Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Detecção de anomalias em imagens de raio-x com aprendizagem não supervisionada

Rafael Silva de Alencar

Monografia apresentada como requisito parcial
para conclusão do Curso de Ciência de Computação

Orientador
Prof. Dr. Díbio Leandro Borges

Brasília
2022

Universidade de Brasília
Instituto de Ciências Exatas
Departamento de Ciência da Computação

Detecção de anomalias em imagens de raio-x com aprendizagem não supervisionada

Rafael Silva de Alencar

Monografia apresentada como requisito parcial
para conclusão do Curso de Ciência de Computação

Prof. Dr. Díbio Leandro Borges (Orientador)
CiC/UnB

Prof. Dr.
Jan Mendonça Correa

Prof. Dr.
Wilson Henrique Veneziano

Prof. Dr. Marcelo Grandi Mandelli
Coordenador do Curso de Ciência de Computação

Brasília, 15 de Setembro de 2022

Dedicatória

Dedico este trabalho a minha família, pela ajuda e imensurável apoio durante toda minha acadêmica. Sem eles não seria possível congretizar o sonho da formação no ensino superior.

Agradecimentos

Agradeço ao Prof. Dr. Díbio Leandro Borges pela imensa paciência, direcionamento e orientação para a realização deste trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), por meio do Acesso ao Portal de Periódicos.

Resumo

O uso de técnicas de inteligência artificial é importante para aplicações no campo da medicina, auxiliando o processo de diagnóstico médico de diversos pacientes. No que diz respeito a imagens médicas, o diagnóstico a partir de radiografías requer o conhecimento técnico de profissionais especializados sobre a área, que são treinados durante anos até obter o nível de expertise necessário para avaliar imagens de raios-x. Além disso, as imagens por si só as vezes não são suficientes para que o profissional de saúde encontre um diagnóstico definitivo, fazendo com que informações adicionais sejam levantadas, como o histórico do paciente e familiar. Todas essas caractéristicas encarecem o processo de obtenção de rótulos para uma base de dados extensa, o que dificulta a aplicação de uma abordagem supervisionada para classificar imagens. Este trabalho aplica técnicas de aprendizagem de máquina utilizando a abordagem não supervisionada para classificar imagens de raios-x, em que diferentes métodos são treinados em imagens que não apresentam anomalias e que podem auxiliar médicos a avaliar imagens de raios-x. Diferentes experimentos foram produzidos e comparados a fim de verificar qual técnica não supervisionada é a mais satisfatória. Ao fim do experimento foi possível criar uma arquitetura genérica para classificar imagens de raios-x como sendo anomalas ou não, com os resultados da ROC-AUC em 0.547 com D *Discrimator probability* e 0.533 com MSE *Mean Squared Error*.

Palavras-chave: Raio-x, aprendizagem de máquina, IA, aprendizado não supervisionado, imagens médicas

Abstract

Artificial intelligence techniques is important for applications in the field of medicine, helping the process of medical diagnosis of several patients. With regard to medical imaging, diagnosis from radiographs requires the technical knowledge of specialized professionals, who are trained for years to obtain the level of expertise necessary to evaluate x-rays images. In addition, sometimes only images are not sufficent to correctly diagnose a pacient, requiring adicional information such as the patient's and family history. These characteristics makes the process of obtaing labels more expensive for an extensive database, which makes it difficult to apply a supervised learning method to classify images. This work applies machine learning tecniques using unsupervised learning approches in order to classify x-rays images, diffent methods are trained on images without anomalies, and can be used to assist doctors in evaluating x-rays images. Different techniques were tested and compared with the objective in verifying which one is better suited. In the end, it was possible to build a generic architecture that classifies x-rays imagens in two categories, with anomaly or without anomaly. The results found based on ROC-AUC were 0.547 with D *Discriminator probability* and 0.533 with MSE *Mean Squared Error*.

Keywords: X-ray, AI, machine learning, unsupervised learning, medical images

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Abreviaturas e Siglas

AM Aprendizagem de Máquinas.

BiGAN Bidirectional GAN.

CAE Computer-aided engineering.

D Discriminator Probabilty.

DCGAN Deep Convolutional GAN.

DNA Deoxyribonucleic acid.

GAN Generative Adversarial Network.

GPU Graphics Processing Unit.

IA Inteligência Artificial.

KLD Kullback-Leibler Divergence.

L1 L1 Regularization.

MSE Mean Square Error.

MURA Musculoskeletal Radiographs.

ROC-AUC Area-under-curve for the Receiver-Operator-Curve.

SSD Single Shot Detector.

TPU Tensor Processing Unit.

VAE Variational Autoencoder.

Capítulo 1

Introdução

O uso de técnicas de aprendizagem de máquina AM tem crescido bastante nos últimos anos, atingido resultados significativos em diversas áreas como o reconhecimento de imagens. Contudo, essa área requer a rotulação de uma vasta quantidade de dados para que o reconhimento de padrões seja uma tarefa de AM bem sucedida.

1.1 Problema

Em aplicações médicas, a rotulação de imagens é uma tarefa muito custosa, levando em consideração a natureza específica dos dados, ou seja, requer o conhecimento de profissionais especializados sobre a área de atuação, como por exemplo, detectar anomalias em imagens radiográficas. Decidir se uma determinada imagem apresenta alguma anormalidade é um problema que requer anos de treinamento. Outro detalhe relevante é o fato da imagem as vezes por si só não revelar um diagnóstico definitivo, sendo necessário ao especialista levar em consideração o contexto do paciente, tais como outros exames auxiliares, histórico médico e familiar. Assim, o uso de técnicas de AM proposto neste trabalho em imagens de raios-x, visa auxiliar o profissional de saúde a encontrar anormalidades nas radiografías, para que então um diagnóstico definitivo seja encontrado.

1.2 Justificativa

No que concerne imagens médicas, o avanço de técnicas de *deep learning* tem facilitado a introdução da AM na medicina (DAVLETSHINA et al., 2020), (LITJENS et al., 2017). Embora a aprendizagem supervisionada seja a abordagem mais difundida, o método não supervisionado está sendo introduzido. (SATO et al., 2018) propóe um método de detecção de anomalia não supervisionada em imagens de tomografía computadorizada utilizando *autoenconder*. O método alcançou um resultado de 0.87 na métrica ROC-AUC.

(UZUNOVA et al., 2019) utiliza o aprendizado não supervisionado VAE *Variational Autoencoder*, em que o método desenvolvido se baseia na aprendizagem de todo o conjunto de variabilidades de dados saudáveis, e a detecçaõ de patologias é caracterizada pela diferença em relação à norma que foi aprendida.

Em (DAVLETSHINA et al., 2020), utilizando a abordagem não supervisionada, foi investigado como diferentes métodos treinados em imagens sem anomalias podem ser usadas para auxiliar profissionais da saúde em avaliar imagens de raio-x. Foi utilizado um subconjunto de dados específico do MURA ***Mu****sculoskeletal* ***Ra****diographs* (GROUP, 2021), consistindo de radiografias essencialmente de mãos humanas. A metodologia aumenta a eficiência em realizar um diagnóstico reduzindo o risco de negligenciar regiões importantes.

1.3 Objetivos gerais

Neste trabalho, o objetivo é aplicar técnicas de AM na detecção de anomalias em imagens de raio-x, em específico, utilizando a abordagem não supervisionada, visando a redução de custos com relação a anotações de imagens radiográficas.

1.4 Objetivos específicos

O estudo é uma prosposta de ampliação do trabalho realizado em (DAVLETSHINA et al., 2020), criando um arquitetura genérica que aborda imagens de diferentes membros do corpo humano.

1.5 Organização do trabalho

O trabalho está estruturado da seguinte maneira: Em fundamentação teórica são discutidos os conceitos e os principais aspectos dos assuntos envolvidos na realização do projeto, essencialmente uma descrição sobre o que é raio-x e seu papel na medicina diagnóstica de hoje em dia. Outro assunto, é a aprendizagem de máquina; é discutido os principais conceitos sobre esse ramo da IA e as métricas que foram utilizadas para a realização do experimento. No tópico materiais e métodos são descritas as tecnologías utilizadas para a realização do experimento juntamente com a base de dados criada. Outro aspecto descrito, são os algorítimos utilizados no treinamento dos modelos de rede. Em experimentos é descrito todos os experimentos realizados, totalizando quatro. Cada experimento é o treinamento de um modelo de rede diferente. E por fim, em resultados e discussão, é descrito o que foi obtido após a realização dos experimentos, consequente, a discussão envolve a comparação dos resultados de cada experimento por meio de métricas específicas.

Capítulo 2

Fundamentação Teórica

Nesta sessão é abordado os conceitos bases para o entendimento da proposta estabelecida neste projeto. É apresentado conceitos sobre imagens radiográficas e sobre AM.

2.1 Radiografia

As imagens radiográficas são geradas através de uma fonte de radiações eletromagnéticas conhecida como raio-x. Seu uso é mais conhecido por ser aplicado no diagnóstico médico, mas também são usadas de forma abrangente em diversas outras áreas, como a astronomia por exemplo (GONZALEZ, 2009).

Origem

A descoberta dos raios-x aconteceu em 1895 pelo físico alemão Wilhelm Konrad Röntgen. Enquanto investigava os feixes de eletrons (também conhecido como raios catódicos) em descargas elétricas através de gases de baixa pressão, Röntgen descobriu que um tela revestida com material fluorescente, colocado fora de um tubo de descarga brilharia mesmo quando estivesse protegido da luz direta visível e ultravioleta da descarga gasosa. O físico deduziu que a radiação invisível do tubo passou pelo ar e fez com que a tela ficasse fluorescente. A radiação responsável pela fluorescência se originou do ponto onde o feixe de elétrons atingiu a parede de vidro do tubo de descarga. Objetos opacos colocados entre o tubo e a tela mostraram-se transparentes à nova forma de radiação; Röntgen demonstrou isso ao produzir uma imagem fotográfica dos ossos da mão humana (STARK, 2022).

Como raio-x é gerado

No que diz respeito a imagens médicas, os raio-x são gerados por meio de um tubo a vácuo com um cátodo e um ânodo. Primeiro, é necessário que o catodo seja aquecido, isso resulta na liberação de eletróns. Os elétrons então se movimentam em alta velocidade na direção do anodo, que está positivamente carregado. Dessa forma, em um momento os eletróns irão atingir um núcleo liberando energia na forma de radiação de raios-x (GONZALEZ, 2009).

Na medicina diagnóstica de imagens, os raios-x são usados para gerar imagens de tecido e estruturas dentro do corpo humano. Os raios passam pelo corpo do paciente e pelo um detector que está próximo do paciente, uma imagem então é formada representado objetos dentro do corpo. Os raio-x são uma forma de radiação eletromagnética, similar a luz. No entanto, ao contrário da luz, os raios-x tem uma quantidade de energia maior e pode atravessar a muitos objetos, incluindo o corpo humano (NIBIB, 2022).

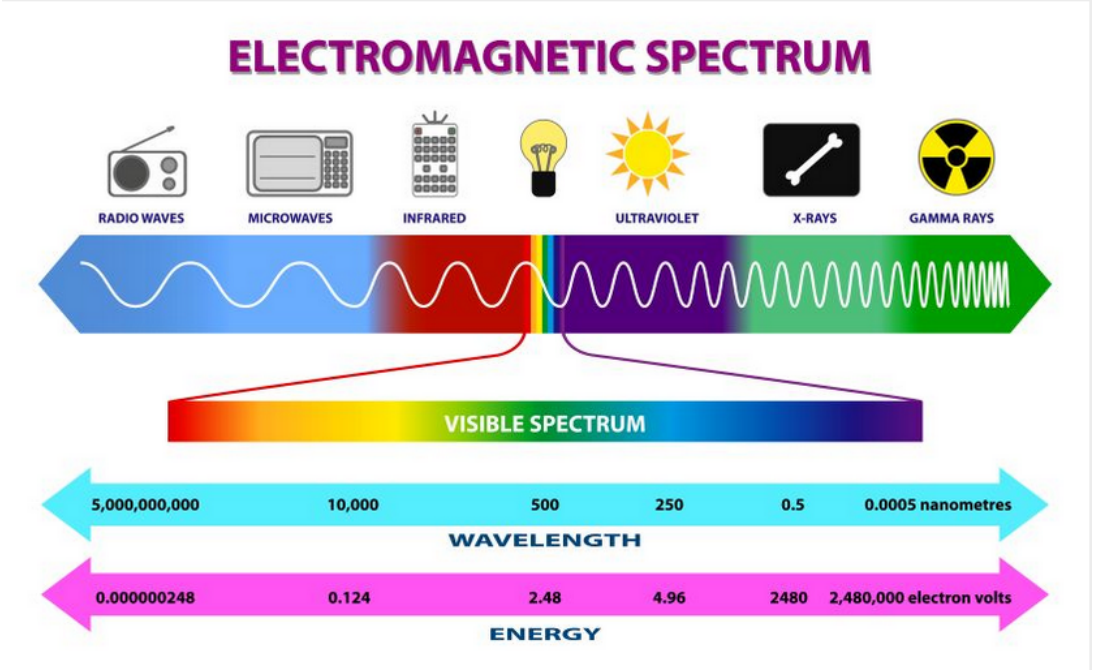

Figura 2.1: O espectro eletromagético e quantidade energética, adaptado de (NIBIB, 2022).

Há duas formas principais de se gerar imagens de raio-x, filme fotográfico, abordagem mais antiga; e digital, que por sua vez possue diversos tipo de detectores para produzir imagens digitais. As imagens de raio-x resultante desse processo são chamadas radiografia (NIBIB, 2022).

A importância das imagens de raio-x

No campo da medicina as imagens de raio-x são utilizadas para auxiliar no diagnóstico médico, como a detecção de fraturas ósseas, detecção de certos tipos de tumores, massas anormais, alguns tipos de lesões, calcificações, objetos no corpo, ou problemas dentais. Além da característica diagnóstica, também é utilizado como uma ferramenta terapêutica (NIBIB, 2022).

No tratamento do cancêr, existe terapia envolvendo radiação. Os raios-x, juntamente com outros tipos de radiação podem ser usados para destruir tumores cancerígenos pela danificação de seu DNA. Neste caso, a dose de raios-x é maior do que a usada para a geração de imagem diagnóstica (NIBIB, 2022).

Digitalização

Atualmente, em diversos departamentos radiográficos, radiografias digitais é a forma mais prática e comum de gerar imagens por raio-x, tomando a frente de radiografia de filme fotográfico. Tal mudança gera benefícos como o armazenamento desses dados e sua distribuição entre os médicos e os envolvidos. Outros benefícios incluem um maior rendimento do paciente, maior eficiência da dose e maior faixa dinâmica dos detectores digitais com possível redução da exposição à radiação do paciente (KORNER et al., 2007).

2.2 Aprendizagem de máquina

Aprendizagem de máquina é um ramo do campo da IA. Se trata da programação de computadores fazendo uso de uma vasta quantidade de dados de exemplo ou experiência passada. A aprendizagem diz respeito a execução de um programa de computador que otimiza parâmetros em um modelo computacional, visando melhorar seu desempenho em classificar dados. O modelo treinado então tem a tarefa de realizar predições futuras sobre novos dados (ALPAYDIN, 2020).

A função da ciência da computação está presente de duas formas: na fase de treinamento é necessário algoritmos eficientes que resolvam problemas de otimização, armazenamento e processamento de uma quantidade imensas de dados. Em um segundo momento, uma vez que o modelo computacional é treinado, sua solução algoritímica precisa ser eficiente sobre a inferência sobre novos dados (ALPAYDIN, 2020).

2.2.1 Tipos de aprendizagem

Visão geral sobre os principais métodos utilizados na aprendizagem de máquina.

Não supervisionada

Neste método, não há uma supervisão sobre os dados, ou seja, não há uma rotulação predefinindo as características dos dados de entrada. O objetivo é encontrar certas regularidades nos dados de entrada, em que certos padrões ocorrem com uma frênquencia maior em alguns tipos de dados do que em outros (ALPAYDIN, 2020).

Supervisionada

Essa abordagem faz uso de rótulos para classificar dados na fase de treinamento, e posteriormente, e feito a predição sobre esses rótulos em novos dados (LIBBRECHT; NOBLE, 2015).

Semi-supervisionada

Neste método é feita uma combinação entre as abordagens não supervisiona e supervisionada (LIBBRECHT; NOBLE, 2015).

2.2.2 Métricas

Nesta seção são decritas as métricas de aprendizagem que são utilizadas na abordagem não supervisionada, e que estão presentes na realização deste projeto.

Mean Squared Error MSE

Dado uma base de dados sempre é necessário avaliar o desempenho do método de aprendizagem estatístico, ou seja, medir o quão preciso as predições do modelo computacional estão próximas dos dados observados. É preciso quantificar através de métricas, o quão próxima está o valor previsto pelo modelo para uma determinada observação verdadeira. Dados essas características, um dos métodos mais comuns para traçar tais medidas é o MSE, dada pela fórmula abaixo 2.1 (JAMES et al., 2013):

$$MSE = \sum_{i=1}^{n} (y_i - \hat{f}(x_i))^2 \tag{2.1}$$

(i) n = pontos de dados

(ii) y_i = valores observados

(iii) $\hat{f}(x_i)$ = valores previstos

Na fórmula acima, $\hat{f}(x_i)$ é a predição que $\hat{f}$ encontra para a i-ésima observação. O valor resultante da MSE 2.1 será pequeno se as respostas previstas estão muito próximas

das respostas verdadeiras, e o valor em 2.1 será alto se para algumas observações, as predições e as respostas verdadeiras diferirem (JAMES et al., 2013).

L1 Regularization

Um questão que está presente na AM e que pode sugir durante o treinamento de um modelo computacional é o *overfitting*. Esse fenômeno é um problema que afeta a qualidade da predição e deve ser evitado, a fim de obter um modelo que seja capaz de realizar predições com o menor erro possível.

Esse problema surge a partir a estrutura da problema de AM em específico. Na fase de treinamento, o algoritimo de aprendizagem é treinado sob um conjunto de dados de treinamento, com o objetivo final de realizar predições sobre um novo conjunto de dados. O objetivo é tentar aumentar ao máximo a precísão com que o modelo realiza a predição sobre esses novos dados. Ao realizar a tentativa de melhorar o desempenho cada vez mais, em algum momento, ocorre o risco de ruídos (poluição nos dados) serem inseridos nos dados pela memorização de várias peculiaridades dos dados de treinamento ao invés de encontrar uma regra geral de predição. Esse fenômeno é conhecido como *overfitting* (DIETTERICH, 1995).

A métrica L1 é uma abordagem utilizada para reduzir o *overfitting*. É um das técnicas utilizadas para penalizar modelos complexos em (AM), em que nas redes neurais os pesos são reduzidos, e também ocorre o melhora do modelo para novas entradas de dados. Essa métrica é um escolha apropriada para dados que possuem um quantidade variada de características (HTTPS://WWW.ANALYTICSSTEPS.COM, 2021).

Discriminator Probability D

A métrica D está relacionada com os modelos de rede GAN *Generative Adversarial Autoencoder* referente a função de perda. Abaixo segue a equação para a métrica 2.2:

$$E_x[log(D(x))] + E_z[log(1 - D(G(z)))] \tag{2.2}$$

(i) *D(x)*, representa a perda no discriminador.

(ii) *G(z)*, representa a perda no gerador.

Nos modelos de rede GANs, as funções de perda podem ser categorizadas em duas partes, discriminador de perda e gerador de perda. O discriminador é treinado para classificar determinada imagem como um dado verdadeiro ou falso. *log(D(x)* se refere a probabilidade de que o gerador está classificando dados em formato imagem como real (DWIVED, 2021).

***Kullback-Leibler Divergence* KLD**

O KLD é uma métrica utilizada para medir a diferença entre duas funções de probabilidade de densidades (ZENG et al., 2014). É usada de forma ampla no campo da estátistica e, dado duas distriuições de densidade,busca medir a similaridade entre elas (HERSHEY; OLSEN, 2007).

Fórmula para o cálculo do KLD (HERSHEY; OLSEN, 2007):

$$D(f||g) = \int f(x) log \frac{f(x)}{g(x)} \, dx \tag{2.3}$$

A divergência é composta por três propriedades (HERSHEY; OLSEN, 2007):

(i) Auto semelhanca: $D(f \mid\mid f) = 0$.

(ii) Auto identificação: $D(f \mid\mid g) = 0$ se somente se f=g.

(iii) Positividade: $D(f \mid\mid g) \geq 0$ para todo f, g.

O KLD 2.3 é aplicado em alguns campos da AM como aspectos do reconhecimento de imagem e voz, determinando por exemplo, se duas imagens ou dois modelos acusticos são similidares (HERSHEY; OLSEN, 2007).

Capítulo 3

Materiais e Métodos

Com a proposta de ser um extensão do estudo (DAVLETSHINA et al., 2020) alguns ajustes e modificações foram realizados nos *scripts* originais, juntamente com a remodelação da base de dados utilizada no experimento.

3.1 Bases de dados

3.1.1 MURA

O MURA ***Mu****sculoskeletal* ***Ra****diographs dataset* é uma extensa base de dados referentes a radiografias de diferentes membros do corpo humano que compreendem respectivamente ao cotovelo, dedos, antebraço, mãos, umero, ombro e punho (GROUP, 2021). Toda a base de dados contém 40.561 imagens de 14.863 estudos, onde cada estudo foi rotulado por profissionais radiologistas como normal ou anormal (RAJPURKAR et al., 2017).

A base de dados do MURA é dividida em sete pastas referentes aos diferentes membros do corpo humano, contendo subpastas de uma extensa lista de pacientes que por sua vez podem possuir radiografias positivas ou negativas (estudos), positiva refere-se a radiografias com alguma anormalidade detectada.

A Figura 3.1 mostra algumas imagens radiográficas da base de dados MURA.

3.1.2 Mini-MURA

No MURA, cada diretório que refere a um determinado membro do corpo humano, possui tamanho em *mega bytes* extramente grandes e indivualmente diferentes; ou seja, a quantidade de dados em cada diretório não são os mesmos. A fim de reduzir a grande quantidade de dados e ao mesmo tempo balanceá-los foi criado um Mini-MURA, uma base de dados que compreende exatamente a mesma subdivisão da base de dados original

Figura 3.1: Exemplo de imagens radiográficas da base de dados do MURA.

mas com uma proporção menor. Foi selecionado 400 pacientes por membro, totalizando um *dataset* de aproximadamente 649 MB.

No total, a base de dados do Mini-Mura (ALENCAR, 2021) contém 7.932 imagens divididas em 2.972 estudos de 2.800 pacientes. Cada estudo é classificado como positivo ou negativo, positivo significa que uma anomalia foi detectada no estudo, e negativo representa a ausência de anomalias. Há um total de 1.736 estudos positivos na base de dados do Mini-Mura.

A quantidade de imagens escolhida para compor o Mini-MURA tem como base (DAVLETSHINA et al., 2020). Neste estudo 5.543 imagens radiográficas de mãos de 2.018 estudos referentes a 1.945 pacientes foram utilizadas no experimento. Com a proposta de extender este estudo, foi selecionado um número aproximada mas abragendo todos os membros do MURA, não somente mãos.

Devido a limitação de recursos computacionais gerenciados pelo Google Colab (GOOGLE, 2021) - ambiente em a implementação do projeto foi realizado - foi necessária a

redução da base de dados comparado a (DAVLETSHINA et al., 2020).

3.2 Software e sistemas

Para a implementação do projeto foi utilizado o Google Colab (GOOGLE, 2021), ambiente em que é possível a qualquer usuário executar códigos em código python através de um navegador de internet. É amplamente utilizado para análise de dados e aprendizagem de máquina. O google Colab não exigi configuções específicas de hardware do computador pessoal do usuário, ao mesmo tempo, fornece acesso a recursos computacionais gratuitos como *Graphics Processing Unit* GPU e *Tensor Processing Unit* TPU.

Tendo em vista que os recursos computacionais fornecidos pelo Google Colab são compartilhados com diversos usuários ao redor do mundo evidentemente há uma limitação quanto ao uso desses recursos. Há limitações quanto tempo de uso da GPU e TPU e também ao tipo de *hardware* utilizado (GOOGLE, 2021).

Os recursos computacionais oferecidos pelo Google Colab na versão gratuita não foi o suficente para pode realizar o treinamento dos modelos de forma adequada, por essa razão, foi optado a versão Google Colab Pro, em que há a possibilidade de usar a GPU por mais tempo. No périodo em que os experimentos foram realizados o custo mensal da versão Pro foi de R$ 58,00.

Todos os modelos foram treinados utilizando os *notebooks* do Google Colab. Configurações: GPU de back-end do *Google Compute Engine* em Python3, memória RAM 12.69 GB e Disco de 156.99 GB.

Para realizar a última etapa do preprocessamento *offline* foi utilizado o software Photoshop™, em que, após as imagens serem processadas nas etapas anteriores a fim de reduzir ruídos, as imagens então são segmentadas. O que for mais predominate como uma mão ou ombro por exemplo, é retirado e colado em um fundo preto. Essa última fase do preprocessamento *offline* é chamado de *segmantic segmantation*.

3.3 Métodos

Descrição dos algoritmos utilizados e etapas da realização do projeto.

3.3.1 Preprocessamento

Os dados extraídos na vida real geralmente são muito ruidosos, e isso é um problema muito grande no que se refere a abordagen de aprendizado de máquina não supervisionado para a detecção de anomomalias. A remoção de ruidos é necessária para que durante

o aprendizado não supervisionado esses ruidos não sejam classificados erroneamente como anomalias. Também é importante que durante a fase de remoção de ruídos anomalias não sejam removidas erroneamente contaminando a real classificação dos dados. O preprocessamente é dividido em duas etapas, *offline* em que o preprocessamento é feito uma vez e armazenado em disco, e o *online*, em que o preprocessamento é feito *on-the-fly* enquanto os dados são carregados (DAVLETSHINA et al., 2020). Todos os passos são descritos detalhadamente abaixo na Figura 3.2:

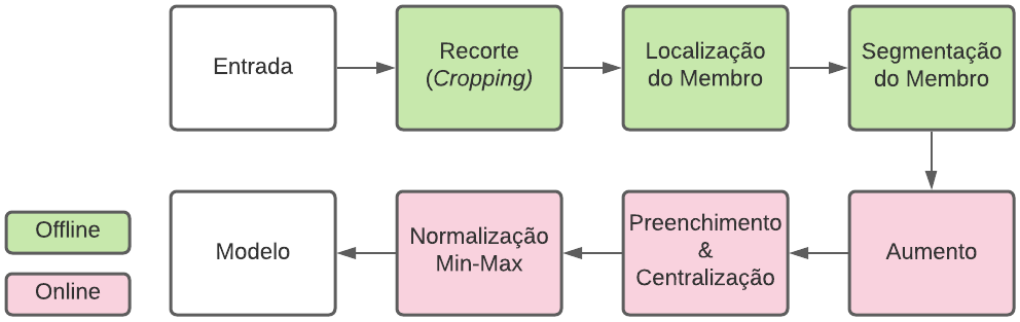

Figura 3.2: Fase do preprocessamento dos dados. Passos destacados em verde são realizados uma vez e armazenados em disco. Passos destacados em rosa são realizados *on the-fly*.

Recorte O primeiro passo do preprocessamento é detectar contornos nas imagens. Para isso é aplicado OpenCV's usando *Otsu binarization*. Funciona relativamente bem, mas pode apresentar falhas para imagens muito inclinadas. O algoritmo utilizado em (DAVLETSHINA et al., 2020) para essa fase do preprocessamento foi adaptado para que todas as radiografias de cada membro do Mini-MURA passase por essa etapa, ao ínvés de somente um membro específico (conjunto de radiografias de um mesmo tipo) conforme em realizado no mesmo (DAVLETSHINA et al., 2020). Um exemplo com os resultados pode ser visto na Figura 3.3.

Localização do Membro Para essa etapa 200 imagens de cada membro foram selecionadas e rotuladas manualmente, totalizando 1.400 imagens. 75% foram separadas para treinamento e 25% para teste. Através dessa pequena base de dados foi ajustado um modelo pré-treinado para detecção de objetos *single shot detector* SSD com *MobileNet* usando TensorFlow (LIU et al., 2016).

Segmentação do membro No último passo do preprocessamento offline a segmentação de primeiro plano é utilizado usando a função *subject select* do Photoshop™, utilizando o modo processamento em lote.

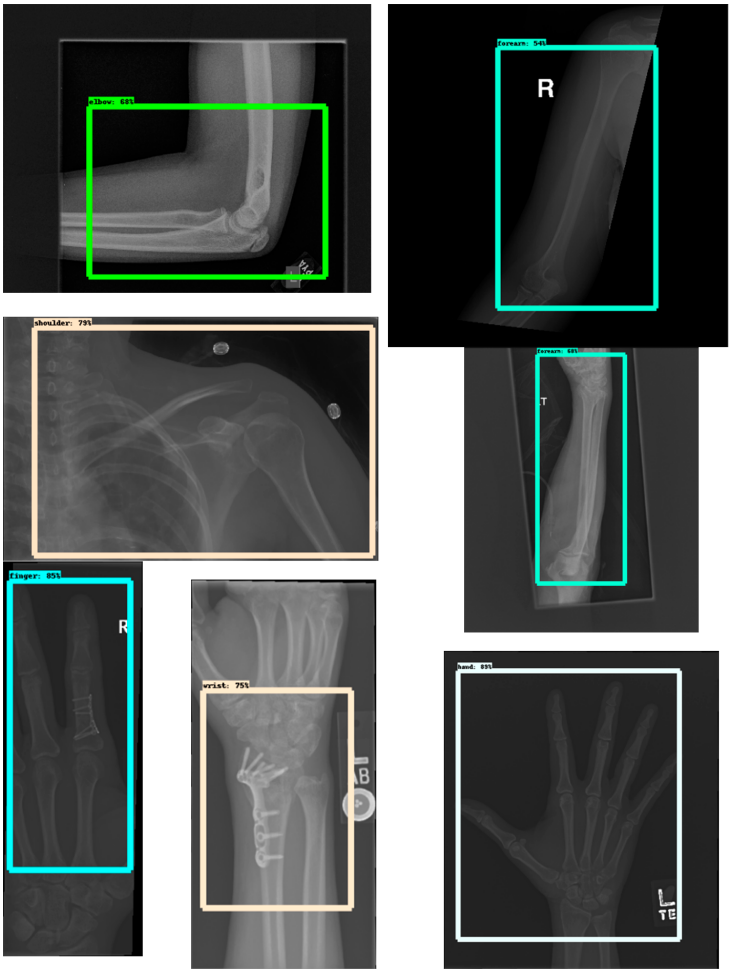

Figura 3.3: Resultado após a execução do detector de objetos em imagens do Mini-MURA.

3.3.2 Modelos de redes utilizadas

Nesta seção é descrita os modelos utilizados no treinamento não supervisionado nos dados do Mini-MURA, que compreende somente pacientes sem anomalias detectadas.

3.3.3 Autoencoders

Os *Autoencoders* são uma abordagem do aprendizado não supervisionado que aplica a técnica conhecida como *backpropagation* onde os valores alvos são configurados como entradas (NG et al., 2011).

Neste modelo de rede, ocorre tentativas de aprendizado de uma aproximação da função indentidade, uma vez que $\hat{x}$ é similar a x Figura 3.4.

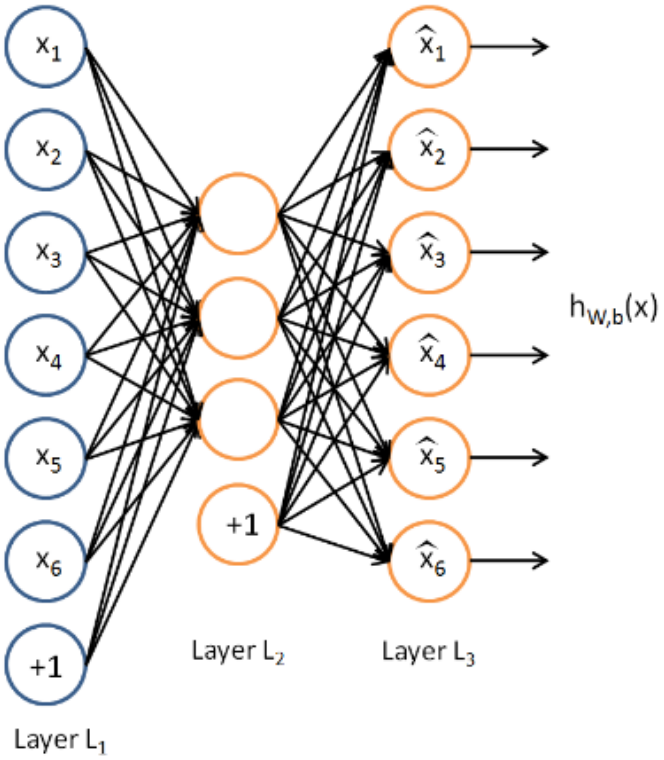

Figura 3.4: Exemplo de uma rede *Autoencoder*, adaptado de (NG et al., 2011).

Neste trabalho é utilizado uma variação de *autoencoder*, chamado *Variational Autoencoder* VAE, em que é realizado o treinamento do modelo de rede através de dados não rotulados. Os *encoders* utilizam um mecanismo de perda de reconstrução em que a entrada na rede também é utilizado como alvo, então é avaliado o quão bem a entrada é reconstruída.

A VAE consiste de uma rede codificadora e outra rede decodificadora. A VAE faz o uso do método gradiente descendente para aprender uma inferência aproximada. A rede codificadora com os parâmetros ϕ aprende uma compressão eficiente do dado em um espaço dimensional menor, que mapeia o dado X em uma variável contínua latente Z. A rede decodificadora com parâmetros θ usa a variável latente para gerar dados que mapeiam Z para um dado reconstruido $\hat{X}$ Figura 3.5.

A ideia central na VAE é usar a probabilidade de distribuição *P(X)* para amostrar pontos de dados que correspondem a essa distribuição, onde X representa uma variável aleatória dos dados. O objetivo é reconstruir o dado de entrada o mais preciso possível, isto é, maximixar a probabilidade de *P(x)* (YANG et al., 2019).

Na VAE, o treinamento é realizado de forma a evitar *overfitting* e garantir que espaços latentes tenham boas propriedades que permitem o processo generativo (ROCCA, 2021). Neste modelo, cada entrada no conjunto de dados é mapeada para uma distribuição

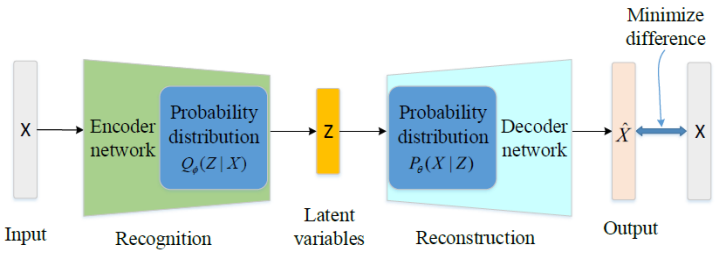

Figura 3.5: Arquitetura do modelo de rede VAE, adaptado de (YANG et al., 2019).

Gaussiana, caracterizado pela sua respectiva média e covariância. (KINGMA; WELLING, 2013).

3.3.4 GAN

O modelo de rede GAN *Generative Adversarial Network*, diz respeito a duas sub-redes, um gerador G, e um discriminador D, que funcionam como antagonistas em um jogo a dois. O gerador obtem ruídos aleatórios como entrada e gera amostras no domínio de destino. O discriminador obtém pontos de dados reais, e ao mesmo tempo dados gerados, e ao final tem como tarefa distinguir entre dados reais e falsos. As sub-redes são treinadas de forma alternada, se bem-sucedidas, o gerador pode posteriormente ser usado para amostrar a partir da distribuição de dados e o discriminador pode ser usada para decidir se uma amostra é retirada da distribuição de dados fornecida (DAVLETSHINA et al., 2020).

O DCGAN *Deep Convolutional GAN* é uma extensão da arquitetura GAN com redes neurais convolucionais. Funciona de forma similiar ao CAE *Computer-aided engineering*, as duas redes contém convoluções (discriminador) e convoluções transpostas (gerador) ao invés de camadas conectadas proposta pela arquitetura GAN (DAVLETSHINA et al., 2020) .

O BiGAN *Bidirectional GAN* extende DCGAN por um *encoder* E, que codifica a imagem real em espaços latentes. O discriminador é fornecido com ambos, a imagem real e falsa, justamente com seus respectivos códigos latentes (DAVLETSHINA et al., 2020).

α - *GAN*, esse modelo compreende quatro sub-redes (DAVLETSHINA et al., 2020):

- Um codificador E(x) que transforma uma imagem real em uma representação latente.
- Um código discriminador CO(z) que distingue entre representações latentes produzidas pelo codificador e o ruído aleatório usado como gerador de entrada.
- Um gerador G(z) que gera uma imagem do amostrado aleatoriamente z, ou da imagem codificada.

- Um discriminador $D(\mathrm{x})$ que distingue entre imagens reais reconstruidas $G(E(\mathrm{x}))$, e imagens geradas $G(\mathrm{z})$.

3.4 Experimentos

A abordagem usada neste projeto é o aprendizado não supervisionado, assim sendo, o treinamento foi realizado somente em imagens negativas, isto é, imagens sem anomalias detectadas. Os dados foram divididos por paciente ao invés de uma divisão por estudo ou imagem, para garantir que não se tenha imagens de pacientes nos dados de treinamento, e outra imagem do mesmo paciente na parte de teste ou validação (DAVLETSHINA et al., 2020).

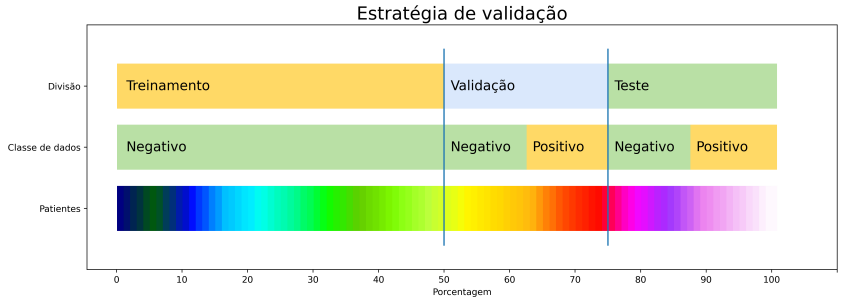

Figura 3.6: Visualização da divisão dos dados. Negativo, representa pacientes com estudos normais, e positivo o contrário. Observe que no treinamento não há imagens com anomalias, imagens anormais não são usadas para o treinamento. O espectro de cores mostram onde a maior parte dos dados estão concentrados (DAVLETSHINA et al., 2020).

Seja P o conjunto de todos os pacientes, e P+ o conjunto dos pacientes com estudos que foram rotulados como anormal, o restante então é representado por P- := P\P+. Para o restante do conjunto de teste e validação, as classes foram balanceadas, assim, P+ foi distribuido uniformemente de forma aleatória entre teste e validação (DAVLETSHINA et al., 2020). No total, foram obtidas 7932 imagens, sendo que 50% (3966) foram utilizadas no treinamento, aproximadamento 25% (3966) foram utilizadas na validação e aproximadamente 25% (1968) utilizadas no teste. Uma visualização do esquema pode ser observado na Figura 3.6.

Foram realizados quatro experimentos utilizando diferentes modelos de redes (VAE, DCGAN, BiGAN e α - *GAN*). Após o treinamento foi realizado uma comparação manual de hiperparâmetro no conjunto de validação e os melhores modelos foram selecioados com respeito ao ROC-AUC *Area-under-Curve for the Receiver-Operator-Curve*.

3.4.1 Experimento 1 Rede VAE

No primeiro experimento foi utilizada a rede VAE para o treinamento dos dados na abordagem de aprendizado não supervisionado. Os dados foram divididos em treinamento, validação e teste. Conforme especificado na seção 3.4, foram 3,966 imagens utilizadas no treinamento, 1,998 imagens utilizadas na validação e 1,968 imagens no teste.

A Rede VAE é um dos modelos mais utilizados se tratando da abordagem não supervisionada, e apresenta bons resultados quando aplicadado a problemas complexos tais como reconhecimento de digitos escritos a mão, faces, segmentação, predição do futuro dado imagens estáticas, entre outros (DOERSCH, 2016). Assim, dado a complexidade do problema tratado neste experimento, foi treinado um modelo de rede VAE, assim como também ocorreu em (DAVLETSHINA et al., 2020).

A duração do treinamento durou uma hora, quarenta e cinco minutos e trinta e sete segundos, sendo executado na plataforma Google Colab (GOOGLE, 2021).

O propósito é comparar o desempenho alcançado pela rede com outras que também são utilizadas na abordagem não supervisionada. Por meio do resultado alcançado na ROC-AUC ao final do experimento será possível analisar o desempenho alcançado por meio de uma comparação com outros modelos de rede.

3.4.2 Experimento 2 Rede DCGAN

No segundo experimento foi utilizada a rede DCGAN para o treinamento dos dados na abordagem de aprendizado não supervisionado. Foi utilizada a mesma quantidade de imagens estabelecida em todos os outros experimentos, assim como descrito na seção 3.4, (3,966 imagens utilizadas no treinamento, 1,998 imagens utilizadas na validação e 1,968 imagens no teste) .

O modelo de rede DCGAN tem resultados consistentes se tratando de treinamento em imagens variadas, em um método em que consiste em apreender uma hierarquia de representações (RADFORD; METZ; CHINTALA, 2015). Considerando que as imagens envolvidas neste projeto apresenta características variadas tais como; radiografias de diferentes membros do corpo humano, diferentes qualidades de imagens; isto faz da rede DCGAN um candidado para o treitamento utilizando esse tipo de dado.

A duração do treinamento durou exatamente duas horas, cinco minutos e vinte e nove segundos, sendo executado na plataforma Google Colab (GOOGLE, 2021).

O propósito, assim como em todos os experimento neste projeto é comparar o resultado encontrado alcançado por meio da ROC-AUC e analisar qual será a rede mais apropriada para a classificação de imagens no contexto de anomalias em radiográfias.

3.4.3 Experimento 3 Rede BiGAN

No terceiro experimento foi utilizado a rede BiGAN para o treinamento a sob a abordagem não supervisionada. Assim como nos experimentos anteriores foi utilizado 3,966 imagens na parte de de treinamento, 1,998 imagens para parte de validação e 1,968 imagens na parte de teste. Em termos de proporção essa divisão dos dados corresponde a aproximadamente 50% de dados utilizados no treinamento, 25% de dados utilizados na validação e 25% de dados utilizados no teste 3.4.

A duração do treinamento durou exatamente uma hora e onze minutos, sendo executado na plataforma Google Colab (GOOGLE, 2021). O tempo de treinamento obtido foi o mais rápido dentre todos os outros modelos utilizados neste projeto, contudo, essa métrica não é utilizada como parâmetro de desempenho entre os modelos de rede.

O objetivo neste experimento é comparar o desempenho do modelo de rede treinada com os outros modelos de rede por meio da ROC-AUC, e estabelecer a viabilidade frente aos modelos concorrentes.

3.4.4 Experimento 4 Rede α - GAN

No quarto experimento foi utilizado a rede α - GAN para o treinamento utilizando a abordagem não supervisionada. Assim como nos experimentos anteriores foi utilizado 3,966 imagens na parte de de treinamento, 1,998 imagens para parte de validação e 1,968 imagens na parte de teste. Em termos de proporção essa divisão dos dados corresponde a aproximadamente 50% de dados utilizados no treinamento, 25% de dados utilizados na validação e 25% de dados utilizados no teste 3.4.

A duração do treinamento durou exatamente três horas e cinquenta e um minutos, sendo executado na plataforma Google Colab (GOOGLE, 2021). O tempo de treinamento foi o mais demorado comparado aos outros experimentos, contudo, essa métrica não é considerada para a comparação de desempenho entre os modelos de rede, somente a curva ROC-AUC.

Ao final do treinamento foi obtido o valor relacionado a curva ROC-AUC, métrica de desempenho que foi comparada com a obtido no experimentos anteriores. Através da comparação é possível estabalecer qual modelo de rede tem o melhor desempenho frente a tarefa específica de classificar imagens de raio-x em, imagens de raio-x com anomalia ou sem anomalia.

Capítulo 4

Resultados e Discussão

Todos os modelos foram treinados utilizando como parâmetro de *performance* o (ROC-AUC). Para cada modelo foi relatado resultados de pontuação para diferentes tipos de métricas: MSE, L1, KLD, D.

A ROC-AUC foi aplicado no conjunto de dados de teste para verificar a precisão com que os modelos do experimento classificam os dados na sua predicão. Posteriormente os resultados encontrados foram comparados com os do (DAVLETSHINA et al., 2020) para verificar se a variedade de dados do experimento afeta os resultados finais, em termos de desempenho da ROC-AUC.

Tabela 4.1: Resultados da ROC-AUC com os diferentes modelos de redes.

	MSE	L1	KLD	D
VAE	0.525	0.525	0.463	-
DCGAN	-	-	-	0.517
BiGAN	0.547	-	-	0.505
α - Gan	0.524	-	-	0.533

O melhor resultado obtido foi através do modelo de rede α - Gan, com o valor da ROC-AUC em 0.533, utilizando a métrica D *Discriminator probability*, como pode ser observado na tabela 4.1. Já com relação a métrica MSE o modelo de rede com o melhor desempenho foi o BiGAN com o valor da ROC-AUC em 0.547, tabela 4.1 .

Através da comparação dos resultados encontrados com os do (DAVLETSHINA et al., 2020) pode-se verificar que os resultados dos modelos de redes são aproximados, inclusive o modelo com o melhor desempenho referente ao valor da ROC-AUC é o mesmo (α - Gan) em ambos os projetos, em relação a métrica D. Mesmo o experimento contendo um

variedade maior de dados com relação a (DAVLETSHINA et al., 2020) não houve uma diminuição significativa da precisão com relação a classificação das imagens.

Através dos experimentos é possivel observar que mesmo com uma variedade de tipos de dados é possível aplicar problemas de classificação. As imagens de raios-x de diferentes membros do corpo humano pode ser classificada binariamente após o treinamento, onde todas essas radiografias formam um único *corpus*.

Capítulo 5

Conclusão

Esse experimento consistiu na expansão do trabalho (DAVLETSHINA et al., 2020) envolvendo uma parcela de todos os tipos de dados do MURA (***Mu****sculoskeletal* ***Ra****diographs*) *dataset*. Expecificamente radiografias do cotovelo, dedos, antebraço, mãos, úmero, ombro e punho. Foram empregados métodos de aprendizagem não supervisionada para a detecção de anomalias nas imagens de raios-x utilizando *auto-encoders* e *GAN*.

Para dar viabilidade ao experimento, foi aplicado uma minunciosa etapa de pré-processamento envolvendo várias camadas, a fim de diminuir a maior quantidade de ruidos possíveis nos dados. O procedimento utilizado foi semelhante ao aplicado em (DAVLETSHINA et al., 2020), mas contendo adaptações algorítimas para realizar o pré-processamento de uma base de dados mais diversa.

O treinamento não supervisionado envolveu modelos *auto-encoders* e *GAN*. Os resultados finais foram comparados utilizando as mesmas métricas em (DAVLETSHINA et al., 2020)

Ao fim do experimento, foi verificado que os resultados encontrados foram próximos, levando a concluir que, no treinamento envolvendo aprendizagem supervisionada com uma base de dados mais diversificada, a precisão da classificação permanece estável, desde que as etapas de pré-processamento sejam preservadas, como pode ser observado pelos resultados encontrados mostrados na tabela 4.1.

Como possíveis melhoras para um trabalho futuro, há o aspecto de sofisticar as etapas de pré-processamento, sendo que uma variedade de dados de qualidades variadas torna-se um desafio para a eliminação de ruídos de forma automatizada, necessário para elevar a precisão da classificação sobre o ROC-AUC. Outro aspecto é o aumento da base de dados para elevar a precisão de classficação dos modelos de redes.

Referências

ALENCAR, R. *Mini-MURA*. 2021. <https://drive.google.com/drive/folders/1q2-4m-im0zSRqalnCK-_1LQT__LLXRu2>. Acessado em: 18-10-2021.

ALPAYDIN, E. *Introduction to machine learning*. [S.l.]: MIT press, 2020.

DAVLETSHINA, D.; MELNYCHUK, V.; TRAN, V.; SINGLA, H.; BERRENDORF, M.; FAERMAN, E.; FROMM, M.; SCHUBERT, M. Unsupervised anomaly detection for x-ray images. *arXiv preprint arXiv:2001.10883*, 2020.

DIETTERICH, T. Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, ACM New York, NY, USA, v. 27, n. 3, p. 326–327, 1995.

DOERSCH, C. Tutorial on variational autoencoders. *arXiv preprint arXiv:1606.05908*, 2016.

DWIVED, A. H. *Understanding GAN Loss Functions*. 2021. <https://neptune.ai/blog/gan-loss-functions>. Acessado em: 13-08-2022.

GONZALEZ, R. C. *Digital image processing*. [S.l.]: Pearson education india, 2009.

GOOGLE. *Google Colab*. 2021. <https://colab.research.google.com/>.

GROUP, S. M. *MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal*. 2021. <https://stanfordmlgroup.github.io/competitions/mura/>. Acessado em: 18-10-2021.

HERSHEY, J. R.; OLSEN, P. A. Approximating the kullback leibler divergence between gaussian mixture models. In: IEEE. *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*. [S.l.], 2007. v. 4, p. IV–317.

HTTPS://WWW.ANALYTICSSTEPS.COM, N. T. *L2 and L1 Regularization in Machine Learning*. 2021. Disponível em: <https://www.deeplearningbook.com.br/?s=l1>.

JAMES, G.; WITTEN, D.; HASTIE, T.; TIBSHIRANI, R. *An introduction to statistical learning*. [S.l.]: Springer, 2013. v. 112.

KINGMA, D. P.; WELLING, M. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

KORNER, M.; WEBER, C. H.; WIRTH, S.; PFEIFER, K.-J.; REISER, M. F.; TREITL, M. Advances in digital radiography: physical principles and system overview. *Radiographics*, v. 27, n. 3, p. 675, 2007.

LIBBRECHT, M. W.; NOBLE, W. S. Machine learning applications in genetics and genomics. *Nature Reviews Genetics*, Nature Publishing Group, v. 16, n. 6, p. 321–332, 2015.

LITJENS, G.; KOOI, T.; BEJNORDI, B. E.; SETIO, A. A. A.; CIOMPI, F.; GHAFOORIAN, M.; LAAK, J. A. V. D.; GINNEKEN, B. V.; SÁNCHEZ, C. I. A survey on deep learning in medical image analysis. *Medical image analysis*, Elsevier, v. 42, p. 60–88, 2017.

LIU, W.; ANGUELOV, D.; ERHAN, D.; SZEGEDY, C.; REED, S.; FU, C.-Y.; BERG, A. C. Ssd: Single shot multibox detector. In: SPRINGER. *European conference on computer vision*. [S.l.], 2016. p. 21–37.

NG, A. et al. Sparse autoencoder. *CS294A Lecture notes*, v. 72, n. 2011, p. 1–19, 2011.

NIBIB. *X-rays*. 2022. <https://www.nibib.nih.gov/science-education/science-topics/x-rays>. Acessado em: 14-08-2022.

RADFORD, A.; METZ, L.; CHINTALA, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

RAJPURKAR, P.; IRVIN, J.; BAGUL, A.; DING, D.; DUAN, T.; MEHTA, H.; YANG, B.; ZHU, K.; LAIRD, D.; BALL, R. L. et al. Mura: Large dataset for abnormality detection in musculoskeletal radiographs. *arXiv preprint arXiv:1712.06957*, 2017.

ROCCA, J. *Understanding Variational Autoencoders*. 2021. <https://towardsdatascience.com/understanding-variational-autoencoders-vaes-f70510919f73>. Acessado em: 18-10-2021.

SATO, D.; HANAOKA, S.; NOMURA, Y.; TAKENAGA, T.; MIKI, S.; YOSHIKAWA, T.; HAYASHI, N.; ABE, O. A primitive study on unsupervised anomaly detection with an autoencoder in emergency head ct volumes. In: SPIE. *Medical Imaging 2018: Computer-Aided Diagnosis*. [S.l.], 2018. v. 10575, p. 388–393.

STARK, G. *X-ray radiation beam*. 2022. <https://www.britannica.com/science/radio-wave>. Acessado em: 14-08-2022.

UZUNOVA, H.; SCHULTZ, S.; HANDELS, H.; EHRHARDT, J. Unsupervised pathology detection in medical images using conditional variational autoencoders. *International journal of computer assisted radiology and surgery*, Springer, v. 14, n. 3, p. 451–461, 2019.

YANG, Y.; ZHENG, K.; WU, C.; YANG, Y. Improving the classification effectiveness of intrusion detection by using improved conditional variational autoencoder and deep neural network. *Sensors*, MDPI, v. 19, n. 11, p. 2528, 2019.

ZENG, J.; KRUGER, U.; GELUK, J.; WANG, X.; XIE, L. Detecting abnormal situations using the kullback–leibler divergence. *Automatica*, Elsevier, v. 50, n. 11, p. 2777–2786, 2014.