

INSTITUTO FEDERAL DE SANTA CATARINA

ANTHONY CRUZ

APLICANDO DEEP LEARNING EM EXAMES LABORATORIAIS DE
SANGUE

Desenvolvimento de Laudos e Hemogramas

Caçador - SC

24 de Março de 2021

ANTHONY CRUZ

APLICANDO DEEP LEARNING EM EXAMES LABORATORIAIS DE SANGUE

Desenvolvimento de Laudos e Hemogramas

Projeto de Pesquisa apresentado à Coordenadoria do Curso de Sistemas de Informação do Câmpus Caçador do Instituto Federal de Santa Catarina para a avaliar a possibilidade de continuidade do Trabalho de Conclusão de Curso.

Orientador: Professor Samuel da Silva Feitosa

Coorientador: Professor Cristiano Mesquita Garcia

Caçador - SC

24 de Março de 2021

ANTHONY CRUZ

APLICANDO DEEP LEARNING EM EXAMES LABORATORIAIS DE SANGUE
DESENVOLVIMENTO DE LAUDOS E HEMOGRAMAS

Este projeto foi julgado adequado para continuidade do Trabalho de Conclusão do Curso de Sistemas de Informação, pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, e aprovado na sua forma final pela comissão avaliadora abaixo indicada.

Caçador - SC, 01 de dezembro de 2020.

Professor Samuel da Silva Feitosa, Dr.
Orientador
Instituto Federal de Santa Catarina

Professor Cristiano Mesquita Garcia, Dr.
Coorientador
Instituto Federal de Santa Catarina

Professor Membro 1, Me.
Banca Avaliadora
Instituto Federal de Santa Catarina

Professor Membro 2, Dr.
Banca Avaliadora
Instituto Federal de Santa Catarina

RESUMO

Deve-se ressaltar de forma clara e sintética a natureza e o objetivo do trabalho, o método que foi empregado, os resultados e as conclusões mais importantes, seu valor e originalidade. O resumo deve conter apenas um parágrafo com no mínimo 150 e no máximo 500 palavras.

Palavras-chave: mínimo três. máximo cinco. separadas por ponto final e iniciadas com letra maiúscula.

ABSTRACT

This is the english abstract. **Keywords:** latex. abntex. text editoration.

LISTA DE ILUSTRAÇÕES

| | |
|--|----|
| Figura 1 – Glóbulos Vermelhos (RBC) | 13 |
| Figura 2 – Glóbulos Brancos (WBC) | 13 |
| Figura 3 – Plaquetas (Platelets) | 14 |
| Figura 4 – Exemplo de Eritrograma e seus Atributos | 15 |
| Figura 5 – Exemplo de Leucograma e seus Atributos | 16 |

LISTA DE TABELAS

| | |
|---|----|
| Tabela 1 – Palavras-Chave e Sinônimos | 19 |
| Tabela 2 – Bases de Dados e Número de Artigos Seleccionados | 19 |
| Tabela 3 – Critérios de Exclusão | 20 |
| Tabela 4 – Artigos Seleccionados | 21 |

LISTA DE ABREVIATURAS E SIGLAS

SUMÁRIO

| | | |
|------------|--|-----------|
| 1 | INTRODUÇÃO | 9 |
| 1.1 | Problema de Pesquisa | 10 |
| 1.2 | Hipótese de Pesquisa | 10 |
| 1.3 | Objetivos | 10 |
| 1.3.1 | Objetivo Geral | 10 |
| 1.3.2 | Objetivos Específicos | 10 |
| 1.4 | Justificativa | 10 |
| 1.5 | Organização do texto | 11 |
| 2 | FUNDAMENTAÇÃO TEÓRICA | 12 |
| 2.1 | Exames Laboratoriais de Sangue | 12 |
| 2.1.1 | Sangue | 12 |
| 2.1.1.1 | Glóbulos Vermelhos | 12 |
| 2.1.1.2 | Glóbulos Brancos | 13 |
| 2.1.1.3 | Plaquetas | 14 |
| 2.1.2 | Hemograma | 14 |
| 2.1.2.1 | Eritrograma | 15 |
| 2.1.2.2 | Leucograma | 15 |
| 2.2 | Inteligência Artificial: Machine Learning e Deep Learning | 16 |
| 2.2.1 | Machine Learning | 16 |
| 2.2.2 | Deep Learning | 17 |
| 2.2.2.1 | Artificial Neural Network (ANN) | 17 |
| 2.2.2.2 | Arquiteturas de Redes Neurais | 18 |
| 2.2.2.2.1 | Deep Neural Networks (DNNs) | 18 |
| 2.2.2.2.2 | Convolutional Neural Networks (CNNs) | 18 |
| 2.2.2.2.3 | Recurrent Neural Networks (RNNs) | 18 |
| 2.2.2.3 | Bibliotecas e Recursos | 18 |
| 3 | ESTADO DA ARTE DA ÁREA PESQUISADA | 19 |
| 3.1 | Mapeamento Sistemático da Literatura | 19 |
| 3.1.1 | Critérios de Exclusão | 20 |
| 3.1.2 | Critérios de Inclusão | 20 |
| 3.2 | Análise dos trabalhos selecionados | 20 |
| 4 | PROCEDIMENTOS METODOLÓGICOS | 22 |
| 4.1 | Recursos | 22 |
| 5 | CRONOGRAMA | 23 |
| 6 | CONSIDERAÇÕES FINAIS | 24 |
| | REFERÊNCIAS | 25 |

1 INTRODUÇÃO

A saúde humana sempre foi uma área pilar de toda a sociedade e vem se tornando ainda mais vital para sustentar as demais. Levando em consideração os problemas e situações advindos da pandemia de COVID-19, é necessário pensar em formas de automatizar e auxiliar os profissionais de saúde em suas tarefas, para que consigam focar em problemas mais graves e urgentes. Também com o avanço da tecnologia e dos meios de comunicação, a automação vem se fazendo presente na vida de todos e cada vez mais se torna indispensável nas mais diversas áreas. Para a área da saúde não é diferente, é preciso pensar em formas de, além de automatizar, também facilitar processos cotidianos para assim garantir um foco maior nos problemas mais críticos.

Também como efeito da pandemia, a demanda por exames laboratoriais vem crescendo, e conforme isso acontece, se necessita cada vez mais de profissionais da saúde especializados em atender, analisar e produzir laudos desses exames. Porém nem sempre existe uma equipe suficiente para isso, e então acontece sobrecarga de funções para dar conta dessa demanda.

Esse trabalho tem como principal objetivo buscar maneiras de facilitar e atender a produção de laudos de exames laboratoriais, com um foco em exames de sangue e na produção de hemogramas. De forma que os profissionais da saúde possam utilizar uma ferramenta para auxiliar nesse procedimento. Atualmente, os hemogramas são realizados por máquinas especializadas nessa tarefa e portanto demandam um alto custo financeiro e de manutenção para isso. Esse processo poderia ser facilitado com o uso de algoritmos de *Deep Learning* para a automatização, como forma alternativa ao maquinário especializado.

Os algoritmos de *Deep Learning* (DL) vêm sendo utilizados nas mais diversas áreas, como na medicina (KRITTANAWONG et al., 2019), na economia (AKANBI et al., 2020), nas áreas da educação (OFFIR; LEV; BEZALEL, 2008), no comércio eletrônico (HA; PYO; KIM, 2016) e até em jogos virtuais (GREENGARD, 2017). Portanto, DL vem se tornando cada vez mais uma alternativa à métodos tradicionais de realizar tarefas e automatizar processos. Podem ser encontrados alguns trabalhos também na área da saúde, que utilizam técnicas de *Deep Learning* como forma de auxiliar os profissionais em suas tomadas de decisão (Ravi et al., 2017) (ZHAO et al., 2019).

As técnicas de *Deep Learning* buscam atingir resultados a partir de um grande conjunto de dados. Esses dados devem ser devidamente coletados e adaptados ou seja, pré-processados de forma adequada para a máxima eficiência, dessa forma, um modelo poderá passar por diversas fases de treino, completando o seu treinamento. Com o modelo treinado, pode-se realizar testes com outros dados para obtenção de resultados, que serão pós-processados para uma melhor visualização e apresentados ao profissional da saúde. Todo este processo pode ser chamado de *Knowledge Discovery in Databases* (KDD), que se refere à extração de conhecimento a partir dos dados (FAYYAD; PIATETSKY-SHAPIO; SMYTH, 1996) (FAYYAD; STOLORZ, 1997).

Nesse trabalho, busca-se analisar dados de exames de sangue através de imagens de placas de Petri, que são recipientes cilíndricos utilizados pelos profissionais para cultura de microrganismos e análise de materiais (WEI et al., 2021), de forma a elaborar hemogramas e laudos a partir dessas informações. Para isso serão utilizados *datasets* de imagens, a fim de detectar diferentes tipos de células do sangue e chegar em resultados assertivos e úteis para auxiliar também os profissionais da saúde.

1.1 Problema de Pesquisa

Pensando nas formas e aplicações dos algoritmos de *Deep Learning*, presentes nas mais diversas áreas, como um modelo computacional pode ser utilizado para a interpretação de imagens de amostras de sangue em placas de Petri a fim de auxiliar profissionais de laboratório e da saúde na elaboração de laudos científicos e também na sua tomada de decisão?

1.2 Hipótese de Pesquisa

A hipótese para o problema apresentado é que modelos computacionais podem ser treinados para a interpretação de imagens de amostras de sangue em placas de Petri com grande eficiência em prover informações úteis na elaboração automatizada de laudos científicos para profissionais de laboratório e da saúde.

1.3 Objetivos

1.3.1 Objetivo Geral

Como objetivo geral deste trabalho, deve-se buscar formas de treinamento de um modelo computacional para interpretação de imagens voltado a prover informações úteis sobre hemogramas, possibilitando a geração de laudos científicos automaticamente de forma a auxiliar os profissionais de laboratório e da saúde.

1.3.2 Objetivos Específicos

- Realizar mapeamento sistemático sobre o tema, a fim de identificar as técnicas/algoritmos de *Deep Learning* mais adequados para o reconhecimento de imagens de exames;
- Buscar dados de imagens de amostras de sangue em bases de dados disponíveis e para esta finalidade;
- Realizar o pré-processamento dos dados a fim de padronizar e preparar todo o conjunto para o treinamento do modelo computacional;
- Desenvolver e treinar modelos computacionais de *Deep Learning* a fim de encontrar informações suficientes na análise de amostras de sangue em placas de Petri;
- Desenvolver um protótipo a partir do modelo computacional pronto e treinado;

1.4 Justificativa

Este estudo busca demonstrar uma forma alternativa de análise das amostras de sangue e na elaboração de laudos, portanto seu principal foco é auxiliar os profissionais da saúde. A contribuição desse estudo poderá ajudar profissionais da saúde a serem mais rápidos em suas decisões sem perder a assertividade, de forma a aumentar a eficiência da análise de exames laboratoriais. Principalmente em momentos de crise, onde a área da saúde é bastante afetada, é necessário ter formas alternativas e associativas em tarefas cotidianas e de extrema importância para a continuidade dos trabalhos. Com esse trabalho, estudiosos da área da computação e também da saúde, poderão ter uma visão muito interessante e associativa de ideias, de forma a auxiliar em novas pesquisas e aplicações.

Outra questão bastante relevante, é em relação aos custos associados, devido ao fato de que o maquinário utilizado hoje para a análise desses exames demanda um custo altíssimo para a sua compra

e manutenção. Esse trabalho também possibilitará a análise laboratorial sem a necessidade de compra dessas máquinas caríssimas, de forma a diminuir custos e gastos nesse aspecto.

Embora já existam estudos utilizando *Deep Learning* e também estudos utilizando esses conceitos na área da saúde, esse trabalho tem como principal diferencial trazer a ideia de associar a análise dos modelos de *Deep Learning* com a elaboração de laudos e hemogramas de uma forma automatizada. Logo, se faz necessária a investigação dos conceitos desse trabalho para essa e futuras pesquisas. Este estudo demonstra viabilidade técnica, onde toda a pesquisa e aplicação das definições desse material podem ocorrer durante todo o projeto de trabalho de conclusão de curso. Os livros, artigos e materiais teóricos podem ser providenciados pela instituição e estão disponíveis para o uso.

1.5 Organização do texto

O restante desse trabalho está organizado da seguinte maneira: No [Capítulo 2](#) são apresentados os principais conceitos relacionados a *Deep Learning*, bem como as técnicas estudadas. No [Capítulo 3](#) são apresentados os resultados do mapeamento sistemático da literatura. No [Capítulo 4](#) são discutidos os procedimentos metodológicos e no [Capítulo 5](#) é apresentado o cronograma para desenvolvimento deste projeto. Por fim, no [Capítulo 6](#) são apresentadas as considerações finais acerca deste trabalho.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo serão abordados os principais conceitos de *Deep Learning*, assim como as técnicas e algoritmos aplicados para auxiliar os profissionais de laboratório a realizarem exames de sangue de uma forma automatizada e eficiente. Além disso, será detalhado sobre os hemogramas e diferentes dados coletados através de uma análise de sangue.

2.1 Exames Laboratoriais de Sangue

Os exames laboratoriais de sangue, principalmente se tratando de hemogramas, são um tipo de exame simples porém de extrema importância para a saúde humana. Através desses exames se pode descobrir diversas informações sobre o organismo da pessoa em questão, inclusive detectar doenças e problemas de forma antecipada, como por exemplo para diagnosticar anemia, deficiências nutricionais, parasitas no sangue, doenças virais e autoimunes. Também é possível se identificar infecções, doenças como leucemia, diagnosticar efeitos de medicamentos e também o efeito de vários tipos de estresse sobre o corpo (LOKWANI, 2013; LONGO, 2013).

Um exame poderá ser solicitado por um médico, ou a partir do interesse do próprio paciente, e realizado em um laboratório de confiança, que será responsável por realizar a coleta, encaminhar para a análise específica e retornar o resultado. Todo esse processo é custoso em tempo de espera e também financeiramente, pois o maquinário para esse tipo de atividade é muito caro para aquisição e manutenção.

2.1.1 Sangue

O sangue é um elemento do corpo humano, que circula em estado líquido através de todo o sistema circulatório do organismo, sendo de importância para o funcionamento correto das células através da entrada e saída de substâncias que podem modificar a sua composição (VIVAS, 2017).

Pode ser dividido em duas principais partes, sendo o plasma ou soro, a parte de transporte das substâncias pelo sistema, formado pela ingestão de água e alimentos, essas duas nomenclaturas existem de acordo com o uso ou não de anticoagulantes para a separação do sangue, onde dependendo do objetivo e foco da análise, se pode adotar uma ou outra (VIVAS, 2017).

A segunda parte do sangue, que será objeto de estudo para este trabalho, é a parte celular que contém todas as células presentes no sangue e se classificam como glóbulos vermelhos, glóbulos brancos e plaquetas. Geralmente observa-se a presença de eritrócitos, vários tipos e classes de leucócitos e as plaquetas como um todo, que serão abordados um a um posteriormente (VIVAS, 2017).

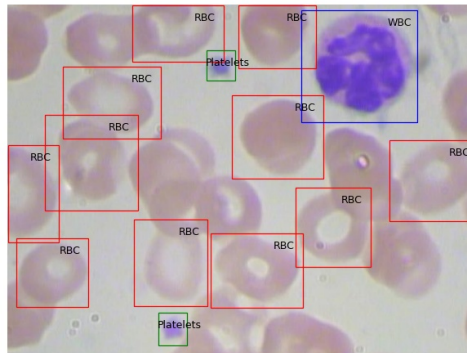
2.1.1.1 Glóbulos Vermelhos

Os glóbulos vermelhos, também conhecidos como *Red Blood Cells (RBC)*, são as hemácias presentes no sangue, também podem ser citadas em exames e registros médicos como eritrócitos, essas células são pequenas e circulares, geralmente em formatos de discos e não possuem núcleo. Estão presentes em grande quantidade, possuindo uma vida útil de aproximadamente 120 dias até que o próprio sistema as elimine (VIVAS, 2017).

É indispensável ao falar sobre a parte vermelha do sangue, citar a hemoglobina que é uma proteína presente nas hemácias e de extrema importância para o funcionamento do sistema, pois através dela é

possível realizar o transporte de oxigênio e gás carbônico pelo sistema sanguíneo, permitindo as trocas gasosas necessárias (VIVAS, 2017).

Figura 1 – Glóbulos Vermelhos (RBC)



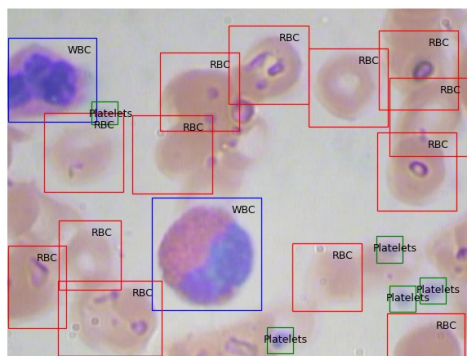
Fonte: (SHENGGAN, 2019)

2.1.1.2 Glóbulos Brancos

Os glóbulos brancos, também conhecidos como *White Blood Cells (WBC)*, são as células brancas do sangue, sendo responsáveis pela defesa do organismo contra as principais ameaças do corpo humano presentes no sistema sanguíneo. Através da fagocitose, que é um processo de englobamento de partículas sólidas pelas células, é realizada ações de defesa contra a invasão de fragmentos estranhos. Os glóbulos brancos são criados na medula óssea e estão presentes em todo o sangue, também em grande quantidade (VIVAS, 2017).

Se faz necessário a classificação dos diferentes tipos de células brancas e de suas importâncias para o sistema de defesa do organismo. É importante frisar que essa classificação se refere aos leucócitos maduros, mas também se pode encontrar presente os leucócitos imaturos (promielócitos, mielócitos, metamielócitos) (VIVAS, 2017).

Figura 2 – Glóbulos Brancos (WBC)



Fonte: (SHENGGAN, 2019)

- **Neutrófilos:** células brancas mais abundantes capazes de entrar nos tecidos, onde conseguem realizar a defesa do organismo, fagocitando partículas estranhas. Essas células são conhecidas como neutrófilos segmentados, pois existe uma célula precursora, que é o bastão, ou também chamado de neutrófilos bastonetes, que possuem essa nomenclatura pois seu núcleo não está amadurecido, ou seja ainda são jovens, e geralmente são identificados quando há infecções em fase aguda.

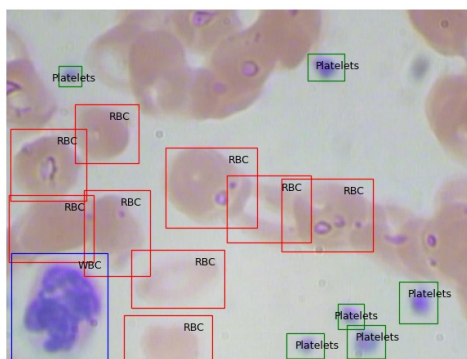
- **Eosinófilos:** células brancas responsáveis na defesa contra parasitas, geralmente estão presentes em grande quantidade no sangue durante reações alérgicas e infestações parasitárias.
- **Basófilos:** células brancas atuantes em respostas alérgicas e na coagulação do sangue. São capazes de liberar histamina, contribuindo para respostas alérgicas ao dilatar e permeabilizar os vasos sanguíneos e também liberam heparina que é capaz de prevenir a coagulação do sangue.
- **Monócitos:** células brancas capazes de entrar no tecido conjuntivo frouxo, onde conseguem se desenvolver em grandes células com grande efeito fagocítico denominadas macrófagos, de forma a ingerir partículas estranhas ao organismo.
- **Linfócitos:** segundo tipo de célula branca mais abundante, são responsáveis e de extrema importância nas respostas imunes específicas do corpo humano, inclusive na produção de anticorpos.

2.1.1.3 Plaquetas

As plaquetas, também conhecidas e citadas como *Platelets*, são os menores componentes do sangue e possuem grande responsabilidade na hemostasia, que é uma resposta fisiológica para a prevenção e interrupção de sangramentos e hemorragias, ou seja, elas atuam na manutenção dos vasos sanguíneos. As plaquetas são fragmentos do citoplasma de megacariócitos, ou seja, elas são produzidas na medula óssea como parte dessas células especializadas que irão se dividir posteriormente e gerar um grande número de plaquetas. Aproximadamente, para cada 1 megacariócito, se pode produzir cerca de 4000 plaquetas (LOKWANI, 2013).

Devido ao fato de serem fragmentos de uma célula, as plaquetas não possuem núcleo e são muito pequenas, com aproximadamente de 1–3 μm de diâmetro, com a coloração azul-acinzentado. A vida útil das plaquetas dura em média de 9 a 12 dias, e elas são removidas pelo baço quando estão velhas ou danificadas (LOKWANI, 2013).

Figura 3 – Plaquetas (Platelets)



Fonte: (SHENGGAN, 2019)

2.1.2 Hemograma

Um hemograma, também conhecido e citado como *Complete Blood Count (CBC)*, é um exame bastante comum e muito utilizado, onde se realiza uma análise de sangue que envolve a contagem das diferentes células sanguíneas. A partir dos números obtidos através dessa contagem e com a comparação desse valor com as faixas de normalidade, é possível chegar a diversas conclusões sobre a saúde do paciente e até mesmo já identificar alguma doença ou problema (VIVAS, 2017; LOKWANI, 2013).

Um hemograma geralmente é realizado em duas principais etapas, sendo a primeira relacionada ao eritrograma que se refere à análise das células vermelhas, de forma a revelar até mesmo alguns tipos essenciais de alterações patológicas do sistema eritropoético, que é o sistema responsável pela produção do material vermelho do sangue, como aumento na produção de glóbulos vermelhos e anemias. A segunda parte está relacionada com o leucograma, que corresponde à contagem global e específica dos leucócitos, a parte branca do sangue. O quadro leucocitário resultante com o exame hematológico, possibilita ao médico tirar importantes conclusões (VIVAS, 2017; LOKWANI, 2013).

2.1.2.1 Eritrograma

O objetivo do eritrograma ao realizar a análise da parte vermelha do sangue, é analisar alguns atributos chave, primeiramente é realizado a contagem geral dos eritrócitos adotando uma escala de milhões/mm³. A hemoglobina também será calculada e registrada em uma escala de g/dl (NAOUM, 2007; VIVAS, 2017).

Depois dessa principal contagem é calculado alguns índices importantes, sendo o primeiro deles o cálculo do volume corpuscular médio (VCM), que é o volume médio das hemácias, calculado pelo quociente de um determinado volume de hemácias pelo número de células contidas no mesmo volume. Outro importante atributo é a hemoglobina corpuscular média (HCM), que semelhante ao VCM, é o conteúdo médio da hemoglobina, calculado pelo quociente de conteúdo de hemoglobina em um determinado volume de hemácias pelo número de células contidas no mesmo volume (NAOUM, 2007; VIVAS, 2017).

Também temos outro índice que é a concentração de hemoglobina corpuscular média (CHCM), sendo a percentagem da hemoglobina em uma amostra de 100ml de hemácias. Por fim temos, a amplitude de distribuição dos glóbulos vermelhos, que em inglês significa *Red Cell Distribution Width (RDW)*, que será responsável por avaliar a variação de tamanho entre as hemácias (NAOUM, 2007; VIVAS, 2017).

Figura 4 – Exemplo de Eritrograma e seus Atributos

| Eritrograma | Valores de referência | | | |
|----------------------|-----------------------|--------|-------|-------------------------|
| Eritrócitos..... | 5,88 | 4,50 a | 5,90 | milhões/mm ³ |
| Hemoglobina..... | 16,6 | 13,5 a | 17,5 | g/dl |
| Volume Globular..... | 49,4 | 41,0 a | 53,0 | % |
| VCM..... | 84,0 | 80,0 a | 100,0 | fl |
| HCM..... | 28,2 | 26,0 a | 34,0 | pg |
| CHCM..... | 33,6 | 31,0 a | 36,0 | g/dl |
| RDW..... | 13,7 | 11,5 a | 15,0 | % |

Fonte: Elaborada pelo autor.

2.1.2.2 Leucograma

O objetivo do leucograma ao realizar a análise da parte branca do sangue, assim como no eritrograma, é analisar alguns atributos chave, porém diferente do processo anterior, essa etapa terá um foco muito maior na classificação e contagem de diferentes células brancas.

Primeiramente é feita uma contagem geral de leucócitos em mm³. Depois é realizada a contagem de forma a classificar cada tipo de leucócito presente, com neutrófilos, eosinófilos, basófilos, linfócitos, monócitos e também os granulócitos imaturos (promielócitos, mielócitos, metamielócitos). Por fim, também é calculado o número presente de plaquetas no sangue em mm³ (NAOUM, 2007; VIVAS, 2017).

Figura 5 – Exemplo de Leucograma e seus Atributos

| Leucograma | | | Valores de referência |
|---------------------------|---------|---------|---|
| Leucócitos..... | 7.520 | | /mm ³ 3.500 a 10.000 /mm ³ |
| Neutrófilos Bastonetes.. | 3 | % 226 | /mm ³ até 840 /mm ³ |
| Neutrófilos Segmentados.. | 43 | % 3.234 | /mm ³ 1.700 a 8.000 /mm ³ |
| Eosinófilos..... | 8 | % 602 | /mm ³ 50 a 500 /mm ³ |
| Basófilos..... | 1 | % 75 | /mm ³ até 100 /mm ³ |
| Linfócitos..... | 35 | % 2.632 | /mm ³ 900 a 2.900 /mm ³ |
| Linfócitos Atípicos..... | 0 | % 0 | /mm ³ |
| Monócitos..... | 10 | % 752 | /mm ³ 300 a 900 /mm ³ |
| Blastos..... | 0 | % 0 | /mm ³ |
| Promielócitos..... | 0 | % 0 | /mm ³ |
| Mielócitos..... | 0 | % 0 | /mm ³ |
| Metamielócitos..... | 0 | % 0 | /mm ³ |
| PLAQUETAS..... | 282.000 | | /mm ³ 150.000 a 450.000 /mm ³ |

Fonte: Elaborada pelo autor.

2.2 Inteligência Artificial: Machine Learning e Deep Learning

A abordagem de Deep Learning que é objeto de estudo desse trabalho é uma das subáreas de Machine Learning que por sua vez é subárea de Inteligência Artificial, portanto antes de abordar cada um desses conceitos, quais suas diferenças e aplicações, se faz necessário explicar e citar sobre a Inteligência Artificial.

A Inteligência Artificial é uma abordagem para a resolução de problemas de diversas naturezas através de uma forma automatizada, ou seja, é uma maneira de se resolver problemas sem a necessidade de um humano ou usuário específico para esse trabalho. Essa área está muito em alta nos dias de hoje, buscando cada vez novas maneiras de automatizar a vida humana. Podemos encontrar essa abordagem em diversos segmentos da indústria e atualmente também vem crescendo o uso dentro das residências para os mais diversos usos.

O intuito dessa área é buscar formas de ensinar máquinas e computadores a serem capazes de ter uma inteligência cada vez mais semelhante aos seres humanos, isso ocorre geralmente através do reconhecimento de padrões por esses dispositivos, onde eles serão ensinados a analisar dados e interpretá-los, muito semelhante ao que acontece para o aprendizado de seres humanos. Porém, diferente de nós, as máquinas geralmente precisam de um volume massivo de dados para que sejam capazes de aprender algo básico.

2.2.1 Machine Learning

A resolução de problemas através de ferramentas da computação é bastante comum e para isso se faz uso de algoritmos programados para essa finalidade específica, porém não é em todos os problemas que podemos aplicar essa abordagem tradicional, devido ao fato de que nem sempre se sabe um caminho único de etapas a serem seguidas para resolução dos problemas. No caso, quando se sabe a entrada de dados e o resultado onde queremos chegar, mas não os meios para se chegar nesse resultado, se deve utilizar um modelo de Machine Learning para realizar a predição. Podemos pensar que na abordagem de algoritmos tradicionais de programação possuímos os parâmetros necessários e conhecemos o método para assim chegar ao resultado, mas na aplicação de machine learning, conhecemos os parâmetros e o resultado, porém o método será aprendido e apresentado pela máquina. (ETHEM, 2016)

Podemos pensar em um modelo de Machine Learning como uma criança que está a aprender algo novo, logo todo modelo passará por três principais etapas, primeiramente o pré-processamento dos dados, depois pela fase de treinamento e por fim será realizado os devidos testes. Dessa forma, um modelo de

Machine Learning deverá ser treinado antes de realizar a predição de qualquer valor ou resultado, onde o responsável pelo treinamento deverá fornecer um grande volume de dados, assim como os resultados esperados por cada um deles. Dessa forma o modelo poderá aprender a entender e interpretar os dados de forma correta, para assim ser capaz de reproduzir esse mecanismo para prever o resultado de futuros dados.

Após o treinamento do modelo como um todo, na fase de teste, para testar e avaliar o treinamento do modelo, também é possível aplicar algumas formas e técnicas para medir o desempenho como um todo, apontando fatores interessantes como o nível de assertividade, a precisão e também o número de acertos e erros relacionados com falsos positivos e falsos negativos através de uma matriz de confusão.

Existem várias abordagens de Machine Learning e elas podem ser classificadas em várias categorias diferentes, porém sempre podemos dividir em duas principais abordagens mais utilizadas, sendo elas a regressão e a classificação. Os modelos de regressão serão responsáveis pela predição de valores reais, enquanto que os modelos de classificação serão responsáveis pela rotulação dos dados em determinadas classes. Além disso, também podemos classificar os algoritmos em supervisionados ou não supervisionados. No aprendizado supervisionado, sabemos o resultado correto, ou seja, onde o modelo deverá chegar a partir dos dados obtidos, porém no aprendizado não supervisionado, o modelo terá que lidar com dados não estruturados e sem um resultado claro. (RASCHKA, 2015)

2.2.2 Deep Learning

Existe alguns casos que o Machine Learning não é suficiente para o aprendizado a partir dos dados, isso ocorre pois o aprendizado acontece através de reconhecimento de padrões com base nos dados e esses dados não podem ser utilizados de qualquer forma, devem ser preparados e adaptados para cada modelo, ou seja, a máquina não será capaz de aprender por conta própria pois sempre precisará de intervenção humana para o processamento dos dados. Logo se necessita de uma abordagem muito mais parecida com a forma de pensar dos seres humanos, como o Deep Learning. (ZOCCA et al., 2017; GIANCARLO; REZAUL, 2018)

O Deep Learning busca uma abordagem bastante parecida com o cérebro humano, isso acontece pois utiliza uma estrutura composta de perceptrons, que são uma versão computadorizada de algo parecido com um neurônio humano. O perceptron será capaz de receber inputs (os dados de entrada) e reproduzir uma saída esperada, porém ao se trabalhar com apenas um perceptron, muitas vezes não possui o processamento necessário para se chegar a uma solução, portanto é necessário utilizar um número maior de perceptrons conectados uns aos outros se assemelhando ainda mais ao cérebro humano, para essa estrutura se dá o nome de rede neural artificial ou como é conhecida, Artificial Neural Network (ANN). (SUGOMORI et al., 2017; GIANCARLO; REZAUL, 2018)

2.2.2.1 Artificial Neural Network (ANN)

Uma Artificial Neural Network é uma estrutura formada por diversos neurônios, conhecidos como perceptrons, organizados de forma a imitar o processamento de um cérebro humano. Sua estrutura é definida em diversas camadas, geralmente adotando uma camada de entrada, outra de saída e camadas intermediárias chamadas de camadas ocultas que serão responsáveis pelo processamento como um todo, onde cada uma delas podem conter um número variável de perceptrons. (GIANCARLO; REZAUL, 2018)

É importante ressaltar que além das informações repassadas de um perceptron para o outro, existe um outro fator importantíssimo para o seu funcionamento, que são os pesos adotados. Os pesos são valores definidos para cada informação recebida pelo perceptron e podemos pensar no seu uso como

um critério de prioridade entre uma informação e outra, que serão ajustados para alcançar o resultado desejado. (GIANCARLO; REZAUL, 2018)

Dessa forma, uma ANN poderá aprender a realizar diversos tipos de tarefas, como classificação por exemplo, através do uso de um algoritmo de backpropagation que significa propagação regressiva, responsável por gerir a taxa de aprendizado do modelo, seguindo alguns passos. (GIANCARLO; REZAUL, 2018)

Primeiramente, o algoritmo inicializará a rede artificial com pesos aleatórios, entrando em modo de treino para realizar o aprendizado. Durante o treinamento, o algoritmo será capaz de realizar previsões de resultados que serão comparados com os valores corretos, para saber se está obtendo sucesso ou não. O algoritmo de backpropagation então irá calcular a diferença do resultado obtido com o real, e essa informação do erro será repassada para todas as camadas anteriores de forma a realizar um ajuste nos pesos e minimizar o erro. (GIANCARLO; REZAUL, 2018)

Todo esse processo é realizado muitas vezes durante o treinamento, e só irá encerrar quando os ajustes levarem ao aumento do erro novamente, indicando que foi os pesos foram ajustados até o limite. Essa etapa de otimização dos pesos utilizados pela rede é muito importante de forma a ser essencial para o funcionamento da rede neural, pois através dela que é possível para o modelo aprender com os seus erros. (GIANCARLO; REZAUL, 2018)

2.2.2.2 Arquiteturas de Redes Neurais

2.2.2.2.1 Deep Neural Networks (DNNs)

As Deep Neural Networks, ou redes neurais profundas, são uma arquitetura de rede neural que são extremamente orientadas ao Deep Learning, ou seja, embora possuam uma arquitetura semelhante às redes neurais tradicionais, são muito mais complexas em seus modelos, com um número maior de neurônios, camadas ocultas e conexões entre neurônios. Além do fato de utilizarem os princípios de aprendizagem do Machine Learning tradicional, como o aprendizado supervisionado, como visto anteriormente.

2.2.2.2.2 Convolutional Neural Networks (CNNs)

teste

2.2.2.2.3 Recurrent Neural Networks (RNNs)

teste

2.2.2.3 Bibliotecas e Recursos

3 ESTADO DA ARTE DA ÁREA PESQUISADA

O processo de pesquisa e seleção dos trabalhos relacionados, foi realizado com base em um mapeamento sistemático sobre as pesquisas com propostas para agilizar a identificação e interpretação de análises de sangue. Esta revisão resultou na identificação e seleção dos principais trabalhos de pesquisa no tema deste Projeto de Trabalho de Conclusão de Curso. Outro objetivo deste mapeamento sistemático foi verificar os métodos utilizados para a aplicação de Deep Learning em imagens de sangue em placas de petri de maneira que possam ser aplicados neste projeto de forma satisfatória.

3.1 Mapeamento Sistemático da Literatura

O mapeamento sistemático da literatura é realizado com base na busca e levantamento de artigos, para isso se utiliza uma string de busca para as principais bibliotecas e repositórios de artigos. Esses artigos serão analisados e selecionados conforme a sua área de pesquisa e a sua temática, para inclusão nesse estudo. Para isso, se é utilizado uma ferramenta para automatização dessa tarefa, que é o Parsifal¹, de modo a definir a string de busca, salvar os artigos necessários e realizar a seleção.

As questões de pesquisas levantadas para isso foram, “Como os algoritmos de Deep Learning podem ser utilizados para a interpretação de exames?” e “Como realizar o tratamento de imagens para reconhecimento por modelos de Deep Learning?”. A partir dessas questões se foram extraídas palavras e termos para o direcionamento da pesquisa. Podemos visualizar estas palavras com seus sinônimos na Tabela 1.

Tabela 1 – Palavras-Chave e Sinônimos

| Palavra-Chave | Sinônimos |
|----------------|--|
| Blood Analysis | Blood Sample |
| Classification | Interpretation, Recognition |
| Deep Learning | Artificial Intelligence, Computer Vision, Machine Learning |

Fonte: Elaborada pelo autor.

Na Tabela 2, é listado as bases de dados onde os artigos foram coletados, a quantidade de cada um de les e a string de busca utilizada na seleção. A mesma string de busca foi utilizado nas três bases de dados, e os artigos encontrados foram dos últimos 5 anos.

Tabela 2 – Bases de Dados e Número de Artigos Selecionados

| Base de Dados | Artigos | String de Busca |
|----------------------|---------|---|
| ACM Digital Library | 37 | (“classification” OR “interpretation” OR “recognition”) AND (“deep learning” OR “artificial intelligence” OR “computer vision” OR “machine learning”) AND (“blood analysis” OR “blood sample”) |
| IEEE Digital Library | 13 | |
| Scopus | 114 | |

Fonte: Elaborada pelo autor.

¹ <https://parsif.al/>

3.1.1 Critérios de Exclusão

Os artigos coletados na pesquisa através da string de busca, passaram por critérios de exclusão por não se adequarem a esta pesquisa, esses critérios podem ser observados na Tabela 3.

Tabela 3 – Critérios de Exclusão

| Critério de Exclusão | Nº de Artigos Recusados |
|--|--------------------------------|
| O estudo não faz parte da área de pesquisa | 101 |
| O estudo apresenta resultados fora da computação | 29 |
| O estudo não é um estudo primário | 6 |
| O estudo é duplicado | 16 |

Fonte: Elaborada pelo autor.

A seleção inciou com 164 artigos no total das três bases de dados buscadas. Com a aplicação dos critérios de exclusão, observa-se um resultante de apenas 14 artigos. Isso ocorreu pois 101 artigos foram eliminados no critério “O estudo não faz parte da área de pesquisa”, que significa que esses artigos tinham alguma relação, porém eram voltados a outras áreas. Outros 29 artigos foram eliminados no critério “O estudo apresenta resultados fora da computação”, que significa que eram da área de pesquisa, porém com resultados e métodos sem conexão com a computação. Foram também encontrados 6 artigos, que entraram no critério “O estudo não é um estudo primário”, o que indica que o artigo pode ser uma revisão sistemática da literatura ou semelhante. Por fim, foram eliminados outros 16 artigos por serem duplicados.

3.1.2 Critérios de Inclusão

Os seguintes critérios de inclusão foram definidos:

- Nova tecnologia para análise de sangue;
- Processo, método ou técnica para contagem de células sanguíneas;
- Sistema para elaboração de hemogramas utilizando Deep Learning;

Na tabela 4, podemos encontrar todos os 14 artigos selecionados com base nos critérios de inclusão, todos eles se enquadram em pelo menos um deles.

Todos os artigos selecionados estão relacionados à maneiras e recursos para auxiliar na interpretação de exames de sangue utilizando conceitos de Deep Learning e Machine Learning.

3.2 Análise dos trabalhos selecionados

Por fim, com os artigos selecionados e classificados, é necessário realizar a extração dos dados desses trabalhos, sendo essa a última etapa desse mapeamento sistemático da literatura. É possível perceber que os algoritmos e abordagens mais utilizados são técnicas de *Deep Learning*, como por exemplo, o uso de *Convolutional Neural Network (CNN)* (A1, A2, A3, A4, A5, A6, A8, A9, A10, A11) e de *Recurrent Neural Network (RNN)* (A6, A11), que são abordagens de redes neurais para a classificação das células sanguíneas.

Outros trabalhos utilizam de algoritmos de *Machine Learning* tradicionais para a classificação, como por exemplo, ocorre com o uso de *Random Forest* ou *Decision Trees* (A2, A4, A7), que são estruturas de árvores de decisão. Também se encontra estudos fazendo uso de *Support Vector Machine (SVM)*

Tabela 4 – Artigos Seleccionados

| ID | Título do Artigo | |
|-----|--|---|
| A1 | Analyzing microscopic images of peripheral blood smear using deep learning | Mundhra, D. and Cheluvaraju, B. and Rampure, J. and Rai Dastidar, T. |
| A2 | Automatic detection and classification of leukocytes using convolutional neural networks | Zhao, J. and Zhang, M. and Zhou, Z. and Chu, J. and Cao, F. |
| A3 | Automatic white blood cell classification using pre-trained deep learning models: ResNet and Inception | Habibzadeh, M. and Jannesari, M. and Rezaei, Z. and Baharvand, H. and Totonchi, M. |
| A4 | Classification of Human White Blood Cells Using Machine Learning for Stain-Free Imaging Flow Cytometry | Lippeveld, M. and Knill, C. and Ladlow, E. and Fuller, A. and Michaelis, L.J. and Saeys, Y. and Filby, A. and Peralta, D. |
| A5 | Blood cell classification using the hough transform and convolutional neural networks | Molina-Cabello, M.A. and López-Rubio, E. and Luque-Baena, R.M. and Rodríguez-Espinosa, M.J. and Thurnhofer-Hemsi, K. |
| A6 | White Blood Cells Image Classification Using Deep Learning with Canonical Correlation Analysis | Patil, A.M. and Patil, M.D. and Birajdar, G.K. |
| A7 | Image processing and machine learning in the morphological analysis of blood cells | Rodellar, J. and Alférez, S. and Acevedo, A. and Molina, A. and Merino, A. |
| A8 | Improving blood cells classification in peripheral blood smears using enhanced incremental training | Al-qudah, R. and Suen, C.Y. |
| A9 | Corruption-Robust Enhancement of Deep Neural Networks for Classification of Peripheral Blood Smear Images | Zhang, S. and Ni, Q. and Li, B. and Jiang, S. and Cai, W. and Chen, H. and Luo, L. |
| A10 | Convolutional neural network and decision support in medical imaging: Case study of the recognition of blood cell subtypes | Diouf, D. and Seck, D. and Diop, M. and Ba, A. |
| A11 | Combining Convolutional Neural Network With Recursive Neural Network for Blood Cell Image Classification | Liang, G. and Hong, H. and Xie, W. and Zheng, L. |
| A12 | Blood diseases detection using classical machine learning algorithms | Alsheref, F.K. and Gomaa, W.H. |

Fonte: Elaborada pelo autor.

(A7) que utilizam vetores de suporte e por fim *K-Means* e *K-Nearest Neighbors (KNN)* (A12), que faz a classificação levando em conta os vizinhos mais próximos.

4 PROCEDIMENTOS METODOLÓGICOS

4.1 Recursos

5 CRONOGRAMA

As atividades propostas neste cronograma podem sofrer leves alterações no decorrer do seu desenvolvimento de acordo com a necessidade.

A forma mais fácil de criar tabelas é através de ferramentas gráficas. Geralmente utiliza-se o site [<https://www.tablesgenerator.com/>](https://www.tablesgenerator.com/) para realizar tal atividade, exportando o código LaTeX e colando na parte do texto que ela deve aparecer.

6 CONSIDERAÇÕES FINAIS

Apresentar as considerações finais do projeto de TCC.

REFERÊNCIAS

- AKANBI, L. A. et al. Deep learning model for demolition waste prediction in a circular economy. *Journal of Cleaner Production*, v. 274, p. 122843, 2020. ISSN 0959-6526. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0959652620328882>>. Citado na página 9.
- ETHEM, A. *Machine Learning : The New AI*. The MIT Press, 2016. (MIT Press Essential Knowledge Series). ISBN 9780262529518. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1369420&lang=pt-br&site=ehost-live>>. Citado na página 16.
- FAYYAD, U.; PIATETSKY-SHAPIO, G.; SMYTH, P. The kdd process for extracting useful knowledge from volumes of data. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 39, n. 11, p. 27–34, nov. 1996. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/240455.240464>>. Citado na página 9.
- FAYYAD, U.; STOLORZ, P. Data mining and kdd: Promise and challenges. *Future Generation Computer Systems*, v. 13, n. 2, p. 99–115, 1997. ISSN 0167-739X. Data Mining. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0167739X97000150>>. Citado na página 9.
- GIANCARLO, Z.; REZAUL, K. M. *Deep Learning with TensorFlow : Explore Neural Networks and Build Intelligent Systems with Python, 2nd Edition*. Packt Publishing, 2018. v. 2nd ed. ISBN 9781788831109. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1789473&lang=pt-br&site=ehost-live>>. Citado 2 vezes nas páginas 17 e 18.
- GREENGARD, S. Gaming machine learning. *Commun. ACM*, Association for Computing Machinery, New York, NY, USA, v. 60, n. 12, p. 14–16, nov. 2017. ISSN 0001-0782. Disponível em: <<https://doi.org/10.1145/3148817>>. Citado na página 9.
- HA, J.-W.; PYO, H.; KIM, J. Large-scale item categorization in e-commerce using multiple recurrent neural networks. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. (KDD '16), p. 107–115. ISBN 9781450342322. Disponível em: <<https://doi.org/10.1145/2939672.2939678>>. Citado na página 9.
- KRITTANAWONG, C. et al. Deep learning for cardiovascular medicine: a practical primer. *European Heart Journal*, v. 40, n. 25, p. 2058–2073, 02 2019. ISSN 0195-668X. Disponível em: <<https://doi.org/10.1093/eurheartj/ehz056>>. Citado na página 9.
- LOKWANI, D. *The ABC of CBC: Interpretation of complete blood count and histograms*. [S.l.]: Jaypee, 2013. Citado 3 vezes nas páginas 12, 14 e 15.
- LONGO, D. L. *Atlas de Hematologia e Análise de Esfregaços do Sangue Periférico*. [S.l.: s.n.], 2013. Citado na página 12.
- NAOUM, F. A. N. P. C. Interpretação laboratorial do hemograma. 2007. Citado na página 15.
- OFFIR, B.; LEV, Y.; BEZALEL, R. Surface and deep learning processes in distance education: Synchronous versus asynchronous systems. *Computers and Education*, v. 51, n. 3, p. 1172–1183, 2008. ISSN 0360-1315. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0360131507001406>>. Citado na página 9.
- RASCHKA, S. *Python Machine Learning*. Packt Publishing, 2015. (Community Experience Distilled). ISBN 9781783555130. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1071004&lang=pt-br&site=ehost-live>>. Citado na página 17.
- Ravi, D. et al. Deep learning for health informatics. *IEEE Journal of Biomedical and Health Informatics*, v. 21, n. 1, p. 4–21, 2017. Citado na página 9.
- SHENGGAN. *BCCD Dataset*. 2019. Citado 2 vezes nas páginas 13 e 14.

- SUGOMORI, Y. et al. *Deep Learning: Practical Neural Networks with Java*. Packt Publishing, 2017. ISBN 9781788470315. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1532297&lang=pt-br&site=ehost-live>>. Citado na página 17.
- VIVAS, W. L. P. *Manual Prático de Hematologia*. [S.l.: s.n.], 2017. Citado 4 vezes nas páginas 12, 13, 14 e 15.
- WEI, J. et al. *A Petri Dish for Histopathology Image Analysis*. 2021. Citado na página 9.
- ZHAO, R. et al. Deep learning and its applications to machine health monitoring. *Mechanical Systems and Signal Processing*, v. 115, p. 213–237, 2019. ISSN 0888-3270. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0888327018303108>>. Citado na página 9.
- ZOCCA, V. et al. *Python Deep Learning*. Packt Publishing, 2017. ISBN 9781786464453. Disponível em: <<http://search.ebscohost.com/login.aspx?direct=true&db=nlebk&AN=1513367&lang=pt-br&site=ehost-live>>. Citado na página 17.