



UEPB

**UNIVERSIDADE ESTADUAL DA PARAÍBA
CAMPUS I**

**CENTRO DE CIÊNCIAS E TECNOLOGIA
DEPARTAMENTO DE COMPUTAÇÃO
CURSO DE BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

JENNIFFER KELLY RUFINO DE ARAÚJO

**SEGMENTAÇÃO AUTOMÁTICA DO RIM EM IMAGENS DE TOMOGRAFIAS
COMPUTADORIZADAS E ULTRASSONOGRAFIAS UTILIZANDO U-NET**

**CAMPINA GRANDE
2022**

JENNIFFER KELLY RUFINO DE ARAÚJO

**SEGMENTAÇÃO AUTOMÁTICA DO RIM EM IMAGENS DE TOMOGRAFIAS
COMPUTADORIZADAS E ULTRASSONOGRAFIAS UTILIZANDO U-NET**

Trabalho de Conclusão de Curso apresentado ao Departamento do Curso Ciência da Computação da Universidade Estadual da Paraíba, como requisito parcial à obtenção do título de bacharel em Ciência da Computação.

Área de concentração: Inteligência Artificial.

Orientador: Prof. Dr. Robson Pequeno de Sousa.

**CAMPINA GRANDE
2022**

É expressamente proibido a comercialização deste documento, tanto na forma impressa como eletrônica. Sua reprodução total ou parcial é permitida exclusivamente para fins acadêmicos e científicos, desde que na reprodução figure a identificação do autor, título, instituição e ano do trabalho.

A663s Araújo, Jenniffer Kelly Rufino de.

Segmentação automática do rim em imagens de tomografias computadorizadas e ultrassonografias utilizando U-Net [manuscrito] / Jenniffer Kelly Rufino de Araújo. - 2022.

38 p. : il. colorido.

Digitado.

Trabalho de Conclusão de Curso (Graduação em Computação) - Universidade Estadual da Paraíba, Centro de Ciências e Tecnologia, 2022.

"Orientação : Prof. Dr. Robson Pequeno de Sousa, Coordenação do Curso de Computação - CCT."

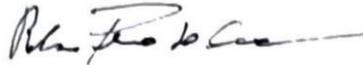
1. Segmentação semântica. 2. Rede Neural Convolucional. 3. U-Net. 4. Função renal. I. Título

21. ed. CDD 005.3

SEGMENTAÇÃO AUTOMÁTICA DO RIM EM IMAGENS DE TOMOGRAFIAS COMPUTADORIZADAS E ULTRASSONOGRAMAS UTILIZANDO U-NET

Trabalho de Conclusão de Curso de Graduação
em Ciência da Computação da Universidade
Estadual da Paraíba, como requisito à obtenção
do título de Bacharel em Ciência da
Computação.

Aprovada em 02 de Dezembro de 2022.



Prof. Dr. Robson Pequeno de Sousa (DC - UEPB)
Orientador(a)



Prof. Dr. Wellington Candeia de Araujo (DC - UEPB)
Examinador(a)



Profa. Dra. Kézia de Vasconcelos de Oliveira Dantas (DC - UEPB)
Examinador(a)

AGRADECIMENTOS

Ao meu pai Lindalberto, a minha avó Lourdes e meu avô Antônio, a minha tia Gildania e meu irmão Francisco, pela compreensão durante todo este período.

A minha mãe (*in memoriam*), embora fisicamente ausente, sentia sua presença ao meu lado, dando-me força.

Ao meu orientador Robson pela ajuda dada durante a realização deste trabalho.

Ao doutorando do NUTES, Alberto, pela presteza e atendimento quando me foi necessário.

À banca examinadora pela aceitação do convite e pelo tempo investido para leitura e avaliação desse trabalho.

Aos colegas de classe pelos momentos de amizade e apoio.

RESUMO

A segmentação do rim é um passo essencial no desenvolvimento de qualquer sistema assistido por computador para avaliação da função renal, todavia, pode ser passível a erros humanos, uma vez que é feito de forma manual e, também, pode ser cansativo, principalmente se for feita em muitas fatias. Em virtude dessa dificuldade, nos últimos anos a inteligência artificial vem obtendo importante papel na segmentação de imagens para inúmeros propósitos e, desfrutando dessas ferramentas, este trabalho faz uso de redes neurais convolucionais, através da arquitetura U-Net, para segmentação de imagens renais. Para alcançar estes objetivos, foi utilizada duas bases de dados, onde uma é referente a tomografias computadorizadas e outra a ultrassonografias. Com base nisso, foram utilizadas técnicas de pré-processamento de imagens para auxiliar na tarefa da segmentação semântica. Por fim, foi avaliada como benéfica à inserção de inteligência artificial na segmentação de imagens, com o modelo alcançando uma métrica de semelhança de Dice para o conjunto de testes de 0,83 para TCs e 0,96 para ultrassonografias.

Palavras-Chave: Segmentação Semântica. Rede Neural Convolucional. U-Net. Rim.

ABSTRACT

Kidney segmentation is an essential step in the development of any computer-assisted system for assessing kidney function, however, it can be subject to human errors, since it is done manually and, also, can be tiring, especially if it is made in many slices. Due to this difficulty, in recent years artificial intelligence has played an important role in image segmentation for numerous purposes and, taking advantage of these tools, this work makes use of convolutional neural networks, through the U-Net architecture, for segmentation of renal images. To achieve these objectives, two databases were used, where one is related to CT scans and the other to ultrasounds. Based on this, image pre-processing techniques were used to assist in the semantic segmentation task. Finally, the insertion of artificial intelligence in image segmentation was evaluated as beneficial, with the model achieving a Dice similarity metric for the test set of 0.83 for CTs and 0.96 for ultrasounds.

Keywords: Semantic Segmentation. Convolutional Neural Network. U-Net. Kidney.

LISTA DE ILUSTRAÇÕES

Figura 1 – Representação visual de uma arquitetura CNN	13
Figura 2 – Representação visual de uma operação de convolução	14
Figura 3 – Representação visual de max pooling e average pooling	15
Figura 4 – Representação visual de uma camada full connected	16
Figura 5 – Representação visual de uma imagem segmentada	18
Figura 6 – Arquitetura de uma rede Encoder-Decoder.....	19
Figura 7 – Arquitetura da U-Net	20
Figura 8 – Dados de entrada: Tomografia, Máscara Renal Original	23
Figura 9 – Dados de entrada: Ultrassonografia, Máscara Renal Original	23
Figura 10 – Ilustração da segmentação de alguns casos da rede proposta para TC.....	31
Figura 11 – Segmentação de melhores casos da rede proposta para ultrassonografias.....	31
Figura 12 – Segmentação do rim onde ocorreu falha.....	32
Figura 13 – Representação do coeficiente de Dice em TC.....	32
Figura 14 – Representação do coeficiente de Dice em ultrassonografias.....	33

LISTA DE TABELAS

Tabela 1 – Parâmetros utilizados na função do aumento de dados.....	25
Tabela 2 – Hiper parâmetros para o treinamento da rede U-Net.....	27
Tabela 3 – Quantidade de parâmetros a cada etapa convolucional da U-Net gerada.....	28

LISTA DE ABREVIATURAS E SIGLAS

CNN	<i>Convolutional Neural Network</i>
DICOM	<i>Digital Imaging and Communications in Medicine</i>
FC	<i>Fully Connected</i>
FCNN	<i>Fully Convolutional Neural Network</i>
GPU	<i>Graphics Processing Unit</i>
ILSVRC	<i>ImageNet Large-Scale Visual Recognition Challenge</i>
PNG	<i>Portable Network Graphics</i>
ReLU	<i>Rectified linear units</i>
RGB	<i>Padrão de cores (Vermelho (Red), Verde (Green) e Azul (Blue))</i>
RNA	<i>Rede Neural Artificial</i>
SBN	<i>Sociedade Brasileira de Nefrologia</i>
TC	<i>Tomografia Computadorizada</i>

SUMÁRIO

1	INTRODUÇÃO	10
1.1	Objetivos	11
1.1.1	Objetivos gerais	11
1.1.2	Objetivos específicos	11
2	FUNDAMENTAÇÃO TEÓRICA	12
2.1	Redes Neurais Convolucionais	13
2.1.1	Camada de Convolução	13
2.1.2	Camada de Pooling	15
2.1.3	Camada totalmente conectada	15
2.1.4	Camada de Saída	17
2.2	Segmentação Semântica	17
2.3	U-Net	19
2.4	Coeficiente de Dice	20
3	METODOLOGIA	22
3.1	Detalhes da Implementação	22
3.2	Conjunto de Dados	22
3.3	Pré-processamento	23
3.3.1	Normalização	24
3.3.2	Aumento de Dados	24
3.4	Arquitetura	26
3.4.1	U-Net	26
3.5	Métrica de desempenho	29
4	RESULTADOS E DISCUSSÕES	30
5	CONCLUSÃO	34
5.1	Trabalhos Futuros	35
	REFERÊNCIAS	36

1 INTRODUÇÃO

A segmentação do rim é de grande relevância para ajudar médicos e especialistas a realizarem a avaliação da função renal, localização do tecido, operação intervencionista, o que é altamente necessário na prática clínica. Muitas vezes, as doenças renais despontam assintomáticas, intensificando-se e, por fim, levando a complicações permanentes. Atualmente, a segmentação do rim é realizada de forma manual, porém, este processo é drasticamente demorado e inclinado a variabilidade entre diversos especialistas. Tendo em vista este trabalho custoso, o desenvolvimento de ferramentas que melhorem a precisão e otimizem o processo de segmentação, tornaram-se imprescindíveis para auxiliar no diagnóstico precoce de doenças renais.

O aprendizado profundo, e em particular as redes neurais convolucionais (CNN), retratam técnicas de ponta para segmentação em imagens médicas. Os métodos CNN extraem um conjunto complexo de recursos da imagem e alcançam resultados superiores em comparação aos métodos tradicionais de aprendizado de máquina. Neste trabalho, foi proposto um método de segmentação automática para delimitar os rins em imagens de tomografia computadorizada (TC) e ultrassonografia, utilizando técnicas de processamento de imagens e redes neurais convolucionais profundas.

A TC, por sua vez, exerce um papel crítico no diagnóstico, avaliação e tratamento de inúmeras doenças referentes aos rins. Contudo, a segmentação precisa de tecidos renais a partir de imagens dinâmicas de TC é um desafio devido a muitas razões, incluindo artefatos de aquisição de dados, grande falta de homogeneidade do rim, grandes diferenças anatômicas entre indivíduos e intensidades semelhantes de órgãos adjacentes, sendo assim, cerceada pela exigência de segmentar manualmente a área de interesse, uma etapa duradoura. Já no que diz respeito a ultrassonografias, a mesma permite procedimentos mais rápidos e precisos devido a sua capacidade em tempo real, além de ter um baixo custo. No entanto, em comparação com imagens de TC, é particularmente difícil de segmentar, uma vez que a qualidade das imagens é baixa, dificultando a aplicação dos métodos de aprendizado profundo.

Para solucionar problemas da área médica, as redes neurais estão sendo aplicadas desde o final da década de 90, mas apenas recentemente vem se difundindo pelos inúmeros avanços tecnológicos, possibilitando aplicar essas redes ao

diagnóstico médico por imagens com desempenhos efetivos. Posto isso, foi sugerida, em 2012, a rede neural AlexNet, pioneira do movimento de CNN's que processam enorme volume de imagens. Depois disso, várias arquiteturas CNN's foram desenvolvidas como a U-Net, VGG16, GoogleNet e todas elas com o propósito equivalente de segmentação e classificação, mas, com estruturas, números de camadas e filtros dispares.

A arquitetura U-Net foi proposta por Ronneberger *et al.* (2015), e no que lhe concerne, se mostrou mais ligeira para trabalhos de segmentação de imagens, por essa razão vem sendo utilizada no mundo inteiro. Essa estrutura consegue treinar uma quantidade de dados substanciais, utilizando-se de uma janela de contração para prever os pixels com maior potencial da classificação desejada (PAN *et al.*, 2020). Além disso, a U-Net também pode ser usada em trabalhos envolvendo poucos treinamentos, sendo, então, uma ferramenta com aplicações para diferentes tipos de segmentação semântica. Na presente pesquisa, é proposto a utilização de uma CNN como elemento primordial em técnicas de segmentação de imagens renais, como a U-Net, é usada vigorosamente em aplicações de áreas médicas.

1.1 Objetivos

1.1.1 Objetivo geral

Este trabalho tem como objetivo principal desenvolver um modelo de segmentação semântica, baseado na arquitetura U-Net para segmentação do rim em imagens de tomografias computadorizadas e ultrassonografias.

1.1.2 Objetivos específicos

- Validar o desempenho do modelo gerado por meio do coeficiente Dice, usado para comparar sobreposições de áreas em imagens;
- Reproduzir uma segmentação coesa a máscara disponibilizada pelo especialista;
- Obter dois modelos de segmentação semântica para cada tipo de imagem e avaliar seus desempenhos de acordo com a métrica definida;

2 FUNDAMENTAÇÃO TEÓRICA

Inicialmente, na presente pesquisa, objetiva-se esclarecer conceitos iniciais acerca do tema. Com foco nisso, este capítulo foi dividido em quatro seções. A primeira seção busca esclarecer a definição de uma Rede Neural Convolutiva e como ela é caracterizada. A segunda seção tem o intuito de fazer uma exposição sobre segmentação semântica. A terceira seção tem por finalidade explicar a U-Net. Por fim, é apresentada a métrica de desempenho.

2.1 Redes Neurais Convolucionais

As redes neurais convolucionais (CNN) são modelos biologicamente inspirados que podem aprender características de forma hierárquica (LECUN; BENGIO; HINTON, 2015).

CNN é um modelo baseado em *deep learning* similar às redes neurais artificiais onde sua aplicação mais comum é no processamento de imagens. Sua criação foi possibilitada através de uma série de avanços em diversos campos de estudos desde o final da década de 1960.

A melhoria da performance e evolução das CNNs desde 1989 a 2015 pode ser atribuída a dois fatores. Primeiro, maiores bancos de imagens diminuíram o grau de generalização das redes. Segundo, as arquiteturas de rede se tornaram maiores, devido a computadores mais poderosos e software com melhores infraestruturas (GOODFELLOW; BENGIO; COURVILLE, 2016).

A primeira realização exitosa de uma CNN se deu por Yann LeCun em 1998, quando o mesmo propôs uma arquitetura de rede baseada em camadas convolucionais e camadas totalmente conectadas. A rede em questão foi treinada mediante o algoritmo de *backpropagation*, sendo usada também para a identificação de dígitos manuscritos alcançando uma imprecisão de 5%, usando duas camadas convolucionais e 2 camadas totalmente conectadas.

Em 1998, foi determinado o modelo padrão das CNNs, com camadas convolucionais alternadas de operação de max pooling(valor máximo) para redução da dimensionalidade e algumas camadas totalmente conectadas.

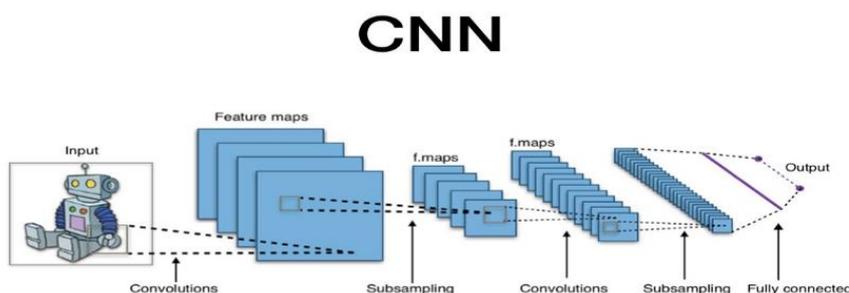
Em 2012, foi apresentada a CNN AlexNet. Esta rede atuou no desafio ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), vencendo o mesmo por uma

margem ampla, diminuindo a taxa de erro do estado da arte de 25,8% para 16,4%. Esse resultado foi responsável por popularizar o uso das redes CNNs na área da visão computacional.

De forma geral, as CNNs foram aplicadas a atividades de visão computacional desde o final da década de 1980. Contudo, existiram apenas algumas aplicações espalhadas, e a pesquisa ficou factualmente paralisada até meados dos anos 2000, quando houve um incremento na quantidade de dados rotulados, completados por algoritmos melhorados e o gradativo poder computacional das GPUs que colaboraram para o seu avanço e um veloz progresso desde 2012.

Atualmente, uma CNN tem sua arquitetura composta por três tipos de camadas: convolução, *pooling* e completamente conectada. Cada camada possui diferentes fundamentos que podem ser otimizados e efetuam funções diversas nos dados de entrada. Na continuação é descrita brevemente cada uma delas. A Figura 1 elucida a arquitetura de uma CNN.

Figura 1 – Representação visual de uma arquitetura CNN



Fonte: (CONVOLUTIONAL, 2015).

2.1.1 Camada de Convolução

Em termos matemáticos, uma convolução é a operação de somatório da área de sobreposição entre duas funções, em razão da variação de posição em um dado intervalo de tempo existente entre elas (JANGIR, 2019). Em imagens digitais, uma convolução é a simples aplicação de um filtro a uma entrada resultante em uma ativação.

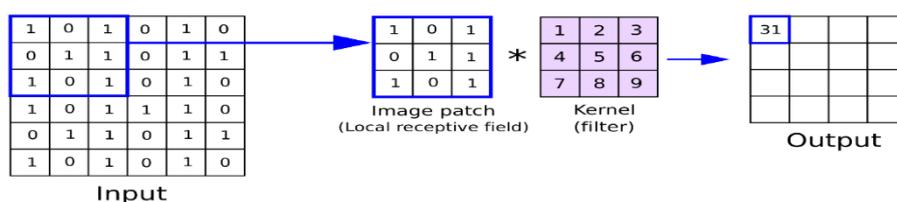
A camada de convolução é o elemento central de uma CNN, é a partir dela que se extrai os recursos da imagem. Os parâmetros desta camada se compõem de filtros, que são conjuntos de pesos inicializados aleatoriamente aplicados em toda a imagem,

durante o processo de *backpropagation*, onde cada fatia 2D dos filtros é chamada de *kernels*. Esses filtros se deslocam ao longo dos eixos x e y da matriz por passos, onde introduzem invariância de tradução e compartilhamento de parâmetros e através da convolução, eles são aplicados.

As camadas de convolução são compostas por filtros treináveis aplicados em toda a imagem de entrada, gerando assim, vários mapas de características (LECUN *et al.*, 2010). Os filtros determinam uma pequena área (3×3 , 5×5 , 7×7 pixels) e cada neurônio é ligado somente aos neurônios nas proximidades da camada anterior. Os neurônios compartilham os pesos entre si, levando os filtros a aprender os padrões frequentes que ocorrem em qualquer parte da imagem. Em seguida ao treinamento da rede, os filtros tornam-se encarregados da detecção de características peculiares que ocorrem na imagem (HAFEMANN, 2014).

Posteriormente a convolução, um conjunto de características, que é a matriz de saída da camada, é extraído da imagem, onde estas nada mais são do que novas imagens procedentes da aplicação dos filtros. A aplicabilidade de vários filtros quanto a uma mesma imagem implica na extração de diferentes características e é com as relações entre esses filtros, a partir do processo de treinamento, que a rede neural vai absorvendo os cruciais conjuntos de características que compõem as imagens a qual a rede foi submetida durante o treino. Uma explicação mais visual pode ser observada na figura 2.

Figura 2 – Representação de uma operação de convolução



Fonte: (DHARMARAJ, 2022).

Após a operação de convolução, onde são gerados os mapas de características, uma função de ativação não-linear é aplicada a cada pixel das imagens resultantes. Introduzir não-linearidade a um modelo torna-o capaz de criar associações mais complexas entre entradas e saídas da rede, o que é essencial para aprendizado de dados complexos, como imagens (GOODFELLOW; BENGIO; COURVILLE, 2016).

Pode-se, inclusive, constatar que conforme se aprofunda na rede, é muito comum ter um aumento gradativo no número de mapas de ativação. Isso é propositalmente projetado, pois, uma vez que os mapas de ativação ficam cada vez mais especializados, representando conceitos mais abstratos, simplesmente são necessários mais mapas de ativação para representar a imagem original (JORDAN, 2017).

2.1.2 Camada de Pooling

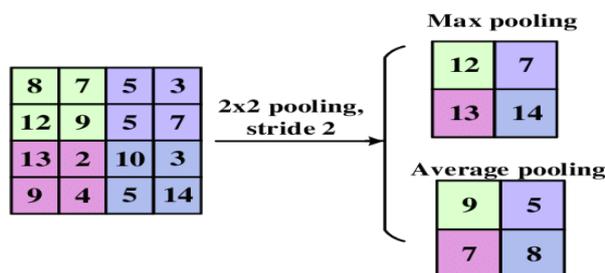
A camada de *pooling* é comumente utilizada entre duas camadas de convolução. Essa camada é responsável por simplificar e reduzir a dimensão de entrada, fazendo com que os parâmetros presentes sejam generalizados e as características dominantes da entrada sejam extraídas. Qualquer mapa de características de uma camada de *pooling* está conectada ao análogo mapa de características da camada convolucional antecedente.

Essa camada pode executar operações típicas de *pooling*:

- *Max pooling*, que percorre a matriz de entrada e obtém o valor máximo de cada região coberta pelo filtro.
- *Average pooling*, que atua de forma similar, contudo, obtém o valor médio dos pixels da região que o filtro cobre.

Um exemplo dos dois tipos de pooling pode ser observado na figura 3.

Figura 3 – Representação visual de *max pooling* e *average pooling*



Fonte: (DHARMARAJ, 2022).

2.1.3 Camada Totalmente Conectada

Uma camada totalmente conectada obtém todos os neurônios na camada antecedente e os conectam a cada neurônio da camada atual. Normalmente, os filtros

da camada convolucional inicial são planejados para detectar características de pequeno nível, enquanto os filtros das camadas superiores identificam características mais abstratas. Logo, ao amontoar várias camadas convolucionais e de *pooling*, é possível extrair, passo a passo, características com mais nível de abstração.

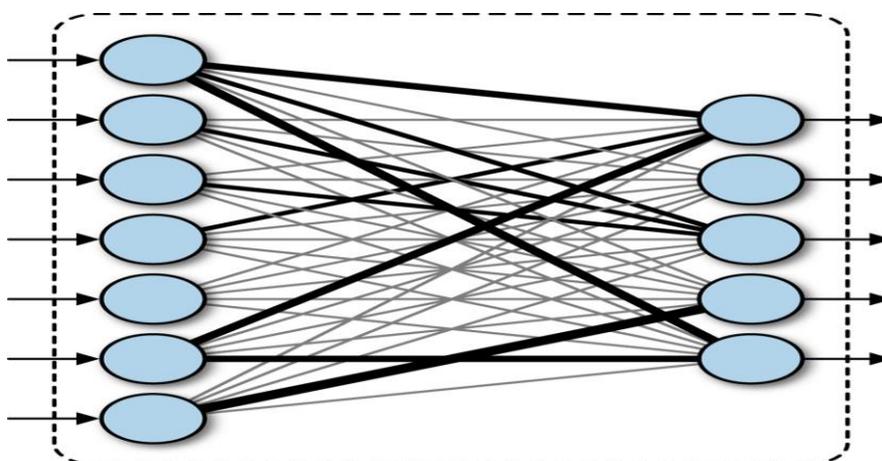
Após várias camadas convolucionais e de *pooling*, uma ou mais camadas completamente conectadas são utilizadas antes da saída de classificação de uma CNN. Esse tipo de camada é utilizada para achatar os resultados, assim como ajuda no aprendizado de combinações não lineares a partir das características da imagem obtidas pelas camadas de convolução. A camada é totalmente conectada, pois, todas as entradas da camada estão interligadas aos neurônios da camada de saída, o que a difere com a fase de extração de traços onde os neurônios de processamento são livres uns dos outros e só têm acesso às informações do campo receptivo que estão processando.

A saída da camada anterior deve passar por um processo de achatamento (*flattening*) antes de entrar na camada totalmente conectada: os mapas de ativação que até então eram representados como matrizes, são transformados em um único vetor que irá alimentar a camada totalmente conectada (JORDAN, 2017).

O objetivo da camada totalmente conectada é interpretar as características de alto nível de abstração efetuando funções de raciocínio complexo, bem como aplicar os recursos que foram extraídos do conjunto de dados de treinamento para a tarefa de classificar a imagem de entrada em várias categorias diferentes.

Um exemplo de uma camada totalmente conectada pode ser visto na figura 4.

Figura 4 – Representação visual de uma camada full connected



Fonte: (CARVALHO, 2021).

2.1.4 Camada de saída

Subsequente às camadas totalmente conectadas, encontra-se a camada de saída. A camada de saída é capaz de interpretar essas características e realizar funções de raciocínios de alto nível como classificação de objetos em uma imagem (RAWAT; WANG, 2017).

2.2 Segmentação semântica

A segmentação de imagens é uma das tarefas mais indispensáveis na visão computacional. A partir da segmentação é possível segregar várias regiões de acordo com propriedades presentes em cada pixel. Para facilitar a análise, o principal objetivo da segmentação é extrair informações significativas (SULTANA; SUFIAN; DUTTA, 2020).

A princípio, a segmentação de imagens é obter uma imagem e dividi-la em vários fragmentos menores, onde o objetivo é associar pixels ou conjunto de pixels de mesmos predicados. Para tarefas de segmentação de imagens, um requisito essencial é o uso de rótulos. Uma vez que descrevemos os constituintes mais essenciais da imagem obtida durante a segmentação com a ajuda das imagens e suas respectivas máscaras, podemos realizar uma infinidade de tarefas futuras com elas.

Atualmente, uma imagem em níveis de cinza pode ser segmentada de duas formas: por descontinuidade e por similaridade. Uma imagem segmentada equivale-se a imagem primária quanto ao seu formato, mas diferencia-se quanto a coloração. Na primeira forma, a abordagem é particionar uma imagem baseando-se nas modificações súbitas nos níveis de cinza. Já a segunda forma, concentram-se em limiarização, crescimento de regiões, divisão, conquista e aglomeração.

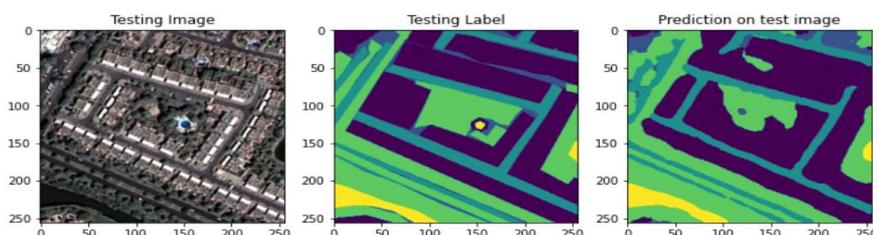
O processo de segmentação de imagem é um processo que aparta a imagem em fragmentos menores, para depois tratá-los, simplificando assim o processo de reconhecimento. Baseado nisso, os algoritmos de segmentação possibilitam achar distinções entre duas ou mais regiões, distinguindo-as das outras. Outro procedimento de segmentação é a combinação de pixels e regiões vizinhas de características visuais semelhantes, que constituem objetos e distribuem rótulo a cada pixel de imagem, caracterizando a qual grupo ele pertence, de acordo com suas características como cor e textura (GONZALEZ; WOODS, 2010).

A segmentação semântica é o processo de atribuição de uma classe de objeto para cada pixel de uma imagem. Isto faz com que a tarefa de segmentação seja complexa e uma das poucas áreas da visão computacional onde o desempenho das redes neurais profundas estão aquém do desempenho humano (PLANCHE; ANDRES, 2019).

A segmentação semântica tem por objetivo classificar cada um dos píxeis de uma imagem em classes, porém, difere de classificação, visto que concede uma classe para cada píxel, ao passo que classificação atribui apenas uma classe para a imagem inteira.

Na imagem segmentada, todo pixel é colorido com uma de Z cores, onde Z é o número de classes que se está segmentando. Ao segmentar uma imagem de satélite, por exemplo, a representação poderia ser simples como $Z = 2$ (“estrada” ou “não estrada”), como também uma série de diversas classes que consigam ser capturadas, exemplo que pode ser visto na figura 5.

Figura 5 – Representação visual de uma imagem segmentada



Fonte: (CHINMAY, 2022).

De modo que a saída resultante da segmentação sejam pixels brutos, pode ser necessário converter esses dados em uma forma adequada, propiciando o processamento. Existem dois tipos de representação que podem ser utilizados: a representação limite, que é aprimorada quando o foco está em característica da forma, e a representação regional, aprimorada quando o foco está em textura e cor.

Para atingir a segmentação, é possível substituir a camada totalmente conectada por convoluções, gerando uma rede neural totalmente convolucional, ou *Fully Convolutional Neural Network* (FCNN). Desta maneira, o modelo pode obter imagens de entrada maiores do que foi treinado e produzir um mapa de probabilidade, ao invés de uma saída para um único pixel. Devido às camadas de *pooling*, é possível obter uma saída com uma resolução muito menor do que a de entrada (LONG; SHELHAMER; DARRELL, 2015).

2.3 U-Net

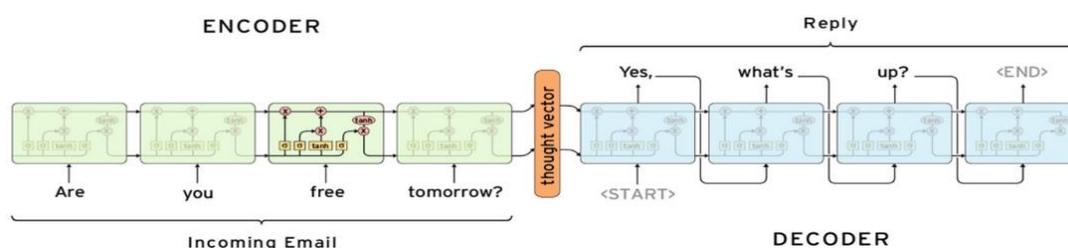
Ronneberger, Fischer e Brox (2015) apresentaram a U-Net, que é uma RNA (Rede Neural Artificial), estendida das redes totalmente convolucionais (FCN). A mesma teve em sua concepção, o objetivo de realizar segmentações mais precisas com conjuntos de treinos pequenos. A U-Net é uma rede totalmente convolucional, cuja estrutura consiste em um caminho de contração para registrar o contexto, e um caminho de expansão semelhante que viabiliza uma segmentação com exatidão.

A U-Net é uma arquitetura de rede convolucional para segmentação rápida e precisa de imagens. A intenção da U-Net é capturar tanto as características do contexto quanto a localização. A ideia indispensável da execução é utilizar camadas de contração sucessivas, que são consecutivamente seguidas pelos operadores de *upsampling* para obter saídas de maior resolução nas imagens de entrada.

Segundo Ronneberger, Fischer e Brox (2015) a principal estratégia que diferencia a U-Net das outras arquiteturas FCN é a combinação entre os mapas de características do estágio de contração e seus correspondentes semelhantes no estágio de expansão, permitindo a propagação de informações de contexto para os mapas de características de alta resolução.

A apresentação de uma rede totalmente conectada foi introduzida por (LONG; SHELHAMER; DARRELL, 2015), em que os mesmos abordaram uma adaptação de redes popularizadas já existentes, modificando as camadas totalmente conectadas em camadas de convolução. Isto propicia o desenvolvimento de mapas de características de segmentação para cada imagem. A U-Net é capaz de juntar as informações contextuais obtidas do caminho de contração (*encoder*) com as informações de localização adquiridas do caminho de expansão (*decoder*) com um bom desempenho. A arquitetura de uma rede *Encoder-Decoder*, a qual a U-Net faz parte, pode ser observada na figura 6.

Figura 6 – Arquitetura de uma rede *Encoder-Decoder*

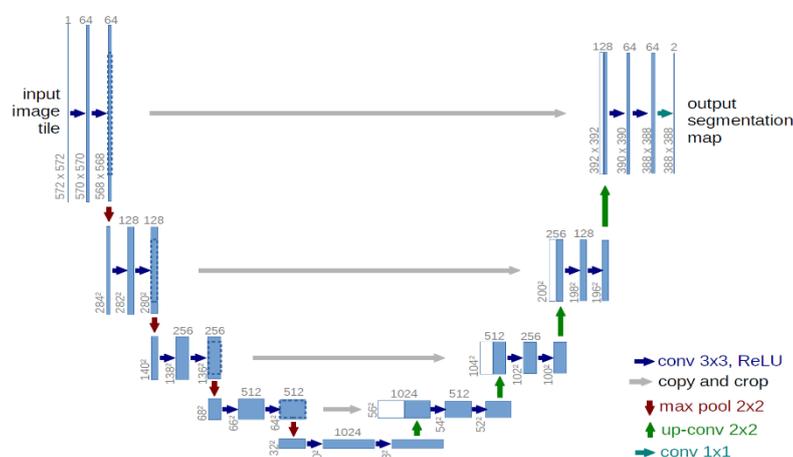


Fonte: (GONÇALVES, 2018).

- *Encoder*: codifica a imagem em uma representação abstrata dos recursos da imagem empregando uma sequência de blocos convolucionais que minimizam gradativamente a altura e a largura da representação, mas um número crescente de canais que correspondem aos recursos da imagem.
- *Decoder*: decodifica a representação da imagem em uma máscara binária aplicando uma sequência de convoluções ascendentes que aumenta gradualmente a altura e a largura da representação para o tamanho da imagem original e diminui o número de canais para o número de classes que está sendo segmentado.

A figura 7 exemplifica a arquitetura da rede U-Net, que possui formato de U e consiste em um caminho de contração e um caminho expansivo. A via de contração segue a arquitetura típica de uma rede convolucional. Ela pode ser dividida em um caminho codificador-decodificador ou caminho expansivo de contração equivalente.

Figura 7 – Arquitetura da U-Net



Fonte: (RONNEBERGER; FISCHER; BROX, 2015).

2.4 Coeficiente de Dice

Para representar a validação dos resultados, ou seja, a capacidade da segmentação, a métrica utilizada foi o Coeficiente de Dice, criado pelos botânicos Thorvald Sørensen e Lee Raymond Dice. Essa métrica dispõe-se a comparar as características comuns a duas amostras, indicando a proporção de pixels

compartilhados entre elas. O cálculo de semelhança se dá através da comparação de áreas segmentadas pelo modelo e pela máscara original.

$$Dice = \frac{2TP}{2TP + FP + FN}$$

A fórmula, representada acima, é igual a duas vezes o número de elementos comuns às duas imagens, dividido pela soma do número total de pixels em uma e outra. A faixa de valores compreendidos por esta medida vão de 0 (quando não existe nenhuma similaridade entre as amostras) a 1 (quando as amostras são completamente iguais).

3 METODOLOGIA

Este capítulo procura abordar características importantes relacionadas a metodologia utilizada nos experimentos. Primeiramente são apresentadas as ferramentas utilizadas para o desenvolvimento do código, depois, definições ligadas ao conjunto de dados, em sequência, o que diz respeito a pré-processamento e, por fim, sobre a arquitetura utilizada.

3.1 Detalhes da Implementação

A modelagem e predição deste estudo foram realizadas utilizando-se de um LENOVO IDEAPAD, CORE I7, 8 GB, com sistema operacional Windows 10, feita em ambiente do *Google Colab* (Processamento em nuvem). A rede neural convolucional, bem como o pré-processamento dos dados, foram implementados em Python, utilizando bibliotecas como Keras/Tensor flow e de outras como Numpy, Pandas, Scikit-learn, OpenCV e Matplotlib.

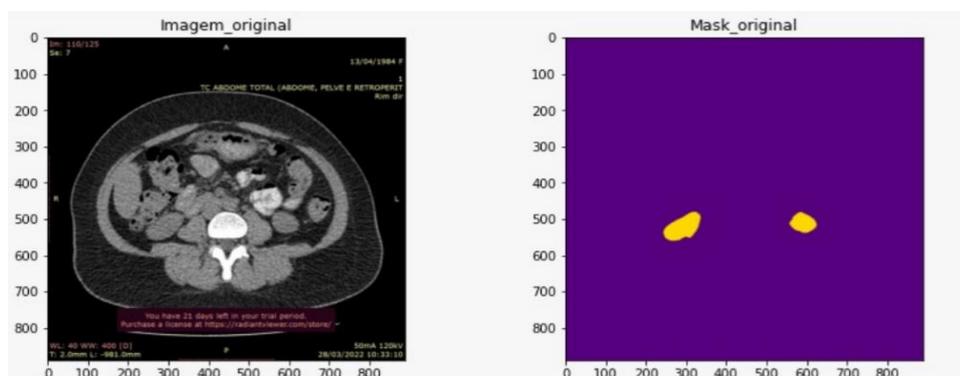
3.2 Conjunto de Dados

Para desenvolvimento do modelo proposto, foram adotados dois *datasets*: um *dataset* público de imagens de ultrassonografias do abdômen em formato PNG, obtidos no portal Kaggle de desafios online, e um *dataset* privado de imagens de tomografias do abdômen em formato DICOM, anonimizadas, obtido no Núcleo de Tecnologia Estratégicas em Saúde da UEPB. Ambos os *datasets* contém uma pasta com as imagens originais e uma pasta com arquivos homônimos contendo a máscara da ROI na respectiva imagem original. A primeira base de dados é composta de 4.586 imagens de ultrassonografias e 4.586 máscaras, das quais foram alocadas 90% para treinamento e 10% para testes. Já a segunda base de dados possui 109 imagens, das quais 96 foram destinadas ao treinamento do modelo e 13 para teste.

Na base de dados de ultrassonografia, as máscaras foram geradas manualmente, em padrão ouro, por médicos profissionais. Similarmente, na base de dados com imagens de tomografia computadorizada, as máscaras geradas para cada imagem foram obtidas a partir de segmentação manual em padrão ouro, considerando apenas região correspondente ao rim, realizada por radiologista em corte axial.

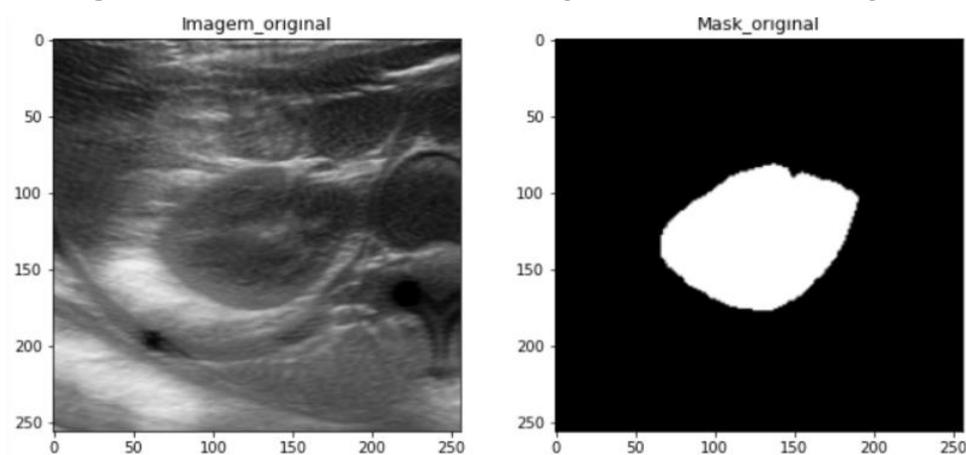
Um exemplo de ambas as segmentações manuais está ilustrado nas figuras 8 e 9.

Figura 8 – Dados de entrada: Tomografia, Máscara Renal Original



Fonte: Autoria própria.

Figura 9 – Dados de entrada: Ultrassonografia, Máscara Renal Original



Fonte: Autoria própria.

3.3 Pré-processamento

O pré-processamento é uma fase absoluta do aprendizado de máquina, pois, a qualidade dos dados e informações pertinentes que podem ser decorrentes deles afetam pontualmente a habilidade de aprendizado do modelo. Assim sendo, é imprescindível trazê-los a um estado que o algoritmo possa facilmente analisá-los, tendo por finalidade a melhoria da qualidade para qualquer tipo de imagem.

As amostras do conjunto de dados estavam em formato PNG, padrão de cores RGB e 8 bits de resolução, cujos valores dos pixels variam entre 0 e 256 níveis de

intensidade. Dadas as características das imagens do conjunto de dados, foi necessário realizar a conversão do padrão RGB para níveis de cinza. Depois, as imagens foram transformadas para o tipo numpy array (uint8) e redimensionadas para o tamanho 256 de largura e 256 de altura. Esse redimensionamento foi realizado para diminuir a carga computacional necessária para realização do treinamento do modelo.

3.3.1 Normalização

A normalização é uma técnica corriqueira na etapa de preparação dos dados e tem como finalidade colocá-los em um intervalo de valores comuns. Esse processo se faz necessário para evitar possíveis erros de valores em ordem de grandezas discrepantes.

No treinamento de modelos de *Machine Learning*, a normalização é fundamental para auxiliar na convergência dos resultados, assim, por meio dela, é possível acelerar o processo de aprendizado da rede. A normalização diminui a importância dos pesos iniciais, haja visto que o intervalo na qual os pesos serão inicializados está dentro da mesma faixa.

É importante destacar que na etapa do pré-processamento é aplicada a quantização não-uniforme da imagem para moralizar os valores originais, ou seja, entre 0 e 255, para a nova faixa de valores, entre 0 e 1. Esse processo é intrínseco ao treinamento da rede neural convolucional e ocorre simplesmente dividindo o valor do pixel por 255, de modo que teremos $\text{imagem}(x,y)/255$.

3.3.2 Aumento de dados

Para apresentar um bom desempenho, as redes precisam de uma grande quantidade de imagens para treino, quando não existem imagens suficientes realiza-se o aumento de dados (SHORTEN; KHOSHGOFTAAR, 2019).

O aumento de dados é um processo utilizado para ensinar à rede a invariância desejada e propriedades de robustez, quando apenas algumas amostras de treinamento estão disponíveis, assim aumentando a quantidade e a diversidade de dados do conjunto de treinamento visando ampliar a generalização do modelo. Especialmente, deformações elásticas aleatórias das amostras de treinamento podem

ser o conceito-chave para treinar uma rede de segmentação com poucas imagens anotadas.

No decorrer do treinamento, a técnica *data augmentation* foi empregada para gerar variações no conjunto de dados.

Dado a elevada porção de parâmetros utilizados por uma CNN, é indispensável mostrar uma parcela proporcional de exemplos para se alcançar uma boa generalização e um bom desempenho. No entanto, a baixa quantidade de amostras de entrada disponíveis em acervos públicos e a dificuldade de um médico liberar muitas imagens, requer o aumento dos dados para aquisição de resultados efetivos.

Com isso, para aumentar a quantidade de dados, foram geradas novas imagens com as transformações a seguir:

- Redimensionamento de 5% das imagens do conjunto de dados, escolhidas aleatoriamente, com a aplicação de *zoom in* e *zoom out*.
- Translação de 5% das imagens do conjunto de dados, escolhidas aleatoriamente, nos eixos x ou y.
- Rotação de 20% das imagens do conjunto de dados, escolhidas aleatoriamente, nos eixos x ou y.
- Cisalhamento de 5% das imagens do conjunto de dados, escolhidas aleatoriamente.

Tabela 1 – Parâmetros utilizados na função do aumento de dados

Operação	Valor atribuído
Faixa de rotação	0,2
Faixa de mudança de largura	0,05
Faixa de mudança de altura	0,05
Faixa de cisalhamento	0,05
Alcance do zoom	0,05
Inversão horizontal	Verdadeiro
Modo de preenchimento	Mais próximo

Fonte: Autoria própria.

3.4 Arquitetura

A próxima etapa da metodologia é o treinamento do modelo, utilizando uma rede neural convolucional treinada para este fim. A arquitetura da CNN escolhida foi a U-Net, proposta inicialmente por Ronneberger, Fischer e Brox (2015). Ela é uma rede completamente convolucional utilizada para a segmentação de regiões específicas em imagens, ou seja, é feita para cada pixel.

3.4.1 U-Net

O treinamento exitoso de redes profundas requer milhares de exemplos de treinamento. A intenção da U-Net, é capturar tanto as características do contexto quanto a localização, ela é uma arquitetura totalmente convolucional voltado a segmentação semântica. A ideia principal da implementação é utilizar camadas de convoluções sucessivas, as quais são imediatamente seguidas pelos operadores de upsampling para obter saídas de maior resolução nas imagens de entrada.

A arquitetura U-Net é uma arquitetura para segmentação semântica, que consiste em um caminho de contração e um caminho expansivo. O caminho de contração segue a arquitetura típica de uma rede convolucional, que é aplicação repetida de duas convoluções 3x3, cada uma seguida por uma unidade linear retificada (ReLU) e uma operação de *pooling 2x2 max* com passo 2 para *downsampling*. Em cada etapa de redução da amostra, dobramos o número de canais de recursos. Cada passo no caminho expansivo consiste em um upsampling do mapa de recursos seguido por uma convolução 2x2 que reduz pela metade o número de canais de recursos, uma concatenação com o mapa de recursos correspondente recortado do caminho de contratação e duas convoluções 3x3, cada uma seguida por um ReLU. O recorte é necessário devido à perda de pixels da borda em cada convolução. Na camada final, uma convolução 1x1 é usada para mapear cada vetor de recursos de 64 componentes para o número desejado de classes. Para criar as camadas de convolução, foram usados os seguintes filtros = [64, 128, 256, 512, 1024].

No decorrer da rede, foi feito o uso da técnica *Dropout*, que consiste na remoção temporária de unidades, dada uma probabilidade P obtida durante o treinamento (SRIVASTAVA et al., 2014). A remoção é efetuada através da atribuição do valor zero a saída do neurônio escolhido aleatoriamente. A vantagem dessa técnica

é o aumento da generalização do modelo, visto que força a rede a extrair características mais robustas. Por sua vez, uma desvantagem é a perda do aprendizado, que se dá pela morte do neurônio.

O treinamento da rede foi fragmentado em ciclos, onde, após o modelo concluir uma época de treinamento, é analisado se houve ou não melhoria em relação ao resultado da época antecedente. A melhora é medida por meio da função de perda destinada às máscaras do modelo geradas pela base. Se a perda apresentar diminuição, os pesos da rede são salvos. Assim sendo, isso assegura que os valores dos parâmetros treináveis sejam os melhores encontrados durante o treinamento.

Para o treinamento da rede usando a base de dados das ultrassonografias, foram utilizadas 500 épocas e 200 por etapa, já para os de tomografia computadorizada, tendo em vista a base de dados pequena, 10 épocas e 100 por etapa. Estes números foram escolhidos com base no histórico de experimentos, que mostrou decaimentos da função de perda durando com poucas épocas em relação à quantidade de amostras.

As técnicas e parâmetros utilizados no treinamento de uma rede neural convolucional são fundamentais para o aprendizado da arquitetura. O modelo foi compilado usando o otimizador Adam, cuja finalidade é aumentar a eficiência da rede neural, diminuindo as perdas ocorridas, com uma taxa de aprendizado $1e-4$. Como métrica adicional, foi utilizado a *accuracy* e, para a função de perda, a *binary cross entropy*, que estima as perdas ocorridas na saída da rede neural durante o treinamento.

Tabela 2 – Hiper parâmetros para o treinamento da rede U-Net.

Hiper parâmetro	U-Net
Otimizador	Adam
Taxa de aprendizagem	$1e-4$
Função de perda	<i>binary cross entropy</i>

Fonte: Autoria própria.

Se o número de épocas sem melhora e o limite superior de épocas ainda não foram atingidos, o modelo continuará com o ciclo de treinamento. Assim que o ciclo for cessado é efetuada a averiguação se o limite inferior da taxa de aprendizado foi atingido. Se atingido, o treinamento é concluído. Em compensação, a taxa de aprendizado é limitada. Muitas vezes essa operação melhora a precisão da predição

da rede, visto que a taxa de aprendizado é responsável pelo tamanho da etapa na descida do gradiente. Ainda, de acordo com (YOU et al., 2019), começar com uma taxa de aprendizado grande impede que a rede memorize dados ruidosos, enquanto a diminuição gradual melhora a aprendizagem de padrões complexos.

Todas as imagens e máscaras passaram pela estrutura U-Net, em um processo iterativo longo, contendo contração e expansão. Sendo que quanto maior o número de imagens e máscaras dispostas ao treinamento do modelo, maior é a precisão da segmentação.

Ao decorrer do processo de cada etapa convolucional, os fragmentos das imagens são multiplicados por filtros, que concedem pesos em cada pixel do recorte e, ao executar o *Max Pooling*, há o espasmo do recorte, sendo escolhido os pixels com maior valor. Quanto mais se prossegue na etapa de contração, maior o número de filtros e, conseqüentemente, de parâmetros gerados. O quadro 3 evidencia a quantia de parâmetros gerados em cada etapa de convolução.

Tabela 3 – Quantidade de parâmetros a cada etapa convolucional da U-Net gerada

Tipo/Stride	Dimensões do filtro	Dimensões de entrada	Dimensões de saída	Número de parâmetros
Conv/ s1	3x3x64	256x256x1	256x256x64	640
Conv/ s1	3x3x64	256x256x64	256x256x64	36928
Max Pooling/ s1	2x2x64	256x256x64	128x128x64	0
Conv/ s1	3x3x128	128x128x64	128x128x128	73856
Conv/ s1	3x3x128	128x128x128	128x128x128	147584
Max Pooling/ s1	2x2x128	128x128x128	64x64x128	0
Conv /s1	3x3x256	64x64x128	64x64x256	295168
Conv /s1	3x3x256	64x64x256	64x64x256	590080
Max Pooling/ s1	2x2x256	64x64x256	32x32x256	0
Conv /s1	3x3x512	32x32x256	32x32x512	1180160
Conv/ s1	3x3x512	32x32x512	32x32x512	2359808
Dropout	3x3x512	32x32x512	32x32x512	0
Max Pooling/ s1	2x2*512	32x32x512	16x16x512	0
Conv /s1	3x3x1024	16x16x512	16x16x1024	4719616
Conv /s1	3x3x1024	16x16x1024	16x16x1024	9438208
Dropout	3x3x1024	16x16x1024	16x16x1024	0
UpSampling/ s1	2x2x1024	16x16x1024	32x32x1024	0
Conv /s1	3x3x512	32x32x1024	32x32x512	2097664

Conv/ s1	3x3x512	32x32x1024	32x32x512	4719104
Conv/ s1	3x3x512	32x32x512	32x32x512	2359808
UpSampling/ s1	2x2x512	32x32x512	64x64x512	0
Conv/ s1	3x3x256	64x64x512	64x64x256	524544
Conv/ s1	3x3x256	64x64x512	64x64x256	1179904
Conv/ s1	3x3x256	64x64x256	64x64x256	590080
UpSampling/ s1	2x2x256	64x64x256	128x128x256	0
Conv/ s1	3x3x128	128x128x256	128x128x128	131200
Conv/ s1	3x3x128	128x128x256	128x128x128	295040
Conv/ s1	3x3x128	128x128x128	128x128x128	147584
UpSampling/ s1	2x2x128	128x128x128	256x256x128	0
Conv/ s1	3x3x64	256x256x128	256x256x64	32832
Conv/ s1	3x3x64	256x256x128	256x256x64	73792
Conv/ s1	3x3x64	256x256x64	256x256x64	36928
Conv/ s1	3x3x2	256x256x64	256x256x2	1154
Conv/ s1	1x1x1	256x256x2	256x256x1	3
Total				31.031.6 85

Fonte: Autoria própria.

3.5 Métrica de desempenho

Depois da segmentação, é imprescindível validar os resultados. Esta pesquisa utiliza uma métrica habitualmente empregada na literatura para verificação de desempenho de sistemas fundamentados em processamento de imagens e generalização de padrões para segmentação. A métrica em questão é o coeficiente de Dice, já apresentado na seção 2.4, onde o objetivo é avaliar a taxa de acerto em relação a sobreposição das imagens, cujo resultado deve estar entre 0 e 1.

4 RESULTADOS E DISCUSSÕES

Este capítulo mostra e discute os resultados obtidos pela metodologia proposta para segmentação do rim em exames de TC e ultrassonografias. Ultimamente, os métodos baseados em aprendizado profundo reproduziram a maior parcela de trabalhos no campo da segmentação de imagens médicas. Inúmeras variedades de redes com muitas novas arquiteturas têm surgido, com propriedades inovadoras e resultados significativamente melhores em vários aspectos quando comparados aos métodos mais tradicionais. Ainda assim, entre todos eles, e até onde se tem conhecimento, a arquitetura inicial U-Net consegue alcançar resultados comparáveis e até superar arquiteturas mais recentes no que se diz respeito à segmentação de imagens médicas.

Um conjunto de dados com imagens de boa qualidade e máscaras criadas com precisão são imprescindíveis para o treinamento adequado do código, pois uma imagem com algum tipo de erro humano na hora da geração do rótulo, poderia induzir o código a erros, que seriam posteriormente passados para a segmentação das imagens preditas.

Para legitimar a segmentação semântica baseada na U-Net, foram realizados experimentos com as bases definidas no conjunto de dados. Todo o treinamento realizado usou a mesma arquitetura e configuração, bem como o aumento da base de dados através de operações aleatórias, a única diferença foi no que se refere ao número de épocas.

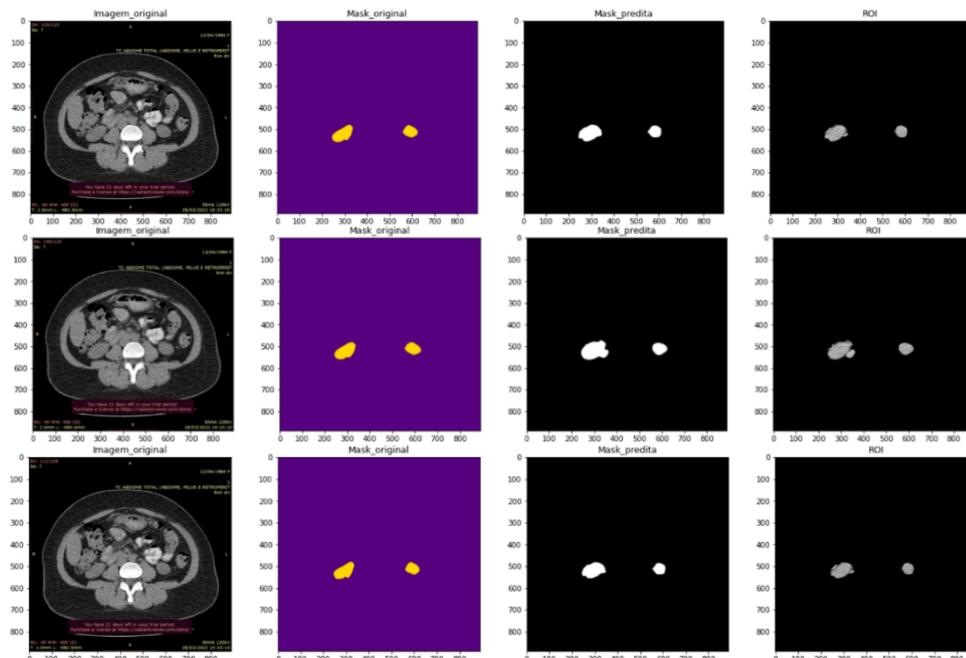
Em resumo, o processo de treinamento de um modelo de aprendizado de máquina supervisionado indica que se deseja obter duas coisas: em primeiro lugar, um modelo que possa prever com sucesso a qual classe uma amostra deve pertencer ou, qual valor deve ser gerado para alguma entrada. Em segundo lugar, embora o poder preditivo seja importante, o modelo também deve conseguir generalizar bem, ou seja, deve prever de forma relativamente correta amostras de entrada que não tenha visto antes.

Após os treinamentos, ficou evidente a dificuldade do modelo em aprender e caracterizar os padrões com uma base de dados pequena, como no caso das tomografias computadorizadas. Isso ocorreu porque a região abdominal possui áreas de muita similaridade com a região do rim. Já com a base de dados das

ultrassonografias, bem mais alimentada, os resultados foram excelentes, conseguindo gerar até as áreas esparsas no ROI.

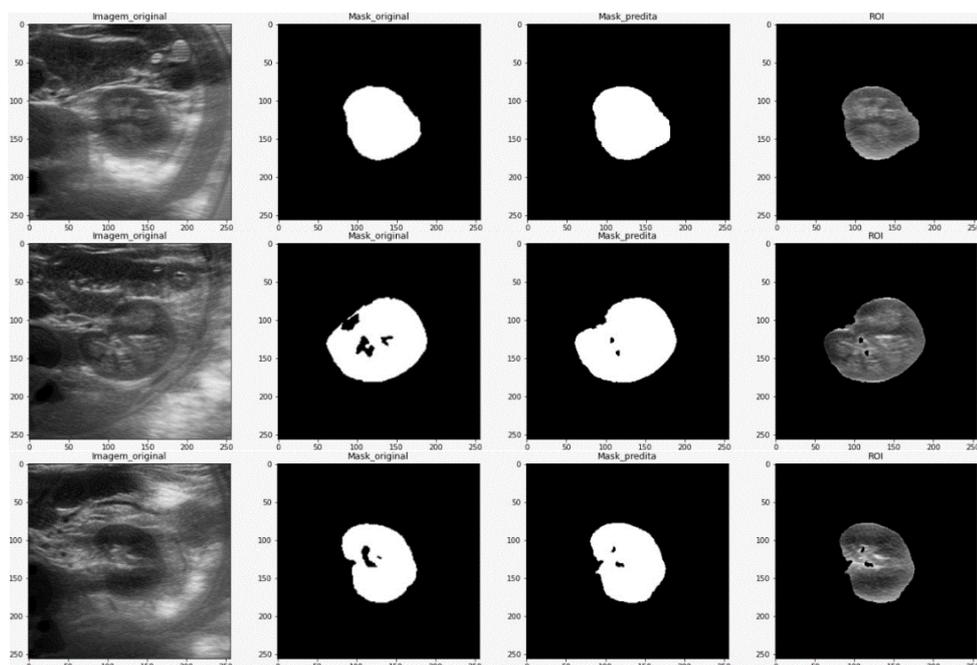
Comumente, a rede se encaminha a gerar máscaras semelhantes às máscaras originais. Abaixo estão algumas imagens com as máscaras originais e as máscaras geradas pelo segmentador.

Figura 10 – Segmentação dos melhores casos da rede proposta para TC



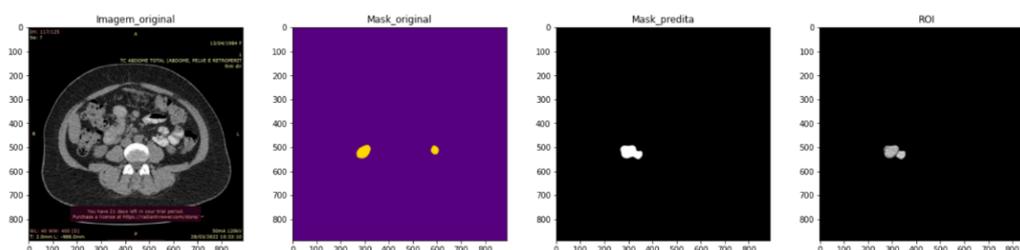
Fonte: Autoria própria.

Figura 11 – Segmentação dos melhores casos da rede proposta para ultrassonografias



Fonte: Autoria própria.

Figura 12 – Segmentação do rim onde ocorreu falha



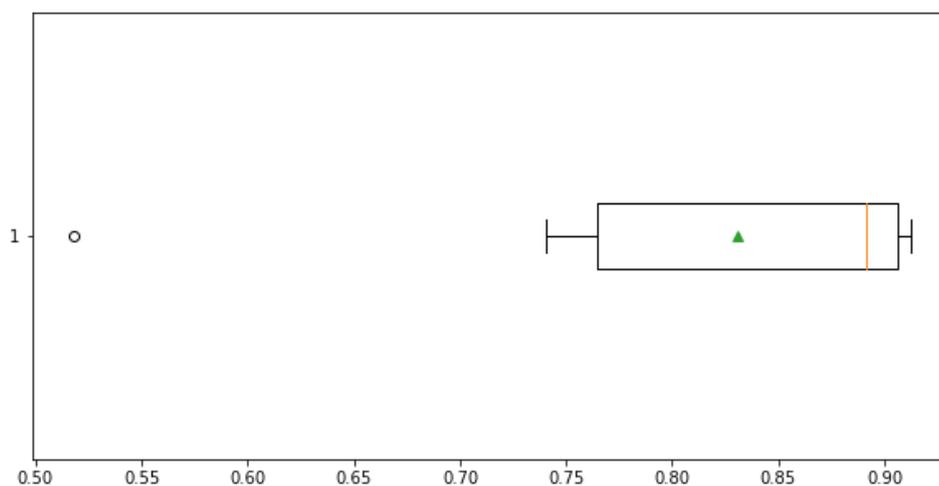
Fonte: Autoria própria.

Em alguns casos, como o da figura 12, o modelo se complicou e a imagem predita identificou apenas um dos rins, diferentemente do que é exposto na máscara original, representando uma falha na segmentação. Um dos motivos para isso se deve ao fato da base de dados das tomografias ser bastante escassa, o que dificultou o aprendizado da rede.

Para verificar e comparar a eficácia do modelo de aprendizado profundo treinado, a análise dos resultados foi fornecida através do coeficiente de Dice. Para isso, foi calculada a média dos coeficientes obtidos no conjunto de testes, que por sua vez registrou uma média de 0,83 para as imagens de tomografias computadorizadas e 0,96 para as imagens de ultrassonografias.

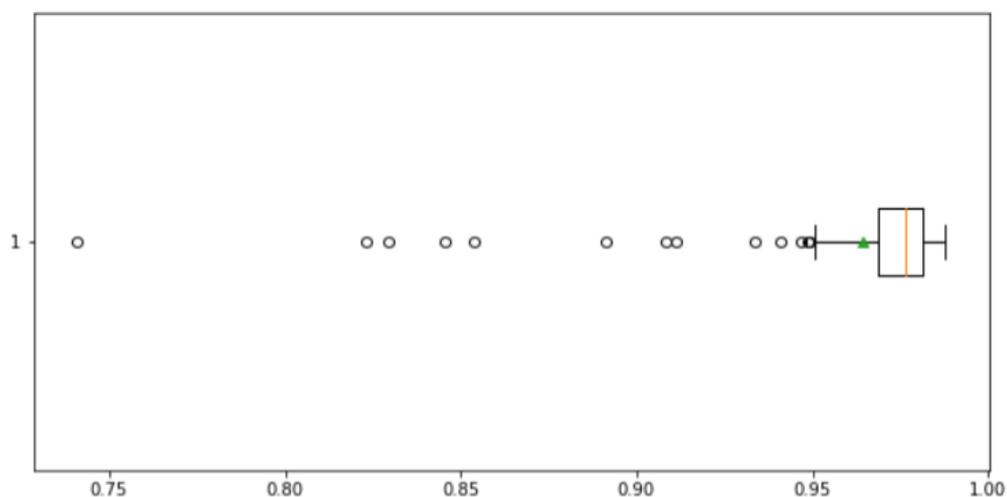
Consoante as precisões, ficou evidente que a segmentação obteve resultados compatíveis com o estado da arte para a métrica escolhida, porém, na base de TCs a segmentação apresentou resultados inferiores. Essa diferença entre resultados mostra que a quantidade de imagens utilizadas no treinamento possui influência na capacidade de generalização da rede.

Figura 13 – Representação do coeficiente de Dice em TC



Fonte: Autoria própria.

Figura 14 – Representação do coeficiente de Dice em ultrassonografias



Fonte: Autoria própria.

As figuras 13 e 14 mostram o *boxplot* do Dice para os 2 conjuntos de testes diferentes, com imagens que ainda não são de conhecimento da rede.

Na figura 14, o *boxplot* indica que os resultados ficaram acima de 95%, representando uma taxa muito alta. As bolinhas, que correspondem aos outliers, são as amostras que tiveram resultados que destoam da distribuição geral, ainda assim, o pior resultado está próximo a 75%. Já na figura 13, o *boxplot* aponta que apesar de a média para as imagens de tomografia estar em 0,83, 50% dos resultados obtidos nessa base está próximo de 0,89.

Os gráficos explicitaram que a capacidade de generalização dessa rede é bastante influenciada pela uniformidade e quantidade de imagens. Essa diferença se deve ao fato da base das tomografias computadorizadas conter poucas imagens e o rim variar de pessoa para pessoa na sua forma e textura, logo, faz-se necessário mais imagens e máscaras para a criação de um modelo com melhor capacidade de generalização.

5 CONCLUSÃO

Essa pesquisa propôs uma rede neural convolucional com arquitetura U-Net para realizar segmentação semântica de rins em imagens de tomografias computadorizadas e ultrassonografias. Considerando que a U-Net é capaz de fazer a localização prevendo a imagem pixel por pixel, o modelo em questão demonstrou a viabilidade da segmentação do rim através de aprendizado profundo. A abordagem utilizada fornece uma solução para segmentação em imagens renais quando o número de máscaras segmentadas manualmente é limitada, mas apesar dos resultados expressivos obtidos pelo modelo, ainda se faz necessário realizar testes em outros conjuntos para avaliar a capacidade de generalização do modelo desenvolvido.

Não foram localizadas na literatura pesquisas que fornecessem base de dados públicas com características semelhantes ao conjunto de dados utilizado, por esse motivo, dois experimentos foram realizados: um com imagens de TCs e outro com imagens de ultrassonografias. Ao todo, os resultados foram aceitáveis para a pesquisa e mostraram que os modelos de segmentação profunda são viáveis e capazes de segmentar com precisão os rins tanto em tomografias computadorizadas quanto em ultrassonografias. Por fim, foi comprovado a importância da U-Net na tarefa de segmentação de imagens, onde a abordagem obteve um coeficiente de Dice para o conjunto de testes de 0,83 para base de dados de TCs, e 0,96 para base de dados das ultrassonografias, que corresponde a uma área de 83% e 96% de sobreposição correta da imagem gerada em relação à máscara original, respectivamente, o que evidencia um resultado oportuno considerando a complexidade do problema a ser resolvido.

Vale salientar que, apesar de termos alcançado um bom desempenho na segmentação renal, todos os experimentos foram realizados no mesmo conjunto de dados, no qual os dados de treinamento foram divididos aleatoriamente e possuem distribuição idêntica aos dados de teste. Contudo, quando aplicado a ambientes clínicos reais, o modelo treinado precisa se adaptar a imagens de diferentes fornecedores e configurações. Isso tem o potencial de trazer uma distribuição diferente e, portanto, degradar o desempenho do modelo.

Segundo um levantamento da SBN (Sociedade Brasileira de Nefrologia), uma em cada dez pessoas corre o risco de ter problemas renais, a estimativa é que em 2040 a doença figure entre às cinco maiores causas de morte do mundo. Sendo assim, a contribuição dessa pesquisa é o desenvolvimento do segmentador, visto que a segmentação é uma das etapas mais trabalhosas do processamento de imagens, bem como uma das mais fundamentais. O uso desse segmentador pode facilitar o trabalho dos radiologistas e tornar o processo de diagnóstico mais veloz, tendo em vista o quanto a segmentação manual é custosa.

5.1 Trabalhos Futuros

Os resultados obtidos nessa pesquisa criam uma baseline para alicerçar novas iniciativas de aplicação da IA na nefrologia. Dentre os possíveis avanços ou rumos que essa pesquisa pode tomar, podem ser listados:

- A contagem de cálculos renais em imagens de tomografias computadorizadas: com o segmentador aqui proposto, é possível automatizar a segmentação dos rins e automatizar o processo de contagem dos cálculos, auxiliando o médico, agilizando processo e tornando-o mais eficiente;
- A segmentação e classificação de tumores renais em imagens de tomografias computadorizadas: de modo análogo ao tópico anterior, o segmentador pode servir de auxílio ao diagnóstico, não só segmentando o rim, mas o tumor também, o que facilita a aplicação de algoritmos de classificação;
- Avaliação do impacto dessa tecnologia na rotina e apoio nos possíveis diagnósticos: que consiste em realizar uma revisão sistemática da literatura embasando para fundamentar o impacto de um segmentador semântico na nefrologia.

REFERÊNCIAS

A BRIEF Guide to Convolutional Neural Network (CNN). [S. l.], 16 jul. 2019. Disponível em: <https://medium.com/nybles/a-brief-guide-to-convolutional-neural-network-cnn-642f47e88ed4>. Acesso em: 15 out. 2022.

CONVOLUTIONAL neural network. [S. l.], 16 dez. 2015. Disponível em: https://en.wikipedia.org/wiki/Convolutional_neural_network. Acesso em: 29 out. 2022.

CLASSIFICAÇÃO de imagens médicas: entendendo a rede neural convolucional (CNN). [S. l.], 16 fev. 2021. Disponível em: <https://www.imaios.com/br/Empresa/blog/Classificacao-de-imagens-medicas-entendendo-a-rede-neural-convolucional-CNN>. Acesso em: 9 nov. 2022.

DEEP Learning: CNN e aplicações. [S. l.], 28 abr. 2021. Disponível em: <https://carvalhodouglas1999.medium.com/deep-learning-cnn-e-aplica%C3%A7%C3%B5es-c25fdffa7089>. Acesso em: 8 nov. 2022.

Em 2040, doença renal crônica pode ser 5º maior causa de morte do mundo. Disponível em: <<https://www.uol.com.br/vivabem/noticias/redacao/2022/03/10/em-2040-doenca-renal-cronica-pode-ser-5-maior-causa-de-morte-do-mundo.htm>>. Acesso em: 18 nov. 2022.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning.** MIT Press. 2016.

GONZALEZ, R. C.; WOODS, R. C. **Processamento digital de imagens.** [S.l.]: Pearson Prentice Hall, 2010.

HAFEMANN, L. G. **An analysis of deep neural networks for texture classification.** 2014.

JANGIR, D. S. k. **Beginner's Guide for Convolutional Neural Network (CNN / ConvNets).** Towards Data Science, 2019. Disponível em: <https://towardsdatascience.com/beginners-guide-for-convolutional-neural-network-cnn-convnets-5a5e725ea581>. Acesso em: 15 nov. 2022.

JORDAN, Jeremy. **Convolutional Neural Networks.** Jeremy Jordan. 16 jul. 2017. Disponível em: <https://www.jeremyjordan.me/convolutional-neural-networks>. Acesso em: 10 fev. 2020.

LECUN, Yann et al. **Backpropagation Applied to Handwritten Zip Code Recognition.** Neural Computation. v. 1. n. 4. p. 541-551. 1989.

LECUN, Yann et al. **Gradient-based Learning Applied to Document Recognition**. Proceedings of the IEEE. v. 86. n. 11. p. 2278-2324. nov. 1998.

LECUN, Y.; BENGIO, Y.; HINTON, G. **Deep learning**. Nature, Nature Publishing Group, v. 521, n. 7553, p. 436, 2015.

LONG, J.; SHELHAMER, E.; DARRELL, T. **Fully convolutional networks for semantic segmentation**. In: Proceedings of the IEEE conference on computer vision and pattern recognition. [S.l.: s.n.], 2015. p. 3431–3440.

O QUE é Sequence-to-sequence em deep learning?. [S. l.], 10 abr. 2019. Disponível em: <https://medium.com/luisfredgs/o-que-%C3%A9-sequence-to-sequence-em-deep-learning-9f8857a423ca>. Acesso em: 9 nov. 2022.

PAN, Z. et al. **Deep learning segmentation and classification for urban village using a worldview satellite image based on U-net**. Remote Sensing, v. 12, n. 10, 1 maio 2020.

PLANCHE, B.; ANDRES, E. **Hands-On Computer Vision with TensorFlow 2**. [S.l.]: Packt Publishing Ltd, 2019. ISBN 978-1788830645.

RAWAT, W.; WANG, Z. **Deep convolutional neural networks for image classification: A comprehensive review**. Neural computation, MIT Press, v. 29, n. 9, p. 2352–2449, 2017.

REDES Neurais Convolucionais (CNN) - Arquitetura Explicada. [S. l.]: Dharmaraj, 1 jun. 2022. Disponível em: <https://medium.com/@draj0718/convolutional-neural-networks-cnn-architectures-explained-716fb197b243>. Acesso em: 13 out. 2022.

RONNEBERGER, O.; FISCHER, P.; BROX, T. **U-net: Convolutional networks for biomedical image segmentation**. In: SPRINGER. International Conference on Medical image computing and computer-assisted intervention. [S.l.], 2015. p. 234–241.

SULTANA, F.; SUFIAN, A.; DUTTA, P. **Evolution of image segmentation using deep convolutional neural network: a survey**. Knowledge-Based Systems, Elsevier, v. 201, p. 106062, 2020.

SRIVASTAVA, N.; HINTON, G.; KRIZHEVSKY, A.; SUTSKEVER, I.; SALAKHUTDINOV, R. **Dropout: a simple way to prevent neural networks from overfitting**. The journal of machine learning research, JMLR. org, v. 15, n. 1, p. 1929–1958, 2014.

SIMPLE Introduction to Convolutional Neural Networks. [S. I.], 26 fev. 2019. Disponível em: <https://towardsdatascience.com/simple-introduction-to-convolutional-neural-networks-cdf8d3077bac>. Acesso em: 10 nov. 2022.

SATELLITE imagery segmentation using U-NET. [S. I.], 20 set. 2022. Disponível em: https://medium.com/@Chinmay_Paranjape/satellite-imagery-segmentation-using-u-net-4ec7f265ddb8. Acesso em: 31 out. 2022.

SHORTEN, C.; KHOSHGOFTAAR, T. M. **A survey on image data augmentation for deep learning.** Journal of Big Data, Springer, v. 6, n. 1, p. 1–48, 2019.

YOU, K.; LONG, M.; WANG, J.; JORDAN, M. I. **How Does Learning Rate Decay Help Modern Neural Networks?.** 2019.