

# LATINOWARE 20 ANOS



20º Congresso Latino-americano de  
Software Livre e Tecnologias Abertas

18 a 20 de outubro de 2023

Foz do Iguaçu | Paraná | Brasil

Realização:



**PTI**

Parque Tecnológico  
Itaipu



## Applying FOSS Support Vector Machine and Rough Sets on COVID-19 Cases Triage

*Autores:*

**Vinícius Hansen, Kalyl Henings  
& Gilmário Barbosa dos Santos**

- A incerteza no diagnóstico médico ganhou destaque durante a pandemia de COVID-19.
- No processo de análise de casos há sempre incerteza diagnóstica e dificuldade de triagem.
- A ideia central é treinar um modelo de *machine learning* (ML) que incorpore o conceito de incerteza diagnóstica.
- No presente trabalho, fez-se a prova empírica dessa técnica, realizando experimentos com *support vector machine* (SVM) e mineração de dados incertos com *rough sets* (RS).

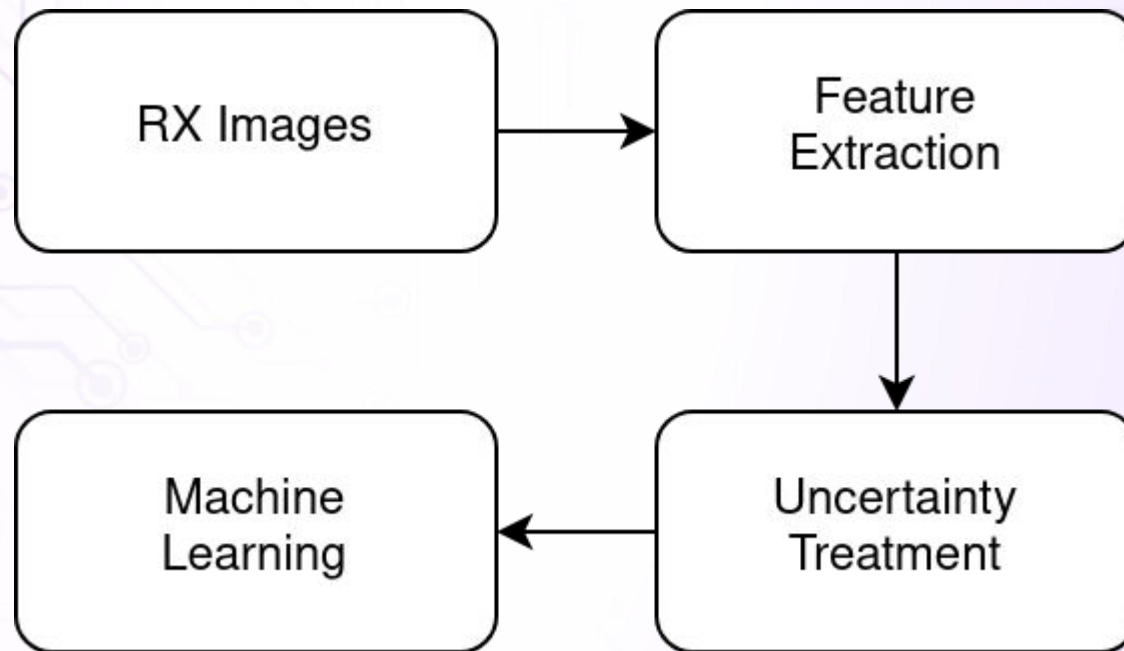


Figura 1: Visão geral conceitual

- Foi conduzida uma pesquisa com o objetivo de explorar bases públicas, gratuitas e bem documentadas.

Casos	Covid Radiography	BIMCV COVID+
COVID-19	3,616	34,829
Saudáveis	10,200	0

Tabela 1: Bases de dados originais



- Foram selecionados os casos BIMCV mais parecidos com a base CRD para completar 10,200 casos de COVID-19, a Figura 2 ilustra esse processo.

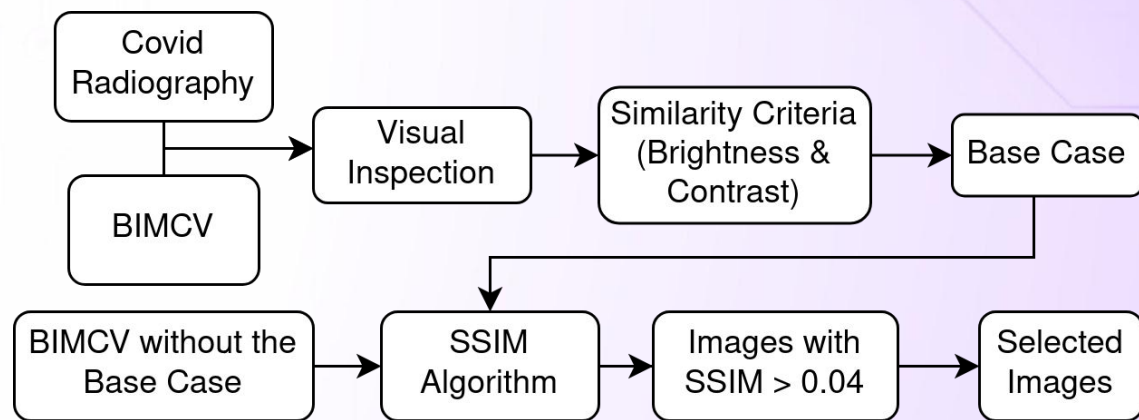


Figura 2: Algoritmo de escolha de casos

- Os dados foram processados utilizando LBP (raio 5) seguido do método *Decision Table* (Tabela 2).
- A base foi analisada pela biblioteca de *Rough Sets* e foi encontrado um *Boundary Region*.
- Os casos excedentes foram cortados e obtivemos a Base Composta (Tabela 3).

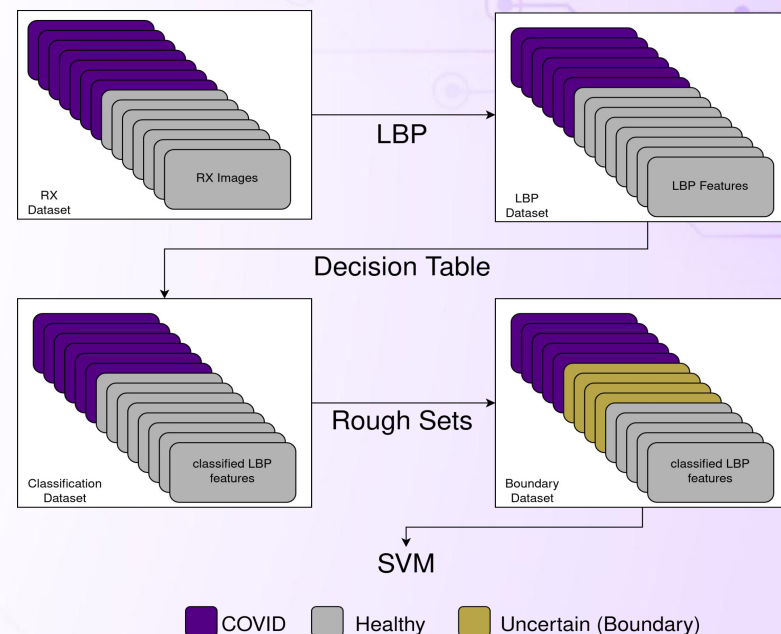


Figura 3: Pipeline geral dos dados

$\delta = (x - \mu) / \sigma \times 100$	
Alvo	Novo Valor
$\delta < -100$	$f(\delta) = 0$
$-100 < \delta \leq -80$	$f(\delta) = 1$
$-80 < \delta \leq -60$	$f(\delta) = 2$
...	...
$60 < \delta \leq 80$	$f(\delta) = 16$
$80 < \delta \leq 100$	$f(\delta) = 17$
$\delta > 100$	$f(\delta) = 18$

Tabela 2: Descrição do método Decision Table

Casos	Base composta
Saudáveis	5301
Incertos	5301
COVID-19	5301

Tabela 3: Base de dados composta



- Foi utilizada a biblioteca de ML open source *scikit-learn*.
- Inicialmente, o modelo SVM foi treinado com os parâmetros padrão.
- Posteriormente, foi feita uma otimização desses hiperparâmetros utilizando o GridSearchCV

Parâmetro	Valor
Kernel	RBF
C	100
Gamma	1

Tabela 3: Hiperparâmetros usados no modelo

# Matriz de Confusão

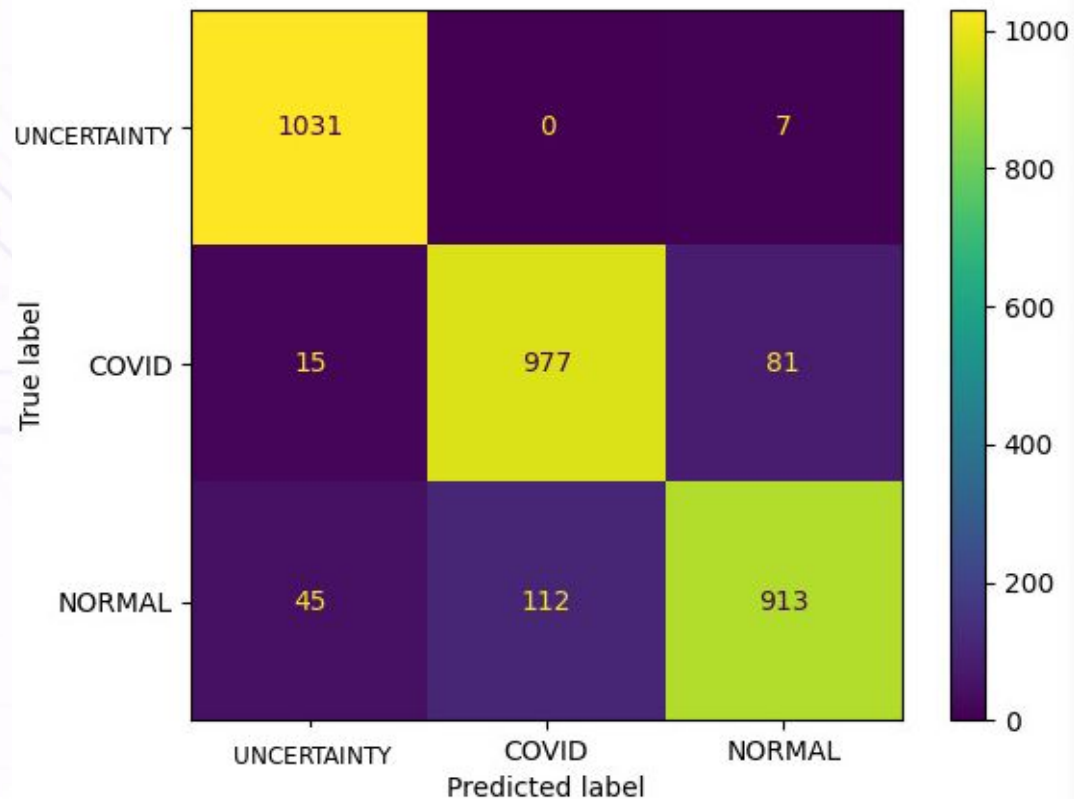


Figura 4: Matriz de confusão

Métrica	Valor
Accuracy	91.82%
Precision	91.81%
Recall	91.90%
F1 Score	91.80%

Tabela 4: Métrica do modelo

- O escopo do trabalho aqui apresentado refere-se a casos de COVID-19 em uma base montada para finalidade experimental (prova de conceito).
- Os resultados demonstram que a abordagem proposta para a modelagem de incerteza é computacionalmente viável.
- Sistemas de apoio à triagem de casos podem ser derivados de trabalhos futuros na linha da proposta apresentada.
- O artigo com sua bibliografia completa e link para nossa implementação no github estão disponíveis nos anais do evento



# Dúvidas?

*Contato:*

[linkedin.com/in/viniciushansen/](https://www.linkedin.com/in/viniciushansen/) ,  
[vinicius.carlo.hansen@gmail.com](mailto:vinicius.carlo.hansen@gmail.com)

[linkedin.com/in/kalyl-henings/](https://www.linkedin.com/in/kalyl-henings/) ,  
[kalyl.henings@edu.udesc.br](mailto:kalyl.henings@edu.udesc.br)