

**FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO**

# **Reidentificação não supervisionada de pacientes em imagens de raio-X**

**Relatório Final do Projeto Integrador**

**Fábio Oliveira Rodrigues**  
**202107030**



Licenciatura em Bioengenharia

**Tutor na U.Porto:** João Pedrosa

**Orientadores no INESC TEC:** João Pedrosa, Sofia Pereira

18 de junho de 2024

# Resumo

O diagnóstico precoce de patologias é essencial, permitindo que o tratamento seja realizado em estágios iniciais da doença. Para tal, existem vários métodos de imagiologia clínica que permitem a extração de imagens que dão informações relevantes sobre o estado de saúde do paciente. As imagens de raio-X do tórax tornam-se indispensáveis no diagnóstico precoce e no acompanhamento de patologias como pneumonia, tuberculose, doenças pulmonares obstrutivas crônicas e condições cardíacas. Atualmente, existem várias ferramentas que recorrem a inteligência artificial de forma a detetar automaticamente estas patologias, facilitando o trabalho realizado pelos profissionais de saúde. Por vezes, a informação que permite associar uma imagem de raio-X a um dado paciente não existe ou encontra-se incompleta, levando a complicações clínicas e impedindo o uso das ferramentas automáticas de deteção. Como tal, este projeto tem como objetivo o treino de uma rede neuronal através de uma abordagem não supervisionada que permite a reidentificação de pacientes, utilizando uma abordagem contrastiva, procedendo-se à comparação dos resultados com uma aprendizagem contrastiva supervisionada. De forma a atingir este fim, estudou-se o efeito do uso de diferentes transformações nas imagens do dataset *ChestX-ray14*, bem como o ajuste de outros parâmetros do treino. Apesar de os resultados da aprendizagem supervisionada serem substancialmente melhores à aprendizagem não supervisionada, esta poder-se-á demonstrar vantajosa em situações cujo o acesso aos rótulos das imagens é limitado ou quando estas apresentam uma baixa qualidade.

**Palavras-chave:** imagiologia, raio-X, reidentificação, rede neuronal, supervisionado, não supervisionado, aprendizagem contrastiva, *ChestX-ray14*.

# Agradecimentos

A realização deste trabalho não teria sido possível sem o apoio e a colaboração de diversas pessoas e instituições, às quais gostaria de expressar a minha profunda gratidão. Primeiramente, gostaria de agradecer aos meus orientadores, João Pedrosa e Sofia Pereira, pela orientação e acompanhamento regular durante todo o período da realização do projeto e pela sua disponibilidade. Gostaria de expressar minha gratidão ao INESC TEC, pela infraestrutura e recursos disponibilizados que permitiram a realização desta pesquisa e a todos os membros do Laboratório de Imagem Biomédica pela partilha de conhecimentos. Por último, queria agradecer à professora Ana Maria Mendonça docente da unidade curricular "Projeto Integrador em Engenharia Biomédica" pela organização da mesma e pela resposta a dúvidas relativas ao formato esperado do relatório.

# Conteúdo

<b>1</b>	<b>Introdução</b>	<b>1</b>
1.1	Enquadramento . . . . .	1
1.2	Motivação . . . . .	1
1.3	Contribuição do projeto para os objetivos de desenvolvimento sustentável . . . . .	2
1.4	Estrutura do trabalho . . . . .	2
<b>2</b>	<b>Conhecimentos de base</b>	<b>4</b>
2.1	<i>Deep learning</i> em imagem médica . . . . .	4
2.1.1	Aprendizagem supervisionada, não supervisionada e semi-supervisionada . . . . .	6
2.1.2	Aprendizagem contrastiva . . . . .	6
2.2	Estado da arte . . . . .	7
<b>3</b>	<b>Metodologia</b>	<b>9</b>
3.1	Descrição do dataset . . . . .	9
3.2	Reidentificação de pacientes . . . . .	10
3.2.1	Arquitetura . . . . .	10
3.2.2	Treino . . . . .	10
3.2.3	Transformações . . . . .	11
3.3	Métricas de avaliação . . . . .	13
<b>4</b>	<b>Resultados e Discussão</b>	<b>15</b>
4.1	Aprendizagem supervisionada e não supervisionada . . . . .	15
<b>5</b>	<b>Conclusão</b>	<b>18</b>
	<b>Referências</b>	<b>20</b>

# Capítulo 1

## Introdução

### 1.1 Enquadramento

Este projeto foi realizado no âmbito da unidade curricular "Projeto Integrador em Engenharia Biomédica" do 3º ano da licenciatura em Bioengenharia da FEUP. O projeto foi desenvolvido no Laboratório de Imagem Biomédica do INESC TEC e teve o acompanhamento por parte do João Pedrosa e Sofia Pereira.

Este trabalho serviu como ponto de partida para a aprendizagem na área de *deep learning*. Assim, o projeto permitiu uma aprendizagem gradual, desde os conceitos mais simples até aos mais complexos. Para além de perceber a parte teórica que existe por detrás do funcionamento das redes neuronais, também se estudou a sua aplicação num caso prático específico.

### 1.2 Motivação

O diagnóstico precoce de patologias é crucial, uma vez que permite a realização de tratamento em estágios não avançados da doença, nos quais esta ainda não provocou muitos danos irreversíveis. Radiografias torácicas (CXR), ecografias e tomografia computadorizada são alguns dos métodos de imagiologia clínica que permitem extrair inúmeras características que podem servir como indicadores para certas patologias. As CXR são particularmente valiosas no diagnóstico e monitorização de doenças como pneumonia, tuberculose, doenças pulmonares obstrutivas crónicas e condições cardíacas. Com os avanços tecnológicos, ferramentas de inteligência artificial (IA) têm sido desenvolvidas para detetar automaticamente essas patologias, aumentando a precisão diagnóstica e aliviando a carga de trabalho dos profissionais de saúde. Para além do baixo tempo de processamento necessário, permitem a identificação de anomalias que podem ser invisíveis ao ser humano, permitindo intervenções mais rápidas e eficazes. Além disso, a utilização de IA leva à redução da variância dos diagnósticos feitos por diferentes profissionais de saúde, o que conduz a maior consistência.

Em vários hospitais e em centros de radiologia, devido ao elevado número de exames de raio-X do tórax que são realizados diariamente, existe uma grande probabilidade de estas serem identificadas erroneamente (haver problemas na atribuição do número de identificação, por exemplo), existindo inúmeras imagens que não podem ser utilizadas para o treino de ferramentas de IA, uma vez que se desconhece a que indivíduo estão associadas.

Como tal, de forma a ser possível resolver este problema e evitar que falhas no sistema e na base de dados levem a consequências severas, este projeto tem como objetivo o treino de uma rede neuronal recorrendo a uma abordagem não supervisionada que permitirá identificar se duas imagens correspondem ao mesmo paciente. Desta forma, caso exista já alguma imagem de um determinado paciente corretamente identificada, será possível encontrar as restantes imagens que foram mal identificadas.

Como um dos maiores problemas no treino de redes neurais é a falta de dados ou a má qualidade destes, decidiu-se resolver este problema utilizando uma abordagem não supervisionada, visto que permite ter um número infinito de imagens de treino através da aplicação de transformações às mesmas.

### 1.3 Contribuição do projeto para os objetivos de desenvolvimento sustentável

Este projeto está principalmente relacionado com a área da saúde, podendo contribuir para a melhoria dos cuidados de saúde prestados ao utente, levando a uma melhor organização e eficiência nos ambientes hospitalares.

Além disso, a redução na necessidade de repetir exames contribui significativamente para a economia de recursos hospitalares. Isso inclui uma menor utilização de equipamentos médicos, menos consumo de energia e uma redução no uso de materiais como filmes de raio-X e produtos químicos para o seu desenvolvimento. A longo prazo, esta melhoria pode levar a uma diminuição dos custos operacionais e a um aumento na sustentabilidade das operações hospitalares. Na Tabela 1.1 encontram-se as contribuições deste projeto para os ODS.

Tabela 1.1: Análise do projeto em relação aos ODS

ODS	Meta Específica	Contribuição do Projeto
ODS 3: Saúde de qualidade	Meta 3.4: Reduzir num terço a mortalidade prematura por doenças não transmissíveis via prevenção e tratamento, e promover a saúde mental e o bem-estar Meta 3.8: Atingir a cobertura universal de saúde, incluindo a proteção do risco financeiro, o acesso a serviços de saúde essenciais de qualidade e o acesso a medicamentos e vacinas essenciais para todos de forma segura, eficaz, de qualidade e a preços acessíveis.	Facilita o diagnóstico precoce de doenças crónicas respiratórias ao garantir a correta identificação dos pacientes.  Permite uma reidentificação de pacientes mesmo em países com piores condições hospitalares, mais sujeitos a problemas de gestão de dados.
ODS 9: Indústria, Inovação e Infraestrutura	Meta 9.5: Fortalecer a investigação científica, melhorar as capacidades tecnológicas de setores industriais em todos os países, particularmente os países em desenvolvimento.	Promove a inovação tecnológica em saúde com a aplicação de <i>deep learning</i> , o que pode ser expandido para outras áreas da indústria.

### 1.4 Estrutura do trabalho

Este relatório está dividido em 5 capítulos. Na presente secção, é feita uma introdução ao projeto, que compreende o enquadramento, a motivação para a sua realização, bem como uma visão geral da estrutura do documento. Para além disto, também é demonstrado como este projeto contribui para os Objetivos de Desenvolvimento Sustentável (ODS) estabelecidos pelas Nações Unidas. No capítulo 2, são apresentados alguns dos conceitos teóricos fundamentais para este trabalho, cuja compreensão é essencial para entender a metodologia e a análise dos resultados obtidos. Abordam-se temáticas como a aplicação de *deep learning* em imagem médica, aprendizagem supervisionada, não supervisionada e

semi-supervisionada, aprendizagem contrastiva e o estado da arte. No terceiro capítulo, é detalhada a metodologia utilizada para resolver o problema. No capítulo seguinte, que corresponde à discussão dos resultados, é realizada uma análise dos resultados obtidos, refletindo sobre quais métodos trouxeram melhores resultados. O capítulo 5 apresenta a conclusão, na qual se abordam as limitações da abordagem adotada e se discutem possíveis melhorias para a continuação do projeto.

## Capítulo 2

# Conhecimentos de base

### 2.1 *Deep learning* em imagem médica

Com o crescimento da inteligência artificial (IA) e o avanço dos métodos de *deep learning*, tem havido um aumento significativo no número de publicações científicas ao longo dos anos nesta área. A Figura 2.1, obtida com os dados disponíveis no website "Scopus", revela um aumento notável no número de artigos da área de *deep learning*, refletindo a tendência exponencial de crescimento desta área.

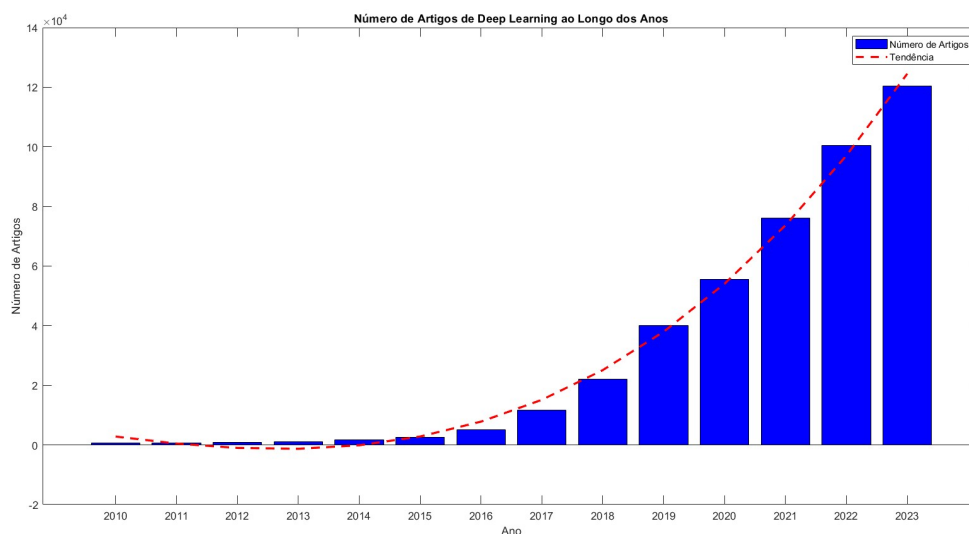


Figura 2.1: Número de artigos de *deep learning* ao longo dos anos.

A aplicação de *deep learning* estende-se por uma ampla gama de áreas, permitindo alcançar resultados que seriam difíceis de obter apenas com a intervenção humana. Este avanço tecnológico tem contribuído significativamente para a automatização de tarefas complexas e para a obtenção de resultados mais precisos e eficientes.

Na área da saúde, o impacto de *deep learning* é particularmente significativo, sendo utilizado em diversas aplicações biomédicas, desempenhando um papel crucial em tecnologias hospitalares modernas. O diagnóstico de doenças é uma das áreas em que o *deep learning* tem mostrado resultados promissores, levando ao desenvolvimento de algoritmos capazes de identificar padrões complexos em imagens médicas, como radiografias, ressonâncias magnéticas e tomografias computadorizadas, o que conduz a diagnósticos mais precisos e a uma maior velocidade de processamento que não seria possível manualmente. Para além disto, esta área tem sido aplicada na otimização do local de trabalho hospitalar, garantindo uma



melhor alocação de recursos e também no planeamento cirúrgico. (1; 2)

*Deep learning* recorre a redes neuronais profundas para modelar e resolver problemas complexos. Uma típica rede neuronal apresenta a arquitetura retratada na Figura 2.2. Esta é constituída por inúmeros elementos base denominados neurónios, os quais se encontram associados a certos parâmetros que vão transformando os dados que servem como *input* à rede neuronal, permitindo a extração de características relevantes para obter o *output* da rede (classificação dos dados, por exemplo). O ajuste dos parâmetros da rede neuronal é realizado através da minimização de uma função de perda (*loss*), que quantifica a diferença entre as previsões da rede e os valores reais. Este ajuste pode ser feito através de algoritmos que recorrem ao cálculo do gradiente. (3)

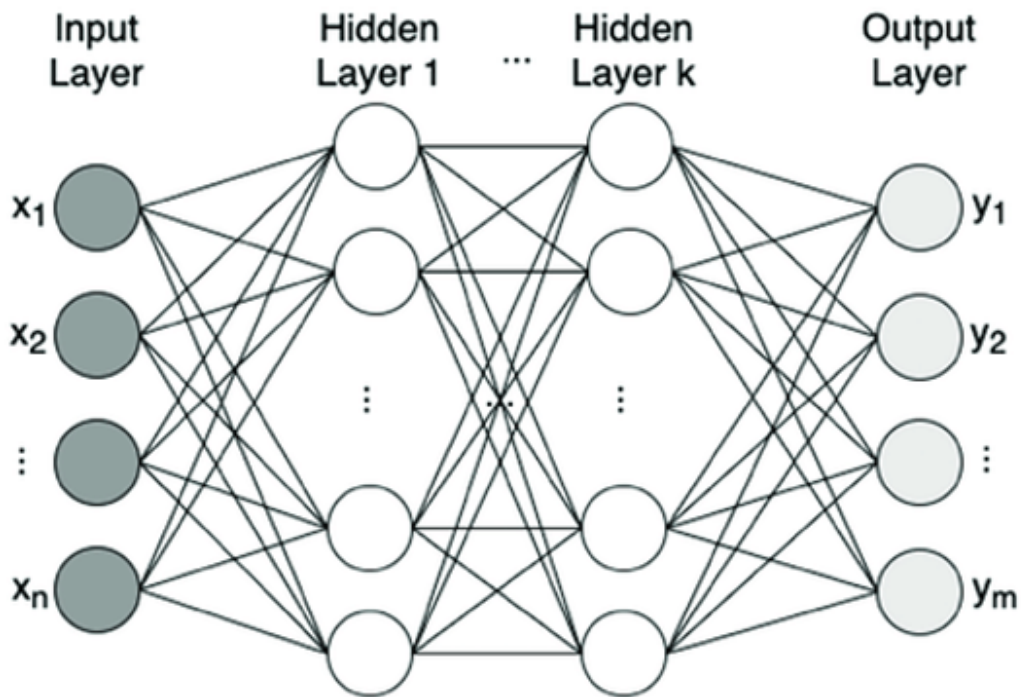


Figura 2.2: Arquitetura típica de uma rede neuronal (retirado de (4))

O processo de treino de uma rede neuronal envolve a passagem dos dados de treino pela rede várias vezes, cada uma dessas passagens é chamada de época. Durante cada época, o conjunto de dados pode ser dividido em pequenos subconjuntos chamados *batches*, usados para treinar a rede em cada iteração. O uso de *batches* permite atualizar os parâmetros da rede várias vezes dentro de uma única época, acelerando o processo de treino e tornando-o mais eficiente em termos de memória e tempo de computação. (3)

Após o treino, é importante avaliar o desempenho do modelo. Para isso, os dados são geralmente divididos em dois conjuntos: o conjunto de treino e o conjunto de teste. O modelo é treinado com o conjunto de treino, ajustando seus parâmetros para minimizar a função de perda. Depois, o modelo é avaliado com o conjunto de teste para verificar sua capacidade de generalização, através da utilização de dados nunca vistos pelo modelo. (5)

As redes neuronais convolucionais (CNNs) são uma arquitetura específica de redes neuronais projetadas para processar dados como imagens. Elas são compostas por camadas convolucionais que aplicam filtros sobre as entradas para extrair características espaciais e hierárquicas dos dados. As CNNs são particularmente eficazes em tarefas de reconhecimento de padrões, devido à sua capacidade de capturar dependências locais e reduzir a dimensionalidade dos dados de entrada. (6)

### 2.1.1 Aprendizagem supervisionada, não supervisionada e semi-supervisionada

Com o decorrer do tempo e surgimento de novos problemas, foram desenvolvidos diferentes algoritmos de *deep learning* que recorrem a diferentes abordagens para treinar as redes neurais, dependendo da disponibilidade de dados e da finalidade do modelo. Assim, é importante fazer a distinção entre aprendizagem supervisionada, não supervisionada e semi-supervisionada.

A aprendizagem supervisionada é a técnica mais utilizada para resolução de problemas de classificação, levando, por norma, aos melhores resultados. Neste tipo de aprendizagem, o modelo é treinado com um conjunto de dados rotulados, onde cada entrada está associada a uma saída desejada. Como o modelo possui exemplos claros do que se pretende obter com os diferentes *inputs*, consegue realizar melhores previsões quando confrontado com novos dados.

Já a aprendizagem não supervisionada, apesar de todo o processo da arquitetura da rede e diminuição da *loss* ser semelhante à aprendizagem supervisionada, apresenta diferenças consideráveis, no que toca ao conjunto de dados que alimentam o treino do modelo. De forma a aplicar uma função de *loss*, os rótulos são gerados indiretamente a partir do conjunto de dados, o que pode ser realizado, por exemplo, através de *clustering* ou da exploração de relações intrínsecas dos dados, como a co-ocorrência de diferentes partes dos dados. Por exemplo, em processamento de linguagem natural, um modelo pode ser treinado para prever palavras ausentes em uma frase com base no contexto fornecido pelas palavras próximas. Essas previsões servem como rótulos implícitos para o treino do modelo. Também pode ser usada uma estratégia que envolve a aplicação de transformações que será abordada posteriormente. (7; 8)

Assim, a principal diferença entre estas duas técnicas está relacionada com o facto da existência de rótulos nos dados (por exemplo, imagens de gato com rótulo "animal") que são essenciais para o treino do modelo. A aprendizagem não supervisionada permite a resolução dos mesmos problemas da aprendizagem supervisionada, lidando com o típico problema da má qualidade e insuficiência dos dados existentes.

Por vezes, quando apenas existe uma pequena quantidade de rótulos num certo conjunto de dados, de forma a obter os melhores resultados possíveis, pode-se combinar estas duas técnicas de aprendizagem, realizando-se uma aprendizagem semi-supervisionada. O processo geralmente envolve o uso inicial de dados rotulados para treinar um modelo, que é então refinado e melhorado com o uso de dados não rotulados. (9)

### 2.1.2 Aprendizagem contrastiva

A aprendizagem contrastiva é uma técnica aplicada no treino de redes neurais, que consiste na aprendizagem de representações que permite a comparação e distinção de dados.

Na aprendizagem supervisionada, amostras que fazem parte da mesma classe são forçadas a aproximarem-se no espaço das características, sendo repelidas relativamente a outras. Desta forma, o processo de separação das amostras pelas diversas classes é facilitado, fazendo com que haja um agrupamento das imagens que apresentam semelhanças dependendo das características que possuem, tal como está representado na Figura 2.3 a). (10)

Já na Figura 2.3 b), está presente um esquema que explica a aprendizagem contrastiva não supervisionada. A partir de uma imagem original, através de aplicação de transformações às imagens, obtêm-se várias imagens que derivaram da original, sendo rotuladas como positivas. As restantes imagens são rotuladas como negativas. Assim, a distância entre pares positivos é minimizada e a distância entre pares negativos é maximizada. (10)



Figura 2.3: a) Aprendizagem contrastiva supervisionada b) Aprendizagem contrastiva não supervisionada

## 2.2 Estado da arte

Relativamente ao estado da arte, já existem inúmeros artigos que incidem sobre a aprendizagem contrastiva, sendo que este projeto foi inspirado principalmente em dois artigos publicados: Chen et al. "A Simple Framework for Contrastive Learning of Visual Representations"(11) e Packhäuser et al. "Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest X-ray data"(12).

O trabalho realizado por Chen et al.(11) tornou-se um dos trabalhos mais relevantes para a aprendizagem contrastiva, servindo como ponto de partida para muitos outros projetos. Neste artigo, eles estudaram a abordagem não supervisionada recorrendo ao método de aprendizagem contrastiva já previamente explicado, analisando a importância das transformações aplicadas às imagens e comparando com os resultados obtidos usando aprendizagem supervisionada em diferentes datasets recorrentemente usados, como o Cifar10, Flowers, entre outros. Tal como está representado na Figura 2.4, parte-se de uma imagem original à qual são aplicadas duas transformações que são posteriormente dadas como *input* à rede neuronal. Procede-se ao cálculo da *contrastive loss* depois de passar a saída da rede por uma cabeça de projeção.

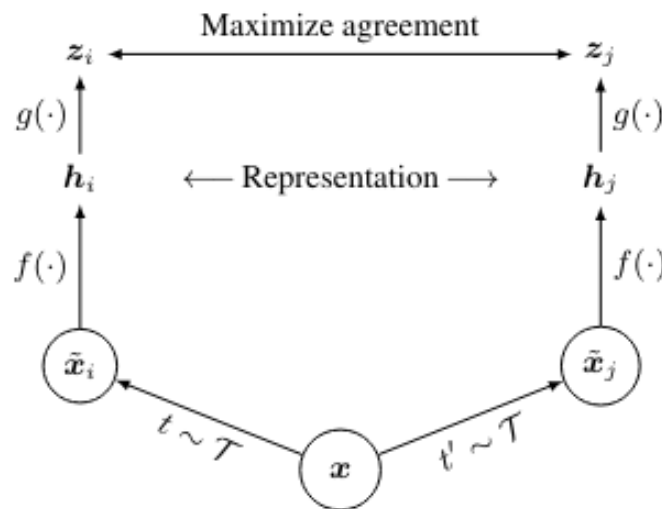


Figura 2.4: Framework da aprendizagem contrastiva (retirada de (11))

Pelo outro lado, Packhäuser et al.(12) comprovaram que a reidentificação de pacientes em imagens de raio-X previamente anonimizadas é uma tarefa possível de alcançar usando redes neurais. Neste estudo eles realizaram uma aprendizagem supervisionada, recorrendo ao treino de uma rede neuronal siamesa que recebe dois objetos de entrada, permitindo a sua comparação. A rede foi treinada utilizando o *dataset ChestX-ray14*(13) e o seu desempenho foi avaliado em diferentes *datasets*. A metodologia que foi realizada neste artigo está representada na Figura 2.5. Obtiveram-se valores de AUC (0.9940) e acurácia (95.55%) bastante elevados no *dataset ChestX-ray14*, sendo importante mencionar que estes valores foram obtidos utilizando um subconjunto artificial balanceado, isto é, com o mesmo número de pares positivos e negativos.

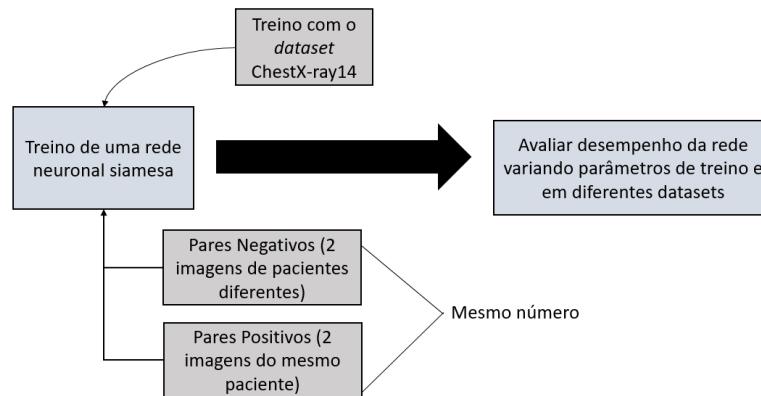


Figura 2.5: Fluxograma da metodologia base do artigo

Este projeto visa integrar os conceitos abordados em ambos os artigos mencionados, com o objetivo de facilitar a identificação de imagens de raio-X do *dataset ChestX-ray14*(13) utilizando uma abordagem de aprendizagem não supervisionada. Pretende-se analisar o impacto das transformações aplicadas às imagens, combinando técnicas de aprendizagem contrastivas. Dessa forma, o projeto busca explorar novas possibilidades na identificação de imagens médicas, potencialmente aprimorando a precisão e a eficiência desse processo. Futuramente, considerando que existem conjuntos de dados de imagens sem identificação do paciente associada, planeja-se treinar e testar nesses conjuntos.

## Capítulo 3

# Metodologia

### 3.1 Descrição do dataset

Recorreu-se à utilização do dataset público *ChestX-ray14* (13) que compreende um conjunto de imagens médicas de raio-X em vista frontal, contendo 112.120 imagens de pacientes únicos, coletadas entre 1992 e 2015, os quais possuem diversas doenças como, por exemplo, edema, enfisema, fibrose, espessamento pleural, entre outras. Através do histograma representado na Figura 3.1, consegue-se perceber que cada paciente possui diversas imagens de raio-X associadas.

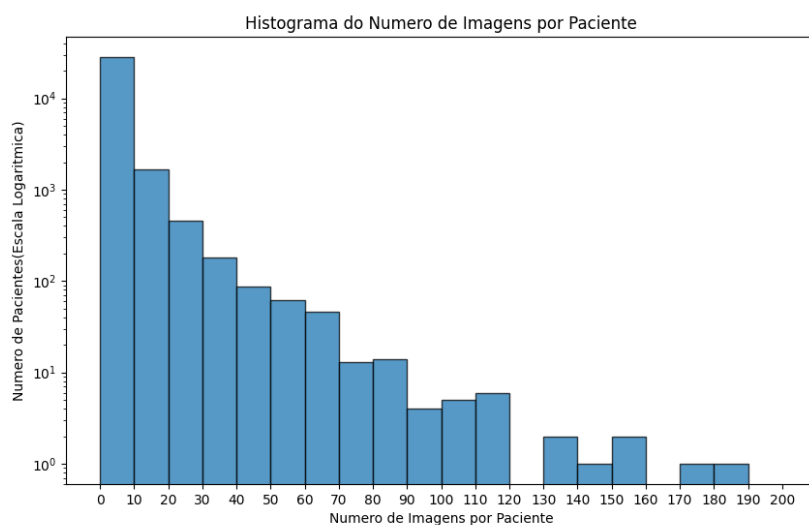


Figura 3.1: Histograma do número de imagens por paciente

Fazendo a análise detalhada do conjunto de dados, verifica-se que existem imagens de 30.805 pacientes diferentes e, em média, cada um tem  $3,64 \pm 7,27$  imagens associadas. Para além disto, verifica-se que a maioria dos pacientes possui uma única imagem (mediana) e o número máximo de imagens associadas a um paciente é 184. O *dataset* foi dividido em conjunto de treino (cerca de 90 %) utilizado durante a etapa de treino do modelo e em conjunto de teste (cerca de 10 %) que permitirá a avaliação do desempenho do modelo. Teve-se especial cuidado em garantir que todas as imagens do mesmo paciente se encontravam no mesmo conjunto, uma vez que isso poderia levar à existência de viés do modelo.

## 3.2 Reidentificação de pacientes

### 3.2.1 Arquitetura

Para o treino da rede neuronal, utilizou-se a arquitetura "ResNet-18". Esta arquitetura é uma versão específica das redes neurais convolucionais residuais, que permite um fluxo contínuo de informações entre diferentes camadas da rede sem a perda de informações. A ResNet-18 é composta por 18 camadas com parâmetros, organizadas em blocos residuais. Esses blocos utilizam conexões de atalho de identidade, onde a entrada original do bloco é adicionada à saída da camada convolucional, garantindo que as informações essenciais sejam preservadas ao longo da rede. Esta estrutura ajuda a mitigar o problema da redução do gradiente, comum em redes profundas, permitindo que a rede seja treinada de maneira mais eficiente. (14)

A arquitetura ResNet-18 inclui várias camadas distintas, conforme representado na Figura 3.2. As camadas convolucionais estão alternadas com camadas que levam à redução da dimensionalidade até chegar ao bloco totalmente conectado (Fc). Neste trabalho, o bloco totalmente conectado foi modificado para incluir uma cabeça de projeção MLP, que consiste em uma camada linear, seguida por uma ReLU, e outra camada linear que leva à saída da rede.

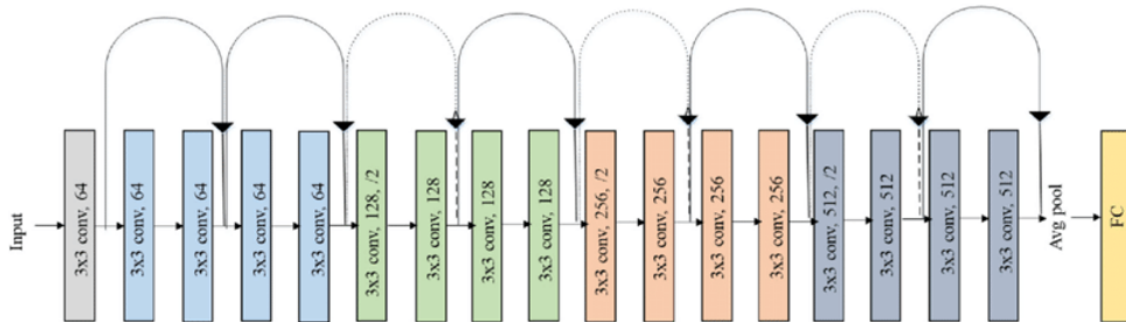


Figura 3.2: Arquitetura ResNet-18 (adaptada de (15))

### 3.2.2 Treino

Na aprendizagem contrastiva não supervisionada, tal como explicado no capítulo 2.1.2, como os rótulos são gerados indiretamente do conjunto de imagens, há necessidade de formar pares positivos e pares negativos (não se utilizaram os rótulos que já vinham associados às imagens, isto é, o ID dos pacientes). Os pares positivos são obtidos através da aplicação de transformações, tal como se encontra ilustrado na Figura 3.3 b), as quais serão mencionadas detalhadamente no capítulo 3.2.3. Após determinar qual conjunto de transformações levava aos melhores resultados, realizaram-se alguns treinos aplicando essas transformações, variando o número de épocas e o *batch size*.

Numa fase inicial, como se pretendia analisar apenas o efeito da aplicação de transformações ao conjunto de treino, todos os treinos foram realizados com os parâmetros que se encontram especificados na Tabela 3.1.

Assim, de forma a ser possível fazer a comparação das duas abordagens, optou-se por também aplicar as transformações às imagens do conjunto de treino e utilizaram-se os parâmetros da Tabela 3.1. Tal como explicado no capítulo 2.1.1, para a aprendizagem supervisionada já se utilizam os rótulos das imagens para criar pares positivos e negativos. No entanto, contrariamente ao que se faz na aprendizagem não supervisionada, já não se utilizam duas visualizações da mesma imagem para criar pares positivos. Como as imagens encontram-se associadas a um rótulo, um par positivo corresponde a duas imagens

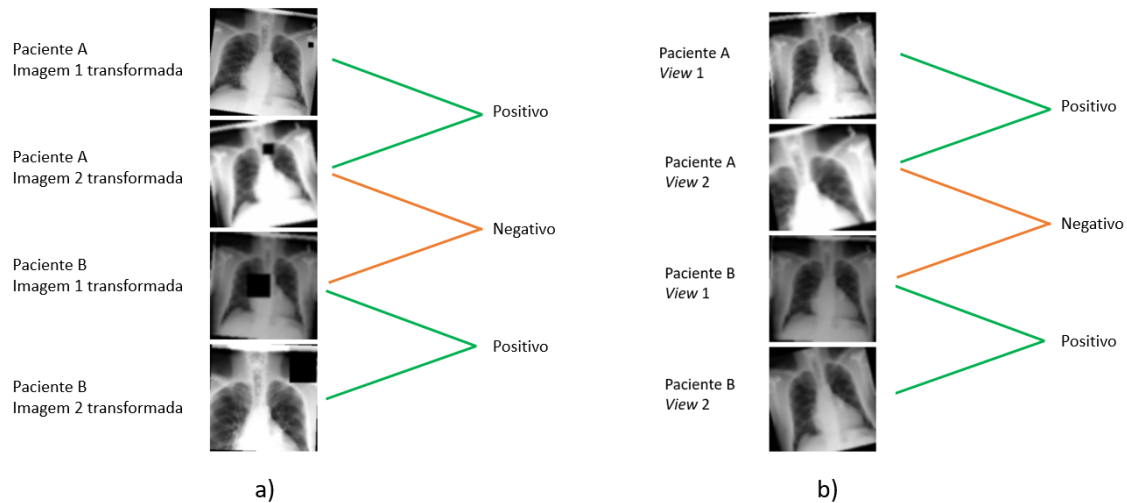


Figura 3.3: a) Aprendizagem contrastiva supervisionada b) Aprendizagem contrastiva não supervisionada

Tabela 3.1: Parâmetros da Rede Neural

Parâmetro	Descrição
Arquitetura	Resnet18
Épocas	100
Tamanho do <i>batch</i>	256
Taxa de Aprendizagem Inicial	0,0003
Número dos <i>features</i>	128
Número de visualizações	2

transformadas do mesmo paciente, tal como está representado na Figura 3.3 a). Como se verificou pela análise do *dataset*, a maioria dos pacientes apenas possui uma única imagem, o que condicionaria esta abordagem. No entanto, nesses casos, como as imagens estão a ser transformadas antes de formar pares positivos e negativos, concluiu-se que não afetaria os resultados, uma vez que as imagens do par positivo seriam distintas devido às transformações. Esta parte do projeto foi realizada após obter os resultados provenientes da abordagem não supervisionada e, como tal, apenas se aplicou o conjunto de transformações do treino que levou aos melhores resultados no conjunto de teste.

Para além disto, também se realizou um treino no qual não foram aplicadas quaisquer transformações nas imagens, de forma a avaliar se as transformações levavam a melhores resultados, devido ao aumento da variância introduzida no conjunto de treino.

O código utilizado foi otimizado de forma a possibilitar o uso da GPU, o que permitiu que os treinos decorressem mais rapidamente, possibilitando a realização de processamento paralelo e a utilização de conjuntos de imagens com dimensões maiores (128x128).

### 3.2.3 Transformações

O treino 0 foi realizado na aprendizagem supervisionada, na qual num dos treinos não se aplicaram transformações, de forma a avaliar se ocorria uma melhoria de resultados com a aplicação das transformações. De forma a avaliar o impacto dos resultados obtidos com a aplicação de diferentes transformações, foram realizados 6 treinos, diferenciando o conjunto e a intensidade das transformações aplicadas ao conjunto de treino. Tendo em conta que os dados correspondem a imagens de raio-X e, como tal, as transformações aplicadas às imagens para gerar as duas visualizações devem fazer sentido, analisou-se



empiricamente as imagens que correspondiam ao mesmo paciente e concluiu-se que se deveriam aplicar as seguintes transformações:

- *Random Resized Crop*: Recorta aleatoriamente uma porção da imagem e a redimensiona para um tamanho específico, permitindo identificar características em diferentes escalas e posições.
- *Gaussian Blur*: Aplica um desfoque gaussiano à imagem, suavizando os detalhes o que pode ajudar a reduzir o ruído.
- *CutOut*: Remove aleatoriamente uma porção da imagem, substituindo-a por um quadrado preto, forçando a rede a aprender com regiões incompletas da imagem.
- *Random Rotation*: Aplica uma rotação aleatória à imagem até um máximo de 15°, garantindo robustez da rede quando sujeita a diferentes orientações das imagens.
- *Random Brightness*: Altera aleatoriamente o brilho da imagem.

As imagens de raio-X podem ser tiradas com pequenas variações nos ângulos, diferenças na luz incidente e no *zoom* da imagem o que explica o uso destas transformações que pretendem simular estas diferenças. Inicialmente, optou-se por aplicar um número reduzido de transformações de baixa intensidade que não deformavam muito as imagens. Analisando os resultados obtidos, novos conjuntos de transformações foram aplicados (Tabela 3.2), visando a melhoria dos resultados.

Tabela 3.2: Transformações aplicadas

Treino	Transformações
0	Sem transformações
1	RandomResizedCrop
	RandomBrightness
2	RandomResizedCrop
	RandomBrightness
	Rotation - 0 a 15°
3	RandomResizedCrop
	RandomBrightness
	Rotation - 0 a 15°
	CutOut - 0 a 1/4
4	RandomResizedCrop
	RandomBrightness
	Rotation - 0 a 15°
	CutOut - 0 a 1/3
5	RandomResizedCrop
	RandomBrightness
	Rotation - 0 a 15°
	CutOut - 0 a 1/3
	Gaussian Blur
6	RandomResizedCrop
	RandomBrightness
	Rotation - 0 a 15°
	Gaussian Blur

Como neste trabalho, as transformações aplicadas às imagens apresentam um papel essencial na capacidade da rede neuronal de determinar se as mesmas imagens pertencem ao mesmo paciente, foram aplicados diferentes conjuntos de transformações, de forma a avaliar qual destes conjuntos permitiria



obter melhor resultados no conjunto de teste. A Figura 3.4 representa as transformações que foram feitas às imagens. É importante realçar, que durante o treino, várias transformações foram aplicadas simultaneamente com parâmetros selecionados de forma aleatória. Por exemplo, para a intensidade do brilho era escolhido aleatoriamente um valor dentro de um intervalo.

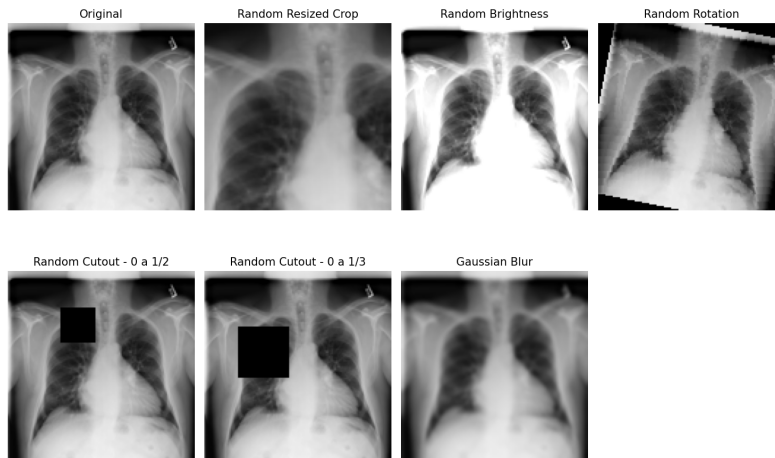


Figura 3.4: Transformações aplicadas à imagem

### 3.3 Métricas de avaliação

De forma a avaliar se as redes neuronais previamente treinadas com diferentes transformações permitiam a resolução do problema inicial, isto é, perceber se duas imagens desconhecidas correspondiam ao mesmo paciente, calcularam-se métricas que permitem avaliar o desempenho do modelo no conjunto de teste. Através da passagem de cada imagem pela rede, obtém-se *output* que corresponde a uma matriz. Para comparar as diversas imagens, pode-se fazer a multiplicação das matrizes resultantes e proceder à sua normalização (garantir que o resultado está entre 0 e 1), obtendo um valor de semelhança (distância) entre duas imagens. Assim, através destes valores, é possível perceber se duas imagens correspondem ao mesmo paciente ou não.

Foram traçadas as curvas ROC (*Receiver Operating Characteristic*) e PR (*Precision-Recall*), o que possibilitou o cálculo dos valores das áreas debaixo da curva (AUC, *Area Under the Curve*), as quais se encontram exemplificadas na Figura 3.5. Para os cálculos de ROCs/PRCs é utilizada a distância como métrica de separabilidade e cada imagem é considerada como positiva/negativa consoante pertença ou não ao mesmo paciente da imagem à qual se está a comparar.

A curva ROC permite avaliar a capacidade do modelo em distinguir entre classes positivas e negativas. O FPR (Taxa de Falsos Positivos) mede a fração de exemplos negativos que são incorretamente classificados como positivos. O TPR (Taxa de Verdadeiros Positivos) mede a fração de exemplos positivos que são corretamente identificados. O eixo do x representa FPR e no eixo do y está a TPR e através da alteração do limiar decisão (valor que separa as classes positivas das negativas), as taxas vão sofrendo alterações. Idealmente, pretende-se que TPR seja 1 e FPR seja 0, o que indicaria que era possível definir um limiar que distingue perfeitamente as classes positivas e negativas, resultando numa AUC de 1. (16)

A curva PR é particularmente útil quando se lida com dados desbalanceados. No eixo do x encontra-se o *recall* e no eixo do y está representada a *precision*. A *precision* é a proporção de verdadeiros positivos entre os exemplos que o modelo classificou como positivos. O *recall* é o mesmo que TPR. Se existir um certo limiar que permita obter valores de *precision* e *recall* iguais a 1 obtém-se uma AUC de 1 que indica que todos os pares positivos foram identificados, não existindo falsos positivos nem falsos negativos. (16)

Para além disto, também foi calculado valores de *precision top* e *recall top*, cujo cálculo está ilustrado

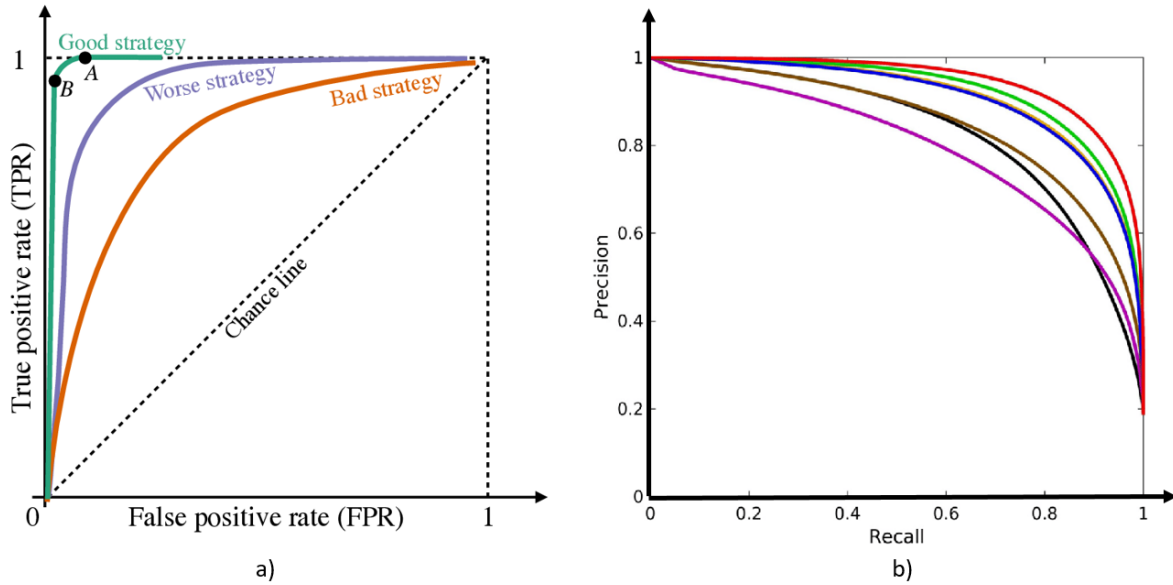


Figura 3.5: a) Típica forma da curva ROC (retirada de (17)) b) Típica forma da curva PR (adaptada de (18))

no exemplo da Figura 3.6. Assumindo que um certo paciente possui 4 imagens, pode-se calcular a semelhança dos valores de saída do modelo de uma imagem com as restantes, obtendo-se um valor que pode ser visto como a probabilidade de as imagens pertencerem ao mesmo paciente. Assim, ordenando estes valores de semelhança, a *precision top-X* representa o número de valores de semelhança de pares positivos que estão nos X primeiros elementos do *array*. Optou-se também por calcular a *precision top-K*, em que K representa o número de pares positivos existentes para aquele paciente. Assim, seria possível avaliar o modelo de forma igual tanto para pacientes com inúmeras imagens como para aqueles que possuem poucas imagens. Já o *recall top-X* representa a percentagem de valores de semelhança de pares positivos que se encontram nos primeiros X elementos. Estes cálculos foram feitos para cada imagem de cada paciente que possuísse mais que uma imagem (só assim é possível calcular valores de semelhança), apresentando nos resultados a média de todos os valores calculados para cada imagem.

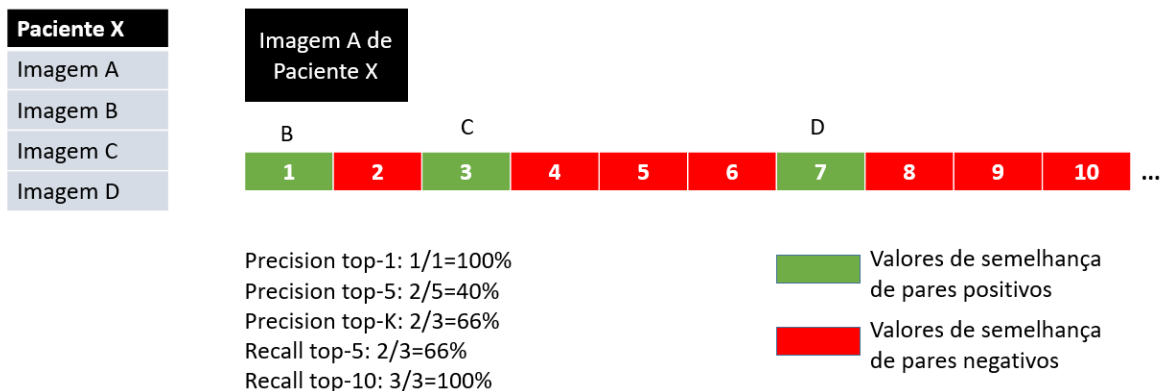


Figura 3.6: Exemplo do cálculo de métricas

## Capítulo 4

# Resultados e Discussão

### 4.1 Aprendizagem supervisionada e não supervisionada

Tal como seria esperado, devido ao grande número de imagens do conjunto de treino, com o decorrer das épocas, há uma convergência do valor da *loss* para valores próximos de zero, tal como se pode verificar pela Figuras 4.1 e 4.2. Com o aumento do número de transformações aplicadas às imagens, o modelo apresenta maior dificuldade a detetar características que permitem identificar corretamente os pares positivos, uma vez que as imagens se encontram mais deformadas, levando a maiores valores de *loss* (treino 6).

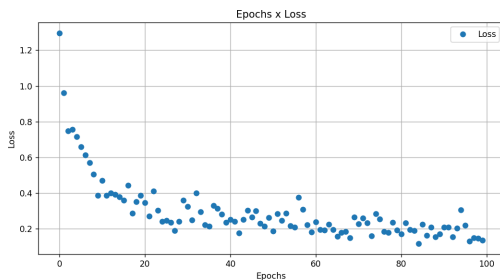


Figura 4.1: Evolução da *loss* durante o treino 1

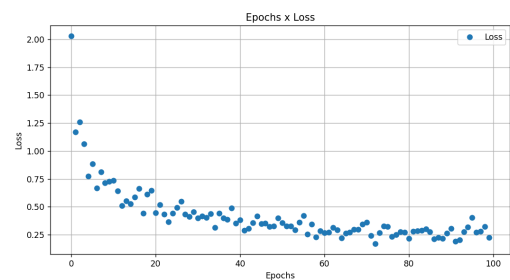


Figura 4.2: Evolução da *loss* durante o treino 6

Os resultados apresentados na Tabela 4.1 mostram os valores das diferentes métricas de avaliação do desempenho do modelo, previamente explicadas na secção 3.3. Assim, foram calculadas os valores de *recall* top 1000, *precision* top1 e topk e os valores de AUC ROC e AUC PR para os diferentes conjuntos de treino usando aprendizagem não supervisionada.

Tabela 4.1: Avaliação da aprendizagem não supervisionada (*Pr-Precision*, *Rc-Recall*). Melhores resultados de cada métrica destacados a negrito.

Treino	Rc Top1000	Pr Top1	Pr TopK	AUC ROC	AUC PR
1	0,760	0,641	0,297	0,908	0,291
2	0,736	0,641	0,289	0,902	0,281
3	0,777	0,650	0,297	0,916	0,294
4	0,778	0,649	0,296	0,916	0,293
5	0,781	0,660	0,308	0,920	0,305
6	<b>0,785</b>	<b>0,667</b>	<b>0,312</b>	<b>0,921</b>	<b>0,311</b>

Através da análise da Tabela 4.1, verifica-se que o aumento do número de transformações aplicadas levou, a nível geral, a melhores resultados, o que pode ser explicado pelo facto de o modelo ser obrigado

a aprender *features* das imagens mais específicos de forma a identificar os pares positivos, uma vez que as visualizações são mais diferentes. No entanto, verificou-se que ao retirar a transformação *CutOut*, obteve-se melhores resultados. Assim, o melhor conjunto de transformações foi o 6: *RandomResizedCrop*, *RandomBrightness*, *Rotation* - 0 a 15° e *Gaussian Blur*.

Os altos valores da AUC ROC podem ser explicados pelo facto de o problema ser desbalanceado e, como tal, o número de exemplos positivos é bastante inferior ao número de exemplos negativos. Na Figura 4.3, para um paciente que possui apenas 3 imagens, ao calcular as métricas para uma dada imagem desse paciente, apenas existem dois pares positivos e milhares de pares negativos. Como tal, é natural que existam vários limiares que permitem a obtenção uma baixa taxa de falsos positivos. Neste caso em específico, a partir de um certo limiar, ambos os dois pares se encontram separados dos pares negativos, o que leva também à existência de uma maior taxa de falsos positivos. Assim, um valor alto de AUC da curva ROC pode levar a falsas conclusões, sendo importante analisar a curva PR.

Na Figura 4.4 encontra-se representada a curva PR. Como a *precision* é a proporção de verdadeiros positivos entre os exemplos que o modelo classificou como positivos, verifica-se que um dos valores de semelhança do par positivo se encontra bastante afastado dos restantes valores. No entanto, a partir de um certo limiar, há uma queda drástica do valor de *precision*, o que indica a existência de inúmeros falsos positivos. Assim, a AUC PR revela-se uma boa métrica para avaliar o desempenho global do modelo na tarefa de determinar os pares positivos.

Tendo em conta que se realizou um treino não supervisionado, os resultados obtidos foram bastante satisfatórios. Comparando uma dada imagem de um certo paciente com todas as outras imagens existentes no conjunto de teste, a imagem mais semelhante a essa correspondeu a uma outra imagem do mesmo paciente 66,7 % das vezes, o que é excelente considerando que o conjunto de teste possui 10.032 imagens.

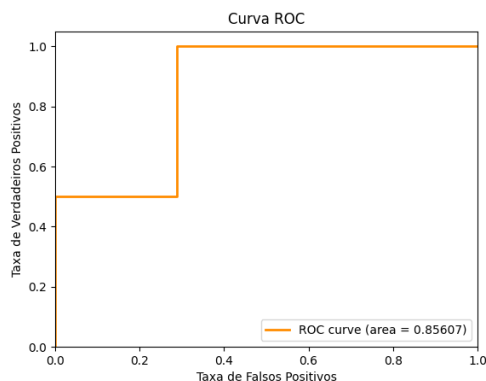


Figura 4.3: Curva ROC para um certo paciente do treino 6

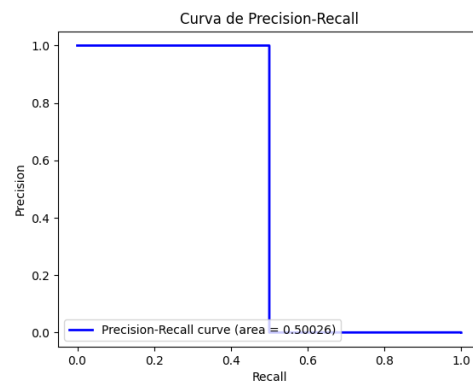


Figura 4.4: Curva PR para um certo paciente do treino 6

Também se fez uma breve análise sobre o impacto da alteração de parâmetros de treino como o número de épocas e *batch size* nos resultados obtidos, os quais se encontram na Tabela 4.2. Como o treino 6 foi aquele que permitiu a obtenção de melhores resultados, as variações foram feitas aplicando as transformações desse treino.

Tabela 4.2: Avaliação da aprendizagem não supervisionada com diferentes parâmetros. Melhores resultados de cada métrica destacados a negrito.

Treino	Rc Top1000	Pr Top1	Pr TopK	AUC ROC	AUC PR
6	0,785	0,667	0,312	0,921	0,311
6 com 300 épocas	<b>0,822</b>	<b>0,711</b>	<b>0,342</b>	<b>0.93351</b>	<b>0.34448</b>
6 com batch size de 512	0.772	0.668	0.308	0.91499	0.30524
6 com batch size de 128	0.766	0.643	0.293	0.91293	0.28767

Assim, tal como seria de esperar, através do aumento do número de épocas, o modelo é capaz de ajustar melhor os parâmetros internos da rede, de forma a levar a uma maior diminuição da *loss*, tal como está representado na Figura 4.5. Como os resultados obtidos foram melhores quando comparados com o treino de 100 épocas, pode-se concluir que o treino ainda não tinha estagnado com as 100 épocas, concluindo que o modelo foi capaz de determinar melhores características que permitem a identificação de pares positivos. Pelo outro lado, a diminuição e o aumento do *batch size* levaram a piores resultados.

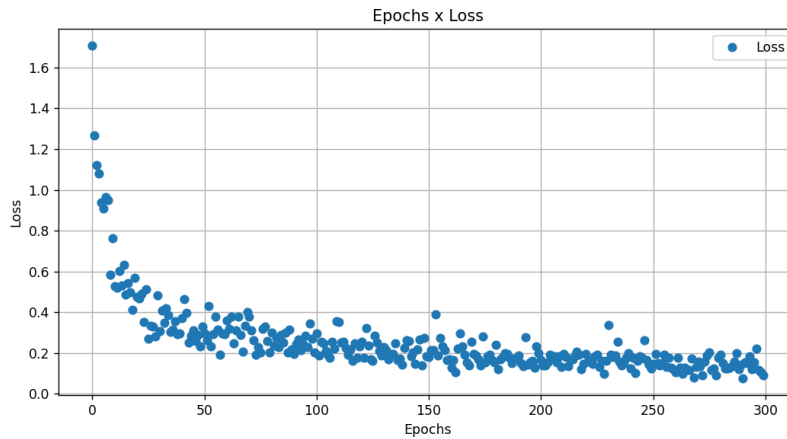


Figura 4.5: Evolução da *loss* durante o treino com 300 épocas

Na Tabela 4.3 estão representados os valores das métricas de avaliação do modelo recorrendo a um treino supervisionado. Tal como explicado na metodologia, foram realizados dois treinos, sendo que num foram aplicadas as transformações do treino 6 da abordagem não supervisionada ao conjunto de treino e noutro não foram aplicadas quaisquer transformações.

Tabela 4.3: Avaliação da aprendizagem supervisionada. Melhores resultados de cada métrica destacados a negrito.

Treino	Rc Top1000	Pr Top1	Pr TopK	AUC ROC	AUC PR
0	0,984	0,928	0,735	0.99315	0.77438
6	<b>0,987</b>	<b>0,935</b>	<b>0,739</b>	<b>0.99444</b>	<b>0.78113</b>

Através da análise da Tabela 4.3, conclui-se que a aplicação de transformações levou a um melhor desempenho do modelo, o que está relacionado com a existência de maior variabilidade entre imagens dos pares, obrigando o modelo a procurar características mais específicas. Os resultados obtidos foram bastante elevados, verificando-se que para o treino com transformações, 93,5% das vezes, a imagem mais parecida à imagem que se pretende comparar, pertence ao mesmo paciente. Todas as outras métricas apresentaram valores bastante elevados, dando especial ênfase à AUC da curva PR, devido à sua importância em problemas desbalanceados. Os resultados obtidos por Packhäuser et al.(12) foram superiores a estes, porém os dois projetos não podem ser comparados diretamente, uma vez que estes usaram um subconjunto artificial balanceado, o que não aconteceu neste projeto.

Os resultados obtidos com a abordagem supervisionada foram melhores quando comparados com os que foram obtidos com a abordagem não supervisionada, obtendo-se valores de *precision*, *recall* e AUC bastante superiores, o que indica um melhor desempenho do modelo na resolução do problema que foi apresentado. No entanto, é importante ter em atenção que perante uma situação de existência de pouca quantidade e qualidade de dados, nos quais estes podem estar indevidamente rotulados, a abordagem não supervisionada permite a resolução deste problema, permitindo a reidentificação de pacientes de imagens de raio-X, o que seria impossível recorrendo a uma abordagem supervisionada.

## Capítulo 5

# Conclusão

Este projeto explorou e comparou as abordagens de aprendizagem supervisionada e não supervisionada para a reidentificação de pacientes a partir de imagens de raio-X, o que se demonstra como um problema pertinente, devido aos casos recorrentes de incorreta colocação de rótulos identificadores em imagens em ambientes hospitalares.

A abordagem supervisionada, que necessita de uma grande quantidade de dados rotulados de qualidade elevada, demonstrou um desempenho superior neste estudo específico, com métricas de *precision*, *recall* e AUC significativamente melhores. No entanto, a dependência de dados rotulados é uma limitação considerável, uma vez que a rotulagem manual de grandes quantidades de dados é um processo demorado e que pode levar à ocorrência de erros.

Por outro lado, a abordagem não supervisionada, que não necessita de dados rotulados, mostra-se particularmente útil em situações onde a obtenção de rótulos precisos é difícil ou inviável. Utilizando dados não rotulados, essa abordagem permite a aprendizagem de representações discriminativas das imagens, permitindo identificar imagens semelhantes correspondentes ao mesmo paciente embora com resultados ligeiramente inferiores em comparação com a abordagem supervisionada.

Os resultados obtidos com a abordagem não supervisionada poderão ser melhorados em projetos futuros, através de um estudo mais detalhado das transformações que são mais adequadas para aplicar nas imagens do treino. Para além disto, como se verificou na discussão de resultados, um aumento do número de épocas poderá ainda levar a uma melhoria do modelo. Pretende-se, no futuro, testar ambas as abordagens em outros *datasets*, uma vez que uma abordagem não supervisionada poderá levar a melhores resultados em *datasets* diferentes do que o do treino relativamente a uma abordagem supervisionada.

A escolha entre as duas abordagens deve levar em consideração as características específicas do problema, a disponibilidade de dados rotulados e os recursos disponíveis para a rotulagem e treino do modelo. Em muitos casos, a combinação de ambas as abordagens pode proporcionar um equilíbrio entre precisão e flexibilidade, maximizando o potencial de reidentificação de pacientes.

No contexto deste projeto, os resultados indicam que, apesar das limitações dos dados, é possível alcançar um bom desempenho na tarefa de reidentificação de pacientes usando uma técnica não supervisionada. Essa flexibilidade metodológica é essencial para aplicações em cenários reais, onde os dados podem ser escassos ou de baixa qualidade.

# Referências

- [1] H.P. Chan, R.K. Samala, L.M. Hadjiiski, and C. Zhou. Deep learning in medical image analysis. *Advances in Experimental Medicine and Biology*, 1213:3–21, 2020. doi: 10.1007/978-3-030-33128-3\_1.
- [2] Mohammed Yasser Ouis and Moulay A. Akhloufi. Deep learning for report generation on chest x-ray images. *Computerized Medical Imaging and Graphics*, 111:102320, 2024. doi:10.1016/j.compmedimag.2023.102320.
- [3] Vladimir Nasteski. An overview of the supervised machine learning methods. *HORIZONS.B*, 4:51–62, 12 2017. doi:10.20544/HORIZONS.B.04.1.17.P05.
- [4] Srinivasan Selvaraj. *Industry 4.0 Interoperability, Analytics, Security, and Case Studies*, page 7. October 2021.
- [5] H Wang, ZeZheZBePJ Lei, X Zhang, B Zhou, and J Peng. Machine learning basics. *Deep learning*, pages 98–164, 2016.
- [6] Keiron O’Shea and Ryan Nash. An introduction to convolutional neural networks. *ArXiv e-prints*, 2015. doi:10.48550/arXiv.1511.08458.
- [7] Dave Bergmann. What is self-supervised learning?, 2023. Acessado em: 8 de junho de 2024.
- [8] Xiao Liu, Fanjin Zhang, Zhenyu Hou, Li Mian, Zhaoyu Wang, Jing Zhang, and Jie Tang. Self-supervised learning: Generative or contrastive. *IEEE Transactions on Knowledge and Data Engineering*, 35(1):857–876, 2023. doi:10.48550/arXiv.2006.08218.
- [9] Padmanabha Y C a, Viswanath Pulabaigari, and Eswara B. Semi-supervised learning: a brief review. *International Journal of Engineering Technology*, 7:81, 02 2018. doi:10.14419/ijet.v7i1.8.9977.
- [10] Ashish Jaiswal, Ashwin Ramesh Babu, Mohammad Zaki Zadeh, Debapriya Banerjee, and Fil-lia Makedon. A survey on contrastive self-supervised learning. *Technologies*, 9(1), 2021. doi:10.3390/technologies9010002.
- [11] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations, 2020. doi: 10.48550/arXiv.2002.05709.
- [12] K. Packhäuser, S. Gundel, N. Münster, et al. Deep learning-based patient re-identification is able to exploit the biometric nature of medical chest x-ray data. *Scientific Reports*, 12:14851, 2022. doi:10.1038/s41598-022-19045-3.
- [13] X. Wang, Y. Peng, L. Lu, Z. Lu, M. Bagheri, and R. M. Summers. Chestx-ray8: Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3462–3471, July 2017.
- [14] Sasha Targ, Diogo Almeida, and Kevin Lyman. Resnet in resnet: Generalizing residual architectures. 2016. doi:10.48550/arXiv.1603.08029.
- [15] Farheen Ramzan, Muhammad Usman Khan, Asim Rehmat, Sajid Iqbal, Tanzila Saba, Amjad Rehman, and Zahid Mehmood. A deep learning approach for automated diagnosis and multi-class classification of alzheimer’s disease stages using resting-state fmri and residual neural networks. *Journal of Medical Systems*, 44, 2019. doi:10.1007/s10916-019-1475-2.

- [16] Jesse Davis and Mark Goadrich. The relationship between precision-recall and roc curves. In *Proceedings of the 23rd International Conference on Machine Learning*, page 233–240, 2006. doi:10.1145/1143844.1143874.
- [17] Kyle Gwirtz, Matthias Morzfeld, Alexandre Fournier, and Gauthier Hulot. Can one use earth’s magnetic axial dipole field intensity to predict reversals?, 2020. doi:10.1093/gji/ggaa542.
- [18] Yongcheng Liu, Bin Fan, Lingfeng Wang, Jun Bai, Shiming Xiang, and Chunhong Pan. Semantic labeling in very high resolution images via a self-cascaded convolutional neural network. *ISPRS Journal of Photogrammetry and Remote Sensing*, page 13, 2018. doi:10.1016/j.isprsjprs.2017.12.007.