

Classification of CT images in COVID-19 and Non-COVID-19 using CNN to extract features and multiple classifiers

Edelson Damasceno Carvalho¹, Edson Damasceno Carvalho²

¹Sistemas de Informação – Universidade Federal do Piauí (UFPI)
Simões – PI – Brasil

²Engenharia Elétrica – Universidade Federal do Piauí (UFPI)
Teresina – PI – Brasil

edelsondamasceno@gmail.com, edsondamasceno@ufpi.edu.br

Abstract. *The COVID-19 is a respiratory disease that has infected more than 12.3 million people worldwide and is responsible for more than 556,300 deaths. The early diagnosis of COVID-19 is essential for the cure and control of the disease. Computed tomography (CT) showed efficient results in the evaluation of patients with suspected COVID-19 infection. CT analysis requires the effort of a specialist, which can lead to diagnostic errors. The use of computer-aided diagnostic systems can minimize the problems generated by CT analysis by specialists. This article presents a methodology for diagnosing COVID-19 using CNN to extract features and multiple classifiers from CT images. The methodology showed an accuracy of 99.79%, recall of 99.79%, accuracy of 99.80%, F-score of 0.997, AUC of 0.997 and kappa index of 0.995. The results obtained show that the proposed methodology can be used as an aid to diagnosis.*

Resumo. *O COVID-19 é uma doença respiratória que já infectou mais de 12.3 milhões de pessoas em todo o mundo e é responsável por mais de 556.300 mortes. O diagnóstico precoce do COVID-19 é essencial para a cura e controle da doença. A tomografia computadorizada (TC) apresentou resultados eficientes na avaliação de pacientes com suspeita de infecção por COVID-19. A análise da TC requer o esforço de um especialista, o que pode levar a erros de diagnóstico. O uso de sistemas de diagnóstico auxiliado por computador pode minimizar os problemas gerados pela análise de TCs por especialistas. Este artigo apresenta uma metodologia para diagnosticar a COVID-19 usando CNN para extração de características e múltiplos classificadores em imagens de TC. A metodologia apresentou uma acurácia de 99,79%, recall de 99,79%, precisão de 99,80%, F-score de 0,997, AUC de 0,997 e índice kappa de 0,995. Os resultados obtidos mostram que a metodologia proposta pode ser utilizada como um sistema de auxílio ao diagnóstico.*

1. Introdução

A doença de coronavírus 2019 (COVID-19) é uma doença respiratória causada pela infecção com o coronavírus da síndrome respiratória aguda grave 2 (SARS-CoV-2) [Gorbalenya et al. 2020]. A COVID-19 teve seus primeiros registros na cidade de Wuhan na China em dezembro de 2019. Em 11 de março de 2020 a COVID-19 foi caracterizada

pela Organização Mundial da Saúde como uma pandemia mundial. A COVID-19 já infectou mais de 12.3 milhões de pessoas em todo o mundo e é responsável por mais de 556.300 mortes [WHO 2020]. No Brasil, a COVID-19 já infectou mais de 1.8 milhões de pessoas e é responsável por mais de 70.300 mortes.

Um dos grandes problemas enfrentados no diagnóstico da COVID-19 é a ineficiência e a escassez de exames médicos, podendo levar as pessoas infectadas a não serem diagnosticadas e a não receber tratamento adequado [Singh et al. 2020]. A tomografia computadorizada (TC) é um exame de imagem que apresentou resultados eficientes no diagnóstico da COVID-19. O principal problema desse método é que depende de um especialista para analisar as imagens de TC, causando fadigas, o que pode levar a erros de diagnóstico.

Para minimizar os problemas gerados pela análise de imagens por especialistas, os sistemas de diagnóstico auxiliado por computador surgem como uma alternativa ao diagnóstico médico. Com os avanços tecnológicos, métodos de *deep learning* foram implementados no desenvolvimento de sistemas de apoio ao diagnóstico. As *convolutional neural networks* (CNN), que são técnicas de *deep learning*, podem interpretar imagens automaticamente [Gozes et al. 2020]. No entanto, a complexidade do modelo, a dificuldade de treinamento, o alto custo computacional e a necessidade de grandes conjuntos de imagens dificultam o desenvolvimento de uma metodologia usando CNN com uma aplicação eficaz.

O trabalho proposto tem como objetivo apresentar o desenvolvimento de uma nova metodologia para classificação de imagens de TC em COVID-19 e Não-COVID-19, usando CNN para extração de características e múltiplos classificadores. Os experimentos mostram que a metodologia proposta apresenta resultados promissores no diagnóstico da COVID-19. O restante deste artigo está organizado da seguinte forma. Na Seção 2, apresentamos os trabalhos relacionados. Na Seção 3, apresentamos a metodologia proposta. Na Seção 4, apresentamos e discutimos os resultados obtidos. Finalmente, na Seção 5, apresentamos as conclusões.

2. Trabalhos Relacionados

Com o surto da COVID-19, houve um crescente esforço para desenvolver métodos automáticos para diagnosticar a COVID-19 em imagens de TC. A Tabela 1 apresenta um resumo dos trabalhos relacionados para o diagnóstico da COVID-19 em imagens de TC.

Tabela 1. Trabalhos relacionados.

| Trabalho | Nº de Imagens | Extração de Características | Classificação | Acurácia |
|-------------------------|----------------------|------------------------------------|----------------------|-----------------|
| [Soares et al. 2020] | 2482 | xDNN | | 97,38% |
| [Ozkaya et al. 2020] | 1500 | CNN | SVM | 98,27% |
| [Barstugan et al. 2020] | 12852 | GLSZM | SVM | 98,71% |
| [He et al. 2020] | 746 | DenseNet | | 86,00% |

3. Metodologia

Para que fosse possível classificar as imagens de TC em COVID-19 e Não-COVID-19, foi empregada a seguinte metodologia: Primeiramente, foi feita a aquisição das imagens

de TC oriundas do SARS-COV-2 CT-Scan Dataset. Em seguida, foi realizada a extração de características utilizando uma arquitetura básica de CNN. Com as características extraídas, foi realizado a classificação usando múltiplos classificadores. Por fim, é feita a validação dos resultados obtidos. A Figura 1 apresenta um resumo da metodologia proposta.

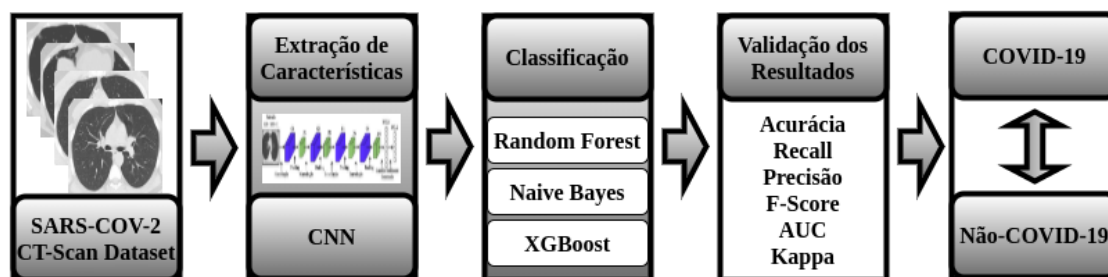


Figura 1. Metodologia proposta.

3.1. Aquisição de Imagens

O SARS-COV-2 CT-Scan Dataset é um conjunto de 2482 imagens de tomografias computadorizadas disponível publicamente, contendo 1252 tomografias positivas para infecção por SARS-CoV-2 (COVID-19) e 1230 tomografias para pacientes não infectados por SARS-CoV-2 (Não-COVID-19) [Soares et al. 2020]. Esses dados foram coletados de pacientes reais em hospitais de São Paulo, Brasil. O objetivo deste conjunto de dados é incentivar a pesquisa e o desenvolvimento de métodos capazes de identificar se uma pessoa está infectada pelo SARS-CoV-2 através da análise de suas tomografias computadorizadas. A Figura 2 apresenta exemplos de imagens do SARS-COV-2 CT-Scan Dataset.

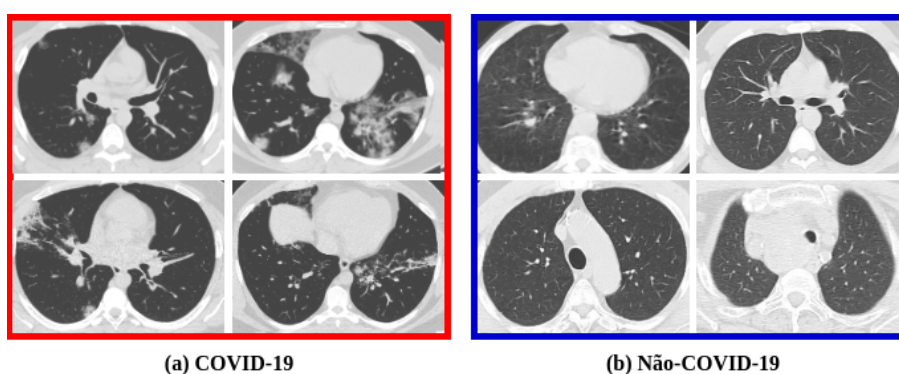


Figura 2. Exemplos de imagens do SARS-COV-2 CT-Scan Dataset.

3.2. Extração de Características

A extração de características é o processo mais importante no desenvolvimento de um sistema automático para classificação de imagens [Ren et al. 2017]. Nos últimos anos, as CNNs, que são modelos de *Deep Learning* foram propostos para etapa de extração de recursos em imagens. O uso de CNN como extrator de características, possui as seguintes vantagens: i) capacidade de extrair características relevantes através de aprendizado de transformações; ii) depende de menor número de parâmetros de ajustes, dado que, cada unidade de uma camada não é conectada com todas as unidades da camada seguinte,

assim, há menos pesos para serem atualizados, facilitando o treinamento; e iii) capacidade de obter informações relevantes sem a necessidade da segmentação de estruturas do objeto.

As CNNs conseguem extrair recursos geralmente úteis de dados, detectar e remover redundâncias de entrada e preservar apenas aspectos essenciais dos dados em representações robustas e discriminatórias [Masci et al. 2011]. Suas camadas semi-conectadas e as totalmente conectadas fornecem um ambiente razoável para o avanço do processo de treinamento e aprendizado [Liu et al. 2018]. Assim, as camadas de convolução servem como extrator de recursos eficiente, especializada na redução da dimensão dos dados e na produção de um conjunto de dados menos redundante.

Uma CNN consiste em camadas alternadas de convolução e *pooling*, depois se transforma em camadas totalmente conectadas à medida que se aproxima da camada de saída. Cada uma das camadas da CNN tem uma função específica na propagação do sinal de entrada. Neste artigo, uma CNN básica foi usada para a extração de recursos das imagens de TC. A CNN consiste em: uma camada de entrada (Entrada), quatro camadas de convolução (C1, C2, C3 e C4), quatro camadas de *pooling* (P1, P2, P3 e P4) e duas camadas totalmente conectadas (FCL1 e FCL2). A arquitetura da CNN para extração de recursos é apresentado na Figura 3. A tabela 2 apresenta um resumo das camadas da CNN.

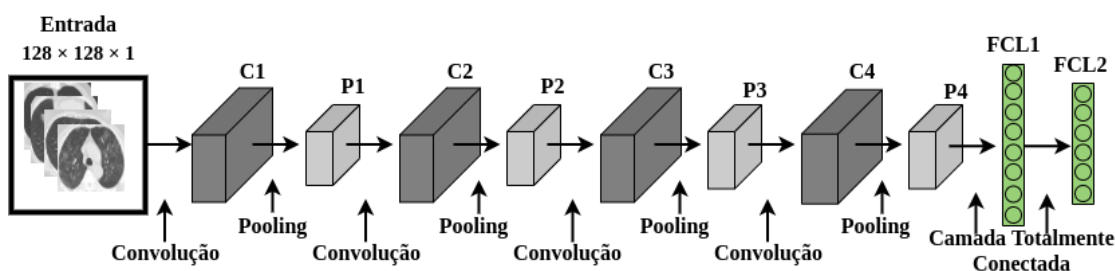


Figura 3. Arquitetura da CNN para extração de recursos das imagens de TC.

Tabela 2. Resumo das camadas da CNN

| Camada | Nº de Kernels | Tamanho do Kernel | Ativação |
|-----------------------------|---------------|-------------------|----------|
| Entrada | 1 | 128×128 | - |
| Convolução (C1) | 32 | 4×4 | ReLU |
| Pooling (P1) | - | 2×2 | - |
| Convolução (C2) | 32 | 4×4 | ReLU |
| Pooling (P2) | - | 2×2 | - |
| Convolução (C3) | 32 | 4×4 | ReLU |
| Pooling (P3) | - | 2×2 | - |
| Convolução (C4) | 32 | 4×4 | ReLU |
| Pooling (P4) | - | 2×2 | - |
| Totalmente conectada (FCL1) | 128 | 1×1 | ReLU |
| Totalmente conectada (FCL2) | 100 | 1×1 | ReLU |

A entrada da CNN é uma imagem em escala de cinza de 128×128 que passa pela primeira camada convolucional (C1) com 32 mapas de recursos com tamanho 4×4 e um

passo de 1. As dimensões da imagem mudam de $128 \times 128 \times 1$ para $128 \times 128 \times 32$. Em seguida, a CNN aplica uma camada de *pooling* (P1) com um tamanho de filtro de 2×2 e um passo de 2. As dimensões da imagem resultantes serão reduzidas para $64 \times 64 \times 32$. Em seguida, uma nova camada convolucional (C2) com 32 mapas de características com tamanho 4×4 e um passo de 1 é aplicada. A quarta camada (P2) é uma camada de *pooling* com tamanho de filtro 2×2 e um passo de 2, desta forma, a imagem de saída será reduzida para $32 \times 32 \times 32$.

Em seguida, uma nova camada de convolução (C3) com 32 mapas de características de tamanho 4×4 e um passo de 1 é aplicada. A sexta camada (P3) é uma camada de *pooling* com tamanho de filtro 2×2 e um passo de 2, para que a saída seja reduzida para $16 \times 16 \times 32$. A sétima camada (C4) é uma convolução com 32 mapas de recursos de tamanho 4×4 e um passo de 1. A oitava camada (P4) é uma camada de *pooling* com filtro 2×2 e um passo de 2, assim, a saída será reduzida para $8 \times 8 \times 32$. A nona camada (FCL1) é uma camada totalmente conectada com 128 mapas de recursos, cada um com tamanho 1×1 . Cada uma das 128 unidades está conectada a todos os 2048 nós ($8 \times 8 \times 32$) na oitava camada (P4). A décima camada (FCL2) é uma nova camada totalmente conectada com 100 unidades, cada uma com tamanho 1×1 . Cada uma das 100 unidades está conectada a todos os 128 nós na camada FCL1.

Para usar a CNN como um extrator de recursos, removeu-se a última camada totalmente conectada da rede e a saída final (FCL2) foi usada como recursos que descrevem a imagem de entrada. As características extraídas das imagens foram usadas em múltiplos classificadores que requerem menos dados para treinamento. Essa estratégia de extração de recursos é amplamente utilizada para aplicações de imagens médicas [Zhu et al. 2015, van Ginneken et al. 2015].

3.3. Classificação

A classificação consiste em reconhecer a qual de um conjunto de categorias pertence uma nova observação, baseado em um treinamento prévio num conjunto de dados cuja categoria é conhecida [Ripley 1996]. Para o desenvolvimento deste trabalho, a classificação foi realizado usando os classificadores *random forest*, *Naive Bayes* e *XGBoost*.

Random Forest é a combinação aleatória de múltiplas árvores de decisão, combinadas para obter uma predição mais estável e com maior acurácia [Breiman 2001]. O *Random Forest* divide cada nó usando o melhor dentre um subconjunto de indicadores escolhidos aleatoriamente naquele nó. Esta estratégia funciona adequadamente em comparação com muitos outros classificadores, além de ser robusto a superajuste nos parâmetros. Os parâmetros utilizados foram: percentual de tamanho da bolsa = 100, tamanho do lote = 100, número de *slots* de execução = 1, profundidade máxima = 0 (ilimitado), número de atributos escolhidos aleatoriamente = 0, número de iterações a serem executadas = 100, número mínimo de instâncias por folha = 1.0, variação mínima para divisão = 0.001 e semente de número aleatório a ser usada = 1.

Naive Bayes é um classificador probabilístico baseado no teorema de Bayes [Yonghong Li and Jain 1998, Smith et al. 2001]. O *Naive Bayes* utiliza dados de treino para formar um modelo probabilístico baseado na evidência das características dos dados. Além disso, o *Naive Bayes* só precisa de um pequeno número de dados para obter uma classificação com uma boa precisão. Os parâmetros usados no *Naive Bayes* foram os

padrões da biblioteca *scikit-learn* [Pedregosa et al. 2011].

eXtreme Gradient Boosting (XGBoost) é um sistema de aprendizado de máquina escalável e eficaz para aumento de árvores [Chen and Guestrin 2016]. O XGBoost é um método de árvore que aplica o princípio de impulsionar aprendizado fraco usando a arquitetura de gradiente descendente. No entanto, o XGBoost melhora a estrutura básica do *Gradient Boosting Machines* por meio da otimização de sistemas e aprimoramentos algorítmicos [Chen and Guestrin 2016]. O XGBoost têm a capacidade de classificar problemas usando uma quantidade mínima de recursos. Os parâmetros usados no XGBoost foram os seguintes: profundidade máxima = 7, taxa de aprendizado = 0.1, iterações = 150, gama = 0, passo delta máximo = 1 e objetivo = “multi:softmax”.

3.4. Validação dos Resultados

Para validar o modelo proposto, foram utilizadas métricas de avaliação baseadas em estatísticas, incluindo acurácia, *recall*, precisão, F-score, AUC e índice kappa. Essas métricas são calculadas com base na matriz de confusão, dado o número de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos.

4. Resultados

Para extrair recursos do modelo de CNN, é necessário treinar a rede. Primeiro, os dados da imagem de entrada são normalizados e transferidos para a camada de entrada da CNN. Após treinar a CNN para obter uma estrutura adequada para a extração de recursos mais robustos das imagens, um conjunto de 100 características foi extraído de cada imagem. Após extrair as características das 2482 imagens de TC, a classificação foi realizada em COVID-19 e Não-COVID-19 usando múltiplos classificadores, com validação cruzada de *k-fold*, sendo $k = 5$. A Tabela 3 apresenta os resultados obtidos.

Tabela 3. Resultados da classificação

| Classificador | Acurácia | Recall | Precisão | F-Score | AUC | Kappa |
|----------------------|-----------------|---------------|-----------------|----------------|--------------|--------------|
| Random Forest | 99,39% | 99,39% | 99,39% | 0,993 | 0,993 | 0,987 |
| Naive Bayes | 99,19% | 99,19% | 99,19% | 0,991 | 0,991 | 0,983 |
| XGBoost | 99,79% | 99,79% | 99,80% | 0,997 | 0,997 | 0,995 |

Como pode-se observar na Tabela 3, a metodologia proposta apresenta resultados bastante eficientes na categorização de imagens de TC em COVID-19 e Não-COVID-19. O XGBoost apresentou os melhores resultados, com uma acurácia de 99,79%, *recall* de 99,79%, precisão de 99,80%, F-score de 0,997, AUC de 0,997 e índice kappa de 0,995. Os resultados apresentados pelos classificadores *random forest* e *naive bayes* foram semelhantes. Isso mostra que a CNN consegue extrair características robustas para categorização das imagens de TC em COVID-19 e Não-COVID-19. Além disso, o uso da CNN como extrator de características não requer treinamento excessivos e nem grandes conjuntos de dados.

A Tabela 4 apresenta uma comparação dos resultados obtidos com a metodologia proposta e os apresentados nos trabalhos relacionados para diagnóstico da COVID-19 em imagens de TC. O estágio de comparação dos resultados é muito complexo, pois muitos fatores podem influenciar uma comparação confiável. Assim, apresentamos um

resumo dos resultados obtidos no método proposto com os apresentados nos trabalhos relacionados.

Tabela 4. Comparação dos resultados.

| Trabalho | Partição | Acurácia | Recall | Precisão | F-score |
|-------------------------|-----------------|-----------------|---------------|-----------------|----------------|
| [Soares et al. 2020] | 80% - 20% | 97,38% | 95,53% | 99,16% | 0,973 |
| [Ozkaya et al. 2020] | 75% - 25% | 98,27% | 98,93% | 98,28% | 0,965 |
| [Barstugan et al. 2020] | 10-fold | 98,71% | 97,56% | 99,62% | 0,985 |
| [He et al. 2020] | 75% - 25% | 86,00% | - | - | 0,850 |
| Método proposto | 5-fold | 99,79% | 99,79% | 99,80% | 0,997 |

Pode-se observar na Tabela 4 que a metodologia proposta apresentou resultados promissores no diagnóstico da COVID-19 em comparação com os trabalhos relacionados. Além disso, os resultados obtidos com a metodologia proposta são estatisticamente mais confiáveis que os apresentados em [Soares et al. 2020, Ozkaya et al. 2020, He et al. 2020] que apresenta resultados em uma única partição.

5. Conclusão

Neste trabalho, apresentou-se uma metodologia para o diagnóstico da COVID-19 em imagens de TC. A metodologia consiste na extração de características usando CNN e classificação usando múltiplos classificadores. Os experimentos foram realizadas em um conjunto de 2482 imagens de TC, sendo 1252 com COVID-19 e 1230 Não-COVID-19. A metodologia proposta apresentou resultados eficientes, com uma acurácia de 99,79%, *recall* de 99,79%, precisão de 99,80%, *F-score* de 0,997, AUC de 0,997 e índice kappa de 0,995. Os recursos extraídos usando a CNN foram robustos para categorização das imagens de TC em COVID-19 e Não-COVID-19. Assim, a metodologia proposta é promissora e pode ser utilizada por um especialista no diagnóstico da COVID-19, fornecendo uma segunda opinião no diagnóstico do paciente.

Referências

- Barstugan, M., Ozkaya, U., and Ozturk, S. (2020). Coronavirus (covid-19) classification using ct images by machine learning methods.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Chen, T. and Guestrin, C. (2016). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '16, page 785–794, New York, NY, USA. Association for Computing Machinery.
- Gorbalenya, A. E., Baker, S. C., Baric, R. S., de Groot, R. J., Drosten, C., Gulyaeva, A. A., Haagmans, B. L., Lauber, C., Leontovich, A. M., Neuman, B. W., Penzar, D., Perlman, S., Poon, L. L., Samborskiy, D., Sidorov, I. A., Sola, I., and Ziebuhr, J. (2020). Severe acute respiratory syndrome-related coronavirus: The species and its viruses – a statement of the coronavirus study group. *bioRxiv*.
- Gozes, O., Frid-Adar, M., Greenspan, H., Browning, P. D., Zhang, H., Ji, W., Bernheim, A., and Siegel, E. (2020). Rapid ai development cycle for the coronavirus (covid-

- 19) pandemic: Initial results for automated detection patient monitoring using deep learning ct image analysis.
- He, X., Yang, X., Zhang, S., Zhao, J., Zhang, Y., Xing, E., and Xie, P. (2020). Sample-efficient deep learning for covid-19 diagnosis based on ct scans. *medRxiv*.
- Liu, X., Hou, F., Qin, H., and Hao, A. (2018). Multi-view multi-scale cnns for lung nodule type classification from ct images. *Pattern Recognition*, 77:262 – 275.
- Masci, J., Meier, U., Cireşan, D., and Schmidhuber, J. (2011). Stacked convolutional auto-encoders for hierarchical feature extraction. In Honkela, T., Duch, W., Girolami, M., and Kaski, S., editors, *Artificial Neural Networks and Machine Learning – ICANN 2011*, pages 52–59, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Ozkaya, U., Ozturk, S., and Barstugan, M. (2020). Coronavirus (covid-19) classification using deep features fusion and ranking technique.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830.
- Ren, X., Guo, H., Li, S., Wang, S., and Li, J. (2017). A novel image classification method with cnn-xgboost model. pages 378–390.
- Ripley, B. D. (1996). *Pattern Recognition and Neural Networks*. Cambridge University Press.
- Singh, D., Kumar, V., Vaishali, and Kaur, M. (2020). Classification of covid-19 patients from chest ct images using multi-objective differential evolution-based convolutional neural networks. *European journal of clinical microbiology infectious diseases : official publication of the European Society of Clinical Microbiology*.
- Smith, I., Lister, R., Ray, M., and Hawson, G. (2001). Naive bayesian prediction of bleeding after heart by-pass surgery. In *The Seventh Australian and New Zealand Intelligent Information Systems Conference, 2001*, pages 317–321.
- Soares, E., Angelov, P., Biaso, S., Higa Froes, M., and Kanda Abe, D. (2020). Sars-cov-2 ct-scan dataset: A large dataset of real patients ct scans for sars-cov-2 identification. *medRxiv*.
- van Ginneken, B., Setio, A. A. A., Jacobs, C., and Ciompi, F. (2015). Off-the-shelf convolutional neural network features for pulmonary nodule detection in computed tomography scans. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 286–289.
- WHO, W. H. O. (2020). Coronavirus disease (covid-19) outbreak situation.
- Yonghong Li and Jain, A. K. (1998). Classification of text documents. In *Proceedings. Fourteenth International Conference on Pattern Recognition (Cat. No.98EX170)*, volume 2, pages 1295–1297 vol.2.
- Zhu, R., Zhang, R., and Xue, D. (2015). Lesion detection of endoscopy images based on convolutional neural network features. In *2015 8th International Congress on Image and Signal Processing (CISP)*, pages 372–376.