

ATUAL CENÁRIO DE PUBLICAÇÕES COM USO DE MACHINE LEARNING PARA DIAGNÓSTICOS EM SAÚDE

PEREIRA, Suzana Maria de Freitas ¹,
MORAIS, Alana Marques ²,
LEITE, Danilo Rangel Arruda ²

1 Discente do curso de Sistemas para Internet – UNIESP

2 PHD em Ciência da Computação

3 Docente do curso de Sistemas para Internet – UNIESP

RESUMO

Machine Learning (ML) possibilita replicar comportamentos, tomar decisões, e utilizar suas técnicas para treinamento. A utilização de ML tem ganhado muito espaço na área da Saúde, apesar de ser uma área bastante complexa. As técnicas de ML podem ajudar a prevenir doenças e na tomada de decisão para pacientes com comorbidades. Este trabalho tem como objetivo entender o atual cenário de publicações de ML na área da Saúde, por meio de um mapeamento sistemático realizado em diferentes bases de dados, no qual foi possível buscar trabalhos em diversas subáreas da Saúde. Os resultados deste estudo apontaram que há um grande volume de publicações nos anos de 2019 e 2020, sendo a China com maior número de publicações, a subárea da Saúde mais estudada foi a de Medicina e a patologia mais explorada foi o Alzheimer. Este trabalho serviu para explorar uma temática muito rica e que muito ainda tem a ser feito e contribuirá para nortear e motivar estudos futuros.

Palavras-chave: Machine Learning. Saúde. Diagnóstico

1. INTRODUÇÃO

Machine Learning (ML) ou Aprendizagem de Máquina é uma área da Inteligência Artificial que, segundo a *International Business Machines Corporation* (IBM), utiliza métodos estatísticos para treinar dados por meio de algoritmos, possibilitando realizar classificações, previsões e insights que geram métricas que contribuirão para a tomada de decisão nas organizações.

Para Schneider (2016) ML faz parte de um subgrupo da ciência da computação, que se preocupa com reconhecimento de padrões e de teorias de aprendizagem de máquina com Inteligência Artificial, caracterizando-se

pelo uso de modelos e algoritmos de aprendizagem automática, que fazem previsões a partir de dados já conhecidos.

Nos últimos tempos, a ciência de dados tem ganhado bastante visibilidade devido à grande quantidade de dados que estamos gerando a todo momento, a exemplo do uso dos diversos tipos de dispositivos eletrônicos que as pessoas têm utilizado, como *smartphones*, *smartwatch*, eletrodomésticos inteligentes, sistemas de prontuários médicos, e outros MOTTA (2008). Esse fenômeno é denominado Internet das Coisas, onde qualquer dispositivo pode estar conectado a alguma rede, gerando assim a hiperconectividade.

Para Lobo (2018) as evoluções da prática médica com o uso de sistemas de processamento que registram dados nos prontuários eletrônicos por linguagem natural, uso de computadores ou *smartphones* durante a consulta, utilização de tecnologias que fazem reconhecimento de imagens, além de sistemas que ajudam a interação do médico com o paciente, permitindo apoio, confiabilidade e acompanhamento dos quadros de saúde do paciente.

A utilização de ML na área da saúde tem ganhado notoriedade, apesar de ser uma área bastante complexa, crítica e que passa por constantes mudanças. As técnicas de ML podem ajudar nos processos de tomada de decisão, seja para diagnosticar, intervir, ou monitorar paciente com comorbidades BATISTA (2019).

Durante a pandemia da Covid-19, a população acompanhou, em tempo real, o crescimento do número de casos, mortes registradas, quantitativo de pessoas vacinadas, graças ao alto poder de processamento de dados, bem como os dados disponibilizados pelo Ministério da Saúde, que podem ser utilizados para diversos estudos, treinar modelos de ML, e assim ter informações relevantes para políticas governamentais e para a ciência como um todo.

No geral, as técnicas de ML têm sido muito utilizadas para trazer qualidade de vida para o paciente por meio do diagnóstico precoce das doenças de forma precisa, além de reduzir o custo benefício dos tratamentos, contribuindo assim para a redução dos índices de mortalidade da população NURMAINI (2019).

Para entender como está o atual cenário de publicações de ML na área da Saúde, este trabalho irá discutir os resultados de um mapeamento sistemático realizado em diferentes bases de dados científicas: *Science Direct*, IEEE, e CAPES. Foi possível buscar trabalhos em diversas subáreas da Saúde, com estudos que utilizem ML para diagnóstico de doenças e/ou comorbidades. Quais áreas da Saúde estão sendo mais estudadas? Por que elas estão sendo mais discutidas?

Ao final deste trabalho, foi possível verificar que a área da Medicina é a que mais tem publicações e o Alzheimer tem sido mais estudado. O presente trabalho foi desenvolvido no período de Outubro de 2020 a Outubro de 2021 e está organizado em tópicos: a contextualização da problemática, em seguida a fundamentação teórica, o método usado para os resultados, os quais serão abordados juntamente com alguns insights obtidos, e ao final concluiremos toda a ideia do trabalho.

2. FUNDAMENTAÇÃO TEÓRICA

Para Tulloch (2020) os constantes avanços na tecnologia que utilizam algoritmos de ML em saúde tem se tornando cada vez mais popular por sua capacidade de diminuir erros e a dependência humana, implicando na redução de custos.

Segundo Rasheed (2020) os avanços tecnológicos têm ajudado aos neurocientistas a utilizar os dados de neuroimagens, e acredita que a utilização de ML e Deep Learning são tecnologias que podem somar ainda mais às comumente usadas para a construção de novas teorias relacionadas ao cérebro.

Sun (2020) menciona que na área médica, há uma grande utilização das técnicas ML em diversas especialidades como cardiologia, oftalmologia, nefrologia, radioterapia, neurologia, endocrinologia, oncologia, estomatologia, psicologia e outras. Essas técnicas tem a capacidade de analisar grandes volumes de dados com a finalidade de obter previsão e diagnóstico de doenças, e sua aplicação visa solucionar alguns problemas da prática médica como baixa eficiência e erros de diagnóstico.

É possível encontrar na literatura diversos métodos isolados ou combinados a outros métodos, como Martinez-Murcia (2018) sugere o uso de Redes Neurais Convolucionais para buscar padrões a fim de identificar o surgimento de doenças como o Parkinson. No geral, métodos baseados em aprendizado profundo, ajudarão a descobrir padrões mais complexos e, superando as metodologias estatísticas clássicas para diagnósticos mais precisos.

Graham (2019) afirma que a Inteligência artificial em saúde, utiliza algoritmos e software que exercem o papel das funções cognitivas de seres humanos para analisar dados médicos complexos, como imagens de exames ou prontuários médicos. As ferramentas de IA usam esses dados para inferir mudanças patológicas no funcionamento cognitivo e deve ser capaz de identificar aos primeiros sinais de problemas.

Para Liu (2009) as imagens de exames são evidências muito importantes para o diagnóstico médico, mas sua interpretação depende de análises de seres humanos e com a grande demanda de exames realizados diariamente se torna um desafio enorme para os especialistas, surgindo assim a necessidade de existir outras formas de realizar diagnósticos.

O presente trabalho foi desenvolvido por meio de um mapeamento sistemático que é um tipo de pesquisa que se baseia em evidências da literatura onde é possível argumentar sobre novas hipóteses. Normalmente é usado para investigar situações na área médica, permitindo classificar os resultados dos dados coletados sobre determinado assunto, mostrando insights para pesquisas futuras, evidenciando possíveis lacunas (ROCHA, 2018).

3. METODOLOGIA

Mediante o cenário de inúmeras publicações disponíveis na literatura relacionado ao uso de ML para realização de diagnósticos, foi possível construir este mapeamento sistemático por meio de algumas etapas para filtragem das publicações e um gráfico para facilitar a compreensão (figura1).

- Etapa I: O ponto de partida do trabalho foi definir as bases de dados a serem utilizadas no mapeamento. Foram selecionadas a *Science Direct*, IEEE (*Institute of Electrical and Electronics Engineers*), e a CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior), por serem bases renomadas, de grande credibilidade e englobam diferentes áreas. O intuito era explorar bases de diversas áreas e grupos de pesquisa do mundo para ter uma visão mais ampla sobre a problemática investigada.
- Etapa II: As palavras chaves utilizadas na expressão de busca dos artigos foram: “Machine Learning”, “Big Data”, e as subáreas buscadas foram “Neurociência”, “Medicina”, “Mental Health”, “Psychology”, “Physiotherapy”, “Dentistry”, “Speech therapy”, “Occupational therapy”, “Neurocomputing”, estas palavras foram buscadas de forma combinada como por exemplo “Machine learning” OR “Big Data” AND “Psychology”, como também foi definido o período de publicações, que foram filtradas entre 2015 e 2020. Tais termos de busca foram extraídos de acordo com as áreas e temáticas afins ao tema da revisão sistemática que estava sendo planejado.
- Etapa III: Foi analisado o título, resumo e palavra-chaves de cada artigo mediante alguns critérios de aceitação como: (i) o artigo precisaria falar sobre Data Science e Saúde, (ii) ser uma publicação do tipo artigo (não postagem, notícia, etc.), (iii) o artigo precisa apresentar resultados sólidos, (iv) deve estar escrito em inglês ou português, e (v) as palavras-chaves precisam ser compatíveis com as buscadas.
- Etapa IV: Foi realizada uma leitura mais detalhada do resumo elencando as razões de exclusão dos artigos que não se encaixam na pesquisa, como não ter objetivo claro, não falar sobre diagnóstico, não falar sobre aprendizado de máquina e saúde, não comentar sobre avaliação dos modelos utilizados.
- Na etapa V: Foi realizada uma leitura otimizada mediante introdução, metodologia e conclusão para entender se o corpo do artigo estava conforme o esperado.

- E na etapa VI: Foram feitos gráficos (figura 2, figura 3, figura 4) para melhor compreender a quantidade de artigos por área e facilitar a visualização.



Figura 1. Etapas da metodologia

4. DESENVOLVIMENTO

Em cada etapa foi feita uma filtragem dos artigos, como descrito no tópico anterior, que resultou nos seguintes quantitativos: Nas etapas I e II, no somatório de todas as áreas, foram encontrados inicialmente 1.205 artigos, na *Science Direct*, 513 na IEEEE, e 232 na CAPES. Já na etapa III: serviram apenas 326 artigos na *Science Direct*, 254 na IEEEE, e 55 na CAPES. Na etapa IV obtivemos 46 artigos na *Science Direct*, 38 na IEEEE, e 7 na CAPES. E na última etapa de filtragem, restaram 28 na *Science Direct*, 8 na IEEEE, e 1 na CAPES.

Para uma visão mais clara da quantidade de estudos disponíveis nas bases de dados usadas neste artigo, é possível entender que as publicações sobre ML e diagnósticos em Saúde se intensificaram no ano 2020 obtendo 0,11% de publicações na *Science Direct* (figura 2), na IEEEE o percentual é de 0,05% também em 2020 (figura 3), e na CAPES temos 0,01% no ano de 2019 (figura 4). Percebemos que a base de dados *Science Direct* reúne um volume maior de publicações dentro do tema deste artigo.

De acordo com a presente pesquisa, foi possível perceber que grande parte dos estudos foram publicados nos anos de 2019 e 2020, os países que mais publicaram foram China, Japão, e Brasil, a sub área da Saúde mais estudada foi a Medicina, seguida da Neurociência e Saúde Mental e as patologias mais exploradas foram o Alzheimer, Câncer de Próstata, Câncer de Mama, Esquizofrenia, Parkinson.

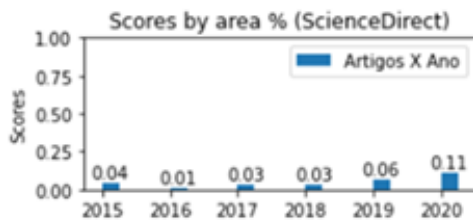


Figura 2. Percentual de artigos (Science Direct)



Figura 3. Percentual de artigos (IEEE)

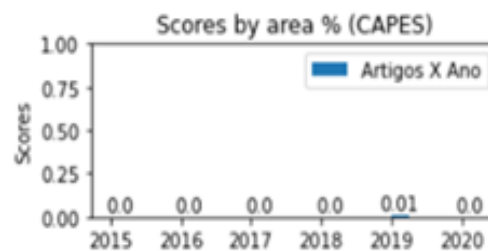


Figura 4. Percentual de artigos (CAPES)

Os gráficos acima foram construídos na ferramenta *Google Colaboratory*, usando a biblioteca *Pandas Numpy*, na linguagem Python.

A patologia com mais publicações foi o Alzheimer, em vários trabalhos foram usadas metodologias diferentes voltadas para processamento e classificação de imagens para diagnóstico, o que torna bastante interessante, pois mostra que diferentes técnicas estão sendo testadas para melhores resultados e mais avanços.

As técnicas de *Machine Learning* podem ser agrupadas em três diferentes categorias: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizado por reforço. Cada grupo possui características próprias a depender da natureza dos dados a serem aprendidos, tarefas e problemas a serem resolvidos.

Normalmente, os modelos de aprendizagem supervisionada são mais usados nas decisões de Saúde, por utilizarem conjuntos de dados etiquetados

por especialistas, onde os algoritmos irão prever ou classificar eventos futuros ou para determinar variáveis mais relevantes (LOBO, 2018).

Segundo Schneider (2016) a classificação de dados, é um modelo muito comum e bastante utilizado quando tratamos de ML, que tem como objetivo categorizar novos eventos que irão somar aos existentes em bancos de dados e resultarão em novos comportamentos, novos padrões.

5. CONSIDERAÇÕES FINAIS

A partir deste mapeamento realizado, foi possível entender que a área da Medicina obtém maior número de publicações e outras áreas publicaram menos ou não publicaram dentro do tema deste artigo, podendo está relacionado ao pouco conhecimento sobre a área de ML e as suas contribuições.

Este estudo foi bastante desafiador diante da vasta quantidade de artigos selecionados nas etapas iniciais e posteriormente na filtragem dos mesmos, pois foram revisados por apenas um colaborador. Porém, irá contribuir para nortear e motivar estudos futuros, proporcionando uma melhor visão de como está o atual cenário das pesquisas na área da Saúde que utilizam ML para diagnóstico médico. Outra contribuição deste trabalho é que ele está explorando uma temática muito rica e que muito ainda tem a ser feito.

6. REFERENCIAS

GRAHAM, S. A; et al. **Artificial Intelligence Approaches to Predicting and Detecting Cognitive Decline in Older Adults: A Conceptual Review.** Journal Pre-proof, 58, 2019.

International Business Machines Corporation (IBM). **Machine Learning.** IBM, 2020. Disponível em: (<https://www.ibm.com/br-pt/cloud/learn/machine-learning>). Acesso em: (24 de Março de 2022).

LIU, X; et al. **A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis.** Science Direct, 1, e271-e297, October, 2019.

LOBO LC. **Inteligência artificial, o Futuro da Medicina e a Educação Médica.** Revista Brasileira de Educação médica, 42 (3), 3-8, Setembro, 2018.

MARTINEZ-MURCIA, F. J; GÓRRIZ J. M; RAMÍREZ, J; ORTIZ A. **Convolutional neural networks for neuroimaging in parkinson's disease: Is preprocessing needed?**, Internacional Journal of Neural Systems, 0, 0, 1-18, July, 2018.

MOTTA GS. **A Mobilidade e a Hiperconexão como Tecnologias de Vigilância na Sociedade de Controle.** ENPAG, 1-13, 2008.

NURMAINI S, et al. **An Automated ECG Beat Classification System Using Deep Neural Networks with an Unsupervised Feature Extraction Technique.** Appl. Sci, 9 (14), 2921, 2019.

RASHEED K, et al. **Machine Learning for Predicting Epileptic Seizures Using EEG Signals: A Review.** IEEE, 14, 139 – 155, July, 2020.

ROCHA F. G, NASCIMENTO B. A. R, NASCIMENTO E. F. V. C. **Um modelo de mapeamento sistemático para a educação.** Cadernos da Fucamp. 17, 29, 1-6, 2018.

SCHNEIDER, P. H. **Análise preditiva de Churn com ênfase em técnicas de Machine Learning: Uma Revisão.** 2016. 82 - Fundação Getúlio Vargas, Escola de Matemática Aplicada. Rio de Janeiro, 2016.

SOCIEDADE BRASILEIRA DE COMPUTAÇÃO. **Machine Learning aplicado à Saúde.** São Paulo, 2019. 42 p.

SUN, M-L; et al. **Application of Machine Learning to Stomatology: A Comprehensive Review.** IEEE, 8, 184360 – 184374, October, 2020.

TULLOCH, J; ZAMANI, R; AKRAMI, A.M. **Machine Learning in the Prevention, Diagnosis and Management of Diabetic Foot Ulcers: A Systematic Review.** IEEE, 8, 198977 – 199000, November, 2020.