

# Análise Comparativa entre Algoritmos de Aprendizagem de Máquina em Classificação de Imagens de Radiografia no Auxílio ao Diagnóstico de Pneumonia

Anderson Silva Ferreira<sup>1</sup>, Rodolfo Carneiro Cavalcante<sup>1</sup>

<sup>1</sup>Núcleo de Ciências Exatas (NCEX) – Universidade Federal de Alagoas (UFAL)  
Caixa Postal 57309-005 – Arapiraca – AL – Brasil

{anderson.ferreira, rodolfo.cavalcante}@arapiraca.ufal.br

**Abstract.** *Pneumonia is a disease that occurs worldwide, with diagnosis made through the analysis of chest x-ray images. The use of machine learning algorithms to perform image classification can assist in the diagnosis of the disease. This work presents a comparative analysis of the performance of the classical techniques  $k$ -Nearest Neighbor, Random Forest, Multi-layer Perceptron, as well as the state of the art, the convolutional neural networks in the classification of 5856 images from the database of chest x-ray images (pneumonia) as normal or pneumonia. Convolutional Neural Network reached the best performance with accuracy of 88.14%.*

**Resumo.** *A pneumonia é uma doença que assola grande parte da população mundial, sendo o diagnóstico realizado por meio da análise de imagens de raio-x do tórax por um especialista. Nesse sentido, o uso de algoritmos de aprendizagem de máquina para realizar a classificação dessas imagens pode auxiliar no diagnóstico da doença. Neste trabalho é apresentado uma análise comparativa do desempenho das técnicas clássicas  $k$ -vizinhos mais próximos, random forest, perceptron multicamadas, bem como do estado da arte, as redes neurais convolucionais na classificação de 5856 imagens da base de dados chest x-ray images (pneumonia) como normal ou pneumonia. A rede neural convolucional apresentou o melhor desempenho com acurácia de 88,14%.*

## 1. Introdução

A pneumonia é uma doença inflamatória aguda causada por micro-organismos (vírus, bactérias ou fungos) ou pela inalação de produtos tóxicos que comprometem os espaços aéreos dos pulmões [FioCruz 2013]. Embora a incidência das infecções respiratórias agudas (IRA) possa ser semelhante em várias regiões, a mortalidade por tais doenças adquire especial importância em países como o Brasil, com dimensões continentais e diferenças inter-regionais marcadas pelas desigualdades sociais, principalmente no acesso aos serviços de saúde [de Olanda Ferraz et al. 2017].

Segundo dados do Sistema de Informações sobre Mortalidade do Ministério da Saúde, apenas em 2010, ocorreram mais de 208 mil internações no Sistema Único de Saúde (SUS) em decorrência da pneumonia e aproximadamente 25 mil idosos morreram devido à enfermidade [DATASUS 2010]. O Brasil é um dos 15 países com maior número de casos anuais de pneumonia clínica em menores de 5 anos (1,8 milhão). Nesse grupo, 30 a 50% das consultas ambulatoriais, mais de 50% das hospitalizações e 10 a 15% dos

óbitos são atribuídos às infecções respiratórias agudas, sendo 80% destes por pneumonia [Cardoso 2010].

O diagnóstico da pneumonia é dado através da realização de uma radiografia do tórax, podendo necessitar ou não de exames laboratoriais complementares, como um hemograma, por exemplo. Entretanto, há uma discordância substancial na literatura no diagnóstico de IRA entre médicos com maior nível de treinamento, que têm em comum o contato frequente com crianças portadoras de doenças respiratórias, levando em consideração apenas a imagem da radiografia do tórax [Sarria et al. 2003]. A partir dessa informação e com o avanço do uso de algoritmos de aprendizagem de máquina para realizar classificação de imagens, o objetivo deste trabalho é efetuar uma comparação entre algoritmos de aprendizagem de máquina e analisar qual deles possui o melhor desempenho para o tipo de imagem do exame de radiografia do tórax, levando em consideração a análise da acurácia e três métricas que indicam a qualidade da classificação: *recall*, *precision* e *f1\_score*.

Foram escolhidos quatro algoritmos de classificação para serem analisados, sendo eles: *k*-vizinhos mais próximos (*k-Nearest Neighbor - kNN*), *Random Forest Classifier - RF*), perceptron multicamadas (*Multilayer Perceptron - MLP*) e rede neural convolucional (*Convolutional Neural Network - CNN*). Os três primeiros são algoritmos clássicos e o último é o estado da arte em problemas de classificação de imagens.

O restante deste trabalho está organizado como se segue: Na Seção 2 são descritos os algoritmos comparados, na Seção 3 são descritos alguns trabalhos relacionados com o assunto, na Seção 4 é descrito como a análise foi realizada, na Seção 5 são descritos os resultados obtidos e efetuadas algumas discussões a respeito das performances alcançadas e por fim na Seção 6 é apresentada a conclusão a partir dos resultados obtidos no estudo e são propostos trabalhos futuros.

## 2. Fundamentação Teórica

O aprendizado de máquina é definido como o campo de estudo que permite que computadores tenham a habilidade de resolver um problema sem terem sido explicitamente programados, ou seja, explora o estudo e construção de algoritmos que podem aprender de seus erros, e por conseguinte efetuar previsões a partir de dados [Provost and Kohavi 1998].

As tarefas do aprendizado de máquina são tipicamente classificadas em 3 categorias: aprendizado supervisionado, não-supervisionado e semi-supervisionado. O aprendizado supervisionado consiste em técnicas onde o algoritmo tenha a capacidade de aprender padrões sobre os dados, a partir de exemplos de entradas e saídas desejadas (dados rotulados) apresentadas por um especialista<sup>1</sup>, adquirindo assim a capacidade de mapear entradas válidas, de um mesmo contexto, para uma saída. Por sua vez, nas técnicas de aprendizado não supervisionado não são fornecidas as saídas desejadas para os dados de entrada correspondentes (dados não rotulados), cabendo unicamente ao algoritmo encontrar padrões nas entradas que gerem uma determinada saída. Algoritmos desse tipo são muito usuais em tarefas de agrupamento. Por fim, no aprendizado

---

<sup>1</sup>Pessoa com conhecimento suficiente para efetuar a classificação do conjunto de entradas e saídas desejadas

semi-supervisionado, são aplicadas combinações de técnicas do aprendizado supervisionado e do aprendizado não-supervisionado. O aprendizado semi-supervisionado, assume que, juntamente com o conjunto de treinamento, há um segundo conjunto de dados não rotulados, também disponível durante o treinamento. Uma das metas do aprendizado semi-supervisionado é o treinamento de classificadores quando uma grande quantidade de dados não rotulados está disponível juntamente com um pequeno conjunto de dados rotulados [Russell and Norvig 2016].

Dentre os algoritmos clássicos de aprendizagem de máquina temos o  $k$ -Vizinhos mais próximos. Este algoritmo apresenta aprendizado baseado em instância, sendo bastante utilizado para reconhecimento de padrões. É um dos métodos mais antigos e simples para classificação e possui performance competitiva quando comparado com outros métodos [Cover et al. 1967]. Neste método um objeto é classificado por uma pluralidade de votos de seus vizinhos, com o objeto sendo atribuído à classe mais comum entre os seus  $k$  vizinhos mais próximos, sendo  $k$  um número inteiro positivo, relativamente pequeno e geralmente ímpar para evitar empates em classificações binárias [Altman 1992].

*Random Forests* são métodos de aprendizado em conjunto que operam construindo várias árvores de decisão no momento do treinamento, onde cada árvore de decisão é inicializada com base nos valores de um vetor aleatório, desde que garantido que todos os vetores sigam a mesma distribuição de probabilidade, garantindo desta forma que o modelo não sofra *overfit*<sup>2</sup> com base na Lei dos Grandes Números<sup>3</sup>. Nestes métodos, a classe de saída é a moda das classes apontadas pelas diversas árvores individuais [Breiman 2001]. As árvores de decisão são modelos estatísticos que utilizam um treinamento supervisionado em tarefas de classificação e regressão, de modo que uma decisão é tomada através do caminhamento a partir do nó raiz até o nó folha.

Uma rede neural artificial (RNA) é um sistema de neurônios interconectados que, a partir de valores de entrada, tem a capacidade de armazenar conhecimento experimental e torná-lo disponível para uso, simulando o comportamento de redes neurais biológicas. Uma RNA se assemelha ao cérebro em dois aspectos: o conhecimento é adquirido pela rede a partir do ambiente, através de um processo de aprendizagem e o conhecimento adquirido é armazenado através de forças de conexões entre neurônios, conhecidas como pesos sinápticos [Haykin 2007].

O Perceptron Multicamadas foi uma solução de aprendizagem de máquina popular na década de 80 [Wasserman and Schwartz 1988], que utiliza a abordagem de redes neurais artificiais e consiste em no mínimo três camadas de neurônios artificiais (nós), sendo uma camada de entrada, pelo menos uma camada oculta<sup>4</sup> e uma camada de saída. Cada nó das camadas ocultas utiliza uma função de ativação<sup>5</sup> não linear para calcular as entradas ponderadas para a saída de cada neurônio. É utilizada uma técnica de aprendizado supervisionado chamada de *backpropagation*, cujo objetivo é calcular um gradiente

---

<sup>2</sup>*Overfitting* é um termo usado em estatística para descrever quando um modelo estatístico se ajusta muito bem ao conjunto de dados anteriormente observado, mas se mostra ineficaz para prever novos resultados.

<sup>3</sup>A lei dos grandes números é um teorema fundamental da teoria da probabilidade, que descreve o resultado da realização da mesma experiência repetidas vezes.

<sup>4</sup>Conjunto de neurônios localizados entre a camada de entrada e a camada de saída.

<sup>5</sup>Função aplicada sobre os valores de entrada nos neurônios das camadas ocultas para simular o impulso elétrico.

necessário para realizar o ajuste dos pesos que são utilizados na rede [Nielsen 2015].

Rede Neural Convolutiva (CNN) é uma classe de rede neural artificial que utiliza técnicas de *deep learning*, ou seja, uma variação de perceptrons multicamadas desenvolvidos a demandar o mínimo de pré-processamento possível. É inspirado na organização do córtex visual dos animais e usado principalmente em reconhecimento de imagens e processamento de vídeo [Sainath et al. 2013]. As CNNs exploram a correlação espacial local impondo um padrão de conectividade local entre os neurônios das camadas adjacentes. Camadas convolucionais aplicam uma operação de convolução à entrada, passando o resultado para a próxima camada. A convolução emula a resposta de um neurônio individual a estímulos visuais [Litjens et al. 2017].

### 3. Trabalhos Relacionados

Na literatura, encontram-se vários trabalhos comparativos entre algoritmos de classificação nos mais diferentes contextos. [Fernández-Delgado et al. 2014] em seu trabalho efetua uma avaliação exaustiva da performance de 178 classificadores distintos em 121 bases de dados. No trabalho de [Sisodia et al. 2014], é efetuada uma análise de vários algoritmos para classificação de imagens de sensoriamento remoto em diferentes sensores e períodos de tempo. Já [Hernández Millán et al. 2018] realizam um estudo comparativo de diferentes técnicas de aprendizado de máquina supervisionado para classificação de imagens de documentos institucionais.

Além disso, é recorrente a busca do aprimoramento de técnicas que possam auxiliar no diagnóstico de IRA. [Oliveira et al. 2008] apresenta uma abordagem no diagnóstico da pneumonia pediátrica usando imagens do raio-x do tórax auxiliado por computador. Nessa abordagem é efetuada a recuperação de imagens semelhantes com base na extração de características usando coeficientes de transformada *wavelet* da imagem em conjunto com o algoritmo de classificação dos  $k$  vizinhos mais próximos. Já [Sousa et al. 2014] apresentou uma evolução do trabalho de [Oliveira et al. 2008] com o objetivo de melhorar a precisão e a robustez do sistema. Para tal, foram testados e comparados cinco classificadores de aprendizado de máquina temporários, sendo eles: *Naive Bayes*, *k-Nearest Neighbor*, *Support Vector Machines*, *Multi-Layer Perceptron* e *Decision Tree*, combinados com algoritmos de redução de dimensionalidade.

Neste trabalho é apresentada uma análise comparativa de diferentes técnicas de aprendizagem de máquinas supervisionado, diferenciando-se dos trabalhos acima apresentados por utilizar técnicas de aprendizagem profunda (*deep learning*) além de algoritmos tradicionais para a classificação de imagens de radiografias do tórax para o auxílio no diagnóstico da pneumonia.

### 4. Estudo Comparativo

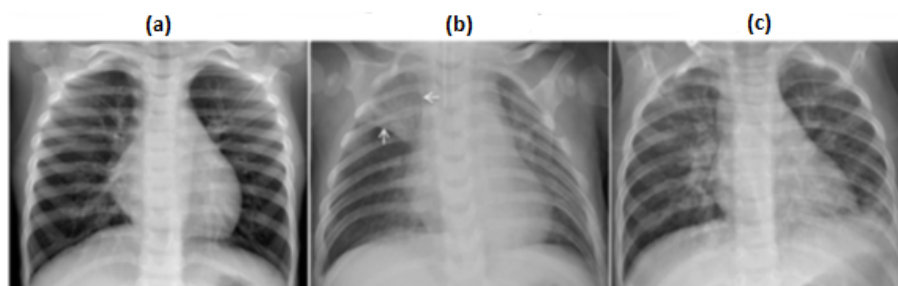
O dataset utilizado para o desenvolvimento dos experimentos, *Chest Ray-X Images* (Pneumonia) [Kermany et al. 2018], é composto de imagens de radiografias do tórax com diversas dimensões. Para viabilizar a experimentação, todas as imagens foram redimensionadas para as dimensões  $150 \times 150$  pixels em 3 canais RGB, utilizando redimensionamento por interpolação. A base de dados é composta de 5856 imagens de radiografias do tórax divididas em 2 categorias, normal (1583) e pneumonia (4273). Essa base de dados foi organizada para o procedimento de treino e teste dos métodos da seguinte forma: 1341

imagens NORMAL e 3875 PNEUMONIA para treino, 234 NORMAL e 390 PNEUMONIA para teste e 8 imagens de cada categoria para validação.

As imagens de radiografia do tórax foram selecionadas de pacientes pediátricos de 1 a 5 anos de idade do Centro Médico de Mulheres e Crianças de Guangzhou, no sul da China. Todas as imagens de radiografia do tórax foram realizadas como parte do atendimento clínico de rotina dos pacientes.

Todas as radiografias foram inicialmente passadas por um controle de qualidade, removendo-se todas as digitalizações de baixa qualidade ou ilegíveis. Os diagnósticos para as imagens foram então classificados por dois médicos especialistas antes de serem liberados para utilização em algoritmos de inteligência artificial. A fim de contabilizar quaisquer erros de classificação, o conjunto de avaliação também foi verificado por um terceiro especialista.

A radiografia do tórax normal (Figura 1a) representa pulmões desobstruídos sem quaisquer áreas de opacificação anormal na imagem. A pneumonia bacteriana (Figura 1b) tipicamente exibe uma consolidação lobar focal, neste caso no lobo superior direito (setas brancas), enquanto a pneumonia viral (Figura 1c) se manifesta com um padrão intersticial mais difuso em ambos os pulmões [Kermary et al. 2018].



**Figura 1. (a) Sem Pneumonia, (b) Pneumonia Bacteriana (c) Pneumonia Viral.**

Os experimentos foram realizados em um microcomputador com processador Intel® Core™ i7 7500U @ 2,7Ghz x86\_64, 8GB de memória RAM e Sistema Operacional Ubuntu 16.04.5 LTS Linux 4.15.0-45-generic. Para implementação dos algoritmos kNN, RF e MLP foi utilizado *Python 2.7.15* e a biblioteca *scikit-learn 0.20.1* [Pedregosa et al. 2011]. Para implementação da CNN foi utilizado *Python 3.6.8* e a biblioteca *Keras 2.2.4* [Chollet et al. 2015]. Em todos os casos foram utilizadas as configurações *default* das bibliotecas, com exceção das alterações informadas particularmente em cada um dos testes explicitados abaixo.

#### **4.1. *k*-Vizinhos Mais Próximos**

Para definir o número de vizinhos a ser considerado no processo de classificação, foi realizada uma análise da acurácia em função do número de vizinhos  $k$  definido por  $\{k = 2x + 1, x \in \mathbb{N} < 10\}$ . Para todos os casos, como métrica de comparação das instâncias, foi considerada a distância euclidiana. Conforme Figura 2, 11 vizinhos obteve a melhor performance.

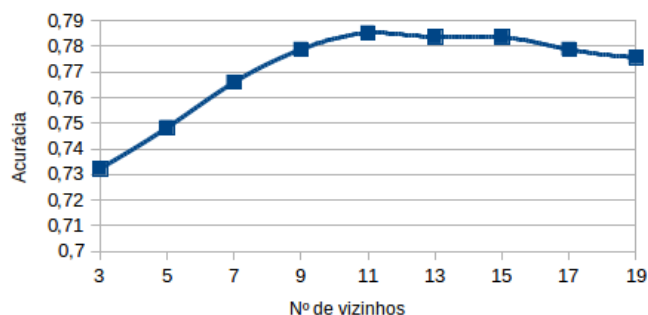


Figura 2. Gráfico da análise de número de vizinhos.

#### 4.2. Random Forest

Para definir o número de árvores para cada floresta a ser considerado no processo de classificação, foi realizada uma análise da acurácia em função do número de árvores  $t$  definido por  $\{t \in \mathbb{N}, 0 < t < 501\}$ . Além disso, todos os demais parâmetros tais quais, *criterion*, número máximo de *features*, número máximo de folhas, etc. foram mantidos com as configurações padrões da biblioteca informada na Seção 4. O número de 100 árvores por floresta obteve a melhor performance.

#### 4.3. Perceptron Multicamadas

Para definir a arquitetura a ser utilizada no processo de classificação, foram avaliados diferentes valores para o parâmetro *hidden\_layer\_sizes*, que indica a quantidade de camadas ocultas e a quantidade de nós por camada, com o objetivo de obter a melhor acurácia dentre os valores testados. Para tal, foram selecionadas 30 configurações distintas com quantidade de camadas variando de 1 a 5 e quantidade de nós variando de 1 a 100. As 5 configurações com melhores desempenhos podem ser vistas na Tabela 1, de modo que a quantidade de elementos entre colchetes representa a quantidade de camadas ocultas e cada elemento corresponde a quantidade de neurônios na respectiva camada.

Tabela 1. Configurações testadas para o *hidden\_layer\_sizes*

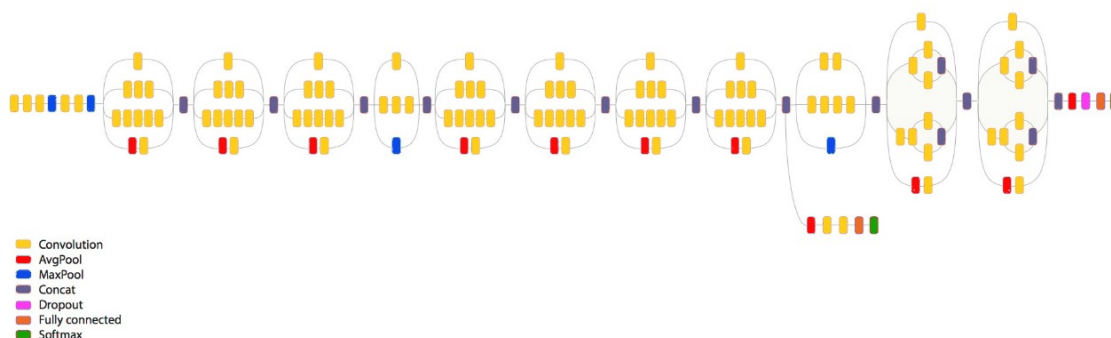
Configurações	Acurácia
1.[4, 16, 8, 12]	0,7917
2.[5, 15, 10, 20]	0,7916
3.[7, 9, 11]	0,7884
4.[9, 2, 7, 12]	0,7884
5.[100]	0,7852

Através desta análise, foi percebido que a configuração de uma rede com 4 camadas ocultas: 4 nós na primeira camada, 16 nós na segunda, 8 nós na terceira e 12 nós na quarta camada, teve o melhor desempenho perante as outras testadas. Além do *hidden\_layer\_sizes*, também foi alterado o número máximo de iterações de 200 para 500.

#### 4.4. Rede Neural Convolutacional

Para o desenvolvimento do experimento, foi utilizado o modelo *InceptionV3* que consiste em uma arquitetura de rede neural convolutacional. Trata-se de um modelo de reconhecimento de imagens amplamente utilizado que demonstrou ter uma precisão superior a 78,1% no dataset ImageNet [Szegedy et al. 2016].

O modelo em si é composto de blocos simétricos e assimétricos, incluindo *convolutions*, *average pooling*, *max pooling*, *concat*, *dropouts* e *fully connected layers* [Szegedy et al. 2016]. A Figura 3 apresenta as camadas presentes na arquitetura do modelo *InceptionV3*.



**Figura 3. Arquitetura do modelo InceptionV3. [Szegedy et al. 2016]**

Foram definidos nos parâmetros do classificador, o formato de entrada das imagens ( $150 \times 150$  pixels em 3 canais) e um número máximo de 100 épocas para treinamento do modelo.

## 5. Resultados e Discussões

A Tabela 2 apresenta os resultados dos experimentos descritos na Seção 4. Para condução dos experimentos foi adotada a semântica de amostra negativa para imagens de radiografia de pacientes sem pneumonia e amostra positiva para imagens de radiografia de pacientes com algum tipo de pneumonia. Cada experimento foi realizado 20 vezes, utilizando-se do método de experimentação *holdout*<sup>6</sup> e calculada a média das métricas de desempenho descritas abaixo:

- **Acurácia** - É a proporção de predições verdadeiras entre o número total de casos examinados .
- **Recall score** - Mede a capacidade do classificador encontrar todas as amostras positivas.
- **Precision score** - Mede a capacidade do classificador de não classificar uma amostra negativa como positiva.
- **F1 score** - Mede a qualidade da classificação, levando em consideração o *recall* e a *precision*.

Como pode ser observado na Figura 4, a rede neural convolucional teve um desempenho superior aos demais métodos, obtendo uma acurácia de 88,14%, sendo superior inclusive nas métricas que avaliam a qualidade da classificação. Os demais classificadores tiveram resultados bastante similares tanto em acurácia quanto em qualidade de classificação, com destaque para o kNN que apesar de sua simplicidade obteve a segunda maior acurácia.

<sup>6</sup>O método *holdout*, consiste em dividir o conjunto de dados em dois conjuntos disjuntos, chamados de conjunto de treino e conjunto de teste.

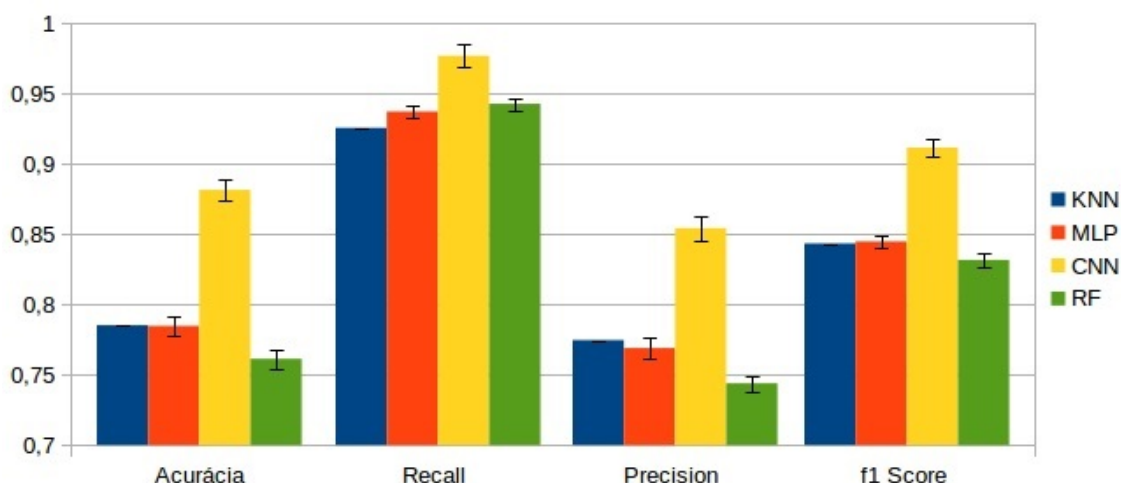


Figura 4. Comparativo das métricas de desempenho dos classificadores.

Tabela 2. Resultados experimentais dos diferentes classificadores analisados

Classificador	Acurácia	Recall	Precision	F1 Score
k-Nearest Neighbors	78,52 %	0,9256	0,7746	0,8434
Random Forest	76,13%	0,9426	0,7439	0,8616
Multi-layer Perceptron	78,47	0,9371	0,7690	0,8448
Convolutional Neural Network	<b>88,14%</b>	<b>0,9769</b>	<b>0,8543</b>	<b>0,9115</b>

Tabela 3. Desvio padrão das métricas apresentada na Tabela 2

Classificador	Acurácia	Recall	Precision	F1 Score
k-Nearest Neighbors	±0%	±0	±0	±0
Random Forest	±0,683%	±0,007	± <b>0,006</b>	±0,005
Multi-layer Perceptron	± <b>0,654%</b>	± <b>0,004</b>	±0,007	± <b>0,004</b>
Convolutional Neural Network	±0,982%	±0,009	±0,008	±0,008

Na Tabela 3 são apresentados os valores de desvio padrão para cada uma das métricas analisadas. Com base nestes valores, pode-se observar que apesar do classificador CNN apresentar melhores resultados médios, é também o método que apresenta maior variação em torno da média, ou seja, os valores obtidos em cada execução diferem-se mais entre si. Apesar de apresentar a maior variação, os menores valores do classificador CNN ainda possuem um valor superior aos melhores valores dos demais classificadores analisados.

Em contrapartida, o classificador MLP é o que apresenta menor variação nas métricas, com exceção da *Precision*, onde o classificador RF apresenta a menor variação de todos os métodos comparados. É importante salientar que como o classificador kNN baseia-se na distância entre os pontos no espaço e essa distância é a mesma em todas as execuções, não há variação em suas métricas nas diferentes execuções.



## 6. Conclusão

Neste trabalho foi realizada uma análise comparativa entre 4 diferentes técnicas de classificação de imagens para auxílio no diagnóstico da pneumonia através do raio-x do tórax. Os resultados apontam que a rede neural convolucional obteve um resultado expressivo em relação às outras técnicas analisadas atingindo uma acurácia de 88,14% com qualidade de classificação acima das demais.

Como trabalhos futuros pode ser feita essa mesma análise considerando outras técnicas de classificação, como também efetuar uma análise comparativa apenas com técnicas de redes neurais convolucionais para analisar qual delas possui a melhor performance para imagens do tipo raio-x do tórax, além disso pode ser feita a aplicação destes algoritmos em outros tipos de imagens médicas para detecção de doenças como câncer de mama, câncer de pulmão, etc.

## Referências

- Altman, N. S. (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Cardoso, A. M. (2010). A persistência das infecções respiratórias agudas como problema de saúde pública.
- Chollet, F. et al. (2015). Keras. <https://keras.io>.
- Cover, T. M., Hart, P. E., et al. (1967). Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27.
- DATASUS (2010). Óbitos por causas evitáveis de 5 a 74 anos - brasil. Disponível em: <http://tabnet.datasus.gov.br/cgi/tabcgi.exe?sim/cnv/evitb10uf.def>. Acesso em: 05 fev 2019.
- de Olanda Ferraz, R., Oliveira-Friestino, J. K., and Francisco, P. M. S. B. (2017). Tendência de mortalidade por pneumonia nas regiões brasileiras no período entre 1996 e 2012. *Jornal Brasileiro de Pneumologia*, 43(4):274–279.
- Fernández-Delgado, M., Cernadas, E., Barro, S., and Amorim, D. (2014). Do we need hundreds of classifiers to solve real world classification problems? *The Journal of Machine Learning Research*, 15(1):3133–3181.
- FioCruz, A. (2013). Pneumonia. Disponível em: <https://agencia.fiocruz.br/pneumonia>. Acesso em: 05 fev 2019.
- Haykin, S. (2007). *Redes neurais: princípios e prática*. Bookman Editora.
- Hernández Millán, R., Mendoza-Moreno, M., Portocarrero López, L. M., and Castro-Romero, A. (2018). Comparative study of machine learning supervised techniques for image classification using an institutional identification documents dataset. In *2018 Congreso Internacional de Innovación y Tendencias en Ingeniería (CONIITI)*, pages 1–6.

- Kermany, D. S., Goldbaum, M., Cai, W., Valentim, C. C., Liang, H., Baxter, S. L., McKeown, A., Yang, G., Wu, X., Yan, F., et al. (2018). Identifying medical diagnoses and treatable diseases by image-based deep learning. *Cell*, 172(5):1122–1131.
- Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., and Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88.
- Nielsen, M. A. (2015). *Neural networks and deep learning*, volume 25. Determination press USA.
- Oliveira, L. L. G., e Silva, S. A., Ribeiro, L. H. V., de Oliveira, R. M., Coelho, C. J., and Andrade, A. L. S. (2008). Computer-aided diagnosis in chest radiography for detection of childhood pneumonia. *International journal of medical informatics*, 77(8):555–564.
- 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.
- Provost, F. and Kohavi, R. (1998). Glossary of terms. *Journal of Machine Learning*, 30(2-3):271–274.
- Russell, S. J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- Sainath, T. N., Mohamed, A.-r., Kingsbury, B., and Ramabhadran, B. (2013). Deep convolutional neural networks for lvcsr. In *2013 IEEE international conference on acoustics, speech and signal processing*, pages 8614–8618. IEEE.
- Sarria, E., Fischer, G. B., Lima, J. A., Barreto, S. S. M., Flôres, J. A., and Sukiennik, R. (2003). Concordância no diagnóstico radiológico das infecções respiratórias agudas baixas em crianças. *J Pediatr (Rio J)*, 79(6):497–503.
- Sisodia, P. S., Tiwari, V., and Kumar, A. (2014). A comparative analysis of remote sensing image classification techniques. In *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, pages 1418–1421. IEEE.
- Sousa, R. T., Marques, O., Curado, G. T., da Costa, R. M., Soares, A. S., Soares, F. A., and de Oliveira, L. L. (2014). Evaluation of classifiers to a childhood pneumonia computer-aided diagnosis system. In *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, pages 477–478. IEEE.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2818–2826.
- Wasserman, P. D. and Schwartz, T. (1988). Neural networks. ii. what are they and why is everybody so interested in them now? *IEEE Expert*, 3(1):10–15.