

# Relatório do Projeto Final de MC886

## MRI Classification and Segmentation for Brain Tumor Detection

Gabriel Braga Proença - 216116  
Lucca Costa Piccolotto Jordão - 182783

Gabriel Teston Vasconcelos - 197550  
Pedro Henrique Rodrigues de Araújo - 223382

## Resumo

Os tumores cerebrais causam milhares de mortes por ano, sendo um dos tipos mais letais de câncer. Entretanto, como toda doença, um diagnóstico precoce é fundamental para a efetividade do tratamento e redução das fatalidades. Dessa forma, a partir de um conjunto de dados composto por imagens em duas dimensões de ressonância magnética, desenvolveu-se uma arquitetura capaz de classificá-las e segmentá-las a fim de detectar regiões cerebrais afetadas por tumores. O modelo criado é baseado no conceito de self attention, utilizando de redes convolucionais para a análise das imagens médicas.

## Introdução

Os tumores cerebrais são doenças extremamente agressivas que ocorrem quando há um crescimento anormal de células em alguma estrutura dentro do crânio [1], acometendo adultos e também crianças. A taxa de sobrevivência em 5 anos para as pessoas com câncer cerebral é de aproximadamente 36% [2]. Apesar disso, o diagnóstico precoce é fundamental para o tratamento e aumento da expectativa de vida do paciente.

Nesse sentido, o uso de Imagens por Ressonância Magnética é a melhor maneira para se detectar um tumor. Porém, para fazer a análise do enorme volume de dados gerados pelas ressonâncias, as imagens são examinadas por radiologistas de forma manual, o que pode acarretar em erros no diagnóstico, já que os tumores são complexos e suas propriedades variam consideravelmente.

Desse modo, a aplicação de técnicas de Aprendizado de Máquina e Inteligência Artificial é de grande relevância para a minimização de erros humanos e, consequentemente, um diagnóstico de alta acurácia e precisão. Nesse intuito, o desenvolvimento do projeto em questão se dá em duas etapas: a classificação de uma imagem de ressonância magnética cerebral, identificando se há ou não um tumor, e a segmentação da imagem, detectando a região específica afetada pela doença.

## Trabalhos Relacionados

A utilização de redes neurais para análise de imagens biomédicas tornou-se muito comum na última década, suscitando em diversas abordagens para os diferentes desafios que surgiram na área. Mais especificamente, o uso de redes neurais

convolucionais, como observado em [3], onde uma rede neural convolucional foi usada como classificador para o mesmo problema abordado neste projeto.

Outro exemplo do uso de redes convolucionais para imagens biomédicas é a U-Net [4], uma arquitetura capaz de realizar a segmentação de imagens de microscopia celular. A arquitetura leva esse nome pelo formato visual, além disso está dividida em duas grandes etapas. A etapa de contração é formada de convoluções com o número de canais crescente e camadas de max pooling, restringindo a informação de “onde” e ampliando a informação de “o que” se trata o problema. Já a etapa de expansão é composta de convoluções transpostas com número de canais decrescente e concatenações de camadas anteriores para a formação de um mapa de segmentação de alta resolução.

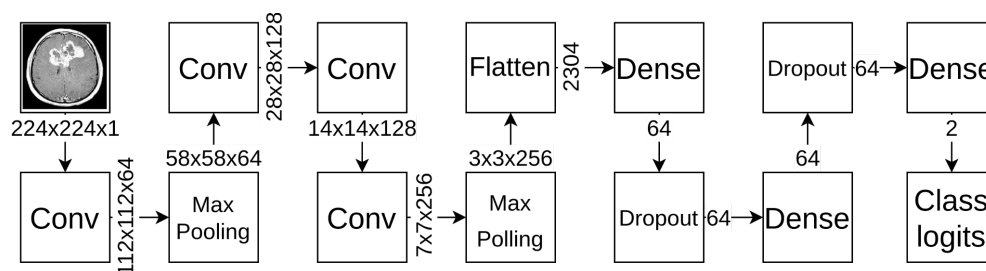
Assim como redes convolucionais tentam simular a percepção visual de humanos, existem arquiteturas que exploram os conceitos de atenção, apresentado durante a realização de uma tarefa arbitrária como a atribuição de um grau alto ou baixo de relevância para certas informações, dependendo do quão pertinente as mesmas são para a realização da atividade em questão. Esse conceito foi explorado em [5] com a proposição de uma arquitetura de redes convolucionais baseada em atenção para a classificação de imagens de raio-x de tórax para a detecção de diversas enfermidades.

## Método

### Dataset

O conjunto de dados utilizado no projeto foi obtido no Kaggle [6], contendo Imagens por Ressonância Magnética cuja fonte não conseguimos determinar. O dataset é dividido entre duas pastas, referentes às classes “no” e “yes”, isto é, uma das pastas contém ressonâncias que não apresentam tumor, enquanto a outra pasta contém somente ressonâncias que apresentam tumor cerebral. Cada uma delas contém 1500 samples, totalizando 3000 imagens. Para 801 imagens dentre as 1500 da classe afetada, temos as anotações das regiões afetadas por tumor, utilizadas para a criação das máscaras de segmentação (obtidas com a união de tais regiões, utilizando o OpenCV). Essas anotações possuem o tipo da região, podendo variar entre uma região elíptica, circular ou poligonal, e também suas respectivas coordenadas. Para as imagens da classe “não afetada”, fizemos nossas próprias anotações, sendo elas vazias, já que não há tumor. É importante ressaltar que as imagens sem máscaras de segmentação foram descartadas por não se adequarem ao problema proposto.

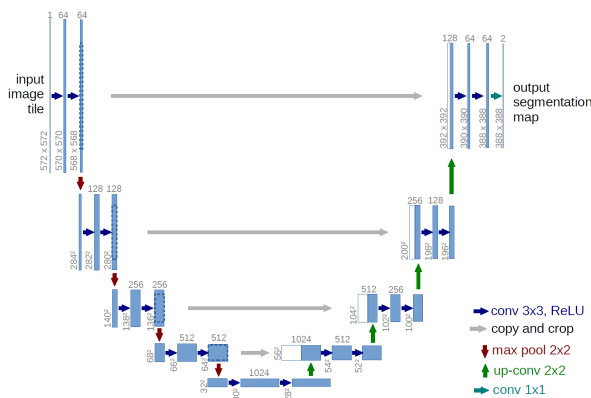
## Rede convolucional para classificação



A primeira arquitetura abordada pode ser vista acima. Tivemos como inspiração um notebook disponibilizado por um usuário [3] do Kaggle junto ao dataset. Essa arquitetura consiste em camadas convolucionais intercaladas com camadas de max pooling, reduzindo as dimensões de altura da imagem e aumentando o número de canais; por fim, temos uma camada de achatamento seguidas de camadas densas e camadas de *dropout* reduzindo o tamanho dos tensores até 2 na saída, onde pode ser aplicada a função *softmax* para encontrar a probabilidade da imagem de entrada pertencer a cada uma das duas classes.

Para o treinamento desse modelo foi utilizada uma saída direta da rede com a entropia cruzada como função de custo, capaz de comparar as probabilidades geradas pela rede com o resultado real esperado.

## Abordagem com U-Net

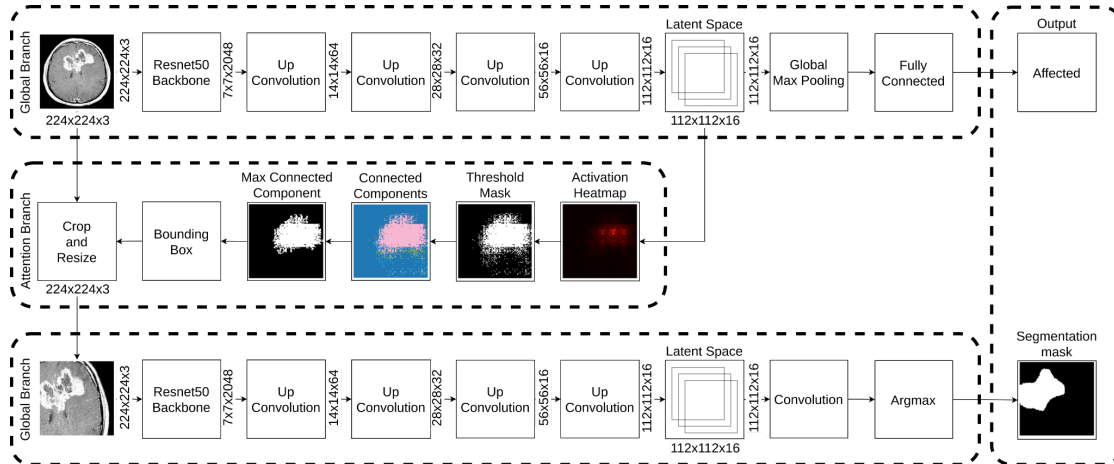


Junto ao notebook [3] desse usuário, encontramos indicativos do uso dessa arquitetura para o problema de segmentação referente a segunda etapa do projeto. Essa arquitetura baseia-se em comprimir a informação espacial e acumular a informação contextual, através de convoluções com número de canais crescentes e camadas de *max pooling*, para em um segundo momento recuperar a informação espacial, através de skip connections provenientes da etapa de

compressão, junto à informação contextual via camadas de *upconv* a fim de aumentar as dimensões da imagem, sendo capaz de gerar um mapa de segmentação de alta resolução.

Analogamente ao modelo anterior que deveria apontar a presença ou ausência de um tumor na imagem, esse modelo deve realizar a mesma tarefa, porém para cada pixel da imagem, usando também a entropia cruzada como função de custo. Ou seja, a saída da rede consiste em uma imagem de mesmas dimensões da imagem de entrada e dois canais, onde cada um representa a probabilidade da presença de uma determinada classe dado um pixel em uma certa posição.

## Abordagem baseada em hard self-attention



O modelo final fundamenta-se nos conceitos de *self-attention* e *hard-attention*, já que não é explicitamente treinado, mas sim baseado na relação entre a tarefa de classificação da imagem, por meio da ativação de neurônios em um espaço latente de uma camada intermediária. Além disso, o mecanismo de *hard-attention* bloqueia totalmente a parte da informação menos relevante através do crop realizado na imagem original.

Para isso, separamos nosso modelo em três componentes:

### Global Branch

A Global Branch tem a função de classificar as imagens (indicando a presença ou ausência de tumor). Nesse sentido, utilizamos uma ResNet50 com Transfer Learning para a extração de features, seguida de camadas de *upconv* para aumentar as dimensões do espaço latente para valores compatíveis à imagem de entrada. Ao fim, aplicamos uma camada de *GlobalMaxPooling*, extraíndo o valor máximo de cada canal deste espaço seguida de uma camada densa para a classificação.

### Attention Branch

Esse componente não treinável tem como objetivo recortar a imagem original na região de maior relevância para a classificação realizada pela Global Branch. Assim, a partir do espaço latente extraído, cujas dimensões são de 112x112x16, são aplicadas diferentes transformações.

A primeira delas obtém o valor absoluto de cada neurônio; a segunda é responsável por normalizar os valores por canal; a terceira tem a função de obter o valor máximo para cada pixel, criando assim um *heatmap* de 112x112.

A partir do *heatmap* gerado, formamos a threshold mask, composta por uns (1) nas posições onde o *heatmap* possui valor superior a um determinado valor de threshold estabelecido. Essa máscara é submetida a um algoritmo de componentes conexas, selecionando apenas a maior delas.

Por fim, extraímos a *bounding-box* dessa maior região conexa, utilizando suas coordenadas para recortar a região relevante da imagem original e redimensioná-la para o mesmo tamanho da imagem que a gerou.

## Local Branch

A Local Branch é responsável pela tarefa de segmentação da imagem. Para essa finalidade, esse componente utiliza a imagem recortada como entrada para a mesma arquitetura da Global Branch (ResNet50 e camadas de *upconv*). Porém, ao final, após formado o espaço latente, é aplicada uma camada de convolução resultando em uma imagem de dois canais com a mesma representação da saída da arquitetura da U-Net, também usando a entropia cruzada.

# Experimentos

## Rede convolucional para classificação

Para essa primeira abordagem tomamos como *baseline* um dado coletado de uma página do Hospital Sírio Libanês [7], na qual afirmava 95% de taxa de detecção (recall) a partir do mesmo tipo de imagem, isto é, imagens por ressonância magnética.

Como descrito anteriormente, essa arquitetura tem a tarefa de classificação das imagens. Para tanto, devemos definir suas métricas para avaliar os modelos gerados, optando pelo uso das seguintes métricas: acurácia, *recall*, precisão e *F1-score*.

Nessa tarefa devemos tentar minimizar a taxa de falsos negativos, já que representam pessoas acometidas por tumores mas que não receberam o diagnóstico, possibilitando à enfermidade alcançar um estágio mais avançado.

Começamos nossos testes separando as imagens e suas labels em conjuntos de treino e teste. Em seguida, o conjunto de treino foi utilizado para treinar K modelos no esquema de K-folds cross-validation [Apêndice], fornecendo a imagem como entrada do modelo e gerando um erro através da função de custo (entropia cruzada) e a label.

Ao fim dos experimentos obtivemos os seguintes resultados:

Metrics	Mean value between validation folds
Accuracy	97.63%
Recall	97.33%
Precision	98.08%
F1-Score	97.42%

## Abordagem com U-Net

Essa arquitetura exigiu que escolhêssemos uma nova métrica capaz de comparar a máscara de segmentação gerada pela rede e a máscara de segmentação real, de maneira que utilizamos a *IoU*, *Intersection over Union*, métrica que indica a razão da intersecção entre a área de segmentação predita e a área de segmentação real pela união dessas duas áreas.

Neste experimento utilizamos a mesma separação dos dados da abordagem anterior, mas ao invés de fornecer uma label para a função de custo, passamos a máscara de segmentação real em forma de imagem, gerada a partir da união das

regiões anotadas (provenientes do próprio dataset). O treino do modelo também foi efetuado usando o protocolo K-folds cross-validation.

Em um primeiro momento, não obtivemos resultados minimamente satisfatórios para a tarefa proposta. Assim, levantamos a hipótese de falta de dados; por ser uma arquitetura bem mais complexa, fizemos um processo de *data-augmentation* baseado em rotações aleatórias, aplicado às imagens originais e à máscara de segmentação. Os modelos gerados por essa etapa também não se mostraram muito eficientes, gerando apenas imagens com listras brancas (positivas) e pretas (negativas). Ao passo que o erro se atenuava durante o treinamento, esse resultado se mantinha.

Levantamos duas hipóteses para esse resultado — a primeira delas colocada antes de seguirmos para o desenvolvimento da arquitetura final e a segunda notada após começarmos a trabalhar no modelo baseado em self-attention.

A primeira hipótese está relacionada com a falta de dados para o treinamento do modelo a partir de um estado randômico. É preciso ressaltar que não usamos *transfer-learning* para essa etapa por motivos de complexidade da tarefa, já que o modelo fornecido para tanto encontrava-se em um formato diferente.

A segunda hipótese diz respeito a um bug na função `tf.keras.preprocessing.image.random_rotation` usada para realizar o *data-augmentation*, no qual criava as linhas pretas nas imagens rotacionadas. Nenhuma das hipóteses foi testada.

Dadas as hipóteses acima, o resultado final [\[Apêndice\]](#) apresentou cerca de 40% de IoU para a validação.

## Abordagem baseada em hard self-attention

O modelo final utilizou-se de todas as métricas apresentadas previamente, devido ao fato que a Global Branch trabalha na classificação, enquanto a Local Branch é responsável pela segmentação.

Separamos o conjunto de dados em treino, validação e teste nas seguintes proporções: 78%, 2% e 20%, respectivamente.

A rotina de treinamento foi dividida em quatro etapas: treino da cabeça de classificação da Global Branch; treino completo da Global Branch (incluindo a ResNet); treino da cabeça de segmentação da Local Branch; e por fim, treino completo da Local Branch (também incluindo a ResNet).

Em uma etapa intermediária entre a Global Branch e a Local Branch, foi determinado um threshold de forma empírica, analisando visualmente os resultados do *heatmap* e *threshold mask* da Attention Branch em alguns batches de treino.

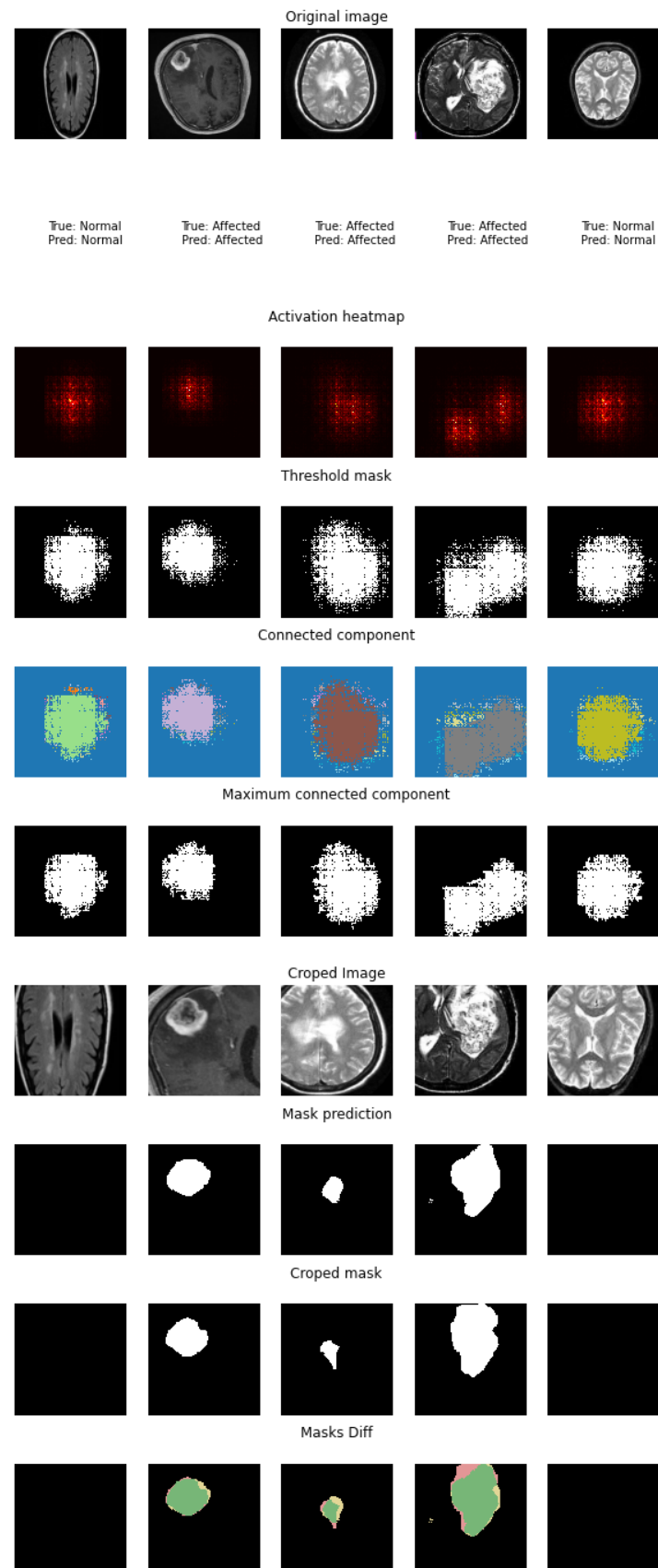
Os resultados obtidos pelo modelo final foram satisfatórios, visto que a baseline de recall é superada e, além disso, as regiões de segmentação estão visivelmente corretas.

Metrics	Train	Validation	Test
Accuracy	100%	99.25%	100%
Recall	98.80%	99.50%	98.80%
Precision	98.08%	98.88%	98.80%
F1-Score	98.80%	99.09%	98.80%
IoU	96.62%	92.03%	94.64%

# Conclusão

Os resultados obtidos foram satisfatórios – desenvolvemos um modelo com base em self-attention que, além de classificar uma imagem de ressonância magnética cerebral indicando a existência de um tumor, o modelo é capaz de realizar a segmentação da imagem na região afetada. Tanto para a tarefa de classificação quanto para a de segmentação, as métricas de acurácia, precisão, recall, F1-Score e IoU (Intersection over Union) tiveram resultados excelentes.

Por último, pensando em aperfeiçoar o modelo no futuro, o próximo passo é a classificação e segmentação de tumores cerebrais em imagens 3D de ressonância magnética, não limitando-se somente a fatias de duas dimensões.





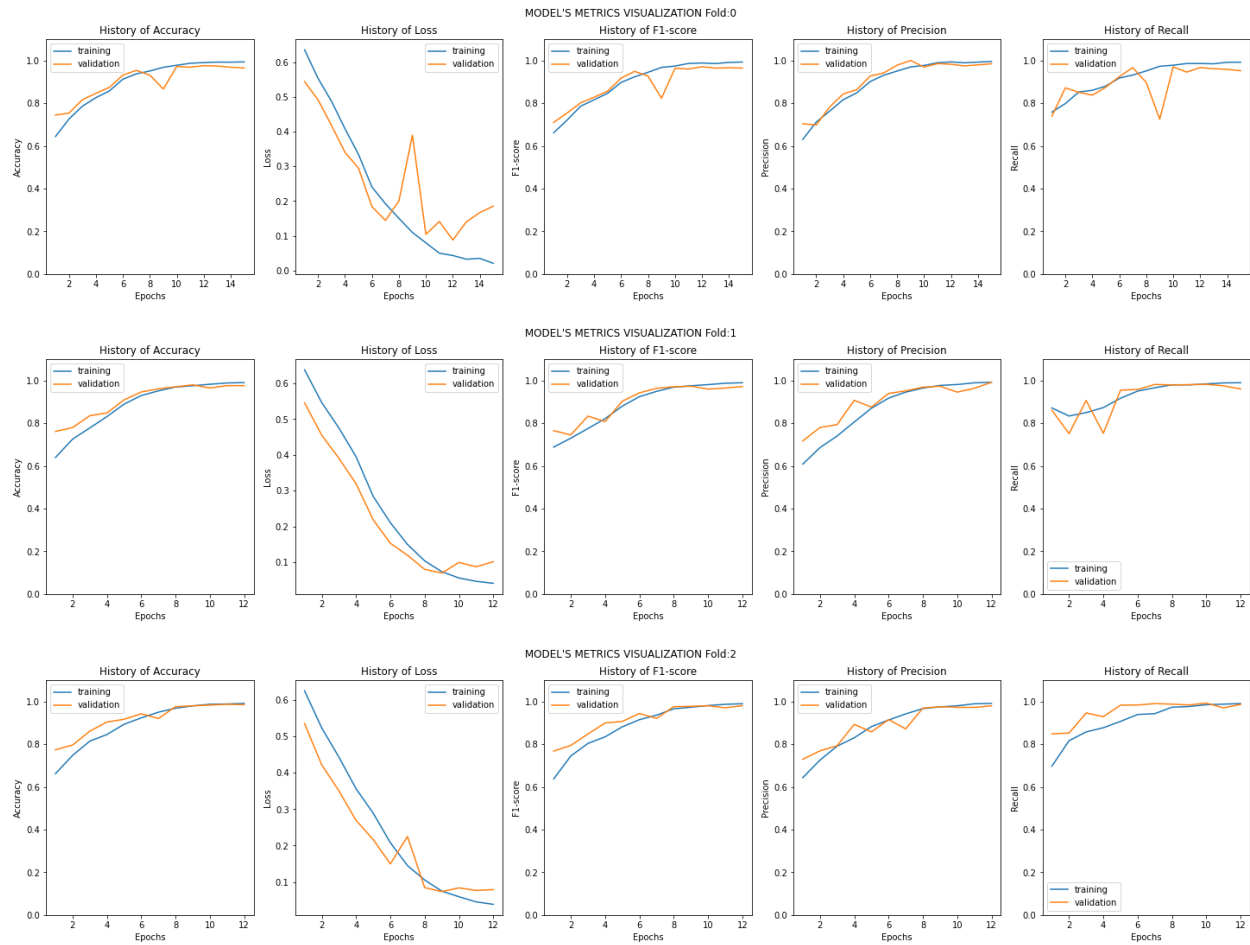
# Referências

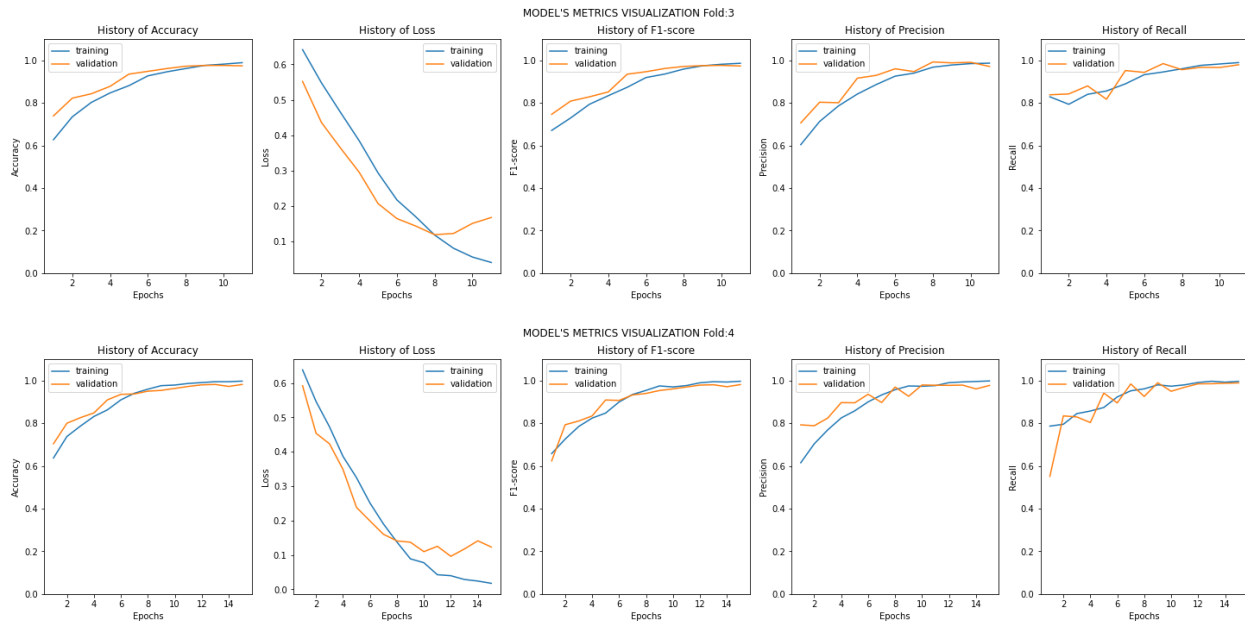
- [1] “O que são tumores cerebrais e principais sintomas” - Disponível em:  
<https://www.funcionalita.com.br/o-que-sao-tumores-cerebrais-e-principais-sintomas>
- [2] “Brain Tumor Statistics” - Disponível em:  
<https://www.cancer.net/cancer-types/brain-tumor/statistics>
- [3] Brain Tumor Detection Notebook by Marto24 on Kaggle - Disponível em:  
<https://www.kaggle.com/marto24/brain-tumor-detection/notebook>
- [4] U-Net Architecture - Disponível em:  
<https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net>
- [5] “Diagnose like a Radiologist: Attention Guided Convolutional Neural Network for Thorax Disease Classification” - Disponível em:  
<https://arxiv.org/pdf/1801.09927.pdf>
- [6] Brain Tumor Br35H Dataset - Disponível em:  
<https://www.kaggle.com/ahmedhamada0/brain-tumor-detection>
- [7] Hospital Sírio Libanês: Página explicativa sobre Câncer cerebral - Disponível em:  
<https://www.hospitalsiriolibanes.org.br/hospital/especialidades/centro-oncologia/cerebral/Paginas/diagnosticos.aspx>



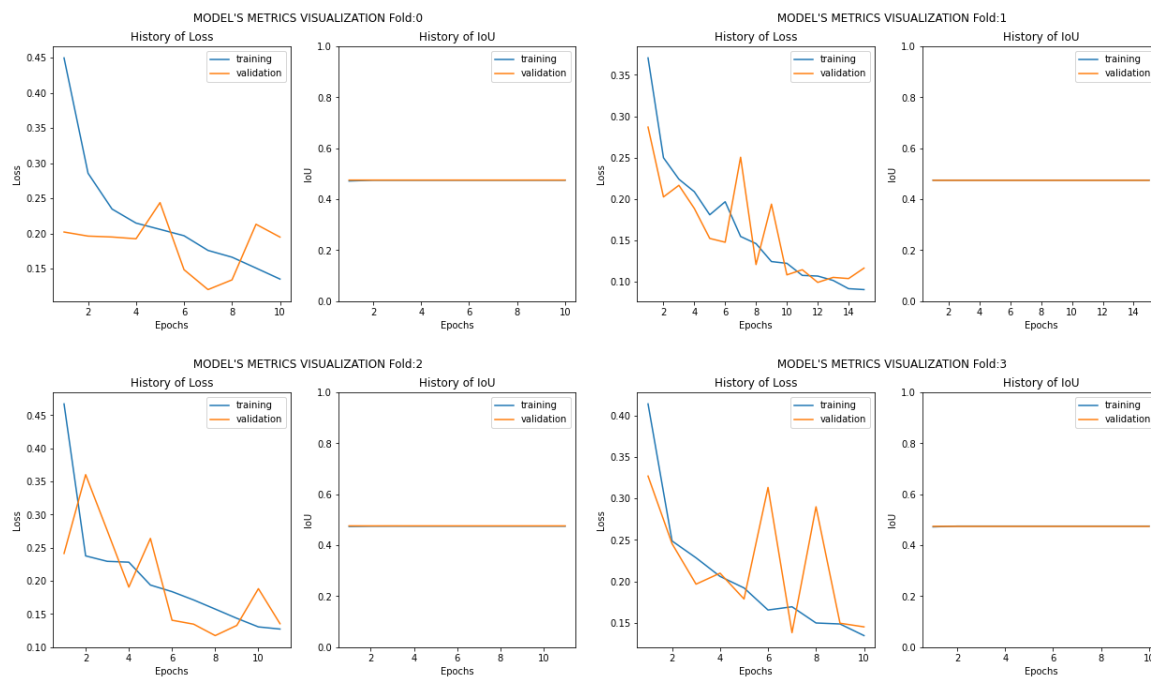
# Apêndice

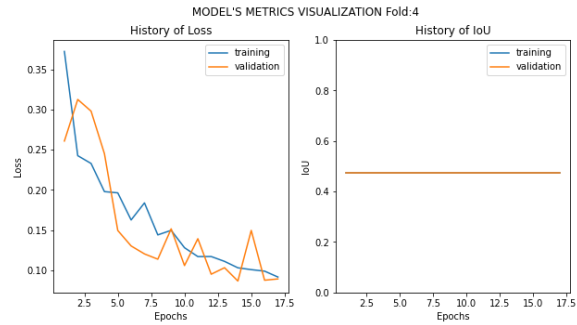
## Rede convolucional para classificação





## Abordagem com U-Net





## Abordagem baseada em hard self-attention

