

**UNIVERSIDADE DE SÃO PAULO**

Instituto de Ciências Matemáticas e de Computação

## Detecção de Hemorragia em Imagens de Tomografia usando Algoritmos baseados em Gradient Boosting

**Khalil Carsten do Nascimento**

Monografia - MBA em Ciência de Dados

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: \_\_\_\_\_

**Khalil Carsten do Nascimento**

## **Deteccção de Hemorragia em Imagens de Tomografia usando Algoritmos baseados em Gradient Boosting**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados.

Concentration area: Ciência de Dado

Orientadora: Prof. Dr. Afonso Paiva Neto

**Versão original**

**São Carlos**

**2021**

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi, ICMC/USP, com os dados fornecidos pelo(a) autor(a)

S856m	Nascimento, Khalil do Detecção de Hemorragia em Imagens de Tomografia usando Algoritmos baseados em Gradient Boosting / Khalil Carsten do Nascimento ; Orientador Afonso Paiva Neto. – São Carlos, 2021. 54 p. : il. (algumas color.) ; 30 cm.  Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados. – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2021.  1. LaTeX. 2. abnTeX. 3. Classe USPSC. 4. Editoração de texto. 5. Normalização da documentação. 6. Tese. 7. Dissertação. 8. Documentos (elaboração). 9. Documentos eletrônicos. I. Neto, Afonso Paiva, orient. II. Título.
-------	--

**Khalil Carsten do Nascimento**

**Deteccção de Hemorragia em Imagens de Tomografia  
usando Algoritmos baseados em Gradient Boosting**

Trabalho de conclusão de curso apresentado ao Centro de Ciências Matemáticas Aplicadas à Indústria do Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, como parte dos requisitos para conclusão do MBA em Ciências de Dados.

Área de concentração: Ciência de Dados

Advisor: Prof. Dr. Afonso Paiva Neto

**Original version**

**São Carlos**

**2021**

## **AGRADECIMENTOS**

Primeiramente aos meus pais, Benivaldo do Nascimento e Germana Carsten, por priorizarem minha educação frente a qualquer diversidade da vida.

Ao meu orientador, professor Afonso Paiva Neto, pelo conhecimento e mentoria acompanhado de muita solicitude e confiança.

Aos meus Irmãos, Carmel Carsten e Caio Lemos, por serem meus companheiros de vida e por me darem forças no meu dia-a-dia.



## ABSTRACT

Nascimento, K. **Hemorrhage Detection in Tomography Images Using Algorithms Based on Gradient Boosting**. 2021. 54p. Monografia - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.

The detection of cerebral hemorrhages in computed tomography images can be very useful in aiding the daily work of medical professionals in hospitals. This study aimed to be a benchmark with an exploratory approach using Haralick features, a traditional and widely used, together with Gradient Boosting algorithms. The data used was from the public RSNA database, also used in the study by Hssayeni *et al.* (2020). Ten experimental tests were carried out using K-fold cross-validation to ensure greater reliability in the performance of the final model. The average of the achieved metrics were 75.80% accuracy, 74,53% F1 score, 77,69% precision, 71,70% sensitivity, and 79,85% specificity. It is important to mention that the availability and reproducibility of this study stands out compared to other similar ones, in which neither the database nor the codes are provided. In addition, analyses were carried out on the decision-making impact that Haralick features obtained alongside the Gradient Boosting model.

**Keywords:** Haralick. Gradient Boosting. hemorrhage. Shapley Values. Computed Tomography. Coursework Final Project





## RESUMO

Nascimento, K. **Detecção de Hemorragia em Imagens de Tomografia usando Algoritmos baseados em Gradient Boosting**. 2021. 54p. Monografia - Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, 2021.

A identificação de hemorragias cerebrais em imagens de tomografia computadorizada pode ser muito útil para ajudar os médicos em hospitais no seu dia-a-dia. Este estudo tem como objetivo ser um ponto de referência com uma abordagem exploratória utilizando as características de Haralick, um método tradicional e amplamente utilizado, em conjunto com algoritmos de *Gradient Boosting*. Os dados utilizados foram da base pública RSNA, também usados no estudo de Hssayeni *et al.* (2020). Foram realizados 10 testes experimentais usando validação cruzada K-fold para garantir maior confiabilidade no desempenho do modelo final. A média das métricas alcançadas foram de 75.80% de acurácia, 74.53% de Escore F1, 77.69% de precisão, 71.70% de sensibilidade e 79.85% de especificidade. É importante mencionar que a disponibilidade e reprodutibilidade deste estudo se destaca em comparação com outros semelhantes, nos quais nem a base de dados nem os códigos são fornecidos. Além disso, foram realizadas análises do impacto das características de Haralick no modelo de Gradient Boosting.

**Palavras-chave:** Haralick. Gradient Boosting. Hemorragia. Valores de Shapley. Tomografia Computadorizada. Trabalho de conclusão de curso (TCC).



## LISTA DE FIGURAS

Figura 1 – Exemplos de hemorragias intracerebrais aparentes em imagens TC. . . . .	20
Figura 2 – Representação formal de uma imagem digital. . . . .	29
Figura 3 – Extração do histograma de uma imagem CT. . . . .	30
Figura 4 – Duas imagens quadradas com um triângulo equilátero desenhado. As imagens possuem padrões de cores diferentes porém seriam a mesma imagem para uma análise de primeira ordem. . . . .	31
Figura 5 – Exemplo de ângulo em relação ao pixel de referência. . . . .	32
Figura 6 – Geração da matriz de coocorrência de uma imagem digital com tons de cinza escalados entre 0 e 3. . . . .	32
Figura 7 – Matriz de probabilidade gerada a partir da GLCM . . . . .	33
Figura 8 – Tabela de Treinamento a esquerda contendo os sintomas e a positividade em relação ao Covid no paciente. A direita é apresentado os dados de teste no qual não vem acompanhado pelos dados de Covid, tendo que assim ser definido pelo modelo. . . . .	35
Figura 9 – Árvore de decisão gerada para identificar a presença de vírus no paciente segundo os sintomas apresentados. . . . .	36
Figura 10 – Representação do aumento da precisão do modelo de gradiente boosting a medida que mais árvores vão sendo adicionadas ao modelo final. . . . .	38
Figura 11 – Diagrama Metodológico . . . . .	40
Figura 12 – Apresentação da tomografia computadorizada à esquerda juntamente com a máscara. A direita encontra-se a sobreposição de ambas. . . . .	41
Figura 13 – Exemplo do recorte feito pelo algoritmo de <i>Threshold</i> com o intuito de eliminar as bordas pretas. . . . .	41
Figura 14 – Gráfico de caixas que apresenta a distribuição das acurácias obtidas nos 10 ensaios do experimento. . . . .	46
Figura 15 – Gráfico de caixas que apresenta a distribuição das métricas de sensibili- dade, precisão e Escore F1 obtidas nos 10 ensaios do experimento. . . . .	46
Figura 16 – Recortes de imagens TC que passaram pela predição do melhor modelo <i>Gradient Boosting</i> desenvolvido no experimento. A cor rosa representa a hemorragia enquanto a cor azul claro a predição positiva para hemorragia naquela área. . . . .	48
Figura 17 – Valores de Shapley demonstram a importância de cada característica na decisão final da classificação . . . . .	50



## LISTA DE TABELAS

Tabela 1 – Compilado de artigos revisados na Seção 2. . . . .	27
Tabela 2 – Valores de distância e ângulos para as características de (HARALICK; SHANMUGAM; DINSTEIN, 1973) . . . . .	42
Tabela 3 – Tabela de métricas adquiridas pelo modelo <i>Catboost</i> durante a fase de treinamento e teste. . . . .	47



## LISTA DE ABREVIATURAS E SIGLAS

DAC	Diagnósticos Asistidos por Computado
HI	Hemorragia Intracerebral
TC	Tomografia Computadorizada
SVM	<i>Support Vector Machine</i>
MLP	<i>Multilayer Perceptron</i>
MDRLSE	<i>Distance Regularized Level set Evolution</i>
GLCM	<i>Gray Level Cooccurrence Matrix</i>





## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	<b>19</b>
1.1	Objetivos Gerais e Específicos	20
1.2	Estrutura do Trabalho	20
<b>2</b>	<b>TRABALHOS RELACIONADOS</b>	<b>23</b>
<b>3</b>	<b>FUNDAMENTACAO TEORICA</b>	<b>29</b>
3.1	Processamento de Imagem	29
3.2	Características de Textura	30
3.2.1	Abordagem Estatísticas	30
3.2.1.1	Abordagem Estatística de Primeira Ordem	30
3.2.1.2	Abordagem Estatística de Segunda Ordem	31
3.2.1.3	Características de Haralick	34
3.3	Aprendizagem de Máquina	34
3.4	Árvores de Decisão	36
3.5	Gradiente Boosting	37
<b>4</b>	<b>METODOLOGIA</b>	<b>39</b>
4.1	Visão Geral	39
4.2	Base de Dados	40
4.3	Pré-processamento dos Dados	40
4.4	Escolha de Parâmetros nas Características de Textura	41
4.5	Treinamento e Validação	42
4.6	Métricas de avaliação	42
<b>5</b>	<b>RESULTADOS</b>	<b>45</b>
5.1	Experimento	45
5.1.1	Hiper-Parâmetros	45
5.1.2	Treinamento e Métricas	46
5.1.3	Importância das Características	49
5.1.4	Considerações Finais	49
<b>6</b>	<b>CONCLUSÃO</b>	<b>51</b>
	<b>REFERÊNCIAS</b>	<b>53</b>



## 1 INTRODUÇÃO

Recentemente, *diagnósticos assistidos por computador* (DAC) se tornaram parte da rotina clínica em hospitais e centros de tratamento. Isso indica uma utilidade visível dos algoritmos computacionais em um rápido e preciso diagnóstico médico em pacientes com anormalidades. A prática de utilização de computadores no meio médico se inicia na década de 60 (LODWICK *et al.*, 1963) utilizando relatórios e leitura de sensores e se desenvolve até os dias de hoje com a automação de diagnósticos e indicações de tratamentos (DOI, 2007).

Dentre as áreas de DAC mais pesquisadas destaca-se a de análise imagens médicas. Com o avanço dos exames de imagem como Raio-X, Ressonâncias Magnéticas e Tomografias Computadorizada mais oportunidades surgiram para aplicação de técnicas de visão computacional com o fim de classificar doenças representadas em tais exames. Tendo isso em vista, o médico através do auxílio dessas soluções computacionais é capaz de diagnosticar em menor tempo e com menor custo uma anormalidade no paciente. (KONONENKO, 2001).

Esse trabalho se dedica a classificar a *Hemorragia Intracerebral* (HI), um evento médico com alta taxa de mortalidade sendo uma das formas mais agressivas de derrame. Entretanto, a detecção precoce dessas ocorrências juntamente com tratamentos intensivos pode acarretar em bons resultados de melhora nos pacientes (GOLDSTEIN; GILSON, 2011). A forma mais usada da detecção de tal hemorragia é a *tomografia computadorizada* (TC) capaz de capturar imagens de Raio-X da região do crânio em alta definição. A utilização rápida e eficiente dessa tecnologia no diagnóstico de Hemorragia Intracerebral pode ter muitos benefícios no dia a dia clínico (KIDWELL; WINTERMARK, 2008). Devido a isso, algoritmos de aprendizagem de máquina vem sendo amplamente usados para fins médicos e auxiliar no diagnóstico de anormalidades. Consequentemente surge a possibilidade de utilizar métodos de classificação e segmentação para identificar tipos de HI em Tomografias Computadorizadas.

Com o intuito de explorar métodos mais rasos de aprendizagem de máquina, esse trabalho se propõe a reproduzir experimentos de segmentação de HI. Dados médicos corretamente rotulados e em grande volume são raros de se encontrar disponíveis para o público o que torna modelos robustos como aprendizagem profunda menos eficazes na generalização dessas bases. Dessa forma, esse trabalho se propõe a reproduzir métodos tradicionais de extração de características texturais juntamente com a utilização de versões mais modernas de algoritmos rasos como *gradient boosting*.

## 1.1 Objetivos Gerais e Específicos

A classificação de Hemorragia Intracerebral é uma tarefa complexa e necessita de sofisticados métodos para alcançar um bom desempenho. Algoritmos de aprendizagem de máquina mais robustos, como redes neurais profundas, exigem muito poder de processamento e uma grande quantidade de dados para alcançarem uma boa generalização nas predições. Porém, os artigos mais modernos revisados na Seção 2 apresentam resultados com dados insuficientes para os modelos suportarem a variância de imagens médicas. Além disso, em sua maioria não fornecem a base de dados utilizada para reprodutibilidade tornando difícil a comparação dos resultados. Tendo isso em vista, esse trabalho foca na classificação de regiões de tomografia computadorizada com o fim de encontrar identificar a posição da hemorragia na tomografia analisada. A Figura 1 Apresenta alguns exemplos de hemorragias identificadas em imagens TC.

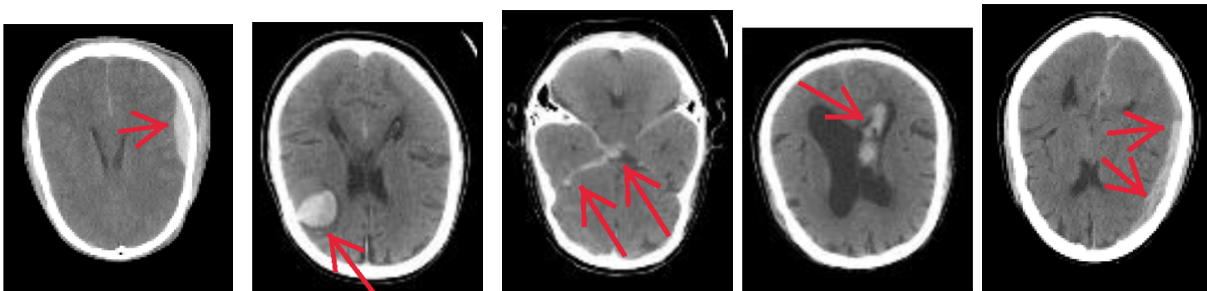


Figura 1 – Exemplos de hemorragias intracerebrais aparentes em imagens TC.

Fonte: (HSSAYENI *et al.*, 2020)

Os objetivos específicos dessa monografia são:

1. Elaborar uma base de pesquisa com revisão de literatura juntamente com um compilado das metodologias mais utilizadas;
2. Aplicar a metodologia escolhida em um conjunto de dados público;
3. Avaliar os resultados e possíveis melhorias no processamento das características e comparar com trabalhos atuais quanto ao desempenho e robustez do modelo *Gradient Boosting*.
4. Fornecer um *benchmark* para reprodução e análise do modelo e resultados obtidos.

## 1.2 Estrutura do Trabalho

Este trabalho está dividido em 5 principais capítulos. Em sequência a introdução, no Capítulo 2 será feito um compilado de trabalhos relacionado que realizaram estudos ao redor de classificação e segmentação de hemorragias intracerebrais. No Capítulo 3 são

abordados alguns conceitos essenciais pra o entendimento do trabalho e das metodologias utilizadas. A seguir, o Capítulo 4 apresenta a metodologia empregada além do algoritmo e características usados no experimento. Logo depois, no Capítulo 5, são descritos mais detalhadamente o experimento, os resultados e suas análises. Por fim, no Capítulo 6 apresenta-se um resumo do trabalho assim algumas observações de melhorias e trabalhos futuros.



## 2 TRABALHOS RELACIONADOS

Este capítulo contempla uma revisão de literatura sobre aplicação de algoritmos de aprendizagem de máquina para detecção de enfermidades em imagens médicas. É feita uma breve comparação entre os diversos métodos e abordagens visando agrupar as contribuições acadêmicas ao redor do problema proposto.

Como primeiro trabalho analisado, Chang *et al.* (2003) realizaram experimentos utilizando imagens de ultrassons de mamas a fim de encontrar sinais de câncer. Foram utilizados 250 imagens, com 8 bits por pixel, de pacientes patologicamente comprovados com tumores malignos e benignos. A extração de característica foi aplicada usando o método de coeficientes de auto covariância semelhante ao método de autocorrelação. Destaca-se que por mais que a autocorrelação seja um bom extrator de textura também possui a desvantagem de ser sensível ao brilho da imagem, podendo duas imagens possuírem a mesma textura mas valores diferentes para a característica. Dessa forma, o coeficiente de auto covariância contorna o problema do brilho e garante um cálculo menos ruidoso. Como modelos de classificação, foram escolhidos uma Máquina de Vetor de Suporte – do inglês *Support Vector Machine (SVM)* – com kernel de base radial e um Perceptron de Múltiplas Camadas – do inglês *Multilayer Perceptron (MLP)*. Como resultados a SVM se destacou conseguindo 86.60% de acurácia, 95.45% de sensibilidade e 77.86% de especificidade.

Padma and Sukanesh (2011) classifica imagens de hemorragias cerebrais utilizando um processo de extração de características de textura, propostas por Haralick, Shanmugam and Dinstein (1973), acompanhado do método estatístico Decomposição *Wavelet* Discreta (DWD). O autor acredita que essa abordagem permite uma maior ênfase nas particularidades das áreas hemorrágicas derivada das várias camadas geradas pelo método DWD. Também é utilizado um algoritmo genético (GA do inglês *genetic algorithm*) para realizar a seleção do melhor conjunto de características. O experimento foi efetuado com 120 imagens e um modelo SVM. Foi comparado o desempenho de três diferentes *kernels*, linear, polinomial e gaussiano utilizando características com e sem as transformações do DWD. Os resultados demonstraram uma melhora na acurácia com a adição do método DWD em comparação com as características de textura propostas em trabalhos anteriores alcançando 96.07% de sensibilidade, 95.91% de especificidade e 96% de acurácia.

Ramteke and Monali (2012) compara um modelo base de K-Vizinhos-mais-próximos – do inglês *K-nearest neighbors (KNN)* – com dois modelos SVM alternando entre o kernel linear e a função de base radial. Como entrada para o modelo é usado características de textura da imagem propostas por Haralick, Shanmugam and Dinstein (1973). São elas: contraste de variações locais de cinza, Homogeneidade, correlação entre os pixel próximos, entropia e homogeneidade local. Tais características foram retiradas de 51

imagens TC classificadas entre imagens normais e anormais de escaneamento do cérebro. Como resultado do experimento o KNN se destacou mostrando resultados superiores aos das SVM conseguindo 80% de acurácia, 81% especificidade e 80% de sensibilidade.

Li *et al.* (2012) aborda o problema de detecção de hemorragia subaracnoidea, um tipo de hemorragia que ocorre entre a camada interna do cérebro e a camada intermediária, mais conhecida como meninge. Os autores utilizam um algoritmo Bayesiano para classificar se um pixel pertence ou não a região da hemorragia, resumindo-se a um problema de segmentação. Como principal característica do pixel foi utilizada um vetor de distância entre o pixel e a marcação da hemorragia, formando assim um vetor de alta dimensionalidade como entrada para o algoritmo. A validação do método proposto foi feita comparando com imagens manualmente segmentadas com as predições do modelo. O algoritmo alcançou uma sensibilidade de 100% e 92% de especificidade. Além disso, o autor comenta dois fatores que os fez optar por características de distância em relação ao pixel no lugar de uma abordagem usando escala monocromática: (1) a grande variação de cinza e o baixo contraste em TC cerebrais; (2) e a considerável variação da área cerebral analisada entre os pacientes. Dessa forma, os autores mostram a possibilidade de elaborar algoritmos de auxílio de diagnósticos para hemorragias cerebrais.

Shahangian and Pourghassem (2016) apresentam uma sequências de características e metodologias para segmentação e classificação de quatro tipos de hemorragia cerebral: Epidural, Subdural, Subaracnoidea e Intraventricular. Esse trabalho se destaca por não utilizar segmentação manual e sim uma segmentação automática através de um método chamado *Distance Regularized Level set Evolution* (MDRLSE), uma ferramenta matemática interativa capaz de isolar o pico de um função. Dessa forma, é possível abstrair um histograma de uma imagem TC para uma função de picos identificando assim as hemorragias em locais de grande amplitude. Após a extração do suposto segmento hemorrágico, informações como formato, área, posição no cérebro, homogeneidade da área, razão do perímetro pela área e outras características com respeito ao formato. A união de todas esses dados formam um histograma que representa um protótipo de uma das classes de hemorragia cerebral. Tal histograma foi chamado pelos autores de Histograma Monocromático Ponderado (HMP). Além do HMP, outras características são extraídas das imagens e são inicialmente passadas em um modelo hierárquico. Segundo os autores, as classes hemorrágicas possuem muitas sobreposições, por consequência, optaram por executar antes um separador determinístico entre as classes Intraventricular e o resto. Desse modo, os outros três tipos de hemorragia restantes são classificados por um segundo modelo usando uma SVM. O primeiro modelo separador alcançou uma acurácia de 92.46% seguido por 94.13% do segundo modelo SVM.

Adentrando em métodos mais modernos, Phong *et al.* (2017) comparam três arquiteturas de redes neurais convolucionais *LeNet*, *GoogLeNet* e *Inception-ResNet* na



classificação de imagens com e sem sinais de hemorragia cerebral. A base de dados foi retirada de 115 hospitais totalizando 100 casos de pacientes. Foi efetuado o enriquecimento dessas imagens utilizando técnicas de rotação e ajuste de brilho. Ao final, foram geradas a base de treino e teste foi composta por 1700 imagens. Todas as redes foram treinadas por 4000 mil épocas utilizando entropia cruzada como função de perda. Na fase de testes a *LeNet* de destacou alcançando 100% de revocação, 99.3% de precisão, obtendo um valor superior de 1% em comparação a *GoogLeNET* e *Inception-ResNET*. Porém, o autor reforça a diferença considerável de tempo de treino da *LeNET* sendo até 9 vezes mais lento que as outras.

Seguindo os avanços da aprendizagem profunda, Majumdar *et al.* (2018) utilizam uma *U-Net* para segmentar entre 6 classes sendo 5 hemorragias cerebrais e uma para casos sem anormalidades. Foram utilizadas um total de 134 Tomografias computadorizadas onde 88 possuíam alguma hemorragia e 46 normais. Além disso, os dados foram enriquecidos com rotações aleatórias. Para a anotação foram acionados 2 especialistas em radiologia com o acompanhamento de um engenheiro. Os autores propõem um pós processo onde é comparado classificações de cortes adjacentes da mesma imagem. Desse modo, uma região somente é classificada como hemorrágica se em ambas forem detectadas hemorragias e se as unidades de *Hounsfield* (escala de radio-densidade utilizadas para interpretar imagens TC) forem maiores que 100. Essa abordagem acarretou em um especificidade de 98% e 81% de sensibilidade.

Adentrando na área de redes neurais recorrentes, Burduja, Ionescu and Verga (2020) abordam o problema de detecção da hemorragia cerebral utilizando uma união da arquitetura *ResNeXt-101*, uma rede neural profunda pré-treinada, e uma rede *Long Short-Term Memory* (LSTM) bidirecional. Os autores encaram o problema como uma série temporal, tendo em vista que uma TC é representada por vários recortes 2D que juntos forma uma imagem 3D do cérebro. Se cada recorte for inserido em sequência o problema se transforma na descoberta de padrões em uma série temporal das imagens. Assim sendo, as imagens são inicialmente passadas pela *REsNExt-101* produzindo um vetor de características de 2048 dimensões para cada recorte da imagem 3D do TC. Tal vetor multiplicado pela quantidade de recortes pode gerar vetores ainda maiores de mais de 100 mil dimensões. Visando otimizar essa entrada para a LSTM foi utilizado o algoritmo PCA reduzindo a dimensionalidade da entrada para 120 componentes. Com as características extraídas pela LSTM sua saída é concatenada com a saída da *REsNExt-101* e passada por uma camada *softmax* classificando a saída entre as 6 classes de hemorragia cerebral. Ao final o modelo resultou em 97% de acurácia, 96% de sensibilidade e 0.98 especificidade. Após isso foi validado e comparado com 3 radiologistas, ficando com resultado melhores, porém não distantes do melhor especialista. Os autores concluem afirmando que os médicos que participaram da rotulação dos dados confirmam a eficiência do sistema e que poderia ser de grande suporte em um ambiente clínico.

Como último trabalho, Hssayeni *et al.* (2020) fazem contribuições com uma revisão de 27 artigos, a disponibilização de uma base dados de TC de hemorragias intracerebrais e uma metodologia de segmentação baseada em redes neurais profundas. O trabalho se compromete a segmentar seis classes de imagens sendo cinco tipos de hemorragia e uma sem sinais de hemorragia. Um modelo estado da arte de rede neurais profundas a *U-Net* foi utilizada para segmentação sendo treinado em 150 épocas com 318 recortes de hemorragias intracraniais. Os resultados finais foram de 97.28% sensibilidade e 50.4% de especificidade.

Apresentado os trabalhos da revisão bibliográfica, fica nítida a dificuldade de adquirir volumosas bases de dados para treino. Pode-se observar na Tabela 1 que a maior quantidade de imagens utilizadas foi de 625 imagens no trabalho de (SHAHANGIAN; POURGHASSEM, 2016). Esse empecilho no encontro de bases ricas e bem rotuladas abre uma brecha no aprendizado profundo para imagens médicas criando uma ineficiência na generalização desses modelos, como demonstrado na especificidade final de (HSSAYENI *et al.*, 2020). Tendo isso em vista, esse trabalho visa reproduzir características propostas por (HARALICK; SHANMUGAM; DINSTEN, 1973) e (SHAHANGIAN; POURGHASSEM, 2016) utilizando modelos modernos de classificação rasa como *Gradient Boosting* (DOROGUSH; ERSHOV; GULIN, 2018). Dessa forma é esperado que seja superado o problema da quantidade de dados uma vez que a engenharia de características não será responsabilidade do modelo.

Tabela 1 – Compilado de artigos revisados na Seção 2.

Referência	Base de Dados	Método	Resultados
Chang <i>et al.</i> (2003)	250 Imagens de Ultrassons	SVM, MLP	86.60% de sensibilidade, 95.91% de especificidade, 96% de acurácia.
Padma and Sukanesh (2011)	120 Imagens TC	SVM, AG, DWD	96,07% de sensibilidade, 95,91% de especificidade, 96% de acurácia
Ramteke and Monali (2012)	51 Imagens TC	SVM, KNN	80% de acurácia, 81% de especificidade, 80% de sensibilidade
Li <i>et al.</i> (2012)	129 pacientes de 5 hospitais da China	Algoritmo Bayesiano	100% de sensibilidade, 92% de especificidade
Shahangian and Pourghassem (2016)	625 imagens JPEG 128x128	SVM, AdaBoost	Primeiro modelo: 92.46% de Acurácia Segundo modelo: 94.13% de Acurácia
Phong <i>et al.</i> (2017)	115 hospitais totalizando 100 casos de pacientes	LeNet, GoogLeNET, Inception-ResNet	100% revocação 99.3% precisão
Majumdar <i>et al.</i> (2018)	134 Imagens TC	U-Net	98% especificidade, 81% sensibilidade
Burduja, Ionescu and Verga (2020)	RSNA Intracranial Hemorrhage Detection	ResNeXt-101, LSTM-Bidirecional	97% acurácia, 96% sensibilidade e 98% especificidade
Hssayeni <i>et al.</i> (2020)	318 imagens TCs	U-Net	97.28% de sensibilidade, 50.4% de especificidade



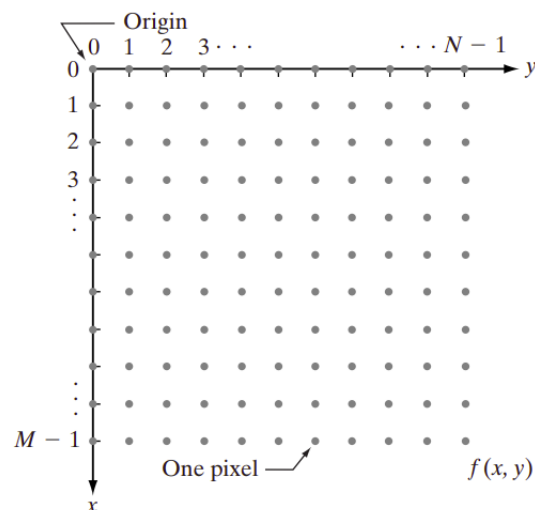
### 3 FUNDAMENTAÇÃO TEÓRICA

Nesse capítulo apresentam-se os conceitos básicos de processamento de imagem, percorrendo sobre representação de imagens no ambiente computacional e como podem ser extraídas características relevantes da mesma. Em seguida discutem-se conceitos relacionados a aprendizagem de máquina juntamente com os algoritmos relevantes para o trabalho.

#### 3.1 Processamento de Imagem

O termo processamento de imagem se refere ao ato de processar uma imagem por meios computacionais. Dessa maneira, computadores somente são capazes de reconhecer imagens do tipo digitais. Uma imagem digital pode ser definida como uma função  $f(x, y)$ , onde  $x$  e  $y$  são coordenadas espaciais e os valores mapeados em  $(x, y)$  são intensidades ou amplitudes de cinza. Contudo, tal função deve ser composta por um conjunto finito de elementos nomeados de *Pixels* (Figura 2) (CASTLEMAN, 1996).

Figura 2 – Representação formal de uma imagem digital.



Fonte: (CASTLEMAN, 1996)

Em suma, uma imagem digital é um conjunto finito de *pixels* representado em um ambiente computacional. Contudo, para o computador, trata-se apenas de uma matriz numérica contendo dados não estruturados. Isto posto, surge a necessidade de aplicar métodos para extrair informação das regiões dessa imagem a fim de identificar características únicas. Uma forma comum e bem conhecida na natureza para diferenciar objetos pela visão são as características de textura.

## 3.2 Características de Textura

Textura é uma propriedade inata de todas as superfícies. Pode ser fina, grossa, lisa, ondulada, irregular e outros padrões diversos a depender do objeto analisado. Quando se trata de reconhecer a texturas em imagens digitais é necessário a utilização e procedimentos e algoritmos para reconhecer os repetições existentes nos aglomerados de pixels (HARALICK; SHANMUGAM; DINSTEIN, 1973). Esse padrões serão reconhecidos em meio aos vários tons de cinza que compõem uma imagem digital em preto e branco. Tendo isso em vista, uma das formas de se abordar a descrição dessas texturas da maneira estatística.

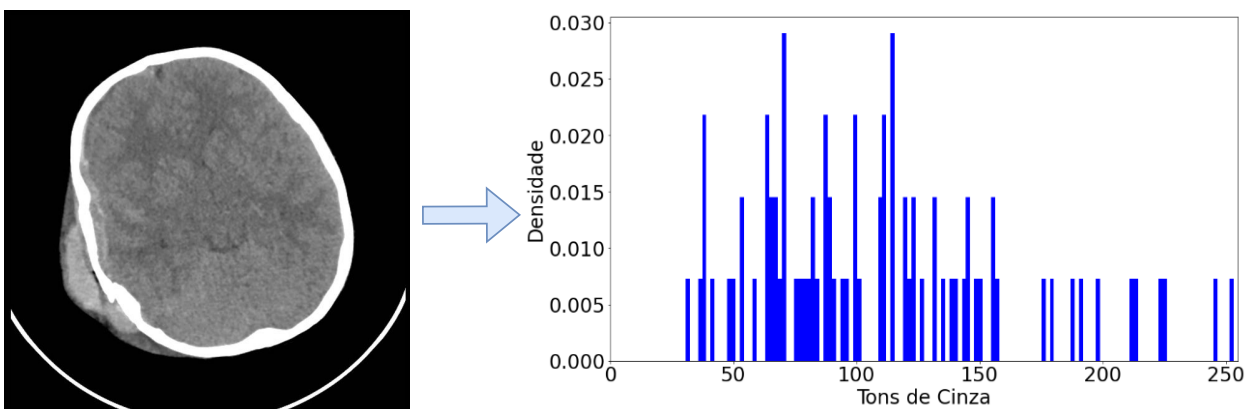
### 3.2.1 Abordagem Estatísticas

Os padrões encontrados em texturas nas imagens podem ser compreendidos através de regras estatísticas revelando as distribuições que regem as relações entre os tons de cinza. Portanto, existem duas classificações quanto a abordagem estatísticas, a de primeira ordem e a de segunda ordem.

#### 3.2.1.1 Abordagem Estatística de Primeira Ordem

A análise estatística de primeira ordem se baseia de forma mais direta nos tons de cinza presentes na imagem analisada. Os tons de cinza, geralmente postos na escala de 0 a 1 ou 0 a 255, formam um histograma de onde podem ser retiradas diversas atributos estatísticos como: média, desvio-padrão e assimetrias. De maneira concisa, pode-se dizer que o histograma forma uma distribuição de probabilidade de cada nível de cinza presente na imagem, desconsiderando assim, a localização dos pixels. Um exemplo dessa abordagem é encontrado na Figura 3, onde à esquerda é apresentada uma imagem CT de um cérebro e a direita um histograma contendo a distribuição de densidade dos tons de cinza.

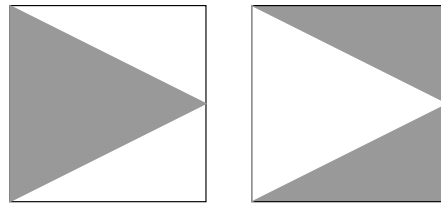
Figura 3 – Extração do histograma de uma imagem CT.



### 3.2.1.2 Abordagem Estatística de Segunda Ordem

Diferentemente da abordagem de primeira ordem a segunda ordem valoriza os quesitos de posicionamento dos pixels na imagem. Desse modo, duas imagens que possuam áreas de cinza iguais mas dispostas de maneira diferente seriam iguais para uma análise em primeira ordem, porém diferentes à vista da segunda ordem, como demonstrado na Figura 4.

Figura 4 – Duas imagens quadradas com um triângulo equilátero desenhado. As imagens possuem padrões de cores diferentes porém seriam a mesma imagem para uma análise de primeira ordem.



Fonte: Autor

Dentre as abordagens de segunda ordem presentes na literatura, as características propostas por Haralick, Shanmugam and Dinstein (1973) se destacam. O autor sugere uma extração de características através de matrizes de coocorrência – do inglês *gray level cooccurrence matrix* (GLCM). Tal método executa uma análise espacial pareada, entre diversos níveis de cinza, a partir de um pixel de referência e de um pixel de comparação. Como resultado final é gerado um histograma bidimensional onde cada eixo contém os  $n$  níveis de cinza presentes na imagem e os valores relativos a quantidade de ocorrências entre um valor referência e um valor de comparação. Dessa forma, dois valores definem qual será a posição do pixel a ser comparado, a distância e o ângulo, em relação ao  $y$ , para o pixel de referência. A distância representa quantas unidades de pixel será considerado a partir da referência e o ângulo qual a direção do pixel de comparação. Como exemplo, a Figura 5 apresenta uma imagem com quatro níveis de cinza (0 a 3). A letra **A** demarca o pixel de referência e as setas vermelhas todos os possíveis ângulos relativos ao pixel assinalado. Como a contagem do par de valores independe da ordem, os quatro ângulos  $0^\circ$  (acima),  $45^\circ$  (diagonal superior),  $90^\circ$  (direita) e  $135^\circ$  (diagonal inferior) são suficientes para extrair os valores em todas as direções.

Para um melhor entendimento da matriz de coocorrência elaborou-se um exemplo apresentado na Figura 6, onde ao lado esquerdo mostra-se uma imagem com quatro tons de cinza (0 a 3) e, ao lado direito, a matriz de coocorrência gerada utilizando a distância igual a 1 e um ângulo de  $90^\circ$  (direita). Em vista disso, é analisado primeiro quantas ocorrências de zero possuem outro zero logo a sua direita, tal comparação é indicada pela letra **A** e gera a contagem de coocorrência **A** na matriz à direita. Em sequência é refeita a

Figura 5 – Exemplo de ângulo em relação ao pixel de referência.

0	0	1	1	1
0	0 <sup>A</sup>	1	1	1
0	2	2	2	2
2	2	3	3	3
2	2	3	3	3

Diagrama de uma matriz 5x5 de pixels. O pixel central (linha 2, coluna 2) é o pixel de referência, rotulado 'A'. Setas vermelhas indicam deslocamentos de 0°, 45°, 90° e 135° a partir deste pixel. Os valores dos pixels são: (0,0)=0, (0,1)=0, (0,2)=1, (0,3)=1, (0,4)=1; (1,0)=0, (1,1)=0, (1,2)=1, (1,3)=1, (1,4)=1; (2,0)=0, (2,1)=2, (2,2)=2, (2,3)=2, (2,4)=2; (3,0)=2, (3,1)=2, (3,2)=3, (3,3)=3, (3,4)=3; (4,0)=2, (4,1)=2, (4,2)=3, (4,3)=3, (4,4)=3.

Fonte: Autor

comparação para os pares (0, 1), (0, 2) e (0, 3) gerando, respectivamente, as coocorrências mostradas pelas letras **B** e **C**. Não houve ocorrências do par (0, 3) logo seu valor será zero. A mesma lógica é refeita para os outros três valores de cinza gerando ao final a matriz de coocorrência, onde  $g_x$  e  $g_y$  representam todos os eixos da matriz contendo todos os valores de cinza presentes na imagem.

Figura 6 – Geração da matriz de coocorrência de uma imagem digital com tons de cinza escalados entre 0 e 3.

0 <sup>A</sup>	0 <sup>B</sup>	1	1	1
0 <sup>A</sup>	0 <sup>B</sup>	1	1	1
0 <sup>C</sup>	2	2	2	2
2	2	3	3	3
2	2	3	3	3

**Imagem**

	$g_x$			
	0	1	2	3
0 <sup>A</sup>	2	2	1	0
1	0	4	3	0
2	0	0	5	2
3	0	0	0	4

**Matriz de Coocorrência**

Diagrama de geração da matriz de coocorrência. À esquerda, a imagem original com deslocamentos de 0°, 45°, 90° e 135° rotulados A, B e C. À direita, a matriz de coocorrência com eixos  $g_x$  e  $g_y$  rotulados de 0 a 3. Os valores na matriz representam o número de ocorrências para cada par de tons de cinza.

Fonte: Autor

De maneira mais formal, pode-se definir uma matriz de coocorrência pela Equ-



ção 3.1, onde  $\Delta i, \Delta j$  representam o par deslocamento  $(\Delta i, \Delta j)$  no qual traduzem formalmente a ideia de distância e ângulo entre o pixel referência da imagem,  $I(x, y)$ , e o pixel de comparação,  $I(i + \Delta i, j + \Delta j)$ . Finalmente,  $g_1$  e  $g_2$  representam o par dos tons de cinza do pixel referência e do pixel de comparação, respectivamente. Reafirma-se que a ordem de  $g_1$  e  $g_2$  não importa podendo alternar a posição de ambas na equação.

$$\text{GLCM}_{\Delta i, \Delta j}(g_1, g_2) = \sum_{i=x-w}^{x+w} \sum_{j=y-w}^{y+w} \begin{cases} 1, \text{ se } I(i, j) = g_1 \text{ e } I(i + \Delta i, j + \Delta j) = g_2 \\ 0, \text{ caso contrário} \end{cases} \quad (3.1)$$

Um modo de interpretar a matriz de coocorrência é por meio da probabilidade conjunta. Assim, gera-se uma compreensão de qual a probabilidade de um par de tons de cinza ocorra na imagem digital analisada. Com esse intuito, divide-se cada valor da matriz GLCM pela soma de todas as coocorrências. Matematicamente, os coeficientes da matriz de probabilidade é dado pela equação abaixo:

$$P(g_1, g_2) = \frac{\text{GLCM}_{\Delta i, \Delta j}(g_1, g_2)}{\sum_{G_1=0}^{g_{\max}} \sum_{G_2=0}^{g_{\max}} \text{GLCM}_{\Delta i, \Delta j}(G_1, G_2)} \quad (3.2)$$

Como exemplo, é mostrado na Figura 7 a matriz de probabilidade da matriz anteriormente citada (Figura 6), nota-se que o par  $(2, 2)$ , quando define-se a distância como 1 e o ângulo como  $90^\circ$ , possui a maior chance de ocorrência na imagem.

Figura 7 – Matriz de probabilidade gerada a partir da GLCM

	<b>0</b>	<b>1</b>	<b>2</b>	<b>3</b>
<b>0</b>	0.1	0.1	0.05	0
<b>1</b>	0	0.02	0	0
<b>2</b>	0	0	0.25	0.1
<b>3</b>	0	0	0	0.2

**Matriz de Probabilidade**

Fonte: Autor

A GLCM gera um histograma que se torna a fonte de cálculo para várias características estatísticas de imagens digitais. Algumas delas definidas pelo próprio Haralick, Shanmugam and Dinstein (1973).

### 3.2.1.3 Características de Haralick

- Segundo Momento Angular (SMA): medida de homogeneidade local dos níveis de cinza. Capaz de identificar entradas de baixa magnitude em imagens muito homogêneas. A equação é dada por Equação 3.3 onde  $n$  representa as dimensões da imagem,  $i, j$  a coordenada do pixel,  $d$  a distância e  $\theta$  o ângulo.

$$SMA = \sum_{g_1} \sum_{g_2} [P(g_1, g_2)]^2 \quad (3.3)$$

- Contraste: mede a quantidade de variações locais de cinza. Segundo Haralick, Shanmugam and Dinstein (1973), valores altos para essa característica representa uma maior distribuição de cinza em seu histograma, portanto um valor menor seria representado por vários valores próximos à média.

$$Contraste = \sum_{g_1} \sum_{g_2} (g_1 - g_2)^2 P(g_1, g_2) \quad (3.4)$$

- Entropia: pode ser entendida como grau de dispersão de cinza, semelhante ao SMA. Valores iguais na matriz geram um baixa entropia, porém um valor alto quando possui muita concentração de valores na diagonal).

$$Entropia = - \sum_{g_1} \sum_{g_2} P(g_1, g_2) \log_2 P(g_1, g_2) \quad (3.5)$$

- Correlação: apresenta a possibilidade de uma linearidade de dependências entre os tons de cinza na imagem. Imagens que possuem padrões locais bem determinados costumam apresentar valores altos de correlação. A Equação 3.6 apresenta o cálculo da correlação onde  $\mu_x$  e  $\mu_y$  representam a médias na direções  $x$  e  $y$ , respectivamente e  $\sigma_x$  e  $\sigma_y$  os desvios padrão dessas direções.

$$Correlação = \sum_{g_1} \sum_{g_2} (g_1 g_2 P(g_1, g_2) - \mu_{P_x} \mu_{P_y}) / \sigma_{P_x} \sigma_{P_y} \quad (3.6)$$

## 3.3 Aprendizagem de Máquina

Na computação, problemas são resolvidos através da escrita de algoritmos, os quais descrevem sequencialmente instruções para gerar um resultado final. De maneira geral, algoritmos tradicionais são descritos por humanos com a tentativa de suprir todos os possíveis casos do problema proposto, entretanto isso cria limitações. Nem todos os problemas podem ser resolvidos por uma sequência determinística de ações. Cobrir todo o espaço de soluções é infazível para certas situações e exigem um inteligência prévia para que seja escolhido uma possível melhor resposta e não necessariamente uma resposta ótima.

Desse modo, surge a necessidade de conduzir o algoritmo a "aprender" com a experiência, introduzindo assim, uma capacidade de escolher uma boa solução ao invés de selecionar deterministicamente a melhor entre todas as possíveis. De acordo com Shalev-Shwartz and Ben-David (2014), de maneira simples, "aprender" é gerar conhecimento a partir de experiências. À vista disso, surge os algoritmos de aprendizagem de máquina (do inglês *Machine Learning*) que de diversas maneiras circundam o problema da rigidez dos algoritmos tradicionais e trazem uma adaptabilidade as soluções.

Dentre os vários tipos de algoritmos de aprendizagem esse trabalho irá focar nos ditos supervisionados. Algoritmos supervisionados são definidos pela forma em que são treinados. O processo de treinamento se descreve pelo processo de extrair "conhecimento" dos dados apresentados ao modelo. No caso do treinamento supervisionado, os exemplos apresentados para o algoritmo são acompanhados pelo seus respectivos rótulos. Em suma, é apresentado, durante o treinamento, o que deve ser aprendido sobre o dado em questão. Como exemplo disso, é apresentado um problema de identificação de Covid baseados em sintomas previamente anunciados por pacientes. No treinamento do modelo os dados são apresentados com os sintomas dos pacientes juntamente positividade do Covid no mesmo. Desse modo, o algoritmo deve aprender os padrões contidos nos sintomas que indicam a presença do vírus. Na sequência, o desempenho do modelo será testado em dados sem a informação da doença para assim ser avaliado em sua corretude. A figura 8 demonstra um diagrama de tabela contendo o exemplo supracitado.

Figura 8 – Tabela de Treinamento a esquerda contendo os sintomas e a positividade em relação ao Covid no paciente. A direita é apresentado os dados de teste no qual não vem acompanhado pelos dados de Covid, tendo que assim ser definido pelo modelo.

Treinamento					Teste			
Temperatura C°	TOSSE	CORIZA	DOR DE GARGANTA	COVID	Temperatura C°	TOSSE	CORIZA	DOR DE GARGANTA
36,7	SIM	NÃO	NÃO	NÃO	37,5	NÃO	NÃO	NÃO
37,5	NÃO	SIM	SIM	SIM	38,0	SIM	NÃO	SIM
38,0	SIM	SIM	SIM	SIM	36,7	SIM	SIM	NÃO
37,0	SIM	SIM	NÃO	NÃO	37,3	SIM	NÃO	NÃO
35,0	NÃO	NÃO	NÃO	NÃO				
37,4	NÃO	SIM	NÃO	SIM				

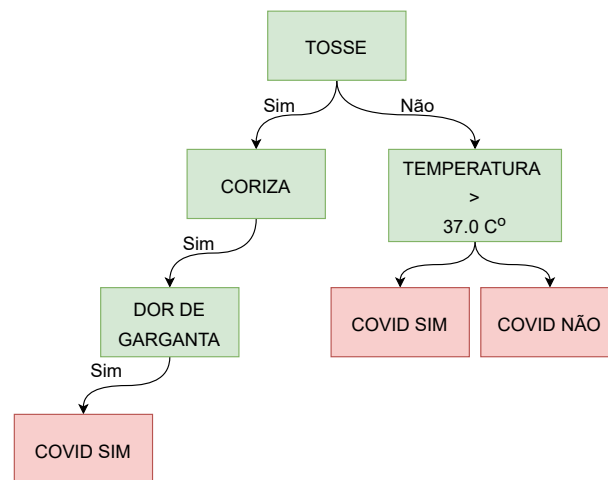
Fonte: Autor

Especificando melhor os algoritmos abordados nesse trabalho adentra-se agora nas árvores de decisão técnica de aprendizado supervisionado muito comum como em problemas referentes ao exemplo da Figura 8.

### 3.4 Árvores de Decisão

Uma abordagem de “dividir em conquistar” sobre dados tabulares, como da Figura 8, guiam de maneira natural até uma representação estrutural de uma árvore. Uma árvore é composta por nós nos quais podem ser ramificar criando nós filhos, criando assim, uma sequência de vários possíveis caminhos em a partir de cada nó. No caso das árvores de decisão cada nó representa um teste de um atributo em particular. Como por exemplo, resgatando novamente a Figura 8, pode-se efetuar um teste do atributo(coluna) “tosse”, criando duas ramificações, um caminho caso possua o sintoma e outra para o caso negativo. Contudo, um único nó não é capaz de prever a presença do vírus, tendo em vista que existem pacientes sem tosse mas positivos para Covid. Desse modo, a árvore de decisão suprime esse problema concatenando vários nós para tentar suprir todos os exemplos presentes nos dados de treinamento. Como ilustração disso é apresentado na Figura 9 uma possível árvore desenvolvida a partir da tabela da Figura 8.

Figura 9 – Árvore de decisão gerada para identificar a presença de vírus no paciente segundo os sintomas apresentados.



Fonte: Autor

Árvores de decisão possuem características importantes que as diferem de outros algoritmos. São capazes de aceitar tanto dados categóricos quanto dados numéricos no mesmo treinamento. Além disso, também toleram ausência de valores nos campos dos atributos. Outro ponto importante nesse método é a explicabilidade do modelo, a montagem da árvore é transparente ao usuário, logo é de fácil entendimento todas as decisões tomadas pela mesma. Tais características são herdadas por algoritmos que utilizam árvores como base, como, por exemplo, o gradiente *boosting*.

### 3.5 Gradiente Boosting

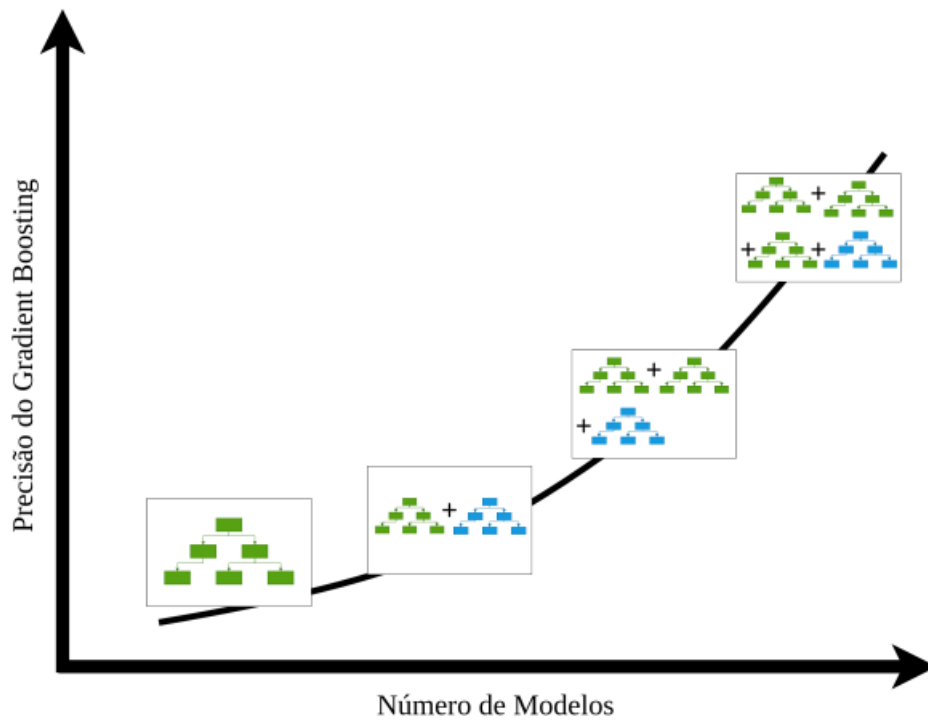
Árvores de decisão possuem limitações quanto a variância dos dados. Qualquer instância adicionada aos dados de treinamento pode gerar árvores completamente novas que reduzam consideravelmente a abrangência do modelo (BENGIO; DELALLEAU; SIMARD, 2010). Como umas das alternativas de solução surge o gradiente *boosting*, um algoritmo que constrói várias pequenas árvores sequencialmente buscando uma redução do erro a cada interação.

Em suma, segundo (FRIEDMAN, 2002) o algoritmo pode ser reduzido a 3 principais etapas. Para uma entrada  $(x_i, y_i)_{i=1}^n$ , onde  $(x_i, y_i)$  representa um exemplo de treino, e uma função de perda diferenciável  $L(y_i, F(x))$ , onde  $F(x)$  é a predição da função a ser otimizada, execute:

1. Inicialize um modelo com valor constante:  $F_0(x) = \operatorname{argmin} \sum_{i=1}^n L(y_i, \gamma)$ .
2. Para  $m = 1$  até  $M$ :
  - a) Compute o valor do resíduo  $r_{im} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x_i)}\right]_{F(x)=F_{m-1}}$  para  $i = 1$  até  $n$ .
  - b) Treine uma árvore de decisão para prever o resíduo  $r_{im}$  e crie folhas  $R_{jm}$  para cada  $j = 1$  até  $J_m$ .
  - c) Para cada folha  $j = 1$  até  $J_m$  compute  $\gamma_{jm} = \operatorname{argmin} \sum_{x_i \in R_{ij}} L(y_i, F_{m-1}(x_i) + \gamma)$ .
  - d) Atualize  $F_m(x) = F_{m-1}(x) + v \sum_{j=1}^{J_m} \gamma_{jm} (x \in R_{jm})$ .
3. Utilize a predição do modelo  $F_M(x)$ .

De maneira mais descritiva, na Etapa 1 o algoritmo estabelece uma árvore inicial  $\gamma$  como um valor constante. Em sequência, na Etapa 2, inicia-se um laço de  $M$  interações, representando a quantidade de árvores que serão construídas até o modelo final. Prosseguindo para a etapa “a”, computa-se a derivada da função de perda em relação ao modelo anterior  $F_{m-1}$  para todas os exemplos de entrada. Tal abordagem realizará a decida no gradiente da derivada da função de perda  $L$ , reduzindo seu valor a cada interação. Por conseguinte na sub-etapa “b”, treinasse uma nova árvore de decisão para prever os resíduos  $r_{im}$ . Desse modo, a cada nova interação os novos modelos estarão tentando prever o erro das suas predecessoras e propagando a diminuição do mesmo através da etapa “a”. Em sequência na etapa “c”, o resultado de cada folha  $\gamma_{jm}$  da atual árvore  $\gamma_j$  é computado para assim ser utilizado na etapa “d”, onde é produzido um novo modelo  $F_m$  igualando-se a soma das predições de  $F_{m-1}$  com a multiplicação de uma taxa de aprendizagem  $v$  e as predições de resíduos da nova árvore criada  $\gamma_{jm}$ , para cada exemplo  $x$  presentes na folhas  $R_{jm}$ . Logo, como apresentado na Figura 10, a união sequencial de várias pequenas árvores podem resultar em um modelo mais robusto ao mesmo tempo que trás todas as vantagens de uma árvore de decisão.

Figura 10 – Representação do aumento da precisão do modelo de gradiente boosting a medida que mais árvores vão sendo adicionadas ao modelo final.



Fonte: (FUHR, 2022)

## 4 METODOLOGIA

Este capítulo se dedica a esclarecer os métodos utilizados para o processo de classificação de hemorragias cerebrais em imagens CTs. Para tanto, os experimentos realizados buscam clarificar as seguintes perguntas:

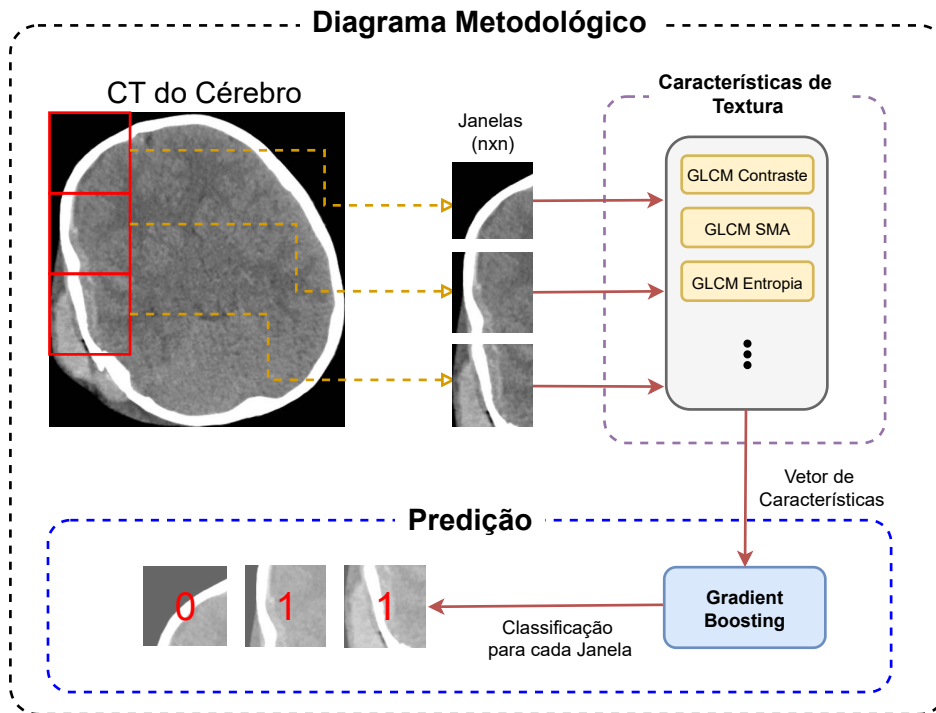
- P1.** Os algoritmos de *boosting* são capazes de obter uma melhor generalização do que os modelos baseados em redes neurais profundas da literatura?
  
- P2.** Dentre as características tradicionais da literatura, quais são as que possuem maior importância na detecção de hemorragias cerebrais?

Com o intenção de responder as perguntas propostas será desenvolvido ensaios de experimentos utilizando imagens de CTs cerebrais. Tais experimentos serão avaliados através de métricas determinadas para que posteriormente seja analisado a importância das características na detecção de hemorragias.

### 4.1 Visão Geral

A tarefa de detectar hemorragias em imagens médicas traz o desafio de incorporar a localização da enfermidade na totalidade da tomografia. Devido a isso, a classificação da imagem inteira se mostra insuficiente para tal finalidade, assim, levanta-se a necessidade de analisar porções menores de pixels gerando uma análise mais específica de cada fragmento da imagem. Porém, as características texturais de Haralick são de grande complexidade e exigem muito tempo computacional para serem computadas. Devido a limitações de tempo e infraestrutura computacional esse trabalho se dispôs a utilizar janelas de tamanho maiores tornando o processamento factível e de fácil reprodução. Baseado nessa estratégia, foi elaborado um procedimento de classificação de hemorragia cerebral nas imagens CTs utilizando uma abordagem de janela móvel. Como demonstrado no diagrama da Figura 11, para cada janela presentes no CT do cérebro é extraída as características de textura e passadas em forma de vetor para o algoritmo de *Gradient Boosting*. Como resposta do modelo, para cada janela é gerado uma classificação binária, sendo 1 a identificação positiva de uma hemorragia e 0 para negativa.

Figura 11 – Diagrama Metodológico



Fonte: Autor

## 4.2 Base de Dados

Para realizar os experimentos supracitados, foi utilizado a base dados *Kaggle RSNA Intracranial Hemorrhage Detection Competition* disponibilizada na plataforma *Kaggle*<sup>1</sup>. Tal base é composta por tomografias computadorizadas de 82 pacientes e 30 recortes de imagens cerebrais para cada um, totalizando 2500 CTs no formato JPG. Cada imagem possui dimensões de  $650 \times 650$  pixels. Dentre esses dados, os recorte CTs que possuem algum tipo de hemorragia vem acompanhados de uma mascara, também no formato JPG, indicando o local exato da enfermidade, como demonstrado na área em vermelho na Figura 12.

## 4.3 Pré-processamento dos Dados

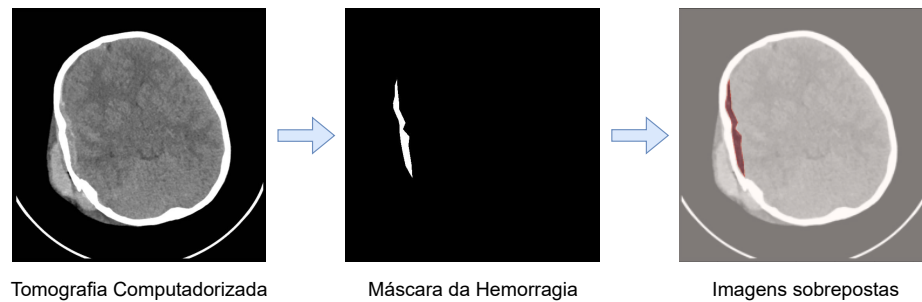
As imagens de TC contidas na base RSNA possuem uma grande quantidade de bordas pretas além de muitos recortes sem a presença de hemorragias. Tendo isso em vista, foi feito uma série de pré-processamentos para garantir o balanceamento e a correta presença de hemorragia dentro de uma mesma janela do recorte

Primeiramente a base foi filtrada para conter somente os recortes de tomografias que possuíam alguma hemorragia. Dessa maneira, a classe negativa é representada por

<sup>1</sup> <https://www.kaggle.com/datasets/vbookshelf/computed-tomography-ct-images>



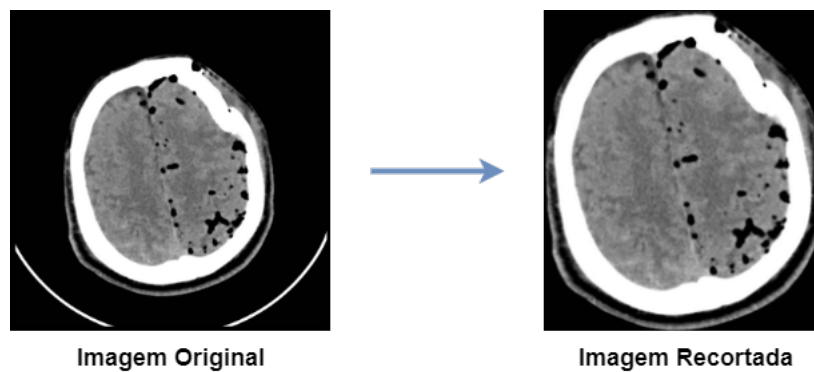
Figura 12 – Apresentação da tomografia computadorizada à esquerda juntamente com a máscara. À direita encontra-se a sobreposição de ambas.



Fonte: Autor

todas as janelas que não possuem hemorragia mas fazem parte do mesmo cérebro. Após essa etapa, com o intuito reduzir a quantidade de janelas pretas presentes nos dados negativos foi utilizado um algoritmo baseado em filtragem por um limiar de *Threshold* que recorta o menor quadrado até as bordas mais externas do crânio, como demonstrado na Figura 13.

Figura 13 – Exemplo do recorte feito pelo algoritmo de *Threshold* com o intuito de eliminar as bordas pretas.



Fonte: Autor

Por fim, as janelas classificadas como positivas para hemorragias foram selecionadas baseado na proporção de pixels com hemorragia presentes na totalidade da janela. Para os experimentos desse trabalho a proporção de pixels positivos escolhido foi de no mínimo 80%. Dessa forma é garantido que o modelo não receberá janelas com pouca informações sobre a hemorragia ocorrendo em possíveis ruídos.

#### 4.4 Escolha de Parâmetros nas Características de Textura

As características de textura agrupadas por esse trabalho utilizam parâmetros previamente estabelecidos para gerar seus valores. As características propostas por Haralick,

Tabela 2 – Valores de distância e ângulos para as características de (HARALICK; SHANMUGAM; DINSTEIN, 1973)

Distâncias	1	3	5	-	Total
Ângulos(Graus)	0	45	90	135	-
Carac. de Textura	Homogeneidade	Contraste	Entropia	Correlação	48

Shanmugam and Dinstein (1973) usam como base uma matriz de coocorrência. Para gerar tal matriz é necessário estabelecer previamente dois parâmetros, a distância e o ângulo entre os pixels que serão comparados. Desse modo, foram estabelecidos arbitrariamente três valores de distancia, 1, 3 e 5 e quatro valores para os ângulos, 0°, 45°, 90° e 135°. Levando isso em consideração, foram gerados 12 histogramas GLCMs sendo que para cada um 4 características são extraídas, totalizando ao final 48 atributos. A tabela 4.4 resume os atributos escolhidos para a geração das características.

#### 4.5 Treinamento e Validação

Após o pré-processamento os dados foram submetidos ao experimento descrito anteriormente na Figura 11. Com o intuito de certificar a robustez do modelo foi estabelecido uma proporção de 70% para treino e 30% de teste. Dentro dos dados de treino será efetuado um treinamento usando validação cruzada de K-Fold, com  $K = 10$ . Além disso, existe um grande desbalanço entre as classes positivas e negativas. Para coordenar essa diferença e averiguar sem erros o experimento foi executado uma amostragem aleatória dos dados negativos na mesma proporção dos dados positivos. Essa amostragem foi feita em um total de 10 ensaios diferentes para garantir uma confiança maior nos resultados obtidos. Importante adicionar que tal método foi escolhido após desempenho ruins de técnicas de balanceamento sintético como Técnica de Sobre-Amostragem de Minoria Sintética – do inglês *Synthetic Minority Over-Sampling Technique* (SMOTE) (CHAWLA *et al.*, 2002).

Como etapa final, foi aproveitado a escolha de algoritmos de *boosting* para realizar uma análise mais aprofundada nas importâncias que cada característica. Essa informação será extraída do cálculo do índice de Shapley (LARUELLE; VALENCIANO, 2001) que será responsável por ordenar e identificar a influência das características na classificação, respondendo assim a segunda pergunta da metodologia.

#### 4.6 Métricas de avaliação

Com a finalidade de analisar os resultados finais do experimento serão utilizadas as métricas de sensibilidade (Equação 4.1), precisão (Equação 4.2), especificidade (Equação 4.3), acurácia (Equação 4.4) e Escore F1 (Equação 4.5). Para o cálculo de tais métricas são utilizados 4 grupos de dados. Os verdadeiros positivos (VP), são os dados corretamente

rotulados como positivos, os falsos positivos (FP), são incorretamente rotulados como positivos, os verdadeiros negativos (VN), corretamente rotulados como negativo e por último o falso negativo, classificado incorretamente como negativo. Cada métrica avalia um aspecto diferente dos resultados. A sensibilidade descreve o quão corretos o modelo consegue prever dentre os resultados positivos. A precisão, expõe a corretude do modelo dentro dos valores que foram julgados positivos pelo mesmo. A especificidade traz o oposto observando os que foram julgados negativos. A acurácia avalia o todo sendo uma média dos acertos perante todas as possibilidades. Por último, o Escore F1 executa uma média harmônica entre a precisão e a sensibilidade criando um olhar mais equilibrado entre as métricas. Tal métrica se apresenta entre valores no intervalo  $[0, 1]$ , sendo 1 o melhor resultado.

$$\text{sensibilidade} = \frac{VP}{VP + FN} \quad (4.1)$$

$$\text{precisão} = \frac{VP}{VP + FP} \quad (4.2)$$

$$\text{especificidade} = \frac{VN}{VN + FP} \quad (4.3)$$

$$\text{acurácia} = \frac{VP + VN}{VP + VN + FP + FN} \quad (4.4)$$

$$\text{Escore F1} = \frac{\text{precisão} * \text{sensibilidade}}{\text{precisão} + \text{sensibilidade}} \quad (4.5)$$



## 5 RESULTADOS

Nessa seção será apresentados os experimentos de classificação de hemorragia cerebral utilizando a base de dados *Kaggle RSNA*. De maneira mais específica, será apresentado os resultados do treinamento do modelo *Gradient Boosting* usando as características de Haralick, citadas na Seção 3. Posteriormente também será avaliado a importância de cada característica e qual variação teve a maior contribuição nas classificações utilizando o método de valores de Shapley. Todos os códigos e dados estão acessíveis no repositório público<sup>1</sup>.

Nota-se a importância de destacar que não foi encontrado nenhum trabalho acadêmico que tenha feito a análise do impacto das características de Haralick na classificação de imagens médicas.

### 5.1 Experimento

Esse experimento apresenta todas as fases de tratamento de dados, treinamento, validação e análise dos resultados finais. Tendo isso em vista, primeiramente foi realizado a extração das características de texturas Haralick usando janelas de 25 pixels de altura e comprimento totalizando 99476, onde 3992 são positivas para hemorragia e 95484 são negativas. Todas as 48 características extraídas são distribuídas em colunas onde cada linha representa as características de uma janela extraída. A implementação do algoritmo *Gradient Boosting* escolhida para o experimento foi o *Catboost*<sup>2</sup>, devido a sua moderna implementação (PROKHORENKOVA *et al.*, 2018) e de fácil interface uso com a linguagem *Python*.

#### 5.1.1 Hiper-Parâmetros

Nesse trabalho foram abordados 4 principais hiper-parâmetros do algoritmo *Catboost*<sup>3</sup> responsáveis por alterar o desempenho do modelo durante o treinamento. O primeiro deles é a quantidade de árvores (ou *Iterations* na nomenclatura da Biblioteca em *Python*) que serão criadas pelo algoritmo. Foi selecionado o valor máximo para o número de árvores de 500, porém o *Catboost* possui um sistema de parada precoce. No momento que é percebido que as métricas obtidas com os dados de validação não se alteraram por um certo número de interações, no caso 5 interações foi selecionado, o algoritmo força a parada.

<sup>1</sup> [https://drive.google.com/drive/folders/1pxnBD0yQiCwZw3jXujmeAfjFT8LzeKnI?usp=share\\_link](https://drive.google.com/drive/folders/1pxnBD0yQiCwZw3jXujmeAfjFT8LzeKnI?usp=share_link)

<sup>2</sup> <https://catboost.ai/en/docs/>

<sup>3</sup> <https://catboost.ai/en/docs/concepts/parameter-tuning>

Outro hiper-parâmetro importante é a taxa de aprendizado – do inglês *learning rate*. A escolha de tal parâmetro foi atribuída ao próprio algoritmo que possui a mecanismos que identificam e alteram o valor da taxa de aprendizado a partir do tamanho da amostra de treino e do avanço das métricas com o passar das iterações.

A contagem de borda (ou do inglês *Board Count*) é o hiper-parâmetro que indica a quantidade máxima de divisões que os nós farão para se adaptar as características numéricas. Tal valor foi escolhido baseado na sugestão da documentação da biblioteca em *Python* que aponta o valor de 254 como um valor ótimo para treino em CPU. Valores maiores foram testados porém não resultaram na melhora do desempenho.

Por último a profundidade das árvores (do inglês *Tree Depth*) é o hiper-parâmetro que define a máxima altura das árvores que o algoritmo irá produzir. A documentação salienta que os valores recomendados estão entre 4 e 10 então foi utilizado 10 como profundidade.

### 5.1.2 Treinamento e Métricas

Ao total foram feitos 10 ensaios treinando 10 modelos diferentes com o intuito de contornar a amostragem dos dados desbalanceados. Dessa forma, para cada ensaio foi treinado um novo modelo utilizando os dados da classe negativa amostrados de maneira aleatória. Como resultado foi gerado um gráfico de caixas, apresentado na figura 5.1.2, no qual são evidenciados a distribuição das acurácias obtidas em cada um das 10 etapas de teste da validação cruzada. Além disso, foram avaliadas as métricas de precisão, sensibilidade e Escore F1, apresentadas na Figura 5.1.2, demonstrando a dispersão das mesmas durante todos os 10 ensaios na fase de teste.

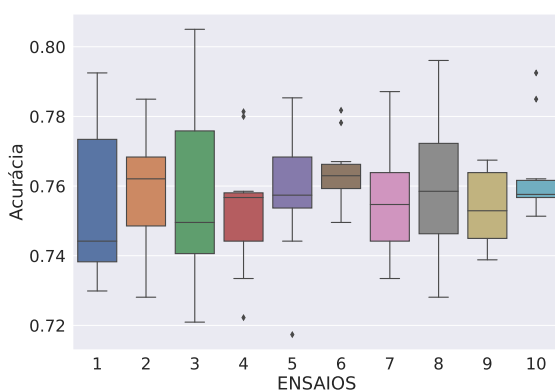


Figura 14 – Gráfico de caixas que apresenta a distribuição das acurácias obtidas nos 10 ensaios do experimento.

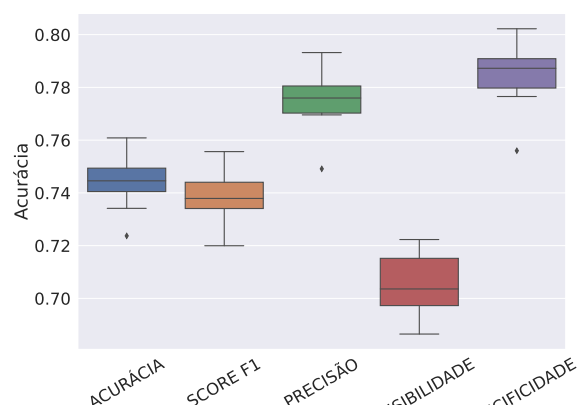


Figura 15 – Gráfico de caixas que apresenta a distribuição das métricas de sensibilidade, precisão e Escore F1 obtidas nos 10 ensaios do experimento.

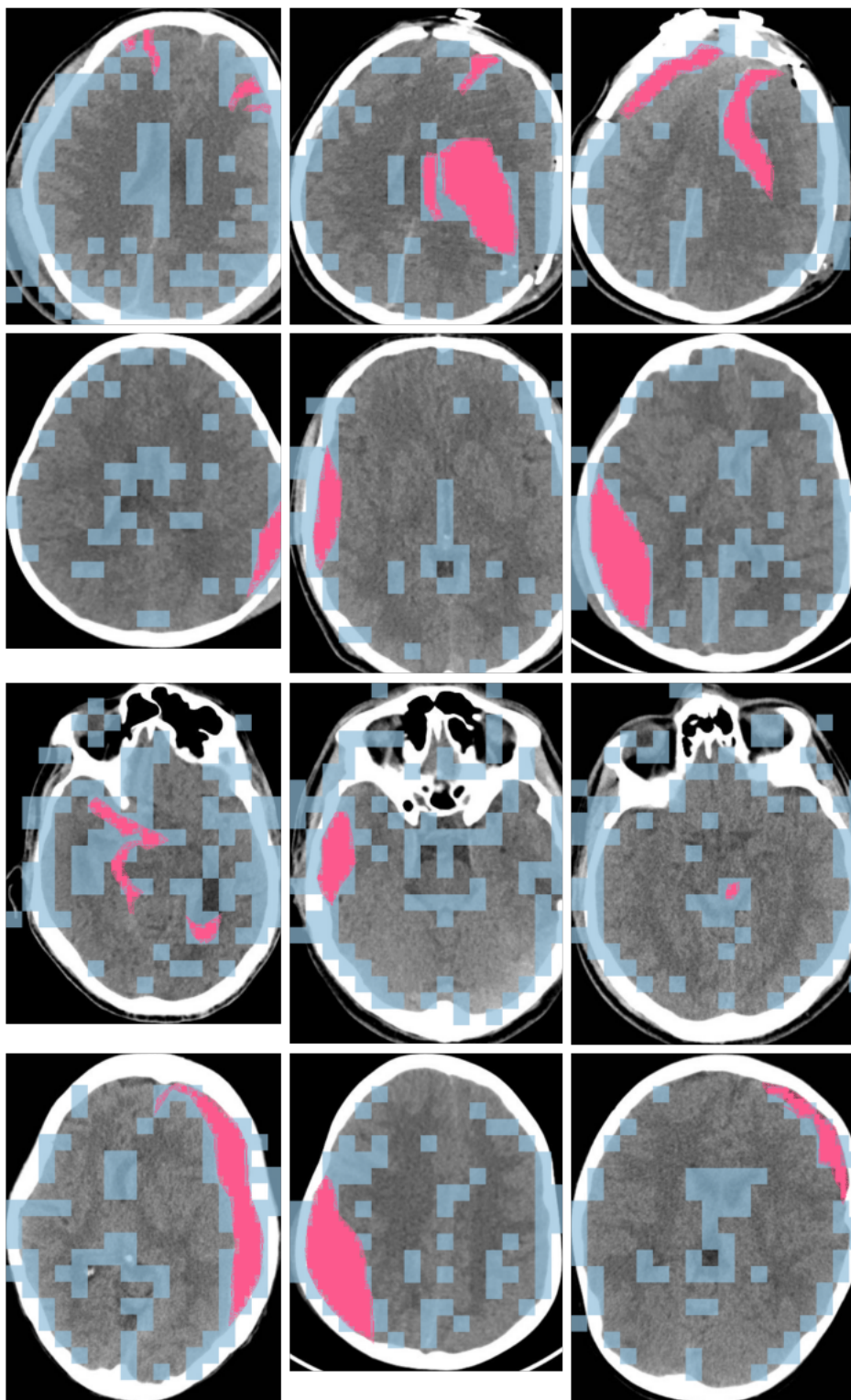
Ao final do treinamento, o modelo com a melhor média alcançou uma acurácia de 75.80%, passível de comparação com alguns resultados dos artigos analisados por Hssayeni *et al.* (2020). Uma sensibilidade de 71.70% indica que o modelo está identificando deixando de identificar áreas que possuem hemorragias. Já uma precisão maior de 77.69% demonstra que os falsos positivos são menos frequentes que os falsos negativos aumentando a confiança ao identificar áreas que realmente possuem a enfermidade. A especificidade alcançou 79.85% sendo a maior métrica evidenciando melhor robustez ao confirmar áreas sem hemorragia. Um Escore F1 de 74.53% se apresenta como uma métrica de equilíbrio entre a sensibilidade e a precisão onde um valor próximo da acurácia mostra que, dentro do limite do modelo, ele se mostra robusto as duas classes que se propõe a identificar. Os valores são comparáveis a alguns dos trabalhos de referência como Ramteke and Monali (2012) que não apresentou a base mas adquiriu resultados semelhantes com algoritmos rasos. Além disso, a especificidade se mostrou comparável a métodos mais custosos como as U-Nets dos trabalhos de Majumdar *et al.* (2018) e Hssayeni *et al.* (2020) que utilizou também o a base de dados RSNA.

Como demonstração do resultado final do algoritmo apresenta-se a Figura 16 com as predições de 12 recortes das imagens TC da base RSNA. A cor rosa representa a real posição da hemorragia no cérebro enquanto os pequenos quadrados azul claros mostram a porção da imagem que o algoritmo acreditou haver hemorragias. Tais exemplos refletem as métricas encontradas no experimento, em todas o algoritmo detectou corretamente a posição da hemorragia, o que explicaria a precisão mais alta que as outras métricas. Porém nota-se também uma grande quantidade de falsos positivos, justificando a baixa sensibilidade. Todas as imagens se encontram no repositório disponibilizado por esse trabalho.

Tabela 3 – Tabela de métricas adquiridas pelo modelo *Catboost* durante a fase de treinamento e teste.

ACURÁCIA	0.758089
ESCORE F1	0.745307
PRECISÃO	0.776936
SENSIBILIDADE	0.717075
ESPECIFICIDADE	0.798545

Figura 16 – Recortes de imagens TC que passaram pela predição do melhor modelo *Gradient Boosting* desenvolvido no experimento. A cor rosa representa a hemorragia enquanto a cor azul claro a predição positiva para hemorragia naquela área.



Fonte: Autor



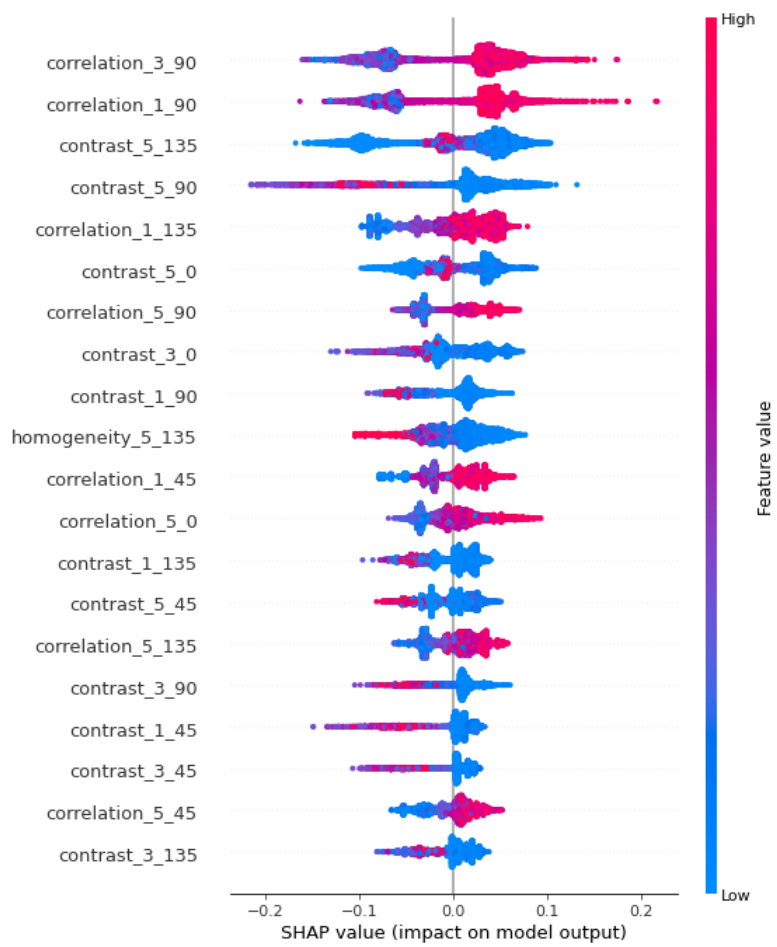
### 5.1.3 Importância das Características

Valor de Shapley se define como um método de análise da importância que cada característica teve na definição das classes durante o processo de treinamento e teste. Na Figura 17 é apresentado em ordem de cima para baixo as características mais impactantes (Mais acima) até as com menor impacto (Mais abaixo). Também é identificado a direita uma barra com os rótulos *high* and *low* representando o valor da característica em relação ao máximo e mínimo. Dessa forma, é possível analisar a relevância de valores específicos para a classificação. Como demonstração disso, na Figura 17, pode-se observar que as duas primeiras características são correlações com distância 3 e 1 e ângulo 90, respectivamente. Tal característica salienta que existem padrões locais bem determinados. Devido a isso, hipotetiza-se que os altos valores dessas variáveis relacionados com a classe a positiva se deve a grande quantidade de hemorragias encontradas na borda com os ossos do crânio. Também pode-se notar que para todas as outras características de correlação presentes na Figura 17 o padrão de alto contraste predispõe a uma classificação positiva. Outra característica que se destaca são as de contraste evidenciando que um valor muito alto ou baixo contraste levou o modelo a tender para classificações negativas. O que reafirma a hipótese de que muitas das hemorragias são encontradas próximas a ossos ou com uma mudança de padrão perceptível entre os tons de cinza. Em suma, pode-se observar que contraste e correlação foram os principais destaques para identificação da hemorragia em meio a massa cinzenta do cérebro deixando características como homogeneidade e energia como as menos importantes para as decisões.

### 5.1.4 Considerações Finais

Os resultados apresentados apontam que o algoritmo de *Gradiente Boosting* teve dificuldades de alcançar, com as características de Haralick, as maiores acurácias da literatura. Porém, se mostrou um algoritmo robusto mantendo as métricas bem equilibradas entre si ao longo de todos os ensaios, além de obter especificidade comparável a tecnologias mais custosas. Essa análise responde a pergunta **P1** realizada no Capítulo 4. Retomando a pergunta **P2**, através dos valores de Shapley foi possível concluir que, para o modelo deste experimento, as características de correlação e contraste se mostraram as com maior influência nas decisões da predição. Ademais as características de energia e homogeneidade não se encontram entre as 20 com maior impacto decisório.

Figura 17 – Valores de Shapley demonstram a importância de cada característica na decisão final da classificação



Fonte: Autor

## 6 CONCLUSÃO

A detecção de hemorragias cerebrais em imagens de tomografia computadorizadas pode ser de grande utilidade para auxiliar o dia-a-dia médico nos hospitais. Este trabalho se propôs a ser um *benchmark* com teor exploratório utilizando características de Haralick, um método tradicional e antigo juntamente com algoritmos de *Gradient Boosting*. Os dados utilizados foi a base pública RSNA também utilizada pelo trabalho de Hssayeni *et al.* (2020). No quesito experimental foram feitos 10 ensaios usando validação cruzada do tipo *K-fold* para obter uma maior confiança no desempenho do modelo final. A média das métricas atingidas foram 75.80% de acurácia, 74.53% de Escore F1, 77.69% de Precisão, 71.70% de sensibilidade e 79.85% de especificidade. Importante mencionar a disponibilidade e reprodutibilidade desse trabalho se destaca quando comparando a outros similares nos quais nem a base de dados e nem os códigos são fornecidos. Além disso, foram feitas também análise do impacto decisório que as características de *Haralick* obteve junto ao modelo de *Gradient Boosting*.

Como trabalhos futuros existem várias opções a serem exploradas. Primeiramente a adição de mais características de textura como dissimilaridade e ASM. Também seria plausível a abordagem multi-classe ao problema de hemorragia, aproveitando os 5 tipos de hemorragias que são rotulados na base RSNA. Outra inclusão seria a possibilidade da utilização de GPUs para a geração das características de textura na tentativa de melhorar o tempo dedicado a essa tarefa permitindo diminuir ainda mais as janelas usadas no experimento.



## REFERÊNCIAS

- BENGIO, Y.; DELALLEAU, O.; SIMARD, C. Decision trees do not generalize to new variations. **Computational Intelligence**, Wiley Online Library, v. 26, n. 4, p. 449–467, 2010.
- BURDUJA, M.; IONESCU, R. T.; VERGA, N. Accurate and efficient intracranial hemorrhage detection and subtype classification in 3d ct scans with convolutional and long short-term memory neural networks. **Sensors**, MDPI, v. 20, n. 19, p. 5611, 2020.
- CASTLEMAN, K. R. **Digital image processing**. [S.l.: s.n.]: Prentice Hall Press, 1996.
- CHANG, R.-F. *et al.* Support vector machines for diagnosis of breast tumors on us images. **Academic radiology**, Elsevier, v. 10, n. 2, p. 189–197, 2003.
- CHAWLA, N. V. *et al.* Smote: synthetic minority over-sampling technique. **Journal of artificial intelligence research**, v. 16, p. 321–357, 2002.
- DOI, K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. **Computerized medical imaging and graphics**, Elsevier, v. 31, n. 4-5, p. 198–211, 2007.
- DOROGUSH, A. V.; ERSHOV, V.; GULIN, A. Catboost: gradient boosting with categorical features support. **arXiv preprint arXiv:1810.11363**, 2018.
- FRIEDMAN, J. H. Stochastic gradient boosting. **Computational statistics & data analysis**, Elsevier, v. 38, n. 4, p. 367–378, 2002.
- FUHR, G. T. Uso de machine learning para a classificação do crédito de empresas por meio de demonstrativos financeiros. **UNIVERSIDADE FEDERAL DE SANTA MARIA CENTRO DE TECNOLOGIA CURSO DE GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO**, UNIVERSIDADE FEDERAL DE SANTA MARIA, 2022.
- GOLDSTEIN, J. N.; GILSON, A. J. Critical care management of acute intracerebral hemorrhage. **Current treatment options in neurology**, Springer, v. 13, n. 2, p. 204–216, 2011.
- HARALICK, R. M.; SHANMUGAM, K.; DINSTEN, I. Textural features for image classification. **IEEE Transactions on Systems, Man, and Cybernetics**, SMC-3, n. 6, p. 610–621, 1973.
- HSSAYENI, M. D. *et al.* Intracranial hemorrhage segmentation using a deep convolutional model. **Data**, MDPI, v. 5, n. 1, p. 14, 2020.
- KIDWELL, C. S.; WINTERMARK, M. Imaging of intracranial haemorrhage. **The Lancet Neurology**, Elsevier, v. 7, n. 3, p. 256–267, 2008.
- KONONENKO, I. Machine learning for medical diagnosis: history, state of the art and perspective. **Artificial Intelligence in medicine**, Elsevier, v. 23, n. 1, p. 89–109, 2001.

- LARUELLE, A.; VALENCIANO, F. Shapley-shubik and banzhaf indices revisited. **Mathematics of operations research**, INFORMS, v. 26, n. 1, p. 89–104, 2001.
- LI, Y.-H. *et al.* Automatic subarachnoid space segmentation and hemorrhage detection in clinical head ct scans. **International journal of computer assisted radiology and surgery**, Springer, v. 7, n. 4, p. 507–516, 2012.
- LODWICK, G. S. *et al.* Computer diagnosis of primary bone tumors: a preliminary report. **Radiology**, The Radiological Society of North America, v. 80, n. 2, p. 273–275, 1963.
- MAJUMDAR, A. *et al.* Detecting intracranial hemorrhage with deep learning. *In: IEEE. 2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*. [*S.l.: s.n.*], 2018. p. 583–587.
- PADMA, A.; SUKANESH, D. R. Automatic diagnosis of abnormal tumor region from brain computed tomography images using wavelet based statistical texture features. **arXiv preprint arXiv:1109.1067**, 2011.
- PHONG, T. D. *et al.* Brain hemorrhage diagnosis by using deep learning. *In: Proceedings of the 2017 International Conference on Machine Learning and Soft Computing*. [*S.l.: s.n.*], 2017. p. 34–39.
- PROKHORENKOVA, L. *et al.* Catboost: unbiased boosting with categorical features. **Advances in neural information processing systems**, v. 31, 2018.
- RAMTEKE, R.; MONALI, K. Y. Automatic medical image classification and abnormality detection using k-nearest neighbour. **International Journal of Advanced Computer Research**, Citeseer, v. 2, n. 4, p. 190, 2012.
- SHAHANGIAN, B.; POURGHASSEM, H. Automatic brain hemorrhage segmentation and classification algorithm based on weighted grayscale histogram feature in a hierarchical classification structure. **Biocybernetics and Biomedical Engineering**, Elsevier, v. 36, n. 1, p. 217–232, 2016.
- SHALEV-SHWARTZ, S.; BEN-DAVID, S. **Understanding machine learning: From theory to algorithms**. [*S.l.: s.n.*]: Cambridge university press, 2014.