U-Net



Fonte: Elaborado pelo autor

A arquitetura R2U-Net demonstrou maior eficácia, atingindo um coeficiente *Dice* de 0,6639 e a IoU de 0,4969, excedendo assim as métricas de desempenho da U-Net e da U-Net++. A proximidade do IoU a 0,5 demonstra uma sobreposição considerável entre os resultados esperados e reais, significando consistência robusta na segmentação da região isquêmica. A acurácia do modelo de 98,84% reflete alto desempenho; no entanto, é essencial reconhecer que essa precisão é afetada por um alto número de negativos verdadeiros em imagens com lesões mínimas. Além disso, uma especificidade de 0,9969 reforça a eficácia do modelo na redução de classificações de falsos positivos de tecido saudável. A R2U-Net exibe uma detecção

aprimorada de lesões, com uma precisão de 0,7894 e uma sensibilidade de 0,5729. Apesar de a sensibilidade continuar abaixo do limite de 0,6, ela significa uma melhoria em relação a U-Net e a U-Net++. A *F1-Score* de 0,6639 reflete o equilíbrio ideal entre precisão e sensibilidade, superando a U-Net e a U-Net++. Por outro lado, a AUC de 0,9865 reflete o excepcional poder discriminatório do modelo em vários níveis de classe, indicando forte eficácia de classificação. Uma AP de 0,7467 demonstra um desempenho superior em vários níveis de confiança. Isso ressalta a confiabilidade e a consistência preditivas do modelo, mesmo em contextos clínicos complexos (Tabela 4).

Tabela 4 – Métricas da arquitetura R2U-Net

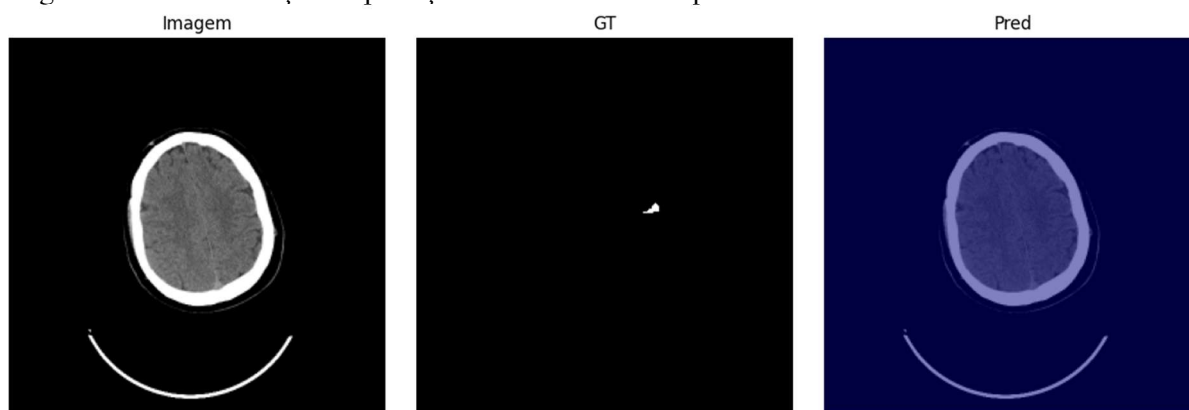
Métrica	Valor
Dice	0,6639
IoU	0,4969
Acurácia	0,9884
Precisão	0,7894
Sensibilidade	0,5729
Especificidade	0,9969
F1-Score	0,6639
AUC	0,9865
AP	0,7467

Fonte: Elaborado pelo autor

A avaliação qualitativa das previsões da R2U-Net é observada nas figuras abaixo. Na Figura 10a, a região segmentada do modelo (à direita) falhou em delinear com precisão a lesão isquêmica, como observado nas arquiteturas anteriores. Essas falhas são esperadas em casos envolvendo lesões diminutas e pouco contrastantes, o que pode sugerir limitações inerentes à arquitetura em relação à detecção de lesões caracterizadas por baixo volume ou baixo contraste e generalização inadequada para apresentações clínicas diferenciadas.

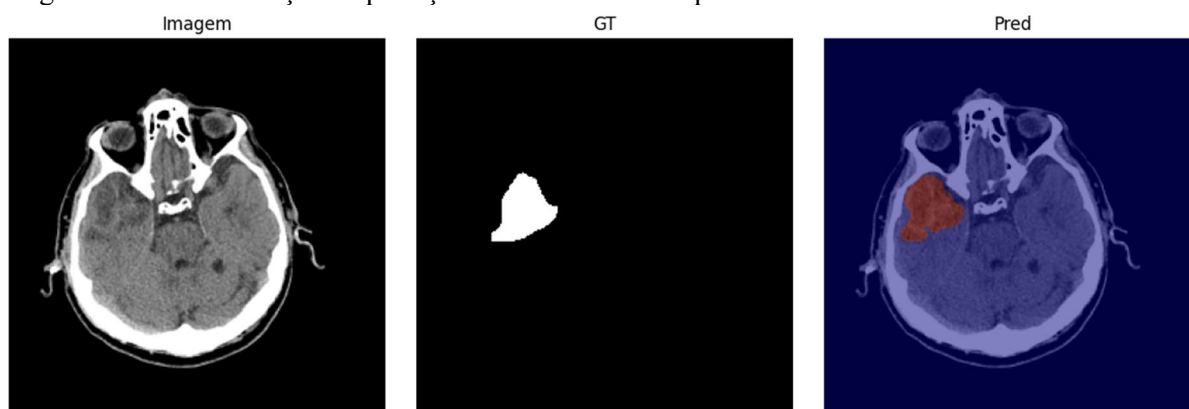
Na segunda imagem, a previsão da R2U-Net se aproxima muito da máscara. A arquitetura R2U-Net exibiu uma eficácia notável neste caso clínico, particularmente em: o delineamento espacial da área afetada; a conformidade com a verdadeira morfologia anatômica da lesão; e a identificação de estruturas de contorno complexas (Figura 10b).

Figura 10a – Visualização da predição em TC de AVC isquêmico R2U-Net – caso 1



Fonte: Elaborado pelo autor

Figura 10b – Visualização da predição em TC de AVC isquêmico R2U-Net – caso 2



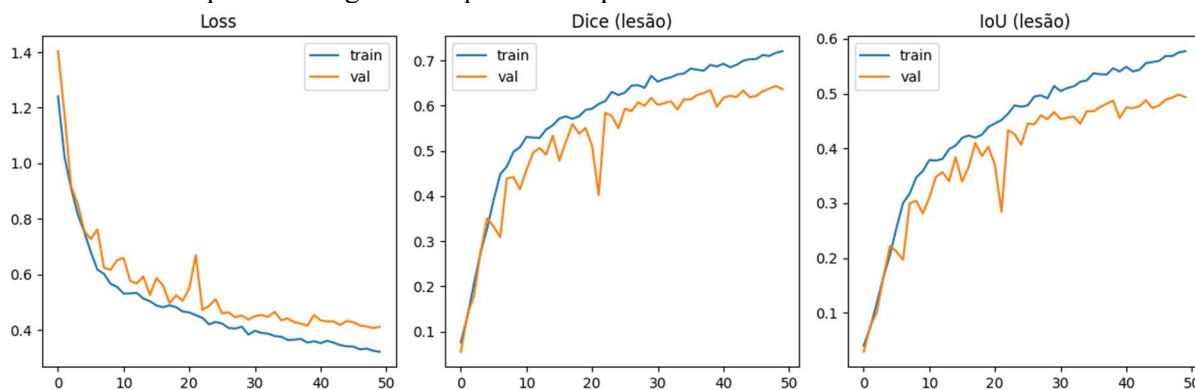
Fonte: Elaborado pelo autor

4.4 V-Net

Na análise do Gráfico 4, um declínio persistente na *Loss* foi observado de forma consistente nos conjuntos de dados de treinamento e validação. Dentro do conjunto de dados de validação, após variações esporádicas, a curva atinge uma condição de equilíbrio subsequente à 35ª época. A métrica de similaridade (*Dice*) exibe uma tendência ascendente consistente no conjunto de dados de treinamento, ultrapassando o limite de 0,70 na 50ª época; inversamente, dentro do conjunto de validação, apesar das variações observadas entre as épocas 15ª e 25ª, ela mostra uma tendência de se estabilizar em torno de 0,60. A proximidade dessas representações gráficas significa uma capacidade robusta de generalização. Vale ressaltar que a métrica IoU apresenta uma forte correlação com o coeficiente de *Dice*, relação que pode ser

exemplificada por meio da curva de treinamento, que gradualmente se aproxima de um valor de 0,58, enquanto a curva de validação mantém a estabilidade na faixa de 0,48 a 0,50.

Gráfico 4 – Aprendizado gerados a partir da arquitetura V-Net



Fonte: Elaborado pelo autor

Com um coeficiente *Dice* de 0,6897 e uma IoU de 0,5264, a arquitetura V-Net exibiu desempenho elevado, superando suas equivalentes (U-Net, U-Net++ e R2U-Net). Isso indica um aumento na robustez e confiabilidade dos resultados da segmentação. A elevada taxa de acurácia de 98,88% e especificidade de 0,9963 indicam a capacidade na identificação precisa de tecidos saudáveis. Além disso, a precisão de 0,7736 revela que quando a V-Net categoriza uma área específica como uma lesão, ela o faz com um alto nível de confiabilidade, aumentando assim a confiança em sua capacidade de diagnóstico. A sensibilidade de 0,6222 foi a mais alta registrada entre os modelos U-Net, U-Net++ e R2U-Net, indicando uma maior habilidade na detecção de áreas isquêmicas. Além disso, a *F1-Score* de 0,6897 ressalta ainda mais o equilíbrio favorável que existe entre sensibilidade e precisão nas métricas de desempenho do modelo. O valor da AUC de 0,9838 indica que o modelo possui uma ótima capacidade de discriminação, enquanto a AP de 0,7502 sugere um nível de estabilidade em seu desempenho preditivo em vários limites de decisão, atributo altamente favorável para sua aplicação e implementação no mundo real (Tabela 5).

Tabela 5 – Métricas da arquitetura V-Net

Métrica	Valor
Dice	0,6897
IoU	0,5264
Acurácia	0,9888
Precisão	0,7736
Sensibilidade	0,6222
Especificidade	0,9963

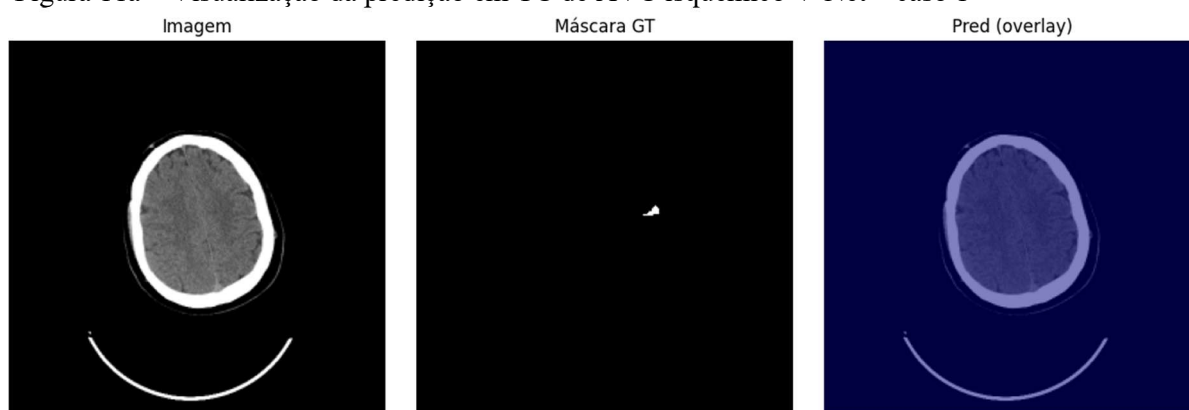
F1-Score	0,6897
AUC	0,9838
AP	0,7502

Fonte: Elaborado pelo autor

Na Figura 11a, a arquitetura V-Net exibiu uma falha na capacidade de detectar e delinear com precisão a lesão localizada no tecido cerebral e a predição foi subestimada, como observado em arquiteturas anteriores.

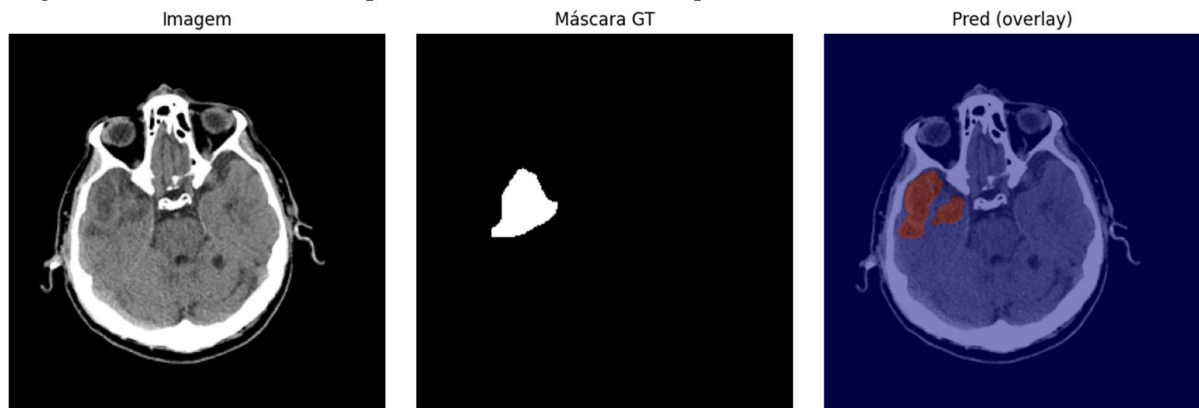
Em contraste com a figura anterior, o modelo V-Net exibiu a capacidade de detectar com sucesso a região da lesão real, alcançando um grau razoável de sobreposição; no entanto, é importante observar que a previsão resultou em um leve grau de hipersegmentação, o que indica que os limites previstos da lesão se estendem além dos limites patológicos reais em certas áreas. Embora possa ser observada a presença de algum grau de excesso de predição, caracterizado como falso positivo leve, vale ressaltar que a rede neural capturou efetivamente a maioria das lesões, o que serve como um indicador positivo do desempenho do modelo em relação a lesões maiores (Figura 11b).

Figura 11a – Visualização da predição em TC de AVC isquêmico V-Net – caso 1



Fonte: Elaborado pelo autor

Figura 11b – Visualização da predição em TC de AVC isquêmico V-Net – caso 2

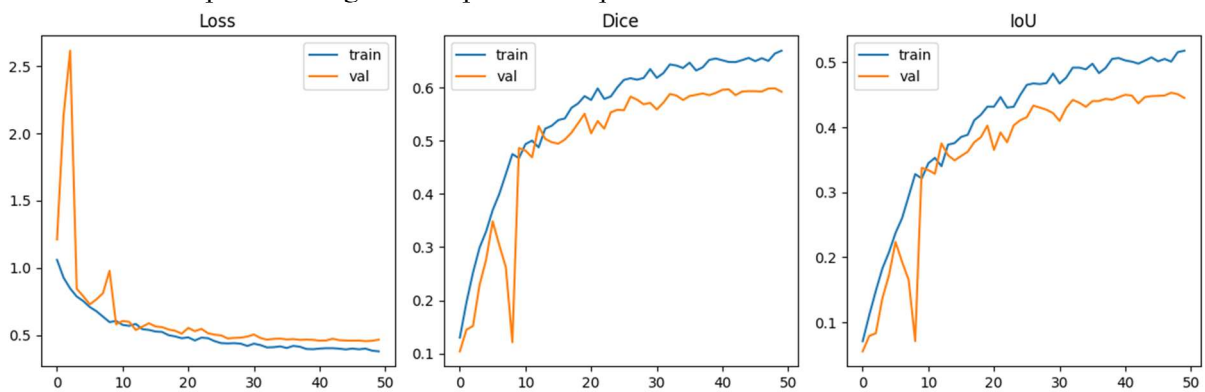


Fonte: Elaborado pelo autor

4.5 Attention U-Net

Na análise do Gráfico 5, um declínio persistente na *Loss* foi observado de forma consistente no conjunto de dados de treino. Dentro do conjunto de dados de validação, após as variações durante as primeiras épocas, a curva atinge uma condição de equilíbrio após a 10^a época em torno do valor de 0,5. A métrica de similaridade (*Dice*) exibe uma tendência ascendente consistente no conjunto de dados de treino, alcançando valores acima de 0,75; a métrica também cresce de forma estável no conjunto de validação, estabilizando-se entre 0,65 e 0,68. A proximidade dessas representações gráficas significa que o modelo não está sofrendo *overfitting* grave. A métrica IoU no conjunto de treino apresenta um crescimento constante até ultrapassar 0,55, enquanto o conjunto de validação evolui de forma semelhantes à métrica Dice, com estabilidade após as primeiras 15 a 20 épocas, ficando em torno de 0,45 a 0,47.

Gráfico 5 – Aprendizado gerados a partir da arquitetura Attention U-Net



Fonte: Elaborado pelo autor

Com um coeficiente Dice de 0,6230 e uma IoU de 0,4524, a arquitetura Attention U-Net exibiu um desempenho moderado e demonstrou a eficácia limitada do mecanismo de atenção. A elevada taxa de acurácia de 98,61% e especificidade de 99,45% indicam a capacidade na identificação precisa de tecidos saudáveis, mas não se correlacionam diretamente com o sucesso da detecção de lesões isquêmicas. A precisão de 0,6819 e a sensibilidade de 0,5735 indicam um desempenho mais equilibrado, porém ligeiramente inferior em comparação com a V-Net e a R2U-Net. O F1-Score de 0,6230 indica um equilíbrio entre precisão e sensibilidade, todavia menor do que o do R2U-Net e do V-Net. O valor da AUC de 0,9802 ilustra a elevada capacidade de discriminação do modelo, enquanto a AP de 0,6930 significa estabilidade em vários *threshold* (Tabela 6).

Tabela 6 – Métricas da arquitetura Attention U-Net

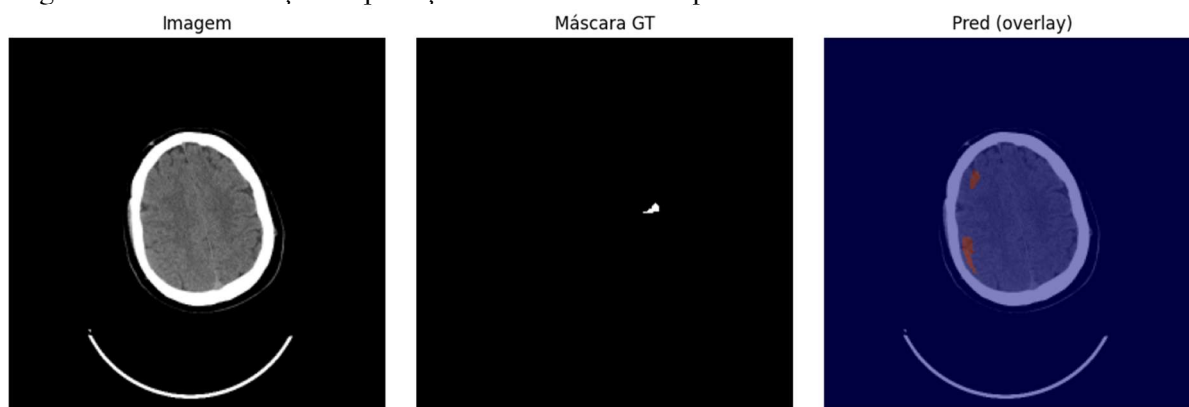
Métrica	Valor
Dice	0,6230
IoU	0,4524
Acurácia	0,9861
Precisão	0,6819
Sensibilidade	0,5735
Especificidade	0,9945
F1-Score	0,6230
AUC	0,9802
AP	0,6930

Fonte: Elaborado pelo autor

Na Figura 12a, a arquitetura Attention U-Net exibiu uma falha na capacidade de detectar e delinear com precisão a lesão localizada no tecido cerebral. O modelo identificou duas áreas em hemisférios opostos, ambas falsamente positivas.

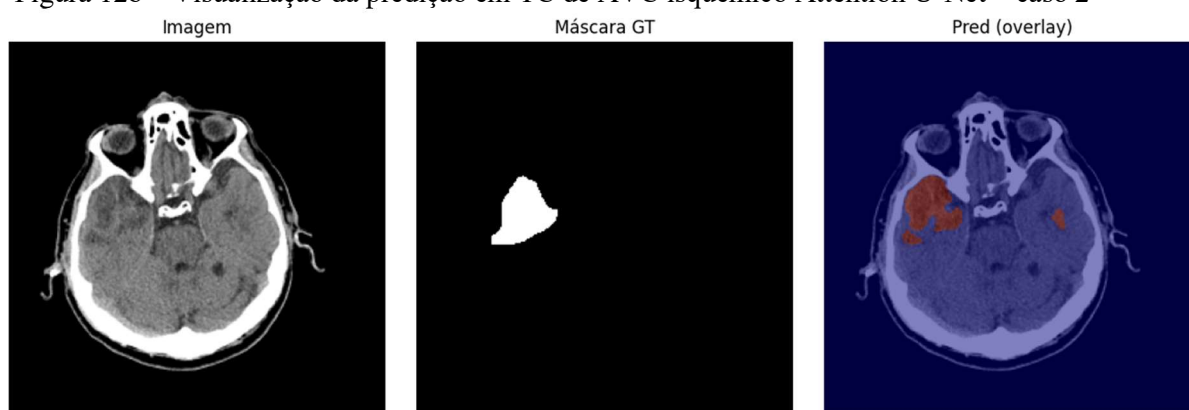
Em contraste com a figura anterior, o modelo Attention U-Net exibiu a capacidade de detectar com sucesso a maior parte da região da lesão real, alcançando um grau razoável de sobreposição; no entanto, existem regiões falso-positivas suplementares (adjacentes à área precisa e no lado contralateral). Esse comportamento é antecipado em arquiteturas que incorporam camadas de atenção, pois existe uma propensão a priorizar regiões próximas visualmente análogas (hipodensas) (Figura 12b).

Figura 12a – Visualização da predição em TC de AVC isquêmico Attention U-Net – caso 1



Fonte: Elaborado pelo autor

Figura 12b – Visualização da predição em TC de AVC isquêmico Attention U-Net – caso 2



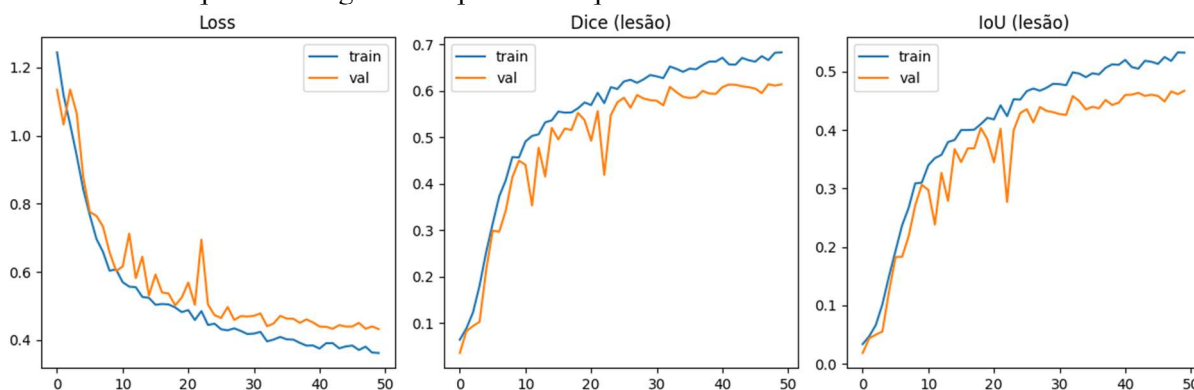
Fonte: Elaborado pelo autor

4.6 ResUNet-a

Na análise do Gráfico 6, uma redução contínua na Loss foi detectada de forma consistente nos conjuntos de dados de treino e validação. O conjunto de dados de validação atinge uma condição de equilíbrio em 0,43, apesar das oscilações, e o conjunto de dados de treino em 0,38, sem caracterizar um overfitting grave e indicando uma boa convergência. A métrica de similaridade Dice demonstra um aumento constante no conjunto de dados de treinamento, acima de 0,65, ao mesmo tempo em que aumenta no conjunto de validação, estabilizando-se acima de 0,55. A diferença é pequena entre as curvas de treino e validação a partir da 25ª época, demonstrando uma boa capacidade de generalização. A métrica IoU Union (IoU) exibe um padrão análogo ao do coeficiente Dice, demonstrando um aumento consistente

e progressivo em todo o conjunto de dados de treinamento até 0,52, ao mesmo tempo em que atinge um estado de estabilização para o conjunto de dados de validação em um nível que gira em torno de 0,45. Vale ressaltar que a curva de validação apresenta maior estabilidade após a 25ª época, permitindo que o modelo mantenha suas métricas de desempenho de forma eficaz ao longo do tempo.

Gráfico 6 - Aprendizado gerados a partir da arquitetura ResUNet-a



Fonte: Elaborado pelo autor

O coeficiente Dice de 0,6699 e a IoU de 0,5036 excederam o desempenho de arquiteturas estabelecidas, como U-Net original, U-Net++ e Attention U-Net. A taxa de precisão de 0,7293, e uma sensibilidade de 0,6194, ilustra um equilíbrio entre as taxas de falsos positivos e falsos negativos. O modelo atingiu um bom equilíbrio entre a seletividade e a identificação de lesões.

Com uma acurácia de 98,78% e uma especificidade de 99,53%, o modelo demonstra sua eficácia na distinção entre tecidos saudáveis e isquêmicos, particularmente em contextos clínicos onde a ocorrência de tecidos saudáveis é frequentemente elevada. A especificidade elevada significa uma taxa mínima de falsos positivos, aumentando assim a confiabilidade do resultado do diagnóstico.

O F1-Score de 0,6194 é comparável com as métricas de desempenho da Attention U-Net; no entanto, permanece inferior aos índices de desempenho registrados pela V-Net (0,6897) e R2U-Net (0,6639). A área sob a curva (AUC) de 0,9835 aumenta significativamente a eficácia discriminativa geral do modelo em múltiplos *thresholds*. Por outro lado, a precisão média (AP) de 0,7314 está entre os valores mais altos alcançados entre todas as arquiteturas avaliadas, o que é fundamental para segmentações clínicas adaptáveis (Tabela 7).

Tabela 7 – Métricas da arquitetura ResUNet-a

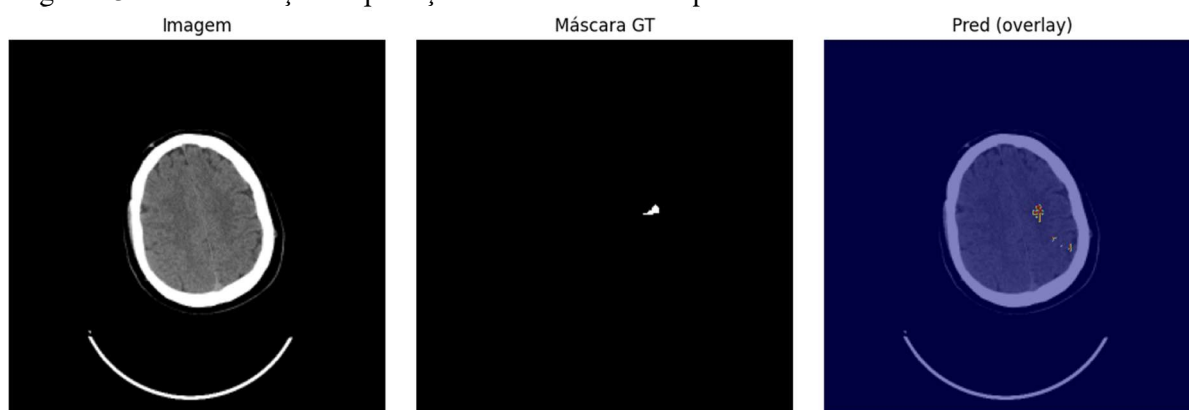
Métrica	Valor
Dice	0,6699
IoU	0,5036
Acurácia	0,9878
Precisão	0,7293
Sensibilidade	0,6194
Especificidade	0,9953
F1-Score	0,6699
AUC	0,9835
AP	0,7314

Fonte: Elaborado pelo autor

Na Figura 13a, o ResUnet-A exibiu uma deficiência em sua capacidade de reconhecer e delinear com precisão a lesão localizada dentro do tecido cerebral. A previsão resultante parece estar subdimensionada e também parcialmente desalinhada, o que levanta preocupações quanto à sua confiabilidade; além disso, existe uma infeliz ocorrência de falsos positivos que estão dispersos pela área parenquimatosa ipsilateral, bem como nas proximidades da lesão real, complicando assim a clareza interpretativa dos resultados.

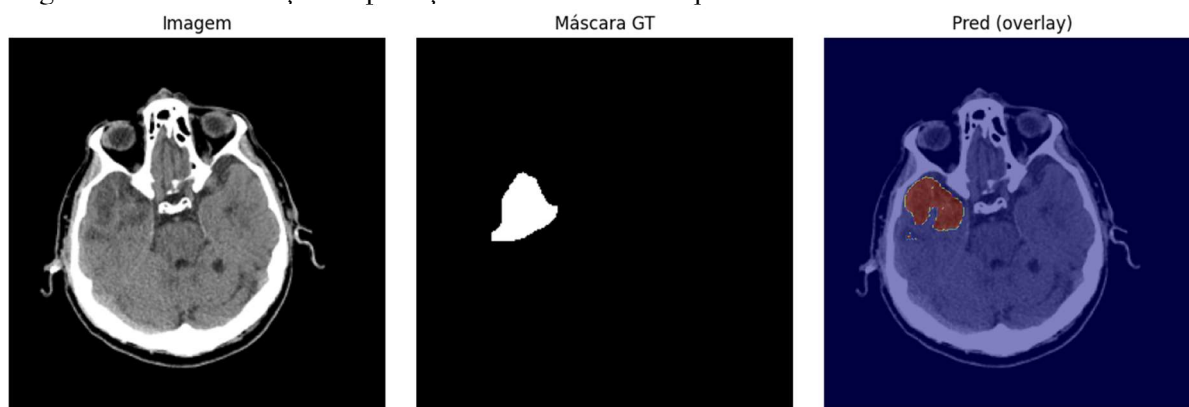
Em contraste com as observações feitas na figura anterior, o modelo ResUnet-A demonstrou um nível elevado de congruência entre a máscara gerada e a predição correspondente. A previsão do modelo abrange com sucesso a maior parte da lesão real, demonstrando uma conformidade louvável das bordas, apesar das imagens de segmentação excessiva no tecido cerebral ipsilateral posterior. Nesse contexto, a ResUnet-A segmenta efetivamente lesões grandes, garantindo integridade morfológica e distribuição espacial realista (Figura 13b).

Figura 13a – Visualização da predição em TC de AVC isquêmico ResUnet-A – caso 1



Fonte: Elaborado pelo autor

Figura 13b – Visualização da predição em TC de AVC isquêmico ResUnet-A – caso 2

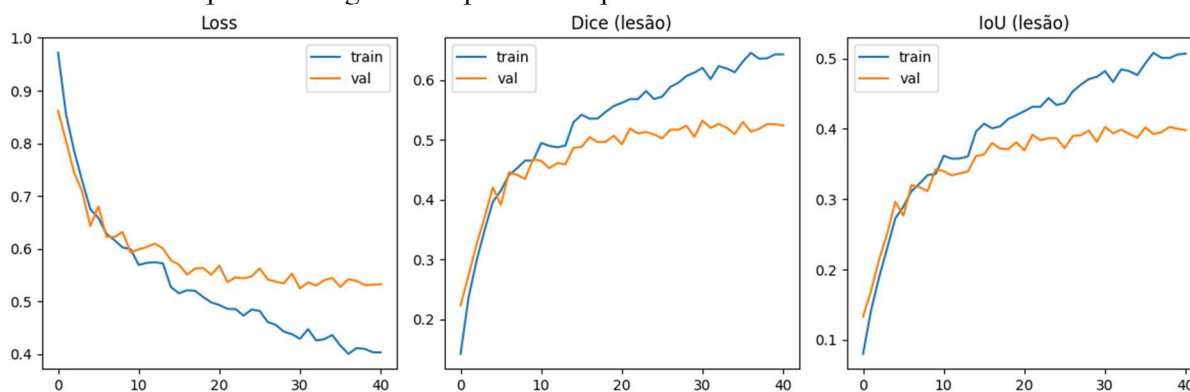


Fonte: Elaborado pelo autor

4.7 Swin U-Net

Na análise do Gráfico 7, uma redução contínua na *Loss* foi detectada de forma consistente nos conjuntos de dados de treinamento e validação, particularmente aparente durante as dez temporadas iniciais. O conjunto de dados de validação atinge uma condição de equilíbrio a partir da 15^a à 20^a época, oscilando em torno de um platô sem demonstrar nenhum avanço considerável. Por outro lado, o conjunto de dados de treinamento continua diminuindo a uma taxa mais pronunciada, o que pode sugerir o início do *overfitting*. A métrica de similaridade *Dice* demonstra um aumento constante no conjunto de dados de treinamento, próximo a 0,68, ao mesmo tempo em que aumenta no conjunto de validação, estabilizando-se entre 0,53 e 0,55. A diferença entre as curvas de treino e validação ocorre a partir da 20^a época, fortalecendo a possibilidade de *overfitting* leve a moderado. A métrica IoU no conjunto de treinamento aumenta consistentemente para 0,52 e se estabiliza para o conjunto de dados de validação em aproximadamente 0,41. A curva de validação é mais estável após a época 15^a, o que é indicativo de robustez, apesar dos valores relativamente baixos.

Gráfico 7 – Aprendizado gerados a partir da arquitetura Swin U-Net



Fonte: Elaborado pelo autor

A arquitetura Swin U-Net demonstrou um índice de sensibilidade de 0,6170, significando uma capacidade moderada na identificação precisa de regiões isquêmicas, com métricas de desempenho que estão no mesmo nível de outros modelos que mostraram eficácia superior na detecção de áreas afetadas. No entanto, a taxa de precisão de 0,6415 indica que uma fração considerável dos pixels classificados como positivos se deve a falsos positivos, destacando assim uma certa vulnerabilidade na seletividade das previsões, o que prejudica a confiabilidade da segmentação em regiões que permanecem inalteradas.

A precisão geral de 98,54% permanece excepcionalmente alta, um fenômeno previsto em conjuntos de dados médicos caracterizados pela predominância de tecidos normais. Por outro lado, a especificidade foi registrada em 0,9929, um valor elevado que ressalta a capacidade robusta do modelo de identificar com precisão regiões sem lesão.

O *F1-Score* de 0,6290, que fica em uma posição intermediária entre sensibilidade e precisão, indica um nível aceitável de desempenho geral; no entanto, permanece inferior ao de modelos como V-Net (0,6897) e R2U-Net (0,6639). Em contraste, a área sob a curva ROC (AUC) foi registrada em 0,9825, significando uma capacidade excepcional de discriminar entre regiões lesadas e não lesadas em vários limites. Esse resultado coloca o modelo gerado pela Swin U-Net entre aqueles mais distintos em relação à eficácia discriminativa, embora com restrições em relação à precisão e ao *F1-Score* (Tabela 8).

Tabela 8 – Métricas da arquitetura Swin U-Net

Métrica	Valor
Dice	0.6290
IoU	0.4588
Acurácia	0.9854
Precisão	0.6415
Sensibilidade	0.6170

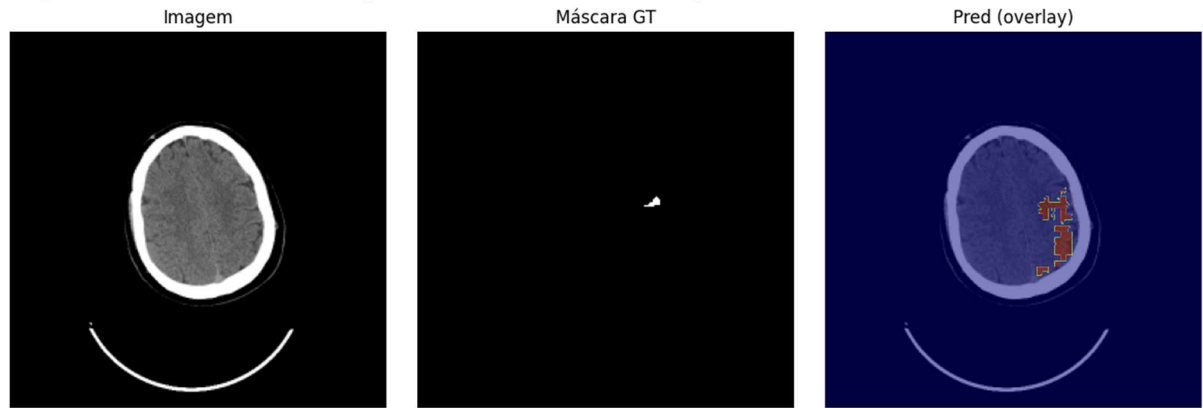
Especificidade	0.9929
F1-Score	0.6290
AUC	0.9825
AP	0.6826

Fonte: Elaborado pelo autor

Na Figura 14a, a arquitetura Swin U-Net exibiu uma falha na capacidade de detectar e delinear com precisão a lesão localizada no tecido cerebral. O modelo identificou áreas no mesmo hemisfério cerebral, que ultrapassam as dimensões da máscara, sugerindo a presença de supersegmentação moderada.

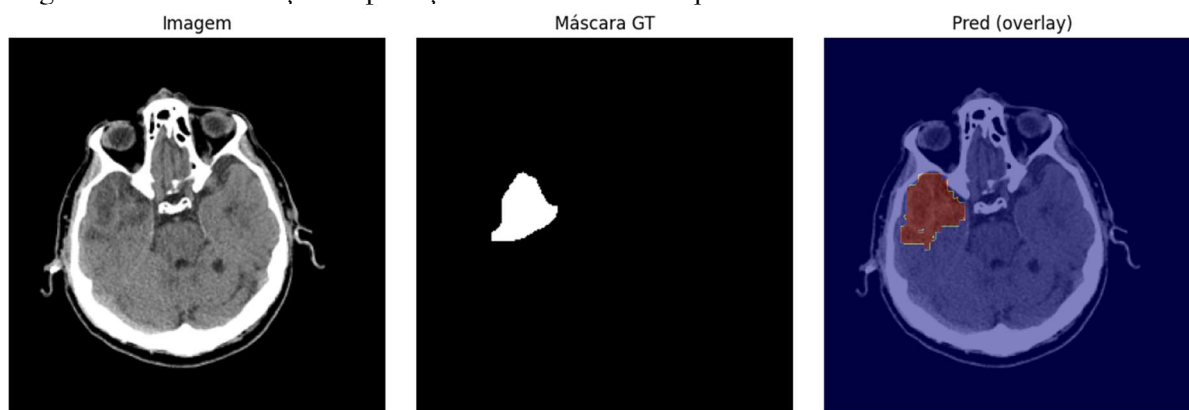
Em contraste com a figura anterior, o modelo da Swin U-Net exibiu uma ótima congruência entre a máscara e a predição, detectando com sucesso a maior parte da região da lesão real, com contornos bem definidos e leve expansão lateral — o que pode representar uma tentativa do modelo em capturar áreas de penumbra ou lesão adjacente; no entanto, existem regiões falso-positivas suplementares (adjacentes à área real). Esse comportamento reflete o desempenho da atenção espacial da Swin U-Net, refletindo sua capacidade de integrar contexto global e detalhes locais com eficácia (Figura 14b).

Figura 14a – Visualização da predição em TC de AVC isquêmico Swin U-Net – caso 1



Fonte: Elaborado pelo autor

Figura 14b – Visualização da predição em TC de AVC isquêmico Swin U-Net – caso 2



Fonte: Elaborado pelo autor

5 DISCUSSÃO

A arquitetura U-Net, apesar de um desempenho inferior em relação às variantes contemporâneas, exibiu métricas satisfatórias: um coeficiente Dice médio de 0,6352, um IoU de 0,4654 e uma acurácia de 0,9873. Essas descobertas estão alinhadas com a literatura existente. Por exemplo, Ronneberger et al. (2015), os criadores da estrutura U-Net original, documentaram coeficientes de Dice superiores a 0,80 para imagens biomédicas segmentadas caracterizadas por uma relação sinal/ruído substancial. No entanto, estudos mais recentes como Zhao et al. (2020) demonstraram que a U-Net tende a apresentar menor desempenho na segmentação de lesões com margens irregulares e baixo contraste, como em AVC isquêmico em TC não contrastada.

No que diz respeito à segmentação precisa das lesões isquêmicas, especialmente nas imagens com áreas hipodensas pequenas ou com contraste reduzido, a arquitetura U-Net obteve um desempenho inferior, o que reforça o alerta feito por Ronneberger et al. (2015) quanto à limitação da U-Net frente a estruturas muito pequenas em imagens com grande proporção de fundo. Por outro lado, sua simplicidade arquitetural permitiu estabilidade de treinamento e baixa oscilação nas curvas de validação, tornando-a ainda uma referência como baseline confiável, especialmente em contextos com recursos computacionais limitados.

Em nossas predições, observamos uma propensão à segmentação centralizada; no entanto, esse fenômeno foi concomitante com uma perda nos detalhes periféricos e uma maior suscetibilidade ao ruído de fundo. A ausência de blocos residuais ou mecanismos de atenção

torna a U-Net cada vez mais propensa a sobreajustes e limitações na delineação proficiente de padrões mais complexos (RONNEBERGER et al., 2015).

O U-Net++ alcançou resultados razoáveis, atingindo um coeficiente Dice de 0,63 e um IoU de 0,46, indicando assim um grau aceitável de generalização e eficácia de segmentação, embora exibindo pequenos casos de sobreajuste durante as épocas iniciais. Na literatura, Zhou et al. (2018), os criadores do U-Net++, alcançaram coeficientes de Dice superiores a 0,89 no contexto da segmentação de neoplasias malignas de pulmão utilizando TC, em conjunto com valores de IoU próximos a 0,80. Além disso, Wang et al. (2020) empregaram o U-Net++ para a avaliação de lesões pulmonares, observando suas características de desempenho robustas, com métricas de sensibilidade superiores a 0,90. Coincidindo com os nossos resultados, o desempenho do U-Net++ é influenciado adversamente pela baixa relação sinal/ruído prevalente em imagens de AVC isquêmico, particularmente em avaliações sem contraste intravenoso, em que a diferenciação de densidade entre tecido saudável e isquêmico se torna cada vez mais sutil.

A U-Net++ se distingue por seus caminhos densamente conectados entre o codificador e o decodificador, para melhorar a coerência dos mapeamentos semânticos em vários níveis de abstração. Zhou et al. (2018) comprovaram que essa metodologia aumenta a precisão da segmentação em relação à U-Net original, particularmente em cenários caracterizados por uma variabilidade substancial na morfologia da lesão. No entanto, neste estudo, apesar das melhorias observadas em comparação com o U-Net original, o U-Net++ exibiu tendências de leve sobreajuste durante as últimas épocas do treinamento (conforme evidenciado pela dinâmica da curva de validação) e uma propensão à supersegmentação em certas instâncias de imagem. Essas observações são corroboradas por Wang et al. (2020), que observaram que o U-Net++ pode experimentar um declínio na especificidade em contextos envolvendo lesões hipodensas e diminutas. Uma inclinação perceptível para a supersegmentação foi registrada em áreas desprovidas de achados patológicos, comprometendo a especificidade.

De acordo com Alom et al. (2019), a integração de conexões recursivas aumenta a profundidade funcional da rede neural sem um aumento significativo dos parâmetros, útil em condições caracterizadas pela escassez de dados. No nosso estudo, o R2U-Net exibiu eficácia moderada dentro da estrutura avaliativa, conforme demonstrado por métricas que incluem um coeficiente Dice de 0,66 e uma IoU de 0,49. Conforme observado por Alom et al. (2019), o R2U-Net atingiu um coeficiente Dice de 0,850 e um IoU de 0,765 quando envolvido na segmentação de estruturas anatômicas cerebrais em conjuntos de dados mais homogêneos que exibem contraste acentuado. As discrepâncias nessas métricas podem ser atribuídas à

complexidade de nosso conjunto de dados, que é derivado de TCSC e abrange uma variabilidade morfológica substancial das lesões isquêmicas.

Nesta investigação, a arquitetura R2U-Net ainda apresentou falsos negativos visíveis em algumas predições, especialmente em lesões muito pequenas. Isso está de acordo com observações feitas por Zhou et al. (2021), que alertam para o desafio das arquiteturas recursivas em captar padrões finos em imagens com ruído ou baixo contraste, como as TC sem contraste. A diferença fundamenta a observação feita por Zhou et al. (2021): redes recursivas, apesar de sua maior profundidade em relação à capacidade representacional, exibem maior sensibilidade a ruídos e variações na densidade, afetando adversamente o desempenho em contextos exigentes

A integração dos mecanismos de atenção na arquitetura U-Net resultou em resultados frutíferos na melhoria da precisão de segmentação de lesões pequenas, conforme delineado por Oktay et al. (2018). Nossos resultados corroboram esse efeito benéfico, evidenciado por uma elevação nos coeficientes de Dice e IoU, particularmente em relação a lesões que apresentam desafios consideráveis de detecção. A capacidade do modelo de se concentrar em regiões mais relevantes se mostra particularmente vantajosa em imagens caracterizadas por contraste heterogêneo, conforme demonstrado por Yang et al. (2021), que notaram um desempenho superior da Attention U-Net em relação à U-Net convencional em tarefas de segmentação cerebral. No entanto, o modelo exibiu durações de treinamento prolongadas e maior utilização da memória.

O Attention U-Net exibiu métricas de desempenho louváveis, alcançando uma pontuação média de 0,62 no Dice, um IoU de 0,45 e uma acurácia de 0,98. Embora essas métricas sejam inferiores as arquiteturas mais modernas (ResUnet-a e Swin U-Net), conforme observado em nossa estrutura experimental, elas acompanham os resultados médios de arquiteturas fundamentais, como a U-Net tradicional. Oktay et al. (2018) documentaram coeficientes de Dice superiores a 0,80 em conjuntos de dados multimodais, enquanto Jeniba et al. (2022) alcançaram coeficientes de Dice flutuando entre 0,70 e 0,85 no escopo dos esforços de segmentação pulmonar, significando assim uma correspondência anatômica substancial.

O V-Net passou por modificações no presente estudo para facilitar sua aplicação na avaliação bidimensional, uma modificação considerada essencial à luz dos atributos do conjunto de dados (MILLETARI et al., 2016). O desempenho desse modelo foi o mais alto entre os avaliados, resultando em um coeficiente de Dice de 0,68 e uma IoU de 0,52. No contexto de nossa investigação, os dados bidimensionais resultaram na perda de uma parte de suas vantagens distintas da arquitetura. Gibson et al. (2018) indicaram anteriormente que a

adaptação direta da V-Net a uma estrutura bidimensional leva a uma diminuição da capacidade de compreensão contextual e a uma diminuição na estabilidade do aprendizado. A configuração volumétrica da V-Net é inerentemente dependente de relações espaciais entre camadas, que não estavam presentes na configuração experimental atual. Além disso, é importante reconhecer que um número considerável de parâmetros pode levar ao *overfitting*, especialmente no contexto de conjuntos de dados restritos. Essas restrições também foram documentadas na literatura, incluindo o estudo conduzido por Mehta et al. (2020), que defende modificações na V-Net antes de sua implantação em contextos bidimensionais.

A implantação da arquitetura ResUnet-a demonstrou maior proficiência de generalização e configurações morfológicamente confiáveis em lesões significativas, fundamentando assim as descobertas de Ibtehaz & Rahman (2020), que enfatizam a robustez da arquitetura em enfrentar os desafios impostos por distribuições de classes desequilibradas. No entanto, existe uma acentuada deterioração na eficácia em relação à segmentação de lesões diminutas, um fenômeno similarmente articulado por Islam et al. (2021), que defendem a incorporação de mecanismos de atenção para aumentar a sensibilidade nesta situação. Dentro do escopo de nosso *dataset*, o ResUnet-a demonstrou ainda uma variabilidade reduzida nas métricas de validação ao longo do tempo, o que pode significar maior robustez às flutuações de contraste, embora acompanhada por resultados preditivos mais conservadores.

O ResUnet-a surgiu como a arquitetura que alcançou os resultados favoráveis em nossa investigação, alcançando um coeficiente médio de Dice de 0,6699, uma IoU de 0,5036, uma acurácia de 0,9878 e um F1-Score de 0,6699. Esse desempenho se correlaciona com estudos anteriores, nos quais Ibtehaz & Rahman (2020) demonstraram as vantagens do ResUnet-a em relação ao U-Net, U-Net++ e Attention U-Net em vários conjuntos de dados, atingindo coeficientes Dice superiores a 0,85 em imagens dermatológicas. Em nosso estudo, a ResUnet-a demonstrou capacidades preditivas superiores, delineando com eficiência áreas extensas e mantendo detalhes finos. A integração de blocos residuais combinados com *atrous convolutions* e *multi-attention* aumenta a adaptabilidade do modelo às variações do AVC isquêmico na TC. A eficácia aumentada destaca o potencial significativo da inteligência artificial em aplicações clínicas.

A arquitetura Swin U-Net exibiu uma eficácia considerável na segmentação de lesões isquêmicas no TCSC, alcançando métricas de validação superiores a 0,60 para Dice e melhorando progressivamente os valores de IoU durante o treinamento. Essas descobertas corroboram as de Cao et al. (2023), que notaram melhorias na segmentação do tecido cerebral por meio de *transformers*, particularmente para morfologias irregulares. No entanto, este estudo

detectou uma propensão limítrofe ao *overfitting* após a 30ª época, juntamente com a instabilidade da curva de validação, ecoando as observações de Zhang et al. (2022) sobre a necessidade de ajuste fino e estratégias de atenção em várias escalas para manter a consistência em imagens clínicas com ruídos. As visualizações de previsão demonstraram uma precisão substancial para lesões de médio a grande porte; no entanto, os problemas permaneceram com falsos positivos e lesões pequenas.

No nosso estudo, a Swin U-Net alcançou um Dice de 0,69 e um IoU de 0,51, enquanto Cao et al. (2023) relataram um Dice de 0,88 para segmentação abdominal, e Tang et al. (2022) se aproximaram-se de 0,75 para detecção de lesões cerebrais isquêmicas na TC. Nossos resultados foram influenciados pelas características do conjunto de dados, como limitações de contraste e variabilidade da lesão. Porém, o modelo demonstrou uma baixa taxa de falsos positivos, indicando sua capacidade de gerenciar dependências espaciais de longo alcance típicas de arquiteturas de *transformers*.

Na Tabela 9, observamos um apanhado de todas os valores médios das métricas avaliadas na implementação de cada arquitetura. Nesta tabela, destacamos (negrito) aquelas que obtiveram os melhores valores médios entre os demais.

Tabela 9 – Valores médios das métricas nas diferentes arquiteturas

Arquitetura	Dice	IoU	Acur	Prec	Sens	Esp	F1	AUC	AP
V-Net	0,6897	0,5264	0,9888	0,7736	0,6222	0,9963	0,6897	0,9838	0,7502
ResUNet-a	0,6699	0,5036	0,9878	0,7293	0,6194	0,9953	0,6699	0,9835	0,7314
R2U-Net	0,6639	0,4969	0,9884	0,7894	0,5729	0,9969	0,6639	0,9865	0,7467
Swin U-Net	0,6290	0,4588	0,9854	0,6415	0,6170	0,9929	0,6290	0,9825	0,6826
U-Net	0,6352	0,4654	0,9873	0,7527	0,5494	0,9963	0,6352	0,9808	0,6938
U-Net++	0,6316	0,4615	0,9873	0,7542	0,5432	0,9964	0,6316	0,9807	0,6987
Attention U-Net	0,6230	0,4524	0,9861	0,6819	0,5735	0,9945	0,6230	0,9802	0,6930

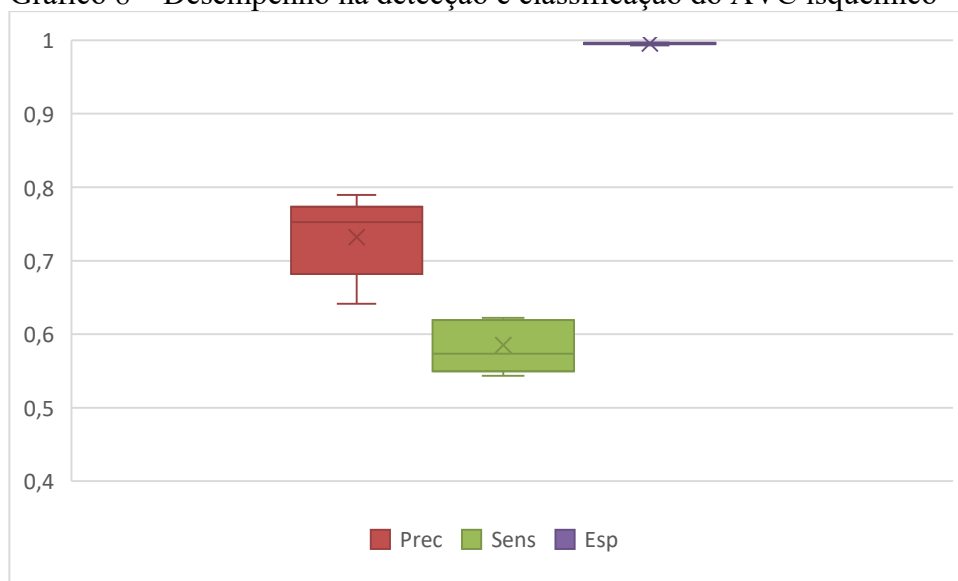
Fonte: Elaborado pelo autor

Apesar dos valores elevados de acurácia em todas as arquiteturas avaliadas, essa métrica deve ser interpretada com cautela em tarefas de segmentação médica com classes desbalanceadas (lesão vs. fundo). Aqui, a alta acurácia reflete mais o número elevado de verdadeiros negativos (fundo corretamente identificado) do que a real capacidade de identificar lesões.

A média de precisão das arquiteturas gira em torno de 0,75, com uma dispersão entre 0,65 e 0,79. Isso indica que a maioria dos modelos teve boa capacidade de evitar falsos positivos (prever lesão onde não há). A sensibilidade teve uma média inferior, próxima de 0,58, com uma

faixa que varia de 0,54 a 0,63. Isso é preocupante em contexto clínico, pois mostra baixa capacidade de detectar todos os verdadeiros positivos (lesões reais), o que pode comprometer a tomada de decisão terapêutica. A especificidade foi consistentemente alta para todas as arquiteturas, com média e dispersão mínima próxima de 0,99 a 1,0. Essa estabilidade indica que os modelos foram altamente eficazes em evitar rotular tecidos saudáveis como lesão, o que é desejável, embora não deva ser priorizado isoladamente em casos de AVC (Gráfico 8).

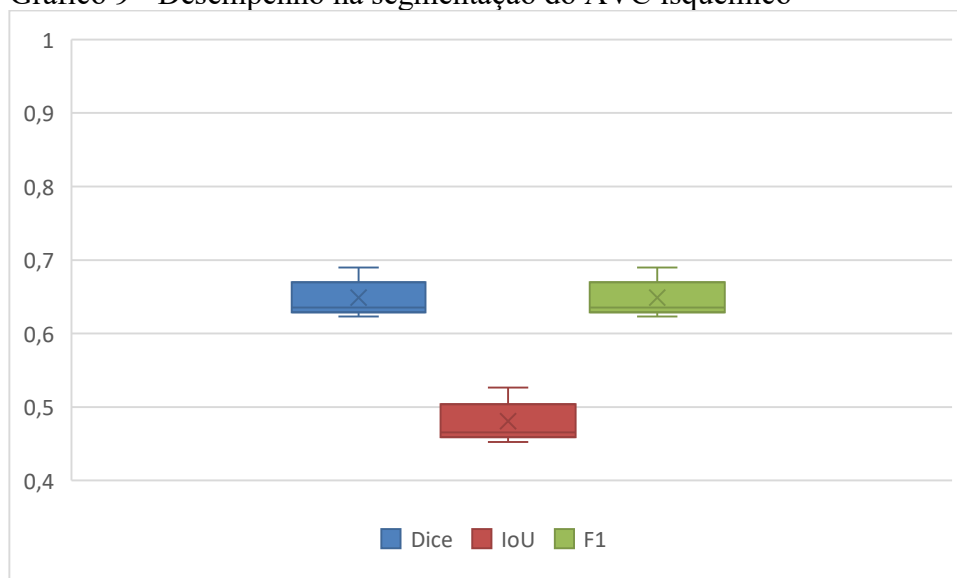
Gráfico 8 – Desempenho na detecção e classificação do AVC isquêmico



Fonte: Elaborado pelo autor

O valor do Dice obteve média em torno de 0,66, com valores oscilando de 0,63 a 0,69. É uma métrica crítica no contexto do AVC, pois indica a qualidade da sobreposição entre as máscaras previstas e as reais. Os modelos apresentaram desempenho razoável, com algumas arquiteturas se destacando (ex: V-Net). A IoU obteve um desempenho mais modesto, com médias entre 0,46 a 0,53. Isso evidencia certa dificuldade dos modelos em delimitar precisamente a área lesada, sendo essa métrica mais penalizadora do que o Dice. Já a *F1-Score* atingiu resultados muito semelhantes ao Dice (o que é esperado, dada sua definição matemática relacionada), com ligeiras variações. O *F1-Score* complementa a análise da sensibilidade e precisão, reforçando a consistência geral dos modelos (Gráfico 9).

Gráfico 9 - Desempenho na segmentação do AVC isquêmico



Fonte: Elaborado pelo autor

Na prática clínica, o desempenho da alta especificidade alcançada entre as arquiteturas utilizadas neste estudo, permite a ocorrência de falhas em detectar uma lesão verdadeira. Isso pode levar à perda da janela terapêutica (ex: trombólise até 4,5h), ao não encaminhamento para centros de referência ou até a alta indevida do paciente. Deste modo, os falsos negativos tornam-se mais graves que os falsos positivos. Estes últimos, apesar de indesejáveis, podem ser filtrados utilizando etapas adicionais como: avaliação por neurologistas, exame físico neurológico detalhado, protocolos clínicos e de imagem complementares.

Com base no exposto, é possível e indicado promover ajustes na metodologia para obter maior sensibilidade clínica, mesmo com uma leve perda de especificidade. Estratégias como *Tversky Loss* com $\beta > \alpha$ (ajustando pesos relativos entre FN e FP), *augmentation* específico para lesões pequenas e *oversampling* balanceado podem ser empregadas para aumentar a sensibilidade e equilibrar os modelos.

6 CONCLUSÃO E PERSPECTIVAS FUTURAS

O estudo demonstrou que arquiteturas modernas de *deep learning* são capazes de segmentar infartos isquêmicos em tomografias sem contraste com desempenho apto a aplicações clínicas reais. Entre os modelos testados, a V-Net destacou-se como a solução mais

eficaz para integração em cenários clínicos, seguida de perto pela ResUnet-a e pela R2U-Net, que também apresentaram métricas robustas. Mesmo arquiteturas mais simples, como a U-Net, mantiveram relevância como referência de comparação, reforçando sua utilidade em ambientes com recursos limitados.

A análise revelou não apenas diferenças técnicas entre as arquiteturas, mas também suas implicações clínicas: a redução de falsos negativos, observada especialmente na ResUnet-a e na R2U-Net, é um fator crítico em contextos de urgência, sugerindo que escolhas arquiteturais não devem ser guiadas apenas por métricas globais, mas também por sua aplicabilidade prática. Além disso, observou-se que a Swin U-Net apresentou estabilidade de aprendizado, mas careceu de consistência em lesões pequenas, reforçando a necessidade de ajustar estratégias de atenção em escalas variadas.

Para trabalhos futuros, algumas direções se mostram promissoras e realistas:

1. Otimização de hiperparâmetros com abordagens automáticas (p.ex., *Bayesian Optimization*) para refinar taxas de aprendizado e tamanhos de *batch*.
2. Funções de perda especializadas, como *Focal Tversky Loss*, mais adequadas a cenários com forte desequilíbrio de classes.
3. Técnicas avançadas de aumento de dados, incluindo GANs, para melhorar a generalização em conjuntos de dados reduzidos.
4. Modelagem volumétrica 3D, explorando arquiteturas como nnU-Net e variantes da V-Net para capturar relações espaciais completas.
5. Validação multicêntrica, ampliando a diversidade dos dados e garantindo maior robustez e aplicabilidade clínica.

Em síntese, este trabalho aponta a viabilidade do uso de redes neurais profundas no suporte ao diagnóstico de AVC isquêmico, ao mesmo tempo em que estabelece um ponto de partida para avanços futuros que conciliem precisão técnica e relevância clínica.

REFERÊNCIAS

AHMED, H. K.; MATHISEN, S. M.; KURZ, K.; DALEN, I.; LOGALLO, N.; THOMASSEN, L.; KURZ, M. **Thrombolysis in wake-up stroke based on MRI mismatch**. *Journal of Neurological Sciences*, v. 466, p. 123265, 15 nov. 2024.

ALOM, M. Z.; HASAN, M.; YAKOPCIC, C.; TAHA, T. M.; ASARI, V. K. **Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation**. arXiv preprint arXiv:1802.06955, 2018.

AMERICAN STROKE ASSOCIATION. **Ischemic Stroke: Pathophysiology and Treatment**. *Stroke*, v. 54, n. 1, p. 12-22, 2023.

AOKI, J. A.; MADIGAN, N. N.; HILL, M. D.; et al. **DWI-FLAIR mismatch guided thrombolysis in patients with unknown onset stroke: real-world data**. *SAGE Journals*, v. 12, 2021.

BENGIO, Y. **Deep Learning of Representations: Looking Forward**. arXiv, 2013.

BORGEFORS, G. **Distance transformations in digital images**. *Computer Vision, Graphics, and Image Processing*, v. 34, p. 344-371, 1986.

CAO, H.; WANG, Y.; CHEN, J.; et al. **Swin-Unet: Unet-like pure transformer for medical image segmentation**. arXiv preprint arXiv:2105.05537, 2021.

CHEN, L.; PAPANDREOU, G.; KOKKINOS, I.; et al. **Deeplab: Semantic image segmentation with deep convolutional nets**. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, v. 40, n. 4, p. 834-848, 2017a.

CHEN, L.; PAPANDREOU, G.; SCHROFF, F.; ADAM, H. **Rethinking atrous convolution for semantic image segmentation**. CoRR abs/1706.05587, 2017b.

DIAKOGIANNIS, F. I.; WALDNER, F.; CACCETTA, P.; WU, C. **ResUNet-a: A deep learning framework for semantic segmentation of remotely sensed data**. *ISPRS Journal of Photogrammetry and Remote Sensing*, v. 162, p. 94-114, 2020.

DOI, K. **Computer-aided diagnosis in medical imaging: historical review, current status and future potential.** Computerized Medical Imaging and Graphics, v. 31, n. 4-5, p. 198-211, jun./jul. 2007.hashi

DOSOVITSKIY, A.; BEYER, L.; KOLESNIKOV, A.; et al. **An image is worth 16x16 words: Transformers for image recognition at scale.** arXiv preprint arXiv:2010.11929, 2021

ERUS, G.; ZACHARAKI, E. I.; DAVATZIKOS, C. **Individualized statistical learning from medical image databases: application to identification of brain lesions.** Medical Image Analysis, v. 18, p. 542–554, 2014.

FAN, J.; LI, X.; YU, X.; et al. **Global Burden, Risk Factor Analysis, and Prediction Study of Ischemic Stroke, 1990-2030.** Neurology, v. 101, n. 2, p. e137-e150, 2023. DOI: 10.1212/WNL.0000000000207387.

GAO, Y.; GERAS, K.; LEWIN, A.; MOY, L. **New Frontiers: An Update on Computer-Aided Diagnosis for Breast Imaging in the Age of Artificial Intelligence.** American Journal of Roentgenology, v. 212, n. 2, p. 300-307, fev. 2019.

GIBSON, E.; GIGANTI, F.; HU, Y.; BONMATI, E.; et al. **Automatic Multi-Organ Segmentation on Abdominal CT with Dense V-Networks.** IEEE Transactions on Medical Imaging, v. 37, n. 8, p. 1822–1834, 2018.

GIGER, M. L.; DOI, K.; MACMAHON, H. **Image feature analysis and computer-aided diagnosis in digital radiography. 3. Automated detection of nodules in peripheral lung fields.** Medical Physics, v. 15, n. 2, p. 158-166, mar./abr. 1988.

GÓMEZ, S.; et al. **An attentional U-Net with an auxiliary class learning to support acute ischemic stroke segmentation on CT.** In: MEDICAL IMAGING 2023: IMAGE PROCESSING, 2023. Proceedings [...]. Bellingham: SPIE, 2023. v. 12464.

GREENSPAN, H.; GINNEKEN, B. V.; SUMMERS, R. M. **Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique.** IEEE Transactions on Medical Imaging, v. 35, n. 5, p. 1153–1159, 2016.

HE, K.; ZHANG, X.; REN, S.; et al. **Deep residual learning for image recognition**. In: IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR 2016), 2016, Las Vegas. Proceedings [...]. Piscataway: IEEE, 2016. p. 770-778.

HE, K.; ZHANG, X.; REN, S.; SUN, J. **Spatial pyramid pooling in deep convolutional networks for visual recognition**. CoRR abs/1406.4729, 2014.

IBTEHAZ, N.; RAHMAN, M. S. **MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation**. Neural Networks, v. 121, p. 74–87, jan. 2020

ISLAM, M.; VAIDYANATHAN, N. R.; JOSE, V. J. M.; REN, H. **Ischemic Stroke Lesion Segmentation Using Adversarial Learning**. *arXiv preprint arXiv:2204.04993*, 2022

JENIBA, J. S.; MILTON, A. **A multilevel self-attention based segmentation and classification technique using Directional Hexagonal Mixed Pattern algorithm for lung nodule detection in thoracic CT image**. *International Journal of Imaging Systems and Technology*, v. 32, n. 5, p. 1496–1510, 2022.

JEON, W.; KO, G.; LEE, J.; LEE, H.; HA, D.; RO, W. W. **Deep learning with GPUs**. In: Advances In Computers, v. 122, p. 167-215, 2021.

JOHANSEN, J.; OFFERSEN, C. M.; CARLSEN, J. F.; et al. **An automatic DWI/FLAIR mismatch assessment of stroke patients**. *Diagnostics*, v. 14, n. 1, art. 69, 2024.

JUNIOR, A. F. R. S.; SANTOS, G. C. D.; KANETO, C. M.; et al. **Hospital service for ischemic stroke patients in Brazilian countryside: are we still in the '80s?**. *Arquivos de Neuro-Psiquiatria*, v. 80, n. 8, p. 770–778, 2022.

KIM, M.; YUN, J.; CHO, Y.; SHIN, K.; JANG, R.; BAE, H. J.; KIM, N. **Deep Learning in Medical Imaging**. *Neurospine*, v. 16, n. 4, p. 657-668, dez. 2019. DOI: 10.14245/ns.1938396.198. Epub 31 dez. 2019. Erratum em: *Neurospine*, v. 17, n. 2, p. 471-472, jun. 2020.

KOÇ, U.; SEZER, E. A.; ÖZKAYA, Y. A.; et al. **Artificial intelligence in healthcare competition (TEKNOFEST-2021): Stroke data set**. *Eurasian Journal of Medicine*, v. 54, n. 3, p. 248-258, 2022.

KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. **Imagenet classification with deep convolutional neural networks.** In: ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS, 2012, Cambridge. Proceedings [...]. Cambridge: MIT Press, 2012. p. 1097-1105.

LEDLEY, R. S.; LUSTED, L. B. **Reasoning Foundations of Medical Diagnosis.** Science, v. 130, p. 9-21, 1959.

LIU, Z.; LIN, Y.; CAO, Y.; HU, H.; et al. **Swin transformer: Hierarchical vision transformer using shifted Windows.** CoRR abs/2103.14030, 2021.

LODWICK, G. S.; HAUN, C. L.; SMITH, W. E.; et al. **Computer diagnosis of primary bone tumor.** Radiology, v. 80, p. 273–275, 1963.

MARTINS, S. C.; et al. **Past, present, and future of stroke in middle-income countries: The Brazilian experience.** International Journal of Stroke, v. 8, suppl. A100, p. 106-111, 2013.

MATSUYAMA, T. **Expert systems for image processing: knowledge-based composition of image analysis processes.** Computer Vision, Graphics and Image Processing, v. 48, p. 22-49, 1989.

MEHTA, S.; MERCAN, E.; BARTLETT, J.; et al. **Y-Net: Joint segmentation and classification for diagnosis of breast biopsy images.** *arXiv preprint arXiv:1806.01313*, 2018.

MILLETARI, F.; NAVAB, N.; AHMED, S.-A. **V-Net: Fully convolutional neural networks for volumetric medical image segmentation.** International Conference on 3D Vision (3DV), p. 565-571, 2016.

MOHAPATRA, S.; LEE, T. H.; SAHOO, P. K.; et al. **Localization of early infarction on non-contrast CT images in acute ischemic stroke with deep learning approach.** Scientific Reports, v. 13, n. 19442, 2023.

MYERS, P. H.; NICE, C. M.; BECKER, H. C.; et al. **Automated computer analysis of radiographic images.** Radiology, v. 83, p. 1029–1033, 1964.

NISHI, H.; ISHII, A.; TSUJI, H.; et al. **Automatic Ischemic Core Estimation Based on Noncontrast-Enhanced Computed Tomography.** Stroke, v. 54, n. 7, p. 1815-1822, jul. 2023.

OKTAY, O.; SCHLEMPER, J.; LE FOLGOC, L.; et al. **Attention U-Net: Learning Where to Look for the Pancreas**. arXiv preprint arXiv:1804.03999, 2018.

OPENAI. **ChatGPT: modelo de linguagem baseado em inteligência artificial**. [S.l.]: OpenAI, 2025. Disponível em: <https://chat.openai.com/>. Acesso em: 7 jul. 2025.

PARK, H.; BLAND, P. H.; MEYER, C. R. **Construction of an abdominal probabilistic atlas and its application in segmentation**. IEEE Transactions on Medical Imaging, v. 22, p. 483–492, 2003.

PEXMAN, J. H.; et al. **Use of the Alberta Stroke Program Early CT Score (ASPECTS) for assessing CT scans in patients with acute stroke**. American Journal of Neuroradiology, v. 22, n. 8, p. 1534–1542, set. 2001.

PINHEIRO, W. M.; et al. **Perfil epidemiológico dos pacientes internados por acidente vascular cerebral no Brasil nos últimos cinco anos**. Revista Ibero-Americana de Humanidades, Ciências e Educação, v. 10, n. 9, p. 3201–3211, 2024.

PNEV, S.; TULUPOV, A.; BERIKOV, V. **Modified U-Net with different attention mechanisms for acute ischemic stroke segmentation using non-contrast CT**. In: INTERNATIONAL CONFERENCE ON ..., 2021. Proceedings [...]. p. 133–136, 2021.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-Net: Convolutional networks for biomedical image segmentation**. MICCAI – Medical Image Computing and Computer-Assisted Intervention, p. 234-241, 2015.

SAHOO, P. K.; MOHAPATRA, S.; WU, C. Y.; et al. **Automatic identification of early ischemic lesions on non-contrast CT with deep learning approach**. Scientific Reports, v. 12, n. 18054, 2022.

SANTOS, J. V.; LEOPOLDINO, D. de J. dos S.; SILVA, A. B. B.; et al. **Acidente Vascular Cerebral no Brasil: aspectos epidemiológicos da mortalidade no período de 2019 a 2023**. Brazilian Journal of Implantology and Health Sciences, v. 7, n. 3, p. 1429–1439, 2025.

STANSFIELD, S. A. **ANGY: a rule-based expert system for automatic segmentation of coronary vessels from digital subtracted angiograms**. IEEE Transactions on Pattern Analysis and Machine Intelligence, v. 2, p. 188–199, 1986.

TANG, Y.; YANG, D.; LI, W.; et al. **Self-Supervised Pre-Training of Swin Transformers for 3D Medical Image Analysis**. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022

TAKÁCS, T. T.; MAGYAR-STANG, R.; SZATMÁRI, S.; et al. **Workload and clinical impact of MRI-based extension of reperfusion therapy time window in acute ischaemic stroke**. GeroScience, 2025.

TAKAHASHI, R.; KAJIKAWA, Y. **Computer-aided diagnosis: A survey with bibliometric analysis**. International Journal of Medical Informatics, v. 101, p. 58-67, maio 2017.

THOMALLA, G.; SIMONSEN, C. Z.; BOUTITIE, F.; et al. **DWI FLAIR mismatch for the identification of patients with acute ischemic stroke within 4.5 h (PRE FLAIR study)**. Lancet Neurology, v. 18, n. 6, p. 475–481, 2019.

VAN LEEMPUT, K.; MAES, F.; VANDERMEULEN, D.; COLCHESTER, A.; SUETENS, P. **Automated segmentation of multiple sclerosis lesions by model outlier detection**. IEEE Transactions on Medical Imaging, v. 20, p. 677–688, 2001.

VAN POPPEL, L. M.; MAJOIE, C. B. L. M.; MARQUERING, H. A.; EMMER, B. J. **Associations between early ischemic signs on non-contrast CT and time since acute ischemic stroke onset: a scoping review**. European Journal of Radiology, v. 155, p. 110455, 2022.

VIEIRA, I. P.; et al. **Funcionalidade e qualidade de vida em pacientes pós acidente vascular cerebral**. Brazilian Journal of Development, v. 6, n. 4, p. 17391-17403, 2020.

VIERGEVER, M. A.; MAINTZ, J. B. A.; KLEIN, S.; MURPHY, K.; STARING, M.; PLUIM, J. P. W. **A survey of medical image registration – under review**. Medical Image Analysis, v. 33, p. 140–144, 2016.

VIRANI, A. A.; et al. **Heart disease and stroke statistics – 2020 update: A report from the American Heart Association**. Circulation, v. 141, n. 9, p. e139-e596, 2020.

WANG, H.; ZHOU, Z.; LI, Y.; et al. **Comparison of machine learning methods for classifying mediastinal lymph node metastasis of non-small cell lung cancer from 18F-FDG PET/CT images**. EJNMMI Research, v. 7, p. 11, 2017.

WARFIELD, S. K.; ZOU, K. H.; WELLS, W. M. **Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation.** 2004.

WORLD HEALTH ORGANIZATION. **Global Health Estimates 2023: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2023.** Disponível em: <https://www.who.int>.

ZHANG, N.; MA, S.; LI, X.; ZHANG, J.; et al. **Multi-scale Neighborhood Attention Transformer on U-Net for Medical Image Segmentation.** In: **IEEE INTERNATIONAL CONFERENCE ON BIOINFORMATICS AND BIOMEDICINE (BIBM)**, 2022. Piscataway, NJ: IEEE, 2022

ZHAO, H.; SHI, J.; QI, X.; WANG, X.; JIA, J. **Pyramid scene parsing network.** In: CVPR, 2017.

ZHOU, Z.; RAHMAN SIDDIQUEE, M. M.; TAJBAKHSI, N.; LIANG, J. **UNet++: A Nested U-Net Architecture for Medical Image Segmentation.** In: STOYANOV, D.; et al. (org.). *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. DLMIA ML-CDS 2018.* Lecture Notes in Computer Science, v. 11045. Cham: Springer, 2018.

APÊNDICE – Notebook

```

pip install tensorflow keras keras-unet-collection scikit-image scikit-learn opencv-python matplotlib
pandas
# =====
# Comparativo: U-Net variants (keras_unet_collection)
# - unet_2d, unet_plus_2d, r2_unet_2d, resnet_a_2d
# - att_unet_2d (Attention U-Net)
# - vnet_2d (2D adaptado p/ 1 canal Sigmoid)
# Dataset: TC crânio (grayscale), máscara binária (0/1)
# =====
import os, glob, cv2, random
import numpy as np
import tensorflow as tf
from tensorflow.keras import backend as K
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping, ReduceLROnPlateau
import matplotlib.pyplot as plt
from sklearn.model_selection import train_test_split
from sklearn.metrics import (
    confusion_matrix, roc_auc_score, average_precision_score,
    precision_score, recall_score, f1_score
)
import pandas as pd
from keras_unet_collection import models

# -----
# Configs gerais
# -----
SEED = 42
np.random.seed(SEED); tf.random.set_seed(SEED)

gpus = tf.config.list_physical_devices("GPU")
for g in gpus:
    try: tf.config.experimental.set_memory_growth(g, True)
    except: pass

IMG_SIZE = (256, 256)
INPUT_SHAPE = (256, 256, 1)
BATCH_SIZE = 2
EPOCHS = 50
LR = 1e-4

PATH_IMG = r"C:\Users\Daniel\Documents\Brain_Stroke_CT_Dataset\Ischemia\PNG"
PATH_MASK = r"C:\Users\Daniel\Documents\Brain_Stroke_CT_Dataset\Ischemia\MASKS"

os.makedirs("Models", exist_ok=True)

# -----
# Utils: parear, carregar, tf.data
# -----
def build_pairs(path_img, path_mask):
    exts = ("*.png", "*.jpg", "*.jpeg", "*.bmp", "*.tif", "*.tiff")
    all_imgs = []
    for e in exts: all_imgs += glob.glob(os.path.join(path_img, e))
    all_imgs = sorted(all_imgs)
    img_files, mask_files = [], []

```

```

for f in all_imgs:
    stem = os.path.splitext(os.path.basename(f))[0]
    found = None
    for e in exts:
        cand = glob.glob(os.path.join(path_mask, stem + e[1:]))
        if cand: found = cand[0]; break
    if not found:
        cand = glob.glob(os.path.join(path_mask, stem + ".*"))
        if cand: found = cand[0]
    if found:
        img_files.append(f); mask_files.append(found)
return img_files, mask_files

def load_and_preprocess(img_paths, mask_paths, img_size):
    X, Y = [], []
    for ip, mp in zip(img_paths, mask_paths):
        im = cv2.imread(ip, cv2.IMREAD_GRAYSCALE)
        ms = cv2.imread(mp, cv2.IMREAD_GRAYSCALE)
        if im is None or ms is None: continue
        im = cv2.resize(im, img_size, interpolation=cv2.INTER_AREA)
        ms = cv2.resize(ms, img_size, interpolation=cv2.INTER_NEAREST)
        im = (im.astype(np.float32)/255.0)[..., None]
        m_bin = (ms > 0).astype(np.float32)[..., None]
        X.append(im); Y.append(m_bin)
    return np.array(X, np.float32), np.array(Y, np.float32)

def tf_augment(img, mask):
    if tf.random.uniform() > 0.5:
        img = tf.image.flip_left_right(img)
        mask = tf.image.flip_left_right(mask)
    if tf.random.uniform() > 0.5:
        img = tf.image.flip_up_down(img)
        mask = tf.image.flip_up_down(mask)
    k = tf.random.uniform(), minval=0, maxval=4, dtype=tf.int32)
    img = tf.image.rot90(img, k)
    mask = tf.image.rot90(mask, k)
    return img, mask

def make_ds(X, Y, batch_size, augment=False, shuffle=True):
    ds = tf.data.Dataset.from_tensor_slices((X, Y))
    if shuffle: ds = ds.shuffle(len(X), seed=SEED, reshuffle_each_iteration=True)
    if augment: ds = ds.map(tf_augment, num_parallel_calls=tf.data.AUTOTUNE)
    return ds.batch(batch_size).prefetch(tf.data.AUTOTUNE)

# -----
# Métricas/Loss binárias
# -----
def dice_coef(y_true, y_pred, smooth=1e-6):
    y_true_f = K.flatten(tf.cast(y_true, tf.float32))
    y_pred_f = K.flatten(tf.cast(y_pred, tf.float32))
    inter = K.sum(y_true_f * y_pred_f)
    return (2.*inter + smooth) / (K.sum(y_true_f) + K.sum(y_pred_f) + smooth)

def iou_coef(y_true, y_pred, smooth=1e-6):
    y_true_f = K.flatten(tf.cast(y_true, tf.float32))
    y_pred_f = K.flatten(tf.cast(y_pred, tf.float32))
    inter = K.sum(y_true_f * y_pred_f)
    union = K.sum(y_true_f) + K.sum(y_pred_f) - inter
    return (inter + smooth) / (union + smooth)

```

```

def dice_loss(y_true, y_pred): return 1.0 - dice_coef(y_true, y_pred)

def bce_dice_loss(y_true, y_pred, bce_w=1.0, dice_w=1.0):
    bce = tf.keras.losses.binary_crossentropy(y_true, y_pred)
    dl = dice_loss(y_true, y_pred)
    return bce_w*bce + dice_w*dl

# -----
# Construtores dos 6 modelos
# (Atenção: nomes de parâmetros podem variar por versão da lib.
# Se aparecer erro de assinatura, me mande a mensagem que eu ajusto.)
# -----
def build_unet_2d():
    return models.unet_2d(
        input_size=INPUT_SHAPE,
        filter_num=[64, 128, 256, 512],
        n_labels=1,
        stack_num_down=2,
        stack_num_up=2,
        activation='ReLU',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,
        unpool='bilinear',
        name='unet2d'
    )

def build_unet_plus_2d():
    return models.unet_plus_2d(
        input_size=INPUT_SHAPE,
        filter_num=[64, 128, 256, 512],
        n_labels=1,
        stack_num_down=2,
        stack_num_up=2,
        activation='ReLU',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,
        unpool='bilinear',
        deep_supervision=False,
        name='unetpp'
    )

def build_r2_unet_2d():
    return models.r2_unet_2d(
        input_size=INPUT_SHAPE,
        filter_num=[64, 128, 256, 512],
        n_labels=1,
        stack_num_down=2,
        stack_num_up=2,
        activation='ReLU',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,
        unpool='bilinear',
        recur_num=2, # se der erro, troque para re_num=2
        name='r2unet'
    )

```

```

def build_resunet_a_2d():
    return models.resunet_a_2d(
        input_size=INPUT_SHAPE,
        filter_num=[64, 128, 256],    # depth=3
        dilation_num=[1, 2, 4],      # << ADICIONADO
        n_labels=1,
        activation='ReLU',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,
        unpool='bilinear',
        name='resuneta'
    )

def build_att_unet_2d():
    return models.att_unet_2d(
        input_size=INPUT_SHAPE,
        filter_num=[64, 128, 256, 512],
        n_labels=1,
        stack_num_down=2,
        stack_num_up=2,
        activation='ReLU',
        atten_activation='ReLU',
        attention='add',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,
        unpool='bilinear',
        name='attunet'
    )

def build_vnet_2d():
    # V-Net original usa PReLU; aqui mantemos PReLU, mas com saída 1 canal Sigmoid para comparação justa.
    return models.vnet_2d(
        input_size=INPUT_SHAPE,
        filter_num=[16, 32, 64, 128, 256],
        n_labels=1,                # << 1 canal (binário)
        res_num_ini=1, res_num_max=3,
        activation='PReLU',
        output_activation='Sigmoid',
        batch_norm=True,
        pool=False,                # down por conv stride
        unpool=False,              # up por conv transpose
        name='vnet'
    )

def build_swin_unet_2d():
    return models.swin_unet_2d(
        input_size=INPUT_SHAPE,
        filter_num_begin=64,      # número inicial de filtros
        n_labels=1,                # binário (1 classe)
        depth=4,                  # número de estágios do encoder/decoder
        stack_num_down=2,         # convs por bloco no encoder
        stack_num_up=2,           # convs por bloco no decoder
        patch_size=(2, 2),        # tamanho do patch de entrada
        num_heads=[4, 8, 16, 32], # cabeças de atenção por estágio
        window_size=(7, 7),       # tamanho da janela de self-attention
    )

```

```

    num_mlp=512,          # tamanho do MLP na atenção
    output_activation='Sigmoid', # saída binária
    shift_window=True,    # janela deslocada (como no paper Swin)
    name='swin_unet'
)

BUILDERS = {
    "resnet_a_2d" : build_resnet_a_2d,
    "vnet_2d"     : build_vnet_2d,
    "att_unet_2d" : build_att_unet_2d,
    "swin_unet_2d" : build_swin_unet_2d,
    "unet_2d"      : build_unet_2d,
    "unet_plus_2d" : build_unet_plus_2d,
    "r2_unet_2d"   : build_r2_unet_2d,
}

# -----
# Treino + avaliação + plots
# -----
def train_and_eval(model_name, X_train, y_train, X_val, y_val, X_test, y_test):
    print(f"\n===== {model_name} =====")
    model = BUILDERS[model_name]()
    model.compile(optimizer=Adam(learning_rate=LR),
                  loss=bce_dice_loss,
                  metrics=['accuracy', dice_coef, iou_coef])
    model.summary()

    ckpt_path = os.path.join("Models", f"{model_name}_best.keras")
    cb = [
        ModelCheckpoint(ckpt_path, monitor="val_loss", save_best_only=True, mode="min"),
        EarlyStopping(monitor="val_loss", patience=10, restore_best_weights=True),
        ReduceLROnPlateau(monitor="val_loss", factor=0.5, patience=4, verbose=1, min_lr=1e-6),
    ]

    train_ds = make_ds(X_train, y_train, BATCH_SIZE, augment=True, shuffle=True)
    val_ds   = make_ds(X_val, y_val, BATCH_SIZE, augment=False, shuffle=False)

    history = model.fit(train_ds, validation_data=val_ds, epochs=EPOCHS, callbacks=cb, verbose=1)

    # Curvas
    plt.figure(figsize=(12,4))
    plt.suptitle(model_name)
    plt.subplot(1,3,1); plt.plot(history.history['loss'], label='train'); plt.plot(history.history['val_loss'],
label='val'); plt.title('Loss'); plt.legend()
    plt.subplot(1,3,2); plt.plot(history.history['dice_coef'], label='train');
plt.plot(history.history['val_dice_coef'], label='val'); plt.title('Dice'); plt.legend()
    plt.subplot(1,3,3); plt.plot(history.history['iou_coef'], label='train');
plt.plot(history.history['val_iou_coef'], label='val'); plt.title('IoU'); plt.legend()
    plt.tight_layout(); plt.show()

    # Teste
    print("📄 Avaliação no TESTE...")
    y_prob = model.predict(X_test, batch_size=BATCH_SIZE, verbose=0) # (N,H,W,1)
    y_pred = (y_prob > 0.5).astype(np.uint8)
    y_true = (y_test > 0.5).astype(np.uint8)

    yt = y_true.flatten().astype(np.uint8)
    yp = y_pred.flatten().astype(np.uint8)
    ypf = y_prob.flatten().astype(np.float32)

```

```

tn, fp, fn, tp = confusion_matrix(yt, yp).ravel()
accuracy = (tp + tn) / (tp + tn + fp + fn + 1e-7)
precision = precision_score(yt, yp, zero_division=0)
sensitivity = recall_score(yt, yp, zero_division=0)
specificity = tn / (tn + fp + 1e-7)
f1 = f1_score(yt, yp, zero_division=0)
try:
    auc = roc_auc_score(yt, ypf)
    ap = average_precision_score(yt, ypf)
except Exception:
    auc, ap = float('nan'), float('nan')

inter = np.sum((yt == 1) & (yp == 1))
dice = (2*inter + 1e-7) / (np.sum(yt) + np.sum(yp) + 1e-7)
union = (np.sum(yt) + np.sum(yp) - inter)
iou = (inter + 1e-7) / (union + 1e-7)

print(f"\n 📊 {model_name} - Métricas no TESTE:")
print(f"Dice = {dice:.4f}")
print(f"IoU = {iou:.4f}")
print(f"Accuracy = {accuracy:.4f}")
print(f"Precision = {precision:.4f}")
print(f"Sensitivity = {sensitivity:.4f}")
print(f"Specificity = {specificity:.4f}")
print(f"F1-Score = {f1:.4f}")
print(f"AUC = {auc:.4f}")
print(f"AP = {ap:.4f}")

# Overlays
def show_examples(num=4):
    num = min(num, len(X_test))
    for i in range(num):
        img = X_test[i, ..., 0]
        gt = y_true[i, ..., 0]
        pr = y_pred[i, ..., 0]
        plt.figure(figsize=(12,4))
        plt.subplot(1,3,1); plt.imshow(img, cmap='gray'); plt.title('Imagem'); plt.axis('off')
        plt.subplot(1,3,2); plt.imshow(gt, cmap='gray'); plt.title('GT'); plt.axis('off')
        plt.subplot(1,3,3); plt.imshow(img, cmap='gray'); plt.imshow(pr, cmap='jet', alpha=0.5);
    plt.title('Pred'); plt.axis('off')
    plt.tight_layout(); plt.show()
print(f"\n 🖼️ Exemplos de predição (overlays):")
show_examples(4)

# Salvar
model.save(os.path.join("Models", f"{model_name}_final.keras"), save_traces=True)
model.save_weights(os.path.join("Models", f"{model_name}_weights.h5"))

return {
    "name": model_name,
    "Dice": dice, "IoU": iou, "Accuracy": accuracy, "Precision": precision,
    "Sensitivity": sensitivity, "Specificity": specificity, "F1": f1, "AUC": auc, "AP": ap
}

# -----
# Carregar dados (uma vez)
# -----
imgs, masks = build_pairs(PATH_IMG, PATH_MASK)

```

```

print(f"🔍 Pares encontrados: {len(imgs)}")
train_img, temp_img, train_msk, temp_msk = train_test_split(imgs, masks, test_size=0.30,
random_state=SEED, shuffle=True)
val_img, test_img, val_msk, test_msk = train_test_split(temp_img, temp_msk, test_size=0.50,
random_state=SEED, shuffle=True)

X_train, y_train = load_and_preprocess(train_img, train_msk, IMG_SIZE)
X_val, y_val = load_and_preprocess(val_img, val_msk, IMG_SIZE)
X_test, y_test = load_and_preprocess(test_img, test_msk, IMG_SIZE)

print("Shapes ->", X_train.shape, y_train.shape, "|", X_val.shape, y_val.shape, "|", X_test.shape, y_test.shape)

# -----
# Execute os 6 modelos
# -----
to_run = ["resnet_a_2d", "vnet_2d", "att_unet_2d", "swin_unet_2d", "unet_2d", "unet_plus_2d",
"r2_unet_2d", "resnet_a_2d"]
results = []
for name in to_run:
    results.append(train_and_eval(name, X_train, y_train, X_val, y_val, X_test, y_test))

# Resumo comparativo
df = pd.DataFrame(results).set_index("name").sort_values("Dice", ascending=False)
print("\n📊 Resultado comparativo (ordenado por Dice):")
print(df.round(4))

```