

## **ROUGH SETS APLICADO A IMAGENS DE RX NO DIAGNÓSTICO DO COVID-19 PARA A DETERMINAÇÃO DE INCERTEZA <sup>1</sup>**

Luciano Wayand de Abreu<sup>2</sup>, Gilmário Barbosa do Santos<sup>3</sup>, Kalyl Henings<sup>4</sup>

<sup>1</sup> Vinculado ao projeto “*Rough Sets e Multilayer Perceptron* no Tratamento de Incertezas em Imagens de RX no Diagnóstico de COVID-19”

<sup>2</sup> Acadêmico do Curso de Ciências da Computação – CCT – Bolsista PROIP

<sup>3</sup> Orientador, Departamento de Ciências da Computação – CCT – gilmario.santos@udesc.br

<sup>4</sup> Acadêmico do Curso de Ciências da Computação – CCT – Bolsista PROIP

Em vista da pandemia do novo Coronavírus (COVID-19), decretada pela OMS em 11 de março de 2020, o projeto focou esforços em contribuir para a detecção de casos através da análise por *machine learning*, em conjunto com a técnica de *rough sets* para tratamento das incertezas, de radiografias de tórax.

A incerteza pode ser caracterizada pela determinação de casos que apresentam um mesmo quadro clínico (representado por descritores extraídos de imagens de RX), porém são diagnosticados diferentemente um do outro. Em termos simples, a incerteza ocorreria quando dois pacientes com o mesmo quadro clínico recebessem diagnósticos diferentes (sadio e enfermo, por exemplo).

O projeto foi dividido em duas frentes (A e B). A frente de trabalho aqui relatada (frente-B) foi desenvolvida pelo bolsista Luciano de Abreu e dedicada ao estudo e aplicação da técnica dos Rough Sets com o objetivo de determinar casos de incerteza quanto ao diagnóstico de Covid, por meio da análise da base de descritores que representam os casos clínicos em uma base de imagens de RX.

Para executarmos a classificação de dados utilizamos de uma base pública de imagens de Raio-X de pacientes acometidos pela doença. A base COVID-19 Radiography Database<sup>5</sup> é amplamente utilizada pela comunidade científica, possui 7232 imagens categorizadas por profissionais de saúde em casos positivos e negativos de COVID-19. Das 7232 imagens, 3616 são saudáveis e 3616 são de pacientes infectados. Objetiva-se futuramente ampliar este banco de imagens através de manipulações de rotação e translação.

Para a categorização de indiscernibilidade das imagens é necessário um pré-processamento de forma a transformar a imagem obtida do banco de imagens em uma lista de descritores (valores numéricos). O método HOG (Histogram of Oriented Gradients) foi utilizado por ter sido selecionado a partir da análise desenvolvida na frente-A de trabalho. Foi criado um dataset baseado no método HOG.

Foi realizada uma busca sistemática que conduziu à biblioteca Python Library for Three Way Decision and Rough Set Theory (TWDRS), a qual pode ser aplicada na análise de um dataset bem definido, particularmente na busca de casos de indiscernibilidade que configurem uma classe de casos de incerteza sobre o dataset. Esta biblioteca é gratuita e acessível via github (pypi.org/project/twd/), sendo uma biblioteca de interesse para os objetivos da frente de trabalho aqui relatada.

Além da indiscernibilidade, a biblioteca TWDRS permite extrair outros metadados sobre o dataset, tais como as chamadas aproximações inferiores e superiores de um conjunto de dados e determinar o índice de roughness para o conjunto de dados. Aqui é dado especial destaque à determinação dos casos pertencentes à região de fronteira ou boundary, os quais pertencem a uma categoria de indiscernibilidade, a modelagem de um sistema classificador que incorpore essa classe de casos conduz ao objetivo maior do projeto em questão, qual seja o tratamento da incerteza no diagnóstico de casos de Covid.

Os testes foram executados em um computador com 16GB de Ram DDR4 de 2666Hz, SSD Team Group GX1 de 120GB, processador AMD Ryzen 7 3800X 3.9Ghz e placa gráfica NVIDIA Quadro P2200.

Foi utilizada a biblioteca TWDRS sobre os datasets, identificando casos de incerteza. Os resultados são exibidos nas tabelas abaixo.

**Tabela 2.** Resultado da Análise via TWDRS (Rough Sets) do dataset HOG

Casos Positivos	Casos Negativos	Incertezas
3309	3266	657

Conforme exibido nas tabelas acima, os resultados demonstram a presença de incerteza. Além do quantitativo de 657 casos de incerteza, o método aplicado permite identificar os próprios casos de incerteza. Esse processo viabiliza a construção de um dataset contendo não apenas os casos positivos para doença e os negativos para doença, como também uma categoria extra de casos de incerteza.

A metodologia e abordagem aplicadas viabilizam o treinamento de um classificador (machine learning) o qual poderá tratar (classificar) casos nas três categorias mencionadas incorporando a incerteza inerente ao processo diagnóstico.

**Palavras-chave:** Machine learning, Rough Sets, Incerteza, COVID-19