

# Machine Learning to Assist in Pneumonia Decision Making: A Systematic Review of the Literature

V. Monteiro Silva<sup>1</sup>, A. Days Ramos Novo<sup>1</sup>, D. Yluska de Souza<sup>1</sup>, A. Sandro da Cunha Rêgo<sup>2</sup>

Programa de Pós Graduação em Tecnologia da Informação – IFPB - Campus JPA  
João Pessoa – PB – Brazil

monteiro.victor@academico.ifpb.edu.br, days.amanda@academico.ifpb.edu.br, damires@ifpb.edu.br,  
alex@ifpb.edu.br

**Abstract.** Clinical decision support systems is a research area in which Machine Learning (ML) techniques can be applied. Nevertheless, specifically in assisting pneumonia decision making, the use of ML has not been so expressive. To help matters, this work aims to contribute to the evolution of the intersection of such areas by presenting a Systematic Review of the Literature. It provides results which may help to identify, interpret and evaluate how ML techniques have been applied and some research enhancements yet to be done.

CCS Concepts: • **Applied computing;**

Keywords: Data Mining, Machine Learning, Pneumonia, Systematic Review

## 1. INTRODUÇÃO

Pneumonia é uma categoria de infecção respiratória que pode atingir pessoas de todas as idades, especialmente crianças e idosos [Respira 2019]. Por ser uma das mais comuns infecções que geram necessidade de internação, casos dessa doença são considerados problemáticos, demandando uma atenção médica especial aos pacientes para evitar complicações, mortalidade e alta prematura durante o tratamento do paciente [Rozenbauma et al. 2015].

A coleta de informações a partir do prontuário do paciente é uma rotina diária de profissionais de saúde para auxiliar no tratamento mais efetivo para pneumonia, assim como para qualquer outra enfermidade. O trabalho consiste em coletar e analisar fontes diversas de informação tais como sinais vitais, históricos familiares, anotações médicas, imagens de raio-x, dentre outras. Nesse contexto, uma parte significativa do treinamento médico é dedicada a aprender como identificar as informações relevantes desse montante, de modo que elas possam guiar a decisão sobre o tratamento [Bezemer et al. 2019]. Sistemas computacionais de suporte à decisão clínica (SAD) vêm sendo amplamente utilizados para auxiliar os profissionais de saúde nessa análise [Horng et al. 2017] [Porat et al. 2016].

A extração de conhecimento a partir do acervo de dados de pacientes pode trazer informações e padrões úteis de modo a agregar valor à tomada de decisão clínica no tratamento de pneumonia. Tendo em vista a inerente complexidade e diversidade em relação a dados e protocolos associados a tratamentos de pneumonia, a aplicação de técnicas de Mineração de Dados (MD) e de Aprendizado de Máquina (AM) junto aos SAD têm sido, cada vez mais, investigadas e utilizadas. É nesse cenário que ocorre esta pesquisa. A presente Revisão Sistemática da Literatura (RSL) objetiva apresentar um panorama acerca do estado da arte da aplicação de MD e AM com vistas à assistência à tomada de decisões em tratamentos de pneumonia. Busca-se assim mapear as evidências metodológicas que

---

Copyright©2020 Permission to copy without fee all or part of the material printed in KDMiLe is granted provided that the copies are not made or distributed for commercial advantage, and that notice is given that copying is by permission of the Sociedade Brasileira de Computação.

concernem à convergência dessas áreas e prover a identificação de oportunidades de aprimoramentos de pesquisas nesse tema.

Este artigo está estruturado da seguinte forma: Na seção 2 são introduzidos alguns conceitos utilizados neste trabalho; na Seção 3 são discutidos os trabalhos relacionados; na Seção 4 é apresentada a metodologia utilizada na condução da RSL; Na Seção 5 é feita uma análise dos resultados obtidos e, na Seção 6, são explanadas as considerações finais acerca do trabalho.

## 2. FUNDAMENTAÇÃO TEÓRICA

Os Registros Médicos Eletrônicos (RME) são informações coletadas e armazenadas diariamente por diversos sistemas hospitalares. Eles incluem resultados de testes, diagnósticos, anotações médicas e medições de sinais vitais. A organização desses dados varia, podendo assumir valores numéricos, imagens de raios-x, textos livres de anotações médicas, entre outros [Ford et al. 2016]. Os dados utilizados nos RMEs podem ser classificados como dados estruturados (DE), não estruturados (DNE) e semi-estruturados (DSE). DE são dados que possuem esquemas de metadados bem definidos, como é o caso de bancos de dados relacionais [Santana 2019]. DNE não apresentam uma estrutura rígida devido à sua dinamicidade e DSE possuem uma estrutura intermediária entre os dois tipos anteriores. Parte considerável do trabalho dos profissionais de saúde é encontrar informação significativa nessa grande quantidade de dados para guiar os tratamentos.

Para facilitar a identificação de padrões, o processo *Knowledge Discovery in Databases* (KDD) dispõe da etapa de MD, que utiliza-se de técnicas computacionais para encontrar correlações, anomalias e padrões em grandes conjuntos de dados. O KDD sistematiza um conjunto de etapas que visa a descoberta de conhecimento potencialmente útil e previamente desconhecido a partir de um conjunto de dados, exibindo o conhecimento de maneira intuitiva. Além da etapa de MD, outras etapas norteiam o processo com atividades de seleção, pré-processamento, transformação, interpretação e avaliação dos resultados [Fayyad et al. 1996].

A MD é fundamentada tecnicamente no aprendizado de máquina (AM), subárea da Inteligência Artificial (IA) que lida com métodos computacionais e estatísticos para adquirir conhecimento de maneira automática [Mitchell 1997]. O AM é realizado a partir de um conjunto de dados cuidadosamente preparado e pode ser efetivado de modo supervisionado (AS) ou não supervisionado (ANS) [Monard and Baranauskas 2003]. O AS é baseado em um conjunto de observações para os quais as saídas desejadas são conhecidas ou em algum outro tipo de informação que represente o comportamento que deve ser apresentado ao sistema [de Souza Gomes 2019]. Exemplos de algoritmos comumente utilizados em AS incluem Redes Neurais Artificiais (RNA), *Support Vector Machine* (SVM), árvores de decisão, *Naive Bayes*, *K-Nearest Neighbors*, entre outros. O ANS, por sua vez, é baseado apenas nas observações do conjunto de dados de entrada cujos rótulos são desconhecidos, produzindo modelos que geralmente se destinam a explorar, agrupar ou identificar relacionamentos em comum entre suas entradas [Monard and Baranauskas 2003]. São exemplos de algoritmos aplicados no contexto de ANS: *K-means*, *K-medoids* e *Apriori*. Uma categoria do AM que vem sendo também utilizada na área de Saúde é a chamada Aprendizado Profundo (AP). O AP possui múltiplas camadas de processamento para extração de características e transformação de modelos de aprendizado. Seus algoritmos são mais complexos e possuem a capacidade de abstrair modelos de aprendizado mais representativos. São exemplos de algoritmos de AP: *Recurrent Deep Neural Networks* e *Convolutional Feedforward Deep Neural Networks*.

## 3. TRABALHOS RELACIONADOS

A busca por trabalhos relacionados a esta RSL objetivou identificar trabalhos secundários associados ao uso de AM ou MD no auxílio à tomada de decisão em quadros de pneumonia. Alguns destes são brevemente descritos a seguir.

[Khan et al. 2020] apresentam uma RSL sobre a identificação de pneumonia a partir de imagens de raios-x utilizando AP. Como resultado, os autores evidenciam uma listagem das principais técnicas de deep learning encontradas e uma análise da qualidade e usabilidade de cada uma delas. Também são discutidos os datasets disponíveis e formas de balanceamento de dados utilizados.

O trabalho de [Chumbita et al. 2020] apresenta um estudo sobre o uso da IA para apoiar decisões clínicas no diagnóstico de pneumonia, utilizando imagens de raio-x. Observou-se que a maioria dos trabalhos empregou redes neurais com altas taxas de precisão nas previsões.

Uma abordagem baseada em modelos de classificação para prever readmissão em 30 dias após a alta hospitalar foi proposta por [Ben-Assuli and Padman 2018]. Os modelos foram treinados com dados dos pacientes e analisados sob a perspectiva de gráficos ROC (*Receiver Operating Characteristic*), tendo o modelo de Árvore de Decisão se destacado dentre os algoritmos analisados.

[Naydenova et al. 2016] apresentam uma abordagem de MD para previsão de pneumonia infantil utilizando de sinais vitais quantificáveis como atributos. Os experimentos demonstraram que o algoritmo Random Forests obteve melhor destaque na predição com métrica da AUC (*Area Under ROC Curve*).

Até onde foi possível identificar, não foram encontradas RSLs ou outros trabalhos secundários que apresentam panoramas da área de AM e MD aplicadas à assistência em quadros de pneumonia, com foco na análise de dados estruturados ou semi-estruturados.

#### 4. METODOLOGIA

Essa RSL foi desenvolvida obedecendo às orientações, políticas e procedimentos estabelecidos por [Kitchenham 2004] e [Dybå and Dingsøy 2008]. O processo é composto pelas seguintes etapas: (i) planejamento da RSL; (ii) condução da RSL, e (iii) resultados da RSL. Nesta seção, são mostrados os passos seguidos nas etapas (i) e (ii).

##### 4.1 Planejamento da Revisão

A principal questão de pesquisa (QP) a ser tratada neste trabalho é: como a MD e o AM estão sendo aplicados no apoio à tomada de decisão clínica em quadros de pneumonia? Essa questão foi decomposta em questões específicas de pesquisa, conforme descrição seguinte:

**QP1:** Quais categorias de assistência à tomada de decisão em quadros de pneumonia são alvo das pesquisas?

**QP2:** Quais tarefas de MD foram identificadas para as categorias de assistência?

**QP3:** Quais modelos de AM encontrados para cada tarefa?

**QP4:** Quais métricas utilizadas para avaliação dos modelos?

Utilizando palavras-chaves relacionadas às questões de pesquisa, a estratégia de busca foi definida a partir de uma string genérica: ("machine learning" OR "data mining" OR "deep learning") AND ("pneumonia"). As bases de dados escolhidas para as pesquisas foram: ACM Digital Library, IEEE, Springer, PubMed, Science Direct e CAPES. As fontes escolhidas estão entre as principais bases relacionadas à QP que disponibilizaram listagens indexadas dos artigos para análise.

Os critérios de inclusão (CI) usados como filtros foram: (i) estudos que apresentem a aplicação de técnicas de MD em tratamento ou prevenção de pneumonia e (ii) estudos que respondem pelo menos uma questão específica de pesquisa. Os critérios de exclusão (CE) utilizados foram: (i) estudos que não provêm relevância científica; (ii) artigos publicados antes de 2010; (iii) estudos secundários ou terciários; (iv) estudos sem experimentação ou avaliação e (v) estudos que não consideram dados estruturados ou semi-estruturados.

Foram adicionalmente definidos critérios de qualidade (CQ) a serem aplicados aos trabalhos selecionados, a saber: (i) definição clara dos objetivos do estudo; (ii) relevância em relação à utilização de técnicas de ML para suporte à decisão clínica; (iii) utilização de métodos de coleta de dados; (iv) apresentação de resultado coerentes com os objetivos e (v) fundamentação teórica sobre os tópicos do estudo. Para cada CQ foi atribuída uma escala de valores: 0, quando o critério é ausente ou não aplicado; 1, quando é parcialmente atendido e 2, quando é totalmente atendido.

## 4.2 Condução da Revisão

A condução da revisão, ou seja, a execução do protocolo planejado, ocorreu em dois meses. A seleção dos trabalhos foi realizada em três fases. Iniciada a partir da seleção de estudos pela string geral de busca, a fase 1 trouxe 563 resultados. Em seguida foi realizado a leitura do título e do resumo dos artigos e, foram excluídas as ocorrências que não respondiam pelo menos uma questão de pesquisa. Na fase 2, foi realizada a leitura das seções de introdução e conclusão dos artigos, resultando em 241 trabalhos selecionados. Esta fase também eliminou os estudos que não estavam alinhados com os objetivos desta RSL. Por fim, os artigos submetidos à terceira fase foram integralmente lidos e analisados, resultando em 34 estudos. Esses trabalhos foram avaliados, quantificados com uma nota conforme os critérios de qualidade definidos e respectivos valores e utilizados para a extração das informações esperadas e definidas no protocolo.

## 5. RESULTADOS

Os 34 artigos selecionados foram avaliados conforme os CQ. Desses, 19 trabalhos atingiram a nota máxima com 10 pontos. A Tabela I mostra os quantitativos dos trabalhos selecionados por fase, separados por fonte de busca. Os resultados particularizados por Questão de Pesquisa são descritos logo adiante.

Table I. Análise Quantitativa.

Fonte	Fase 1	Fase 2	Fase 3
ACM Digital Library	8	8	2
IEEE Xplore	42	29	5
Springer	59	55	15
PubMed	250	70	2
Science Direct	92	33	8
CAPES	112	46	2
<b>TOTAL</b>	<b>563</b>	<b>241</b>	<b>34</b>

### 5.1 Quais categorias de assistência à tomada de decisão em quadros de pneumonia são alvo das pesquisas?

As categorias de acordo com as principais contribuições identificadas como resultado do planejamento da RSL foram:

- **C1.Deteção e Diagnóstico.** Trabalhos com foco em detecção e diagnóstico de pneumonia se concentram na aplicação de técnicas de AM que utilizam dados clínicos dos pacientes para prever a ocorrência da doença;
- **C2.Mortalidade e Complicações.** Essa categoria se refere aos trabalhos que tinham como objetivo sinalizar os pacientes indicativos de risco de falecimento ou de complicações que possam conduzi-los à admissão na UTI ou à necessidade de ventilação mecânica;

- C3. Atributos e Biomarcadores.** Referente a trabalhos exploram a análise de atributos e biomarcadores, com a finalidade de identificar dados clínicos que têm forte correlação com o diagnóstico, tratamento ou complicações do quadro dos pacientes com pneumonia;
- C4. Tempo de Internação e Readmissão.** Trabalhos enquadrados nesta categoria têm como objetivo avaliar o tempo de internação necessário do paciente ou prever um possível risco de readmissão hospitalar.

Os dados apresentados na tabela II expõem o quantitativo das categorias de assistência mais recorrentes .

Table II. Quantitativo de trabalhos por Categoria.

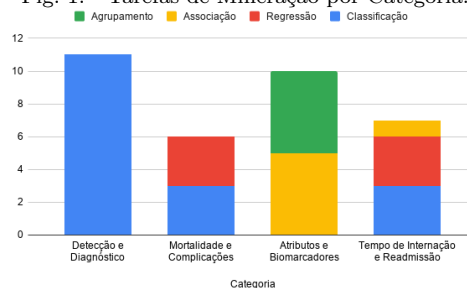
Categoria	Quantitativo
Detecção e Diagnóstico	11
Mortalidade e Complicações	6
Atributos e Biomarcadores	10
Tempo de Internação e Readmissão	7
<b>TOTAL</b>	<b>34</b>

Pode-se observar que a maioria dos trabalhos selecionados (32.4%) tem como alvo a realização de estudos acerca da Detecção e Diagnóstico de Pneumonia ou de patologias relacionadas. Categorias com menor representatividade são relacionadas a problemas de tempo de hospitalização e risco de readmissão, juntamente com a identificação de mortalidade e complicações nos pacientes já diagnosticados. Uma lacuna encontrada no estado da arte foi na estimativa numérica para o tempo de internação, aonde dentre diversas condições de pacientes, requerem diferentes estudos específicos para melhor assertividade. Ainda, na categoria C1, destacam-se em número os trabalhos no diagnóstico em crianças e idosos, duas faixas etárias com maior incidência de casos de pneumonia. Exemplos de trabalhos que se destacam com maior nota nos CQ neste quesito incluem [Naydenova et al. 2015], [DeLisle et al. 2013] e [Liao et al. 2020]. Outros trabalhos que se destacaram nas demais categorias com maior nota nos índices de qualidade foram: [Baechle et al. 2017], [Wu et al. 2014] e [Caruana et al. 2015].

## 5.2 Quais tarefas de MD foram identificadas?

O gráfico ilustrado na Figura 1 sumariza a parcela de ocorrência das tarefas de MD por categoria indicada na Subseção 5.1.

Fig. 1. Tarefas de Mineração por Categoria.



É perceptível na Figura 1 que os trabalhos com foco no diagnóstico preventivo de pneumonia utilizaram apenas tarefas de classificação. Conforme comentado pelos autores, a identificação da patologia no início do tratamento é um fator essencial para um bom prognóstico e por isso a tarefa de classificação se destaca entre as demais [DeLisle et al. 2013] [Ge et al. 2019]. Para trabalhos focados nas

categorias de Mortalidade e Complicações e Tempo de Internação e Readmissão, existe uma divisão dos resultados entre Tarefas de Classificação, em que os autores focam apenas na sinalização desses riscos à equipe médica [Wu et al. 2014][Shimizu et al. 2019][Lai et al. 2018] e Regressão, para estimar a probabilidade de tais riscos ocorrerem [Caruana et al. 2015][Villiers et al. 2018]. No que se refere à identificação de Atributos ou Biomarcadores que influenciam no tratamento de pneumonia, autores utilizaram unicamente técnicas de Agrupamento e Regras de Associação, incluindo entre os dados analisados *scores* médicos e comparando ocorrências de atributos com Diagnósticos [Baechle et al. 2017] [Lin et al. 2010] [Ubaid et al. 2010].

### 5.3 Quais os modelos de AM encontrados para cada tarefa?

Identificadas as tarefas de MD mais recorrentes nos resultados, foram destacados os modelos de AM mais utilizados em cada uma das tarefas. De acordo com a Tabela III, a tarefa de classificação tem o destaque de utilização do algoritmo SVM na maioria dos trabalhos. Os algoritmos LR e RF também ficaram entre os três modelos mais utilizados. O consenso entre os trabalhos com maior nota nos CQ que utilizaram esses algoritmos é de que eles são comumente aplicados e na área de saúde e apresentam bom desempenho na tarefa de classificação, a exemplo do SVM em [Caruana et al. 2015][Lai et al. 2018][Wu et al. 2014]. Na tarefa de Regressão, os modelos RF, LR e AD se destacam dentre os demais modelos pelos mesmos motivos citados [Caruana et al. 2015] [Chmielewska 2016].

Table III. Utilização dos modelos por Tarefa.

Modelos	Classificação	Regressão	Associação	Agrupamento
SVM	12	1	0	0
RNA	5	1	0	0
RF	7	2	0	0
LR	7	2	0	0
AS	0	1	3	4
AD	3	2	0	1
NB	3	0	0	0
<b>TOTAL</b>	<b>37</b>	<b>9</b>	<b>3</b>	<b>5</b>
Legenda: SMV - Support Vector Machine; RNA - Rede Neural Artificial; RF - Random Forests; LR - Logistic Regression; AS - Análise Estatística; AD - Árvore de Decisão; NB - Naive Bayes;				

Nas Tarefas de Associação e Agrupamento, utilizadas para identificar atributos e biomarcadores, os modelos de análise estatísticas tiveram forte predominância, ressaltando lacunas no estado da arte aonde aplicação de algoritmos de AM focados nessas tarefas, poderiam apresentar resultados melhores. [Chen et al. 2016][Ubaid et al. 2010]. Outros modelos como *k-nearest neighbors*, algoritmo genético, Apriori, *k-means* e AP, de modo geral, também foram observados nos trabalhos, entretanto, todos tiveram uma representatividade baixa, com uma ou duas ocorrências, em comparação com os demais.

### 5.4 Quais as métricas mais utilizadas para avaliação dos modelos?

As principais métricas de avaliação dos modelos encontrados na análise dos resultados são destacadas na Tabela IV. Na maioria dos resultados, a plotagem da curva ROC foi o método gráfico mais empregado para avaliar o desempenho dos modelos de classificação.

Considerando que a análise da curva ROC é baseada na Sensitividade e Especificidade, representando a relação bidimensional entre os casos verdadeiros positivos e falso positivos, justifica-se sua utilização pela maioria dos modelos considerando que erros na previsão desses casos causam maior impacto os pacientes, representando doentes, corretamente classificados como doentes [Caruana et al. 2015] [Naydenova et al. 2015] [Lai et al. 2018] [Ge et al. 2019]. Acurácia geral e Score Kappa obtiveram a menor representatividade nos trabalhos, resultado justificável visto que ambas as medidas focam na

Table IV. Principais Métricas de avaliação por Modelo.

Modelos	ROC	ACC	SEN	ESP	KAPP
SVM	10	8	9	8	2
RNA	6	4	4	3	1
RF	9	5	6	6	2
LR	9	3	4	4	0
AS	0	0	0	0	0
AD	5	4	3	3	3
NB	3	2	2	2	2
<b>TOTAL</b>	<b>42</b>	<b>26</b>	<b>28</b>	<b>26</b>	<b>10</b>
Legenda: ROC - Curva ROC; ACC - Acurácia Geral; SEN - Sensitividade; ESP - Especificidade; KAPP - Score Kappa;					

performance genérica dos modelos, enquanto os modelos utilizados na área médica se concentram em errar o menos possível na classe de interesse.

## 6. CONSIDERAÇÕES FINAIS

Nesta pesquisa foi apresentado o planejamento, a condução e os resultados obtidos de uma RSL sobre a utilização de AM e MD para auxiliar na tomada de decisão clínica no tratamento de pneumonia. Ao final de cada etapa de seleção, foi realizada a revisão dos estudos aceitos garantindo assim a confiabilidade e reprodutibilidade da revisão. Foi feita uma categorização dos trabalhos encontrados, destacando diferentes aplicabilidades de pesquisas no tratamento de pneumonia. O trabalho também apresentou uma análise quantitativa destacando tarefas de MD e modelos de AM que foram mais utilizados, juntamente com as métricas de desempenho mais recorrentes para cada algoritmo. Observa-se, a partir dos trabalhos analisados e resultados, que a aplicação de AM e MD possui um alto potencial para fornecer conhecimento e valor importantes e úteis para assistência à tomada de decisão em quadros de pneumonia. A grande diversidade de condições clínicas de pacientes, cada um com cenários distintos de tratamento efetivo, mostra que essa é uma área a ser investigada e aprofundada. Uma lacuna encontrada inclui a falta de trabalhos para estativa de tempo de internação recomendado. Verificou-se também que muitos dos trabalhos avaliados possuem datasets com dados de poucos pacientes, devido a falta de registros médicos eletrônicos ou a restrição pela sensibilidade da informação, o que foi destacado pelos autores como limitações nos seus estudos. As evidências obtidas neste trabalho servirão de norte para a evolução das pesquisas neste contexto.

## REFERENCES

- BAECHLE, C., AGARWAL, A., AND ZHU, X. Big data driven co-occurring evidence discovery in chronic obstructive pulmonary disease patients. *Journal of Big Data* vol. 4, pp. 9, 03, 2017.
- BEN-ASSULI, O. AND PADMAN, R. Analysing repeated hospital readmissions using data mining techniques. *Health Systems* 7 (2): 120–134, 2018.
- BEZEMER, T., DE GROOT, M., BLASSE, E., TEN BERG, M., KAPPEN, T. H., BREDENOORD, A. L., VAN SOLINGE, W. W., HOEFER, I. E., AND HAITJEMA, S. A Human(e) Factor in Clinical Decision Support Systems. *Journal of Medical Internet Research* 21 (3): e11732, 2019.
- CARUANA, R., LOU, Y., GEHRKE, J., KOCH, P., STURM, M., AND ELHADAD, N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. KDD '15. Association for Computing Machinery, New York, NY, USA, pp. 1721–1730, 2015.
- CHEN, C., SHI, L., LI, Y., WANG, X., AND YANG, S. Disease-specific dynamic biomarkers selected by integrating inflammatory mediators with clinical informatics in ards patients with severe pneumonia. *Cell Biol Toxicol*, 2016.
- CHMIELEWSKA, M. Clostridium difficile infection due to pneumonia treatment: Mortality risk models. in: *Pokorski M. (eds) Pathobiology of Pulmonary Disorders*, 2016.
- CHUMBITA, M., CILLÓNIZ, C., PUERTA-ALCALDE, P., MORENO-GARCÍA, E., SANJUAN, G., GARCIA-POUTON, N., SORIANO, A., TORRES, A., AND GARCIA-VIDAL, C. Can artificial intelligence improve the management of pneumonia. *Journal of Clinical Medicine* 9 (1): 248, Jan, 2020.

- DE SOUZA GOMES, E. A. Implementando um Mural Eletrônico em PHP: Uma Aplicação Voltada a uma Instituição de Ensino Superior. Aplicabilidade de Algoritmos de Aprendizado de Máquina para Detecção de Intrusão e Análise de Anomalias de Rede, 2019. Dissertação (especialização) — UFMG, Brasília, DF.
- DELISLE, S., BERNARD, K., DEEPAK, J., AND SIDDIQUI, T. Using the electronic medical record to identify community-acquired pneumonia: toward a replicable automated strategy. *PLoS One* vol. 8, 08, 2013.
- DYBÅ, T. AND DINGSØYR, T. Empirical studies of agile software development: A systematic review. *Information and Software Technology* vol. 50, pp. 833–859, 08, 2008.
- FAYYAD, U., PIATETSKY-SHAPIO, G., AND SMYTH, P. From data mining to knowledge discovery in databases. *AI magazine* 17 (3): 37, 1996.
- FORD, E., CARROLL, J., SMITH, H., SCOTT, D., AND CASSELL, J. Extracting information from the text of electronic medical records to improve case detection: A systematic review. *Journal of the American Medical Informatics Association* vol. 23, pp. ocv180, 02, 2016.
- GE, Y., WANG, Q., WANG, L., WU, H., PENG, C., WANG, J., XU, Y., XIONG, G., ZHANG, Y., AND YI, Y. Predicting post-stroke pneumonia using deep neural network approaches. *International Journal of Medical Informatics* vol. 132, pp. 103986, 2019.
- HORNG, S., SONTAG, D., HALPERN, Y., JERNITE, Y., SHAPIRO, N., AND NATHANSON, L. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. Vol. 12, 2017.
- KHAN, W., ZAKI, N., AND ALI, L. Intelligent pneumonia identification from chest x-rays: A systematic literature review. *medRxiv* 12 (5): p–p, 2020.
- KITCHENHAM, B. *Procedures for Performing Systematic Reviews*, 2004.
- LAI, H., CHAN, P., LIN, H., CHEN, Y., LIN, C., AND HSU, J. A web-based decision support system for predicting readmission of pneumonia patients after discharge. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. pp. 2305–2310, 2018.
- LIAO, Y.-H., SHIH, C.-H., ABBOD, M., AND SHIEH, J.-S. Development of an e-nose system using machine learning methods to predict ventilator-associated pneumonia. *Microsyst Technol* 2020, 03, 2020.
- LIN, W.-T., WANG, S.-T., CHIANG, T.-C., SHI, Y.-X., AND YU CHEN, W. Abnormal diagnosis of emergency department triage explored with data mining technology: An emergency department at a medical center in taiwan taken as an example. *Expert Systems with Applications* 37 (4): 2733 – 2741, 2010.
- MITCHELL, T. *Machine Learning*. McGraw-Hill International Editions. McGraw-Hill, 1997.
- MONARD, M. C. AND BARANAUSKAS, J. A. Conceitos sobre aprendizado de máquina. In *Sistemas Inteligentes Fundamentos e Aplicações*, 1 ed. Manole Ltda, Barueri-SP, pp. 89–114, 2003.
- NAYDENOVA, E., TSANAS, A., CASALS-PASCUAL, C., AND DE VOS, M. Smart diagnostic algorithms for automated detection of childhood pneumonia in resource-constrained settings. In *2015 IEEE Global Humanitarian Technology Conference (GHTC)*. pp. 377–384, 2015.
- NAYDENOVA, E., TSANAS, A., HOWIE, S., CASALS-PASCUAL, C., AND DE VOS, M. The power of data mining in diagnosis of childhood pneumonia. *Journal of The Royal Society Interface* vol. 13, pp. 20160266, 07, 2016.
- PORAT, T., KOSTOPOULOU, O., WOOLLEY, A., AND DELANEY, B. C. Eliciting user decision requirements for designing computerized diagnostic support for family physicians. Vol. 10, 2016.
- RESPIRA, A. A pneumonia. <http://www.respira.pt/content/docs/pneumonia.pdf>, 2019.
- ROZENBAUMA, M., MANGENC, M.-J., HUIJTS, S., VAN DER WERF, T., AND POSTMA, M. J. Incidence, direct costs and duration of hospitalization of patients hospitalized with community acquired pneumonia: A nationwide retrospective claims database analysis. *Vaccine* 3 (28): 3193 – 3199, 2015.
- SANTANA, B. S. *Extração e Aplicação de Indicadores no Processo de Recomendação de Recursos Urbanos Utilizando Dados Estruturados e Não-Estruturados*. M.S. thesis, Universidade Federal do Rio Grande do Sul, <https://lume.ufrgs.br/handle/10183/193897>, 2019.
- SHIMIZU, S., HARA, S., AND FUSHIMI, K. Prs55 predicting the risk of in-hospital mortality in adult community-acquired pneumonia patients with machine learning: A retrospective analysis of routinely collected health data. *Value in Health* vol. 22, pp. S882, 2019. ISPOR Europe 2019.
- UBAID, A., MIRZA, F., BAIG, M., AND MIRZA, F. Identifying the relationship between unstable vital signs and intensive care unit (icu) readmissions. *Expert Systems with Applications* 37 (4): 2733 – 2741, 2010.
- VILLIERS, L., CASPAR, Y., MARCHE, H., BOCCOZ, S., MAURIN, M., MARCHE, P., MORAND, P., MARQUETTE, C., AND CORGIER, B. Resynplex: Respiratory syndrome linked pathogens multiplex detection and characterization. *IRBM* 39 (5): 368 – 375, 2018.
- WU, C., ROSENFELD, R., AND CLERMONT, G. Using data-driven rules to predict mortality in severe community acquired pneumonia. *PLOS ONE* 9 (4): 1–9, 04, 2014.