

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

Exame de Qualificação de Mestrado

18 de Outubro de 2024

CLASSIFICAÇÃO DE PADRÕES DE LOCALIZAÇÃO DE PROTEÍNAS
SUBCELULARES USANDO MÉTODOS DE SEGMENTAÇÃO

Candidata: Juliana Midlej do Espírito Santo

Orientador: Prof. Dr. Zanoni Dias

Coorientador: Prof. Dr. Hélio Pedrini

Sumário

1	Introdução	1
1.1	Descrição do Problema	1
1.2	Objetivos	2
1.3	Hipóteses do Trabalho	2
1.4	Organização do Texto	3
2	Revisão Bibliográfica	4
2.1	Fundamentação Teórica	4
2.1.1	Rede Neural Artificial	4
2.1.2	Rede Neural Profunda	6
2.1.3	Rede Neural Convolucional Profunda	6
2.1.4	Aprendizado Fracamente Supervisionado	7
2.1.5	Aumentação de Dados	8
2.1.6	Segmentação de Imagens Fracamente Supervisionada	8
2.2	Trabalhos Correlatos	9
2.2.1	Modelos com Padrões Únicos	9
2.2.2	Modelos com Padrões Mistos	10
3	Material e Métodos	12
3.1	Metodologia	12
3.1.1	Pré-Processamento	12
3.1.2	Seleção de Modelos	13
3.1.3	Treinamento	13
3.1.4	Teste	13
3.2	Base de Dados	13
3.3	Métricas de Avaliação	15
3.4	Recursos Computacionais	17
4	Plano de Trabalho e Cronograma de Execução	19
	Bibliografia	21

Resumo

Este trabalho aborda a classificação de padrões de localização de proteínas em células individuais usando imagens de microscopia fluorescente. As proteínas, compostas por aminoácidos, desempenham funções vitais no organismo, como transporte de oxigênio, proteção contra patógenos e catálise de reações químicas. A função dos sistemas celulares é predominantemente definida pela estrutura, quantidade, localização espacial e interações das proteínas individuais que, em conjunto, formam o proteoma. O aumento dos dados de imagens fluorescentes demanda modelos computacionais avançados para investigar como a regulação espacial das proteínas contribui para a função celular em condições de saúde e doença. Este estudo propõe o uso de redes neurais convolucionais profundas (DCNNs) com aprendizado fracamente supervisionado para segmentar as células e classificar os padrões de localização das proteínas. Os experimentos utilizarão a base de dados do Atlas de Proteínas Humanas (HPA).

Capítulo 1

Introdução

Quando observamos uma célula através de um microscópio ou analisamos sua atividade elétrica ou bioquímica, estamos, em essência, observando proteínas. As proteínas constituem a maior parte da massa seca de uma célula. Elas não são apenas as unidades fundamentais das células, mas também executam a maior parte das funções celulares. As proteínas que são enzimas fornecem complexas superfícies moleculares no interior das células para a catálise de diversas reações químicas [3].

A função dos sistemas celulares é predominantemente definida pela estrutura, quantidade, localização espacial e interações das proteínas que compõem o proteoma. A localização proteica subcelular é fundamental para entender as redes biológicas, pois a expressão e localização das proteínas variam entre diferentes tipos de células humanas e até mesmo dentro de populações celulares geneticamente idênticas [19]. Encontrar essas variações proteicas é uma atividade crucial para compreender o funcionamento celular, o desenvolvimento de doenças e, em última análise, desenvolver melhores tratamentos.

Com o aumento da quantidade de dados de imagens de microscopia fluorescente, torna-se necessário o desenvolvimento de modelos computacionais mais avançados, capazes de classificar a distribuição espacial de proteínas em células individuais. Esses modelos são fundamentais para investigar como a regulação espacial das proteínas contribui para a função celular em condições de saúde e doença [19].

Este trabalho propõe a aplicação de modelos de aprendizado de máquina capazes de classificar padrões de localização de proteínas subcelulares em células individuais a partir de imagens de microscopia fluorescente no conjunto de dados do Atlas de Proteínas Humanas (HPA)¹.

1.1 Descrição do Problema

Frequentemente, muitas proteínas diferentes se reúnem em um local específico para realizar uma tarefa, e o resultado exato dessa tarefa depende de quais proteínas estão presentes. Diferentes distribuições subcelulares de uma proteína podem dar origem a uma grande heterogeneidade funcional entre as células.

¹<https://www.proteinatlas.org>

Para analisar como a regulação espacial das proteínas influencia a função celular, é necessário classificar rótulos de localização de organelas celulares para cada célula em uma imagem. Esses rótulos indicam onde, dentro da célula, as proteínas estão localizadas, como no núcleo, nas mitocôndrias ou no citoplasma.

O conjunto de dados de treinamento foi anotado de maneira que um ou múltiplos rótulos foram atribuídos a cada imagem ao avaliar os padrões de localização de todas as células. Por isso, os rótulos atribuídos a cada célula em uma imagem podem não ser totalmente precisos e o aprendizado será fracamente supervisionado. Anotações individuais para cada célula em uma imagem estão disponíveis apenas no conjunto de dados de teste.

Historicamente, a classificação de proteínas foi limitada a padrões únicos em alguns poucos tipos de células específicas. Como a arquitetura celular possui estruturas dinâmicas e é refinada com compartimentos de subestruturas especializadas, esse limite de padrões fornece apenas uma descrição superficial da célula. Assim, para compreender detalhadamente a biologia celular, os modelos devem ser capazes de classificar padrões mistos, ou seja, realizar classificação multirrótulo, em uma grande variedade de células diferentes.

É importante destacar ainda que a classificação da distribuição espacial de proteínas em células individuais possui desafios técnicos significativos, como a classificação multirrótulo, o desequilíbrio extremo de classes na base de dados e a necessidade de anotações de células individuais, que são caras e demoradas.

1.2 Objetivos

Este trabalho tem como objetivo principal classificar padrões de localização de proteínas subcelulares em células individuais a partir de imagens de microscopia fluorescente.

Para atingir esse propósito, alguns objetivos específicos devem ser alcançados:

- Realizar a revisão bibliográfica e o estudo das principais abordagens utilizadas na classificação de proteínas subcelulares em imagens de microscopia fluorescente;
- Aplicar métodos de segmentação com aprendizado fracamente supervisionado para segmentar as células e classificar os padrões de localização de proteínas subcelulares;
- Avaliar e comparar o desempenho da metodologia proposta com outras abordagens disponíveis na literatura, identificando suas vantagens e limitações;
- Documentar e publicar os resultados obtidos.

1.3 Hipóteses do Trabalho

As seguintes hipóteses serão investigadas neste trabalho:

- É possível classificar com precisão padrões de localização de proteínas subcelulares em células individuais a partir de imagens de microscopia fluorescente usando métodos de segmentação fracamente supervisionados;

- Modelos que empregam segmentação multirrotulo com aprendizado fracamente supervisionado podem aumentar a precisão na classificação da localização de proteínas subcelulares em comparação com abordagens tradicionais e recentes;
- A aplicação de técnicas de aumento de dados resultará em modelos mais robustos e generalizáveis.

1.4 Organização do Texto

Este projeto está organizado em quatro capítulos. No Capítulo 1, são descritos o problema de pesquisa a ser abordado, assim como os principais objetivos e contribuições esperadas. No Capítulo 2, são introduzidos conceitos importantes para o entendimento do projeto, além da apresentação de alguns trabalhos relacionados relevantes existentes na literatura. No Capítulo 3, são detalhadas a metodologia, as bases de dados e as métricas que serão utilizadas para avaliação dos modelos. No Capítulo 4, são apresentados o plano de trabalho e o cronograma de execução.

Capítulo 2

Revisão Bibliográfica

Este capítulo se divide em duas partes: inicialmente, são apresentados conceitos e técnicas fundamentais para problemas de classificação e segmentação de imagens; em seguida, são apresentados trabalhos correlatos que buscam resolver o problema da classificação de padrões de localização de proteínas em imagens de microscopia fluorescente.

2.1 Fundamentação Teórica

A fundamentação teórica é essencial para auxiliar na compreensão do problema abordado neste trabalho e na seção de trabalhos correlatos. Portanto, são apresentados nos tópicos subsequentes conceitos e técnicas utilizadas nos processos de classificação e segmentação de imagens.

2.1.1 Rede Neural Artificial

Redes Neurais Artificiais (*Artificial Neural Networks*, ou ANNs) são modelos computacionais inspirados na estrutura e funcionamento do cérebro humano [24, 29, 30]. Elas são compostas por unidades de processamento simples, denominadas neurônios artificiais, que são organizadas em camadas [27]. As ANNs são capazes de aprender e generalizar a partir de exemplos, tornando-as ferramentas poderosas para tarefas de reconhecimento de padrões, classificação e predição.

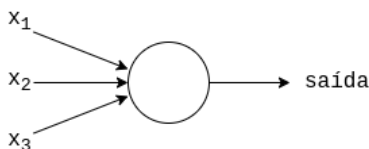


Figura 2.1: Exemplo de um perceptron.

Um dos primeiros modelos de neurônio artificial foi o perceptron. Ele foi desenvolvido nas décadas de 1950 e 1960 pelo cientista Frank Rosenblatt [29], inspirado em trabalhos anteriores de McCulloch e Pitts [24]. A Figura 2.1 mostra um exemplo de um perceptron

que recebe três entradas e produz uma saída. A fórmula matemática para a saída y de um perceptron é dada por:

$$y = \begin{cases} 0, & \text{se } \sum_j w_j x_j \leq T \\ 1, & \text{se } \sum_j w_j x_j > T \end{cases}$$

onde x_j são as entradas, w_j são os pesos associados às entradas, T é o limiar. Os pesos são números reais que expressam a importância das respectivas entradas na saída. Assim como os pesos, o limiar é um número real, que é um parâmetro do neurônio. Variando os pesos e o limiar, podemos obter diferentes modelos de tomada de decisão.

Hoje, é mais comum usar outros modelos de neurônios artificiais, sendo o principal deles o neurônio sigmoide. Os neurônios sigmoides são semelhantes aos perceptrons, mas modificados para que pequenas mudanças em seus pesos e viés causem apenas uma pequena mudança em sua saída [27].

A função de ativação de um neurônio artificial é uma função que é aplicada após a soma ponderada das entradas com o viés e define como o neurônio produz sua saída [10]. Nos perceptrons, a função de ativação utilizada é a função degrau. Nos neurônios sigmoides, utiliza-se a função de ativação sigmoide. A fórmula matemática para a saída y de um neurônio sigmoide é dada por:

$$y = \sigma(z) = \frac{1}{1 + e^{-z}}$$

onde $\sigma(z)$ é a função de ativação, e é a base do logaritmo natural e z é a soma ponderada das entradas com o viés, dada por:

$$z = \sum_{j=1}^n w_j x_j + b$$

na qual x_j são as entradas, w_j são os pesos associados às entradas e b é o viés [27].

Como demonstrado por Rumelhart, Hinton e Williams [30], a vantagem crucial de usar a função de ativação sigmoide em vez da função degrau do perceptron é que a função sigmoide é contínua e diferenciável. Dessa forma, o aprendizado é realizado através de um processo de ajuste dos pesos da rede neural, minimizando a diferença entre a saída prevista e a saída desejada, utilizando algoritmos como o gradiente descendente e a retropropagação do erro.

Outras funções de ativação, como ReLU (*Rectified Linear Unit*) e \tanh (tangente hiperbólica) são amplamente utilizadas em modelos de redes neurais atualmente. Essas funções de ativação têm características diferentes que podem melhorar o desempenho da rede neural em diversas tarefas. No entanto, o modelo do neurônio permanece o mesmo que o do neurônio sigmoide; a única diferença é a função de ativação utilizada.

De forma mais detalhada, uma ANN é um modelo computacional que consiste em unidades de processamento chamadas neurônios, conectadas entre si em camadas. Cada neurônio recebe entradas, processa essas informações através de uma função de ativação e transmite uma saída para os neurônios na camada seguinte. A Figura 2.2 mostra um exemplo de uma rede neural artificial. A primeira camada em uma rede é chamada de camada de entrada

(*input layer*), enquanto a última é chamada de camada de saída (*output layer*). Todas as camadas entre as duas são referidas como camadas ocultas (*hidden layers*). Cada camada é formada por um ou mais neurônios [27].

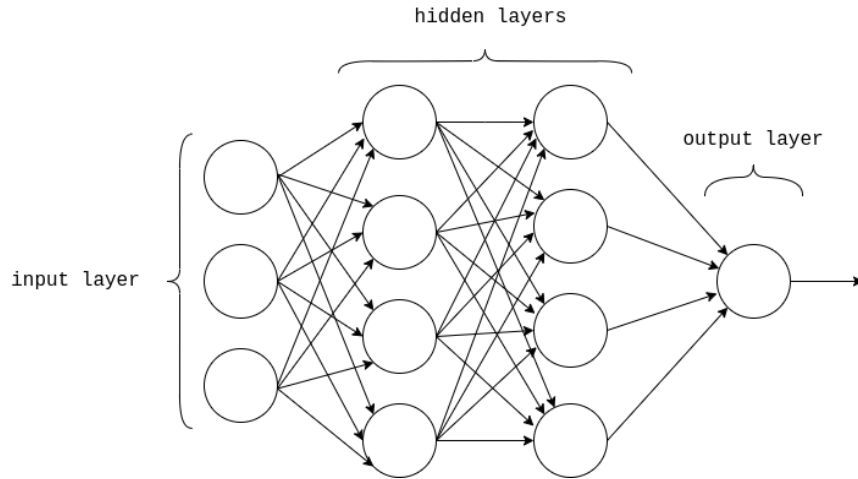


Figura 2.2: Exemplo de uma rede neural artificial.

2.1.2 Rede Neural Profunda

O aprendizado profundo permite que modelos computacionais compostos de múltiplas camadas de processamento aprendam representações de dados com múltiplos níveis de abstração [20]. Uma Rede Neural Profunda (*Deep Neural Network*, ou DNN) é uma rede com uma estrutura de muitas camadas, com as primeiras extraindo características mais simples e específicas sobre o dado de entrada e, posteriormente, criando uma hierarquia de características cada vez mais complexas e abstratas.

Rumelhart, Hinton e Williams [30] desempenharam um papel fundamental no desenvolvimento e popularização do algoritmo de retropropagação do erro durante os anos 1980, um método que revolucionou o treinamento de redes neurais com múltiplas camadas. Antes desse avanço, a dificuldade em ajustar os pesos limitava a eficácia desses modelos. Com a retropropagação, foi possível calcular de forma eficiente os gradientes do erro através das camadas da rede, permitindo o ajuste de pesos e o treinamento de redes neurais profundas.

Muitos dos problemas estudados pela comunidade de inteligência artificial que não possuíam soluções obtiveram grande avanços com o uso de redes neurais profundas e, considerando-se tarefas com imagens, com os modelos de redes neurais convolucionais profundas [20].

2.1.3 Rede Neural Convolucional Profunda

As Redes Neurais Convolucionais Profundas (*Deep Convolutional Neural Networks*, ou DCNNs) foram inspiradas pelo trabalho de Fukushima, que propôs o “Neocognitron” [9], baseado no modelo de campo receptivo hierárquico do córtex visual proposto por Hubel e

Wiesel [16]. Subsequentemente, Waibel et al. [37] introduziram redes com pesos compartilhados entre campos receptivos temporais e treinamento por retropropagação para reconhecimento de fonemas, e LeCun et al. [21] desenvolveram uma arquitetura de DCNN para o reconhecimento de documentos.

As DCNNs geralmente consistem em camadas convolucionais que aplicam filtros para extração de características locais em diferentes partes da imagem, seguidas por camadas de agrupamento (*pooling*) que reduzem o tamanho da entrada e o número de parâmetros do modelo enquanto preservam informações importantes [27]. Os filtros, também chamados de *kernels*, são aprendidos pela rede durante o processo de treinamento. Além disso, há um compartilhamento de pesos nos filtros, ou seja, os mesmos filtros são aplicados em diferentes regiões da imagem, o que permite que a rede seja mais eficiente e capture padrões similares em diferentes partes da entrada.

Após as camadas de convolução e *pooling*, as DCNNs frequentemente incluem camadas totalmente conectadas (*fully connected*), onde todos os neurônios estão conectados a cada unidade da camada anterior. Por fim, a última camada, chamada camada de saída (*output layer*), é também totalmente conectada e, em tarefas de classificação, atribui uma probabilidade para cada classe possível [27].

A Figura 2.3 apresenta um modelo de uma DCNN contendo camadas ocultas de extração de características usando camadas convolucionais e de agrupamento, e camadas densas para a classificação usando camadas totalmente conectadas. Atualmente, algumas das arquiteturas de DCNNs mais destacadas para tarefas de classificação de imagens em *benchmarks* como ImageNet [31] incluem ResNet [12], InceptionV4 [35], EfficientNet [25], DenseNet [14] e SENet [13]. Cada uma dessas arquiteturas traz inovações específicas que aprimoram seu desempenho em tarefas de classificação.

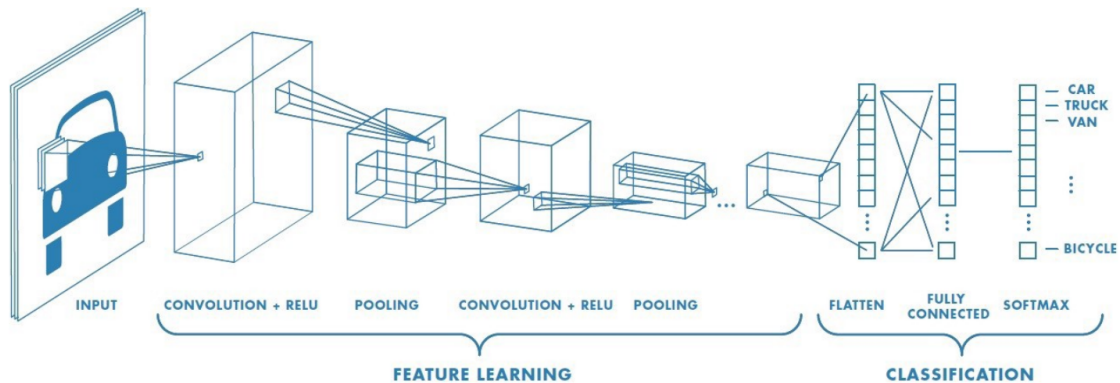


Figura 2.3: Exemplo de uma rede neural convolucional profunda [6].

2.1.4 Aprendizado Fracamente Supervisionado

Supervisão forte de dados refere-se ao uso de conjuntos de dados de treinamento que são completamente rotulados, ou seja, cada exemplo de treinamento tem um rótulo preciso e confiável associado a ele. Supervisão fraca de dados refere-se ao uso de conjuntos de dados de treinamento que são parcialmente rotulados, têm rótulos imprecisos ou incompletos.

Técnicas de aprendizado de máquina bem-sucedidas, como o aprendizado profundo, geralmente requerem rótulos para um grande conjunto de dados de treinamento. No entanto, em muitas tarefas, pode ser difícil obter informações de supervisão forte devido ao alto custo do processo de rotulagem dos dados. Portanto, é desejável que as técnicas de aprendizado de máquina sejam capazes de trabalhar com supervisão fraca [42].

Tipicamente, existem três tipos de supervisão fraca. O primeiro é a supervisão incompleta, onde apenas um subconjunto (geralmente pequeno) dos dados de treinamento é rotulado enquanto o restante permanece não rotulado. O segundo tipo é a supervisão inexata, onde são fornecidos apenas rótulos de baixa precisão. Por exemplo, considerando a tarefa de categorização de imagens é desejável que cada objeto nas imagens seja anotado; no entanto, geralmente temos apenas rótulos ao nível da imagem em vez de rótulos ao nível do objeto. O terceiro tipo é a supervisão imprecisa, onde os rótulos fornecidos nem sempre são a verdade absoluta. Essa situação ocorre, por exemplo, quando o anotador de imagens é descuidado ou está cansado, ou algumas imagens são difíceis de categorizar [42].

2.1.5 Aumentação de Dados

Aumentação de Dados (*Data Augmentation*) é um método utilizado para mitigar o problema de *overfitting* em modelos de aprendizado de máquina. A técnica consiste em aumentar artificialmente o tamanho e a diversidade do conjunto de dados de treinamento. Com isso, o modelo é exposto a uma maior variedade de exemplos, o que melhora seu desempenho e sua capacidade de generalização [38].

A aumento de dados ocorre por meio de técnicas de distorção de dados ou superamostragem. As aumentações de distorção de dados transformam imagens existentes de maneira que seus rótulos sejam preservados. Isso inclui aumentações como transformações geométricas e de cor, apagamento aleatório, treinamento adversarial e transferência de estilo neural. Já as aumentações de superamostragem criam instâncias sintéticas e as adicionam ao conjunto de treinamento, incluindo técnicas como mistura de imagens, aumentações no espaço de características e Redes Generativas Adversariais (GANs) [33].

2.1.6 Segmentação de Imagens Fracamente Supervisionada

Segmentação instancial de dados é uma tarefa complexa no campo da visão computacional, que envolve a identificação e a separação de cada instância de um objeto em uma imagem, atribuindo um rótulo a cada pixel pertencente a uma instância específica. Isso é crucial para aplicações em áreas como a robótica, a condução autônoma e a análise de imagens médicas, onde é importante distinguir entre diferentes objetos ou partes de um mesmo objeto.

Class Activation Mapping (CAM) é uma técnica que permite visualizar as regiões da imagem que mais contribuíram para a decisão do modelo [41]. Ele gera um mapa de ativação que destaca as áreas importantes para a classificação. *Gradient-weighted Class Activation Mapping* (Grad-CAM) melhora o CAM ao utilizar os gradientes dos neurônios de uma determinada classe para produzir um mapa de calor ponderado, que indica as regiões da imagem mais relevantes para a predição [32]. Puzzle-CAM é uma variante mais recente que melhora a

explicabilidade ao segmentar a imagem em regiões e aplicar CAM em cada região, permitindo uma interpretação mais detalhada e granular [17].

Dado que dados rotulados são caros, há uma tendência crescente entre os pesquisadores de aprendizado profundo de desenvolver modelos que possam aproveitar dados não rotulados ou rotulados de forma fraca. Por exemplo, abordagens guiadas por mapas de ativação de classe (CAM) [7, 8] e adaptações supervisionadas de forma fraca de Mask R-CNNs (redes DCNNs baseadas em regiões) [11, 39] foram projetadas para localizar regiões de sinal e atribuir rótulos de objetos precisos, dados rótulos de baixa precisão em nível de imagem.

David et al. [7] analisaram técnicas complementares da literatura e propuseram duas novas abordagens: P-NOC e C²AM-H. O P-NOC promove o treinamento conjunto de duas redes adversárias de geração de CAMs, onde o gerador aprende a apagar progressivamente regiões com características específicas de classe e o discriminador refina-se para focar em novas características discriminantes. Já o C²AM-H utiliza os *pseudo-priors* de segmentação gerados pelo P-NOC para guiar a aprendizagem de informações de saliência de forma fracamente supervisionada, resultando em máscaras de segmentação pseudo-semântica de maior qualidade. O estudo conclui que a completude da predição, a sensibilidade aos contornos semânticos dos objetos e a robustez contra ruídos são componentes fundamentais na segmentação de dados fracamente supervisionados, com as estratégias propostas produzindo *priors* de segmentação semântica competitivos e superiores ao *baseline*.

2.2 Trabalhos Correlatos

Nesta seção, são apresentados trabalhos que propõem métodos para segmentar e/ou classificar imagens de microscopia de células. Os trabalhos foram escolhidos por suas contribuições e relevância para a classificação de padrões de localização subcelular de proteínas.

2.2.1 Modelos com Padrões Únicos

O artigo de Boland, Markey e Murphy [4] descreve métodos para a descrição numérica e classificação de padrões de localização de proteínas em imagens de microscopia de fluorescência, utilizando momentos de Zernike e características de textura de Haralick. As imagens foram processadas para remover fluorescência fora de foco e de fundo, e classificadores baseados em árvores de decisão e redes neurais foram aplicados, alcançando até 88% de precisão. Estes métodos demonstraram a viabilidade de aplicar técnicas de reconhecimento de padrões para localização subcelular, abrindo possibilidades para a descrição automatizada de proteínas e automação de funções de microscopia.

Os resultados de Boland, Markey e Murphy [4] mostraram que é possível automatizar a classificação de padrões de localização de proteínas e serviram de motivação para dezenas de trabalhos que surgiram desde então. Com esforços para essa tarefa, surgiram propostas de métodos usando classificadores como ANNs [5, 15], Máquinas de Vetores de Suporte (*Support Vector Machines*, ou SVMs) [15, 22, 26], *Random Forests* [26] e regressão logística [23]. Recentemente, DCNNs têm sido bem sucedidas na tarefa de classificação. Essa abordagem foi utilizada nos trabalhos correlatos que serão citados na sequência.

Kraus et al. [18] propuseram combinar DCNNs com aprendizado de múltiplas instâncias (*Multiple Instance Learning*, ou MIL) para segmentar e classificar imagens de microscopia celular. Métodos com aprendizado profundo eram abordagens que estavam dominando os desafios de reconhecimento de objetos, porém, os modelos não eram aplicáveis diretamente ao conjunto de dados de imagens de microscopia. Assim, os autores apresentaram uma nova arquitetura de rede neural profunda que usa aprendizado de múltiplas instâncias para segmentar e classificar imagens de microscopia fracamente rotuladas. O modelo proposto foi capaz de localizar regiões com células nas imagens de microscopia de alta resolução.

2.2.2 Modelos com Padrões Mistos

Os trabalhos citados anteriormente focam em padrões únicos ou em um único tipo de célula. No entanto, devido à natureza dinâmica e complexa da arquitetura celular, que inclui compartimentos especializados com subestruturas distintas, aquelas abordagens fornecem apenas uma visão superficial da biologia celular. Para aprofundar o entendimento da biologia celular, é essencial realizar uma classificação mais ampla e detalhada. Portanto, assim como o presente trabalho, os estudos correlatos descritos a seguir propõem métodos que abordam essa complexidade ao realizar a detecção de padrões mistos com a classificação multirrotulo de vários tipos de células.

Sullivan et al. [34] foram pioneiros em trabalhar com classificação múltipla de proteínas. Eles propuseram combinar duas abordagens para a classificação em larga escala de imagens de microscopia fluorescente. Primeiro, integrar a tarefa de classificação de imagens em um videogame (EVE Online), como um minijogo chamado *Project Discovery*. Segundo, usar o aprendizado profundo para criar uma ferramenta automatizada de anotação celular (Loc-CAT). Eles então combinaram as anotações dos jogadores e a ferramenta com aprendizado profundo, aplicando transferência de aprendizado para criar um método aprimorado para classificação de padrões de localização de proteínas subcelulares.

Combinando-se as anotações dos jogadores e o aprendizado profundo, eles aplicaram transferência de aprendizado para criar um “jogador” aprimorado que conseguiu caracterizar a distribuição de proteínas subcelulares. Sullivan et al. [34] mostraram como os jogadores e as redes neurais profundas se destacam em diferentes tipos de classificações e que a saída dos jogadores pode ser usada para aumentar e melhorar os modelos de aprendizado profundo.

Xiang et al. [40] propuseram um *framework* que tem como objetivo a extração automática de características e a classificação multiclasse de proteínas em imagens de células humanas. Eles criaram o modelo denominado rede neural convolucional assimétrica e em várias escalas (*Asymmetric and Multi-scale Convolutional Neural Network*, ou AMC-Net). A arquitetura proposta combina três ramos da DCNN. Especificamente, cada ramificação possui diferentes tamanhos de núcleo convolucional e da camada de agrupamento (*pooling*), a fim de gerar mapas de características com vários campos receptivos. Além disso, a escala das imagens que são alimentadas em cada ramo pode ser diferente. O método proposto alcançou taxa F_1 de 0,823 para a tarefa de classificação de padrões de localização de proteínas subcelulares em imagens de células humanas da base de dados do HPA.

Ouyang et al. [28] analisaram os resultados de uma competição de classificação de proteí-

nas realizada no Kaggle. A competição foi feita para abordar os dois principais desafios do desenvolvimento de modelos computacionais para a classificação de padrões de localização subcelular de proteínas em imagens de microscopia confocal do Atlas de Proteínas Humanas (HPA): o desequilíbrio de classes e a classificação multirrótulo.

Todas as equipes, exceto uma, usaram modelos de aprendizado profundo. A maioria utilizou variações das arquiteturas ResNet, DenseNet ou Inception como *backbone*, conhecidas por sua eficácia em tarefas de reconhecimento de imagens. Para resolver o problema multirrótulo, a maioria das equipes utilizou a entropia cruzada binária. Muitos lidaram com o desequilíbrio de classes aplicando pesos de classe ou usando *focal loss* para treinar os modelos, além de empregarem a estratificação multirrótulo para gerar conjuntos de validação. A maioria das equipes usou, adicionalmente ao conjunto de treinamento disponibilizado para a competição, a base de dados pública HPAv18 e estratégias de aumento de dados, como recorte aleatório, rotação e inversão. A equipe vencedora gerou modelos com pontuações macro F_1 variando de aproximadamente 0,56 a 0,59, mais de 20% melhores do que os resultados do Loc-CAT.

Al-joudi e Arif [2] utilizaram um modelo de DCNN chamado GapNet-PL para a tarefa de classificação multirrótulo no conjunto de dados do Atlas de Proteínas Humanas (HPA) atingindo taxa F_1 de 0,54. Aggarwal et al. [1] utilizaram uma técnica de *ensemble* para melhorar o desempenho das redes neurais convolucionais de última geração existentes e dos modelos pré-treinados. Eles apresentaram um modelo de aprendizado profundo baseado em *ensemble* empilhado, que oferece um classificador mais confiável e robusto. Os resultados mostram que a estratégia de *ensemble* proposta teve um bom desempenho na classificação multirrótulo de imagens do HPA, com revocação, precisão e taxa F_1 de 0,70, 0,72 e 0,71, respectivamente.

Le et al. [19] apresentaram um artigo sobre a competição realizada no Kaggle para o desenvolvimento de soluções de aprendizado de máquina capazes de classificar padrões de localização de proteínas subcelulares em células individuais a partir de imagens de microscopia fluorescente.

A competição apresentou desafios técnicos significativos, como a classificação multirrótulo, o desequilíbrio extremo de classes e a necessidade de anotações de células individuais, que são caras e demoradas. Assim, modelos fracamente supervisionados que utilizam anotações de nível de imagem foram utilizados para superar essas limitações. O uso de redes neurais profundas, especialmente variações de redes neurais convolucionais, foi fundamental para abordar os problemas de classificação e segmentação.

A maioria das soluções utilizou uma de duas abordagens principais, referidas como “modelos a nível de célula” e “modelos a nível de imagem”. Modelos a nível de imagem recebem imagens inteiras e prevêm localizações a nível de imagem. Para obter rótulos de células individuais, as equipes combinaram mapas de ativação de classes individuais com regiões segmentadas de células ou usaram uma abordagem de “*bag-of-cells*”, que concatena aumentações de uma única célula em uma entrada de imagem. A abordagem mais popular para os modelos a nível de imagem foi o Puzzle-CAM [17]. Modelos a nível de célula exigem entradas de células segmentadas para produzir rótulos de células individuais. O melhor resultado obtido por um time foi *mean Average Precision* (mAP) de 0,5667.

Capítulo 3

Material e Métodos

Este capítulo descreve a metodologia, as métricas de avaliação, a base de dados e os recursos computacionais que serão utilizados no desenvolvimento do projeto.

3.1 Metodologia

A metodologia a ser adotada neste projeto está dividida em quatro etapas principais: pré-processamento da base de dados, seleção de modelos de classificação e segmentação, treinamento e teste.

3.1.1 Pré-Processamento

Na etapa de pré-processamento dos dados, serão feitos os processos de redimensionamento das imagens, análise e correção de rótulos incorretos, estratificação multirrótulo para gerar um conjunto de validação, aumento de dados e aplicação de estratégias de amostragem e balanceamento multirrótulo.

O redimensionamento das imagens garante que todas as amostras tenham a mesma dimensão, possibilitando o processamento subsequente pelo modelo. Adicionalmente, reduzir as imagens para um tamanho padrão ajuda a otimizar o tempo de processamento e a memória necessária durante o treinamento.

A análise e correção de rótulos têm o intuito de garantir que cada imagem esteja associada ao rótulo correto. Isso pode incluir a remoção de rótulos inconsistentes ou a atualização de rótulos onde necessário. Essa etapa pode aumentar significativamente o desempenho do modelo. No artigo de Li et al. [22], que utilizou a base de dados do HPA, os autores implementaram uma estratégia de ranqueamento de proteínas com base na confiança do classificador, gerando uma lista de proteínas que eram fortes candidatas para reexame, visando à correção de possíveis erros de rotulagem. Esta abordagem é uma forma de *prediction conformity estimation*, que avalia a conformidade das previsões para identificar e corrigir possíveis erros de rotulagem.

Parte dos dados de treinamento será reservada para a validação do modelo, permitindo avaliar seu desempenho em dados não vistos. Além disso, a aumento de dados será

aplicada para enriquecer o conjunto de treinamento. A aumento de dados expande artificialmente o conjunto, gerando novas imagens a partir das existentes. Particularmente, o *oversampling* mitiga o desbalanceamento de classes, garantindo uma distribuição mais equilibrada.

Outra estratégia importante para lidar com o desbalanceamento é buscar em outras bases de dados mais exemplos de classes minoritárias, enriquecendo ainda mais o conjunto de dados de treinamento e aumentando a representatividade dessas classes.

3.1.2 Seleção de Modelos

Serão analisados e escolhidos os modelos para classificação e segmentação das imagens, com destaque para o Puzzle-CAM [17], que já apresentou bons resultados no artigo apresentado por Le et al. [19], assim como para modelos mais novos, como o P-NOC e o C²AM-H [7].

3.1.3 Treinamento

Caso o conjunto de treinamento ainda apresente classes desbalanceadas após o pré-processamento dos dados, pode-se utilizar uma função de perda ponderada (*weighted loss*), a qual atribui pesos diferentes às classes de acordo com sua representatividade. Essa abordagem ajuda a lidar com o desbalanceamento de classes, garantindo que o modelo aprenda de forma adequada tanto as classes majoritárias quanto as minoritárias.

3.1.4 Teste

Na fase de teste, cada amostra do conjunto de teste será submetida ao modelo de classificação e segmentação. O modelo será avaliado de acordo com as métricas descritas na Seção 3.3.

3.2 Base de Dados

A base de dados que será utilizada é a do Atlas de Proteínas Humanas (HPA), que faz parte da competição pública de Aprendizado de Máquina para classificação de padrões de localização de proteínas em células humanas¹. Adicionalmente ao conjunto de treinamento disponibilizado para a competição, também será utilizado o conjunto de amostras da Seção Subcelular HPA (HPAv20) [36].

O HPA possui milhares de imagens de alta resolução. Cada amostra da base de dados que será utilizada corresponde a um campo de visão contendo múltiplas células (em média, 20 células por amostra).

Existem, ao todo, 19 rótulos diferentes presentes no conjunto de dados (18 rótulos para localizações específicas e um rótulo para sinal negativo e não específico). Todas as amostras são representadas por quatro filtros (armazenados como imagens individuais): a proteína de interesse (verde), os núcleos (azul), os microtúbulos (vermelho) e os retículos endoplasmáticos (amarelo). O filtro verde deve, portanto, ser usado para prever o rótulo, e os outros filtros

¹<https://www.kaggle.com/c/hpa-single-cell-image-classification>

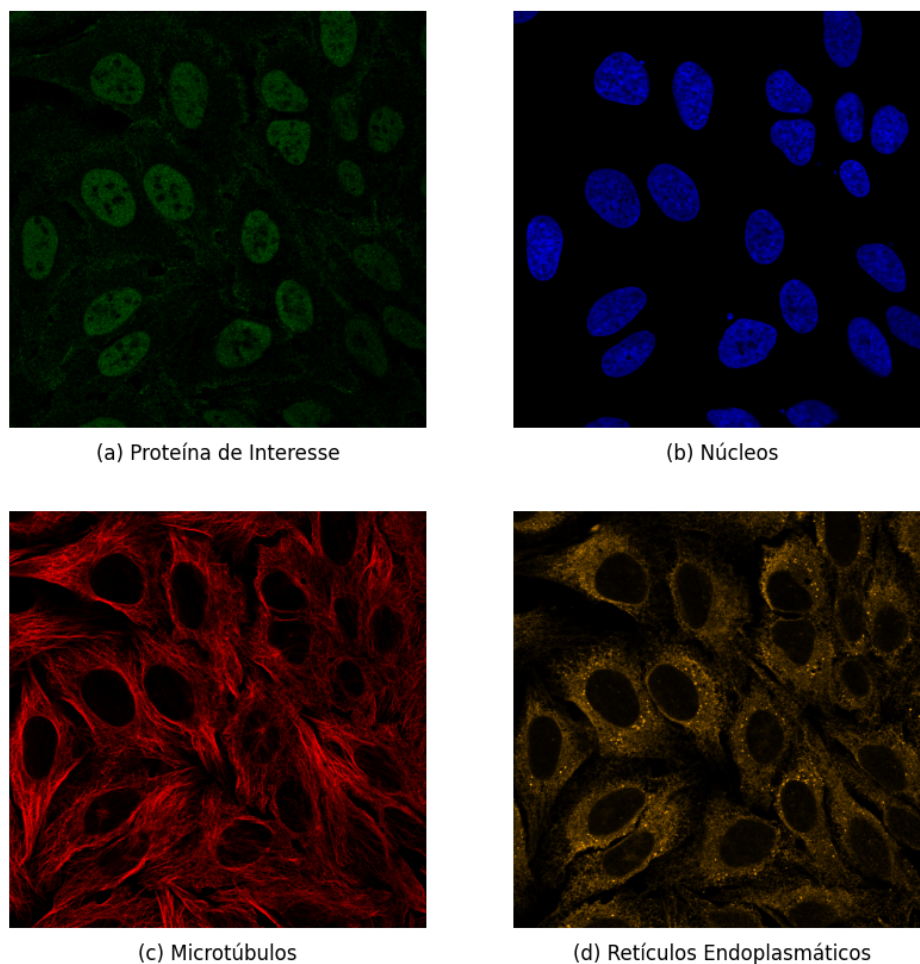


Figura 3.1: Exemplo dos filtros de uma amostra.

são usados como referências. As imagens para treinamento do modelo terão 4 canais para representar os filtros existentes. A Figura 3.1 exemplifica os quatro filtros de uma amostra.

As imagens do conjunto de dados de treinamento foram anotadas de maneira que um ou múltiplos rótulos foram atribuídos a cada imagem ao avaliar os padrões de localização de todas as células. No entanto, devido à heterogeneidade celular, os rótulos atribuídos a cada célula em uma imagem podem não ser totalmente precisos. Em uma mesma imagem de uma população geneticamente idêntica, células individuais podem apresentar diferentes padrões de localização de proteínas. Isso resulta em dados fracamente rotulados e torna a tarefa de classificação uma forma de aprendizado fracamente supervisionado. A Figura 3.2 mostra a imprecisão dos rótulos devido a heterogeneidade celular em uma amostra da base de dados.

O conjunto de dados de treinamento da competição consiste em 21.806 amostras. Para avaliar o desempenho da solução proposta, será utilizado um conjunto de dados de teste contendo 41.597 células em 1.776 amostras anotadas manualmente para rótulos a nível de célula. O conjunto de teste é dividido em um conjunto de teste público com 31% das amostras ($n = 559$) e um conjunto de teste privado com 69% das amostras ($n = 1.217$). O conjunto

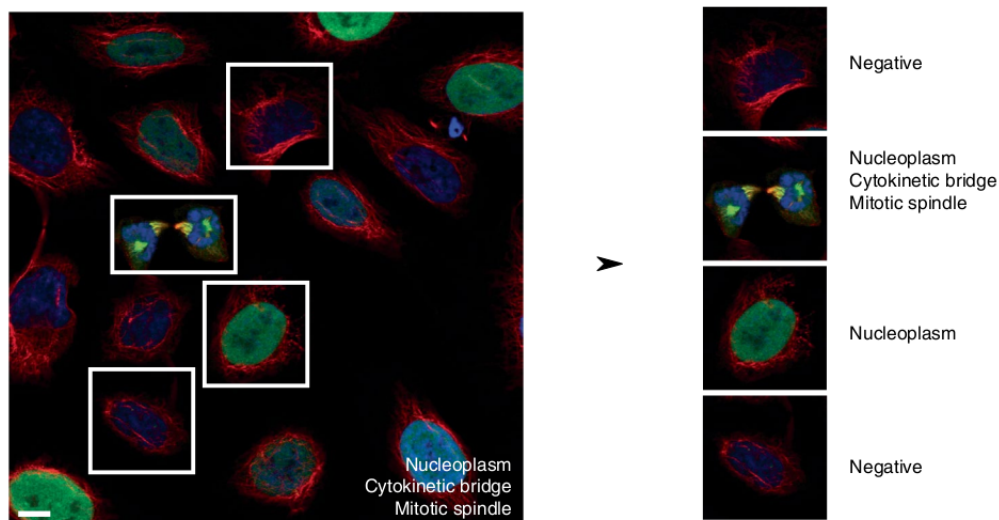


Figura 3.2: Exemplo da heterogeneidade celular em uma amostra. À esquerda a imagem em resolução total e rotulada. À direita, algumas células segmentadas da imagem e seus rótulos individuais. Rótulos a nível de imagem e a nível de célula não são iguais para todas as células de uma imagem [19].

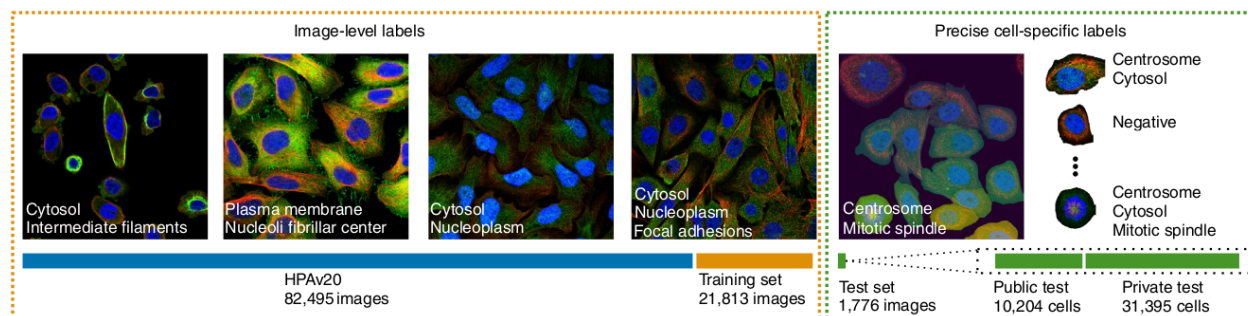


Figura 3.3: Divisão das imagens entre treino e teste na base de dados do HPA [19].

de dados da Seção Subcelular HPA (HPAv20) possui 82.495 amostras. Os conjuntos de treinamento e teste são ilustrados na Figura 3.3.

A Figura 3.4 mostra o número de imagens e células por classe no conjunto de treinamento e no conjunto de teste. A base de dados é altamente desbalanceada.

3.3 Métricas de Avaliação

Para avaliar a classificação serão utilizadas as métricas acurácia, precisão, revocação e a taxa F1. As quatro métricas são usadas para medir a eficácia do modelo proposto para o conjunto de dados do HPA. Antes de definir a fórmula para calcular cada métrica, é importante definir alguns termos que serão usados nas equações:

- P: o número de casos positivos;
- N: o número de casos negativos;

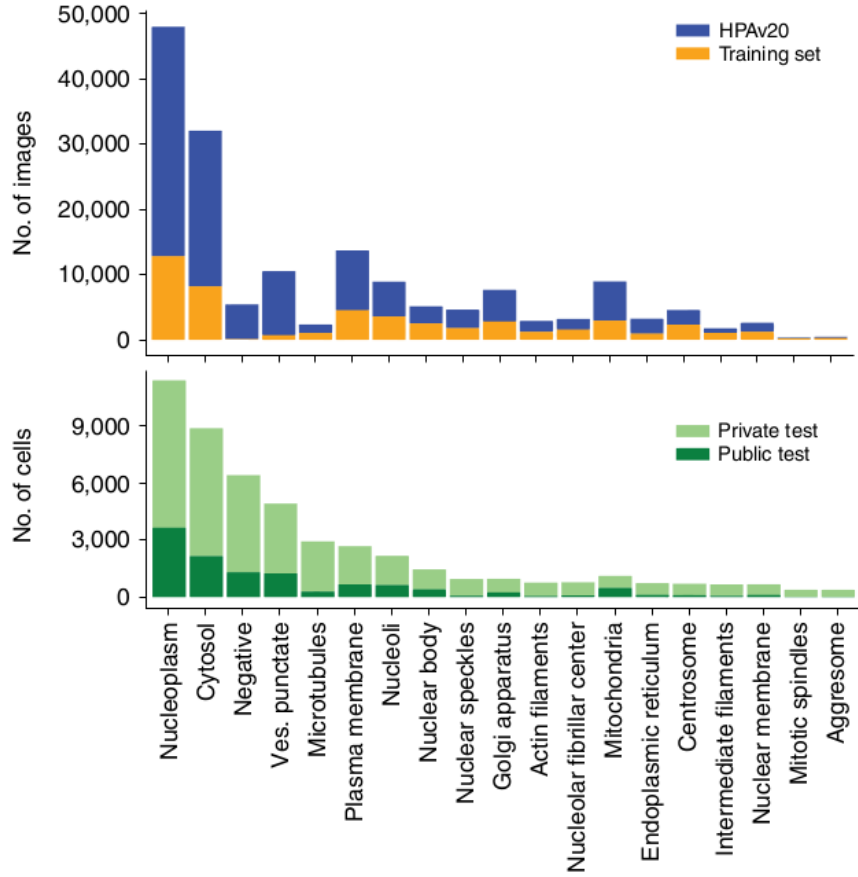


Figura 3.4: Número de imagens e células por classe nos conjuntos de treinamento e teste [19].

- VP: o número de verdadeiros positivos;
- VN: o número de verdadeiros negativos;
- FP: o número de falsos positivos;
- FN: o número de falsos negativos.

A acurácia é uma métrica de avaliação muito usada em problemas de classificação. Seu valor é calculado como apresentado na Equação 3.1. A acurácia de uma classe é o número de casos preditos corretamente com relação a todos os casos existentes.

$$\text{Acurácia} = \frac{VP + VN}{P + N} \quad (3.1)$$

A precisão é calculada pela Equação 3.2. Essa métrica indica a capacidade de classificar como positivo um dado que realmente é positivo.

$$\text{Precisão} = \frac{VP}{VP + FP} \quad (3.2)$$

A revocação é representada pela Equação 3.3. A revocação verifica a capacidade de classificar corretamente dados positivos.

$$\text{Revocação} = \frac{\text{VP}}{\text{VP} + \text{FN}} \quad (3.3)$$

A taxa F1 é dada pela Equação 3.4. Essa métrica corresponde à média harmônica da precisão e da revocação, de modo a obter um valor único que indique a qualidade geral do modelo.

$$F_1 = 2 \times \frac{\text{Precisão} \times \text{Revocação}}{\text{Precisão} + \text{Revocação}} \quad (3.4)$$

Average Precision (AP) é uma métrica usada para avaliar a precisão de um modelo em cada classe individualmente. No contexto de segmentação instancial, ela é calculada pela área sob a curva de precisão-revocação, considerando como VP as detecções que satisfazem um critério mínimo de Interseção sobre União (IoU). O IoU é definido como a razão entre a área da interseção e a área da união das máscaras preditas e de referência, conforme mostrado na Equação 3.5.

$$\text{IoU} = \frac{\text{Área da Interseção}}{\text{Área da União}} \quad (3.5)$$

A métrica AP é definida pela Equação 3.6, onde P_n e R_n são, respectivamente, a precisão e a revocação no ponto n da curva de precisão-revocação, e c representa uma classe específica.

$$\text{AP}_c = \sum_n (R_n - R_{n-1}) P_n \quad (3.6)$$

Para avaliar a segmentação, será utilizada a métrica *mean Average Precision at 60%* (mAP@60). A mAP@60 é a média das precisões médias (AP) para todas as classes no conjunto de dados, considerando um limiar de IoU de 60%, conforme definido na Equação 3.7, onde C é o número total de classes e AP_c é a média das precisões para a classe c .

$$\text{mAP@60} = \frac{1}{|C|} \sum_{c \in C} \text{AP}_c \quad (3.7)$$

3.4 Recursos Computacionais

A implementação deste projeto será feita usando a linguagem de programação Python, dado o grande número de ferramentas desenvolvidas com suporte para essa linguagem. O projeto utilizará bibliotecas de aprendizado de máquina, aprendizado profundo, funções científicas e numéricas e visualização de dados. Algumas bibliotecas que podem ser destacadas são: NumPy², scikit-learn³, PyTorch⁴ e Matplotlib⁵. Os experimentos deste projeto serão reali-

²<https://numpy.org>

³<https://scikit-learn.org>

⁴<https://pytorch.org>

⁵<https://matplotlib.org>

zados no supercomputador Santos Dumont⁶.

⁶<https://sdumont.lncc.br>

Capítulo 4

Plano de Trabalho e Cronograma de Execução

O plano de trabalho é composto pelas seguintes atividades:

1. Obtenção dos créditos obrigatórios em disciplinas do programa de mestrado;
2. Participação no Programa de Estágio Didático (PED);
3. Seleção de uma base de dados;
4. Revisão bibliográfica;
5. Exame de Qualificação do Mestrado (EQM);
6. Pré-processamento da base de dados;
7. Seleção dos modelos de classificação e segmentação;
8. Treinamento dos modelos;
9. Realização de testes e análise dos resultados;
10. Documentação e publicação dos resultados;
11. Escrita do documento da dissertação;
12. Defesa da Dissertação de Mestrado.

O cronograma de execução das atividades propostas, em um período de 24 meses, é apresentado na Tabela 4.1.

Tabela 4.1: Cronograma de atividades dividido em bimestres.

Atividades	1º ano						2º ano					
	1	2	3	4	5	6	1	2	3	4	5	6
1	•	•	•	•	•	•						
2	•	•	•	•	•	•						
3		•	•									
4		•	•	•								
5				•								
6				•	•	•						
7					•	•						
8							•	•	•			
9										•		
10							•	•	•	•		
11										•	•	
12												•

Bibliografia

- [1] S. Aggarwal, S. Gupta, D. Gupta, Y. Gulzar, S. Juneja, A. A. Alwan, and A. Nauman. An artificial intelligence-based stacked ensemble approach for prediction of protein subcellular localization in confocal microscopy images. *Sustainability*, 15(2):1695, 2023.
- [2] L. Al-joudi and M. Arif. Classification of human protein in multiple cells microscopy images using CNN. *Computers, Materials & Continua*, 76(2), 2023.
- [3] B. Alberts, A. Johnson, J. Lewis, D. Morgan, M. Raff, K. Roberts, P. Walter, J. Wilson, and T. Hunt. *Biologia Molecular da Célula*. Artmed Editora, 2017.
- [4] M. V. Boland, M. K. Markey, and R. F. Murphy. Automated recognition of patterns characteristic of subcellular structures in fluorescence microscopy images. *Cytometry: The Journal of the International Society for Analytical Cytology*, 33(3):366–375, 1998.
- [5] M. V. Boland and R. F. Murphy. A neural network classifier capable of recognizing the patterns of all major subcellular structures in fluorescence microscope images of hela cells. *Bioinformatics*, 17(12):1213–1223, 2001.
- [6] C. C. Chatterjee. Basics of the classic CNN, 2019. Disponível em: <https://towardsdatascience.com/basics-of-the-classic-cnn-a3dce1225add>. Acesso em: 12/09/2024.
- [7] L. David, H. Pedrini, and Z. Dias. P-NOC: Adversarial training of CAM generating networks for robust weakly supervised semantic segmentation priors. *Journal of Visual Communication and Image Representation*, page 104187, 2024.
- [8] L. Fang, H. Xu, Z. Liu, S. Parisot, and Z. Li. EHSOD: CAM-guided end-to-end hybrid-supervised object detection with cascade refinement. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 34, pages 10778–10785, 2020.
- [9] K. Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, 36(4):193–202, 1980.
- [10] S. Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1998.
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick. Mask R-CNN. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 2961–2969, 2017.

- [12] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, 2016.
- [13] J. Hu, L. Shen, and G. Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 7132–7141, 2018.
- [14] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4700–4708, 2017.
- [15] K. Huang and R. F. Murphy. Boosting accuracy of automated classification of fluorescence microscope images for location proteomics. *BMC Bioinformatics*, 5:1–19, 2004.
- [16] D. H. Hubel and T. N. Wiesel. Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *The Journal of Physiology*, 160(1):106, 1962.
- [17] S. Jo and I.-J. Yu. Puzzle-CAM: Improved localization via matching partial and full features. In *Proceedings of the IEEE International Conference on Image Processing (ICIP)*, pages 639–643. IEEE, 2021.
- [18] O. Z. Kraus, J. L. Ba, and B. J. Frey. Classifying and segmenting microscopy images with deep multiple instance learning. *Bioinformatics*, 32(12):i52–i59, 2016.
- [19] T. Le, C. F. Winsnes, U. Axelsson, H. Xu, J. M. Kaimal, D. Mahdessian, S. Dai, I. S. Makarov, V. Ostankovich, Y. Xu, E. Benhamou, C. Henkel, R. A. Solovyev, N. Banić, V. Bošnjak, A. Bošnjak, A. Miličević, W. Ouyang, and E. Lundberg. Analysis of the human protein atlas weakly supervised single-cell classification competition. *Nature Methods*, 19(10):1221–1229, 2022.
- [20] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.
- [21] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.
- [22] J. Li, J. Y. Newberg, M. Uhlén, E. Lundberg, and R. F. Murphy. Automated analysis and reannotation of subcellular locations in confocal images from the human protein atlas. *PLOS ONE*, 7(11):e50514, 2012.
- [23] J. Li, L. Xiong, J. Schneider, and R. F. Murphy. Protein subcellular location pattern classification in cellular images using latent discriminative models. *Bioinformatics*, 28(12):i32–i39, 2012.
- [24] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5:115–133, 1943.
- [25] T. Mingxing and V. L. Quoc. Efficientnet: Rethinking model scaling for convolutional neural networks. In *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.

- [26] J. Y. Newberg, J. Li, A. Rao, F. Pontén, M. Uhlén, E. Lundberg, and R. F. Murphy. Automated analysis of human protein atlas immunofluorescence images. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 1023–1026. IEEE, 2009.
- [27] M. A. Nielsen. *Neural networks and deep learning*, volume 25. Determination Press, 2015.
- [28] W. Ouyang, C. F. Winsnes, M. Hjelmare, A. J. Cesnik, L. Åkesson, H. Xu, D. P. Sullivan, S. Dai, J. Lan, P. Jinmo, S. M. Galib, C. Henkel, K. Hwang, D. Poplavskiy, B. Tunguz, R. D. Wolfinger, Y. Gu, C. Li, J. Xie, D. Buslov, S. Fironov, A. Kiselev, D. Panchenko, X. Cao, R. Wei, Y. Wu, X. Zhu, K.-L. Tseng, Z. Gao, C. Ju, X. Yi, H. Zheng, C. Kappel, and E. Lundberg. Analysis of the human protein atlas image classification competition. *Nature Methods*, 16(12):1254–1261, 2019.
- [29] F. Rosenblatt. *The perceptron, a perceiving and recognizing automaton Project Para*. Cornell Aeronautical Laboratory, 1957.
- [30] D. E. Rumelhart, G. E. Hinton, and R. J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [31] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, 115:211–252, 2015.
- [32] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-CAM: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017.
- [33] C. Shorten and T. M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):1–48, 2019.
- [34] D. P. Sullivan, C. F. Winsnes, L. Åkesson, M. Hjelmare, M. Wiking, R. Schutten, L. Campbell, H. Leifsson, S. Rhodes, A. Nordgren, K. Smith, B. Revaz, B. Finnbogason, A. Szantner, and E. Lundberg. Deep learning is combined with massive-scale citizen science to improve large-scale image classification. *Nature Biotechnology*, 36(9):820–828, 2018.
- [35] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*, volume 31, 2017.
- [36] P. J. Thul, L. Åkesson, M. Wiking, D. Mahdessian, A. Geladaki, H. Ait Blal, T. Alm, A. Asplund, L. Björk, L. M. Breckels, A. Bäckström, F. Danielsson, L. Fagerberg, J. Fall, L. Gatto, C. Gnann, S. Hober, M. Hjelmare, F. Johansson, S. Lee, C. Lindskog, J. Mulder, C. M. Mulvey, P. Nilsson, P. Oksvold, J. Rockberg, R. Schutten, J. M. Schwenk, Sivertsson, E. Sjöstedt, M. Skogs, C. Stadler, D. P. Sullivan, H. Tegel, C. Winsnes, C. Zhang, M. Zwahlen, A. Mardinoglu, F. Pontén, K. von Feilitzen, K. S. Lilley,

- M. Uhlén, and E. Lundberg. A subcellular map of the human proteome. *Science*, 356(6340):eaal3321, 2017.
- [37] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, and K. J. Lang. Phoneme recognition using time-delay neural networks. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(3):328–339, 1989.
- [38] J. Wang and L. Perez. The effectiveness of data augmentation in image classification using deep learning, 2017. Disponível em: <https://arxiv.org/abs/1712.04621>. Acesso em: 12/09/2024.
- [39] X. Wang, J. Feng, B. Hu, Q. Ding, L. Ran, X. Chen, and W. Liu. Weakly-supervised instance segmentation via class-agnostic learning with salient images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10225–10235, 2021.
- [40] S. Xiang, Q. Liang, Y. Hu, P. Tang, G. Coppola, D. Zhang, and W. Sun. AMC-Net: Asymmetric and multi-scale convolutional neural network for multi-label HPA classification. *Computer Methods and Programs in Biomedicine*, 178:275–287, 2019.
- [41] B. Zhou, A. Khosla, A. Lapedriza, A. Oliva, and A. Torralba. Learning deep features for discriminative localization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2921–2929, 2016.
- [42] Z.-H. Zhou. A brief introduction to weakly supervised learning. *National Science Review*, 5(1):44–53, 2018.