



ALISSON ALLEBRANDT
DIEGO HENRIQUE SCHMIDT

**SIMPLIFICANDO A INTERPRETAÇÃO DE ANÁLISE DE SOLO COM
INTELIGÊNCIA ARTIFICIAL**

Três de Maio
2019

**ALISSON ALLEBRANDT
DIEGO HENRIQUE SCHMIDT**

**SIMPLIFICANDO A INTERPRETAÇÃO DE ANÁLISE DE SOLO COM
INTELIGÊNCIA ARTIFICIAL**

Trabalho de Conclusão do Curso
Bacharelado em Sistemas de Infor-
mação da Faculdade Três de Maio -
SETREM.

Orientador:

Dr. Dalvan Griebler

Co-orientador:

M. Sc Priscila Guarienti

Três de Maio

2019

TERMO DE APROVAÇÃO

ALISSON ALLEBRANDT
DIEGO HENRIQUE SCHMIDT

SIMPLIFICANDO A INTERPRETAÇÃO DE ANÁLISE DE SOLO COM INTELIGÊNCIA ARTIFICIAL

Relatório aprovado como requisito parcial para obtenção do título de **Bacharel em Sistemas de Informação** concedido pela Faculdade de Sistemas de Informação da Sociedade Educacional Três de Maio, pela seguinte Banca examinadora:

Orientador: Prof. Dalvan Jair Griebler, Dr.
Faculdade de Sistemas de Informação da SETREM

Co-orientador: Prof. Priscila Guarienti, M.Sc
Faculdade de Sistemas de Informação da SETREM

Prof. Samuel Camargo de Souza, M.Sc.
Faculdade de Sistemas de Informação - SETREM

Prof. Tiago Luis Cesa Seibel, M.Sc.
Faculdade de Sistemas de Informação - SETREM

Prof. Marcos Caraffa, M.Sc.
Faculdade de Agronomia – SETREM

Profa. Vera Lúcia Lorenset Benedetti, M.Sc.
Coordenação do Curso Bacharelado em Sistemas de Informação
Faculdade de Sistemas de Informação da SETREM.

Três de Maio, 17 de Junho de 2019.

RESUMO

Um dos aspectos que interfere em uma boa produção e safra agrícola é o solo. Desta forma, a sua conservação por meio da aplicação correta de nutrientes e adubação é de suma importância. Com base neste cenário e com o crescimento acelerado da tecnologia agrícola, propõe-se um aplicativo capaz de interpretar as análises de solo que são geradas pelos laboratórios de solo resultante de uma amostra de terra que é colhida pelo produtor rural. Após a análise, sugerir a quantidade adequada de fertilizantes e nutrientes agrícolas que o produtor deverá aplicar em sua lavoura. Atualmente, este processo de recomendação ainda é feito de maneira manual pelos engenheiros agrônomos ou por softwares Desktops que se utilizam de manuais de calagem e adubação como base para os cálculos de recomendação. Visando mudar esse cenário, buscou-se utilizar a tecnologia de aprendizado de máquina, no qual foram extraídos mais de 30 mil registros de análises de solos do laboratório de solos da SETREM. Com base no estudo e análise desses dados, foi proposto uma arquitetura de solução, visando a criação de um modelo de treino e uma forma genérica de receber análises de solo, normalizar as mesmas, interpretar, gerar recomendações, e salvar em um banco de dados único, para que esses dados possam ser utilizados futuramente para utilização em BI e mineração. Diante da arquitetura de solução proposta, foi desenvolvido um aplicativo mobile capaz de interpretar uma foto retirada da análise de solo e então transformar os valores dos elementos químicos presentes na imagem em informações digitais que poderão ser consultadas e compartilhadas de uma forma mais rápida entre as pessoas interessadas. Além disso, iniciar com o processo de alimentação de um banco de dados com informações sobre análises de solo. Obteve-se resultados relativos a necessidade atual de ainda utilizar manuais de calagem e de adubação, bem como a aplicação de inteligência artificial diante dessa área. Por fim, foram estudadas ferramentas de processamento de imagens e interpretação de caracteres com o uso de *Machine Learning* como o Tesseract OCR e Google Vision, tendo como resultado uma comparação entre as duas ferramentas de interpretação testadas.

Palavras-chave: Sistemas de Informação, Aprendizado de máquina, Recomendação de nutrientes, Processamento de imagens, Reconhecimento de caracteres.

ABSTRACT

One of the aspects that interferes in a good production and agricultural crop is the soil. Its conservation through the correct application of nutrients and fertilization is of paramount importance. Based on this scenario and with the accelerated growth of agricultural technology, an application capable of interpreting the soil analyzes that are generated by soil laboratories resulting from a sample of land collected by the rural producer is proposed. After the analysis, the idea is to suggest the appropriate amount of fertilizers and agricultural nutrients that the producer should apply in his crop. Currently this recommendation process is still manually done by agronomists or desktop software that uses liming and fertilization manuals as the basis for recommendation calculations. This work aims to use machine learning technology, in which more than 30,000 soil analysis records were extracted from the SETREM soil laboratory. Based on the study and analysis of these data, a solution was proposed aimed at creating a training model and a generic way to receive soil analyzes, normalize them, interpret, generate recommendations, and save to a single database so that this data can be used for BI and mining in the future. Therefore, it was developed a mobile application capable of interpreting a photo taken from the soil analysis and then transform the values of the chemical elements present in the image into digital information, which can be consulted and shared in a faster way among the interested people, in addition to starting with the process of feeding a database with information on soil analysis. We obtained results regarding the current need to still use manuals of liming and fertilization as well as the application of artificial intelligence in front of this area. Also, we studied tools of image processing and interpretation of characters with the use of Machine Learning, such as Tesseract OCR and Google Vision, which resulted in a comparison of the two interpretation tools tested.

Keywords: Information systems, Machine Learning, Nutrient recommendation, Image processing, Character Recognition.

LISTA DE FIGURAS

1	Ferramentas para a amostragem do solo	32
2	Análise de solo	40
3	Recomendação de adubação e calagem	41
4	Interpretação de teores de argila e de matéria orgânica e da capacidade de troca de cátions (CTC) a pH 7,0	44
5	Desenvolvimento computacional difuso	49
6	Arquitetura de um sistema especialista	51
7	Modelo básico de um neurônio	53
8	Rede neural com 3 camadas	54
9	Funcionamento de uma rede LSTM	55
10	A interação agente-ambiente na Aprendizagem por Reforço	57
11	Diferenças entre React e React Native	67
12	Atuação do NodeJS	69
13	Laços de evento do Node JS	70
14	Fluxo atual de recomendação de adubação por meio de uma análise de solo	89
15	Exemplo de análise de solo gerada pelo laboratório de solos da Setrem	91
16	Sugestão de fluxo para recomendação de análise de solo	93
17	Passo a passo para criação do processo de recomendação de análise de solo	94
18	Modelo CSV para importação de análises no <i>software</i> CADUB	97
19	Tela de início do CADUB	99
20	Cadastro de análise de solo	100
21	Realização de cálculo de recomendação no CADUB - Selecionar Análise	101

22	Realização de cálculo de recomendação no CADUB – Selecionar grupo de cultura	102
23	Realização de cálculo de recomendação no CADUB – Ignorar cálculo de calagem	103
24	Realização de cálculo de recomendação no CADUB – Selecionar cultura	103
25	Realização de cálculo de recomendação no CADUB – Selecionar fórmula recomendada	104
26	Realização de cálculo de recomendação no CADUB – Concluir recomendação	105
27	Exportação das recomendações - Etapas para geração do arquivo XLS	106
28	Exportação das recomendações - Estrutura gerada em formato XLS	106
29	Fluxo de leitura dos dados de análise de solo.	111
30	Fluxo da arquitetura de uma solução para melhorar o processo de análise e recomendação	112
31	Ações do usuário no aplicativo	114
32	Tela inicial do aplicativo	114
33	Captura de foto para interpretação - alinhamento correto	115
34	Processamento da interpretação	116
35	Detalhes da análise de solo interpretada	116
36	Listagem de análises de solo interpretadas	117
37	Fluxo do algoritmo de interpretação de análise de solo	118
38	Testes com a ferramenta Tesseract OCR	122
39	Resultado do teste com a API do Google Vision	124
40	Json de retorno com o conteúdo dos blocos agrupado	124
41	Antes e depois da correção de níveis da imagem	128
42	Resultado do tratamento de ruído	129
43	Aplicação de alinhamento em análise de solo	130
44	Estrutura JSON retornado pelo algoritmo	133
45	Aplicação do método de erosão e dilatação	134
46	Aplicação de erosão e dilatação Eixo X vs Eixo Y	135
47	Cruzamento das linhas verticais e horizontais	136
48	Separação das tabelas com os elementos químicos	137
49	Extração da linha com os valores dos componentes químicos da tabela	138
50	Testes da implementação com Tesseract OCR e Google Vision . . .	142

51	Formulário utilizado para criação do <i>dataset</i>	144
52	Amostras de análise de solo	145
53	Resultado do algoritmo utilizando Google Vision	149
54	Resultado do algoritmo utilizando Tesseract OCR	150
55	Resultado para recomendação utilizando Google Vision	151
56	Resultado para recomendação utilizando Tesseract OCR	151
57	Acurácia da interpretação das análises de solo	152

LISTA DE QUADROS

1	Trabalhos relacionados.	85
2	Nome das colunas x Significados da base do laboratório de solos SETREM.	95
3	Códigos do elemento x Elemento químico associado	96
4	Valores aceitos para o parâmetro –OEM	120
5	Valores aceitos para o parâmetro –PSM	120
6	Parâmetros de entrada do algoritmo	132
7	Tipos de imagens do <i>dataset</i>	146
8	Especificações da máquina virtual utilizada	147

LISTA DE SIGLAS

AM	Aprendizado de máquina
API	Application Program Interface
APP	Aplicativo
AS	Análise Sumária
BSON	<i>Binary JSON</i>
CTC	Capacidade de troca de cátions
ESALQ	Escola Superior de Agricultura Luiz de Queiroz
GIS	<i>Geographic Information System</i>
GPL	<i>General Public Licence</i>
HTML	<i>Hypertext Markup Language</i>
IA	Inteligência Artificial
IDE	Ambiente de desenvolvimento integrado
IMEA	Instituto Mato Grossense de Economia Aplicada
JSX	<i>XML Like Syntax</i>
MIT	<i>Massachusetts Institute of Technology</i>
LSTM	<i>Long Short Term Memory</i>
NDVI	Índice de Vegetação por Diferença Normalizada
NRS	Núcleo regional Sul
OpenCV	<i>Open Source Computer Vision Library</i>
RDBMS	<i>Relational Database Management Systems</i>
RNA	Redes Neurais Artificiais
ROLAS	Rede Oficial de Laboratórios de Análise de Solo e de Tecido Vegetal dos Estados do Rio Grande do Sul e de Santa Catarina

SEAF	Seguro da agricultura familiar
SETREM	Sociedade Educacional Três de Maio
TAWC	<i>Total Available Water Capacity of the Soil</i>
TCC	Trabalho de conclusão de curso
TI	Tecnologia da Informação
UNSODA	<i>United States Department of Agriculture</i>
VBA	<i>Visual Basic for Applications</i>
ABC	Agricultura de Baixo Carbono
Cu	Cobre
Fe	Ferro
Mn	Manganês
Zn	Zinco
Ca	Cálcio
Mg	Magnésio
K	Potássio
Na	Sódio
P	Fósforo
EMEG	Éter Mono etílico do etileno Glicol
ROLAS	Rede Oficial de Laboratórios de Análise de Solo e de Tecido Vegetal dos Estados do Rio Grande do Sul e de Santa Catarina

SUMÁRIO

INTRODUÇÃO	15
CAPÍTULO 1: CONTEXTUALIZAÇÃO	18
1.1 TEMA	18
1.2 DELIMITAÇÃO DO TEMA	18
1.3 OBJETIVOS	20
1.3.1 Objetivo geral	20
1.3.2 Objetivos específicos	20
1.4 JUSTIFICATIVA	21
1.5 PROBLEMA	23
1.6 HIPÓTESES	23
1.7 METODOLOGIA	24
1.7.1 Métodos de abordagem	25
1.7.2 Métodos de procedimento	26
1.7.2.1 <i>Validação das Hipóteses</i>	26
1.7.3 Técnicas	27
CAPÍTULO 2: FUNDAMENTAÇÃO TEÓRICA	28
2.1 AGRICULTURA	28
2.1.1 Agricultura de precisão	29
2.1.2 Agronegócio	29
2.1.3 Análise de solo	30
2.1.3.1 <i>Amostragem de solo</i>	30
2.1.3.2 <i>Processo de análise de solo</i>	34
2.1.3.3 <i>CADUB</i>	44
2.1.4 Custeio agrícola	45
2.2 INTELIGÊNCIA ARTIFICIAL	46
2.2.1 Teste de Turing	47
2.2.2 Lógica difusa	48
2.2.3 Sistemas especialistas	49
2.2.4 Redes Neurais Artificiais	52
2.2.4.1 <i>Redes Neurais LSTM</i>	54
2.2.5 Aprendizagem de máquina	55
2.2.5.1 <i>Aprendizagem por reforço</i>	56

2.2.5.2	<i>Aprendizagem supervisionada</i>	59
2.2.5.3	<i>Aprendizagem não supervisionada</i>	59
2.3	TECNOLOGIAS	60
2.3.1	Aplicações móveis	60
2.3.1.1	<i>Características de um ambiente mobile</i>	61
2.3.2	Bancos de dados NOSQL	61
2.3.2.1	<i>MongoDB</i>	61
2.3.3	Cloud computing	62
2.3.4	Tratamento de dados e modelagem preditiva	63
2.3.4.1	<i>Weka</i>	63
2.3.5	Desenvolvimento	66
2.3.5.1	<i>React Native</i>	66
2.3.5.2	<i>NodeJS</i>	68
2.3.5.3	<i>Python</i>	70
2.3.6	Automatização de processos	71
2.3.6.1	<i>Autolt</i>	71
2.4	TRABALHOS RELACIONADOS	72
2.4.1	CADUB GHF: um programa computacional para cálculo da quantidade de fertilizantes e corretivos da acidez do solo para culturas produtoras de grãos, hortaliças e forrageiras	72
2.4.2	<i>Large scale data processing in real world: From analytics to predictions</i>	73
2.4.3	<i>Heuristic prediction of rainfall using Machine Learning techniques</i>	74
2.4.4	<i>Artificial calf weaning strategies and the role of machine learning: A review</i>	75
2.4.5	<i>Machine Learning Approaches for Soil Classification in a Multi-agent Deficit Irrigation Control System</i>	75
2.4.6	<i>Crop Selection Method to maximize crop yield rate using machine learning technique</i>	76
2.4.7	<i>Intelligent system for measuring soil compaction on croplands</i>	77
2.4.8	<i>Low-cost multispectral imaging system for crop monitoring</i>	77
2.4.9	<i>Interconnection of a crop growth model with remote sensing data to estimate the total available water capacity of soils</i>	78
2.4.10	<i>Design and development of M3SS: A Soil Sensor Node for precision agriculture</i>	79
2.4.11	<i>Machine Learning for the Management of Agricultural Soil Data</i>	79
2.4.12	<i>Developing fertilizer recommendations for agriculture</i>	80
2.4.13	<i>Using Machine Learning for land suitability classification</i>	81
2.4.14	<i>Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction</i>	81
2.4.15	<i>Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks</i>	82
2.4.16	Table Recognition in Heterogeneous Documents using Machine Learning	83

2.4.17	DeCNT: Deep Deformable CNN for Table Detection	83
2.5	TABELA DE COMPARAÇÃO DE TRABALHO RELACIONADOS	84
CAPÍTULO 3: RESULTADOS OBTIDOS		88
3.1	LABORATÓRIO DE ANÁLISE DE SOLOS	88
3.2	CENÁRIO ATUAL	89
3.3	SOLUÇÃO PROPOSTA	91
3.4	ESTUDO E ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL PARA RECOMENDAÇÃO	93
3.4.1	Extração e tratamento de dados	95
3.4.2	Utilização do software CADUB	96
3.4.2.1	<i>Importação de análises de solo</i>	97
3.4.2.2	<i>Processo de recomendação</i>	100
3.4.2.3	<i>Exportação das Recomendações</i>	105
3.4.3	Avaliação do uso de inteligência artificial	107
3.5	DESENVOLVIMENTO DA SOLUÇÃO	110
3.5.1	Arquitetura da solução	111
3.5.2	Aplicativo mobile	113
3.5.3	Algoritmo para detecção e reconhecimento de análises de solo	118
3.5.3.1	<i>Escolha da ferramenta de interpretação e considerações</i>	118
3.5.3.2	<i>Tratamento da imagem</i>	125
3.5.3.3	<i>Pré-processamento</i>	131
3.5.3.4	<i>Interpretação</i>	138
3.5.4	Api	142
3.6	EXPERIMENTOS	143
3.6.1	Dataset	143
3.6.2	Execução do algoritmo de interpretação de análises de solo	147
3.6.2.1	<i>Validação da interpretação</i>	148
3.6.3	Sugestões para implementação da recomendação	153
CONCLUSÃO		155
REFERÊNCIAS		164

INTRODUÇÃO

A aceleração no crescimento da população mundial juntamente com a importância de uma alimentação saudável irá gerar muitos desafios para o setor agrícola, no qual a segurança por uma produção diversificada e autossuficiente e o interesse cada vez maior pelo consumo de frutas, legumes e hortaliças terá impacto direto nessa situação.

Com o intuito de contribuir com a tecnologia aplicada à agricultura, o (SEBRAE, 2017) elaborou um relatório denominado “O futuro do desenvolvimento tecnológico da agricultura brasileira” onde é exposta uma prospecção do setor para o período de 2014 até 2034. O relatório, além de fortalecer a importância da tecnologia na área, também aborda a abertura de novas vertentes que serão capazes de agregar valor e competitividade no setor agroindustrial.

Vista a grande importância que a tecnologia está tendo e ainda terá na área da agricultura, esse trabalho visa criar um meio para auxiliar os produtores rurais e engenheiros agrônomos, em um dos aspectos mais importantes da produção, o solo, buscando obter uma maior produtividade nas plantações através da adubação correta e do uso de quantidades ideais de nutrientes que deverão ser aplicados na lavoura, por meio do entendimento no processo de análise de solos.

Dentre as bases do trabalho estão os dados de análises de solos já realizadas pelo Laboratório de Solos da SETREM. Posteriormente esta base de dados foi submetida ao *software* CADUB, *software* desenvolvido pela Universidade Federal de Santa Maria (UFSM) e que auxiliou no processo de geração de recomendações

de adubação e calagem para cada análise submetida. Com a conclusão desse processo, tem-se as entradas e saídas necessárias para realizar a criação do modelo de treino.

Diante desta iniciativa se teve como objetivo, inicialmente, avaliar o uso da tecnologia de inteligência artificial nessa área, visto que este tipo de tecnologia ainda é pouco aplicado em análises de solo. Por meio dos estudos realizados e da avaliação se teve então a criação de uma arquitetura de solução, elencando processos manuais que podem ser automatizados, desde a disponibilização das análises de solo pelos laboratórios até uma melhor interpretação dos valores de elementos químicos presentes nas análises.

Com a avaliação de uso da inteligência artificial também verificou-se que não seria possível desenvolver a proposta de se criar um modelo de treino capaz de aprender automaticamente as referências de recomendações de adubação, visto que hoje, esta referência ainda é originada dos manuais de calagem e adubação, sendo esses, livros físicos, com valores estáticos que são atualizados em intervalos de 8 a 10 anos, por meio de um amplo estudo de diversos solos, realizado por uma rede de laboratórios parceiros do grupo ROLAS. Com isso, visto essa dependência, toda vez que os manuais de adubação e calagem fossem modificados, seriam preciso gerar dados atualizados para que o algoritmo de aprendizado de máquina se adequasse ao cenário.

Com a arquitetura da solução desenhada e a primeira proposta não corroborada, se realizou o desenvolvimento de um aplicativo mobile capaz de interpretar análises de solo com o uso de inteligência artificial, extraído então os valores dos elementos químicos presentes na análise por meio da captura de uma foto pelo dispositivo móvel. Com isso a informação é salva em um banco de dados e pode ser rapidamente compartilhada com as pessoas interessadas.

Dentre as contribuições do trabalho estão:

- Uma arquitetura de solução para simplificar o processo de interpretação e recomendação de análise de solos.
- Avaliação do uso de inteligência artificial para ser aplicada na área de análise

de solos e recomendação de adubação

- Implementação da interpretação de análises de solo por meio de um aplicativo mobile, com comparação entre as ferramentas Tesseract OCR e Google Vision
- Criação de um dataset com 296 imagens de análises de solo e realização de experimentos para avaliação da acurácia atingida pelo algoritmo

O trabalho é composto de 3 capítulos, onde no primeiro capítulo é abordada e apresentada a estrutura metodológica do estudo feito, juntamente com os objetivos gerais e específicos, hipóteses, métodos e técnicas utilizados na pesquisa e a justificativa do projeto.

No segundo capítulo é apresentado o referencial teórico com todo o embasamento da pesquisa onde foram levantadas informações sobre os principais pilares do projeto, *Análise de Solos*, *Machine Learning*, *Análise de Sistemas e Tecnologias de desenvolvimento de software*. Além desse conteúdo também foram abordados os trabalhos relacionados ao projeto desenvolvido.

No terceiro capítulo é apresentado os resultados obtidos com base nos estudos realizados na área de solos e *Machine Learning*, que vão desde a coleta e tratamento de dados coletados do laboratório até uma análise do uso de inteligência artificial. Com isso se propõe uma arquitetura de solução e uma possível implementação de um algoritmo de interpretação de imagens de análises de solo por meio de ferramentas como Tesseract OCR e Google Vision, juntamente com um aplicativo mobile para se comunicar com esse algoritmo e enviar as fotos retiradas pelo dispositivo para interpretação.

Por fim, é criado um dataset com 296 imagens de análises de solo para testar a solução, realizando assim uma comparação entre as ferramentas de interpretação de caracteres escolhidas e um cálculo de acurácia atingido pelo algoritmo. Com os resultados apresentados se tem o fechamento do trabalho por meio da conclusão e discussão de trabalhos futuros.

CAPÍTULO 1 CONTEXTUALIZAÇÃO

1.1 TEMA

Simplificando a interpretação de análise de solo com Inteligência Artificial.

1.2 DELIMITAÇÃO DO TEMA

Desenvolvimento de um aplicativo mobile para interpretação de análise de solos, por meio da utilização de tecnologias de aprendizagem de máquina e processamento de imagens, tendo como resultado final, recomendações sobre a quantidade de fertilizantes e adubos que precisam ser aplicados na lavoura de acordo com as substâncias químicas encontradas no solo.

Uma análise de solo, nada mais é do que um documento gerado pelo laboratório de solo como resultado da análise realizada diante da porção de terra que foi trazida pelo produtor rural. Este documento traz como resultados informações químicas, como a quantidade de potássio, fósforo, ferro, magnésio, alumínio e água presentes no solo, onde cada uma dessas informações interfere direta ou indiretamente na qualidade do solo.

Com a análise de solos em mãos, o produtor rural geralmente o utiliza como forma de solicitar uma requisição de custeio de sua safra e também como forma de saber como está a qualidade do solo, porém, visto que os resultados da análise são muito técnicos é necessário que o produtor solicite uma segunda análise com um engenheiro agrônomo.

A partir de então o engenheiro agrônomo verifica os resultados da análise e cria uma lista de recomendações de aplicação de adubação e fertilizantes levando em consideração cálculos específicos dos manuais de adubação da região do Rio Grande do Sul e Santa Catarina gerando um problema de falta de agilidade, visto

que deve ser feito mais um encaminhamento, e que geralmente demora em média uma semana.

Os manuais de calagem são livros criados para auxiliar o engenheiro agrônomo quanto a tomada de decisão para realizar a recomendação de adubos e fertilizantes para o solo, desta forma o mesmo contém métricas e intervalos de sugestão de nutrientes para cada componente químico encontrado no solo, onde cada métrica interfere diretamente entre si, por exemplo, se o valor de cálcio resultante da análise de solo for maior que 1,5 e a quantidade de fósforo estiver entre 0,8 e 1,2 deve-se aplicar quantidade X de adubação, caso contrário já é outra quantidade.

Para que seja possível o desenvolvimento do aplicativo foi necessário a coleta de informações de análise de solo já realizadas, informações essas, obtidas juntamente com o Laboratório de Análises de Solo da SETREM, que disponibilizou acesso ao seu banco de dados e anualmente realiza em torno de 5000 análises de solos. Desta forma foi possível obter mais de 30 mil registros de análises para se trabalhar na criação de um modelo de treino.

Além do auxílio do laboratório de solos também foi feito o uso do programa CADUB, *software* esse desenvolvido pela UFSM e disponibilizado gratuitamente para que partes interessadas submetam um arquivo de análise de solo e com base nos manuais de calagem e adubação da região do Rio Grande do Sul é retornado a quantidade necessária de adubação e nutrientes do solo.

Após a obtenção dos dados necessários foram iniciados os processos de tratamento das informações, eliminando assim dados de interesse pessoal e deixando apenas a quantidade de componentes químicos encontrados em cada solo e a localização em que a coleta da amostra de solo foi realizada. Com os dados devidamente tratados se deu início ao processo de criação de um modelo de treino onde para isso foi realizado o estudo de ferramentas que possam apoiar nesse processo, como o *software* Weka, a biblioteca TensorFlow, Tesseract OCR e a linguagem de programação Python juntamente com a biblioteca Open CV.

Atualmente, as recomendações de adubação e calagem com base em aná-

lises de solo ainda são realizadas por *softwares* específicos onde é preciso informar manualmente as informações sobre os nutrientes do solo, como por exemplo o *software* CADUB já citado, ou até mesmo pelo próprio engenheiro agrônomo que com base em consultas em livros de recomendação de adubação e calagem calcula a porcentagem recomendada de adubação, atividade essa que tem impacto direto na agilidade do processo, pois um aplicativo para realização desse processo torna o processo mais ágil.

Dessa forma, a criação de um modelo de dados baseados em análises de solo anteriores juntamente com o auxílio do *software* CADUB disponibiliza a criação de uma ferramenta mais inteligente, que com o uso de aprendizado de máquina poderá sugerir recomendações de adubação de novas análise de solo somente a partir de dados anteriores já analisados, sem contar que a interpretação automática de uma foto da análise de solo facilita o processo de extração da informação, possibilitando um rápido compartilhamento da mesma com o engenheiro agrônomo.

O desenvolvimento do projeto foi realizado pelos acadêmicos do 8^o semestre do Curso de Bacharelado em Sistemas de Informação da Sociedade Educacional Três de Maio - SETREM, como Trabalho de Conclusão de Curso (TCC), no período de julho de 2018 à junho de 2019, envolvendo as áreas de Inteligência artificial *Machine Learning*, Processamento de imagens, Engenharia de *software* e Sistemas de Informação.

1.3 OBJETIVOS

1.3.1 Objetivo geral

Desenvolver um aplicativo móvel que realize a interpretação e recomendação de adubação de análise de solo disponibilizados pelos laboratórios.

1.3.2 Objetivos específicos

- Aprofundar os conhecimentos na área da agricultura no ramo da ciência e análise dos solos.
- Aprofundar os conhecimentos nas áreas da tecnologia da informação, principalmente em áreas como Inteligência Artificial, *Machine Learning* e proces-

samento de imagens.

- Avaliar o uso de inteligência artificial.
- Implementar um modelo/ algoritmo de inteligência artificial como forma de auxiliar na interpretação de análise de solo e possível recomendação de adubação.
- Comparar os resultados obtidos com os dados do laboratório de solos da SETREM

1.4 JUSTIFICATIVA

Atualmente o agronegócio vem ganhando grande espaço no mercado de tecnologia por ser uma área ainda defasada de aplicações que auxiliam o dia a dia do produtor no campo. Segundo pesquisa do Sebrae (2017) 95,8% dos produtores rurais já utilizam smartphones e ou celulares, sendo destes, 51,9% com acesso à internet.

Dentro da ampla variedade de problemas que se busca solucionar no ramo do agronegócio, um deles é a busca por uma maior produtividade nas plantações, sendo que um dos fatores que interfere diretamente neste parâmetro é o solo, dessa forma, sua qualidade deve ser preservada a cada novo ciclo.

Conforme Goulart (1997), o solo é um reservatório natural de nutrientes, porém o nível e volume desse reservatório podem variar devido a vários fatores, podendo existir solos bem supridos (férteis) como também solos mal supridos (pobres). Atualmente é praticamente impossível, que a agricultura comercial se sustente sem a aplicação de nutrientes, como a adição de corretivos e fertilizantes.

Para garantir este cenário, um dos métodos utilizados é o de análise de nutrientes de solo que é realizado por meio da coleta de uma amostra de terra que posteriormente é processada em um laboratório especializado. Conforme Watanabe and Bessa et al. (2002), a interpretação da análise de solo de uma lavoura é um requisito essencial para que o agricultor obtenha uma boa produção, visto que os dados alcançados apenas através da observação visual são insuficientes para

se apontar problemas referente à nutrição das plantas, tornando difícil a recomendação de adubação ideal nas plantações.

Atualmente o processo de interpretação de análise de solo ainda é feito de forma manual ou com a utilização de *softwares* pouco intuitivos e que possuem uma vida útil muito curta, tornando-se desatualizados muito rapidamente. Isso se deve pelo fato de que os manuais de calagem e adubação, os quais são levados como base para os cálculos sobre as recomendações de solo são reescritos de tempos em tempos, no qual uma informação sobre nutrientes sugerida em 2011 por exemplo, já se torna insignificante em 2018, devido a atualização dos manuais e a mudança das propriedades do solo.

As instituições financeiras fazem uso dessas recomendações para requisição de custeios. Devido esse fato, o aplicativo também poderá ser utilizado por essas organizações, como meio de garantir uma maior coerência entre as interpretações realizadas pelo agrônomo e pelos resultados gerados no próprio aplicativo.

Visto este cenário, o presente projeto tem como objetivo aplicar o *Machine Learning* para buscar aprender como esses cálculos funcionam e dessa forma ir se adequando automaticamente com o passar dos anos, sem ser preciso alimentar novas tabelas para realização de cálculos e interpretações futuras, dessa forma o sistema não se torna obsoleto e vai sendo aprimorado de forma autônoma com o auxílio da inteligência artificial.

Outra proposta de automatização, diz respeito a interpretação de fotos de análise de solo, por meio de uma rápida extração das informações contidas na análise com o uso de processamento de imagens e reconhecimento de caracteres, facilitando assim um rápido compartilhamento da informação do produtor rural para engenheiros agrônomos ou até mesmo bancos.

Para o desenvolvimento do aplicativo foi utilizado os dados de análise de solos já realizadas, para que se inicie o estudo para criação de um possível modelo de treino juntamente com os resultados do *software* CADUB, onde quanto mais dados for possível obter melhor. A obtenção dessas informações se dará por meio do laboratório de solos da SETREM que permitiu extrair mais de 30 mil registros

de análises de solo da sua base de dados.

O aplicativo resulta um laudo contendo recomendações para correção do solo, e conseqüentemente os ajustes das quantidades dos componentes químicos obtidos com base em um ponto de coleta específico, sendo que esta interpretação ainda é feita de forma manual pelos produtores rurais juntamente com um engenheiro agrônomo que se baseia nos manuais de calagem e adubação para realizar a recomendação de adubação.

O presente trabalho irá proporcionar melhorias para o atual mercado, por não possuir um produto semelhante ao proposto, assim como também para o ambiente acadêmico, visto que a utilização de tecnologias como o Machine Learning para resolução desse tipo de problema ainda é pouco aplicada na área da agricultura, podendo servir como referência para futuros trabalhos nesta área.

É notável perceber que a tecnologia da informação está cada vez mais presente no meio agrícola atingindo não só os grandes produtores, mas também os pequenos. Dessa forma, o presente trabalho traz contribuições ao pequeno, médio e grande produtor pois o mesmo poderá obter de forma fácil, ágil e na palma de sua mão uma interpretação completa da quantidade de insumos agrícolas que deverá aplicar em sua terra para seguir com uma alta produtividade a cada novo plantio, além de facilmente compartilhar a informação presente na análise de solo com pessoas interessadas.

1.5 PROBLEMA

Como a análise de solo, através da utilização de técnicas de Machine Learning auxilia no processo de interpretação e recomendação de correção do solo?

1.6 HIPÓTESES

A solução proposta possui um grau de acurácia de 95% na interpretação de análise de solo.

A utilização de Machine Learning para realizar a recomendação das análises de solo elimina a necessidade de utilização diretamente dos manuais de calagem e adubação.

1.7 METODOLOGIA

Dentre os aspectos relacionados na elaboração de projetos de conclusão de curso está a metodologia, de acordo Lovato (2013) este item é visto como uma ciência empregada por pesquisadores, tendo como base a utilização de meios que visam atingir teorias que acrescentam conhecimento em um determinado assunto.

Um dos métodos aplicado foi a pesquisa exploratória, visto que a mesma geralmente é aplicada a partir de um problema ou questão que é pouco discutido ou ainda não tem nenhum estudo a respeito, cenário este encontrado no projeto atual. Por meio de consultas em bases de artigos científicos como IEEE e CAPES, não foram encontrados muitos trabalhos relacionados exatamente no mesmo ramo ao qual este trabalho se propõe, que é o uso de Machine Learning (aprendizagem de máquina) para realizar a interpretação de análises de solo.

Deste modo, no âmbito exploratório serão obtidas maiores informações sobre o ramo da agricultura responsável pela análise e ciências de solos, estudando os principais componentes químicos e fatores que influenciam na qualidade do mesmo, bem como a aplicação e funcionamento da tecnologia de Machine Learning dentro da área de inteligência artificial.

Como início do estudo foi analisadas as informações da base de dados do Laboratório de Solos da SETREM para que seja realizada a interpretação e tratamento destes dados para definir quais detalhes precisam ser melhores analisados no desenvolvimento do aplicativo mobile.

Outro método de pesquisa utilizado neste trabalho foi o descritivo, que segundo TEIXEIRA (2003) é uma pesquisa que busca descrever uma realidade ou necessidade existente, como meio de comprovar uma proposta ou até uma hipótese. Neste sentido, este método foi aplicado para constatar a importância do desenvolvimento da aplicação para interpretação de análises de solo e como a mesma poderá facilitar o dia a dia das pessoas e organizações que irão utilizá-la com base nas necessidades existentes hoje no mercado.

1.7.1 Métodos de abordagem

Conforme MARCONI and LAKATOS (2007), os métodos de abordagem são caracterizados por um conjunto de atividades racionais e sistemáticas que auxiliam no alcance dos objetivos, proporcionando mais proteção e economia, além de ser fundamental na identificação de erros e no auxílio na tomada de decisões.

A abordagem que foi utilizada no projeto foi a quali-quantitativa sendo que serão usadas tanto análise de características, como a identificação dos nutrientes que o solo possui em sua composição e também a análise de números, esses que serão obtidos através de análise de solo e servirão de modelo para geração de estatísticas da qualidade dos solos cadastrados no aplicativo.

Segundo Creswell (2007), a abordagem quali quantitativa se popularizou com o passar dos anos devido a legitimidade e desenvolvimento encontrado nos métodos quantitativos e qualitativos visto que proporciona uma maior compreensão dos problemas estudados.

Já Neves (1996), afirma que a utilização deste modelo se baseia da comparação dos dados obtidos com a forma qualitativa e quantitativa onde essa combinação pode acontecer de forma simultânea ou alternada com o propósito de conseguir responder o problema da pesquisa. Este tipo de método é indicado para que a abstração seja diminuída e o pesquisador se aproxime ainda mais do objeto estudado, proporcionando uma maior credibilidade nas informações obtidas.

No método qualitativo foi realizada uma abordagem por meio de entrevistas com consultores e engenheiros agrônomos bem como produtores rurais como forma de compreender suas principais necessidades e dificuldades enfrentadas nas áreas de análise de solo.

Já no aspecto quantitativo foi analisadas as porcentagens e graus de erros e acertos gerados pelo aplicativo de Machine Learning que se pretende desenvolver. Além disso foi útil para apresentar indicadores como eficiência, desempenho, veracidade e grau de acerto de cada modelo.

1.7.2 Métodos de procedimento

De acordo com Gil (1994), os métodos de procedimentos são um plano de ação, formado por um conjunto de etapas ordenadamente dispostas, destinadas a realizar e antecipar uma atividade na busca de uma realidade. Os métodos se baseiam em dois motivos: natureza do objetivo a que se aplica e objetivo que se tem em vista. Por se tratar de um trabalho em que a abordagem se dará de forma quali quantitativa serão utilizados os procedimentos documental, estatístico e comparativo.

O método de procedimento documental foi utilizado para obtenção das informações da base de dados do Laboratório de Solos da SETREM, com esses dados foi possível obter análises com componentes químicos variados para cada tipo de solo e servirá de entrada para gerar recomendações utilizando o *software* CADUB, obtendo assim a saída esperada, então as entradas e saídas serão submetidas para treinamento do modelo de Machine Learning, onde com o modelo criado foi possível gerar novas recomendações, dessa forma também será possível a criação de um conjunto de imagens de análises de solo para testes, onde essas imagens serão tiradas de forma que tenha sombra na imagem, sem a sombra, com *flash*, sem *flash*, imagens inclinadas, tremidas e algumas fotos de forma livre, para que fotos diferentes possam ser tiradas e tratadas pela API.

A utilização do método estatístico se dará por meio da análise das porcentagens e graus de erros e acertos gerados pelo aplicativo de aprendizagem de máquina que se pretende desenvolver.

O método comparativo foi utilizado como forma de determinar qual o algoritmo mais adequado para elaboração e aplicação neste projeto, comparando indicadores como eficiência, desempenho, veracidade e acurácia da aprendizagem e dos resultados obtidos.

1.7.2.1 Validação das Hipóteses

A primeira hipótese “A solução proposta possui um grau de acurácia de 95% na interpretação de análise de solo.”, foi validada por meio da criação de um dataset contendo várias imagens de análises de solo nos mais diferentes cenários.

Diante disso, foi realizado o cálculo de acurácia, de acordo com a quantidade de elementos químicos interpretados corretamente para cada análise de solo deste dataset.

A segunda hipótese “A utilização de Machine Learning para realizar a interpretação das análises de solo elimina a necessidade de atualização dos manuais de calagem e adubação”, foi validada por meio do estudo das fórmulas e métodos utilizados pela Embrapa para geração dos valores base que são inseridos nos manuais de calagem. Dessa forma, após entender os fatores que influenciam na geração desses dados estatísticos, foi verificada a possibilidade de simular estas informações para novos manuais de calagem. Isso por meio da utilização de Machine Learning aplicada as informações de análise de solos obtidas no laboratório de solos da SETREM.

1.7.3 Técnicas

De acordo com MARCONI and LAKATOS (2007), as técnicas são definidas por serem o aglomerado de processos que tem como base a ciência e a forma que se tem de chegar a um propósito específico.

A técnica de pesquisa bibliográfica foi abordada como forma de adquirir conhecimentos sobre os assuntos discutidos, principalmente nas áreas de análise de solos e na área de tecnologia voltada a Inteligência Artificial e Machine Learning. Com base nisso também foram realizadas pesquisas de trabalhos relacionados ao tema para se encontrar um referencial teórico competente e suficiente para se desenvolver o aplicativo proposto.

Outra técnica empregada neste trabalho foi a de entrevista não estruturada onde se buscará conversar com engenheiros agrônomos e profissionais da área de solos para buscar maiores informações sobre o cenário atual dentro da área de interpretação de análises de solo e com base nisso recolher feedbacks para implementar as sugestões no aplicativo proposto.

O cronograma do presente trabalho está localizado no apêndice A e o orçamento no apêndice B.

CAPÍTULO 2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo é apresentada a fundamentação teórica do trabalho, as pesquisas já realizadas por outros autores sobre os temas tratados e os conceitos que serviram de base para o trabalho proposto.

2.1 AGRICULTURA

Conforme Diehl (1989) a agricultura pode ser definida pelo esforço desempenhado para deixar a planta cultivada nas melhores condições de clima e de solo visando o máximo proveito e rendimento em qualidade e quantidade. Já Barros (1975) diz que a agricultura é a artificialização do meio natural realizada pelo homem com o objetivo de tornar o solo mais apto ao desenvolvimento de espécies vegetais e animais.

A atividade agrícola passou por uma série de transformações nos últimos anos, sendo necessário um gerenciamento cada vez maior de seus processos produtivos. Junto com isso, o crescimento acelerado do desenvolvimento de novas técnicas vinculadas ao manejo das culturas, novas ferramentas e insumos mais eficientes proporcionou aumentos significativos no rendimento das culturas.

Conforme Soglio and Kubo (2009), o século XX ficou marcado pela importância e crescimento da agricultura com um grande impacto como atividade econômica, cujo principal objetivo era o lucro e sua maximização e extrema valorização. Desta forma a função da agricultura passou a ser não só um meio de produção de alimentos mas muito mais uma atividade comercial, como um negócio a ser controlado pelo mercado. Este cenário alcançou tal ponto que atualmente muitos setores e ramos da sociedade já estão tratando a agricultura como sinônimo de agronegócio.

Como visto acima, a agricultura é uma área de vital importância, dentro dela existem várias outras áreas ligadas a ela, podemos citar, a agricultura de precisão e o agronegócio.

2.1.1 Agricultura de precisão

De acordo com Molin and Rabello (2011), o conceito de agricultura de precisão (AP) é pouco conhecido entre os produtores rurais brasileiros, gerando dessa forma bastante dúvidas quanto a utilização de suas técnicas de manejo. Atualmente ainda é bastante frequente a associação do termo AP como sendo um pacote de metodologias para auxiliar na resolução de todos os problemas presentes na agricultura nacional.

Conforme Coelho (2004), a agricultura de precisão implica diretamente em vários fatores, como na análise da variabilidade espacial, etapas de coleta de dados, gerenciamento de informações, aplicação de insumos e também na avaliação econômica e ambiental dos resultados.

2.1.2 Agronegócio

Para ZYLBERSZTAJN, NEVES and CALEMAN (2015), o termo agronegócio no âmbito global, se caracteriza por toda a relação comercial e industrial que envolve a cadeia de produção agrícola ou pecuária. Já no Brasil, o conceito é usado para dar referência a grandes propriedades de monocultura atual que utilizam tecnologias de ponta e muito pouca mão de obra, sendo que a produção é geralmente voltada para o mercado externo ou diretamente para as agroindústrias, visando o lucro.

O ramo do agronegócio tem sido por muito tempo um dos principais influenciadores na sustentação econômica e social do Brasil visto que as circunstâncias de solo, clima e extensão territorial, juntamente com a grande quantidade de produtores com alto poder produtivo são diferenciais competentes para diferenciar o Brasil de seus concorrentes, tanto no mercado de produção quanto no de exportação agrícola.

De acordo com QUEIROZ and Zuin (2015) um dos maiores desafios en-

contrados recentemente na área do agronegócio é o de implementar, planejar e conduzir padrões produtivos inovadores e que tragam benefícios ao meio ambiente e aos territórios rurais. Nesta mesma área, a agricultura de precisão entra como meio de auxiliar na preparação do solo por meio da inserção de quantidades ideais de insumos, diminuindo assim o impacto no meio ambiente e também o desperdício de nutrientes.

2.1.3 Análise de solo

O processo de análise de solo é um dos únicos métodos capazes de determinar a capacidade que o solo terá para suprir os nutrientes necessários pelas plantas, sendo uma das formas mais econômicas, simples e eficientes usadas na obtenção de indicadores sobre fertilidade das terras e sobre as devidas aplicações de corretivos e fertilizantes como objetivo de aumentar a produtividade das culturas e lavouras.

Dentre os muitos aspectos positivos da utilização das técnicas de análise de solo vale destacar o seu baixo custo e rapidez para obtenção de resultados, o alto grau de acerto quanto a quantidade de fertilizantes e corretivos que deverão ser aplicados no solo, evitando assim gastos desnecessários com insumos e mão de obra, além de ainda minimizar danos ao meio ambiente evitando a contaminação de águas por excesso de fertilizantes.

De acordo com FURTINI NETO and VALE et al. (2001), o processo de análise de solos pode ser separado em três passos, o de amostragem do solo, análise dos dados obtidos em um laboratório especializado de solos e posteriormente a realização da leitura e interpretação dos resultados que foram gerados pelo laboratório.

2.1.3.1 Amostragem de solo

Segundo FURTINI NETO and VALE et al. (2001), dentro do processo de análise do solo, uma das etapas mais críticas é a de amostragem, visto que para a mesma é utilizada somente uma pequena porção de terra para representar alguns hectares, não havendo meios para se corrigir um possível erro durante a coleta.

Dessa forma, para que os resultados da análise realmente tenham uma representatividade confiável e possam servir como base para uma futura recomendação de adubação é necessário que a amostragem da área seja feita corretamente.

Diante disto, para que a amostragem de solo seja realizada realmente de forma correta e se obtenha uma representação realmente útil da terra ao qual se deseja analisar é preciso seguir alguns passos.

2.1.3.1.1 Subdivisão da propriedade em pedaços de terra homogêneos

Antes do início da coleta da amostra é necessário que a propriedade ao qual se deseja realizar a análise seja igualmente dividida em pedaços de terra menores com área máxima de 10 hectares, onde nessa subdivisão deve ser levado em consideração a posição do solo em relação ao relevo, ou seja se está em um morro, encosta ou baixada, a cor e textura do solo, histórico de uso e manejo como culturas anteriores, calagens e adubações já realizadas e também a verificação se existe ou não presença de erosão na área.

2.1.3.1.2 Coleta das amostras de solo

Após a divisão da área em pedaços de terra menores devem ser coletadas um número mínimo de 20 amostras de forma aleatória, ou seja, caminhando-se em ziguezague buscando percorrer toda a área, onde serão retirados aproximadamente 300g de solo que deverá ser guardado imediatamente em um saco plástico e devidamente identificado com o nome do proprietário, data da coleta, profundidade da coleta e qual a cultura.

2.1.3.1.3 Profundidade de amostragem

Este tipo de medida varia de acordo com o tipo de cultura, que pode ser anual ou perene e também conforme o sistema de cultivo sendo ele convencional ou plantio direto. Em geral para a maior parte das culturas anuais a profundidade indicada é de 0 a 20 cm, já para culturas perenes essa medida sobe um pouco podendo chegar de 20 a 60 cm. Quanto ao sistema de plantio direto onde não se tem um revolvimento do solo é recomendado que se aplique uma profundidade de 10 cm.

2.1.3.1.4 Ferramentas de amostragem

Atualmente existem diversos amostradores no mercado que servem para se realizar a coleta da amostra, dentre os principais está o trado de rosca, trado holandês, trado caneca, sonda e pás, conforme é possível ver na figura 1. É muito importante lembrar a importância da limpeza dos equipamentos e recipientes utilizados nas misturas das amostras quando for mudar de pedaço de terra.

Figura 1: Ferramentas para a amostragem do solo



Fonte: SBCS (2004)

2.1.3.1.5 Época e frequência de amostragem

O período de coleta ideal também varia de acordo com as culturas, para culturas anuais é recomendado o início do período de seca, sendo cerca de 3 a 4 meses antes do plantio, já para as culturas perenes se recomenda fazer a amostragem logo após a colheita da última cultura plantada. Quanto a frequência em que se deve repetir o processo de amostragem, o mesmo varia de 1 a 4 anos dependendo da intensidade de uso e manejo do solo.

2.1.3.1.6 Análise em laboratório

A análise em laboratório acontece assim que a amostra dos pedaços de terra é coletada na propriedade rural, dessa forma as amostras são submetidas a uma série de análises químicas, em laboratórios públicos ou privados. Para se avaliar a fertilidade do solo, uma análise completa compreende o levantamento de

informações como pH, fósforo, potássio, cálcio, magnésio, enxofre, zinco, manganês, cobre, ferro, boro, alumínio, hidrogênio e outros componentes.

De acordo com RAIJ (1991), as análises químicas e físicas de amostras de solo eram feitas por laboratórios anexados juntamente a instituições de pesquisa e faculdades de Agronomia, porém com o passar dos anos a quantidade de laboratórios nesse ramo teve um grande crescimento e a análise de amostras de terra passou a ser realizada por várias organizações, observando ainda um grande aumento de serviços prestados por laboratórios privados que além da análise do solo também disponibilizam de serviços de consultoria e assistência técnica.

Para efeitos de recomendação os laboratórios realizam um conjunto de análises de solo de rotina muito solicitadas por agricultores e técnicos que se referem a esse grupo de análise como Análise Sumária (AS), onde a mesma engloba o teor extraível dos nutrientes principais do solo bem como a verificação da presença de 3 parâmetros que interferem diretamente na qualidade desses nutrientes, visando assim obter uma nutrição eficaz das culturas.

De acordo com TOMÉ JÚNIOR (1997), para determinar os nutrientes disponíveis nas amostras, como Potássio, Fósforo, Magnésio, Cálcio e outros micronutrientes é realizado um processo químico de duas etapas, onde primeiramente ocorre a extração e homogeneização da amostra com soluções extratoras apropriadas e posteriormente se aplica a decantação ou filtração, fazendo com que as formas extraíveis ou trocáveis do nutriente sejam deslocadas para a solução ou sobrenadante.

Outro fator importante para obtenção de sucesso em uma análise de solo em laboratório é a utilização de equipamentos corretos como os para realização de medidas de massa e volume. Segundo TOMÉ JÚNIOR (1997), dentre os principais equipamentos utilizados nos processos químicos da análise estão o calorímetro, fotômetro de chama, espectrofotômetro de absorção atômica e medidores de pH. Conforme o mesmo, além dos equipamentos adequados também é preciso seguir determinados padrões e especificações de rotina, que vão desde a lavagem do material utilizado até o processamento das informações referentes aos resultados recebidos na análise.

Quanto aos parâmetros químicos levados em consideração pelos laboratórios na realização da análise de solo esses podem variar dependendo da região em que a amostra foi recolhida. Conforme FURTINI NETO and VALE et al. (2001), os critérios de interpretação não são únicos e variam muito notadamente entre os estados, porém, se o método de análise for o mesmo não existe problema pois a interpretação dos resultados será realizada com base na consulta a tabelas de diferentes estados.

2.1.3.2 *Processo de análise de solo*

O processo de análise de solo, segundo Embrapa (2011) é composto de várias etapas, a coleta do solo pelo produtor, a preparação da amostra, análise física e análise química. A coleta do solo é uma etapa de extrema importância nesse processo, como já vimos em 2.1.3.1 - Amostragem de solo, que existem formas corretas da coleta de análise de solo, assim como também, ferramentas específicas que podem ser utilizadas. Na próxima etapa é a análise física do solo, ou seja, atributos como umidade atual do solo, terra fina, cascalho e calhaus, etc. Por fim, é realizada a análise química do solo, nessa etapa é realizado o levantamento das quantidades, por exemplo, Nitrogênio, Cálcio e Potássio, etc.

2.1.3.2.1 *Análise física*

- O início da análise física do solo é dado pela preparação da amostra, com o objetivo de realizar a separação das frações de solo, ou seja, realizar a separação da terra fina, cascalhos e calhaus. Para que essa separação seja feita é necessário realizar a dessecação ao ar em uma tabua 40x60x8cm com todos os torrões desmanchados manualmente, pesando e anotando o peso da amostra que vai ser preparada, onde geralmente uma parte dessa amostra é submetida para duas peneiras, uma de 20mm e outra de 2mm. Os materiais retidos na primeira peneira será os calhaus e o material retido na peneira de 2mm será os cascalhos, após essa etapa será feita a pesagem de cada amostragem e utilizada a seguinte fórmula para quantificação da terra fina: (terra fina (g/kg) = Peso total da amostra (g/kg) - (g/kg calhaus + g/kg cascalho))

- A definição da umidade atual presente na amostra de solo, porém somente se transportada em embalagem impermeável e vedada. Uma parte dessa amostra é submetida para a uma estufa a 105-110°C por um período de 24 horas e então é realizado um cálculo com o peso da terra úmida versus terra seca, para alcançar o teor de umidade do solo.
- Quando a etapa de umidade do solo é concluída, a próxima etapa é a a verificação da densidade do solo, que é feito utilizando um balão volumétrico que é preenchido com solo seco em estufa e completado com álcool etílico, após então é realizado um cálculo dividindo o peso da amostra utilizada seca em estufa dividido pelo volume de álcool utilizado.
- Para definição do volume de poros totais do solo ocupados por água e/ou ar. Para realizar o cálculo da porosidade total do solo pode ser através da seguinte formula: Porosidade total = $100 (a - b) / a$, onde "a"significa a densidade real e "b" a densidade aparente.
- Análise granulométrica é realiza na velocidade de queda das partículas que compõem o solo, para isso é realizado a suspensão do solo com água, adicionado com um dispersante químico(soda ou calgon) e com isso é pipetado um volume da suspensão para ocorrer um deslocamento vertical. As frações grosseiras (areia fina e grossa) são divididas por tamisação, que são secadas em estufa e pesadas para obter os percentuais respectivos. Já para a definição do silte, é necessário a realização da diferença das outras frações em relação a amostra original.
- Para quantificação da argila dispersa em água, é necessário deixar uma amostra de solo em repouso durante a noite em um copo de plástico, após retirar o material e submete-lo a uma peneira, lavar a areia, completar o volume para 1000ml, realizar a agitação por suspensão por 20 segundos, sifonar a suspensão após 90 minutos e realizar a leitura do densímetro. Junto com essa etapa é necessário realizar um prova em branco e fazer a leitura do densímetro, para somente após isso realizar o cálculo Teor de argila = $(a + b) \times 20$, onde "a"é a leitura de amostra e "b"é a leitura da prova em branco.

- No cálculo do grau de flocculação que é realizada através da fórmula: grau de flocculação = $100 (a - b) / a$, onde "a" é a argila total e "b" a argila dispersa em água, por meio deste é indicado qual a proporção de fração de argila que está flocculada.
- Com a determinação da análise granulométrica concluída, pode se obter a relação de silte/argila, essa relação é utilizada para verificação do grau de intemperismo do solo e permite também a verificação da existência de movimentação da argila.
- Após este processo é verificado o grau de condutividade hidráulica do solo, ou seja, a velocidade que a água se movimenta no solo, essa velocidade pode ser medida após 7 - 8 horas, ou então quando os valores estabilizam.
- A métrica utilizada para verificação da capacidade de retenção de água do solo é a porcentagem de saturação, ou seja, é a razão entre o volume de água e espaço poroso.
- Na porcentagem de agregados, existe duas formas de realização desse cálculo, um pelo método de via úmida e outro pelo método de via seca, no primeiro método, o objetivo é medir a quantidade e distribuição do tamanho dos agregados que são estáveis em água, que não são desmanchados, ou seja, são os sólidos contidos nessa solução que após são separados por tamisação. Já no segundo método, o solo é submetido a agitação mecânica do solo já seco, ou seja, com a utilização de uma jogo de peneiras que são movimentados de forma circular com vibração e existir total estabilização, não irá passar nenhuma partícula, então após isso é realizado a pesagem na estufa de 105 °C da amostra submetido a peneira e da amostra sem submeter a peneira, então utilizando a fórmula: Teor de agregados = $1.000 (a / b)$, onde "a" é peso dos agregados secos a 105°C e "b" é o peso da amostra seca a 105°C.
- A etapa de cálculo da limite de liquidez, consiste na capacidade de mudança do estado líquido para o de plástico, ao que se refere o teor de umidade do solo, através da técnica de resistência ao cisalhamento, para auxiliar nesse processo é utilizado o aparelho de casagrande, para contagem de pancadas

dadas em uma massa de solo que é feita com silte e água, então será possível alcançar o limite de liquidez.

- Para determinação do limite e índice de plasticidade, é realizada a verificação da taxa de umidade do solo, referente a troca do estado plástico para o estado de semi-sólido, através de ensaio mecânico. Para essa etapa, é preciso retirar em torno de 10 a 15g da massa de solo criada na etapa anterior e formar um bastão de 3 a 4 mm de diâmetro até que comece a ficar quebradiço, então deve ser pesado e determinada a umidade. Esse processo é repetido por 3 ou mais vezes e o limite de plasticidade é determinado pela média das umidades de cada amostra.
- O próximo passo da análise física é o limite de pegajosidade do solo, nessa etapa é necessário a criação de uma massa, com uma amostra do solo e água, então com a ajuda de uma espatula, deve se comprimir essa massa até que fique grudada na espatula, enquanto isso não acontecer deverá ser adicionado mais água, ao ficar uma quantidade de solo grudada na espatula, entende-se que o solo atingiu o estado plástico, então essa massa gruda na espatula deve ser retirada, medido a umidade contida e repetir o processo por 3 vezes, o limite é alcançado pela média da umidade da repetição do processo.
- A determinação do limite de contração, é realizado por meio do teor de umidade de solo saturado com água, dessa forma, no período certo o solo irá parar de se contrair e assim irá controlar o nível de perda de água deste solo.
- O grau de contração é utilizado um cálculo que expressa a porcentagem, da diferença de volumes finais e volumes iniciais, assim como o volume inicial do objeto utilizado para a etapa anterior, que resultou no limite de contração. A fórmula utilizada pe $GC = (a - b)/a \times 100$, onde "a"é volume inicial da pastilha e "b"é o volume final da pastilha.
- No cálculo da superfície específica é baseada na retenção dos líquidos polares, ou seja, a quantidade de éter mono etílico do etileno glicol (EMEG),

no momento que não existir mais EMEG livre e somente aquele absorvido, irá formar uma camada monomolecular.

- A separação de argila e silte para posterior análise de raio x, é determinada pela velocidade da queda de partículas em meio agregada há uma dada densidade, então as partículas de areia são separadas por tamização, uma espécie de peneira utilizada para separar partículas de diferentes tamanhos. Já a argila por tempo de sedimentação, que consiste em deixar a mistura em repouso para que todas suas fases sejam separadas por meio da diferença de densidade.

2.1.3.2.2 Análise química

Segundo Laborsolo (2019), a análise química do solo é subdivida em várias vários tipos de análises, são elas, a análise básica, alumínio, enxofre, fósforo remanescente, matéria orgânica, troca de cátions(CTC) determinada, carbono total, silício solúvel, micronutrientes, boro, sódio, condutividade elétrica.

- O tipo de análise básica consiste na análise dos nutrientes básicos de fertilidade do solo, onde o são contemplados os macronutrientes (Ca, Mg, Na e P) e micronutrientes (Cu, Fe, Mn, Zn). Esse tipo de análise química já é o suficiente financiamentos de custeios agrícolas, permitindo também a avaliação da calagem e adubação com os parâmetros de potássio e fósforo.
- Outro tipo de análise é a de alumínio trocável presente, ou também chamada de acidez trocável, sendo imprescindível para auxiliar na tomada de decisão da necessidade de aplicar a gessagem, pois o alumínio inadequado pode coibir o crescimento radicular e dificulta o desenvolvimento das plantas.
- A análise do enxofre parametriza a quantidade de sulfato presente, possibilita a avaliação da falta de adubação sulfatada, importante ressaltar, que esse tipo de análise deve ser o monitoramento principalmente das camadas mais profundas do solo.
- A determinação de níveis preocupantes de fósforo, enxofre e zinco no solo, é realizada pela análise do fósforo remanescente, esse relatório é imprescindível

vel para avaliação com segurança da escassez da adubação desses nutrientes.

- A avaliação do teor de carbono orgânico no solo manifesta-se pela matéria orgânica do solo, onde a análise da matéria orgânica tem vital importância para monitoramento da degradação ou até a desertificação do solo.
- No próximo tipo de análise tem a importância para gerar resultados de recomendações mais precisa e assertivas, pois a análise de CTC determinada, é a mais correta, visto que, a CTC apresentada em algumas análises é definida através de um cálculo e esse por sua vez pode conter erros por ter fatores externos que podem influenciar no resultado.
- Já para linhas de financiamento da ABC (Agricultura de Baixo Carbono) é necessária a análise do carbono total, que define todos os conjuntos de carbono, sejam eles orgânicos e inorgânicos.
- O silício solúvel é o tipo de análise que define o parâmetro do teor solúvel de silício, a verificação desse parâmetro tem como o objetivo garantir a absorção do fósforo.
- Os micronutrientes são importantes para definição dos valores presentes de cobre, ferro, manganês e zinco, para que dessa forma a recomendação para correção desses nutrientes possa ser realizada.
- A definição do boro, tem como objetivo avaliar a presença e a quantidade desse nutriente e posterior correção em caso de escassez.
- A análise de sódio presente no solo, é responsável pela salinização do solo e está presente em diversos corretivos de solo, devido à isso, quando está presente em grandes quantidades, estimula a deficiência de potássio.
- E por fim a análise de condutividade elétrica, é necessária para definir indiretamente o nível de salinidade do solo.

2.1.3.2.3 Interpretação dos resultados

De acordo com FURTINI NETO and VALE et al. (2001), os critérios de interpretação não são únicos e variam de acordo com o estado ao qual se solicitou a análise de solo, porém, dependendo do método de análise pode acontecer situações em que a interpretação dos resultados aconteça com base na consulta de tabelas de diferentes estados obtendo dessa forma um média.

A interpretação dos resultados se dá através da análise das quantidades de nutrientes químicos que foram encontrados nas amostras analisadas e pela quantidade de nutrientes que é absorvida pela planta que será plantada no solo. Após realizar a análise do solo o laboratório disponibiliza os resultados ao produtor, onde são descritos todos os componentes químicos encontrados no solo e suas respectivas quantidades, conforme figura 2.

Figura 2: Análise de solo

LAUDO DE ANÁLISE DE SOLO

Nome: _____
Localidade: _____
Município: _____

Data Recebimento: 29/03/2019
Data de Expedição: 05/04/2019

Amostra Nº	Área (ha)	Argila %	Classe Textural	pH H ₂ O	Índice SMP	M.O. %	P	K	Ca	Mg	Al	H+Al
42424	21,00	81	1	5,7	5,5	3,7	22,5	94	5,8	2,0	0,0	7,7

Amostra Nº	CTC (Cmol L ⁻¹) Efetiva	Saturação (%)	S	B	Cu	Zn	Mn	Mo	Fe	Na
42424	8,0	15,7	50,9	0,0	11,7	0,4	7,0	4,3	33,6	N.D

Amostra Nº: 42424 | AD1 - Mat 12334

Identificação

Observações

Responsável Técnico

Fonte: Laboratório de solos SETREM, 2018

A interpretação da análise de solo geralmente é realizada pelos profissionais agrônomos, os mesmos utilizam manuais de calagem e adubação para se basear na quantidade de nutrientes e adubação necessária que será necessário aplicar na lavoura de acordo com os nutrientes químicos resultantes da análise de solo.

Na figura 2 percebe-se todos os nutrientes que são verificados pelo labo-

ratório e seus valores resultantes encontrados na amostra de solo, dentre eles está o Fósforo (P), Potássio (K), Cálcio (Ca), Magnésio (Mg), Alumínio (Al), onde com base no manual de calagem e adubação o valor de cada nutrientes é separado em diferentes intervalos conforme a sua qualidade, que podem ser, muito baixo (MB), baixo (B), médio (M), alto (A) e muito alto (MA).

Com a análise de solos e o manual de calagem e adubação em mãos o engenheiro agrônomo realiza um diagnóstico das informações verificando a quantidade de nutrientes encontrados no solo e a partir disso gera um relatório sobre recomendações de adubação, neste relatório são apresentados resultados em 3 tópicos principais, calagem, adubação de base e adubação de cobertura, conforme figura 3.

Figura 3: Recomendação de adubação e calagem

Cooperativa de Técnicos do Noroeste do Estado do RS- UNITEC														
Rua: Av. Santa Rosa		Nº: 301		Cidade: Três de Maio				Estado: RS						
Bairro: centro		CEP: 98910-000		Telefone: 0 55 99962 0573		E-mail: klarmann@san.psi.br								
Visite nossa empresa no seguinte endereço eletrônico: www.unitec-tm.com.br														
INDICAÇÃO DE ADUBAÇÃO E CALAGEM														
Proprietário: Sociedade Educacional Três de Maio				Área da gleba (ha): 15										
Local: Esq. Mota - Independência				Cultivo após análise: Segundo										
Fazenda: matr. 3149 - Rapachi centro				Expectativa de rendim. (t/ha): 4										
Gleba: lavoura				Registro da análise de solo: 29867										
Cultura: Soja				Data de semeadura ou plantio: 01/11/2017										
DIAGNÓSTICO PARA ADUBAÇÃO E CALAGEM														
pH	Ca	Mg	Al	H+Al	CTCpH7	Al	V	SMP	MO	Argila	P	K		
---	---	---	cmolc/dm ³	---	---	---	---	---	---	---	---	---	---	---
5,6	4,8	2,0	0,0	4,3	11,6	0,0	62,8	6,0	3,0	63,0	12,7	186,0		
M	A	A	---	---	M	MB	B	---	M	---	MA	MA		
MB = Muito Baixo		B = Baixo		M = Médio		A = Alto		MA = Muito Alto						
CALAGEM														
Fonte		PRNT (%)	Preço (R\$/t)	Dosagem (t/ha)	Custo (R\$/ha)	Quantidade (t)	Custo total (R\$)							
Exemplo		80	133,00	0,0	0,00	0,0	0,00							
ADUBAÇÃO DE BASE														
Fórmula		Preço (R\$/t)	Dosagem (kg/ha)	Custo (R\$/ha)	Quantidade (kg)	Custo total (R\$)								
0-20-25		1330,00	309	410,97	4635	6164,55								
ADUBAÇÃO DE COBERTURA														
Fonte		Preço (R\$/t)	Dosagem (kg/ha)	Custo (R\$/ha)	Quantidade (kg)	Custo total (R\$)								
0		0,00	0	0,00	0	0,00								

Fonte: Laboratório de Solos Setrem, 2019

A primeira parte dos resultados gerados pelo engenheiro agrônomo são referentes a calagem necessária para ser aplicada no solo, ou seja, a dosagem de calcário que deve ser aplicado, geralmente medida por (t/ha), tonelada por hectare. Além disso, também pode ser informado o preço que será gasto por tonelada aplicada.

A segunda parte dos resultados se refere a adubação de base, que engloba os nutrientes da fórmula NPK, Nitrogênio (N), Fósforo (P) e Potássio (K),

sendo estes os nutrientes mais importantes para a qualidade das plantações, já que juntos formam um fertilizante poderoso se aplicado corretamente no solo. A disposição dos resultados também se dá por tonelada por hectare, e além dessas propriedades também são apresentadas as quantidades totais que devem ser aplicadas e o custo de cada aplicação.

Por fim, o último resultado apresenta a situação de adubação de cobertura, essa responsável por registrar a necessidade de uma adubação recorrente em meses futuros, dessa forma a quantidade de adubo a ser aplicado é dividido em partes, para situações em que se acabou de fazer uma colheita naquele solo, dessa forma, a adaptação para uma nova plantação acontece com uma melhor absorção de nutrientes pelo solo.

2.1.3.2.4 Manuais de calagem

Os manuais de calagem são livros criados por profissionais e estudiosos da área agrícola com o intuito de orientar o uso apropriado de fertilizantes e corretivos de acidez no solo sendo que variam de acordo com a região e o tipo de solo, dessa forma existe por exemplo um manual específico do estado do Rio Grande do Sul e Santa Catarina e outro para o estado do Rio de Janeiro, visto que o relevo e o solo sofrem grandes alterações.

Conforme SBCS (2016) o manual de calagem contém os últimos avanços no ambiente agrônomo obtidos com base em pesquisas e apoio de instituições de pesquisa, ensino, assistência técnica, extensão rural e inclusive o setor privado. Devido à rápida evolução da tecnologia este tipo de manual deve ser aperfeiçoado continuamente. Desta forma é papel dessa sociedade atualizar o manual quando necessário com os novos parâmetros de recomendação.

No Rio Grande do Sul, a primeira recomendação de adubação com base em análises de solo e manuais de calagem foi feita por Mohr em 1950, o mesmo dividiu o estado em quatro regiões fisiográficas estabelecendo graus de referência para as análises de solo de acordo com as regiões e seus 4 tipos de solos, solos formados sobre basalto, solos formados sobre arenito Botucatu, solos formados sobre rochas graníticas e solos formados sobre areias e sedimentos recentes.

Para levantamento e atualização dos dados dos manuais de calagem da região do Rio Grande do Sul existe a organização ROLAS, (Rede Oficial de Laboratórios de Análise de Solo e de Tecido Vegetal dos Estados do Rio Grande do Sul e de Santa Catarina). Esta organização é vinculada ao setor de fertilidade do solo e nutrição de plantas do NRS (Núcleo regional Sul), sendo que a rede que a contempla possui aproximadamente 25 laboratórios onde anualmente são feitas reuniões para analisar informações de pesquisa sobre metodologias de análise de solo e controle de qualidade das análises de solo e de micronutrientes.

Dessa forma, várias amostras são submetidas a análise por todos os laboratórios que fazem parte da organização e a partir disso são gerados dados estatísticos para se caracterizar uma exatidão analítica sobre a quantidade de fertilizantes e adubação necessários para determinada quantidade de nutrientes encontrados no solo.

Os parâmetros base de NPK, níveis recomendados de cada componente químico e demais resultados encontrados no manual de calagem para cada tipo de colheita e solo, são dessa forma oriundos de vários estudos e calibrações referentes ao rendimento de cada nutriente quando aplicado ao solo, no qual desde 1985 novos parâmetros de correção de solo e adubação são definidos. Isso ocorre principalmente por meio de ensaios e experimentos de campo feitos em conjunto com as instituições participantes do ROLAS e também pelos resultados encontrados por meio da publicação de artigos referente ao assunto e congressos de ciência do solo.

Com base nesse conceito, para cada tipo de nutriente existente no solo há uma tabela no manual de calagem demonstrando em qual faixa determinado resultado da análise de solo se encaixa, podendo ser representado por um número de 1 a 10 ou por meio de porcentagem e sendo classificado em muito baixo, baixo, médio, alto e muito alto, variando de acordo com o tipo de nutriente.

Na figura 4 é possível ver a tabela de índices com base na porcentagem de argila, matéria orgânica e troca de cátions encontrados no solo, onde cada faixa de valores possui a sua determinada classe que pode ir de 1 a 4. Essa classe é utilizada posteriormente para calcular o teor de fósforo do solo, visto que

ambos possuem relações nos resultados e interferem diretamente nos índices de recomendação.

Figura 4: Interpretação de teores de argila e de matéria orgânica e da capacidade de troca de cátions (CTC) a pH 7,0

Argila		Matéria orgânica		CTC _{pH 7,0}	
Faixa	Classe	Faixa	Interpretação	Faixa	Interpretação
%		%		cmol _c /dm ³	
≤ 20	4	≤ 2,5	Baixo	≤ 5,0	Baixo
21 - 40	3	2,6 - 5,0	Médio	5,1 - 15,0	Médio
41 - 60	2	> 5,0	Alto	> 15,0	Alto
> 60	1	-	-	-	-

Fonte: SBCS (2016)

Tabelas como a da figura 4 são encontradas por todo o manual de calagem e adubação, visto que existe uma para cada tipo de nutriente de solo. Além disso, as mesmas possuem relações entre si, que interferem nos valores finais das faixas de classificações. Dessa forma, para auxiliar os agricultores e engenheiros agrônomos na tomada de decisões existem alguns *softwares* disponíveis que recebem como entrada uma análise de solo e resultam na recomendação de nutrientes baseando-se nessas tabelas do manual de calagem e adubação, dentre esses *softwares*, um deles é o CADUB.

2.1.3.3 CADUB

O CADUB é um *software* de plataforma *desktop* distribuído pelo Departamento de Solos da Universidade Federal de Santa Maria - RS e desenvolvido na época em um Trabalho de Conclusão de Curso pelo atual doutor em Ciência do Solos Paulo Ivonir Gubiani juntamente com os professores orientadores Leandro Souza da Silva e Dalvan José Reinert.

Segundo Gubiani, Silva and Reinert (2015), o *software* foi desenvolvido com o intuito de proporcionar um auxílio aos agrônomos para realizar o cálculo de fertilizantes e corretivos para adequar o solo da área em questão para que dessa forma se obtenha o máximo de aproveitamento do solo e conseqüentemente uma maior produtividade. Ele foi desenvolvido utilizando o *software* Microsoft Excel e

para sua interface foi utilizado o VBA (*Visual Basic for Applications*).

Tomando como base os Manuais de calagem e adubação dos estados do Rio Grande de Sul e Santa Catarina, o mesmo gera as devidas recomendações para o solo de acordo com as entradas de uma análise de solo.

O funcionamento do *software* se dá por meio de condições lógicas de *IF ELSE*, tendo como base o manual de calagem de adubação do ano de 2014. Para realizar uma recomendação de solo primeiramente é preciso cadastrar um cliente e então uma análise de solo, por meio da entrada manual de valores de nutrientes químicos encontrados no solo. A recomendação de adubação gera como resultado a fórmula do adubo NPK (Nitrogênio, Fósforo e Potássio) e sua respectiva quantidade que deve ser aplicada ao solo.

Depois de sua origem, o *software* sofreu uma única atualização, sendo disponibilizada por meio da versão 2.1 e tendo como principais melhorias a possibilidade de recomendação de adubos orgânicos, uma seleção mais equilibrada de formulas NPK no cálculo da necessidade de adubação e também a possibilidade de importação de análises por meio de uma planilha do Excel, permitindo que mais de uma análise seja importada ao mesmo tempo.

2.1.4 Custeio agrícola

O custeio agrícola, segundo Abramovay (2004) tem como intuito auxiliar nos custos de produção que são de vital importância para cumprimento do calendário agrícola, como por exemplo, a contratação de serviços de terceiros para preparar o solo antes das primeiras chuvas.

Este é voltado principalmente para oferecer um crédito para o ressarcimento de despesas do dia a dia, como por exemplo, a colheita. De acordo com Brasil (2018) quem pode se beneficiar desse custeio é o produtor rural, que pode realizar a solicitação diretamente com o banco ou então por meio de cooperativas de crédito rural, assim como, as cooperativas de produtores rurais e suas centrais, quando for feito para interesses próprios e ainda agroindústrias que possuem suas atividades voltadas para o suinocultura e avicultura de corte.

Esse custeio pode ser renovado de forma simples, em até cinco safras,

podendo ser efetuada em até 180 dias antes do término da data de vencimento do custeio anterior, limitado a 70% da receita bruta deduzida para o empreendimento a ser financiado, tendo como prazo para realização em até 2 anos, levando em conta a cultura ou a atividade explorada.

Alguns itens que são de vital importância para aquisição do custeio, como por exemplo o SEAF (Seguro da agricultura familiar), que segundo CIVIL (2018) para realizar a contratação ou renovação do financiamento, o mesmo deve-se atentar para a parte de zoneamento agrícola, solo, semente, sistema de produção/nível de tecnologia, croqui, orçamento, entre outros. Importante destacar o item de solo, onde para estar adequado nesse quesito, é necessário realizar a análise de solo da área a ser cultivada, assim como também deverá acompanhar a recomendação de uso de insumos para correção do solo.

2.2 INTELIGÊNCIA ARTIFICIAL

De acordo com LUGER (2004), a Inteligência Artificial é um setor da ciência da computação que tem por objetivo criar dispositivos capazes de simular a capacidade humana de raciocínio, percepção, tomada de decisão e resolução de problemas, buscando aprender seus erros e acertos e melhorar seus resultados gradativamente.

Durante muitos anos a IA (Inteligência artificial) foi vista como uma área bastante teórica onde suas aplicações aconteciam apenas em pequenos problemas curiosos e desafiadores, porém de pouco valor prático. Dessa forma, muitos dos problemas maiores eram resolvidos diretamente por meio da codificação em uma linguagem de programação específica, onde se juntavam estudiosos da área ao qual o problema de propunha e desenvolviam a lógica do sistema.

Foi a partir da década de 1970 que a IA começou a ser mais utilizada, principalmente em conjunto com o uso de técnicas de computação para solução de problemas reais, onde inicialmente os problemas recebiam conhecimento de especialistas para sua solução e posteriormente essa informação era codificada.

Com o crescimento cada vez maior dos problemas computacionais e da grande quantidade de dados gerados pelos diferentes setores se tornou fundamen-

tal a necessidade de ferramentas computacionais mais robustas e que agissem por conta própria de forma autônoma, buscando reduzir ao máximo a necessidade de intervenção humana e dependência de especialistas da área. Dessa forma buscou-se técnicas capazes de por si só, criar uma hipótese ou função capaz de resolver determinado problema que se deseja tratar.

A área de IA também pode ser classificada em IA fraca e IA forte. A IA fraca aborda o desenvolvimento de sistemas como algo inteligente, mas que não tem capacidade suficiente para raciocinar por conta própria, como exemplo desse tipo de inteligência pode-se imaginar um chatbot, que nada mais é do que várias expressões de decisões encadeadas, não possuindo nenhum raciocínio ou vontade própria, visto que nesta situação a máquina ainda se baseia na informação que o humano lhe deu.

Enquanto isso, a IA forte tem por objetivo a criação de máquinas que tenham consciência e possam pensar por si próprias e não apenas simular raciocínios. Como exemplo deste tipo de inteligência é possível citar um *software* que seja capaz de escrever uma música ou uma poesia.

2.2.1 Teste de Turing

Dentre um dos principais nomes da história da IA está Alan Turing, um matemático e cientista da computação que depois da guerra iniciou com sua ideia de criar um computador que pensasse, e logo em seguida criou o teste de Turing como forma de avaliar o sucesso ou não de um computador pensante.

O teste de Turing, criado em 1950, se baseia na ideia de uma pessoa interrogar um computador e com base nisso definir se o computador é inteligente ou não. Inicialmente o interrogador é colocado no meio de dois indivíduos, onde um é humano e o outro, um computador. Após isso, o interrogador começa a fazer perguntas para ambos os indivíduos, mas sem nenhuma interação direta. Com o início da conversa ambos os indivíduos podem tentar enganar o interrogador para provar quem é humano e quem é a máquina.

Este tipo de teste já foi utilizado por vários cientistas de TI para treinar seus programas de inteligência artificial e processamento de linguagem natural, porém

até hoje nenhum programa conseguiu passar pelo teste e prêmios em dinheiro não oferecidos regularmente para quem conseguir criar um programa que se passe realmente por um humano.

2.2.2 Lógica difusa

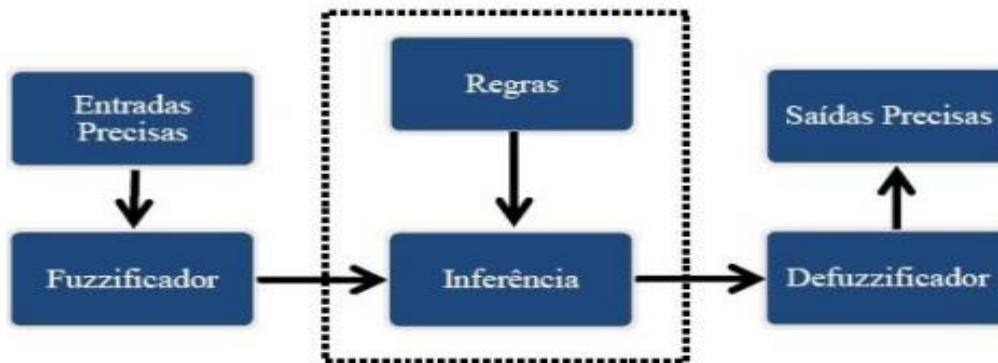
Apresentada inicialmente por Zadeh em 1965 como forma de pesquisa para o desenvolvimento da inteligência computacional, a lógica difusa, ou teoria dos conjuntos nebulosos aborda a implementação de graus de incerteza, ou seja, defende a teoria de que um elemento possa pertencer de forma parcial a um conjunto, onde os mesmos são definidos por meio de uma função de pertinência, essa, estando associada ao grau de confiança sobre um determinado número.

Conforme SIMÕES (2007), trata-se de uma técnica que se associa a natureza humana de pensamento como forma de controlar um sistema, sendo que é possível criar um modelador difuso particular para trabalhar de forma similar ao raciocínio dedutivo, ou seja, a forma com que o humano se utiliza para corroborar suas conclusões fundamentadas em informações já estudadas e conhecidas.

Dentro desse conceito, as funções e conjuntos difusos de pertinência são declarados para cada variável, tanto de entrada, quanto de saída, onde para que o sistema funcione é preciso converter os valores reais de entrada em valores difusos para que assim, seja possível realizar a ativação e combinação das informações conforme o conjunto de dados que foi fornecido pelo sistema, para que, em seguida os conjuntos difusos de saída se transformem em uma saída de sistema precisa.

De acordo com Mendel (1995), para se criar uma modelagem difusa existem duas formas, pode ser desenvolvido um novo sistema, ou pode-se utilizar um *software* já desenvolvido que tenha incluso o modelo matemático como ferramenta. A construção do modelo acontece por meio de três etapas, a primeira, chamada de fuzzificação, é responsável por converter os valores numéricos de entrada em valores difusos, enquanto isso, a segunda etapa, também chamada de inferência tem por papel definir os valores de saída com base nos sistemas de regras pré-estabelecidas. Por fim, ocorre a defuzzificação, onde os valores difusos são convertidos novamente em valores numéricos, fluxo esse, ilustrado na figura 5.

Figura 5: Desenvolvimento computacional difuso



Fonte: Mendel (1995)

Segundo KLEIN and WEBER (2003), os sistemas difusos hoje são empregados em várias áreas, como em *softwares* para controle de aeronaves e na área de biomedicina, também é aplicado como forma de detectar a presença de vacas leiteiras no cio na área da agropecuária. O autor Ferreira (2018) afirma os benefícios de uso dessa técnica para auxiliar na tomada de decisões que possuem incertezas nas afirmações de “sim” e “não”.

2.2.3 Sistemas especialistas

Sistemas especialistas são sistemas que surgiram na década de 70 por meio do esforço exercido por pesquisadores, que tinham o intuito de criar programas computacionais que fossem capazes de ter um raciocínio semelhante ao ser humano. Conforme WELBANK (1983), trata-se de um *software* de computador que possui um amplo conhecimento sobre uma área específica, e que, por meio da utilização do raciocínio consegue realizar atividades, até então, executadas por especialistas humanos.

Para que este tipo de sistema atinja o objetivo ao qual é proposto o mesmo deve interagir com o usuário da mesma forma que um especialista humano o faria, por exemplo, escutando o usuário, e suspendendo perguntas cuja a resposta pode ser deduzida. Sistemas especialistas também devem ser capazes de tirar conclusões, mesmo que as informações fornecidas a ele não sejam totalmente completas.

De acordo com RICH and KNIGHT (2016), dentre as principais caracterís-

ticas que um sistema especialista deve ter em relação a interação com o usuário, destacam-se:

- Explicação do raciocínio: trata-se de explicar de forma clara e concisa qual a solução para o problema e garantir que a mesma é adequada ao problema, ou seja, o sistema deve ser capaz de explicar o raciocínio utilizado que o fez chegar aos resultados apresentados.
- Obtenção e atualização de conhecimento: esta característica defende a ideia de que o sistema especialista precisa estar sempre atento a novas informações que possam modificar ou complementar o seu conhecimento, mantendo dessa forma, a sua base de dados sempre atualizada.
- Manter interações entre o especialista humano: trata-se de submeter dados brutos utilizados como base para o especialista humano e fazer com que o sistema especialista tente aprender com ele.

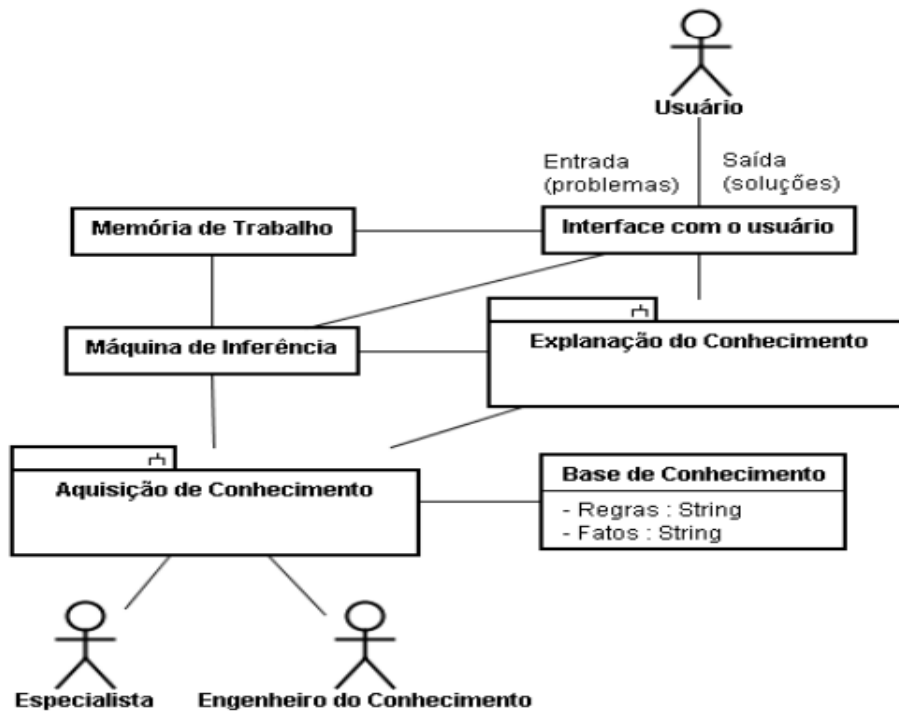
É importante frisar, que o que torna um sistema especialista vivo é a sua base de conhecimento, pois é a partir da mesma que o conhecimento é fornecido para a resolução de problemas. Este conhecimento pode estar em vários formatos, como fatos, heurísticas, experiências, opiniões, julgamentos e previsões e conforme Metaxiotis and Ergazakis et al. (2003) geralmente é coletado por meio de métodos de entrevista e análise de protocolos e questionários.

Além das características citadas acima, outra que chama atenção neste tipo de sistema é a presença de um mecanismo de raciocínio incerto que é capaz de representar a incerteza em relação ao conhecimento de determinado assunto. Com isso, como forma de expressar esse tipo de incerteza foram desenvolvidos diversos métodos para representação do conhecimento, são eles, o método de lógica, redes semânticas e quadros ou frames.

Para melhorar o entendimento deste tipo de sistema, é possível ver uma representação do mesmo na figura 6.

Conforme figura 6, é possível perceber a arquitetura de subsistemas que fazem o papel de aquisição e explanação do conhecimento, bem como os módulos,

Figura 6: Arquitetura de um sistema especialista



Fonte: Adaptado de Fowler (2000)

esses responsáveis pela base de conhecimento, inferência, memória de trabalho e pôr fim a interface do usuário.

Inicialmente a base do conhecimento extrai o conhecimento do especialista conforme o método de representação definido. Em seguida a etapa de inferência fica responsável por examinar o conteúdo da base de conhecimento e então definir qual a melhor ordem em que as inferências acontecem. Desta forma, com base em uma consulta de usuário, a etapa de inferência realiza a transferência dos fatos e regras para a memória de trabalho, que por fim, armazena os fatos e regras mais recentes.

Enquanto isso, a atualização da base de conhecimentos é realizada pelo subsistema de aquisição por meio da interação direta com o especialista, e o subsistema de explicação descreve o raciocínio do sistema para o usuário, detalhando as informações para que ocorra a obtenção de um resultado satisfatório.

2.2.4 Redes Neurais Artificiais

Desenvolvidas na década de 40 pelo neurofisiologista Warren McCulloch, do MIT, e pelo matemático Walter Pitts, da Universidade de Illinois, as Redes Neurais Artificiais (RNA), de acordo com Medeiros (1999) trazem a ideia de tratar o cérebro como um organismo computacional, por meio da tentativa de simulação do cérebro humano e seu comportamento quanto a eventos, como o aprendizado, erros cometidos e novas descobertas.

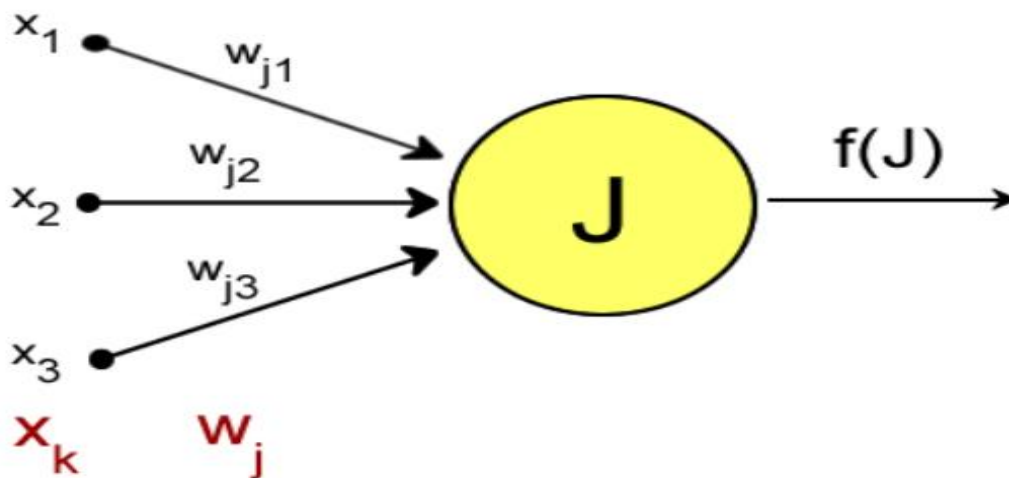
Totalmente diferente da abordagem de sistema especialistas, este tipo de tecnologia não usa uma base de conhecimento explícita para aprender, mas sim um conjunto de relações derivadas entre dados, dessa forma, conforme Metaxiotis and Ergazakis et al. (2003), afirmar que as RNA possuem conhecimento sobre um domínio específico é uma mentira.

A composição das RNAs se dá por meio de um aglomerado de neurônios artificiais interconectados entre si formando uma rede capaz de resolver problemas, onde a técnica para resolução é parecida com a usada pelos humanos. Cada neurônio possui uma estrutura lógico-matemática análoga com base em uma unidade de processamento capaz de comportar e combinar estímulos de muitos outros neurônios artificiais, procurando simular a forma, comportamento e funções de um neurônio biológico.

Na figura 7 é ilustrado o modelo básico de um neurônio j com entrada x_k , pesos sinápticos w_j , nível de ativação J e saída $f(J)$. Nesse caso, cada neurônio j possui um vetor com dados de entrada $x_k = [x_1, x_2, x_3, \dots, x_n]^T$, uma ativação interna J , $f(J)$ como sendo a função de ativação e os pesos sinápticos representados por $W_j = [W_{j1}, W_{j2}, W_{j3}, \dots, W_{jd}]^T$, conectando os elementos de X_k ao neurônio j .

Outro ponto que merece atenção é quanto ao aprendizado das redes neurais; o mesmo acontece quando existem modificações significantes nas sinapses entre os neurônios. Uma sinapse é caracterizada pela conexão existente entre os neurônios, onde as mesmas recebem valores chamados de pesos sinápticos, utilizados para se armazenar conhecimento. Para saber se determinada modifica-

Figura 7: Modelo básico de um neurônio



Fonte: Metaxiotis and Ergazakis et al. (2003)

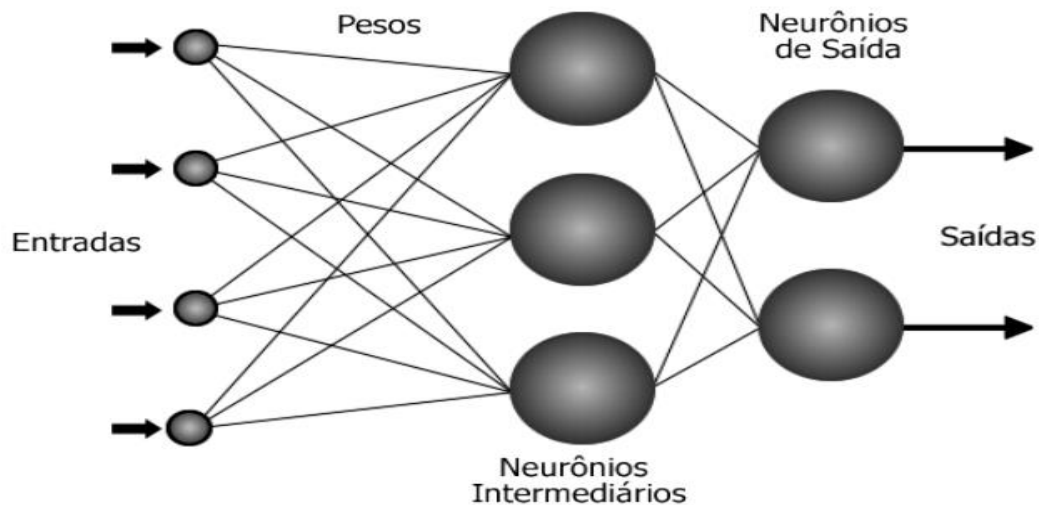
ção ocorrida é significativa ou não é verificado quais as conexões que são mais utilizadas, essas então são fortalecidas enquanto que as menos utilizadas são enfraquecidas.

De acordo com Metaxiotis and Ergazakis et al. (2003), logo após a implantação de uma rede neural artificial em uma aplicação é preciso que se espere um período para que esta rede seja treinada, já que assim como acontece nos sistemas biológicos, para o aprendizado são necessários ajustes nas conexões que existem entre os neurônios. Em outra abordagem, entende-se que a informação que foi aprendida é alocada em forma de valores numéricos denominados pesos, que são então designados para as conexões entre o processamento dos neurônios da rede.

As RNAs também podem ser distinguidas quanto a sua topologia, ou seja, pelo número de conexões, elementos de processamento e camadas existentes bem como também pelas leis de aprendizagem a qual foram expostas. Na figura 8 é apresentado uma rede neural com 3 camadas.

Atualmente, segundo Metaxiotis and Ergazakis et al. (2003), a principal vantagem obtida por meio dessa tecnologia é a possibilidade que a mesma tem para trabalhar com dados incompletos para gerar e demonstrar a sua intuição aparente. Além disso, com a utilização de RNAs torna-se possível trabalhar de forma

Figura 8: Rede neural com 3 camadas



Fonte: Metaxiotis and Ergazakis et al. (2003)

análoga com o cérebro, podendo ser potencializado dentro de um arcabouço de gerenciamento de conhecimento.

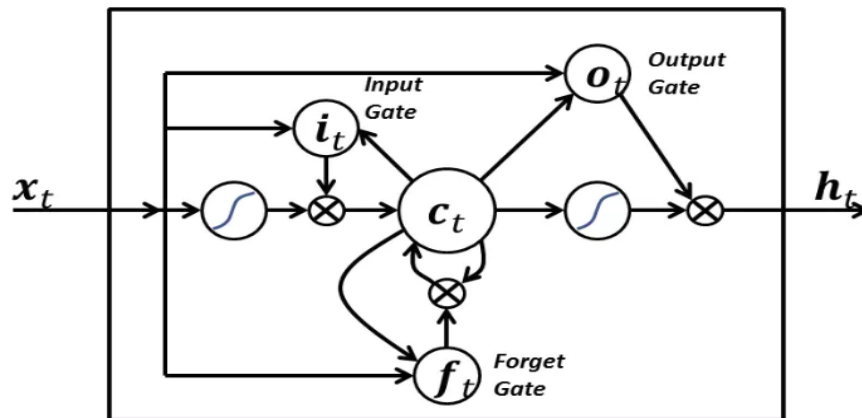
2.2.4.1 Redes Neurais LSTM

As redes neurais *Long Short-Term Memory* se baseiam em um modelo profundo e recorrente de redes neurais, sendo introduzidas na área da Inteligência Artificial em 1997 por Hochreiter e Schmidhuber por meio da motivação em oferecer um melhor desempenho para resolver o problema de desaparecimento de gradiente, que até então as redes neurais recorrentes sofriam, ao lidar com uma grande massa de dados sequenciais.

Para buscar resolver esse problema essas redes fazem com que o fluxo de erro seja constante por meio de unidades especiais intituladas de "portões" que tem por objetivo ajustar os pesos da mesma maneira que um truncamento da sequência no momento em que a informação não é mais necessária, caracterizando assim um esquecimento, na figura 9 é possível ver uma ilustração da lógica de uma rede LSTM.

Conforme é possível ver na figura 9 estes portões atuam sobre os sinais recebidos e decidem se bloqueiam ou transmitem a informação recebida levando em consideração a sua força e importância que são filtrados com seu próprio conjunto

Figura 9: Funcionamento de uma rede LSTM



Fonte: Greff and Srivastava et al. (2017)

de pesos. Estes pesos são adaptados por meio do processo de aprendizagem das redes recorrentes. Em suma, as células buscam aprender quando podem permitir que os dados entrem, saiam ou sejam removidos por meio de um processo totalmente iterativo de fazer suposições, calculando o erro durante a retropropagação e assim ajustando os pesos da rede durante a descida do gradiente.

De acordo com Graves (2012) este tipo de rede neural está sendo muito utilizado, sendo capaz inclusive de obter um dos melhores resultados se comparado com demais métodos, tendo maior destaque especialmente em áreas como processamento de linguagem natural e no processamento de caligrafia.

2.2.5 Aprendizagem de máquina

A aprendizagem de máquina ou (AM) é um ramo dentro da inteligência artificial que se compreende pelo processo de indução de uma possível hipótese ou aproximação de uma função a partir de uma experiência passada e que busca com isso solucionar um determinado problema que recebeu como entrada. Segundo Mitchell (1997), a AM se define pela capacidade de aprimorar o desempenho na execução de determinada tarefa por meio da experiência.

De acordo com Sen and Weiss (1999) o aprendizado e a inteligência são duas áreas intimamente conectadas uma a outra, sendo que geralmente há uma concordância de que um sistema que tenha capacidade de aprender possa ser atribuído como inteligente e também de forma recíproca que um sistema que seja

inteligente tenha capacidade de aprender.

Na AM os computadores são treinados para aprender com base em uma experiência anterior e dessa forma empregam um princípio de inferência chamado de indução, onde são obtidas conclusões genéricas a partir de uma coleção de exemplos. Dessa forma os algoritmos de AM aprendem a realizar a indução de uma hipótese ou atividades capazes de solucionar um problema por meio de dados que descrevem instâncias do problema ao qual se quer resolver, formando assim um conjunto de dados.

A capacidade de aprendizado é considerada uma tarefa fundamental para que se tenha um comportamento inteligente, dessa forma atividades de memorização, observação e exploração de situações são essenciais para aprender fatos, melhorar habilidades e organizar o conhecimento recém aprendido.

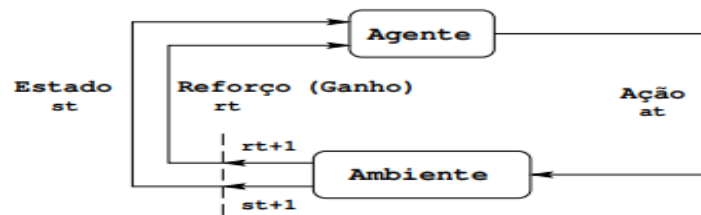
2.2.5.1 Aprendizagem por reforço

De acordo com Bellman (2010), a aprendizagem por reforço é uma função de aprendizado de máquina não supervisionado que tem como principal objetivo realizar a construção de um de algoritmos capazes de realizar o aprendizado por meio da interação de determinado agente com um ambiente, baseando para isso, nos conceitos matemáticos de desenvolvimento dinâmico.

A utilização desta metodologia é recomendada quando não existem modelos de dados disponíveis ou quando não é possível adquirir exemplos apropriados das situações pelas quais o agente irá enfrentar. Conforme Júnior and Chagas Araújo de (2009), o agente tenta aprender de maneira autônoma e ativamente a melhor política de atuação, usando para isso técnicas de experimentação direta, ou seja, o mesmo não é ensinado por meio de exemplos fornecidos por um supervisor. Conforme pode ser visto na figura 10.

O retorno das ações não pode ser perfeitamente antecipado, visto isso é importante que o agente monitore o ambiente frequentemente e toma uma atitude apropriada. Desta forma, em um sistema de AR, é possível representar o estado do ambiente por vários fatores, dentre eles pelo conjunto de variáveis de estado percebidas pelo agente; pelo conjunto de ações discretas, que ao serem escolhidas

Figura 10: A interação agente-ambiente na Aprendizagem por Reforço



Fonte: Sutton and Barto (1998)

por um agente tem o poder de mudar o estado do ambiente; e também pelo valor das transições de estados, estes que são passados ao agente por meio de um sinal de reforço, também chamado de ganho.

O principal objetivo do método é possibilitar que o agente escolha a sequência de ações que tendem a aumentar a soma dos valores de reforço, ou seja, realizar o encontro da política definida como o mapeamento de estados através de ações, com o intuito de maximizar as medidas de reforço que são acumuladas de tempo em tempo. Diante disso, existem cinco partes fundamentais da aprendizagem por reforço que devem ser utilizadas para que esse cenário se contemple:

- O ambiente: os sistemas de AR têm como característica aprender um mapeamento de cenários e situações por meio da experiência em um ambiente dinâmico, dessa forma o ambiente ao qual o mesmo está inserido precisa ser ao menos parcialmente observável, seja por meio de sensores e descrições simbólicas ou até situações mentais. Também existe a possibilidade de que toda a informação de valor do ambiente já esteja disponível, neste caso, o agente poderá definir ações baseadas em características reais do ambiente.
- A política: essa parte tem por papel representar o comportamento que o sistema AR usa para alcançar determinado objetivo, dessa forma caso um agente mudar sua política então também acontecem mudanças nas probabilidades de seleção de ações, afetando também o comportamento do sistema conforme o agente for acumulando experiência a partir das interações com o ambiente.
- Reforço e retorno: trata-se de um sinal do tipo escalar devolvido ao agente pelo ambiente logo após a execução de uma ação e uma transição de estado.

A forma de definição do reforço para cada transição varia, gerando assim funções de reforço que buscam expressar o objetivo que o sistema AR deve alcançar. Outro papel do agente nesta etapa é o de maximizar a quantidade total de reforços recebidos, também chamado de retorno. É importante frisar que essa maximização de esforço nem sempre se refere ao reforço imediato a receber, mas também pode ser o reforço acumulado durante uma execução total do algoritmo por exemplo.

- Função de reforço: geralmente essas funções são um tanto complicadas de interpretar, porém dessas, geralmente existem 3 classes de problemas mais utilizadas para criar funções adequadas.

Reforço só no estado final: nesta classe de funções todas as recompensas são zeradas, exceto no estado final, onde o agente passa a receber uma recompensa real (+1) ou uma penalidade (-1). Visto que o objetivo é maximizar o reforço, o agente irá passar a aprender que a relação entre os estados e as recompensas são coisas boas e os que levam a uma penalidade precisam ser evitados.

Tempo mínimo ao objetivo: as funções de retorno dessa classe possibilitam que o agente faça ações que criem o caminho mais curto para determinado estado objetivo. Nesta classe também toda a ação tem penalidade (-1), porém o estado final é (0). Desta forma o agente aprende a escolher ações que minimizem o tempo gasto para alcançar o estado final.

Minimizar reforços: não é sempre que ocorre a maximização da função de reforço pelo agente, pois também podem haver situações que o agente deve aprender a minimizá-las. Esta técnica se torna útil no momento que o reforço é uma função que possui recursos limitados e o agente deve aprender uma forma de conservá-los, caso contrário não atingirá seu objetivo.

Função valor-estado: este tipo de função tem por objetivo realizar um mapeamento de estado, ou par estado-ação por meio de um valor obtido a partir do reforço atual e reforço futuro.

2.2.5.2 *Aprendizagem supervisionada*

De acordo com Osório (1999) o fundamento básico desse tipo de aprendizagem é o de conhecer por meio de entradas e saídas quais respostas devem ser fornecidas ao sistema para cada entrada ou impulso externo. Dessa forma, quem terá esse conhecimento será um supervisor (humano), que será capaz de verificar os erros produzidos nas saídas da rede e a alimentando com os valores corretos para cada entrada ajustando assim os parâmetros da rede.

Nesta técnica é possível admitir que o aprendizado estará completo no momento em que não existir mais erros nas saídas, ou o tamanho do erro for muito pequeno, a partir desse momento passa a se dizer que o sistema conseguiu obter o conhecimento ensinado pelo supervisor, estando então preparado para o problema que lhe foi apresentado.

2.2.5.3 *Aprendizagem não supervisionada*

Neste modelo de aprendizagem não existem saídas desejadas e por esse motivo, diferente do modelo supervisionado, o mesmo não precisa de ‘professores’ para o seu treinamento visto que o aprendizado da rede acontece somente através dos valores de entrada, onde a rede processa essas entradas e então realiza a organização em categorias, usando para isso, critérios próprios. Neste caso, para cada entrada aplicada a rede, será retornada uma resposta apontando a classe à qual a entrada faz parte, onde caso o padrão de entrada não pertença as classes existentes será gerada uma nova classe.

2.2.5.3.1 *Árvores de decisão*

De acordo com Su and Zhang (2006), esta técnica de aprendizado permite a modelagem de sistemas mais discretos onde o modelo resultante é uma árvore de decisão. Este tipo de método é um dos mais utilizados hoje para realizar a inferência indutiva, já que se trata de um componente robusto quanto a questão de ruído de dados, além de ser de fácil compreensão, já que permite uma representação gráfica do modelo gerado. Outros pontos que reforçam a popularidade do método é a sua simplicidade na montagem de árvores e a eficiência nos algorit-

mos de escalamento que conseguem lidar facilmente com um aumento na base de dados.

A estrutura de uma árvore de decisão consiste em vários nodos internos que se correspondem a testes sobre determinado atributo, onde cada ramo descendente tem a representatividade de uma possibilidade para cujo teste e cada folha que contém sua respectiva classe é a decisão obtida após testar os atributos em forma sequencial.

2.3 TECNOLOGIAS

Para atingir os objetivos e desenvolver o algoritmo de reconhecimento proposto, se buscou estudar as principais ferramentas da área de processamento de imagens e reconhecimento de imagens por meio da utilização de tecnologias de *Machine Learning*. Diante disso, se aprofundou o estudo da área de visão computacional por meio das ferramentas OpenCV e da área de *Machine Learning* e reconhecimento de texto por meio das ferramentas TensorFlow, Tesseract e Google Vision API. Além disso, também se buscou estudar o funcionamento das aplicações móveis bem como tecnologias de armazenamento como banco de dados NOSQL MongoDB e processamento e armazenamento *Cloud Computing*.

2.3.1 Aplicações móveis

A computação móvel, segundo Garg (2010) é a computação que possibilita o acesso contínuo a recursos remotos por meio da utilização de dispositivos como smartphones, tablets e celulares tendo como característica principal o uso da tecnologia de transmissão de dados sem fio.

Conforme o autor, um dos principais objetivos de todo dispositivo móvel é permitir que sua utilização aconteça em qualquer lugar que o usuário se encontre, diminuindo assim a necessidade do mesmo se deslocar até um ponto de conexão cabeada e para que isso seja possível o dispositivo precisa fazer o uso de sinais eletromagnéticos que se propagam pelo ar para se conectar a uma rede.

A característica de mobilidade também é defendida por Talukder and Yavagal (2007), onde o mesmo afirma a importância do usuário conseguir se mover de

um lugar para o outro sem nenhuma interferência e podendo utilizar os mesmos serviços em lugares diferentes. Com base nessa teoria o autor abordar algumas das características que juntas formam um cenário ideal para um ambiente mobile.

2.3.1.1 Características de um ambiente mobile

Dentre as principais características para se obter um ambiente mobile de qualidade está a mobilidade do usuário, que compreende a possibilidade de um determinado usuário se mover de um local físico a outro e continuar usando os mesmos serviços. Outro aspecto importante é a mobilidade da rede, que permite que o celular consiga se conectar em outra rede quase que instantaneamente e de forma transparente, sem perder a cobertura de sinal.

Além das duas características já abordadas, também existe a mobilidade de serviço e a mobilidade de sessão, onde a primeira garante que determinado serviço utilizado no celular também possa ser utilizado em um computador e a segunda controla situações para que um usuário não perca o seu histórico de trabalho já realizado, ou seja, se determinado usuário está editando um texto online pelo celular e acaba por perder a sua conexão com a internet, caso o mesmo abra esse mesmo arquivo de texto em seu computador, ele deve continuar a edição de onde parou com o celular ao invés de precisar criar um novo arquivo em branco.

2.3.2 Bancos de dados NOSQL

Um novo conceito de armazenamento de dados, surgiu devido ao crescimento da quantidade de dados que precisam ser gerenciados, o chamado *Big Data*, com uma diferença fundamental entre os bancos relacionais, que é o processamento em memória ram, por esse motivo, tendem a ser banco de dados muito mais performática.

2.3.2.1 MongoDB

O MongoDB é um banco de dados NOSQL escrito na linguagem C++ e disponibilizado pela empresa 10gen por meio de licença GPL. De acordo com Cattel (2011) o banco pode ser considerado similar ao CouchDB, porém com uma grande diferença de possuir suporte nativo a sharding, o que permite a distribuição

de dados em várias máquinas, por exemplo através de *clusters*. Outro diferencial deste banco é a sua utilização de índices de consulta de forma automática, similar a um RDBMS, e também ao seu efetivo controle de concorrência, fornecendo operações atômicas nos campos que realizam alteração na memória local, dessa forma, são verificados se os mesmos documentos não sofreram alteração por outros usuários. Segundo MongoDB (2018), O MongoDB é reconhecido por ser o representante mais conhecido dentro da categoria de bancos de dados NOSQL.

De acordo com Yuhanna, leganza and Austin (2016), os documentos submetidos a base de dados são salvos em formato BSON porém os clientes controlam a conversão para o formato JSON. Fowler and Sadalage (2012) destaca, que o mongoDB segue à risca o conceito de NOSQL e dessa forma não oferece nenhum suporte padrão a relacionamentos entre documentos e coleções, defendendo sempre a de-normalização, paradigma esse, em que, segundo Copeland (2013), os dados são salvos da mesma forma como serão necessário para uma posterior consulta, buscando assim maximizar o desempenho das buscas.

Segundo Fowler and Sadalage (2012), os bancos do tipo documento se assemelham a bancos do tipo chave-valor, a única diferença é que, neste caso, e ao contrário dos bancos SQL tradicionais, estes não têm um esquema ou estrutura pré-definida. Além disso, Strauch (2011) destaca que o banco não tem chaves estrangeiras, onde então o relacionamento é feito de forma manual através de um campo de referência que serve para informar o caminho completo do documento atual até o documento de destino. Mesmo assim, um dos jeitos mais eficiente usado para definição de relação entre documentos é o de aninhamento, onde um documento é colocado dentro de outro.

2.3.3 Cloud computing

Considerada como a nova arquitetura de TI, segundo Veras (2013), o *cloud computing*, é a evolução dos conceitos baseados em mainframe e cliente/servidor. Antes era necessário realizar grandes investimentos para aquisição de servidores poderosos, assim como também possuía altos custos para manutenção, com o novo conceito, a forma de compra de servidores, se ajusta de acordo com a ne-

cessidade de utilização, assim como o seu custo é somente calculado em torno da sua utilização, tornando assim o serviço muito mais dinâmico e flexível.

Sua configuração pode ser feita também de forma totalmente livre, podendo escolher de acordo com a necessidade, assim como, de forma fácil, realizar alterações na infraestrutura somente com configurações, não dependendo somente de especialistas na área.

A vantagem da utilização da nuvem não fica restrita somente para compartilhamento de arquivos, mas também significa a utilização de algo que estaria restrito a apenas um servidor, que poderia ser utilizado por um grupo de pessoas específicas, porém com o conceito de computação em nuvem, esse cenário muda, pois agora tudo pode ser compartilhado de forma fácil e rápida, podendo com isso a troca de ideias com pessoas ou grupos diferentes que pode estar do outro lado do planeta.

Quando a empresa, por exemplo, decide que irá utilizar a computação na nuvem, ela precisa realizar alguns planejamentos, de que forma ela vai utilizar a nuvem, se irá decidir a utilização de algum *software* que está na nuvem, alguma plataforma de desenvolvimento na nuvem, ou então, sua infraestrutura inteira de TI em forma de serviços na nuvem.

2.3.4 Tratamento de dados e modelagem preditiva

O tratamento de dados juntamente com a modelagem preditiva são áreas responsáveis pela extração, manipulação e modelagem visando transformar os dados em informações capazes de ser interpretadas visualmente, eliminando dessa forma dados desnecessários ou que não trarão nenhum benefício para a criação de um modelo preditivo.

2.3.4.1 Weka

De acordo com Eibe Frank and Witten (2016), o Weka é uma ferramenta *open source*, utilizado para mineração de dados, esta ferramenta foi desenvolvida em Java, dentro de regras que a GPL propõe, que ganhou espaço principalmente entre professores e estudantes de faculdades. Essa ferramenta auxilia no processo

de aprendizado de profissionais sobre as teorias básicas sobre mineração de dados. Ele tem uma interface gráfica para tornar a mineração de dados de modo que seja o mais simples possível, oferecendo a funcionalidade de comparar algoritmos utilizados e ter um relatório de dados resultantes obtidos.

O formato de dados mais utilizado para se trabalhar no Weka, é o formato de texto, normalmente estruturados em arquivos nos formatos ARFF ou CSV, porém é possível a utilizar a ferramenta para realizar conexão com SGBD's relacionais, tornando-se possível a mineração de dados diretamente de tabelas. Esse método para trabalhar com os dados no Weka, chama bastante a atenção, pois torna o processo muito mais fácil, por não necessitar mais realizar conversões de dados, muitas vezes já em banco de dados relacionais, para outros formatos como CSV ou ARFF, que normalmente esse tipo de atividade consomem bastante tempo.

2.3.4.1.1 TensorFlow

O *software* TensorFlow disponibilizado de forma gratuita pela Google, a equipe responsável por esse projeto é a mesma que lançou o Google Tradutor em 2010. De acordo com Reads (2017), por cerca de dois anos, a partir da publicação do livro, esse *software* já usado pelo Google, em várias ferramentas de informática, como por exemplo, tradutores automáticos, em fotos, no motores de busca ou até no Gmail, assim como por empresa de grande porte que estão ligadas a IBM, a eBay, UBER, Intel, o Twitter, etc. De acordo com a própria Google, essa ação foi realizada com o propósito de fomentar a pesquisa na área dos processos de *Machine Learning*, disponibilizando assim meios tecnológicos para todos.

Esta ferramenta foi utilizada como forma de se tentar detectar na imagem os componentes químicos e seus valores por meio da submissão de um modelo treinado para detecção de números como o MNIST ou de letras como o EAST. Serão realizados testes com base nesses dois modelos juntamente com o auxílio da biblioteca OpenCV que possui um método específico que recebe um modelo treinado como parâmetro e retorna os pontos encontrados em uma imagem.

2.3.4.1.2 Google Vision

O *software* Google Vision, é um poderosa ferramenta para análise avançada de imagens, esta trabalha com modelo de *Machine Learning* pré treinadas, oferecendo uma API para fácil utilização pelos desenvolvedores, onde oferece várias funcionalidades, como por exemplo, o reconhecimento de objetos contidos na imagem, interpreta também categorias de objetos contidos nas imagens e também extrai letras e palavras contidas na imagem.

Segundo Cloud (2019), o reconhecimento de texto em imagens, pode ser subdividida em detecção de texto e documento de detecção de texto, onde basicamente a primeira busca encontrar texto em imagens do dia a dia, como por exemplo, uma placa na rua, ela irá retornar a descrição encontrada na placa e retornar a posição, os textos extraídos e as palavras separadas. Já na segunda, seu foco está voltado para a detecção extração de textos providos de documentos.

2.3.4.1.3 OpenCV

Conforme Vinay (2017) é uma biblioteca originalmente desenvolvida pela Intel que traz consigo funções que podem ser aplicadas ao ramo de visão computacional para sistemas Windows, Linux e Mac OS x. Dentre as aplicações da biblioteca estão o tratamento de imagens e aplicação de filtros corretivos, identificação de objetos, segmentação, reconhecimento de imagens e reconhecimento de bordas.

Neste projeto foi utilizada a ferramenta para tratar a imagem capturada pelo aplicativo mobile e aplicar filtros corretivos buscando identificar a tabela que contém os nutrientes químicos e seus devidos valores. Para tornar isso possível utilizaremos as funções de dilatação, erosão, limiarização e extração de contornos.

2.3.4.1.4 Tesseract

O Tesseract é uma biblioteca OCR estruturada com base na linguagem C/C++ e inicialmente desenvolvida pela HP de 1984 a 1994 em um laboratório de pesquisa para ser um possível substituto de scanners nas impressoras da companhia. Mesmo ganhando uma fama no início o projeto não se desenvolveu e

acabou sendo cancelado, dessa forma, em 2005 a HP liberou o *software* como *open Source* para a comunidade e conforme Smith (2007), desde então o mesmo é mantido pela Google.

Hoje em dia o Tesseract-OCR possui a licença Apache 2.0 estando disponível para sistemas Windows, Linux e MacOS X, além disso existe uma completa documentação sobre a forma de uso e configuração em cada tipo de sistema. O idioma padrão do *software* é o inglês porém hoje já existe a opção de configurar outras línguas, através do download de um pacote separado, inclusive o português, possibilitando assim a utilização de dicionários para auxiliar no reconhecimento de palavras que são correspondentes a cada idioma específico.

Neste projeto utilizou-se essa ferramenta para identificar o nome de cada componente químico presente na análise de solo, e com base nisso extrair o seu valor, a detecção da tabela se dará por meio da ferramenta OpenCV e após essa detecção será submetida coluna por coluna para reconhecimento pelo Tesseract.

2.3.5 Desenvolvimento

Este tópico tem por objetivo o estudo das ferramentas e processos de desenvolvimento de *software* existentes no mercado juntamente com as novas metodologias de desenvolvimento e linguagens de programação para a parte backend e frontend.

2.3.5.1 React Native

O React Native é uma biblioteca javascript baseada na biblioteca React do Facebook e é muito utilizada para desenvolver aplicativos móveis nativos para IOS e Android. De acordo com Eisenman (2016), com a ferramenta é possível criar soluções realmente autênticas e nativas, onde o desenvolvedor só precisa escrever código javascript que o compilador do React Native irá gerar o restante do código nativo tanto para Android quanto para IOS, aumentando dessa forma a produtividade da equipe e a facilidade de manutenção do código.

A criação das telas e interfaces de usuários por meio da biblioteca se dá através de uma tecnologia chamada de JSX (*XML LIKE SINTAX*), uma mistura de

javascript com XML. Após a desenvolvimento, o React Native utiliza de APIs de renderização nativas gerando então os objetos originais de cada linguagem, ou seja, Objective-C para ios e Java para android. Além disso também é possível acessar interfaces de javascript para APIs de plataforma, conseguindo com isso acessar recursos do hardware do telefone, como a câmera ou o microfone por exemplo.

A tecnologia JSX utilizada no React Native possui algumas diferenças quanto a nomenclatura de suas tags se comparado com linguagens de marcação como o HTML ou o React padrão. Na figura 11 é possível ver algumas das mudanças entre o React e o React Native.

Figura 11: Diferenças entre React e React Native

React	React Native
<code><div></code>	<code><View></code>
<code></code>	<code><Text></code>
<code>, </code>	<code><ListView></code>
<code></code>	<code><Image></code>

Fonte: Eisenman (2016)

2.3.5.1.1 Contribuições da tecnologia

O principal destaque do React Native quanto aos seus concorrentes é que o mesmo consegue gerar aplicações para duas plataformas distintas com o mesmo código fonte e ainda de forma nativa, sem a utilização de WebViews, o que acaba aumentando muito a performance da aplicação se comparado com outras tecnologias como Ionic ou Cordova que não trabalham com a geração nativa dos componentes, mas sim híbrida.

Outro fator importante é a forma de funcionamento do processamento das interações do usuário com o aplicativo, onde o React Native funciona totalmente separado da *thread* principal da interface, fazendo com que o app tenha um alto desempenho sem sacrificar todo o recurso.

A forma de desenvolver com React Native também é um ponto que agrada muitos desenvolvedores, primeiramente por causa da livre escolha da IDE de desenvolvimento por se tratar de código em javascript e por outro lado devido a sua funcionalidade padrão chamada de *hot reloading* onde a cada alteração no código o aplicativo é recarregado automaticamente no celular ou no emulador em que se está executando a aplicação como *debug*.

2.3.5.2 NodeJS

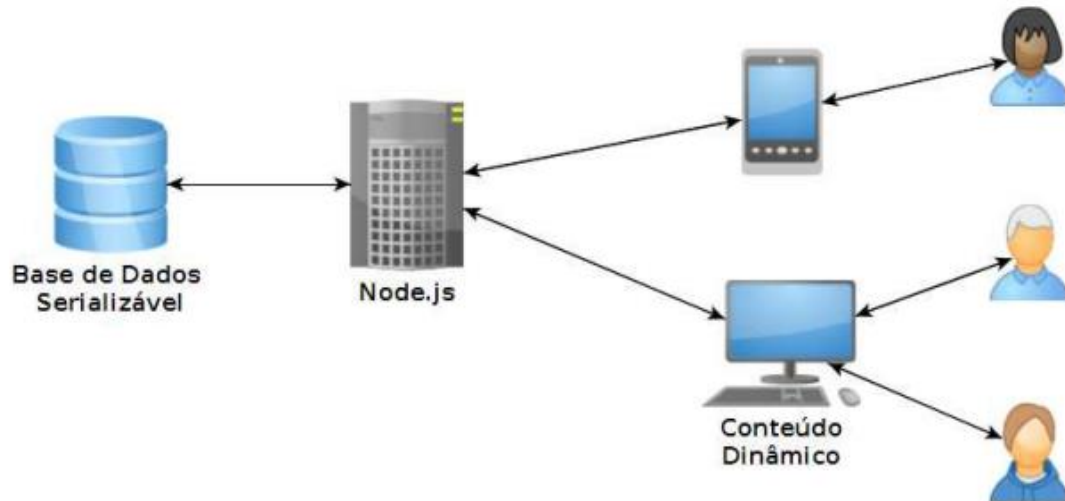
O Node.js pode ser caracterizado como um ambiente runtime por meio da utilização da linguagem JavaScript rodando sobre uma engine conhecida como Google V8, engine essa que foi criada pela Google e que é utilizada no navegador da web Google Chrome. Segundo PEREIRA (2013) o V8 é capaz de realizar a interpretação e compilação para a linguagem de máquina buscando otimizar a aplicação com o uso de heurísticas, além disso, o autor ainda destaca de que a arquitetura do Node.js é caracterizada por não bloqueante, fazendo com que ela tenha uma ótima performance quanto ao consumo de memória ram e poder de processamento.

Conforme PEREIRA (2013), dentre as principais características presentes na linguagem Node está a sua separação de processos em *single thread* onde cada instância terá instância de apenas um processo e também a programação assíncrona e compartilhamento de recursos para que se tenha um melhor proveito da *thread* que está sendo usada. Além disso, a linguagem ainda conta com operações de entrada e saída totalmente assíncronas e não bloqueantes facilitando dessa forma a execução em forma paralela e com total aproveitamento de recursos.

Outro ponto que se destaca quanto a utilização dessa tecnologia é o da facilidade de inicialização de um servidor sem exigir recursos de hardware excessivos, como espaço em disco e memória. Geralmente, a linguagem é muito utilizada para criações de *API Rest Full* auxiliando o desenvolvedor a fazer uma ligação consistente entre *front-end* e *back-end*. Na figura 11 é possível ver como um servidor em Node Js se enquadra em uma estrutura dinâmica com *front-end* web e *mobile*

incluídos.

Figura 12: Atuação do NodeJS



Fonte: Wilson (2013)

2.3.5.2.1 Forma de funcionamento

Para entender o funcionamento e o poder de processamento do node js, é possível realizar a comparação entre ele e uma plataforma web tradicional, onde inicialmente, em plataformas tradicionais é necessário que para cada requisição realizada seja criada uma nova *thread*, onde cada *thread* consome uma quantidade considerável de memória ram. Dessa forma caso uma aplicação tenha um grande fluxo de requisições seria necessário adquirir uma infraestrutura bastante consistente para que ela possa funcionar sem interrupções.

Por outro lado, o Node js funciona de uma maneira diferente, operando em uma única *thread* e utilizando chamadas de entrada e saída que não sofrem bloqueio, fazendo com que seja possível comportar milhares de requisições concorrentes e ainda diminuir pela metade o consumo dos recursos computacionais.

Conforme Wilson (2013), o trabalho que o Node JS exerce acontece através de um laço de eventos, também chamado de *Event Loop*, e pode ser descrito por uma sequência de passos. Primeiramente o Node JS é iniciado e o programa é carregado, em seguida é inicializado o laço de eventos e o processo fica espe-

rando um evento ser disparado, na sequência o node executa os manipuladores de evento e caso tiver um processo ocioso o mesmo é encerrado. No momento em que um evento entrar no *loop* de eventos o Node JS irá executar os *callbacks* que estão escutando aquele evento. Na figura 13 é possível entender um pouco melhor todo o fluxo de eventos dessa tecnologia.

Figura 13: Laços de evento do Node JS



Fonte: Wilson (2013)

2.3.5.3 Python

No ano de 1982, Guido Van Rossum que trabalhava no Centro de Matemática e Ciência da Computação do Centrum Wiskunde e Informatica, da Universidade de Middenmeeer, que está situado na cidade de Amsterdã, Holanda, criou o Python, uma linguagem de programação de alto nível interpretada, onde os programas escritos são apenas executados pelo sistema operacional, sem haver a necessidade da compilação de todo o código. Por se tratar de uma linguagem simples de escrita, ela aceita os paradigmas de programação orientada a objetos e também o paradigma de programação estruturada e funcional.

Por trabalhar com tipo de dados dinâmicos de dados, ou seja, não é necessário realizar a tipagem das variáveis, sendo considerada uma sintaxe que facilita na leitura do código e ainda é bastante concisa, ela é operada com um conjunto de códigos implícitos, o que significa que diferente das demais linguagens, não é necessário criar blocos de comandos como o BEGIN e END, citadas pelo autor

Manzano (2018)

2.3.6 Automatização de processos

Conforme De SORDI (2009), a automação de processos é um meio que torna possível eliminar o trabalho manual em tarefas repetitivas do dia a dia, diminuindo dessa forma erros comuns que podem vir a acontecer caso a tarefa seja realizada por pessoas, como o lançamento errado de um registro de uma tabela para outra tabela, ou a geração de um relatório gerencial de determinada data, visto que a máquina não sofre interferência de fatores externos como falta de atenção ou perda de foco, o que poderia acontecer com uma pessoa.

Atualmente com inúmeras possibilidades na área da tecnologia, onde diariamente novos *softwares* são criados, automatizar processos se tornou uma prática necessária para que as pessoas consigam focar em tarefas não repetitivas e que demandam de mais tempo e raciocínio lógico.

Diante disso, novos *softwares* e linguagens de programação são criados para este fim, de automatizar processos, muitas vezes buscando simular a ação de um usuário interagindo com o computador, tornando assim processos que eram totalmente manuais em processos ágeis.

2.3.6.1 Autolt

Dentre os *softwares* de automatização e linguagens existentes, um deles bastante conhecido é o Autolt, onde é possível desenvolver scripts por meio da escrita em Visual Basic e então executar ações que simulam um usuário. O mesmo foi criado em 1998 com uma simples funcionalidade de enviar comandos de teclas e foi evoluindo até chegar na versão 3 em 2014, além disso o seu fórum na internet é muito forte, e muitas dúvidas sobre a linguagem de programação e seu uso podem ser vistos na documentação também disponível no site.

O Autolt é um *software* voltado para o objetivo de automação de tarefas em ambientes Windows, sua implementação se dá por meio de uma linguagem interpretada. De acordo com Mar-elía (1998) o mesmo teve início por Jonathan Bennett, que até então utilizava uma antiga ferramenta para criação de scripts chamada de

ScriptIt e com isso sentiu uma grande necessidade de criar algo novo nesta área de automação, visto que a atual ferramenta não era suficiente para suprir todas as suas necessidades por não possuir muitas funcionalidades com o teclado e mouse, e opções de gerenciamento de janelas.

A base do *software* foi feita na linguagem C e posteriormente foi reescrito para C++ onde desde 1998 novas funcionalidades foram sendo adicionadas, porém passado alguns anos, em 2004, Bennett começou a perder o interesse pelo *software* devido a grande complexidade de se dar manutenção e desta forma, também sendo pressionado pelos usuários que utilizam a ferramenta, Bennett reescreveu o fonte transformando-o em uma linguagem de script, onde esta versão ficou conhecida como Autoit V3.

2.4 TRABALHOS RELACIONADOS

Nesta seção serão apresentados os trabalhos relacionados à pesquisa por meio da descrição de cada um visando o ramo ao qual se aplica e seus objetivos, para assim obter um comparativo em relação ao presente estudo.

2.4.1 CADUB GHF: um programa computacional para cálculo da quantidade de fertilizantes e corretivos da acidez do solo para culturas produtoras de grãos, hortaliças e forrageiras

De acordo com Gubiani and Silva et al. (2007), no Rio Grande do Sul e Santa Catarina, a recomendação de adubação e calagem é realizada de acordo com dados oficiais, onde é suportada por histórico de manejo, análises de solo e na própria experiência técnica. Com a constante modificação dos manuais de calagem e adubação, requer a criação de novos programas.

Esse artigo demonstra o CADUB, como uma ferramenta computacional, para realizar recomendações, de maneira digital, para os cultivos de grãos, forrageiras e hortaliças, onde retorna as quantidades de nitrogênio, fosforo e potássio para correção do solo para a adubação base, assim como também a quantidade necessária de calcário para cultivos como forrageiras, hortaliças e de grãos.

Com as informações obtidas pelos clientes e análise de solo, o CADUB

GHF, pois dessa maneira é possível criar um banco de dados para realizar comparações, análise da evolução da fertilidade do solo para um mesmo local e também para aumentar a performance de novas análises de solo para a mesma recomendação. Nesse *software* ainda, o usuário, pode ainda definir quais os fertilizantes irão ser cadastrados no *software*, por meio de inclusão e exclusão de fórmulas e fontes.

A exigência de calcário é definida por de acordo com um sistema de produção, de acordo com a área. O cálculo da necessidade de calcário tem como base o pH com a referência de 5,5; 6,0 e 6,5, esse é o valor a ser alcançado do pH da água contida no solo depois da calagem, que podem ser verificadas pelo índice de SMP das amostras e demais informações coletadas sobre o cultivo. Outra forma para realização do cálculo de calcário é com o uso do teor de alumínio trocável e a quantidade de matéria orgânica do solo, ou então, pelo processo de avaliação da saturação do CTC por bases.

O presente trabalho, busca solucionar o problema da criação de novos programas devido a alterações nos manuais de calagem, como o próprio autor Gubiani, Silva and Reinert (2015), comenta, devido a esse fato novos programas precisam ser desenvolvidos, mas com a utilização de Machine Learning, quando novos manuais serem lançados, basta realizar a atualização do modelo do Machine Learning.

2.4.2 *Large scale data processing in real world: From analytics to predictions*

A análise de processamento de dados em grande escala, está várias vezes encontrando muitos casos sob o tema de "*Bigdata*", onde o próprio Google, o IBM Watson e carros autônomos são algumas das histórias de sucesso. Quanto ao tipo de análise varia de estatísticas básicas, como a média, até modelos complicados de previsão baseados em *Machine Learning*.

Em Perera (2014) o uso do *Machine Learning* mostra a importância da tecnologia para resolver modelos complicados de previsão, assim como, a utilização de *Map Reduce* para o processamento de dados em lotes e também o uso de

Processamento de Eventos Complexos e Processamento de Fluxos para uso em tempo real.

Com base nisso, é correto destacar que a utilização do *Machine Learning* busca resolver problemas complexos, assim como o presente trabalho, também busca utilizar essa tecnologia para resolver o problema de recomendações de solo, pois dessa forma, talvez não seja mais necessário a utilização do manual de adubação e calagem pelo agrônomo, ficando por conta do aplicativo móvel.

2.4.3 *Heuristic prediction of rainfall using Machine Learning techniques*

Com a principal fonte da economia da Índia vinda da agricultura e vindo de encontro com a quantidade de chuva escassa, a necessidade de criação de uma forma de visualizar a previsão de chuva, se torna de vital importância para o crescimento da economia do país.

Para realizar previsão pode ser utilizada várias técnicas de inteligência artificial, como por exemplo, redes neurais, máquinas de vetor de suporte, análise de regressão linear e agrupamento.

A regressão linear é uma técnica para prever os vários valores de uma variável dependente a partir de uma variável independente quando a relação entre essas variáveis pode ser descrita por um modelo linear, sendo considerado uma forma de estimar o valor esperado condicional.

O autor Thirumalai and Harsha et al. (2017) buscou encontrar com o *Machine Learning* a previsão heurística de chuva, medindo através de dados dos anos anteriores de acordo com as estações de cultivo e prevê a precipitação futuras. Com a utilização de regressão linear, o aplicativo apresenta aos agricultores uma decisão mais assertiva para colher uma cultura específica.

Esse trabalho mostra o uso do *Machine Learning* para cálculo de previsão de chuva, onde o presente trabalho também fez o uso da tecnologia de *Machine Learning* para realizar a interpretação das análises de solo, deixando de realizar esse processo manualmente, assim como apresentar resultados bem próximos da correção ideal do solo para ter maior produtividade nas próximas plantações.

2.4.4 *Artificial calf weaning strategies and the role of machine learning: A review*

Nessa pesquisa, o autor Katamreddy, Riordan and Doody (2017) relata um pouco sobre o aumento drástico da utilização do *Machine Learning*, porém a utilização dessa tecnologia ainda não é muito usada na área da agricultura e na criação de animais, comparada com as demais áreas. Essa pesquisa teve como enfoque a utilização de *Machine Learning* para revisar estratégias passadas e atuais de desmame de bezerros, com intuito de buscar a integração de sensores mais relevantes, para melhorar a precisão da previsão.

Esta revisão é limitada à estratégias de desmame de bezerros que poderiam beneficiar a produção de carne (carne bovina) e leite, adicionando técnicas de *Machine Learning*, com o propósito de aumentar a quantidade de aplicações de *Machine Learning* no setor agrícola.

Para que o objetivo dessa revisão fosse alcançado, foi criada uma forma de detecção da sucção em bezerros pré-desmamados durante o processo de desmame, onde com a observação de padrões de movimento da cabeça e pescoço dos bezerros é possível alcançar esse objetivo, porém o autor buscou a utilização de câmeras para coleta de informações.

Então esses dados são analisados em comparação com observações físicas feitas e depois são extraídos e rotulados, para que dessa forma possam ser pré-processados e alimentados para classificadores de *Machine Learning*. Com esse trabalho é relevante destacar a escassez de trabalhos voltado para a agricultura e criação de animais, onde a utilização de tecnologias como o *Machine Learning* para solucionar problemas dessa área.

2.4.5 *Machine Learning Approaches for Soil Classification in a Multi-agent Deficit Irrigation Control System*

A Austrália vem sofrendo com graves faltas de água nos últimos anos, onde o principal agente responsável pelo volume total é a agricultura, chegando a 65% do consumo de água. Esse alto volume de água foi o fato que necessitou-se encontrar uma forma mais eficiente do uso da água para irrigação.

Essa prática ainda assim conta com um custo reduzido de sensores de umidade do solo, eles são implantados para monitorar o status de água no solo, onde as estratégias de controle acionados por liminares são usados para manter os níveis de umidade do solo.

Esse trabalho de Smith and Peng (2009) buscou através do uso de *Machine Learning*, o nível de retenção de água do solo para medição da textura do solo, dessa forma, pode-se alcançar um nível de acuracidade de 85,11 % em comparação com o banco de dados do UNSODA.

Com a utilização de seis características do solo, o mesmo é comparado com o presente trabalho pela utilização do *Machine Learning*, onde relacionando entre si, consegue atingir uma alta acuracidade.

2.4.6 *Crop Selection Method to maximize crop yield rate using machine learning technique*

Conseguir alcançar taxas máximas de rendimento da cultura usando espaços de terras limitadas, ou seja, buscando realizar o planejamento agrícola de um país. Neste são encontrados dois fatores que influenciam diretamente, o primeiro a taxa de rendimento da cultura: primeiro é a qualidade das sementes que podem ser melhorados pelo desenvolvimento, já o segundo é a seleção de culturas com base em condições favoráveis de ambos os lados.

Em Kumar and Singh et al. (2015), demonstra que vários estudos são feitos nas mais diversas áreas da agricultura, como, previsão do tempo, classificação do solo, classificação das culturas e previsão da taxa de produtividade da cultura. Porém, este buscou com a utilização do *Machine Learning* de melhorar a lucratividade líquida da cultura, onde, com isso resolve o problema de seleção de culturas.

Com o estudo do presente trabalho é possível visualizar que diversas áreas são estudadas no ramo da agricultura, porém a área de interpretação da análise de solo, ainda é uma parte da agricultura pouco explorada para automatização e melhoria no processo de interpretação que atualmente é realizada manualmente.

2.4.7 *Intelligent system for measuring soil compaction on croplands*

De acordo com Furriel and Calixto et al. (2015), a agricultura de precisão possui um grande impasse que é a análise de dados, o custo dos instrumentos e qualificação dos colaboradores, dificultando o uso dessa técnica de correção de solo. O desenvolvimento de um processo inteligente através da tecnologia, pode facilitar a captura de informações essenciais para a agricultura de precisão, como a captura da compactação do solo, utilizando o Índice de Cone, substituindo a forma tradicional e obtendo dados com uma precisão muito maior.

A forma de coleta dos dados também é inédita, onde era utilizado células de carga, agora é utilizado cargas elétricas para definir a compactação do solo. O sistema foi criado com uma inteligência, onde o mesmo especifica os locais que devem ser a realizada as coletas das terras e compila essas informações em imagens, melhorando a visualização e interpretação dos dados coletados.

A proposta de aprimoramento de aprimoramento na análise de dados na agricultura de precisão vem de encontro direto com o presente trabalho, no qual as recomendações de análise de solo, buscam também auxiliar no processo de correção do solo, podendo assim ser uma ferramenta de apoio à agricultura de precisão.

2.4.8 *Low-cost multispectral imaging system for crop monitoring*

Este trabalho de Oca and Arreola et al. (2018) tem como intuito a solução de problemas voltado para tarefas de agricultura de precisão, através de um sistema de multiespectral. Com a utilização de um drone, que possui duas câmeras acopladas é o responsável por realizar a captura de imagens, onde uma delas foi modificada com o objetivo para ser sensível a radiação infravermelha próxima refletida pela vegetação, enquanto a segunda câmera é acoplada como uma câmera comum em formato de RGB.

Com as informações coletas foi desenvolvido um *software* que realiza alguns tratamentos na imagem aérea obtida, como a remoção de distorções e inserção de metadados para geração do ortomosaico que servirá para calcular o Índice de Vegetação por Diferença Normalizada (NDVI) que avalia a saúde da cultura

analisada.

O trabalho descrito é comparado com o presente projeto desenvolvido com o enfoque para solucionar tarefas da agricultura de precisão, onde o direcionador é a recomendação da análise do solo.

2.4.9 *Interconnection of a crop growth model with remote sensing data to estimate the total available water capacity of soils*

Segundo Todoroff, De Robillard and Laurent (2010), para auxiliar no gerenciamento da produção dos mais diversos cultivos, várias ferramentas são usadas, onde em sua maioria é baseada no *Geographic Information System* (GIS), que visa a heterogeneidade geográfica para as condições de crescimento da cultura. Outros aspectos também são levados em conta, como a capacidade total de água disponível do solo (TAWC) que é um fator que também tem grande impacto no desenvolvimento da cultura. Com a utilização de tecnologias como o NDVI e a captura de imagens feitas pelo satélite SPOT, através dessas tecnologias é possível verificar a condição da cultura.

Os resultados foram validados de acordo com uma plantação de cana-de-açúcar, onde conseguiram atingir valores bem próximos do ideal de acordo com a bibliografia, podendo assim ser uma técnica robusta e barata. Também para a validação da medição do TAWC foi usado uma área onde durante o período de 2 anos fosse relativamente seco, onde a cultura dependeu relativamente da capacidade de solo para retenção de água.

O trabalho dos autores acima, podem ser comparados ao presente projeto, pois buscaram novas tecnologias para aprimorar o desenvolvimento do dia a dia na agricultura, pois com a utilização de imagens é possível realizar a avaliação da saúde da cultura, assim como o uso do *Machine Learning* reduz o trabalho manual para interpretação de análises de solo.

2.4.10 *Design and development of M3SS: A Soil Sensor Node for precision agriculture*

De acordo com Kumar and Reddy (2016), a crescente demanda pela população por alimentos os agricultores necessitam que cada vez mais suas produções tenham um maior rendimento, utilizando produtos químicos, maquinários com altas tecnologias, assim como muitas técnicas especiais sem entender muitas vezes a real necessidade do solo.

Portanto, para isso é necessário o monitoramento do solo em tempo real, então para isso o sensor Multi-modo, Multi-parâmetro, Multi-Aplicação *Soil Sensing* (M3SS) foi desenvolvido, oferecendo alertas e conselhos sobre o solo para o agricultor obter o maior rendimento da sua cultura. Para desenvolvimento do M3SS foi criado vários algoritmos, como o algoritmo baseado em eventos, algoritmos de compreensão de dados e ainda algoritmos para o controle da eficiência energética.

Pode-se observar que o trabalho acima criou o M3SS para auxiliar o agricultor no monitoramento em tempo real do solo e posterior ajustes necessários para correção. O presente trabalho também busca auxiliar o agricultor na correção do solo, tomando como base informações extraídas da análise de solo, sendo possível realizar a correção do solo.

2.4.11 *Machine Learning for the Management of Agricultural Soil Data*

De acordo com Sirsat and Fernández-Delgado (2016), a agricultura na Índia possui terras de baixa produtividade e devido a esse fato é ainda mais dependente de nutrientes contidos no solo, dessa forma a qualidade do solo é de vital importância, onde que através da avaliação de propriedades físicas, químicas e biológicas é possível criar planos de produção e prever sua produtividade.

Em função disso, a tecnologia de *Machine Learning* supre a necessidade e facilitar a classificação de vários parâmetros presentes no solo, entre eles, o índice de fertilidade do solo, recomendações para correção de N, P, K, PH do solo e tipo de cultura, todos usando como base dados oferecidas pela análise de solo.

Para criação do modelo utilizado pelos autores, foi desenvolvido um total de

5016 experimentos, subdivido em 66 conjuntos de dados contendo cada um com 76 item, para que seja possível processar essa quantidade de dados, foi utilizado um *cluster* CITIUS, com processador Intel Xeon E5-2650L, com 64 GB de memória RAM.

Importante ressaltar aqui que alguns conjuntos de dados não foram processados devido a erros ocorridos durante o processamento, como por exemplo, a enorme utilização de memória (mais de 128 GB) ou tempo de computação (um limite superior de 150 horas ou 6,25 dias foi definido). O conjunto de dados que não foi processado foi de 46 experimentos do total de 5016, representando 0,92% do total.

O trabalho desenvolvido por Sirsat and Fernández-Delgado (2016), teve como objetivo atender uma necessidade do governo indiano, onde foi criado um sistema desktop para utilização. E o trabalho aqui proposto busca criar um aplicativo móvel que abrange outros elementos contidos na análise de solo. Importante destacar ainda que este trabalho se diferencia por ser voltado para a Região Sul do Brasil, pois será baseado indiretamente no manual de calagem e adubação oferecido para essa região.

2.4.12 *Developing fertilizer recommendations for agriculture*

Em Dinkins and Jones (2007), especifica que normalmente uma análise de solo vem acompanhado de recomendações de análise de solo, baseado em safras anteriores, que tipo de cultura irá se cultivar e as metas de rendimentos que devem ser alcançadas. Porém, as recomendações podem também ser adequadas de acordo com as propriedades do solo contidas no ponto de coleta, assim como também, podem ser ajustadas com base no clima.

Devido a esse fato, para os produtores, ter consultores de agrícolas que conhecem a região para recomendações tendem a ter um maior êxito na recomendação. O trabalho desenvolvido, tem como intuito deixar documentado como realizar a recomendações de forma simples, onde o produtor mesmo consegue realizar uma breve recomendação.

O autor Dinkins and Jones (2007), criou uma forma manual de realizar a

recomendação, se tornando o oposto do presente trabalho, que busca justamente automatizar a recomendações de análise de solo.

2.4.13 *Using Machine Learning for land suitability classification*

A agricultura é fundamental como fonte de alimento e renda, a adequação da terra é feita para estimar a adequação da terra para uso específico, como por exemplo, a agricultura arável ou agricultura irrigada.

A utilização de sistemas únicos de classificação nem sempre são a melhor opção quando a alta precisão precisa ser levada em conta. Já sistemas de classificação múltipla ou métodos ensemble, por exemplo, estão crescendo rapidamente por serem dois métodos que tem se mostrado mais precisos e robustos do que métodos de classificação única.

Os métodos de classificação múltipla, normalmente trabalham de maneira que primeiro aplicam um conjunto de classificadores de base aplicando algum algoritmo de *Machine Learning* para diferentes conjuntos de treinamento alternativos e, em seguida, as saídas de cada membro do conjunto são combinadas de uma forma adequada para criar a previsão do conjunto.

No trabalho desenvolvido por Mokarram and Hamzeh et al. (2015), comenta que a inteligência artificial, mais especificamente o *Machine Learning*, podem ser usados para a classificação do solo, onde é realizado utilizando uma técnica de classificação de conjuntos, denominada de RotBoost, que através do método de FAO, pode se comprovar a utilização desse algoritmo para resolver o problema de classificação de solo. O presente trabalho procura solucionar o mesmo problema que é de classificação da terra.

2.4.14 *Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction*

Data Mining é um domínio de pesquisa crucial em um mundo de pesquisa recente, a utilização de estatísticas padrão, são forma demoradas e caras, técnicas mais eficientes podem ser desenvolvidas para resolver conjuntos complexo de dados do solo, melhorando a eficácia e precisão na classificação de grandes

conjuntos de dados.

Em Gholap and Ingole et al. (2012), com a crescente demanda de pesquisas na área agrícola, o uso de processos automatizados, como a mineração de dados, que atualmente é utilizada em várias áreas diferente, assim como produtos voltados para a agricultura que já trazem a mineração de dados.

O objetivo dessa pesquisa é realizar a análise de dados de solo, utilizando da técnica de mineração de dados, onde também em alguns casos ele utiliza métodos de *Machine Learning*, como por exemplo, o método de regressão linear pra realizar previsões.

O autor Gholap and Ingole et al. (2012), resolve o mesmo problema que o presente trabalho, com a utilização de algoritmos de mineração de dados, mas utilizou técnicas de regressão linear, para resolver problemas de premeditação de análise de dados.

2.4.15 Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks

Os nutrientes contidos no solo são propriedades crucial que contribui para a fertilidade do solo e outro fatores naturais. Portanto, uma metodologia eficaz é necessário para realizar a avaliação dos nutrientes do solo por uma questão de gestão científica e utilização racional dos nutrientes do solo por uma questão.

De acordo com Li and Leng et al. (2014), um aspecto muito importante que impacta diretamente em efeitos ambientais e na fertilidade do solo são os nutrientes e as formas atuais existentes são difíceis de manusear, então nesse trabalho foram utilizadas diversas áreas da inteligência artificial para solucionar o problema, como algoritmos de regressão linear, máquinas de vetores de suporte e redes neurais, onde através do primeiro algoritmo é possível alcançar um 92,86% de acuracidade, assim como 77,87% e 88% respectivamente para os demais algoritmos.

O trabalho dos autores Li and Leng et al. (2014), demonstra que é possível realizar a avaliação dos nutrientes contidos no solo somente com a utilização da inteligência artificial, o presente trabalho também busca resolver o problema de recomendação com a utilização de uma área da inteligência artificial.

2.4.16 Table Recognition in Heterogeneous Documents using Machine Learning

De acordo com Rashid and Akmal et al. (2017) a detecção de tabelas em documentos ainda é um desafio na área da tecnologia, devido a grande variedade de layouts existentes atualmente. Diante disso, este trabalho tem por objetivo reconhecer tabelas em documentos heterogêneos, por meio da utilização de uma rede neural pré treinada. Durante os testes da solução foi atingida uma precisão de aproximadamente 95% de tabelas detectadas.

Os pesquisadores fizeram o uso de um repositório com aproximadamente 2889 páginas que continham estruturas de tabelas de uma variedade de fontes, como revistas, notícias, documentos e carta de negócios para treinar o algoritmo. O treinamento é realizado por meio de um perceptron multicamada sintonizável (AutoMLP) usado para evitar a otimização da rede neural e problemas durante o treinamento.

Para validação da solução foram realizados testes com 225 imagens de documentos onde como primeiro resultado se obteve uma precisão de 64%, porém ao aplicar uma etapa de pré-processamento a essas imagens, a precisão melhorou para 95%.

O trabalho em questão demonstra que atualmente já existem muitas formas de se realizar detecção de tabelas em documentos, porém nenhuma delas ainda pode ser utilizada de forma genérica, porém com a utilização de Machine Learning esse cenário começa a mudar, tendo suporte a reconhecimento de layout mais variados. Além disso, foi possível identificar a importância da etapa de pré-processamento das imagens antes de se submeter ao processamento, onde houve um aumento de 31% na acurácia total atingida.

2.4.17 *DeCNT: Deep Deformable CNN for Table Detection*

Este trabalho apresenta uma nova abordagem para a detecção de tabelas visto que faz a utilização de redes neurais profundas. O uso da tecnologia de *Deep Learning* para resolução do problemas de detecção de tabelas proposto no trabalho se defende visto que a abordagem comum utilizada para detecção de

tabelas como detecção de linhas, métodos de busca de contorno são propensos a erros e geralmente são aplicados somente em uma situação ou estrutura de tabela específica.

Conforme Siddiqui and Malik et al. (2018), a abordagem proposta no trabalho é baseada em uma nova combinação de CNN deformável visto que a CNN convencional possui um campo receptivo fixo e limitado, não sendo viável para a detecção de tabelas, já que tabelas podem estar presentes em escalas arbitrárias, e sofrerem inclinações e mudanças de orientação e neste caso a CNN convencional não é suficiente.

Para realização dos experimentos foram utilizados 4 datasets, são eles ICDAR-13, ICDAR-17 POD, MORMOT, UNLV, onde foi realizado testes com ambos os datasets como forma de definir a acurácia de cada um separadamente. Destes testes quem se saiu melhor foi o dataset ICDAR-13, com uma precisão de 99,4 % na identificação das tabelas.

Com o estudo deste trabalho é possível perceber que já existem alguns datasets disponíveis para auxiliar na detecção de tabelas em imagens, porém sua implementação ainda depende de um conhecimento maior na área de Deep Learning, visto que trata-se de uma área bastante avançada do conhecimento técnico, e seu tempo de implementação pode ser maior.

2.5 TABELA DE COMPARAÇÃO DE TRABALHOS RELACIONADOS

Para concluir o estudo realizado sobre os trabalhos relacionados, o quadro 1 demonstra os principais assuntos apresentados, onde por meio desse é possível visualizar em aspecto mais amplo a contribuição de cada trabalho.

Quadro 1: Trabalhos relacionados.

Título do trabalho	Agricultura	Recomendações	<i>Machine Learning</i>	Processamento de imagens
CADUB GHF: um programa computacional para cálculo da quantidade de fertilizantes e corretivos da acidez do solo para culturas produtoras de grãos, hortaliças e forrageiras	X	X		
<i>Large scale data processing in real world: From analytics to predictions</i>			X	
<i>Heuristic prediction of rainfall using machine learning techniques</i>	X		X	
<i>Artificial calf weaning strategies and the role of machine learning: A review</i>			X	
<i>Machine Learning Approaches for Soil Classification in a Multi-agent Deficit Irrigation Control System</i>	X		X	

Título do trabalho	Agricultura	Recomendações	<i>Machine Learning</i>	Processamento de imagens
<i>Crop Selection Method to maximize crop yield rate using machine learning technique</i>	X		X	
<i>Intelligent system for measuring soil compaction on croplands</i>	X			
<i>Low-cost multispectral imaging system for crop monitoring</i>	X			
<i>Interconnection of a crop growth model with remote sensing data to estimate the total available water capacity of soils</i>	X			
<i>Design and development of M3SS: A Soil Sensor Node for precision agriculture</i>	X			
<i>Machine Learning for the Management of Agricultural Soil Data</i>	X	X	X	
<i>Developing fertilizer recommendations for agriculture</i>	X	X		

Título do trabalho	Agricultura	Recomendações	<i>Machine Learning</i>	Processamento de imagens
<i>Using machine learning for land suitability classification</i>	X		X	
<i>Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction</i>	X		X	
<i>Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks</i>	X		X	
<i>Table Recognition in Heterogeneous Documents using Machine Learning</i>			X	X
<i>DeCNT: Deep Deformable CNN for Table Detection</i>			X	X

CAPÍTULO 3 RESULTADOS OBTIDOS

O presente capítulo tem como objetivo apresentar um estudo da viabilidade de utilização de *Machine Learning* para geração de recomendação de análises de solo.

Dentre as etapas abordadas estarão as de tratamento de dados para possível criação de um modelo de treino com o auxílio do *software* CADUB, o estudo do uso de inteligência artificial para recomendação, a criação de um algoritmo capaz de reconhecer as informações presentes em uma análise de solo por meio de testes com as ferramentas Tesseract OCR e Google Vision e a implementação de um aplicativo mobile que irá realizar a integração com o algoritmo de reconhecimento desenvolvido.

3.1 LABORATÓRIO DE ANÁLISE DE SOLOS

Para dar início ao desenvolvimento do projeto tornou-se preciso encontrar uma base de dados com uma grande quantidade de análises de solo cadastradas para, então, a partir destas, realizar a interpretação dos resultados por meio do *software* CADUB e posteriormente estudar a possibilidade de criar um modelo de treino com os dados obtidos.

Diante disso, se obteve auxílio dos profissionais do Laboratório de Solos da SETREM, por meio de conversas direcionadas especialmente ao professor responsável Paulo André Klarmann teve-se acesso a base de dados do laboratório juntamente com a ajuda do, até atual, coordenador de TI da SETREM, Edinei Stefens.

O Laboratório de Solos da SETREM recebe por semana aproximadamente 250 amostras de solo para serem analisadas, sendo que a procura pelos serviços

parte de toda a região noroeste do estado do Rio Grande do Sul. De acordo com, Paulo André Klarmann, o laboratório possui um alto controle de qualidade seguindo os padrões da Rede Oficial de Laboratórios de Análise de Solo e de Tecido Vegetal dos Estados do Rio Grande do Sul e de Santa Catarina (ROLAS).

O mesmo está situado no Campus da SETREM em Três de Maio e atualmente conta com cinco profissionais que são responsáveis por realizar o processo de análise dos componentes químicos encontrados nas amostras de solo recebidas, sendo eles, uma laborista formada em química industrial, um engenheiro agrônomo e três auxiliares de laboratório.

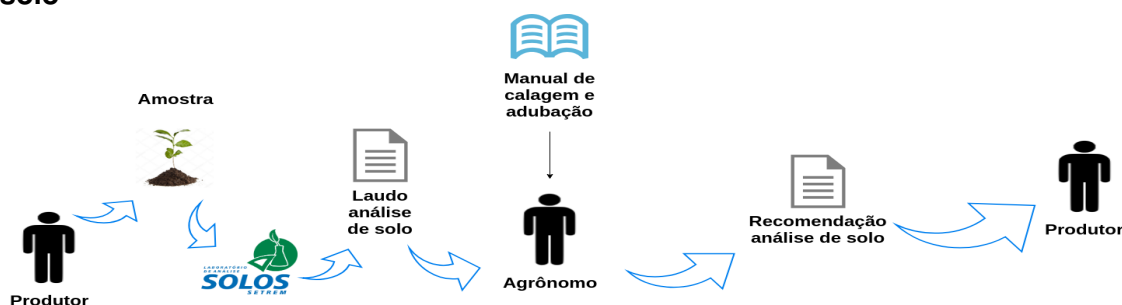
3.2 CENÁRIO ATUAL

O maior interesse em desenvolver esse tipo de projeto se deu após ter conhecimento da forma de trabalho extremamente mecânica que o produtor rural ou agrônomo ainda exerce para obter as informações de uma análise de solo e então sugerir uma recomendação de adubação para determinada gleba de terra.

Por meio do estudo realizado foram identificados processos muito manuais que acabam demandando um maior esforço e tempo, principalmente por parte do engenheiro agrônomo que é responsável por consultar as tabelas de adubação e calagem e, então estruturar uma possível recomendação de aplicação de adubos e fertilizantes.

Para compreender o cenário atual foi desenvolvido um fluxo, representado pela figura 14, que demonstra as etapas necessárias para se chegar a uma recomendação de análise de solo, desde a submissão da amostra de solo ao laboratório até o processo de recomendação de adubação pelo engenheiro agrônomo.

Figura 14: Fluxo atual de recomendação de adubação por meio de uma análise de solo



Inicialmente o produtor rural realiza a coleta de uma amostra de solo que servirá de referência para a quantidade de hectares que o mesmo deseja realizar a plantação futuramente, seja ela de milho, trigo, soja ou outro tipo de cultura. Com a amostra coletada a mesma é encaminhada para o Laboratório de Solos da SETREM.

Chegando no laboratório a amostra de solo é processada, secada e tratada para que se iniciem os experimentos químicos, essa tarefa ainda depende de pessoas, ou seja, não existe nenhum processo que automatize essas atividades. Após passado o período de 7 a 10 dias, a análise é disponibilizada ao solicitante. Vale lembrar que os laboratórios de análise de solo não podem recomendar nenhum tipo de adubação ou calagem com base nos resultados obtidos na análise de solo, estes devem ser feitos posteriormente por um engenheiro agrônomo formado.

Após a finalização do processo de geração da análise de solos pelo laboratório, o produtor rural ou solicitante recebe uma folha em papel A4 conforme a da figura 15 contendo as informações referentes aos nutrientes químicos encontrados na amostra de solo anteriormente submetida. Com esse papel em mãos é preciso que o produtor rural procure algum engenheiro agrônomo para que o mesmo realize a interpretação dos resultados e indique qual a melhor adubação ou correção a se aplicar na lavoura caso seja necessário.

Em casos onde a análise de solo é utilizada para comprovante de custeio agrícola, quem concretiza este tipo de solicitação é o banco, para dessa forma, identificar o quão fértil aquele solo é e dependendo dos resultados, disponibilizar ou não algum benefício monetário, como um financiamento ao produtor rural. No momento em que o engenheiro agrônomo recebe a análise de solo pelo produtor rural ou pelo banco, cabe ao mesmo analisar a quantidade de nutrientes e componentes químicos encontrados naquela amostra de solo e com base nos manuais de calagem e adubação do estado do Rio Grande do Sul e Santa Catarina fazer a devida recomendação.

Este processo exige bastante atenção, visto que, conforme já explanado no capítulo 2, o manual de calagem e adubação é composto de várias tabelas de referências, onde um elemento químico influencia em outro, abrindo uma ramificação

Figura 15: Exemplo de análise de solo gerada pelo laboratório de solos da Setrem

LAUDO DE ANÁLISE DE SOLO

Nome: _____ Data Recebimento: 29/03/2019
 Localidade: _____ Data de Expedição: 05/04/2019
 Município: _____

Amostra Nº	Área (ha)	Argila %	Classe Textural	pH H ₂ O	Índice SMP	M.O. %	P	K	Ca	Mg	Al	H+Al
				l:l			mg L ⁻¹				Cmol _c L ⁻¹	
42424	21,00	81	1	5,7	5,5	3,7	22,5	94	5,8	2,0	0,0	7,7

Amostra Nº	CTC (Cmol _c L ⁻¹)	Saturação (%)	S	B	Cu	Zn	Mn	Mo	Fe	Na		
	Efetiva	pH _e	Bases	Al	mg L ⁻¹							
42424	8,0	15,7	50,9	0,0	11,7	0,4	7,0	4,3	33,6	N.D	N.D	N.D

Amostra: 42424 - Mat 12334

Identificação

Observações

Responsável Técnico

Fonte: Laboratório de solos SETREM, 2018

de condições a serem analisadas para só então ser possível recomendar alguma fórmula de adubo básica no formato NPK (Nitrogênio, Fósforo e Potássio), bem como a quantidade a ser aplicada por hectare, geralmente medida em toneladas.

3.3 SOLUÇÃO PROPOSTA

Diante do cenário identificado durante o estudo de todas as etapas pertinentes ao processo de recomendação de análises de solo, que vai desde a coleta da amostra pelo produtor rural até a recomendação gerada pelo engenheiro agrônomo, se identificou alguns pontos de melhorias que podem ser aplicados ao processo.

No processo existente hoje pode-se identificar alguns problemas de lentidão e dificuldade para se obter resultados quanto a informações sobre o solo. Dessa forma, o produtor rural precisa submeter a amostra de solo com uma certa antecedência e ainda ter algum engenheiro agrônomo disponível para lhe auxiliar na interpretação da análise de solo para uma possível recomendação de adubação. Outro ponto identificado é que o tempo que esse processo dura varia muito,

principalmente devido ao fato de depender de fatores humanos, como a consulta ao engenheiro agrônomo, e de fatores manuais, como a consulta ao manual de calagem e adubação.

A disponibilização do resultado pelo Laboratório de Solos da SETREM também é um fator que se deve levar em consideração, visto que hoje o principal meio de demonstrar a resposta da análise é através de uma folha de papel, fazendo com que o produtor rural novamente compareça até o laboratório para retirar determinado documento, sendo que com as tecnologias existentes hoje isto se torna totalmente desnecessário, podendo ser disponibilizado alguma consulta *online* dos resultados.

Seguindo para o aspecto da recomendação de adubação, quando o produtor rural já tem a análise em mãos, é necessário um profissional formado para interpretar os resultados. Esse profissional precisa se basear em um manual de adubação e calagem, que conforme já explicado, trata-se de um livro físico e não digital que o engenheiro agrônomo precisa adquirir de tempos em tempos para se manter atualizado sobre as referências indicadas de nutrientes encontrados no solo e com isso conseguir interpretar e disponibilizar uma recomendação ao produtor rural ou banco quando solicitado para custeio.

Outro ponto que deve ser levado em consideração é uma maior possibilidade de gerar uma recomendação de solo incorreta, visto que por se tratar de algo manual e necessitar de apoio humano, na consulta do manual de calagem, podem acontecer situações em que o profissional cometa algum descuido e lance uma quantidade diferente de adubação para determinado componente químico, alterando assim a fórmula NPK no final, além da possibilidade de prejudicar toda uma safra devido a aplicação incorreta do adubo pelo produtor rural. Esse tipo de erro pode acontecer principalmente pela grande quantidade de tabelas que o engenheiro agrônomo precisa consultar e calcular.

Diante do cenário apresentado, foi proposto um novo fluxo, conforme figura 16 para ser utilizado na etapa de análise de solos e recomendação de adubação, já levando em consideração o objetivo do trabalho, de criar um aplicativo mobile com a utilização de *Machine Learning*.

Figura 16: Sugestão de fluxo para recomendação de análise de solo



Conforme é possível ver na figura 16, em relação a figura 14, a diferença está na forma com que a análise de solo será interpretada, mudando o que antes era manual para um *software*.

Analisando a situação atual do fluxo de recomendação de análises de solo, o principal ponto de lentidão acontece devido a necessidade de se consultar um manual que possui parâmetros responsáveis por definir a quantidade de adubação necessária. Dessa forma, a proposta do presente trabalho é automatizar esse processo com a utilização de um aplicativo mobile capaz de interpretar a análise de solo que o agricultor possui. Em seguida, com base nas análises de solos já existentes no laboratório da SETREM, buscar entender como os parâmetros desses manuais são gerados, tentando, assim, aprendê-los e simulá-los para novas análises de solo por meio do uso de *Machine Learning*.

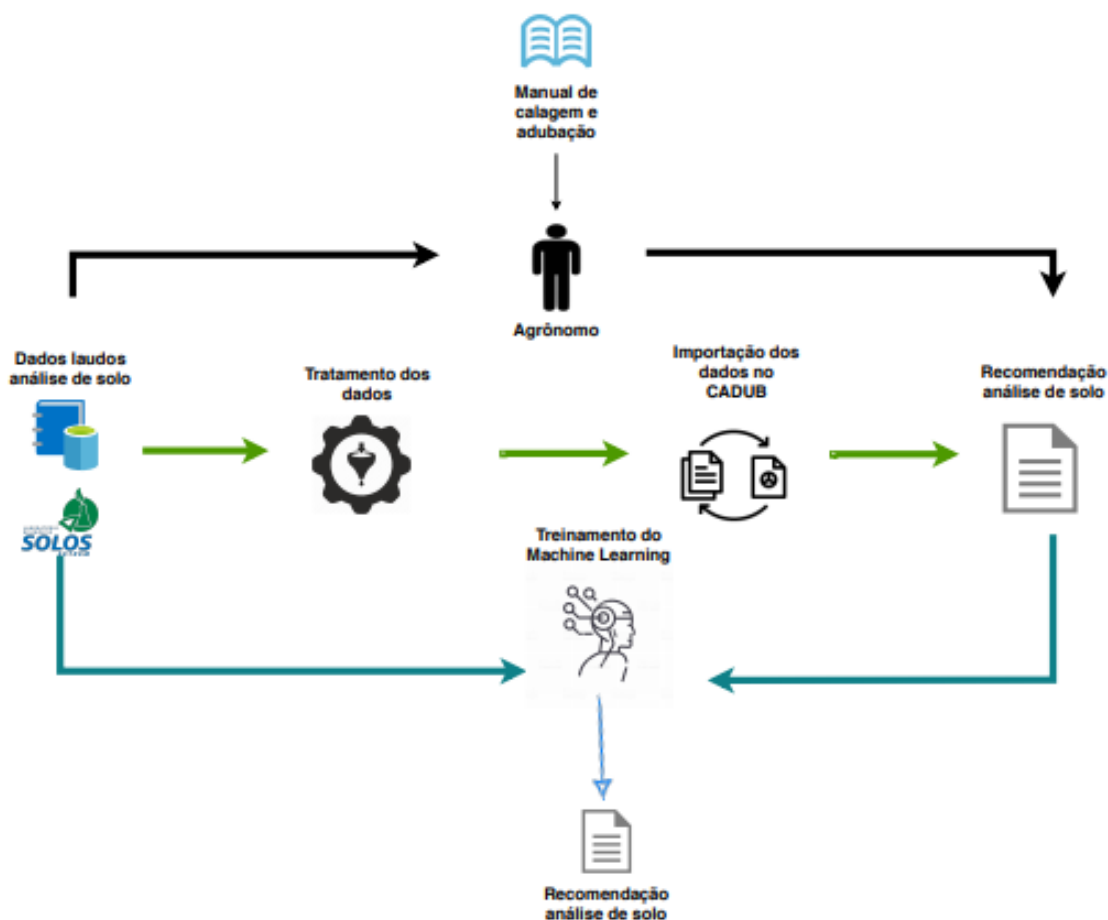
3.4 ESTUDO E ANÁLISE DO USO DA INTELIGÊNCIA ARTIFICIAL PARA RECOMENDAÇÃO

Para que as recomendações de análise de solo fossem geradas, foram estudadas duas formas de obtenção desses dados. A primeira seria por meio de um agrônomo gerar as recomendações. No entanto, o agrônomo precisaria gerar trinta mil recomendações, provenientes dos dados fornecidos pelo laboratório de análise de solos da SETREM, para que dessa forma o *Machine Learning* fosse alimentado. Por exemplo, se um agrônomo gerar dez recomendações em um dia, seria em torno de dez meses para geração dessas recomendações, assim como, a chance de equívocos acontecerem durante esse processo.

Com o intuito de agilizar esse processo, a automatização desta etapa foi a solução mais viável encontrada. Por meio do tratamento dos dados e utilização do *software* CADUB, foi possível reduzir esse tempo para menos de uma semana, realizando essa mesma etapa do processo para ter as recomendações de todas as análises geradas.

Na figura 17 está ilustrado passo a passo as etapas que precisam ser realizadas para se atingir o objetivo desejado, que vão desde a obtenção das análises de solo no Laboratório de Solos da SETREM até o treinamento desses dados com o algoritmo de *Machine Learning*.

Figura 17: Passo a passo para criação do processo de recomendação de análise de solo



Como forma de demonstrar o passo a passo de cada etapa a figura 17 mostra de forma mais clara as etapas que foram seguidas para alcançar os dados que são necessários para treinamento do *Machine Learning*. A seguir será descrito detalhadamente cada etapa seguida nesse processo.

3.4.1 Extração e tratamento de dados

Com os dados disponibilizados pelo laboratório de solos foi possível obter informações referentes a amostra de solo, a cidade de origem, o tipo de solo, o nome da gleba, profundidade da coleta e o mais importante, os componentes químicos encontrados no solo e seus respectivos valores.

Todas as informações foram disponibilizadas através de um arquivo CSV que precisou ser tratado para se realizar a importação em uma base de dados relacional (MySQL). Dessa forma, foi possível iniciar o tratamento dos dados, por meio do agrupamento de informações e retirada de colunas que não serão utilizadas por questões de privacidade, como por exemplo os dados pessoais da pessoa responsável pela solicitação de análise de solo.

Após realizar a importação do arquivo CSV no banco de dados relacional a tabela ficou com a estrutura de campos conforme o quadro 2, onde é possível ver quais campos foram ignorados e o que cada coluna armazena.

Quadro 2: Nome das colunas x Significados da base do laboratório de solos SETREM.

Coluna	Descrição
PRT_CODIGO	Código do protocolo
PRT_PRODUTO	Código do produto
PRT_SOLICITANTE	Nome completo da pessoa que solicitou a análise de solo
PRT_DATA_PEDIDO	Data em que a análise foi solicitada
PRT_PAGO	Informa se a análise já foi paga, Sim ou Não
PRT_VALOR	Informa qual o valor que o solicitante pagou pela análise
PRT_STATUS	Informa se a análise já está concluída ou está em andamento
PRT_NOMELAUDO	Nome descritivo da análise
PRT_OBSERVAÇÃO	Nenhuma informação na coluna
PRT_LOCALIDADE	Informa a localidade dentro do município, por exemplo, Linha Alma
PRT_MUNICIPIO	Informa o município de onde foi realizada a coleta da amostra de solo
SOL_CODIGO	Informa o tipo de solo
AMO_CODIGO	Código da amostra de solo
ANA_CODIGO	Código da análise de solo
AMO_NOME_GLEBA	Informa o nome da gleba/amostra
AMO_AREA	Informa a área em hectares em que a amostra foi colhida
AMO_PROFUNDIDADE	Informa a profundidade ao qual a amostra foi coletada
AMO_DATA	Informa a data em que a amostra foi coletada
DAC_CODIGO	Código referente as amostras da coleta
ELE_COD_ELEMENTO	Código numérico do elemento químico associado a análise
DAC_VALOR	Valor referente a amostras de coleta
DAL_VALOR	Valor referente a amostras da análise final

Durante a análise da base de dados do laboratório de solos, percebe-se que para uma única análise haviam aproximadamente sete registros na tabela, sendo um registro para cada componente químico, este, correspondente a coluna ELE_COD_ELEMENTO. Esta coluna apresentava em números sequenciais o elemento correspondente, dessa forma foi preciso interpretar o que cada número significava entre os elementos químicos, tendo como resultado a combinação “DE-PARA” descrita no quadro 3.

Quadro 3: Códigos do elemento x Elemento químico associado

Código	Descrição
1	P
2	M. O.
3	pH H ₂ O
4	Índice SMP
5	Argila
6	K
7	Ca
8	Mg
9	Al
10	H+Al

Para que fosse possível utilizar estes dados, foi necessário importa-los para um banco de dados relacional MySQL, possibilitando dessa forma o manuseio dessas informações de forma mais rápida, onde se necessitava realizar o agrupamento das informações e tratar campos vazios ou não padronizados, pois esses causavam erros não tratados no *software* CADUB.

3.4.2 Utilização do *software* CADUB

Como forma de obter os resultados e recomendações de adubação com base nos dados obtidos da base de dados do Laboratório de Solos da SETREM foi utilizado o *software* CADUB, visto que o mesmo permite a submissão de mais

de uma análise de solo por meio da importação de um arquivo CSV com os dados de cada análise, facilitando assim a integração e obtenção dos resultados de adubação desejados.

3.4.2.1 Importação de análises de solo

Após a conclusão do processo de tratamento de dados foi realizado um estudo da ferramenta CADUB para identificar qual a estrutura do arquivo XLSX do Excel de importação esperada, para então importar os dados das análises de solo. Ao acessar a opção de importação de uma planilha do Excel o próprio *software* gerou um modelo de como deve ser a estrutura das colunas, conforme figura 18.

Figura 18: Modelo CSV para importação de análises no software CADUB

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Cliente	Local	Fazenda	Gleba	Área	Data (dd/mm/aaaa)	Nº_Análise	pH	Ca (cmolc/Mg	cmolc/Al	cmolc/H + Al	cn	SMP	MO (%)	Argila (%)	P (mg/dm ²)	K (mg/dm	Condição	Extração_P	
2	SETREM	Três de Maio	Fazenda TCC	Coxilha	35	22/06/2011	3245	5,2	2,4	1,3	0,1	5,2	5,5	2,6	46	6,7	45	1	1	
3																				

Visto que o *software* CADUB foi criado em linguagem VBA (Visual Basic) foi preciso realizar alguns testes antes de se iniciar a importação dos dados propriamente dita para se entender qual a performance do *software*. Este processo seguiu de três tentativas. São elas.

3.4.2.1.1 Importação de uma planilha com 30 mil registros

Foi gerada a estrutura esperada conforme o modelo de importação proposto com as 30 mil análises de solo e então dado início ao processo de importação. Esta tentativa foi falha, pois, o processo de importação dessa quantidade de registros era muito lento, sendo cadastradas em média somente 3 análises a cada 1 minuto. Pelo que verificamos o *software* CADUB abria a planilha Excel selecionada em segundo plano, e a partir disso percorria os valores linha a linha, tornando o processamento mais lento. Mesmo com a lentidão do processo, aconteciam situações em que algum registro estava com dados incorretos na planilha importada e dessa forma a aplicação disparava uma exceção em tela, abortando todo o processo de importação.

3.4.2.1.2 Divisão da planilha em planilhas menores

Visto o insucesso obtido na primeira tentativa de importação dos dados das análises, foi realizada a quebra dos registros em partes menores, dessa forma, se gerou 30 planilhas do Excel cada uma contendo 1000 análises de solo. O processo de importação nesse caso se tornou mais efetivo e eficaz porém ainda não estava em um padrão aceitável, visto que no meio do processo, em alguns arquivos aconteciam erro de conversão de valores, disparando novamente uma exceção em tela e encerrando a aplicação, além disso, a velocidade de importação de cada registro ainda estava lenta, tendo uma média seis análises importadas a cada um minuto percorrido.

3.4.2.1.3 Cadastro das análises via automatização com o AutoIt3

Em virtude do tempo considerável na tentativa de importar as análises de solo no *software* CADUB e pela quantidade de registros importados até o momento (3000 das 30000 análises de solos) se optou por adotar um fluxo diferente, deixando de lado a importação de planilha disponibilizada pelo CADUB e passando a estudar a criação de um script na ferramenta AutoIt3 como forma de simular o cadastro manual de uma análise no *software*.

Esta necessidade surgiu após realizar os testes acima descritos e identificar com o estudo realizado que a ferramenta de importação de análises disponibilizada pelo CADUB não foi criada para realizar a importação de uma massa tão grande de dados e por esse motivo não se tornou performática para a necessidade desse trabalho.

Diante disso, com a nova metodologia de importação dos dados definida se deu início a criação do script responsável por simular o cadastro de uma análise de solo por um usuário. Uma vez que esse fluxo foi definido se integrou ao script a leitura da planilha em Excel com os trinta mil registros criada na primeira tentativa, e então para cada registro foi cadastrada a análise no *software* CADUB.

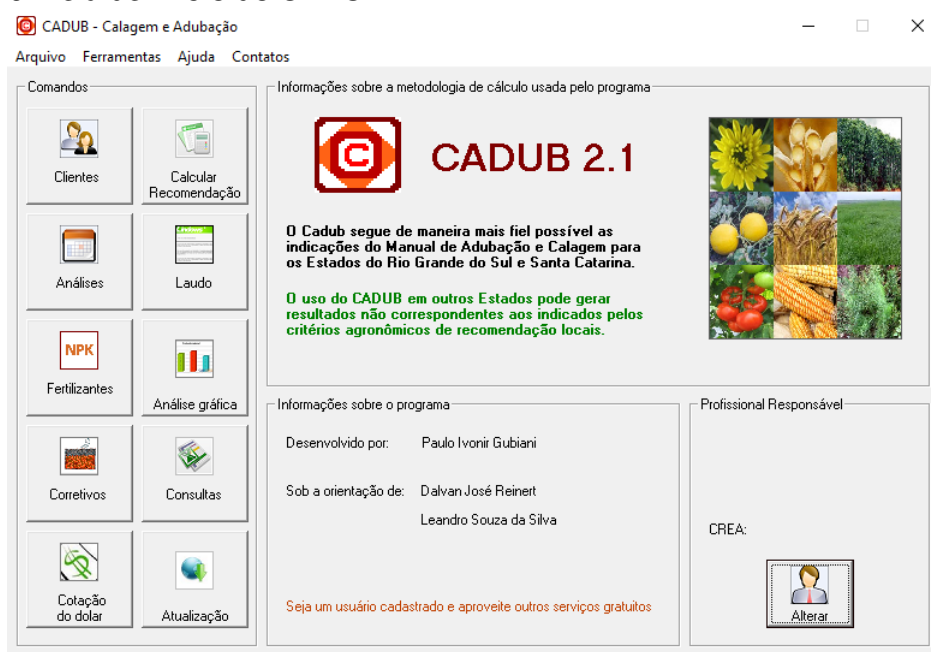
Dentre todas as tentativas testadas essa foi a mais performática visto que foi possível cadastrar aproximadamente doze análises por minuto o que levou em média 42 horas até se concluir todo o processo de cadastro. Outro ponto que

influenciou na efetividade do cadastro foi que por meio do Autolt3 foi possível realizar alguns controles para campos com valores vazios ou incorretos antes mesmo de serem submetidos ao cadastro no *software* CADUB, dessa forma evitou que registros corrompidos ou sem as informações necessárias fossem submetidos ao cadastro.

3.4.2.1.4 Cadastro de análises via automatização

A realização do cadastro de uma nova análise de solo no *software* CADUB acontece por meio do acesso a opção Análises na tela inicial do *software*, conforme figura 19.

Figura 19: Tela de início do CADUB



Fonte: Gubiani and Silva et al. (2007)

Após clicar nessa opção uma nova tela é apresentada com um botão contendo a opção de “Cadastrar Novo”, onde ao clicar nesse botão são habilitados os campos para informar o cliente ao qual a análise se destina, como apresentado na figura 20 na coluna da esquerda, e também todas as informações de componentes químicos encontrados no solo, apresentado na coluna à direita.

Além do preenchimento das informações do cliente e dos componentes químicos encontrados no solo também se marcou para todas as análises cadastradas o método de extração de fósforo como Extração por Mehlich, método esse

Figura 20: Cadastro de análise de solo

Análises [X]

Calculadora Conversor de unidades Importar análises

Dados cadastrais

Cliente:

Local/Cidade:

Fazenda/Propriedade:

Gleba/Lavoura:

Área (ha): Data (dd/mm/aaaa): N° da Análise:

Método de extração de fósforo e condição de área

Extração por Mehlich Extração por resina

Esta área será utilizada para arroz irrigado

Dados da análise de solo

Inseridos

pH (água)	<input type="text" value="4"/>
Ca (cmolc/dm³)	<input type="text" value="5"/>
Mg (cmolc/dm³)	<input type="text" value="6"/>
Al (cmolc/dm³)	<input type="text" value="3"/>
H+Al (cmol/dm³)	<input type="text" value="3"/>
Índice SMP	<input type="text" value="6"/>
MO (% m/v)	<input type="text" value="8"/>
Argila (% m/v)	<input type="text" value="3"/>
P (mg/dm³)	<input type="text" value="4"/>
K (mg/dm³)	<input type="text" value="8"/>

Calculados

Saturação por bases (%)	<input type="text" value="78,60"/>
CTCpH7 (cmolc/dm³)	<input type="text" value="14,02"/>

Os dados da análise devem ser fornecidos nas unidades requeridas pelo programa. Caso precisar, acesse o conversor de unidades no topo.

Fonte: Gubiani and Silva et al. (2007)

utilizado por todos os laboratórios que fazem parte da rede ROLAS. Conforme informado pelo Engenheiro agrônomo Paulo André Klarmann, o método de extração por resina somente é utilizado no estado de São Paulo e como todas as análises coletadas eram do estado do Rio Grande do Sul não foi preciso alterar essa informação. Estes métodos dizem respeito a forma com que o laboratório irá realizar a extração do componente químico P (fósforo), na amostra de solo coletada.

Ao finalizar o preenchimento das informações solicitadas é necessário clicar no botão executar que então registra a análise de solo no banco de dados do sistema CADUB. Todo esse processo foi realizado aproximadamente 30 mil vezes através de um script ¹ de automatização construído por meio do Autolt e no final do processo ter as análises do Laboratório de Solos da SETREM cadastradas no *software* CADUB, com exceção de algumas que possuíam informações faltantes. Essas foram ignoradas.

3.4.2.2 Processo de recomendação

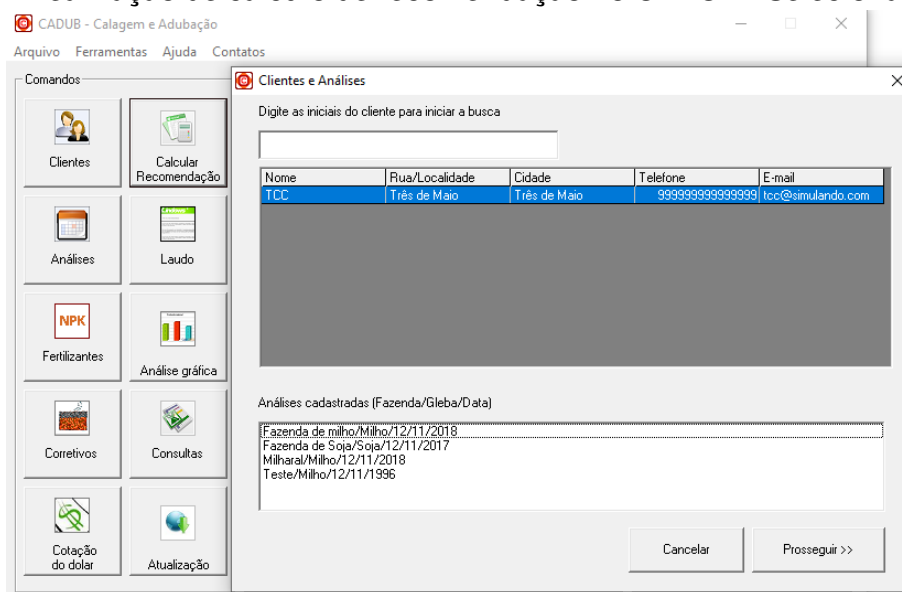
Após a importação das análises no *software* CADUB foi preciso gerar a recomendação de adubação para cada análise. Dessa forma, visto que a geração

¹Código fonte da solução: <https://github.com/larcc-group/tcc-si-ad.git>

manual de cada análise seria impossível para ser feito por um usuário novamente a ferramenta Autolt3, a partir da mesma foi possível criar um script de automatização capaz de simular a ação de um usuário na frente da tela do computador e gerar a recomendação das análises.

O processo de recomendação de uma análise de solo se inicia quando o usuário clica na opção "Calcular Recomendação", em seguida é aberta uma tela para se selecionar o cliente responsável, e ao clicar no cliente são carregadas todas as análises cadastradas para aquele cliente. Como a utilização do *software* CADUB neste trabalho servirá apenas para gerar as recomendações para as análises com base nos manuais de calagem de 2014, apenas um único cliente, com todas as análises importadas em seu nome, conforme demonstra a figura 21.

Figura 21: Realização de cálculo de recomendação no CADUB - Selecionar Análise



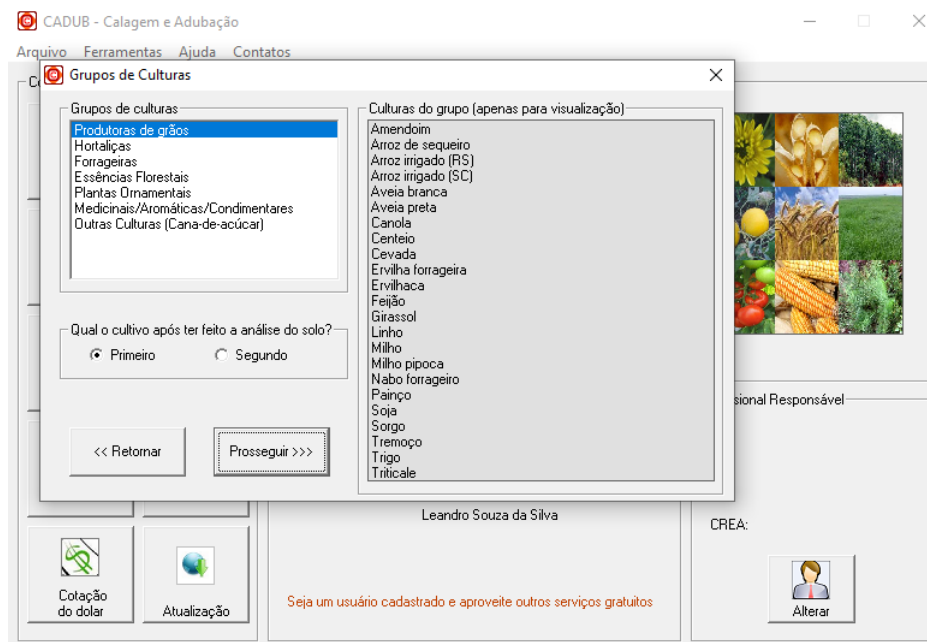
Fonte: Gubiani and Silva et al. (2007)

Na figura 21 é possível ver a lista de análises cadastradas, o que foi feito nessa situação foi programar o script para percorrer cada linha da lista de análises cadastradas, no caso, eram trinta mil e então para cada linha clicar no botão prosseguir para dar seguimento ao processo de recomendação.

Ao clicar nesta opção é aberta uma nova janela para informar qual o grupo de cultura desejada, conforme figura 22. Para criação do modelo proposto foi baseado na cultura de Soja que se encaixa dentro do grupo de grãos, visto que se

tornaria inviável gerar recomendações para todas as culturas disponíveis devido a demora para processar as 30000 análises para cada tipo de cultura.

Figura 22: Realização de cálculo de recomendação no CADUB – Selecionar grupo de cultura

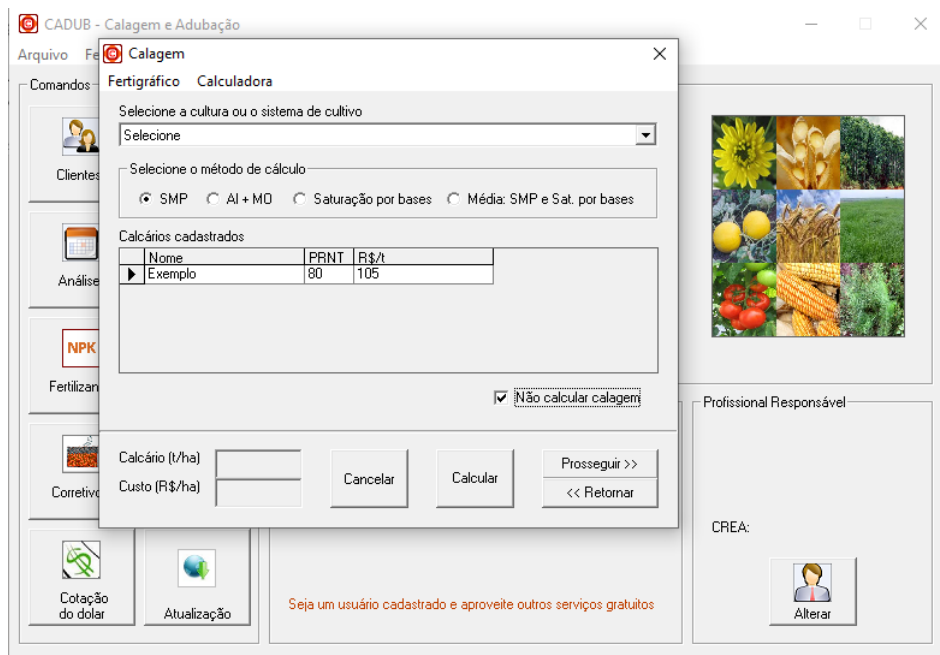


Fonte: Gubiani and Silva et al. (2007)

Outra opção que deve ser informada nessa etapa é a opção de primeiro ou segundo cultivo. Para este caso, inicialmente criamos a recomendação simulando situações de primeiro cultivo e posteriormente situações de segundo cultivo. Nesta etapa, após selecionar a cultura já é realizado um cálculo base e então segue-se para a próxima etapa que é referente a informações de calagem. Como neste trabalho o foco será na recomendação de adubação e não calagem, nesta etapa marcarem a opção "Não calcular calagem" conforme ilustrado na figura 23, seguindo para a próxima etapa.

Após ignorar a etapa de cálculo de calagem é necessário seguir para a tela ilustrada na figura 24, para então informar a cultura à qual se deseja gerar a recomendação e também a expectativa de rendimento em toneladas por hectare. Na expectativa de rendimento, se gerou números aleatórios com rendimentos variados entre um intervalo de 1 a 20 hectares, para se ter uma variância nos dados gerados na recomendação. Após selecionar essas duas informações a próxima etapa é realizar o clique no botão "Calcular" para gerar o cálculo NPK base e posteriormente

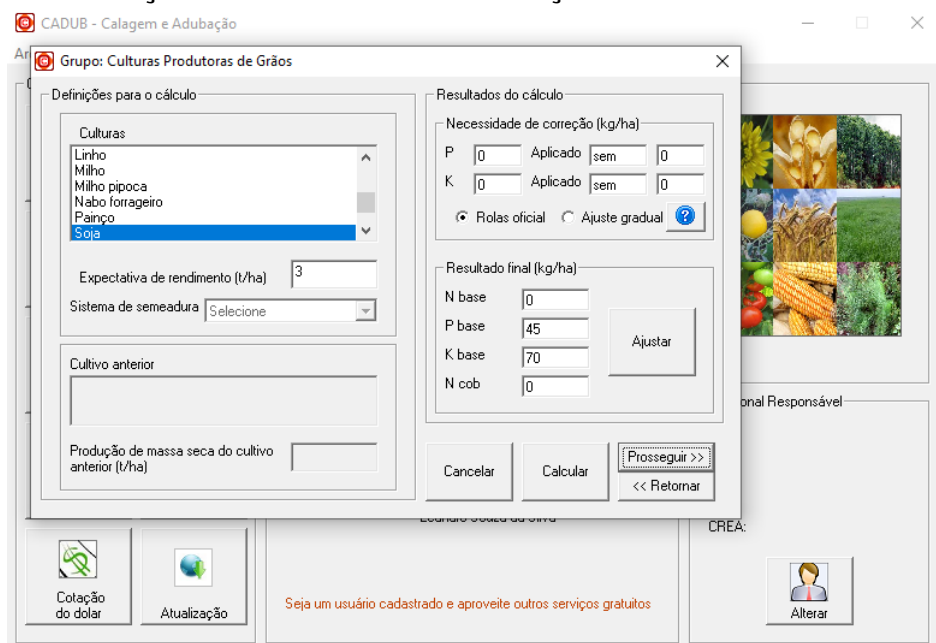
Figura 23: Realização de cálculo de recomendação no CADUB – Ignorar cálculo de calagem



Fonte: Gubiani and Silva et al. (2007)

se executa o clique no botão “Prosseguir” para ir para a próxima etapa.

Figura 24: Realização de cálculo de recomendação no CADUB – Selecionar cultura

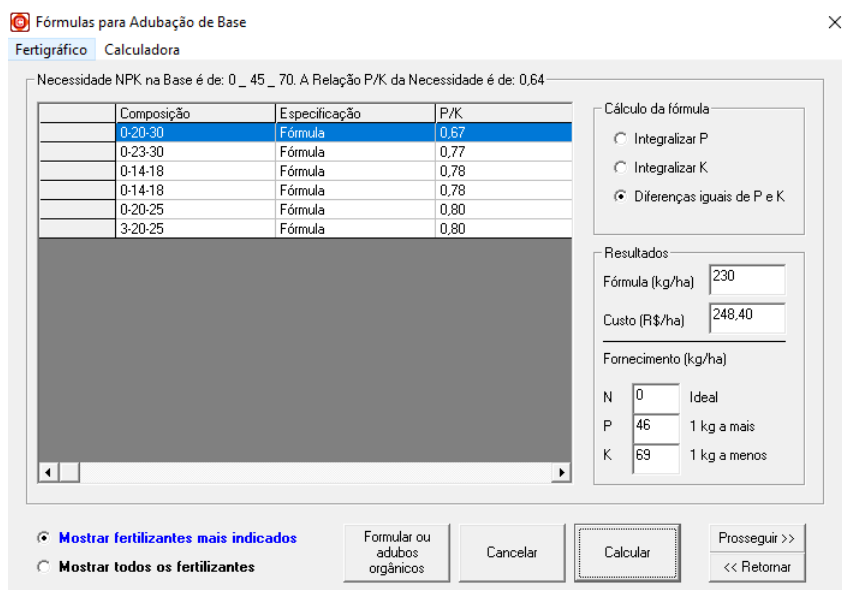


Fonte: Gubiani and Silva et al. (2007)

Com a cultura selecionada, na próxima etapa, são apresentados os fertilizantes recomendados de acordo com as informações inseridas nos passos an-

teriores, como a cultura, a expectativa de rendimento e, principalmente, por meio do cálculo base de NPK realizado no passo anterior, conforme ilustrado na figura 25. Nesta etapa são apresentados vários fertilizantes, o primeiro da lista sempre é o mais indicado. Desse modo, em consideração a primeira fórmula recomendada pelo *software* CADUB com base nos manuais de calagem.

Figura 25: Realização de cálculo de recomendação no CADUB – Selecionar fórmula recomendada



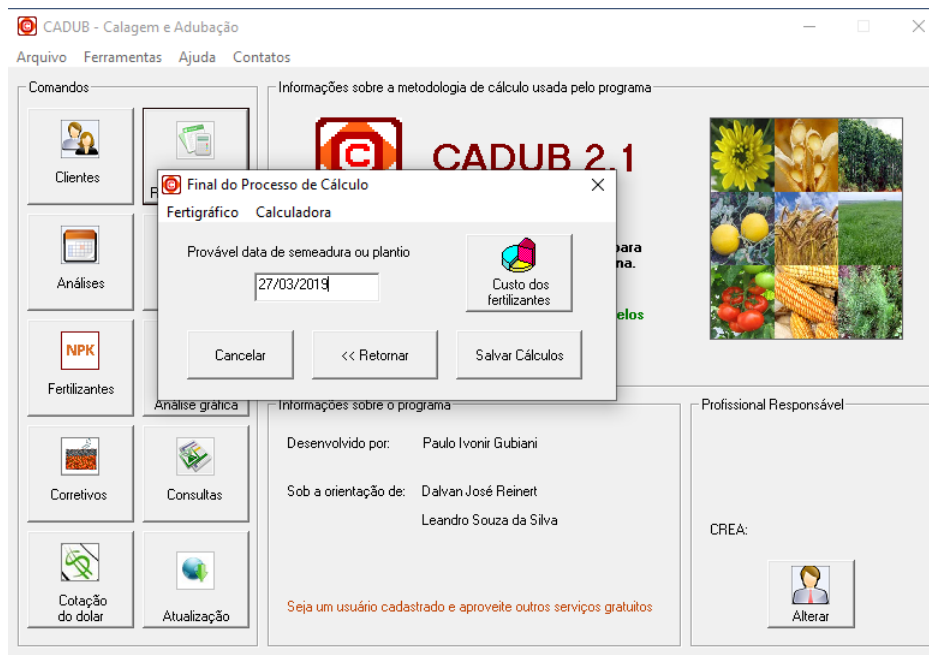
Fonte: Gubiani and Silva et al. (2007)

Após realizada a escolha da fórmula mais indicada segue-se com o cálculo diante da base na fórmula NPK para gerar a quantidade de kg por hectare que deve ser aplicado de determinado adubo. Para isso é realizado um clique no botão calcular e por fim no botão “Prosseguir”, apresentada por final a tela de finalização da recomendação conforme figura 26.

A última etapa do processo compreende a gravação da recomendação na base de dados do *software* CADUB, para isso é necessário preencher uma data referente ao dia da semeadura ou plantio, nesse caso, foi acrescentado uma data padrão para todas as análises.

Vale lembrar que esse processo foi realizado para todas as análises que foram cadastradas na primeira etapa de cadastro das análises, explicadas no tópico 3.4.2.2. O script criado para realização das recomendações de forma automatizada pode ser visto no apêndice D.

Figura 26: Realização de cálculo de recomendação no CADUB – Concluir recomendação



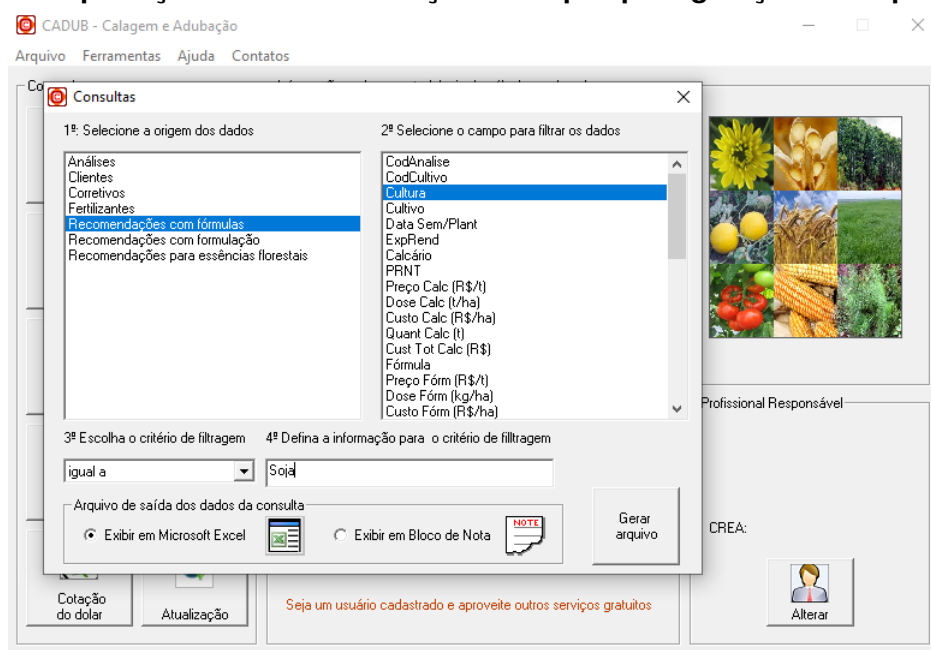
Fonte: Gubiani and Silva et al. (2007)

3.4.2.3 Exportação das Recomendações

Com as análises cadastradas e suas devidas recomendações geradas restou realizar a exportação desses dados, tendo assim o código da análise e sua devida recomendação, dado esse que será utilizado como entrada e saída para o treinamento de um modelo de *Machine Learning*. A entrada terá os componentes químicos encontrados no solo, e como saída (resultado) haverá a fórmula NPK recomendada pelo *software* CADUB e a quantidade em kg por hectare que deve ser aplicada.

Diante disso, para exportar essas informações, também foi utilizado o *software* CADUB, por meio do acesso ao menu Consultas. Após carregar essa opção foi preciso selecionar a origem dos dados. Nesse caso foi selecionada a opção de exportar as recomendações com suas fórmulas, na sequência foi definido o campo “Cultura” que seria o utilizado para realizar o filtro na base de dados e, por fim, definiu-se a busca pela Cultura de “Soja”, que foi a cultura de referência nos cadastros dos passos anteriores e das recomendações conforme ilustrado na figura 27.

Figura 27: Exportação das recomendações - Etapas para geração do arquivo XLS



Fonte: Gubiani and Silva et al. (2007)

Após realizar esta configuração, bastou clicar na opção “Gerar Arquivo” para o *software* carregar os dados solicitados em um arquivo no formato XLS com uma estrutura igual à da figura 28.

Figura 28: Exportação das recomendações - Estrutura gerada em formato XLS

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	CodAnálise	CodCultiv	Cultura	Cultivo	Data Sem/ExpR	Ca	PRN	Preço Calc	Dose Calc	Custo Calc	Quan	Cust	Fórmula	Preço Fórm (R	Dose Fórm (kg/ha)	Custo Fórm (R\$/ha)	Quant Fórm (kg)	
2	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	148	159,8400116	740
3	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	148	159,8400116	2368
4	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	230	248,4000092	230
5	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	230	248,4000092	230
6	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-23-30	1050	349	366,4499817	349
7	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-10-30	1020	248	252,9599915	11160
8	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	148	159,8400116	6364
9	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	148	159,8400116	2812
10	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-10-30	1020	362	369,2399902	362
11	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	3-25-25	960	290	278,3999939	5800
12	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	3-25-20	1080	267	288,3600159	267
13	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	3-30-20	1110	260	288,6000061	3900
14	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	3-25-25	960	310	297,6000061	2170
15	Sítio da INOVAÇÃO	Soja-Prim	Soja	Primeiro	#####	3	0	0	0	0	0	0	0	0-20-30	1080	148	159,8400116	1480

Fonte: Gubiani and Silva et al. (2007)

Como é possível observar, a estrutura gerada contém o código da análise e sua devida fórmula NPK por meio da coluna “Fórmula” bem como a quantidade em kg/hectare que deve ser aplicada por meio da coluna “Dose fórmula”. Por meio desses resultados é possível iniciar a avaliação da viabilidade de uso de *Machine Learning*, como o treinamento de um modelo de treino e a simulação de futuros valores com base nas informações obtidas.

3.4.3 Avaliação do uso de inteligência artificial

Conforme a sugestão proposta, de uma aplicação mobile para se realizar a interpretação da análise de solo e então sugerir a quantidade e tipo de adubação ideal por meio da utilização de inteligência artificial, se iniciou o estudo dos processos para aplicação dessa tecnologia, que são o processo de recomendação de adubação e por segundo o processo de interpretação de imagem das análises de solo e extração dos dados nela contidos com o uso de processamento de imagens.

Diante do primeiro objetivo, o de criar uma inteligência capaz de recomendar a quantidade e tipo de adubação necessária para cada análise de solo submetida foi realizada a divisão das tarefas necessárias para se atingir o resultado esperado. Dentro dessa árvore de atividades foram executadas as tarefas de extração, importação e recomendação das análises por meio do *software* CADUB, conforme citado na seção 3.4.

Durante a execução dessas tarefas, juntamente com um maior conhecimento sobre o funcionamento das recomendações de adubação, bem como, das tabelas de adubação e calagem e de várias conversas com o engenheiro agrônomo Paulo André Klarmann, se teve um debate da real necessidade de utilização de inteligência artificial no primeiro cenário proposto, que diz respeito a parte de geração de laudos de análise de solo.

Visto isso, para dar continuidade ao trabalho por meio da execução do primeiro processo proposto, foi necessário levantar algumas questões a respeito do atual processo de recomendação e definir se a ideia inicialmente proposta por este trabalho poderia ser realmente aplicada para esta área. Dessa forma, foram analisados alguns pontos.

O primeiro ponto a ser destacado, tem referência ao esforço exercido para se gerar os dados de entradas e saídas desejados para criação de um modelo de *Machine Learning*, por meio do uso do *software* CADUB, e o quão trabalhoso e massivo é esse processo, visto que foi preciso tratar cada uma das 30 mil análises separadamente, não sendo possível utilizar o mecanismo de importação de XLS disponibilizado pelo *software* devido a ocorrências de erros durante o processo.

Dessa forma, quando foi realizado a importação de grandes arquivos XLS, o *software* apresentou problemas de lentidão. Para contornar isso, foi preciso gerar diversos arquivos contendo 1000 linhas de dados de análise de solo e mesmo assim o processo foi demorado, levou em torno de 1 dia somente para realizar a importação dos dados.

Quando a importação dos dados foi concluída, foi necessário a geração das recomendações para 30 mil análises uma a uma, pois o *software* não oferece a opção de realizar a recomendação para várias análises de uma só vez. Então, para isso, foi criado um processo em AutoIT, para que o processo de recomendação no CADUB fosse agilizado, porém mesmo assim, a geração levou em torno de três dias para conclusão.

Com maior aprofundamento do manual de calagem e adubação e da forma de disponibilização do mesmo, concluiu-se que hoje não há possibilidade de simular as referências de nutrientes que poderão servir como base para futuros manuais de calagem e adubação. Inicialmente a proposta prévia era de calcular esse tipo de informação com base nas análises de solo coletadas no laboratório, dessa forma o algoritmo poderia ser autossustentável e não depender de futuros manuais de calagem e adubação.

A medida que o aprofundamento do conhecimento na área aumentava, percebeu-se que a obtenção desse tipo de referência é muito mais complexa, já que a atualização da mesma se dá por meio de vários estudos em laboratórios de toda a região que fazem parte da rede de laboratórios ROLAS juntamente ao Núcleo Regional Sul de Ciência de Solo, conforme estudo levantado no capítulo 2 durante a pesquisa bibliográfica.

Outro ponto identificado é que para manter uma maior assertividade em um algoritmo inteligente seria preciso muito mais do que apenas 30 mil análises de solo para se ter um resultado coerente e suficiente para que a simulação com base em um aprendizado de máquina aconteça com um grau de acurácia aceitável, pois nessa situação, poderiam acontecer casos de o aplicativo não retornar nenhum resultado na recomendação.

Além disso, também se identificou, que além da referência nos manuais de calagem e adubação, muitas vezes o engenheiro agrônomo se baseia em seu conhecimento próprio para alterar alguns resultados na recomendação do solo, visando assim, com que o produtor rural consiga, reduzir custos desnecessários com certos nutrientes.

Com isso, percebe-se que além dos manuais de calagem e de adubação também seria preciso recolher informações externas de um grande grupo de produtores rurais, para dessa forma, criar um algoritmo mais assertivo e que melhor se adequa com as mudanças que possam vir a ocorrer no solo em determinadas regiões de acordo com o tempo, já que conforme já dito, os manuais de calagem e adubação são atualizados em períodos maiores de tempo, de 8 a 10 anos.

Conforme visto acima, existem vários pontos que dificultam a utilização do *Machine Learning* para recomendação de análise de solo e além disso, existe o sistema CADUB, que pode ser utilizado para geração de recomendações, pois também tem o mesmo propósito.

Além disso, seria necessário um esforço bastante representativo para utilização de *Machine Learning* para geração de recomendações, pois, é necessário o levantamento de informações de dados estáticos contidos nas tabelas dos manuais de adubação e calagem, assim como também aspectos físicos do solo, químicos, e histórico do solo.

A recomendação de calagem a adubação pode hoje ainda ser atingida com softwares menos modernos e desktops que se baseiam em tabelas estáticas presentes nos manuais, onde neste caso, não se necessita do uso de *Machine Learning* no presente momento, principalmente devido a falta de informações suficientes a respeito, para se criar um sistema mais inteligente, como por exemplo um sistema especialista.

Diante dessa conclusão se teve como seguimento a construção de uma arquitetura de solução, tendo como obtido elencar os processos na área de análise de solos e recomendação que poderiam sofrer melhorias, e diante desses processos, elencar um deles para se aplicar neste trabalho, deixando os demais para

trabalhos futuros, visto a grande complexidade de aplicação de ambos nesta área de solos.

Com isso e como forma de oferecer mais praticidade ao produtor e agrônomo, bem como aos futuros usuários do aplicativo, identificou-se a necessidade do uso de processamento de imagens utilizando a inteligência artificial, pois dessa forma, é possível agilizar o processo de cadastro de informações de análises no aplicativo, que apenas com uma foto irá realizar o reconhecimento das informações para posteriormente oferecer a recomendação.

A utilização de inteligência artificial para reconhecimento de imagem, facilita a identificação automática de textos contidos nas análises, pois se a tecnologia de OCR simples, sem nenhuma adição de tecnologias de *Machine Learning* ou *Deep Learning*, o aplicativo poderia ficar restrito a modelos de análises específicas.

Com base em informações do Sebrae (2017), 95,8% dos produtores rurais já utilizam *smartphone*, demonstrando que os produtores rurais já estão ambientados com a tecnologia, sendo assim o uso da câmera para leitura dos dados de análise de solo tem maior ênfase, pois além de já utilizar a tecnologia, o uso da câmera para preenchimento das informações, agiliza o processo de recomendação.

Ainda de acordo com o Sebrae (2017), 51,9 % dos produtores rurais possuem acesso à internet e 60,5% não utilizam computadores, isso demonstra que muitos produtores tem acesso a internet através do celular, tornando ainda mais viável a leitura de análise de solo e armazenamento dessas informações na nuvem, possibilitando assim a criação de estatísticas sobre a qualidade do solo, os tipos de solos, que ainda podem ser delimitados por região ou até por municípios.

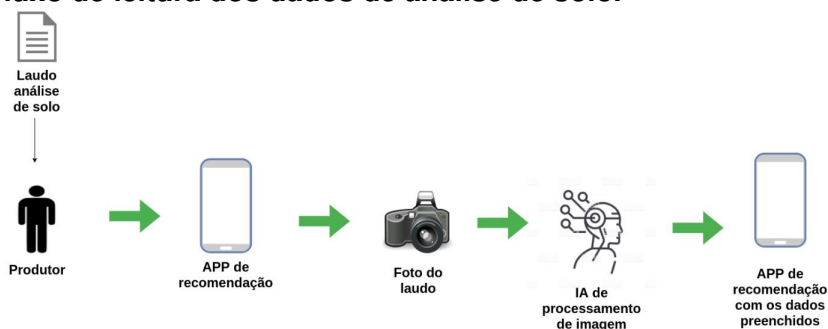
3.5 DESENVOLVIMENTO DA SOLUÇÃO

Com o intuito de agilizar o processo do produtor na geração de dados, o reconhecimento de imagem para preenchimento de informações da análise do solo. Como vimos na figura 15, o produtor necessita preencher várias informações para que a geração de recomendação seja realizada.

Dessa forma, a figura 29 demonstra o fluxo a ser seguido pelo produtor

para utilização do aplicativo. Com a análise de solo em mãos e o aplicativo de recomendação de análise de solo instalado no celular, basta abrir o aplicativo e escolher a opção de tirar uma foto da análise de solo. Com isso, a inteligência artificial do aplicativo já irá gerar as informações na tela do aplicativo sem que seja necessário digitar essa informação para salvar a análise de solo.

Figura 29: Fluxo de leitura dos dados de análise de solo.

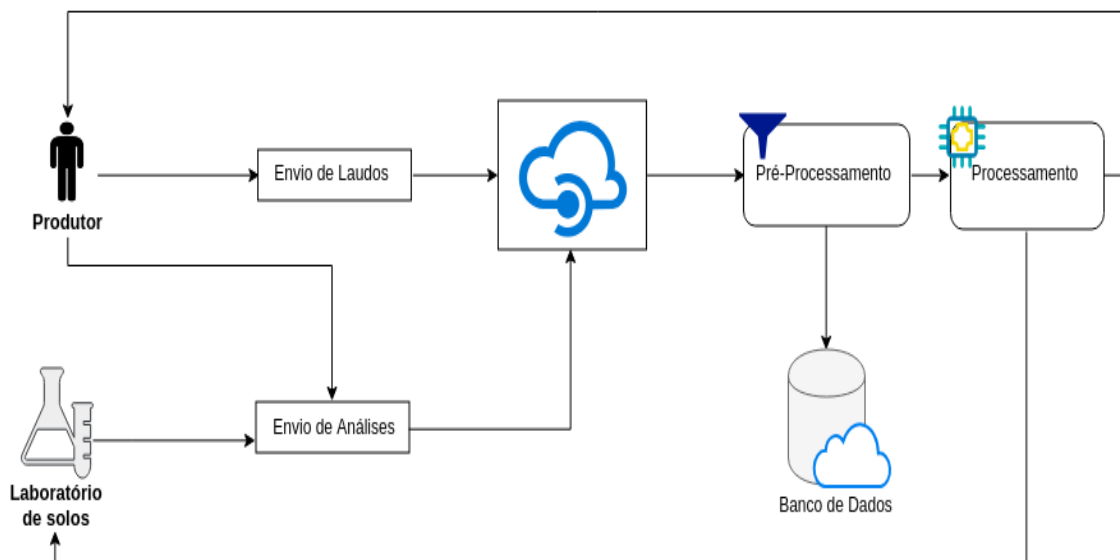


O desenvolvimento na prática do fluxo da figura 29 se dará por meio de três etapas, que serão; desenvolvimento do algoritmo de detecção e reconhecimento de imagens de análises de solo por meio da linguagem Python; desenvolvimento de um aplicativo mobile que será responsável por capturar uma foto da análise de solo e então armazenar as informações obtidas para um uso futuro; e, por fim, a criação de uma API que irá realizar a ligação entre o aplicativo mobile e o algoritmo de reconhecimento de imagem desenvolvido.

3.5.1 Arquitetura da solução

Diante dos estudos realizados e do melhor entendimento do cenário atual ao qual o produtor rural está inserido e também da forma de geração de análises de solo pelos laboratórios se desenvolveu um fluxo que contempla a arquitetura de uma solução que visa agilizar o processo de geração de resultados para o produtor bem como armazenar estatísticas em uma base de dados que poderá ser usada futuramente para mineração e análise dos dados. Este fluxo está descrito na figura 30.

Figura 30: Fluxo da arquitetura de uma solução para melhorar o processo de análise e recomendação



O fluxo se inicia por meio da disponibilização dos dados pelo produtor que pode enviar os dados do solo já analisado por um laboratório, para que a recomendação dessa análise seja realizada.

Esses dados são enviados para o *cloud*, onde serão tratados para normalização e posterior armazenamento no banco de dados. Após isso, eles serão processados resultando em uma interpretação da análise de solo para o produtor ou o laboratório de solo.

No pré-processamento, os dados serão tratados, de maneira que possam ser armazenados de forma normalizada para posteriores consultas. Esse tratamento prepara os dados da análise de solo, que podem estar em diversos formatos, por exemplo, png, jpeg, pdf, xlsx, txt, entre outros. Somente após a execução desse processo os dados estarão prontos para persistência no banco de dados.

Tendo os dados tratados, é possível processá-los a fim de gerar a recomendação da análise de solo, a qual é calculada com base nas tabelas dos manuais de adubação e calagem. Concluída a execução do cálculo, irá resultar no retorno da requisição para o produtor ou o laboratório de solo.

Com o uso contínuo da ferramenta, o produtor poderá utilizar os dados históricos para auxiliar na gestão da propriedade, extraíndo, por exemplo, informações

relacionadas ao histórico de fertilidade de áreas já analisadas, acompanhando a evolução de cada área, deste modo fornecendo uma potente ferramenta para auxiliar na tomada de decisões.

Além da aplicabilidade da solução para o produtor rural e laboratório de análise de solos, os dados coletados podem ser utilizados para analisar a qualidade de solo em diferentes áreas de municípios e regiões, isso fará com que a tomada de decisão do produtor rural seja mais efetiva, quanto a quantidade de adubos e calcário a ser aplicado.

O aplicativo também pode disponibilizar para empresas de vendas de insumos, como por exemplo, dados atualizados de como está a fertilidade dos solos na área abrangida pela empresa. Dessa forma, o vendedor poderá concentrar suas forças em locais que a fertilidade das áreas está mais baixa, aumentando assim a chances de aquisição de insumos junto a empresa.

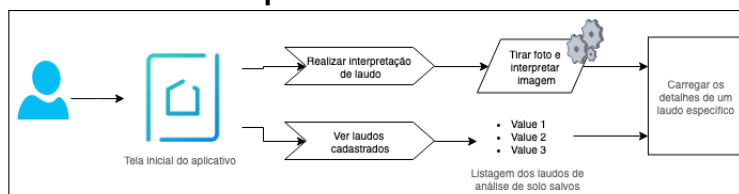
Outra funcionalidade que pode ser disponibilizada no aplicativo é um mapa com os tipos de solos nas diferentes regiões, como por exemplo, se o solo é mais arenoso, ou se é um solo de terra vermelha, entre outros, podendo assim servir inclusive para pesquisas com foco nesse segmento.

Vista toda a complexidade presente na área de Machine Learning, esta pesquisa limitou-se na interpretação e reconhecimento de imagens de análise de solo por meio da utilização de processamento de imagens e reconhecimento de caracteres através da tecnologia de Machine Learning. As outras propostas citadas na arquitetura de solução poderão ser desenvolvidas em trabalhos futuros.

3.5.2 Aplicativo mobile

Este tópico tem por objetivo apresentar o protótipo de aplicativo mobile desenvolvido para realizar a integração com o algoritmo de interpretação de análise de solos. Para o desenvolvimento do aplicativo utilizou-se o framework React-Native, possibilitando gerar versões nativas do aplicativo tanto para a plataforma Android quanto para IOS.

A figura 31 demonstra as ações que podem ser executadas pelo usuário ao abrir o aplicativo.

Figura 31: Ações do usuário no aplicativo

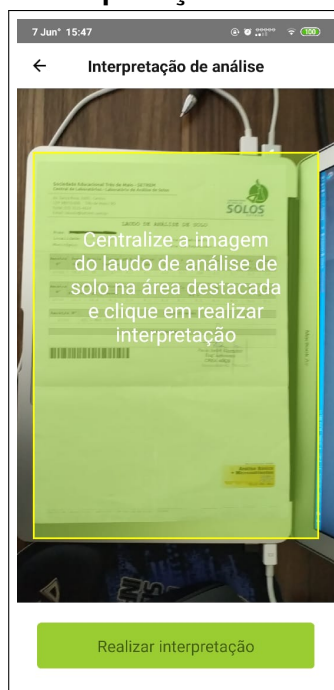
Como é possível ver na figura 31, ao abrir o aplicativo o usuário é direcionado para a tela de início, possuindo duas opções, realizar a interpretação de análises e ver os mesmos já cadastrados. A tela inicial do aplicativo, bem como as opções disponíveis para o usuário podem ser vistas na figura 32.

Figura 32: Tela inicial do aplicativo

Diante disso, caso o usuário escolha a opção de realizar a interpretação de análises, o mesmo é direcionado para uma tela onde o aplicativo passa a ter acesso a câmera do dispositivo móvel, conforme é possível acompanhar por meio da figura 33.

O objetivo da etapa ilustrada na figura 33 é que o usuário tenha a possibilidade de retirar uma foto da análise de solo que possui e, com base na foto capturada, o aplicativo envie a mesma para uma API que irá executar o algoritmo de processamento e interpretação de imagens.

Para que a foto seja capturada de uma maneira mais efetiva e correta foi adicionado uma máscara, destacada em amarelo, que sobrepõe a imagem da câ-

Figura 33: Captura de foto para interpretação - alinhamento correto

mera, indicando ao usuário o espaço exato e o orientando que o mesmo não tire fotos de ponta-cabeça, para que dessa forma a análise de solo seja alinhado corretamente e obtenha uma melhor interpretação.

Com a imagem da análise devidamente alinhada, basta que o usuário clique na opção Realizar Interpretação para que o processo de interpretação se inicie. Neste momento a foto é capturada e enviada para a API que irá seguir com o processo de interpretação por meio do uso da ferramenta Tesseract OCR. Este processo pode demorar até 30 segundos e, enquanto isso, o usuário recebe um aviso de que o processo está em andamento, conforme figura 34.

Após o término do processamento, o usuário é direcionado para uma tela que contém os detalhes de todos os elementos químicos encontrados na imagem e seu devido valor. Caso algum elemento não tenha sido reconhecido corretamente ou não ter algum valor numérico válido o mesmo é retornado como N.D (Não definido). Na figura 35 é possível ver os detalhes da análise que havia sido submetido para interpretação.

De acordo com a figura 35 nota-se que todos os elementos químicos presentes na imagem foram identificados com sucesso, visto que os elementos da segunda tabela que estão como N.D já se encontravam assim na análise. Caso

Figura 34: Processamento da interpretação

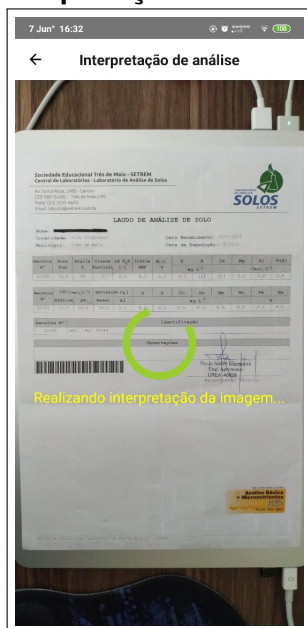


Figura 35: Detalhes da análise de solo interpretada

7 Jun° 15:48

← Detalhes da análise de solo

Análise lavoura milho

Amostra
20352

Tabela 1

Área	Argila	Classe textural	pH _{H2O}
19.0	90	1	6.0
Índice SMP	M.O	P	K
6.1	2.2	9.0	126
Ca	Mg	Al	H+Al
9.1	3.1	0.0	3.8

Tabela 2

CTC Efetiva	CTC PH	Saturação Base	Saturação Al
12.9	17.3	74.8	0.0
S	B	Cu	Zn
N.D	N.D	N.D	N.D
Mn	Mo	Fe	Na
N.D	N.D	N.D	N.D

Salvar análise

algum elemento não esteja com o valor correto, é possível que o usuário ajuste o valor clicando no campo correspondente e, então, após corrigir, basta clicar na opção salvar análise, que o registro será armazenado no aplicativo para posterior consulta e compartilhamento com pessoas interessadas, como um engenheiro agrônomo ou algum banco.

Outra opção possível através do aplicativo é a visualização das análises registradas. Para isso, seguindo da tela inicial, basta clicar na opção, Ver análises já importadas, que serão listados todas as análises de solo já submetidos e interpretados, conforme figura 36, podendo, a partir de então, compartilhar os mesmos, excluir ou editar informações dos elementos químicos encontrados.

Figura 36: Listagem de análises de solo interpretadas



É importante destacar que como trata-se de um protótipo de aplicativo e não um produto final, foram implementadas apenas as partes básicas necessárias para realização da interpretação de imagens de análise de solo.

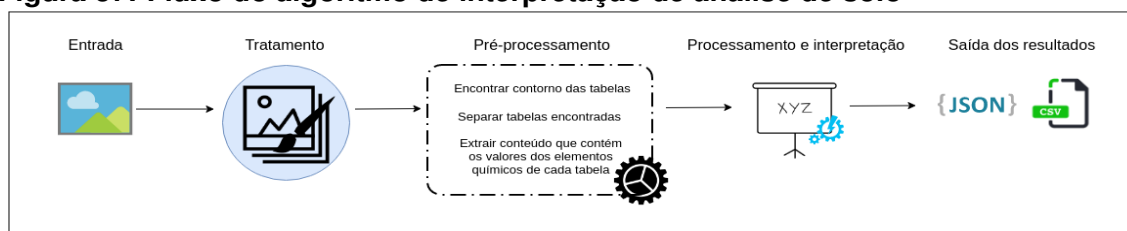
3.5.3 Algoritmo para detecção e reconhecimento de análises de solo

Esta seção tem por objetivo descrever as etapas realizadas e métodos utilizados para a criação de um algoritmo capaz de detectar as informações presentes em uma análise de solo através da captura de imagem por um dispositivo móvel. Os métodos descritos e códigos implementados na linguagem de programação Python estão disponíveis online².

É abordado a etapa de tratamento da imagem, processamento da imagem tratada, interpretação da imagem processada e, por fim o processo de verificação de acurácia diante de cada imagem de análise de solo submetida para interpretação.

Para uma melhor interpretação do que será desenvolvido, é possível ver a ilustração da figura 37

Figura 37: Fluxo do algoritmo de interpretação de análise de solo



3.5.3.1 Escolha da ferramenta de interpretação e considerações

Esta seção tem por objetivo descrever as ferramentas e serviços estudados para auxiliar na resolução do problema de interpretação de uma imagem de uma análise de solo e, com base nisso, extrair os valores dos elementos químicos presentes na análise.

Para resolução do problema foram estudadas 3 possibilidades de uso, tendo em vista a relação da ferramenta com a tecnologia de Machine Learning. Os estudos se desenvolveram diante das ferramentas TensorFlow e Tesseract OCR e da API Vision, ambos projetos mantidos pelo Google.

²Código fonte da solução: <https://github.com/larcc-group/tcc-si-ad.git>

3.5.3.1.1 TensorFlow

Inicialmente analisou-se a utilização da biblioteca TensorFlow juntamente com algum *dataset* já existente que fosse capaz de interpretar números digitais. Juntamente com isso verificou-se a complexidade e tempo de implantação de uma funcionalidade desse nível, onde com várias etapas, como verificação de um *dataset* existente já compatível com o problema, treinamento por épocas exigindo um maior poder computacional, tempo para treinamento do modelo entre outros fatores.

Dentre os datasets verificados para utilização com a biblioteca TensorFlow está o MNIST ³, sendo esse um *dataset* de dígitos manuscritos com mais de 60 mil exemplos para treinamento e 10 mil exemplos para teste e que se mostrou o mais promissor para uma possível implementação para resolução do problema. De qualquer forma, a maioria dos números que representam os valores dos elementos químicos da análise de solo são números com vírgula e não foi encontrado nenhum caso de uso deste *dataset* para números com vírgula, o que reduziu a confiança de implementação nele baseada.

Outro ponto que foi levado em consideração na escolha da ferramenta foi o tempo que levaria para se implementar determinada solução e, neste caso, não se utilizou esta biblioteca, visto que iria demandar de um grande aprendizado técnico e, o tempo poderia não ser suficiente ficando, assim, a implementação com o uso do TensorFlow para um possível trabalho futuro.

3.5.3.1.2 Tesseract OCR

Com a decisão de não utilizar o TensorFlow, se iniciou o estudo e testes com a ferramenta Tesseract OCR, ferramenta esta, disponibilizada pela Google em código aberto no Github. Diante disso, buscou-se entender o seu funcionamento e seus casos de uso para reconhecimento de caracteres em imagens e qual a relação da mesma com a área de Machine Learning.

A ferramenta Tesseract OCR possui como entrada três parâmetros principais, sendo eles o `-OEM`, o `-PSM` e o `-l`. O parâmetro `-OEM` diz respeito ao motor

³MNIST *dataset*: <http://yann.lecun.com/exdb/mnist/>

de processamento que será utilizado para buscar interpretar os caracteres. Neste caso existem quatro opções disponíveis, conforme quadro 4.

Quadro 4: Valores aceitos para o parâmetro –OEM

Valor	Descrição
0	Utilização do motor legado
1	Utilização do motor com redes neurais LSTM
2	Utilização do motor legado em conjunto com LSTM
3	Utilização do motor padrão, com base na disponibilidade

Fonte: OCR (2019)

O outro parâmetro a ser passado para a ferramenta é o –PSM, o qual diz respeito ao método que será utilizado para realizar a segmentação de página, ou seja, qual será a região de busca de conteúdo que a ferramenta deverá levar em consideração ao realizar a interpretação. Existem 13 opções de valores disponíveis para esse parâmetro conforme quadro 5.

Quadro 5: Valores aceitos para o parâmetro –PSM

Valor	Descrição
0	Apenas orientação e detecção de scripts (OSD).
1	Segmentação de página automática com o OSD.
2	Segmentação de página automática, mas sem OSD ou OCR.
3	Segmentação de página totalmente automática, mas sem OSD. (Padrão)
4	Suponha uma única coluna de texto de tamanhos variáveis.
5	Suponha um único bloco uniforme de texto alinhado verticalmente.
6	Suponha um único bloco uniforme de texto.
7	Trate a imagem como uma única linha de texto.
8	Tratar a imagem como uma única palavra.
9	Tratar a imagem como uma única palavra em um círculo.
10	Trate a imagem como um único caractere.
11	Texto escasso. Encontre o máximo de texto possível em nenhuma ordem específica.
12	Texto escasso com o OSD.
13	Tratar a imagem como uma única linha de texto Ignorando hacks que são específicos do Tesseract OCR.

Fonte: OCR (2019)

O último parâmetro a ser passado para a ferramenta é o `-l`, este parâmetro diz referência a linguagem que deve ser considerada na interpretação. Atualmente já existem 102 linguagens disponíveis no repositório do projeto. Dentre elas o inglês, alemão e português.

De acordo com OCR (2019), a partir da versão 4.0 foi implementado um mecanismo de reconhecimento baseado em redes de memória de longo prazo, LSTM que se trata de uma rede neural recorrente (RNN), utilizada no campo de aprendizagem profunda. Esta nova opção é disponibilizada pelo Tesseract OCR por meio do parâmetro `-OEM = 1`.

Conforme Colah (2019), a principal diferença entre uma rede neural recorrente comum e uma rede neural recorrente LSTM está na capacidade de persistência de dados, permitindo assim recordar determinada informação por um longo período de tempo não sendo necessário executar todo processo de aprendizado novamente a cada novo ciclo.

Vista essa possibilidade de utilização de redes neurais LSTM por meio do Tesseract OCR, realizam-se alguns testes visando identificar se a ferramenta realmente atenderia as necessidades de interpretação de análise de solo e também comparar entre o antigo motor de legado já existente e o novo motor disponibilizado na versão 4.0. Para realização dos testes foi utilizada a análise de solo da figura 15 por meio de testes com o parâmetro `-OEM = 0` e `-OEM = 1` para três diferentes tipos de segmentação de página e para a linguagem português, `-l = por`. Os resultados dos testes podem ser vistos na figura 38.

Na figura 38 pode ser vista a diferença entre os dois motores. Como vemos o motor do legado, de uma forma geral, reconheceu menos informações que o motor de redes neurais LSTM e também teve maiores dificuldades para interpretar os números referente aos elementos químicos. Outro ponto que merece destaque é quanto ao tempo médio de processamento: em todas as situações o motor de redes neurais se saiu melhor, conseguindo executar o processamento da imagem em menos de 5 segundos, diante de 10 segundos do motor legado.

Em ambos os casos, mesmo o motor de redes neurais se saindo melhor,

é possível notar que nenhum dos testes retornou sucesso absoluto, muitos valores ficaram quebrados ou incoerentes com o real valor representado na imagem que foi submetida, porém isso se justifica porque a imagem que foi submetida não passou por nenhum tipo de tratamento de cores, contraste, luminosidade, fazendo com que a precisão exata das informações fosse comprometida.

Tendo isso em vista, a ferramenta Tesseract OCR como um possível candidato para resolução do problema de interpretação por meio do motor de redes neurais LSTM, visto que o processamento é rápido e a integração com a mesma é simples e prática, sendo necessário apenas, realizar testes com uma imagem já tratada.

3.5.3.1.3 Google Vision

A última solução a ser estudada foi a API de reconhecimento de imagens Google Vision. A mesma disponibiliza uma API Rest onde é possível submeter arquivos de imagens e receber como retorno todas as informações presentes na imagem, desde *insights* sobre objetos da imagem, imagens relacionadas, cores provenientes e também a detecção de textos, tudo isso através de modelos pré-treinados.

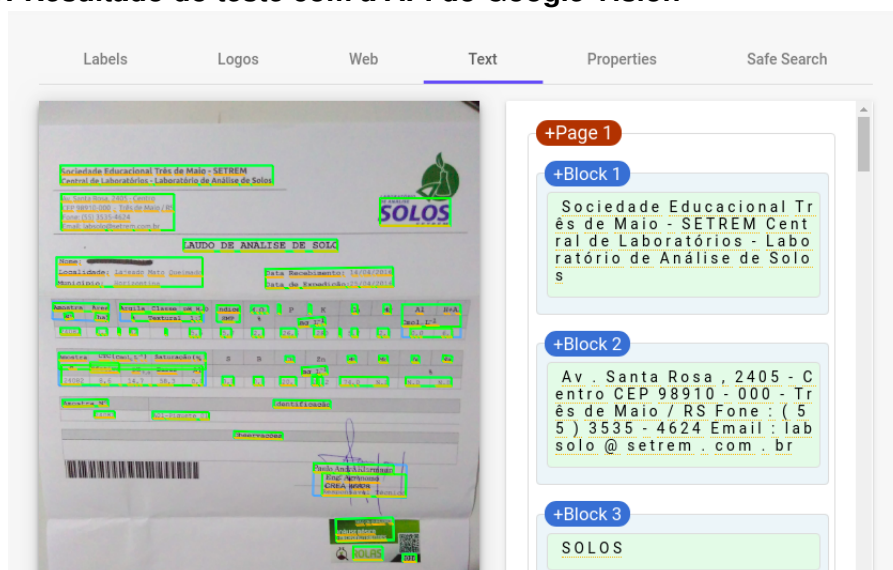
A Google disponibiliza um ambiente online ⁴ para se realizar testes com imagens, sem ser necessária realizar uma integração com a API. Diante disso, se submeteu a imagem da figura 15 para analisar o resultado do processamento obtido. Os detalhes desse processamento podem ser vistos na figura 39.

Na figura 39 é possível ver que o algoritmo de detecção de texto da API Google Vision conseguiu identificar grande parte das informações textuais presentes na imagem, separando os registros por blocos de texto. Cada bloco é representado por um quadrado amarelo presente na imagem.

A API retorna as informações no formato JSON, sendo possível, assim, percorrer todos os blocos de texto encontrados ou também recuperar uma concatenação com todas as informações dos blocos já agrupadas, conforme é possível ver na figura 40

⁴<https://cloud.google.com/vision/>

Figura 39: Resultado do teste com a API do Google Vision



Fonte: Cloud (2019)

Figura 40: Json de retorno com o conteúdo dos blocos agrupado

```
{
  "cropHintsAnnotation": { "cropHints": [] },
  "fullTextAnnotation": {
    "pages": [ { "text": "Sociedade Educacional Três de Maio - SETREM\nCentral de Laboratórios - Laboratório de Análise de Solos\nAv. Santa Rosa, 2405 - Centro\nCEP 98910-000 - Três de Maio/RS\nFone: (55 ) 3535-4624\nEmail: labsolo@setrem.com.br\nSOLOS\nLAUDO DE ANÁLISE DE SOLO\nNome :Lajeado Mato Queimado\nMunicípio: Horizontina\nData Recebimento: 14/04/2016\nData de Expedição:25/04/2016\nAmostra Área\n(ha)\nArgila Classe pH H,\n% Textural 1:1\nM.0 .\nÍndice\nSMP\nP\nK\nCa\nMg\nNO\nmg L-1\nAl H +AL\nCmol Li\n0,0 6,1\n24082\n0,5\n1\n42\n2\n5 ,6\n5,7\n2.4\n26.6\n200\n5\n,7\n2 ,2\nCu\nMn\nMo\nFe\nNa\nAmostra CTC (Cmol L 1) Saturação (%) \nNº \nEfetiva \nTeata \npH Bases AI\n24082 8,6 | 14,7 | 58,3 0,0\nmg L-1\n3,0\n0 ,5\n20,7\n24,\n2\n74,0 ND\nND ND\nAmostra Nº \nIdentificação\n24082\nA01-Piquete 01\nObservações\nPaulo And alarmann\nEng Agrônomo\nCREA 6828\nResponsável técnico\nSELO DE PUNTO DE\nANAUSE BASICA\n+MICRONUTRIENTES\nROLAS\n2016\n" } ],
  }
}
```

Fonte: Cloud (2019)

Analisando os resultados do retorno em formato JSON percebeu-se que muitos valores, principalmente os presentes na tabela de elementos químicos não eram retornados em ordem, o que dificultaria o processo de análise, visto que não seria possível saber que determinado valor fazia referência a determinado componente químico.

Outro fator levado em consideração é que a qualidade da imagem e seu conteúdo interferem diretamente na qualidade dos resultados, conforme comentado na análise da ferramenta Tesseract OCR na seção 3.5.3.1.2. Diante disso, acredita-se que uma imagem tratada poderia retornar resultados mais satisfatórios.

Visto que a API do Google Vision conseguiu interpretar boa parte das informações de texto presentes na imagem, a mesma como uma possível candidata para resolução do problema de interpretação. Dessa forma foram realizados testes nas próximas etapas, após o tratamento da imagem da análise, juntamente com a ferramenta Tesseract OCR para se ter um comparativo de qual a melhor ferramenta a ser utilizada.

3.5.3.2 *Tratamento da imagem*

Por meio de uma breve análise da imagem que foi definida como entrada para os testes e desenvolvimento do algoritmo verificou-se que existem muitas informações paralelas que devem ser separadas, como as informações do laboratório de solo e informações de localidade, município, data de recebimento e expedição, conforme é possível ver na figura 15 (p. 91).

Além disso, para buscar aumentar a acurácia com as ferramentas de reconhecimento Tesseract OCR e Google Vision, era preciso aplicar filtros de ajustes na imagem. Diante disso, o primeiro passo executado foi o tratamento da imagem da figura para buscar ajustar os níveis de luminosidade e contraste, reduzir ruídos, redimensioná-la e alinhá-la, bem como converter a imagem em uma escala de cinza para ser melhor trabalhada no restante do processo.

Grande parte do tratamento da imagem foi realizado por meio da utilização do *software* ImageMagick que tem por objetivo realizar edições em imagens por

meio de chamadas personalizadas via linha de comando ou através de scripts já existentes, criados pela comunidade para executar determinada tarefa de melhoria em uma imagem.

O *software* foi criado em 1987 por John Cristy, com o objetivo inicial de converter imagens de 24 bits em imagens menores de 8 bits, diminuindo assim a quantidade de cores possíveis presentes na imagem, para que fosse possível exibir essas imagens na maioria das telas existentes na época. Após 1990 o *software* que era de código fechado, passou a ser livre e John Cristy transferiu os direitos para o ImageMagick Studio LLC, existente até hoje.

Além do *software* ImageMagick também se utilizou a biblioteca OpenCV conforme citado na pesquisa bibliográfica no tópico 2.3.4.1.3.

3.5.3.2.1 Correção de níveis, luz e contraste

Um dos principais ajustes a ser realizado em uma imagem é o controle de níveis, visto que grande parte das fotos profissionais hoje capturadas recebem um ajuste deste tipo. A aplicação da correção de níveis geralmente é mais utilizada em fotos com tons de cinza ou fotos coloridas que foram convertidas para escala de cinza, se adequando exatamente à situação do problema, visto que as fotos serão convertidas para uma escala de cinza para serem processadas.

De um modo geral esse tipo de ajuste tem por objetivo criar um equilíbrio na luminosidade da foto, alternando os níveis de intensidade entre preto, cinza-médio e branco, fazendo com que a imagem tenha, assim, alguns tons realçados e outros suavizados.

No algoritmo proposto a correção dos níveis da imagem por meio do ImageMagick, já que o mesmo possui este tipo de ajuste de forma nativa. A chamada da função de ajuste de níveis por meio da linguagem Python pode ser vista no código 1 que também está disponível por meio do arquivo tools.py⁵.

Código 1: Método de tratamento de níveis da imagem

```
1 def levelTreatment(inputoutput):  
2     command = 'convert'
```

⁵Código fonte da solução: <https://github.com/larcc-group/tcc-si-ad.git>

```

3     params = '-colorspace Lab -channel 0 -auto-level +channel -colorspace
4     sRGB'
5     finalCommand = command + " " + inputoutput + " " + params + " " +
inputoutput
os.system(finalCommand)

```

Na função em questão inicialmente é definido o espaço de cores que será utilizado na aplicação do controle de níveis por meio do parâmetro `-colorspace` recebendo como propriedade o valor `Lab`, que é um espaço de cor amplamente utilizado neste tipo de tratamento, correlacionando de forma muito consistente os valores de cor com a percepção visual.

Conforme Jun (1997), a Comissão Internacional em Iluminação (Commission Internationale Declairage – CIE), o espaço de cor LAB lida com 3 canais diferentes. O canal L que armazena as informações de luminosidade e os canais a e b, que comportam a informação de cor. Dentre os dois, o canal 'a' caracteriza valores positivos que simbolizam a cor magenta, e negativos, que simbolizam a cor verde, enquanto o canal 'b' caracteriza valores positivos que simbolizam azul, e negativos, amarelo.

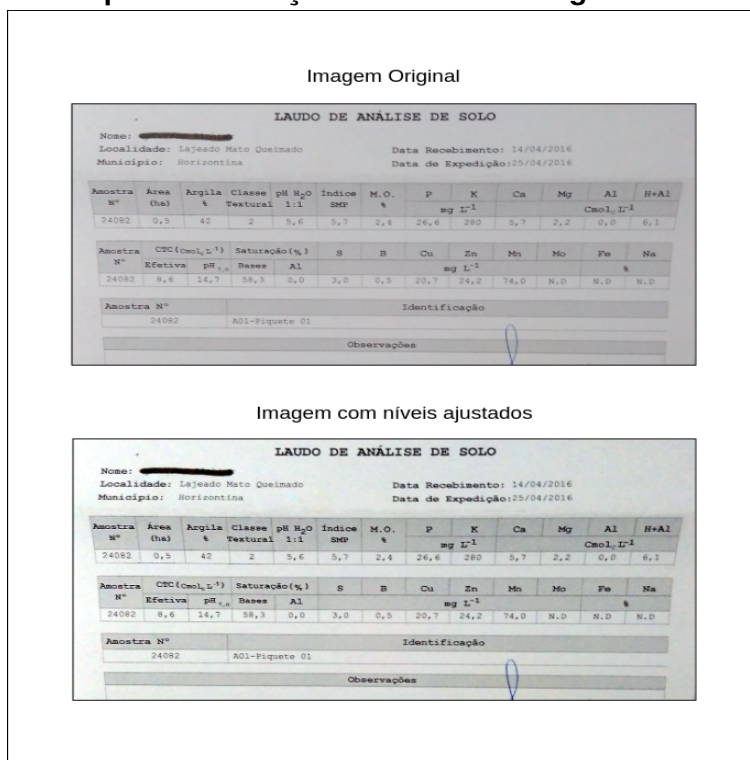
Outro parâmetro é o `-channel` que significa os canais de cores que serão utilizados para determinação do ajuste de nível. Neste caso se informou como 0, significando que serão utilizados os 3 canais, R, G e B, ou seja, vermelho, verde e branco.

Pode-se perceber que ainda existem outros três parâmetros, o `-auto-level` que diz respeito a qual tipo de tratamento realizar, nesse caso o tratamento de níveis, sendo logo após, informado um canal de filtragem baseado no espaço de cores sRGB. Por meio da execução da função de auto correção de níveis se obteve o resultado da figura 41.

3.5.3.2.2 Redução de ruído

Conforme Lagendijk and Biemond (2009), o ruído são pixels espalhados aleatoriamente, que são facilmente perceptíveis já que não fazem parte da imagem em si, sendo oriundos de uma exposição muito alta da imagem ou também devido

Figura 41: Antes e depois da correção de níveis da imagem



a captação de uma imagem muito escura em uma exposição muito longa. Dessa forma, a redução de ruído tem por objetivo remover esse efeito granulado, bem como algum conteúdo indesejado ou alguma sujeira que possa existir na imagem.

Dentro do *software* ImageMagick foi utilizado o script `textCleaner`, criado especificamente para realizar a redução de ruído em imagens e trata-las com o objetivo de otimizar a leitura das mesmas em um leitor de OCR e detecção de texto. Diante disso, se executou o script por meio da linguagem Python na chamada do método `textCleaner()` apresentado no código 2.

Código 2: Método de tratamento de ruído da imagem

```


1 def textCleaner(inputPath, outputPath):
2     command = './textcleaner'
3     params = '-g -e stretch -f 20 -o 2 -t 7 -s 1 -T -p 20'
4     finalCommand = command + " " + params + " " + inputPath + " " +
5     outputPath
6     os.system(finalCommand)

```

Basicamente o que é solicitado para o script é que inicialmente o mesmo

converta a imagem recebida em uma escala de cinza, por meio do parâmetro -g e após isso aprimore o brilho da imagem por meio do parâmetro -e, aplicando um filtro de tamanho 20 por meio do parâmetro -f que significa o grau de limpeza que será aplicado na imagem. Os demais parâmetros dizem respeito a suavização da imagem e tamanho total de deslocamento do filtro. Após a execução do script se obteve o resultado apresentado na figura 42.

Figura 42: Resultado do tratamento de ruído

LAUDO DE ANÁLISE DE SOLO														
Nome:														
Localidade:		Data Recebimento: 29/03/2019												
Município:		Data de Expedição: 05/04/2019												
Amostra NR	Área (ha)	Argila %	Classe Textural	pH H ₂ O 1:1	Índice SNP	M.O. %	P		K		Ca		Mg	
							mg L ⁻¹		mg L ⁻¹		mg L ⁻¹		mg L ⁻¹	
42424	21,00	81	1	5,7	5,5	3,7	22,5	94	5,8	2,0	0,0	7,7		
Amostra NR	CTC (cmol _c L ⁻¹)		Saturação (%)		S	B	Cu		Zn	Mn	Mo	Fe	Na	
	Efetiva	pH _{7,0}	Bases	Al			mg L ⁻¹							%
42424	8,0	15,7	50,9	0,0	11,7	0,4	7,0	4,3	33,6	N.D	N.D	N.D	N.D	
Amostra NR	Identificação													
42424	A01 - Mat 12334													
Observações														
														
_____ Responsável Técnico														

Na figura 42 é bastante visível perceber que todo o fundo da imagem passou a ser branco, principalmente no cabeçalho da tabela. Além disso, também houve a redução de ruídos existentes, o aumento do contraste do texto existente, a intensificação das bordas e cantos da tabelas e um ajuste na iluminação da imagem.

3.5.3.2.3 Alinhamento

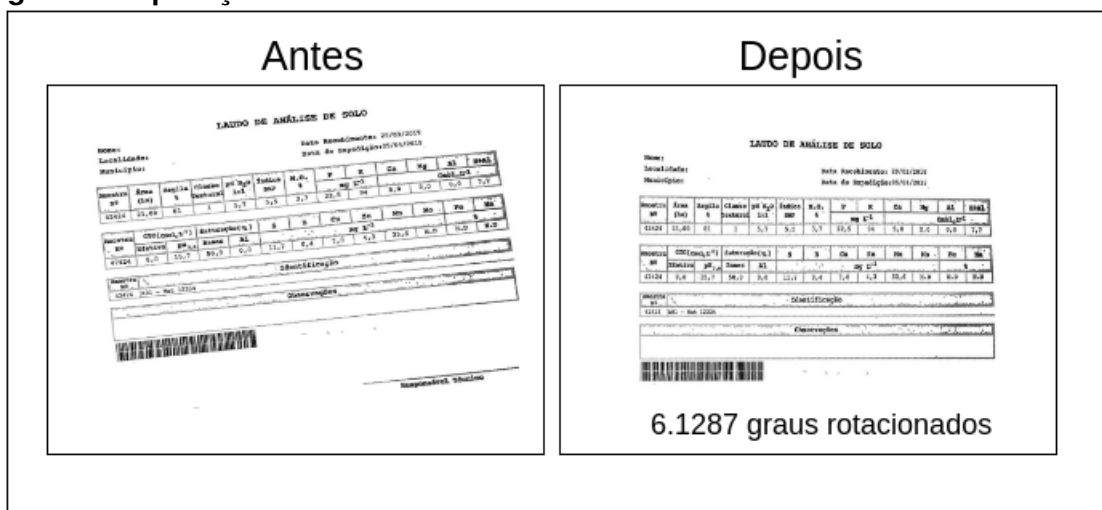
O alinhamento tem por objetivo rotacionar uma imagem em sentido horário ou anti-horário para que o conteúdo da imagem esteja uniforme e linearmente igual, podendo se basear em padrões comuns da imagem ou em uma imagem modelo, cópia da anterior, porém já alinhada.

Este processo surge visto a possibilidade do usuário retirar uma fotografia desalinhada da análise de solo em seu dispositivo móvel, interferindo diretamente na qualidade do processo de extração dos dados desejados por algoritmos de detecção de texto ou leitura OCR.

Para realizar o alinhamento da imagem da análise de solo se utilizou outro script da família ImageMagick, este agora específico para este fim, chamado de textdeskew, tendo por principal objetivo captar padrões lineares que se encontram na imagem, por exemplo uma linha de texto, verificando o grau de desalinhamento e rotacionando a imagem para o grau 0.

O script não precisa de nenhum parâmetro obrigatório, mas é possível configurar a cor do plano de fundo aplicado ao pedaço da imagem que foi rotacionado e também o grau de preenchimento nos arredores da imagem rotacionada. Na figura 43 é possível ver uma imagem com as informações desalinhadas à esquerda, e uma imagem com as informações alinhadas após a execução do script à direita.

Figura 43: Aplicação de alinhamento em análise de solo



3.5.3.2.4 Redimensionamento

Após realizar o processo de alinhamento da imagem da análise de solo se iniciou o processo de redimensionamento para uma resolução específica. Dessa forma, independente do tamanho de entrada recebido a saída sempre terá a mesma resolução e quantidade de pixels por altura e largura, tornando possível uma melhor manipulação da imagem nas próximas etapas da interpretação.

Buscou-se então estudar uma resolução ideal para tal fim, e por meio da documentação da API do Google Vision ⁶ se identificou que o tamanho ideal para um melhor processamento dos resultados e reconhecimento de textos em imagens seria 1024x768.

Para realizar o redimensionamento da imagem utilizou-se o parâmetro `resize` nativo do *software* ImageMagick, o qual recebe por parâmetro a largura da imagem em pixels e com base nessa informação calcula a altura automaticamente, sem que haja perda de proporção e escala da imagem.

3.5.3.3 Pré-processamento

Este tópico tem por objetivo descrever o processo de codificação do algoritmo implementado, responsável por pré-processar as informações contidas na imagem e, diante disso, disponibilizar uma saída que será utilizada para a etapa de interpretação dos dados contidos na imagem.

Com o tratamento da imagem concluído na etapa anterior foi possível obter uma imagem padrão, alinhada, de mesmo tamanho, proporção e com o mínimo de ruídos para assim facilitar esta próxima etapa. Diante disso, se iniciou o desenvolvimento do algoritmo responsável por realizar o pré-processamento na linguagem de programação Python, o qual tem por objetivo separar as tabelas encontradas na imagem a fim de retornar uma imagem somente com os valores dos elementos químicos que precisam ser interpretados pela ferramenta Tesseract OCR ou Google Vision.

Inicialmente, o algoritmo construído recebe por instrução de linha de comando 4 parâmetros, que são responsáveis por determinar algumas configurações da interpretação da imagem, como o caminho de leitura, e de saída, e a forma de extração de dados, bem como, um parâmetro que determina o arquivo de *layout* padrão que será utilizado. No quadro 6 é possível ver a relação dos parâmetros aceitos.

⁶<https://cloud.google.com/vision/docs/supported-files>

Quadro 6: Parâmetros de entrada do algoritmo

Parâmetro	Descrição
1 - Pasta de leitura	Caminho onde se encontram uma ou mais imagens que serão lidas para a interpretação
2 - Arquivo de <i>layout</i>	Caminho onde se encontra o arquivo de <i>layout</i> da análise correspondente
3 - Forma de extração de dados	Define quais dados o algoritmo deverá tentar extrair da imagem, podendo ser, 'simple', 'default', ou 'full'
4 - Pasta de saída	Caminho onde serão salvos os resultados da interpretação

Destes parâmetros descritos no quadro 6 apenas os dois primeiros são obrigatórios, sendo a pasta de leitura das imagens e o arquivo de *layout*. Se o parâmetro de forma de extração for deixado em branco se assumirá o valor 'default' e se o parâmetro 4, referente ao caminho de saída dos resultados, for deixado em branco, ele irá salvar os arquivos na raiz onde se encontra o script em Python.

3.5.3.3.1 Criação e leitura de *layout*

Inicialmente a aplicação do método de interpretação da análise de solo está se detendo apenas a análises do Laboratório de Solos da SETREM, porém nada impede que futuramente o algoritmo possa processar análises de outros laboratórios. Diante disso, foi preciso criar uma estrutura de *layouts*, com as propriedades existentes em cada análise de solo.

Levando como exemplo a análise de solo da SETREM se obteve os principais dados presentes na análise, desde dados do proprietário até os componentes químicos presentes e, por esse meio, se criou um arquivo JSON de *layout* conforme figura 44 que será utilizado como referência no preenchimento dos valores de retorno resultantes da interpretação e criação dos arquivos de resultados em formato JSON e CSV. O algoritmo neste caso tem conhecimento de qual arquivo de *layout* deve usar por meio do parâmetro 2, que recebe o *layout* a ser utilizado durante toda execução do processamento e posterior interpretação da imagem.

3.5.3.3.2 Identificação de tabelas presentes na imagem

Após receber a localização das imagens que devem ser interpretadas pelo parâmetro 1 e o *layout* que deverá ser utilizado pelo parâmetro 2, é aplicado o método `AdaptiveThreshold` disponível na biblioteca OpenCV, tendo por objetivo transformar a imagem recém tratada para pixels com cor 0 e 255, ou seja, preto e

Figura 44: Estrutura JSON retornado pelo algoritmo

```

{
  "Nome" : "",
  "Localidade" : "",
  "Município" : "",
  "DataRecebimento" : "",
  "DataExpedicao" : "",
  "Confidencia" : 0,
  "FilePath" : "",
  "dadosExtraídos" : {
    "amostra" : "",
    "area ha" : "",
    "argila" : "",
    "classe textural" : "",
    "ph_H2O" : "",
    "indice_SMP" : "",
    "m o" : "",
    "p" : "",
    "k" : "",
    "ca" : "",
    "mg" : "",
    "al" : "",
    "h al" : "",
    "ctc_efetiva" : "",
    "ctc_ph" : "",
    "saturacao_bases" : "",
    "saturacao_al" : "",
    "s" : "",
    "b" : "",
    "cu" : "",
    "zn" : "",
    "mn" : "",
    "mo" : "",
    "fe" : "",
    "na" : ""
  }
}

```

branco.

Além da aplicação do método AdaptiveThreshold também se aplicou os métodos de erosão e dilatação que auxiliam na redução de ruído da imagem e no destaque de elementos específicos desejados. Ambos possuem funções semelhantes, se diferenciando de acordo com o tipo de imagem que se deseja processar. A figura 45 detalha seu funcionamento.

Figura 45: Aplicação do método de erosão e dilatação

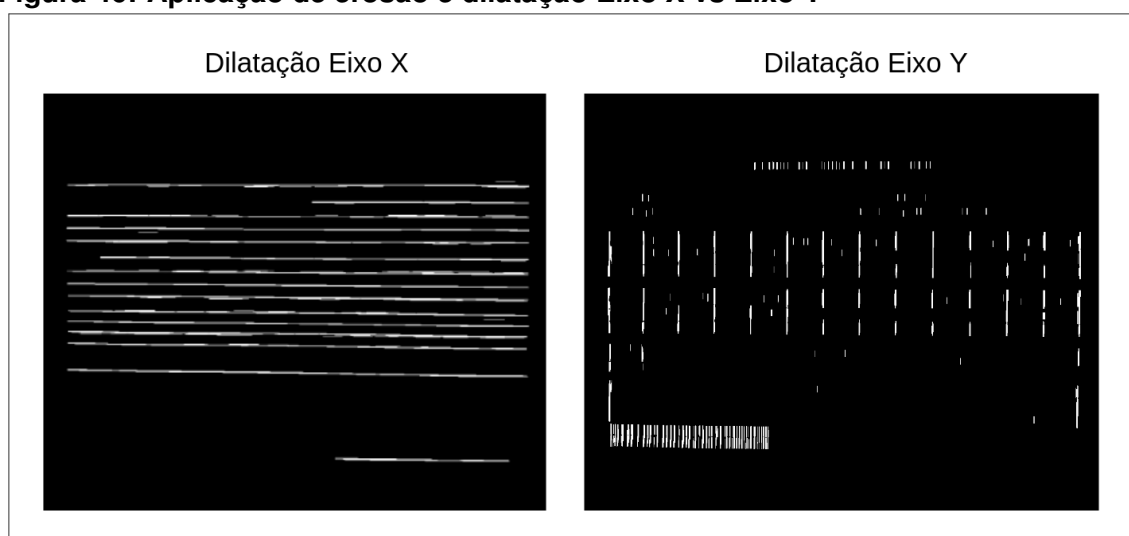


Na figura 45 pode-se perceber que o fundo de toda a imagem é preto e a escrita é branca. O método de erosão sempre intensifica as áreas escuras da imagem, dessa forma a escrita ficou mais fina. Já o método de dilatação destaca as partes claras da imagem, por esse motivo a escrita ficou mais grossa. Levando em consideração o caso acima, se a imagem fosse invertida, ou seja, a escrita fosse preta e o fundo fosse branco o método de erosão iria destacar mais a escrita e o de dilatação iria afinar mais a escrita.

Como no caso da análise de solo a maior parte do fundo é branca se utilizou o método de erosão juntamente com um *kernel* representando um retângulo. Dessa forma, a erosão foi aplicada somente em lugares onde se tinha retângulos na imagem, para destacar esses pontos. Após isso, se utilizou o método de dilatação para suavizar as demais ocorrências da imagem e retirar linhas e textos que não seriam utilizados na extração dos dados.

A dilatação e erosão aplicada na imagem aconteceu nos dois eixos separadamente, tanto o eixo X, para capturar apenas as linhas horizontais presentes na imagem, quanto o eixo Y, para capturar linhas verticais. Na figura 46 é possível ver a aplicação do método no eixo X e eixo Y.

Figura 46: Aplicação de erosão e dilatação Eixo X vs Eixo Y

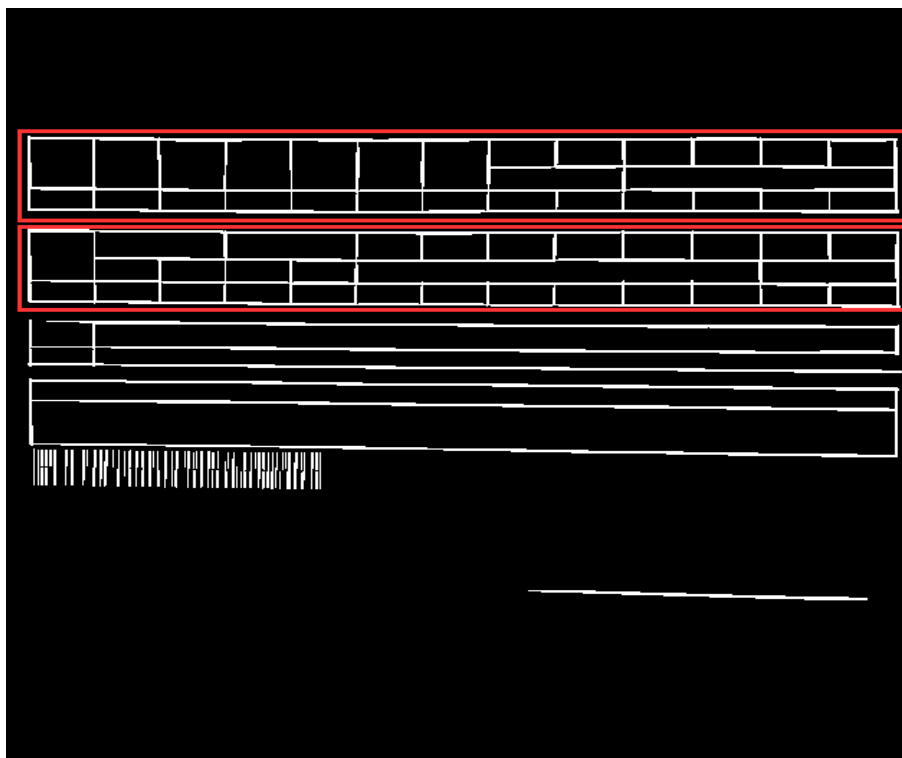


Após a identificação das linhas horizontais e verticais da imagem se realizou o processo de cruzamento dos dois eixos por meio de uma função de soma, onde ($M = h_dilata + v_dilata$), sendo M o resultado do cruzamento. Com esse cálculo se obteve a saída ilustrada na figura 47.

A figura 47 apresenta o resultado, destacado em vermelho, sendo possível perceber a formação de duas tabelas maiores distintas. Essas duas tabelas são utilizadas para extrair as informações dos componentes químicos e seus relativos valores.

Após identificar as duas tabelas principais que são utilizadas para a interpretação de informações, se iniciou o processo de extração das colunas de cada tabela, afim de separar cada componente químico e seu valor para ser interpretado posteriormente por meio do uso da ferramenta Tesseract OCR ou API do Google Vision.

Basicamente o algoritmo identifica as tabelas principais, e a partir dessas identifica as colunas de cada tabela. Esse processo foi realizado por meio da utilização do método *findContours* disponível na biblioteca OpenCV, o qual tem

Figura 47: Cruzamento das linhas verticais e horizontais

por objetivo buscar todos os contornos presentes na imagem e retornar uma lista contendo o eixo X, o eixo Y, a largura e a altura de cada contorno encontrado.

Uma das dificuldades encontradas na utilização do método *findContours* é que o mesmo não ordena os contornos encontrados por ordem de direção, ou seja, da direita para a esquerda, ou de baixo para cima. Diante disso, foi preciso criar uma função para desempenhar essa tarefa, onde é passada a ordem de ordenamento e a lista de contornos desordenada e a mesma retorna à lista de contornos já ordenada. O nome da função criada é *sort_contours()* e pode ser encontrada no arquivo *tools.py* ⁷.

Diante disso, inicialmente busco-se encontrar na imagem somente os contornos responsáveis pelas tabelas que possuem elementos químicos e valores. Dessa forma, é realizado um filtro pelo tamanho de área ocupada pelo contorno específico. A área nesse caso é calculada por meio da execução do método da biblioteca OpenCV *contourArea*, responsável por receber as coordenadas de um contorno e retornar à respectiva área ocupada.

⁷Código fonte da solução: <https://github.com/larcc-group/tcc-si-ad.git>

Caso o tamanho da área do contorno estiver dentro do intervalo solicitado de 65000 a 120000, que foram testados empiricamente, então é realizada a chamada ao método da biblioteca OpenCV *boundingRect*, passando como parâmetro as coordenadas do contorno e retornando novas coordenadas referentes ao quadrante encontrado. O método *boundingRect*, dada as coordenadas por parâmetro, tem por objetivo calcular o retângulo de limite mínimo superior direito.

Após realizar o cálculo do retângulo disponível na imagem se realiza o corte da imagem para se dividir as duas tabelas correspondentes aos elementos químicos, como é possível ver na figura 48.

Figura 48: Separação das tabelas com os elementos químicos

Amostra N°	Área (ha)	Argila %	Classe Textural	pH H ₂ O 1:1	Índice SMP	M.O. %	P	K	Ca	Mg	Al	H+Al
							mg L ⁻¹		Cmol _c L ⁻¹			
24082	0,5	42	2	5,6	5,7	2,4	26,6	280	5,7	2,2	0,0	6,1

Amostra N°	CTC (Cmol _c L ⁻¹)		Saturação (%)		S	B	Cu	Zn	Mn	Mo	Fe	Na
	Efetiva	pH _{7,0}	Bases	Al	mg L ⁻¹						%	
24082	8,6	14,7	58,3	0,0	3,0	0,5	20,7	24,2	74,0	N.D	N.D	N.D

Com a identificação das duas tabelas principais realizada, é executado novamente o mesmo método de busca de contornos, porém agora com base nessas duas tabelas menores e tendo como base da busca apenas as duas últimas linhas horizontais, para, dessa forma, se extrair o conteúdo que possui os valores que deseja-se interpretar, já sem a informação do cabeçalho da tabela.

O cabeçalho foi retirado visto que a informação dos elementos possíveis que podem ser encontrados em uma tabela já foram definidos no arquivo JSON responsável pelo *layout* da análise de solo e, para realizar a interpretação com as ferramentas Tesseract OCR ou Google Vision, quanto mais definido estiverem os valores que desejou-se buscar, melhor será a interpretação das informações pelo algoritmo, visto que não haverá perigo de embaralhar o retorno dos dados.

Por meio do cálculo de diferença entre a altura da última linha com a penúltima linha da tabela, foi possível cortar a imagem como era desejado e, após, se aplicou o método *drawContours* da biblioteca OpenCV para substituir os contornos

que antes possuíam a cor preta, para a cor branca. O resultado deste processo poder ser visto na figura 49.

Figura 49: Extração da linha com os valores dos componentes químicos da tabela

Tabela 1												
24082	0,5	42	2	5,6	5,7	2,4	26,6	280	5,7	2,2	0,0	6,1
Tabela 2												
24082	8,6	14,7	58,3	0,0	3,0	0,5	20,7	24,2	74,0	N.D	N.D	N.D

Como pode-se ver na figura 49, como resultado final, tem-se os valores dos elementos das duas tabelas presentes na análise, sendo possível perceber que os valores numéricos que na foto inicial estavam fracos e apagados agora estão destacados, para assim se ter um melhor grau de acurácia na próxima etapa, de interpretação por meio das ferramentas Tesseract OCR ou Google Vision.

Além disso, foi possível filtrar toda uma imagem da análise, extraíndo somente a informação desejada, que são os valores dos elementos, evitando assim que se retornem valores embaralhados ou fora de ordem e que o algoritmo de interpretação detecte os números de uma forma mais efetiva.

3.5.3.4 Interpretação

Com a imagem da análise de solo tratada e pré-processada, resultando em uma imagem contendo somente os valores que se deseja extrair, conforme figura 49, deu-se início ao desenvolvimento do processo de interpretação, onde foram realizados testes com as ferramentas Tesseract OCR e Google Vision, separadas anteriormente durante seu processo de estudo.

3.5.3.4.1 Implementação com Tesseract OCR

Conforme os testes realizados anteriormente com a ferramenta Tesseract OCR, a mesma havia se mostrado bastante eficiente na interpretação da imagem completa da análise. Diante disso, criou-se um método na linguagem Python para interpretar agora a imagem final da análise de solo, já tratada.

Para utilização do leitor de OCR Tesseract OCR em conjunto com a linguagem Python foi preciso usar o pacote pytesseract, sendo possível assim chamar

o comando de interpretação e ler o retorno do reconhecimento diretamente via código.

O método responsável por realizar a interpretação com o uso do Tesseract OCR é o *readImage* conforme apresentado no código 3 e mais informações sobre o mesmo podem ser encontradas no repositório do projeto no GitHub.

Código 3: Método de interpretação de imagem com o Tesseract OCR

```

1 def readImage(path):
2     imagePath = path
3     config = ('-l por --oem 1 --psm 4')
4     im = cv2.imread(imagePath, cv2.IMREAD_COLOR)
5     text = pytesseract.image_to_string(im, config=config)
6     newValues = []
7     values = text.split(" ")
8     # --- trecho de conversão para número ocultado, onde newValues = values
9     return newValues

```

Conforme o código 3, inicialmente uma imagem é recebida por parâmetro e armazenada na variável 'imagePath', logo em seguida, na linha 4 é criada uma segunda variável, 'config', para armazenar a configuração do Tesseract OCR que será utilizada para realizar o reconhecimento. Neste caso está sendo passada a linguagem português, o motor de processamento de redes neurais LSTM e a segmentação de página padrão, realizando assim uma busca por inteiro em toda a imagem recebida.

Dando seguimento ao código, na linha 6 a imagem é carregada da pasta para a memória por meio da variável 'im' e em seguida é passada para o método próprio do pytesseract *image_to_string(image)* que tem por objetivo receber uma imagem, executar o processamento por meio do Tesseract OCR e retornar um texto com os valores reconhecidos na imagem.

O retorno dos valores é então armazenado em uma variável do tipo array de nome 'values' por meio da função *split* que tem por objetivo quebrar o texto, neste caso, a cada espaço encontrado.

3.5.3.4.2 Implementação com Google Vision

Inicialmente realizou-se uma integração com a API do Google Vision. Por meio da documentação disponível *online* foi possível ver exemplos de comunicação com a API em diversas linguagens de programação, como C#, Java, Go, NodeJS e Python, sendo essa última, a utilizada neste projeto.

Diante da documentação disponível foi possível criar um método na linguagem Python que tem por objetivo receber um arquivo de imagem, enviar o mesmo para a API do Google Vision e retornar o texto interpretado. O método de nome *SendToGoogleVision()* pode ser visto no código 3.

Código 4: Método de interpretação de imagem com a API do Google Vision

```

1 def sendToGoogleVision(urlPath, convertToNumber=False):
2     os.environ["GOOGLE_APPLICATION_CREDENTIALS"] = "key.json"
3     client = vision.ImageAnnotatorClient()
4     with io.open(urlPath, 'rb') as image_file:
5         content = image_file.read()
6     image = vision.types.Image(content=content)
7     resp = client.document_text_detection(image=image)
8     requestConfidence = 0
9     newValues = readBlocks(resp.full_text_annotation)
10    # – trecho oculto -etapa de conversão dos valores para números
11    if len(newValues) > 0:
12        requestConfidence = getConfidence(resp.full_text_annotation)
13
14    return newValues, requestConfidence

```

No código 3 se ocultaram algumas linhas para facilitar a explicação, porém o código completo pode ser encontrado no repositório do GitHub⁸. No fluxo básico de funcionamento do algoritmo é executada uma autenticação com o ambiente do Google, carregamento da imagem e, por fim, o envio da imagem para API e leitura do retorno, conforme melhor descrito a seguir.

Toda comunicação com o Google necessita de uma chave de autenticação para que seja possível verificar a autenticidade e origem de determinada re-

⁸Código fonte da solução : <https://github.com/larcc-group/tcc-si-ad.git>

quisição. Desta forma, na linha 2, carrega-se o arquivo `key.json` que contém as credenciais necessárias para abrir uma comunicação com a API do Google Vision.

Com a comunicação aberta instancia-se o cliente de detecção de imagens, *ImageAnnotatorClient*, disponível no pacote `google.cloud.vision` importado no início do script. Em seguida, nas linhas 4 e 5 é realizada a leitura do arquivo de imagem que deseja-se interpretar por meio da passagem de seu caminho físico na máquina.

Com a imagem física carregada na memória cria-se então uma instância da classe `vision.types.image` e converte-se a imagem física para um array de bytes requerido pela api do Google Vision. Após isso, com a imagem instanciada corretamente é realizada a chamada ao cliente de detecção de texto em documentos, passando a imagem em bytes por parâmetro e recebendo o retorno do processamento na variável `resp`, conforme linha 7.

A API do Google Vision retorna um arquivo no formato JSON contendo diversas informações sobre os resultados da interpretação, dentre eles, os principais são os blocos de texto encontrados na imagem, as coordenadas `x`, `y` e largura e altura para contornar o os blocos, dados encontrados na própria imagem, e também um grau de confiança por bloco, que tem por objetivo determinar a certeza de que aquela informação daquele bloco está correta. Este grau varia de 0 a 1.

Logo após receber a resposta da API é efetuada a chamada ao método *readBlocks* que tem por objetivo percorrer todos os blocos retornados e adicionar o valor do texto de cada bloco em uma variável do tipo array, que conterà cada valor encontrado em um índice separado.

3.5.3.4.3 Observações

Como é possível notar, ambas as implementações tem em comum o retorno de um *array* com os resultados encontrados. Com base nesse retorno o arquivo de *layout* inicialmente passado por parâmetro no início do processo é alimentado com seu respectivo valor encontrado. Caso não se identifique um valor válido, ou seja, não seja possível converter o valor retornado para um número, o algoritmo irá salvar o valor 'N.D' no atributo referente ao elemento químico.

Caso algum valor de elemento químico não seja retornado pelo algoritmo de interpretação, pode acontecer uma situação em que as referências do elemento químico com seu respectivo valor não fiquem de acordo, podendo ocorrer o embaralhamento dos valores. Este tipo de situação não aconteceu durante os testes, sempre havia um valor retornado, mesmo que não fosse numérico.

Na figura 50 é possível acompanhar o resultado de alguns testes que foram realizados diante de duas análises de solo como forma de testar a implementação desenvolvida.

Figura 50: Testes da implementação com Tesseract OCR e Google Vision

Tabela 1												
24082	0,5	42	2	5,6	5,7	2,4	26,6	280	5,7	2,2	0,0	6,1
24082	8,6	14,7	58,3	0,0	3,0	0,5	20,7	24,2	74,0	N.D	N.D	N.D
Tesseract OCR												
[24082', '0,5', '42', '2', '5,6', '5,7', '24', '26,6', '280', '5,7', '2,2', '0,0', '6,2']												
[24082', '8,6', '14,7', '58,3', '0,0', '3,0', '?', '20,7', '24,2', '74,0', 'N.D', 'N.D', 'N.D', 'N.D']												
Vision API												
[24082', '0,5', '42', '2', '5,6', '5,7', '24', '26,6', '280', '5,7', '2,2', '0,0', '6,1']												
[24082', '8,6', '14,7', '58,3', '0,0', '3,0', '0,5', '20,7', '24,2', '74,0', 'N.D', 'N.D', 'N.D', 'N.D']												

Tabela 2												
21831	50,0	90	1	5,1	5,9	2,3	11,8	72	5,0	1,4	0,4	4,8
21831	6,9	11,4	57,3	5,7	N.D	N.D	N.D	N.D	N.D	N.D	N.D	N.D
Tesseract OCR												
[21831', '50,0', '90', '1', '5,1', '5,9', '2,3', '11,8', '72', '5,0', '1,4', '0,4', '4,8']												
[21831', '6,9', '11,4', '57,3', '5,7', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D']												
Vision API												
[21831', '50,0', '90', '1', '5,1', '5,9', '2,3', '11,8', '72', '5,0', '1,4', '0,4', '4,8']												
[21831', '6,9', '11,4', '57,3', '5,7', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D', 'N.D']												

Por meio da figura 50 percebe-se que a API do Google Vision reconheceu todos os elementos das duas análises que foram submetidas, enquanto que o Tesseract OCR não reconheceu o valor da segunda linha da sétima coluna, onde foi colocado um caractere de ponto de interrogação (?) no lugar para fins de identificação. Mesmo assim, não é possível obter um parâmetro ou média diante de duas análises de solo, sendo preciso criar um *dataset* com mais imagens para testar as duas implementações efetivamente.

3.5.4 Api

Como forma de auxiliar o processo de submissão de imagens para o algoritmo, foi criada uma api em Python, utilizando um pacote python denominado de

Flask. Dessa forma, quando a requisição é recebida a mesma é submetida para o algoritmo de interpretação de análise de solo para processamento, retornando a estrutura JSON vista na figura 44.

3.6 EXPERIMENTOS

Este tópico tem por objetivo apresentar as etapas executadas para validar o processo de interpretação dos dados das análises de solo. Para a validação foram utilizados cinquenta análises de solo impressos que foram disponibilizados pelo Laboratório de Solos da SETREM e com base nos mesmos foi criado um *dataset* com variadas amostras destas análises para posterior interpretação pelo algoritmo de interpretação desenvolvido.

3.6.1 *Dataset*

Para validação do processo de reconhecimento de análises de solo foi preciso criar uma base de dados contendo imagens de análise de solos dos mais diversos tipos, dentre eles, imagens com flash e sem flash, com sombra e sem sombra, inclinadas, tremidas e nítidas.

Diante disso, se utilizou a ferramenta Google Forms, onde foi possível configurar um formulário com as perguntas desejadas e assim compartilhar o mesmo com um grupo de pessoas, disponibilizando para cada uma 4 análises de solo impressos, para que as mesmas capturassem fotos conforme o que foi solicitado no formulário.

O formulário é composto de duas etapas principais, na primeira o usuário preenche o seu nome e algumas informações sobre o seu dispositivo móvel, são essas: marca, modelo, memória e resolução da câmera. Na segunda etapa é preciso tirar as fotos das análises anteriormente entregues de acordo com o tipo solicitado na pergunta, conforme é possível ver na figura 51

Nesta etapa do formulário exibida na figura 51 existem 8 tipos de imagens que devem ser submetidas, cada uma com suas características específicas, dentre elas as imagens sombreadas, com flash e imagens livres.

Por meio da pesquisa realizada foi possível recolher 256 amostras de aná-

Figura 51: Formulário utilizado para criação do *dataset*

The image shows a vertical list of eight form fields, each with a label and a sub-label. The labels are: 'Imagem com sombra *', 'Imagem sem sombra *', 'Imagem com flash *', 'Imagem sem flash *', 'Imagens inclinadas *', 'Imagens tremidas *', 'Imagem correta (com informações nítidas e legíveis) *', and 'Imagem livre do laudo *'. Each label is followed by a blue link that says 'ADICIONAR ARQUIVO'. At the bottom of the form, there are two buttons: a grey one labeled 'VOLTAR' and a purple one labeled 'ENVIAR'. To the right of these buttons is a green progress bar and the text 'Página 3 de 3'.

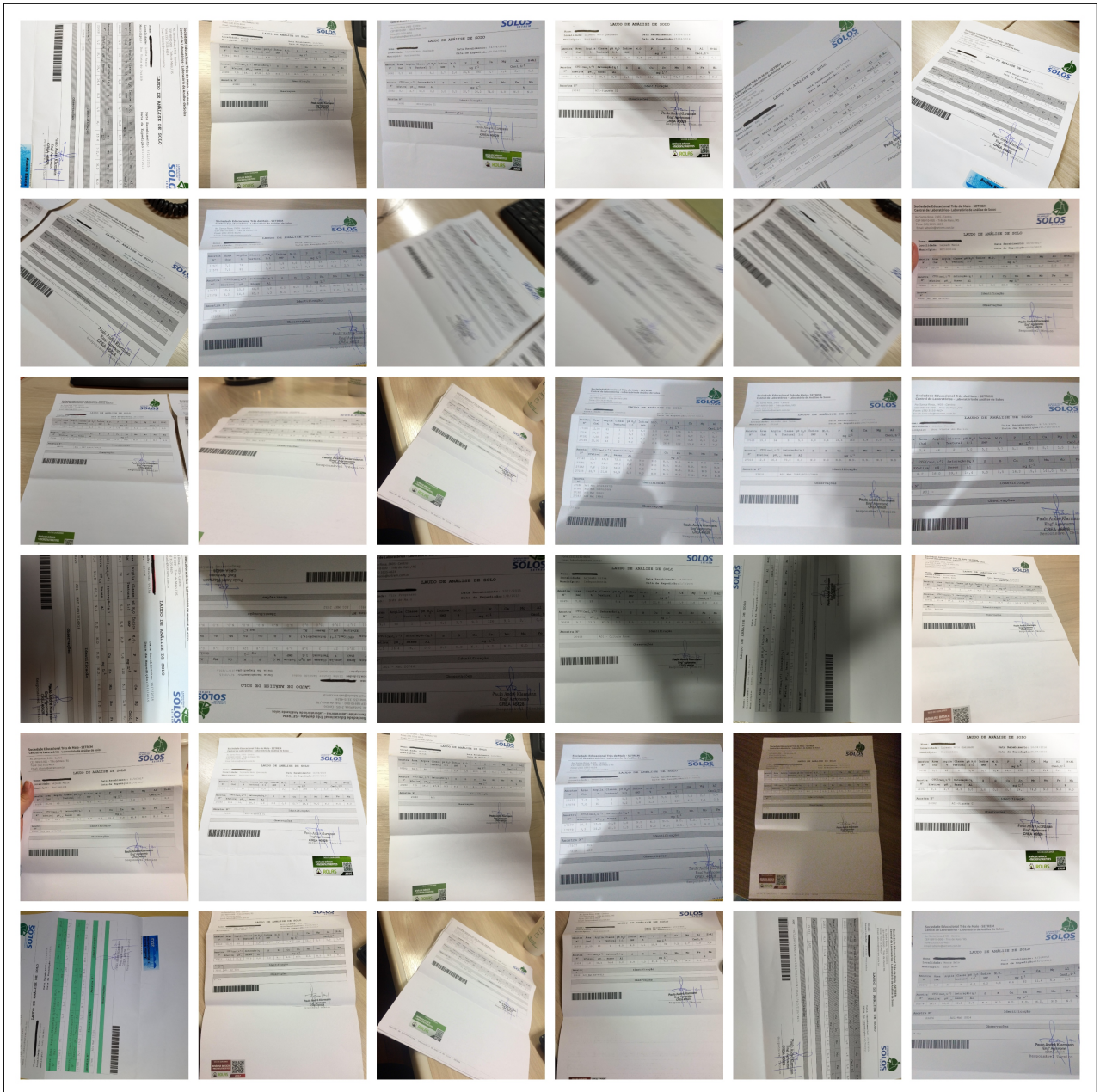
lise de solo de 8 dispositivos diferentes, dentre eles, dispositivos da Xiaomi, Motorola e Samsung. Estas amostras foram então utilizadas para testar o algoritmo desenvolvido e analisar o grau de acurácia do algoritmo para os diferentes cenários e possibilidades de fotos. A autorização para uso de dados e imagens pode ser vista em anexo no apêndice C. Na figura 52 é possível ver um conjunto de imagens coletadas, com diferentes características.

Como é possível perceber na figura 52, foi possível coletar dados bastante variados diante das fotos, possuindo situações de fotos sombreadas, muito iluminadas, inclinadas e também tremidas.

No momento em que todos os usuários preencheram o formulário com as imagens solicitadas, foi realizado um processo manual de verificação em cada tipo de imagem. Neste processo, foi verificado se a imagem que havia sido solicitada realmente possuía as características de seu tipo. Com essa análise foi possível encontrar imagens sombreadas que também estavam inclinadas, ou imagens inclinadas e tremidas, nítidas porém com sombra, entre outras situações.

Diante dessa divergência encontrada em alguns grupos de imagens, se realizou uma reestruturação do *dataset*, se dividindo as imagens em apenas 5

Figura 52: Amostras de análisis de solo



tipos, conforme o quadro 7.

Quadro 7: Tipos de imagens do *dataset*

Tipo	Quantidade de imagens
Imagem alinhada e clara	47
Imagem inclinada	95
Imagem inclinada com sombra	32
Imagem sombreada	41
Imagem tremida	41

De acordo com o quadro 7 é possível perceber que a maior quantidade de imagens pertence ao grupo de "inclinadas", sendo que muitas dessas imagens, antes da revisão realizada estavam junto ao grupo de imagens alinhadas, o que poderia acarretar em divergências durante a execução do algoritmo.

Para cada tipo de imagem existe um critério de classificação: as imagens alinhadas e claras são definidas com ótima qualidade de visualização das informações contidas na análise, assim como também são imagens que estão pouco desalinhadas e com iluminação ótima.

Já, as imagens inclinadas, são as imagens que possuem uma iluminação ótima, porém são imagens que foram fotografadas em um ângulo diferente, ocorrendo assim, uma taxa de proporção da imagem distorcida, entre as extremidades das tabelas que serão lidas. Essa mesma situação ocorre para imagens inclinadas com sombra, porém a diferença está nas iluminação das imagens, que além de estarem inclinadas, possuem variação de iluminação, sendo que na mesma imagem existe parte da iluminação clara e outra escura.

As imagens alinhadas e claras são definidas com ótima qualidade de visualização que possuem um grau de desalinhamento mínimo ou nenhum, porém sua iluminação é totalmente ou em grande parte escurecida. Por fim as imagens tremidas são imagens onde as informações contidas na análise não são visíveis em sua totalidade.

3.6.2 Execução do algoritmo de interpretação de análises de solo

Após o recolhimentos das 256 amostras de análise de solos, se iniciou o processo de tentativa de interpretação dessas imagens. Para isso se criou e utilizou uma máquina virtual da plataforma em nuvem do Google, com as especificações conforme quadro 8.

Quadro 8: Especificações da máquina virtual utilizada

Sistema Operacional	Ubuntu 18.04
Processador	Intel Xeon com 6 núcleos de 2Ghz
Disco	SSD de 30Gb
Memória	15 GB

Após a configuração da máquina virtual, conforme quadro 8, se realizou o download do projeto do algoritmo de interpretação de análises na máquina e em sequência o *download* das dependências necessárias para o algoritmo ser executado no servidor. Com as dependências instaladas e o algoritmo funcionando corretamente, se realizou o *upload* do *dataset* para o servidor.

Com o *dataset* criado se iniciou a execução do algoritmo para cada tipo de imagem separadamente, onde a cada execução, caso a mesma tenha sido executada com sucesso para aquela imagem de análise de solo, os valores extraídos foram sendo adicionados em um arquivo JSON com os resultados de todas as interpretações realizadas com sucesso.

A estrutura desse arquivo JSON é igual a já descrita anteriormente e pode ser vista por meio da figura 44. Além do arquivo JSON também é gerado um arquivo CSV com as informações básicas da análise de solo, semelhante ao que foi gerado para importação no sistema CADUB, apresentado no tópico 3.4.2.1. Além disso, a cada processamento a imagem da análise é movida para uma pasta de destino, podendo essa ser uma pasta de erro ou de sucesso, dependendo do resultado do processamento.

3.6.2.1 *Validação da interpretação*

Para validação dos resultados gerados após a execução do algoritmo usaram-se os dados de análises de solo que já haviam sido disponibilizados pelo Laboratório de Solos da SETREM por meio de um arquivo CSV, conforme melhor comentado no tópico 3.4.1. Por meio disso, foi criado um programa auditor, o qual realiza a comparação de elemento por elemento de cada análise de solo e quantifica os acertos de acordo com as informações retornadas pelos dados das análises fornecidas pelo Laboratório de Solos da SETREM.

O desenvolvimento do programa auditor foi realizado na linguagem NodeJS visto que a mesma possui mais suporte para trabalhar com estruturas JSON, já que seu motor é baseado em javascript.

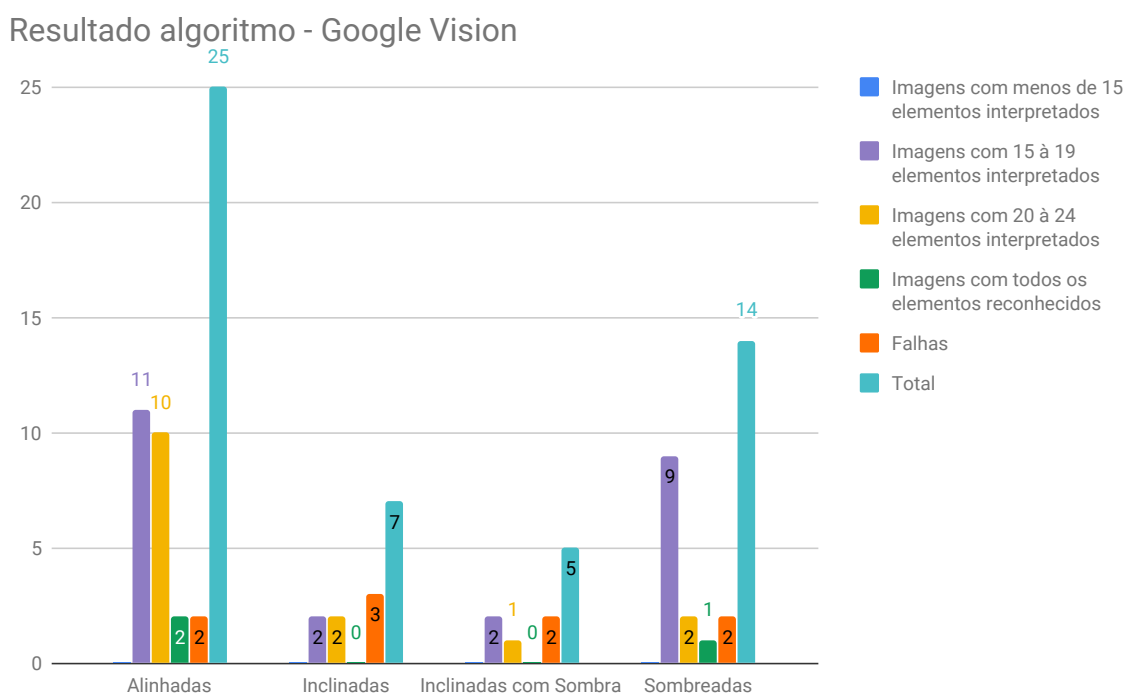
A execução do algoritmo foi realizada em duas etapas, a primeira submetendo a imagem para a API do Google Vision e a segunda utilizando o Tesseract OCR. Em ambos os casos, foram utilizadas todas as 256 imagens separadas pelos grupos já vistos para, dessa forma, realizar uma comparação entre as duas ferramentas, onde nos dois casos o pré-processamento da imagem é o mesmo. Para os dois testes sua execução é em torno de 1 hora para interpretação dessas imagens, importante ressaltar, que essa demora se deve ao pré-processamento, onde várias tratativas são feitas.

Quando a execução do algoritmo é concluída, as imagens as quais o algoritmo não consegue realizar os tratamentos descritos com maiores detalhes na seção 3.5.3.3.2, são ignoradas, retornando somente com sucesso as imagens em que é possível a interpretação de análises de solo. Os resultados obtidos que retornaram com sucesso foram submetidos ao auditor das informações da imagens que o algoritmos de interpretação da imagem obteve êxito para retornar a informação.

Com o intuito de verificar a eficiência do algoritmo, os resultados obtidos foram separados por faixas de validação com os dados fornecidos pelo Laboratório de Solos da SETREM, que com base no número da amostra, que é o primeiro elemento da tabela interpretada. As demais informações são buscadas e comparadas entre os dados do laboratório e os dados retornados pelo algoritmo.

São divididas pela quantidade de elementos que são interpretados para cada imagem da análise e os casos de falhas são imagens que podem ter vários elementos interpretados, porém, o elemento do número da amostra que é tomada como base para buscar as informações do laboratório não foi reconhecido. No gráfico da figura 53 é possível verificar os resultados retornados utilizando a API do Google Vision.

Figura 53: Resultado do algoritmo utilizando Google Vision



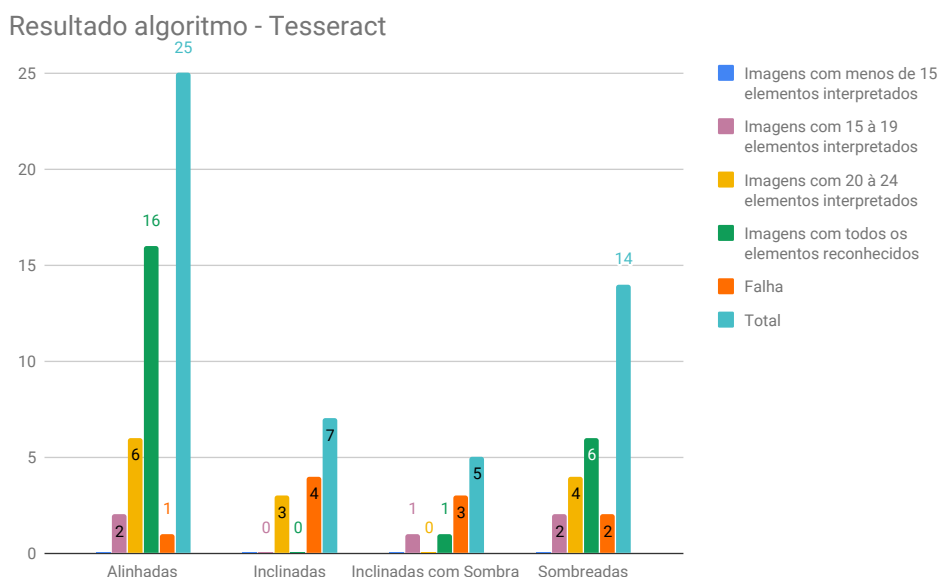
É possível verificar que os resultados obtidos utilizando o Google Vision, no grupo de imagens alinhadas, as maiores quantidades de imagens estão na faixa de 15 à 19 elementos interpretados e a faixa de 20 a 24 elementos interpretados e as imagens que alcançaram a interpretação de todos os elementos, são somente 2.

Quanto aos demais grupos, é possível verificar que somente no grupo de imagens sombreadas o algoritmo conseguiu realizar a leitura completa de toda a análise, os demais não conseguiram realizar a leitura completa.

Realizada novamente a execução do algoritmo, porém substituindo o a API do Google Vision pelo Tesseract OCR, é possível verificar na figura 54 os resulta-

dos obtidos no mesmo cenário de execução com o Google Vision.

Figura 54: Resultado do algoritmo utilizando Tesseract OCR

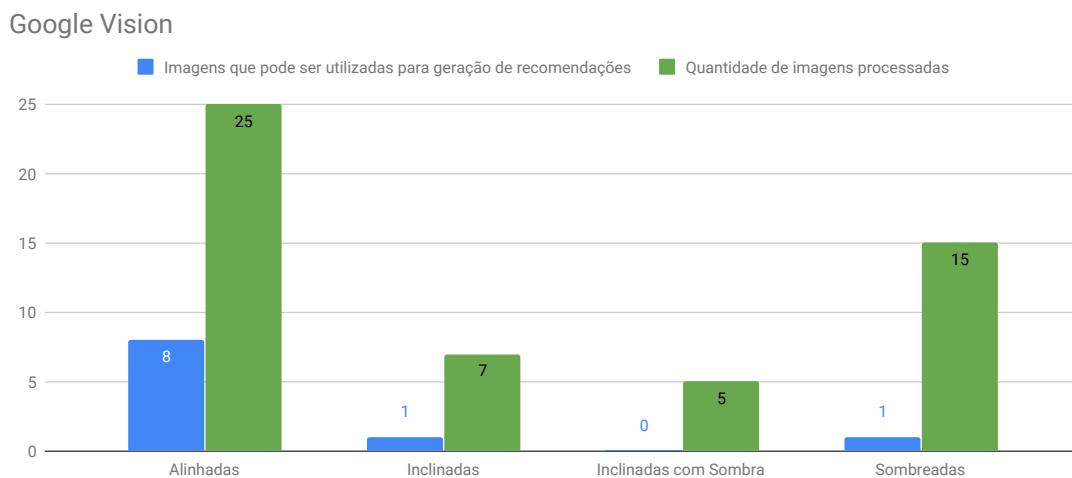


No gráfico apresentado na figura 54 é possível verificar que no grupo das imagens alinhadas, o algoritmo obteve sucesso na interpretação completa de 16 análises das 25 e ainda nesse mesmo grupo 6 análises foi possível realizar a interpretação de 20 a 24 elementos.

Verificando os demais grupos, pode-se destacar que no grupo de imagens sombreadas a quantidade de análises interpretadas com sucesso foi 6, e nesse mesmo grupo 4 análises foram interpretadas de 20 a 24 elementos.

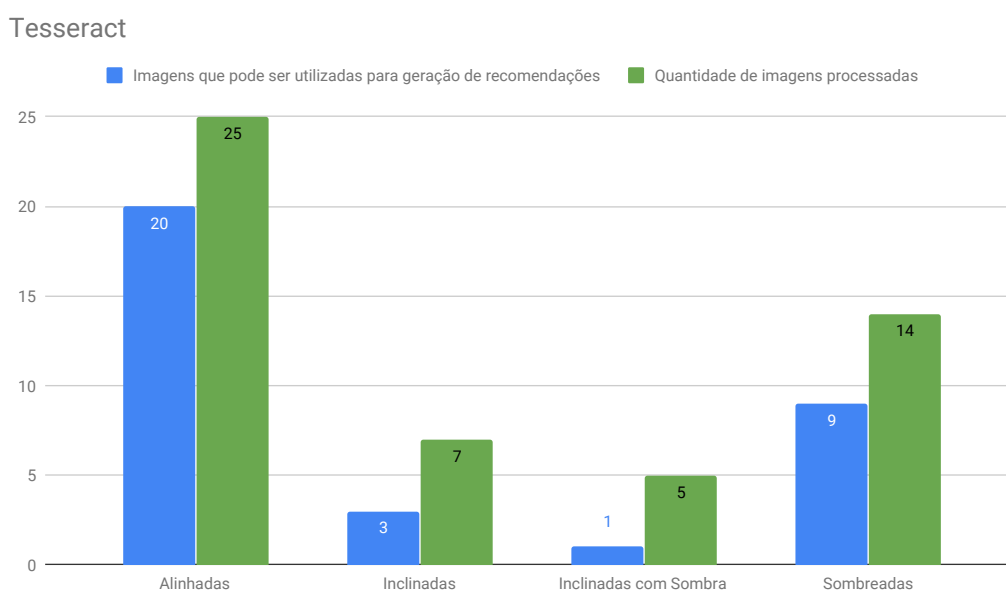
Em comparação utilizando o Google Vision e o Tesseract OCR, é possível notar que a quantidade de análises que todos os elementos foram interpretados, é possível visualizar que os resultados obtidos utilizando o Tesseract OCR a quantidade é bem maior em comparação com o Google Vision.

Para recomendação de análises de solo é importante ressaltar que a quantidade de elementos necessários diminui para 12, ou seja, mesmo que a execução do algoritmo tenha realizada a interpretação correta de todos os elementos, estes podem ser o suficiente para realização das recomendações. Devido a isso, o gráfico da figura 55 foi criado com o intuito de demonstrar a quantidade de análises interpretadas que possuem os requisitos mínimos para realizar a recomendação da análise de solo.

Figura 55: Resultado para recomendação utilizando Google Vision

Nesta figura 55, a quantidade de análises interpretados para realização das recomendações são de somente 6, utilizando a API da Google Vision para o grupo de imagens alinhadas. Nos grupos de imagens inclinadas e sombreadas, a quantidade de análises interpretadas é de somente 1 para cada grupo.

Realizando a comparação com os resultados obtidos utilizando o Tesseract OCR, demonstrado na figura 56, é possível visualizar grandes diferenças de quantidade de análises interpretadas para realização da recomendação da análise de solo.

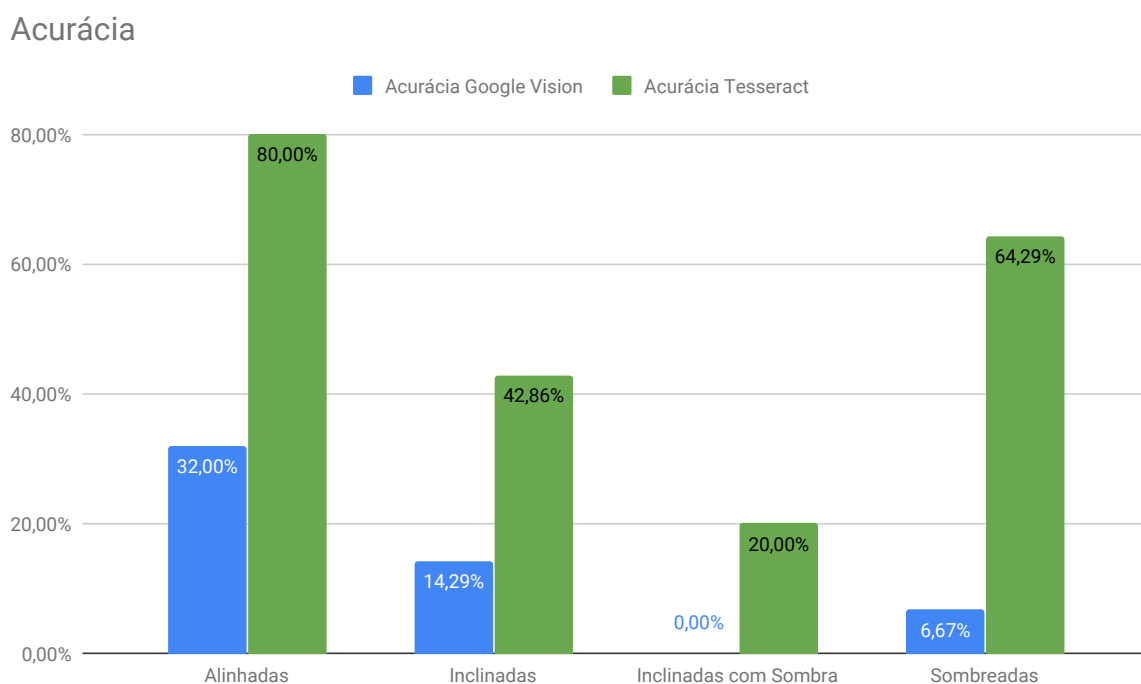
Figura 56: Resultado para recomendação utilizando Tesseract OCR

Na figura 56, a quantidade de análises interpretadas com mínimo de elementos para geração de recomendações, no grupo de imagens alinhadas chega a 20.

Para os demais grupos é possível notar que a quantidade de análises interpretadas é maior, como por exemplo, o grupo de imagens sombreadas, que interpretar 9 imagens corretamente para realizar a recomendação. Então, com o cenário de comparação com entre o Google Vision e o Tesseract OCR, é possível ver uma eficiência muito melhor nos resultados retornados pelo Tesseract OCR.

Com os resultados obtidos, foi possível a realização do cálculo da acurácia obtida para cada grupo de imagens, divididas pelas duas ferramentas utilizadas para a interpretação de análise. O gráfico da figura 57 apresenta a acurácia obtida, realizando a divisão da quantidade de análises interpretadas para recomendação pelo total de análises submetidas ao algoritmo.

Figura 57: Acurácia da interpretação das análises de solo



Conforme visto acima na figura 57, o grupo de imagens com a maior acurácia atingida foi as imagens alinhadas, atingindo 80% de acurácia utilizando o Tesseract OCR, porém para as imagens alinhadas utilizando o Google Vision, foi

de 32%.

Logo em seguida, o grupo de imagens sombreadas conseguiu uma acurácia de 64% com o Tesseract OCR e somente 6,57% utilizando o Google Vision. Para o grupo de imagens inclinadas é possível verificar que a acurácia atingida pelo Tesseract OCR foi de 42,86% , que contrapartida utilizando o Google Vision foi de 14,29%.

Por fim, no grupo de imagens inclinadas e com sombra, o Tesseract OCR consegue atingir uma acurácia de 20%, enquanto nesse grupo o Google Vision não consegue realizar a interpretação de nenhuma análise, portanto a acurácia para esse grupo é zero.

Analisando o gráfico da figura 57, é importante destacar a superioridade da acurácia em todos os grupos de imagens utilizando o Tesseract OCR, alcançando até 80% no grupo de imagens alinhadas. Pode-se dessa forma validar que a melhor solução para a interpretação de análises de solo é o Tesseract OCR.

3.6.3 Sugestões para implementação da recomendação

Com o aplicativo criado para realizar a interpretação das análises de solo e retornando um JSON e CSV com os dados, a criação de um processo de recomendação baseado nos manuais de adubação e calagem é o próximo passo, conforme a proposta 3.5.1, detalhado nas etapas de pré-processamento e processamento. Então, somente após essa etapa, toda a arquitetura da solução estará concluída.

Durante o desenvolvimento do trabalho percebeu-se que não seria possível implementar a proposta de recomendação de adubação com a utilização de *Machine Learning* visto que as tabelas de referência que o engenheiro agrônomo utilizada para calcular uma recomendação ainda são estáticas e disponibilizadas apenas em um livro físico. Além disso, ainda não se tem uma quantidade suficiente em variedade de dados de análises de solo, para se iniciar o treinamento de um modelo de aprendizado de máquina.

Diante disso, uma possibilidade é a criação de um sistema estatístico, baseado em condições IF e ELSE, porém que se atualize de uma forma mais dinâmica, no momento em que um novo manual de adubação e calagem seja dispo-

nibilizado. Além disso, o novo sistema deve permitir integração por meio de uma API Rest, para que análises de solo sejam rapidamente submetidas, resultando em recomendações de adubação. Este tipo de integração ainda não existe hoje e seria um grande diferencial em um *software* dessa área.

CONCLUSÃO

No presente trabalho apresentou-se um estudo voltado a área de ciência de solos, por meio da interpretação de análise de solo e posterior recomendação de adubação, juntamente com o estudo de tecnologias de *Machine Learning* e processamento de imagens que podem ser aplicadas a esse cenário. Com base no estudo obteve-se a criação de uma arquitetura de solução para auxiliar tanto o produtor rural quanto o engenheiro agrônomo.

Inicialmente foi proposta uma solução voltada ao contexto de recomendação de adubação, por meio da extração de informações de análises de solo do Laboratório de Solos da SETREM. Dessa forma, buscou-se com o estudo dos manuais de calagem e adubação e do *software* CADUB, desenvolver um modelo treinado com a capacidade de simular recomendações de adubação ideais para cada nova análise de solo que fosse submetida.

Para treinar um modelo de Machine Learning é necessário entradas (*inputs*) e saídas (*outputs*), diante disso, o *software* CADUB foi utilizado para gerar as saídas necessárias por meio da recomendação de adubação ideal. Este processo foi automatizado por meio do uso da ferramenta Autolt. Dessa forma, as informações obtidas no laboratório de solos serviriam como entradas para serem submetidas ao *software* CADUB, para que, com esses resultados o dataset de treinamento de *Machine Learning* pudesse ser criado juntamente com os dados fornecidos pelo Laboratório de Solos da SETREM.

A etapa de criação do dataset de treinamento não ocorreu como esperado,

visto que foram encontradas algumas dificuldades para realizar esse processo. Dentre elas, estava a dependência direta dos manuais de adubação e calagem, permitindo que o problema fosse resolvido diretamente com a criação de um sistema estatístico, com base nos manuais de adubação e calagem, como o próprio CADUB, não havendo uma real necessidade de implementar um algoritmo de inteligência artificial.

Além disso a quantidade de dados de análise de solo obtida se fez muito inferior após o tratamento dos dados disponibilizados pelo laboratório, onde os 300 mil registros iniciais do arquivo CSV se transformaram em apenas 30 mil registros, diminuindo assim a variedade de dados e confirmando que a utilização de Machine Learning nesse quesito não seria realmente eficaz.

Diante das dificuldades enfrentadas com a primeira ideia proposta, se propôs uma arquitetura de solução completa, de interpretação e recomendação de análises de solo, mantendo o foco na melhoria do processo do dia a dia do produtor. Dessa forma, no presente trabalho foi concluída a etapa de leitura dos dados de análises de solo, que visa agilizar o processo manual de preenchimento de informações dos 25 elementos contidos nas análises de solo, convertendo os mesmos para o meio digital, através da utilização de tecnologias de processamento de imagens e reconhecimento de caracteres.

O problema foi solucionado, com a utilização de um algoritmo que utiliza inteligência artificial para leitura de informações contidas em imagens, onde é realizada a submissão da imagem para o algoritmo, recebendo como retorno os valores dos elementos químicos identificados. Além disso, com esta integração fica possível a criação de um banco de dados global, possibilitando armazenar dados de análises de solo de várias regiões em que o aplicativo poderia ser utilizado.

O trabalho em questão tem grande importância no ramo da agricultura, visto que busca entender as necessidades e problemas existentes na área de interpretação e recomendação de análise de solo. Por meio dos estudos realizados, foi possível perceber a dependência do manual de calagem e adubação para realização dos cálculos necessários para se gerar uma recomendação de adubação ideal, como os nutrientes NPK esperados.

Quanto ao desenvolvimento do trabalho, foi possível com base no estudo, entender o real uso de uma análise de solo e todas as etapas executadas pelo laboratório de solos para chegar aos resultados dos elementos químicos. Com base neste conhecimento, foi possível buscar formas de se aplicar a tecnologia da informação nos processos de interpretação de análise de solo e recomendação de adubação, visando auxiliar o produtor rural e engenheiro agrônomo na tomada de decisão.

Dentre os pontos positivos gerados pelo trabalho, está a criação de uma arquitetura de solução que visa facilitar a forma de acesso à informação quando uma análise de solo é finalizado, por meio de possibilidades de integração entre o laboratório de solos e uma API capaz de integrar diretamente com um aplicativo móvel. Esta arquitetura é útil pois permitiu expandir as possibilidades de aplicação da tecnologia na área da agricultura, especificamente no ramo de ciência e análise de solos, gerando, assim, ideias de trabalhos futuros que podem ser desenvolvidos nesta área.

Outra área bastante discutida e aplicada neste trabalho, é a de processamento de imagens, onde foram abordados métodos de tratamento, como aplicação de filtros, alinhamento, redução de ruído e controle de níveis visando facilitar o reconhecimento dos de análise de solo. Juntamente a isso, são abordados métodos de pré-processamento de imagem por meio da utilização da biblioteca OpenCV, que foi importante para extrair as informações realmente importantes da imagem, como as tabelas principais e seus respectivos valores, antes de ser submetida ao algoritmo de interpretação.

Além disso, também se destaca o estudo de tecnologias de Machine Learning, exclusivamente na área de processamento de imagens, que foi responsável por uma das aplicações práticas realizadas. Neste quesito, foi possível testar as ferramentas Tesseract OCR e Google Vision visando comparar qual das duas ferramentas tem maior potencial para reconhecer informações de análises de solo tratadas.

Como objetivo do trabalho a proposta de desenvolver um aplicativo móvel capaz de realizar a interpretação de análise de solos e a recomendação de aduba-

ção das análises disponibilizadas pelos laboratórios de solo foi apresentada. Este objetivo foi parcialmente atingido visto que não foi possível implementar a recomendação de adubação, já que por meio dos estudos realizados viu-se que este processo ainda depende exclusivamente dos manuais de calagem e adubação e a atualização dos dados desses manuais não sofre alteração com base em apenas alguns fatores específicos, como clima, tipo de solo ou umidade.

Com o estudo realizado percebeu-se que os valores existentes como referência nos manuais de calagem e adubação, usados para se gerar uma recomendação, tem sua origem em diversos estudos que são realizados tanto por acadêmicos da área da agricultura, quanto por uma rede de laboratórios participantes do grupo ROLAS, que realizam uma série de experimentos assistidos, visando calcular a produtividade das plantações e, com base nos resultados recolhidos, os manuais são atualizados.

Outro ponto que reforça o atingimento parcial do objetivo do trabalho diz respeito a falta de variedade de análise de solos de outros laboratórios, visto que neste caso, o método de interpretação de valores foi aplicado e testado apenas em análises de solo do laboratório da SETREM, fazendo que, possivelmente seja preciso realizar alguns ajustes no algoritmo de interpretação para que funcione corretamente em análises de solos de outros laboratórios.

Por meio dos estudos realizados obteve-se resultados significativos, principalmente quanto a questão de funcionamento do processo de análise de solos, processos envolvidos e limitações que hoje ainda impedem a criação de um sistema de recomendação auto sustentável por meio do uso de *Machine Learning*.

Dentre os objetivos específicos estava o aprofundamento nos conhecimentos na área da agricultura destinada à área de ciência e análise de solos. Este objetivo foi alcançado, visto que foram realizadas pesquisas bibliográficas sobre o assunto e também reuniões com o engenheiro agrônomo Paulo André Klarmann para buscar entender o funcionamento do processo de análise de solos, que vai desde a coleta da amostra de solo pelo produtor, o processamento da mesma pelo laboratório e a apresentação dos resultados e geração das recomendações de adubação pelo engenheiro agrônomo.

Outro objetivo alcançado diz respeito ao aprofundamento na área da tecnologia da informação referente a *Machine Learning*. É possível afirmar que este objetivo também foi atingido visto que o assunto foi abordado por meio de pesquisa bibliográfica no capítulo 2. Além disso, foram realizados estudos em 5 ferramentas que utilizam a tecnologia de inteligência artificial e processamento de imagens, para criação do algoritmo de reconhecimento e interpretação de análise de solos desenvolvidos. São elas: o OpenCV, ImageMagick, TensorFlow, Tesseract OCR e Google Vision.

O terceiro objetivo que diz respeito a avaliação do uso de inteligência artificial também foi alcançado. O mesmo se atingiu, visto que após a coleta das informações do Laboratório de Solos da SETREM e tratamento dos dados com o uso do *software* CADUB, se executou uma avaliação do uso da tecnologia de inteligência artificial diante dos resultados até então obtidos.

Foi nesta etapa que percebeu-se que o uso dessa tecnologia para aprender a realizar recomendações de adubação de forma automática era falha, pois, conforme já destacado, existem fatores externos, como o manual de adubação e calagem e a quantidade insuficiente de variedade de dados, que acabam impossibilitando o uso de *Machine Learning* na área de recomendação. Foi com base nesta avaliação, que foi desenvolvida a segunda proposta, ao qual tem por objetivo interpretar as análises de solo, por meio do processamento de imagens e reconhecimento de caracteres.

Outro objetivo presente no trabalho é referente a implementação de um modelo/algoritmo de inteligência artificial para auxiliar na leitura dos dados de análise de solo e realizar a recomendação da adubação. Neste objetivo, foi atingido apenas o processo que diz respeito a leitura dos dados da análise, por meio da utilização das ferramentas ImageMagick, OpenCV e Tesseract OCR, sendo criado um algoritmo para interpretação das imagens de análises de solo submetidas.

O processo de recomendação da adubação exposto no objetivo não foi atingido, pois, conforme o estudo realizado durante o trabalho, principalmente na análise do uso da inteligência artificial, percebeu-se que as tabelas que são usadas como referências para se recomendar a quantidade de nutrientes e adubação

ideal do solo são disponibilizadas apenas nos manuais de calagem e adubação, sendo essas, informações estáticas e que são atualizadas em intervalos de 8 a 10 anos por uma equipe de cientistas juntamente com uma parceria entre laboratórios que fazem parte do grupo ROLAS. Além disso, não existia muita variedade de informação nos dados extraídos do Laboratório de Solos da SETREM, pois, após o tratamento dos dados, os 300 mil registros iniciais se transformaram em 30 mil registros.

Quanto aos dois últimos objetivos, que dizem respeito ao desenvolvimento de um aplicativo mobile para auxiliar na interpretação de análises de solo e do processo de comparação dos resultados interpretados pelo algoritmo com os dados do Laboratório de Solos da SETREM, ambos foram atingidos. Foi desenvolvido um aplicativo mobile com a utilização do framework React-Native que se comunica diretamente com o algoritmo de interpretação desenvolvido. Além disso, a comparação dos resultados com o laboratório de solos, se fez necessária para avaliar a acurácia do algoritmo, para assim, ser possível validar a primeira hipótese, a qual atingiu-se a acurácia de 80%.

Este trabalho deixa como contribuições o estudo para utilização de *Machine Learning* para geração de recomendações das análises de solo. Baseado em estudos de trabalhos relacionados, notou-se que a existência de trabalhos que buscam solucionar esse problema através de inteligência artificial para geração de recomendação é bastante escassa, com o presente trabalho, o estudo de viabilidade para utilização de *Machine Learning* para resolver esse problema até o desenvolvimento desse estudo, não é possível, pois as recomendações atualmente são baseadas nos manuais de adubação e calagem e com variáveis físicas do solo.

Quanto a análise das hipóteses levantadas no início do projeto, nenhuma delas foi corroborada. A hipótese "A solução proposta possui um grau de acurácia de 95% na interpretação de análise de solo" não foi consolidada pois, conforme os testes de acurácia realizados e apresentados em gráficos, o nível máximo de acurácia atingido foi de 80% em imagens de análise de solos alinhadas.

Acredita-se que para obter 95% de acerto nas interpretações, seja necessário treinar um modelo próprio de reconhecimento de padrões em imagens, por

meio da coleta de uma grande variedade de imagens de análises de solo. Porém, de qualquer forma, a acurácia de 80% atingida por meio da utilização da ferramenta Tesseract se mostra bastante válida, em se tratar de um modelo genérico, não treinado especificamente para este fim.

Em relação a segunda hipótese, "A utilização de Machine Learning para realizar a recomendação das análises de solo elimina a necessidade de utilização diretamente dos manuais de calagem e adubação.", a mesma não foi corroborada já que, conforme a pesquisa bibliográfica realizada no capítulo 2 e o conhecimento obtido com o engenheiro agrônomo Paulo André Klarmann, percebeu-se que hoje ainda existe uma total dependência de uso de manuais de calagem e adubação para se gerar as recomendações.

Além disso, não existem dados referente aos elementos químicos e seus valores disponibilizados *online* para uma se realizar uma possível interpretação com base neles. Esta hipótese também é inválida porque a atualização dos manuais de calagem e adubação dependem de vários fatores que são levantados por uma rede de laboratórios que fazem parte do grupo ROLAS e também por pesquisas científicas na área da agronomia e ciência de solos, no qual são realizados experimentos de plantações observadas por períodos, para se medir o nível de produtividade atingido em determinado solo.

Com base no estudo realizado foi possível notar que o processo de análise de solos e recomendação de adubação ainda é muito mecânico e manual, possuindo total dependência dos manuais de calagem e adubação, que segundo estudo, somente são disponibilizados em formato impresso, não digital. Além disso, não foi encontrada nenhuma base de dados pública para consultar valores de elementos químicos encontrados em determinadas regiões do estado do Rio Grande do Sul, muito menos uma média de valores por região, dificultando assim a proposta inicial de se criar um algoritmo que fosse capaz de gerar recomendações de adubação com base em um estudo anterior, principalmente pela insuficiência de dados.

Dentre as limitações encontradas no decorrer do trabalho a mais visível é a dependência do engenheiro agrônomo com o manual de calagem e adubação

de seu respectivo estado, onde para gerar uma recomendação de adubação ao produtor rural ainda são utilizados métodos manuais de consulta e verificação de nutrientes em tabelas.

Diante disso, no decorrer do estudo concluiu-se que não seria possível criar uma inteligência capaz de simular recomendações de solo, visto que não foi encontrada nenhuma base de dados pública disponibilizada por algum órgão de controle de solos, como a EMBRAPA, sendo necessário recorrer a livros físicos, como o manual de calagem e adubação para se extrair resultados de recomendação de adubação.

Outro ponto identificado tem referência a quantidade de amostras de dados de análises de solo que haviam sido coletados no laboratório da SETREM. Inicialmente o arquivo de análises disponibilizado tinha 300 mil linhas, porém a desestruturação do banco de dados em um arquivo CSV fez com que cada elemento de uma determinada análise estivesse separado em uma nova linha. Com isso o agrupamento das informações retornou um resultado final de aproximadamente 30 mil análises, quantidade bastante pequena para se iniciar um treinamento com *Machine Learning*

Este trabalho permitiu abordar uma área ainda pouco explorada pela tecnologia e pouco estudada por artigos científicos, a área de interpretação e recomendação de adubação por meio de análises de solo. Diante disso, novos trabalhos podem ser desenvolvidos para se dar seguimento a pesquisa.

Um desses estudos poderia abordar a interpretação de análise de solo por meio de um modelo treinado especificamente para isso, por meio da coleta de análises de solo de diferentes laboratórios da região e com a utilização de um framework de *Machine Learning* como o TensorFlow, que foi um dos frameworks que debateu-se neste trabalho.

Outra sugestão para desenvolvimento de um novo trabalho é a criação de um sistema que possibilite a geração de recomendações com base nos manuais de adubação e calagem, que ofereça uma forma de integração via API, para incremento do aplicativo desenvolvido no presente trabalho, alcançando assim uma

solução completa de interpretação e recomendação de análise de solo para produtores e engenheiros agrônomos.

O presente trabalho buscou realizar a interpretação de análises de solo visando facilitar a coleta de dados. Então, é possível desenvolver um novo trabalho visando a mineração de dados, que possibilite o levantamento de informações estatísticas referentes a qualidade e histórico de produtividade do solo em regiões específicas do estado.

Para dar continuidade ao presente trabalho desenvolvido, é possível realizar a expansão de interpretações de análises de solo visando abranger a compatibilidade com diferentes laboratórios da região, aumentando, assim, a área de coleta de dados para as demais localidades.

REFERÊNCIAS

- ABRAMOVAY, R. **Laços Financeiros na Luta Contra a Pobreza**. São Paulo: Anablume, 2004.
- BARROS, H. d. **Os grandes sistemas de organização da economia agrícola**. Lisboa: Livraria Sá da Costa, 1975.
- BELLMAN, R. **Dynamic Programming**. Princeton, NJ, USA: Princeton University Press, 2010.
- BRASIL, B. d. **Custeio agropecuario**. 2018. Disponível em: [https://www.bb.com.br/pbb/pagina-inicial/agronegocios/agronegocio---produtos-e-servicos/credito/credito-para-custeio/custeio-agropecuario#/>](https://www.bb.com.br/pbb/pagina-inicial/agronegocios/agronegocio---produtos-e-servicos/credito/credito-para-custeio/custeio-agropecuario#/). Acesso em: 6 novembro 2018.
- CATTELL, R. Scalable SQL and NoSQL Data Stores. **SIGMOD Rec.**, New York, NY, USA, v.39, n.4, p.12–27, Maio 2011.
- CIVIL, C. **Seguro da Agricultura Familiar - SEAF**. Secretaria Especial de Agricultura Familiar e do Desenvolvimento Agrário. Disponível em: <http://www.mda.gov.br/sitemda/secretaria/saf-seaf/cuidados-para-garantir-cobertura>>. Acesso em: 8 novembro 2018.
- CLOUD, G. **Produtos de aprendizado de máquina e IA**. 2019.
- COELHO, A. M. Agricultura de Precisão - manejo. **Embrapa**, Sete lagoas - MG, v.1, p.1–60, Dezembro 2004.

- COLAH. **Understanding LSTM Networks**. 2019.
- COPELAND, R. **MongoDB Applied Design Patterns**. Sebastopol - CA: O'Reilly Media, 2013.
- CRESWELL, J. W. **Projeto de Pesquisa - Métodos qualitativo, quantitativo e misto**. Porto Alegre: Bookman, 2007.
- DE SORDI, J. O. **Gestão por processos: uma abordagem da moderna administração**. São Paulo: Saraiva, 2009.
- DIEHL, R. **Agricultura Geral**. Lisboa: Clássica Editora, 1989.
- DINKINS, C. P.; JONES, C. Developing Fertilizer Recommendations for Agriculture. **Montana State University Extension**, Bozeman - EUA, Junho 2007.
- EIBE FRANK, M. A. H.; WITTEN, I. H. **Data Mining: practical machine learning tools and techniques**. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/Witten_et_al_2016_appendix.pdf>. Acesso em: 12 novembro 2018.
- EISENMAN, B. **Learning React Native - Building native mobile apps with javascript**. Estados Unidos: O'REILLY, 2016.
- EMBRAPA. **Manual de Métodos de Análise de Solo**. Rio de Janeiro: Embrapa, 2011. v.2.
- FERREIRA, D. L. **Análise das potencialidades e importância da soja na Baixada Cuiabana**. Cuiabá - MT: [s.n.], 2018.
- FOWLER, A. The role of AI-based technology in support of the knowledge management value activity cycle. **The Journal of Strategic Information Systems**, Netherlands, v.9, p.107–128, Setembro 2000.
- FOWLER, M.; SADALAGE, P. Polyglot Persistence. **ThoughtWorks**, Chicago - EUA, Fevereiro 2012.
- Furriel, G. P.; Calixto, W. P.; Alves, A. J.; Campos, P. H. M.; Domingues, E. G. Intelligent system for measuring soil compaction on croplands. In: IEEE 15TH

- INTERNATIONAL CONFERENCE ON ENVIRONMENT AND ELECTRICAL ENGINEERING (EEEIC), 2015. **Anais...** [S.l.: s.n.], 2015. p.1357–1361.
- FURTINI NETO, A. E.; VALE, F. R.; RESENDE, A. V.; GUILHERME, L. G. R.; GUEDES, G. A. A. **Fertilidade do solo**. Lavras: UFLA/FAEPE, 2001.
- GARG, K. **Mobile Computing**. New Delhi: Pearson Education India, 2010.
- GHOLAP, J.; INGOLE, A.; GOHIL, J.; GARGADE, S.; ATTAR, V. Soil Data Analysis Using Classification Techniques and Soil Attribute Prediction. **N**, Nova Iorque - EUA, v.abs/1206.1557, 2012.
- GIL, A. C. **Métodos e técnicas de pesquisa social**. São Paulo: Atlas, 1994.
- GOULART, C. P. A. **Tratamento de sementes de soja com fungicidas**. São Paulo: EMBRAPA-CPAO, 1997.
- GRAVES, A. **Supervised Sequence Labelling with Recurrent Neural Networks**. New York, Londres: Springer Berlin Heidelberg, 2012.
- GREFF, K.; SRIVASTAVA, R. K.; KOUTNÍK, J.; STEUNEBRINK, B. R.; SCHMIDHUBER, J. LSTM - A search space odyssey. , EUA, p.12, Outubro 2017.
- GUBIANI, P. I.; SILVA, L. S. d.; REINERT, D. J. 2015, Disponível em: <http://coral.ufsm.br/fisicadosolo/index.php/pt_br/publicacoes/14-fisica-do-solo/software/59-cadub>. Acesso em: 7 novembro 2018.
- GUBIANI, P. I.; SILVA, L. S. d.; REINERT, D. J.; REICHERT, J. M. CADUB GHF - um programa computacional para cálculos da quantidade de fertilizantes e corretivos da acidez do solo para culturas produtoras de grãos, hortaliças e forrageiras. **Cia Rural**, Santa Maria, v.37, p.1161 – 1165, 08 2007.
- JUN, S. Communication is vital to produce natural looking metal ceramic crowns. , US, 1997.
- JÚNIOR, L.; CHAGAS ARAÚJO DE, F. das. Algoritmo Q-learning como estratégia de exploração e/ou exploração para metaheurísticas GRASP e algoritmo genético. In: 2015 , Natal - RN. **Anais...** [S.l.: s.n.], 2009. p.1–140.

- Katamreddy, S.; Riordan, D.; Doody, P. Artificial calf weaning strategies and the role of machine learning: a review. In: IRISH SIGNALS AND SYSTEMS CONFERENCE (ISSC), 2017. **Anais...** [S.l.: s.n.], 2017. p.1–6.
- KLEIN, P. A. T.; WEBER, L. **Aplicações de lógica fuzzy em software e hardware**. Canoas: ULBRA, 2003.
- KUMAR, P.; REDDY, S. R. N. Design and development of M3SS - A Soil Sensor Node for precision agriculture. In: INTERNATIONAL CONFERENCE ON RECENT ADVANCES AND INNOVATIONS IN ENGINEERING, Jaipur, India. **Anais...** IEEE, 2016. p.1–10.
- Kumar, R.; Singh, M. P.; Kumar, P.; Singh, J. P. Crop Selection Method to maximize crop yield rate using machine learning technique. In: INTERNATIONAL CONFERENCE ON SMART TECHNOLOGIES AND MANAGEMENT FOR COMPUTING, COMMUNICATION, CONTROLS, ENERGY AND MATERIALS (ICSTM), 2015. **Anais...** [S.l.: s.n.], 2015. p.138–145.
- LABORSOLO. **Tipos de Análises Químicas do Solo**. 2019.
- LAGENDIJK, R. L.; BIEMOND, J. Basic Methods for Image Restoration and Identification. In: 2015 , Netherlands. **Anais...** [S.l.: s.n.], 2009.
- LI, H.; LENG, W.; ZHOU, Y.; FUDI, C.; XIU, Z.-L.; YANG, D. Evaluation Models for Soil Nutrient Based on Support Vector Machine and Artificial Neural Networks. **TheScientificWorldJournal**, [S.l.], v.2014, p.478569, 12 2014.
- LOVATO, A. **Metodologia da Pesquisa**. Três de Maio: Setrem, 2013.
- LUGER, G. F. **Inteligência Artificial**: estruturas e estratégias para a solução de problemas complexos. Porto Alegre: Bookman, 2004.
- MANZANO, J. A. N. G. **Introdução à linguagem Python**. São Paulo: Novatec Editora, 2018.
- MAR-ELIA, D. 1998. Disponível em: <<https://www.itprotoday.com/devops-and-software-development/scriptit>>. Acesso em: 6 novembro 2018.

- MARCONI, M. D. A.; LAKATOS, M. E. **Fundamentos de metodologia científica**. São Paulo: Atlas, 2007.
- MEDEIROS, J. S. de. **Bancos de Dados Geográficos e Redes Neurais Artificiais - Tecnologias de Apoio à Gestão do Território**. 1999. Tese (Doutorado em Ciência da Computação) — Universidade Federal de São Paulo, São Paulo, Brasil.
- Mendel, J. M. Fuzzy logic systems for engineering: a tutorial. **Proceedings of the IEEE**, [S.l.], v.83, n.3, p.345–377, March 1995.
- METAXIOTIS, K.; ERGAZAKIS, K.; SAMOUILIDIS, E.; PSARRAS, J. Decision support through knowledge management: the role of the artificial intelligence. **Inf. Manag. Comput. Security**, [S.l.], v.11, p.216–221, 12 2003.
- MITCHELL, T. M. **Machine Learning**. [S.l.]: McGraw-Hill Science, 1997. v.1.
- MOKARRAM, M.; HAMZEH, S.; AMINZADEH, F.; ZAREI, A. Using machine learning for land suitability classification. **West African Journal of Applied Ecology**, [S.l.], v.23, p.63–73, Setembro 2015.
- MOLIN, J. P.; RABELLO, L. M. **Estudos sobre a mensuração da condutividade elétrica do solo**. Piracicaba - SP: Engenharia Agrícola, 2011. v.31.
- MONGODB. **What is MongoDB?** 2018. Disponível em: <<https://www.mongodb.com/what-is-mongodb>>. Acesso em: 12 novembro 2018.
- NEVES, J. L. Pesquisas Qualitativas. **Caderno de pesquisa de administração**, [S.l.], 1996.
- OCA, A. M. de; ARREOLA, L.; FLORES, A.; SANCHEZ, J.; FLORES, G. R. Low-cost multispectral imaging system for crop monitoring. **2018 International Conference on Unmanned Aircraft Systems (ICUAS)**, [S.l.], p.443–451, 2018.
- OCR, G. T. **Tesseract OCR**. 2019.
- OSÓRIO, F. **Redes Neurais Artificiais**: do aprendizado natural ao aprendizado artificial. Canoas: I fórum de inteligência artificial, 1999.

- PEREIRA, C. R. **Node.js: aplicações web real-time com node.js**. São Paulo: Casa do Código, 2013.
- Perera, S. Large scale data processing in real world: from analytics to predictions. In: INTERNATIONAL CONFERENCE ON ADVANCES IN ICT FOR EMERGING REGIONS (ICTER), 2014. **Anais...** [S.l.: s.n.], 2014. p.8–8.
- QUEIROZ, T. R.; ZUIN, L. F. S. **Agronegócios - Gestão, Inovação E Sustentabilidade**. São Paulo: Saraiva, 2015.
- RAIJ, B. V. **Fertilidade do solo e adubação**. Piracicaba - SP: Agronômica Ceres, 1991.
- Rashid, S. F.; Akmal, A.; Adnan, M.; Aslam, A. A.; Dengel, A. Table Recognition in Heterogeneous Documents Using Machine Learning. In: IAPR INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION (ICDAR), 2017. **Anais...** [S.l.: s.n.], 2017. v.01, p.777–782.
- READS, S. **Inteligência Artificial: compreender em que consiste a i.a. e o que implica a aprendizagem das máquinas**. Canadá: KOBO EDITIONS, 2017.
- RICH, E.; KNIGHT, K. **Inteligencia artificial**. Espanha: McGraw-Hill Interamericana, 2016. v.2.
- SBCS. **Manual de adubação e de calagem**. Porto Alegre: [s.n.], 2004.
- SBCS. **Manual de adubação e de calagem**. Porto Alegre: [s.n.], 2016.
- SEBRAE. **Tecnologia no meio rural**. Disponível em: <<http://datasebrae.com.br/tecnologia-no-meio-rural/#celularinternet>>. Acesso em: 11 novembro 2018.
- SEN, S.; WEISS, G. Multiagent Systems. In: 2017 14TH , Cambridge, MA, USA. **Anais...** MIT Press, 1999. p.259–298.
- SIDDIQUI, S. A.; MALIK, M. I.; AGNE, S.; DENGEL, A.; AHMED, S. DeCNT: deep deformable cnn for table detection. **IEEE Access**, [S.l.], v.6, p.74151–74161, 2018.

- SIMÕES, M. G. **Controle e modelagem fuzzy**. São Paulo: E. Blücher, 2007. v.2.
- SIRSAT, M.; FERNÁNDEZ-DELGADO, M. Machine learning for the management of agricultural soil data. In: WORKSHOP MACHINE LEARNING EN GALICIA 2016. **Anais...** [S.l.: s.n.], 2016. p.1.
- Smith, D.; Peng, W. Machine learning approaches for soil classification in a multi-agent deficit irrigation control system. In: IEEE INTERNATIONAL CONFERENCE ON INDUSTRIAL TECHNOLOGY, 2009. **Anais...** [S.l.: s.n.], 2009. p.1–6.
- Smith, R. An Overview of the Tesseract OCR Engine. In: NINTH INTERNATIONAL CONFERENCE ON DOCUMENT ANALYSIS AND RECOGNITION (ICDAR 2007). **Anais...** [S.l.: s.n.], 2007. v.2, p.629–633.
- SOGLIO, F. D.; KUBO, R. R. **Agricultura e Sustentabilidade**. Porto Alegre: UFRGS, 2009.
- STRAUCH, C. NoSQL Databases. **A. Further Reading, Listening and Watching**, Stuttgart - Alemanha, 2011.
- SU, J.; ZHANG, H. A Fast Decision Tree Learning Algorithm. In: NATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE - VOLUME 1, 21. **Proceedings...** AAAI Press, 2006. p.500–505. (AAAI'06).
- SUTTON, R. S.; BARTO, A. G. **Reinforcement Learning**: an introduction. Cambridge: A Bradford book, 1998.
- TALUKDER, A. K.; YAVAGAL, R. P. **Mobile Computing**: technology applications and service creation. New York: McGraw Hill, 2007. v.1.
- TEIXEIRA, E. B. A análise de dados na pesquisa científica. **Desenvolvimento em questão**, Itajaí, n.2, p.177–201, Dezembro 2003.
- Thirumalai, C.; Harsha, K. S.; Deepak, M. L.; Krishna, K. C. Heuristic prediction of rainfall using machine learning techniques. In: INTERNATIONAL CONFERENCE ON TRENDS IN ELECTRONICS AND INFORMATICS (ICEI), 2017. **Anais...** [S.l.: s.n.], 2017. p.1114–1117.

- TODOROFF, P.; DE ROBILLARD, F.; LAURENT, J.-B. Interconnection of a crop growth model with remote sensing data to estimate the total available water capacity of soils. In: 2017 . **Anais...** [S.l.: s.n.], 2010. p.1641–1644.
- TOMÉ JÚNIOR, J. B. **Manual para interpretação de análise de solo**. Guaíba - RS: Livraria e Editora Agropecuária, 1997.
- VERAS, M. **Arquitetura de Nuvem (AWS): amazon web services**. São Paulo: Brasport, 2013. v.1.
- VINAY, V. **What is opencv?** 2017. Disponível em: <<https://software.intel.com/en-us/articles/what-is-opencv>>. Acesso em: 29 março 2019.
- WATANABE, A. M.; BESSA, L. P. D.; CORRADINI, R. A.; MARTINS, T. G. d. M.; SERRAT, B. M.; LIMA, M. R. d. Por que fazer análise de solo? **Projeto de Extensão Universitária Solo Planta**, Curitiba, p.143, 2002.
- WELBANK, M. **A Review of Knowledge Acquisition Techniques for Expert Systems**. [S.l.]: British Telecom Research Laboratories, 1983.
- WILSON, J. R. **Node.js 8 the Right Way Practical, Server-Side JavaScript That Scales**. United State: The Pragmatic programmers, 2013.
- YUHANNA, N.; LEGANZA, G.; AUSTIN, c. The Forrester Wave™: big data nosql, q3 2016 scalability, multimodal, flexibility, and lower cost drive enterprise adoption of nosql. **Forrester**, [S.l.], p.18, 2016.
- ZYLBERSZTAJN, D.; NEVES, M. F.; CALEMAN, S. M. D. Q. **Gestão de Sistemas de Agronegócios**. São Paulo: Atlas, 2015.