

PREPARAÇÃO DE BASES DE IMAGENS DIGITAIS RX DE COVID E ANÁLISE USANDO ROUGH SETS¹

Vinicius Hansen², Gilmário Barbosa do Santos³

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

² Acadêmico do Curso de Ciências da Computação – CCT – Bolsista PROIP

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

O novo Coronavírus (COVID-19) teve origem em Wuhan, na China, em dezembro de 2019, de onde vem se espalhando com alta transmissibilidade. O surto foi declarado uma Emergência de Saúde Pública de Âmbito Internacional em 30 de janeiro de 2020 e declarado pela OMS (Organização Mundial da Saúde) como pandemia em 11 de março de 2020. O COVID-19 pode causar doenças no sistema respiratório, febre e tosse e, em alguns casos extremos, pode levar a pneumonia grave e a morte. De fato, mesmo nos dias atuais, ainda não há nenhum medicamento licenciado pela OMS para a cura da infecção por COVID.

Imagens digitais de RX do tórax são um meio não invasivo, de baixo custo e alta sensibilidade para anomalias na região pulmonar, podendo assim ser usado para diagnosticar diferentes anormalidades pulmonares. Porém, a precisão do diagnóstico por meio de RX depende fortemente do especialista que está analisando a imagem em questão. A fadiga visual dos radiologistas aumenta os riscos, potenciais erros de diagnóstico especialmente quanto a detecção de sinais ou lesões mais sutis apresentados no RX. Dessa forma, há um campo para desenvolvimento de aplicações computacionais visando o aproveitamento do potencial das imagens geradas pelos sistemas radiológicos, presentes praticamente em qualquer hospital, e o aumento da acurácia na análise dessas imagens. Esse sistema apoiaria o especialista médico na inspeção das imagens reduzindo erros de diagnóstico.

É de conhecimento geral que a medicina não é uma ciência exata, cujos processos podem ser analisados e modelados sem primordialmente levar-se em conta o fator humano envolvido. A inexatidão também está presente no próprio processo de diagnose médica, o qual consiste num complexo processo dialógico entre médico e paciente. Esse diálogo pode apresentar conflitos resultantes do fato do profissional da saúde não compartilhar dos mesmos valores ou da mesma linguagem dos pacientes e de seus familiares.

Em termos de soluções computacionais na área médica, a teoria Rough Set (RS) pode ser uma ferramenta para acomodar essa inexatidão inerente aos processos médicos. A teoria RS está associada ao conceito de ambiguidade e aproximações entre objetos em um conjunto.

Supondo um universo de casos clínicos, dois pacientes recebem diagnósticos diferentes, porém, os dois apresentam as mesmas características descritivas. Por exemplo, ambos apresentam respectivamente os mesmos sinais que descrevem o estado de saúde de cada paciente nos atributos de “enxaqueca”, “mialgia” e “temperatura”. Esses dois casos clínicos estão em uma região de incerteza, sob o conjunto de atributos a eles associados e o diagnóstico.

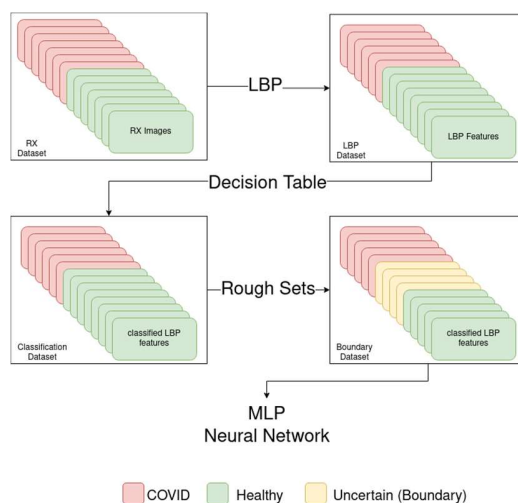


Figura 1. Processo utilizado nos experimentos

O presente relatório encontra-se no contexto do projeto que busca a modelagem de incerteza médica por meio da aplicação de Machine Learning, sendo essencial a configuração de bases de imagens para treinamento do modelo, por conta disso foram necessárias as seguintes etapas:

1. Levantamento de bases bem documentadas e gratuitas contendo imagens de digitais de RX de casos de COVID;
2. Preparação das bases finais (seleção de casos, processamento de imagens para realce de detalhes, etc);
3. Levantamento, configuração e operacionalização de bibliotecas gratuitas e bem documentadas voltadas à análise usando teoria Rough sets.

As bases selecionadas foram a COVID-19 Radiography (extraída da plataforma Kaggle), que contém 3616 exemplares de COVID e 10.200 exemplares de pulmões Sadios e a base BIMCV COVID+ que contém 34.829 exemplares de pulmões com COVID-19. Foram extraídas imagens da Base BIMCV com semelhança SSIM maior que 0.04 pontos em relação as imagens do Covid Radiography até totalizar 10.200 casos de COVID. As imagens com as mesmas foram processadas e preparadas para a fase de análise (mineração de dados de incerteza) por meio de RS. As bibliotecas RS selecionadas foram TWD e RoughSetsBase (Pandas).

A partir da análise com a ferramenta RS foram extraídos os casos incertos do restante dos casos de pulmões saudáveis e acometidos com COVID. Foi utilizado o descritor LBP com configuração que permitiu a obtenção da base final com um número significativo de casos incertos, e assim foram montadas bases com 3 classes (casos saudáveis, de COVID e incertos) para treinamento, testes e extração de métricas do modelo. O fluxo de trabalho é descrito na Figura 1, uma rede neural multicamadas foi treinada e testada com a base constituída.

Os resultados encontrados demonstram a prova a viabilidade do modelo proposto, a precisão geral do modelo é de 88.02%.

Palavras-chave: machine learning, processamento de imagens médicas, COVID-19, tratamento de incerteza, rough sets