

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 – Exemplo de uso de imagens no \LaTeX 12

LISTA DE TABELAS

Tabela 1 – Cronograma das atividades previstas.	16
---	----

LISTA DE ABREVIATURAS E SIGLAS

ABNT Associação Brasileira de Normas Técnicas

abnTeX ABsurdas Normas para TeX

AC Autoridade Certificadora

AES *Advanced Encryption Standard*

TLS *Transport Layer Security*

TPC Terceira Parte Confiável

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	Conceito 1	12
2.2	Conceito 2	12
2.3	Conceito 3	12
3	ESTADO DA ARTE DA ÁREA PESQUISADA	13
3.1	Mapeamento Sistemático da Literatura	13
3.1.1	Critérios de Exclusão	14
3.1.2	Critérios de Inclusão	14
3.2	Análise dos trabalhos selecionados	14
4	PROCEDIMENTOS METODOLÓGICOS	15
4.1	Recursos	15
5	CRONOGRAMA	16
6	CONSIDERAÇÕES FINAIS	17
	REFERÊNCIAS	18

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 apresentados os principais tópicos relacionados ao <Assunto Estudado>, seu conceito e seus impactos na sociedade, bem como as motivações para suas publicações e formas de identificá-las. Além disso, serão abordadas técnicas que permitem <Descrever as técnicas utilizadas>, que serão aplicados para <Tema Proposto>.

2.1 Conceito 1

Abaixo é apresentada uma figura com o logotipo do Instituto Federal de Santa Catarina. Para inserir uma figura usando o LaTeX, utilizamos a diretiva *figure*. Normalmente referenciamos a figura a partir do seu label, conforme segue. A Figura 1 mostra o exemplo de uso de imagens no L^AT_EX.

Figura 1 – Exemplo de uso de imagens no L^AT_EX.



Fonte: Elaborada pelo autor.

Observe todos os detalhes utilizados. A diretiva *centering* é utilizada para deixar a imagem centralizada. A diretiva *caption* é utilizada para adicionar a legenda na parte superior da imagem. A diretiva *includegraphics* serve para adicionar a imagem propriamente dita, estando neste caso, localizada dentro da pasta *img*. Na mesma diretiva, é possível notar o código `width=0.40`, que significa que a imagem vai utilizar 40% da largura do texto. Por fim, a diretiva *legend* é utilizada para indicar a fonte da imagem, e a diretiva *label* para criar uma referência.

2.2 Conceito 2

2.3 Conceito 3

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 - Tabela com 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.

¹ <https://parsif.al/>

Tabela 2 - Bases de Dados e Quantidade de Artigos Seleccionados

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

3.1.1 Critérios de Exclusão

3.1.2 Critérios de Inclusão

3.2 Análise dos trabalhos seleccionados

4 PROCEDIMENTOS METODOLÓGICOS

4.1 Recursos

5 CRONOGRAMA

A Tabela 1 apresenta o cronograma de atividades propostas para o desenvolvimento deste projeto de trabalho de conclusão de curso, de forma a viabilizar <Falar sobre o que se pretende atingir com o projeto>.

Tabela 1 – Cronograma das atividades previstas.

Etapa	Meses									
	Fev	Mar	Abr	Mai	Jun	Ago	Set	Out	Nov	Dez
Fundamentação Teórica	X	X								
Mapeamento Sistemático da Literatura			X	X						
Escrita do Projeto de TCC e Defesa			X	X	X					
Atividade a ser desenvolvida 1						X				
Atividade a ser desenvolvida 2							X			
Atividade a ser desenvolvida 3							X	X		
Verificação de Aceitação dos Resultados								X		
Comparação dos Resultados com a Literatura								X	X	
Exposição dos Resultados									X	
Escrita do TCC									X	X
Defesa do TCC										X

Fonte: Elaborada pelo autor.

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/>> para realizar tal atividade, exportando o código LaTeX e colando na parte do texto que ela deve aparecer (TABLESGENERATOR.COM, 2021).

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.
- 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.
- 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.
- 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.
- 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.
- TABLESGENERATOR.COM. *Create LaTeX tables online - TablesGenerator.com*. 2021. Disponível em: <https://www.tablesgenerator.com/>. Acesso em: 09 mar 2021. Citado na página 16.
- 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.