

# Processamento de Linguagem Natural em Imagiologia: uma revisão sistemática

No Author Given

Universidade do Minho, Braga, Portugal

**Abstract.** O crescimento exponencial da quantidade de informação disponível ao nível da imagem médica vivido nos últimos anos tornou indispensável a criação de mecanismos apoiados na Inteligência Artificial para um mais rápido processamento dos dados, mas também para se conseguir explorar as novas potencialidades que estes acarretam. O presente estudo pretendeu recolher e sistematizar as potencialidades documentadas do Processamento de Linguagem Natural (PLN) na área da Imagiologia, bem como as aplicações registadas e fundamentadas na literatura. Estas variam desde a deteção de erros nos relatórios médicos até à proteção de dados sensíveis dos pacientes em estudos que utilizam grande quantidades de documentos médicos.

**Keywords:** Processamento Linguagem Natural · Imagem Médica · Imagiologia · Inteligência Artificial.

## 1 Introdução

Nos dias de hoje, a Imagiologia tem um papel fundamental na área médica, auxiliando os profissionais médicos a diagnosticar e tratar doenças e lesões. Tradicionalmente, a Imagiologia assentou sempre sobre a produção de diferentes tipos de imagem para diagnóstico. Contudo, a sua versão moderna inclui não só a produção, mas também o processamento, apresentação, armazenamento e transmissão das imagens médicas, constituindo uma ciência em si só.[1]

O Processamento de Linguagem Natural (PLN) é uma área da Inteligência Artificial (IA) responsável por fazer com que os computadores sejam capazes de compreender a linguagem humana escrita. Falar do processamento de linguagem em Imagiologia pode parecer contraditório, mas todos os exames imagiológicos são geralmente acompanhados de um relatório médico, o que abre a porta à procura de informação nestes e a sua correlação com os dados imagiológicos em si.

Esta é uma área vista como em expansão transversalmente, não estando confinada a sua utilidade ao setor clínico e, devido aos seus possíveis benefícios, tem potencial interesse, o que leva a uma quantidade considerável de investigação acerca desta nos mais diversos ramos.

Deste modo, o PLN surge como uma ferramenta promissora para melhorar a eficácia e a Imagiologia, fornecendo um meio de extração de conhecimento a

partir de uma grande quantidade de dados, sejam eles não estruturados ou estruturados, podendo ser auxiliada por outros métodos, tais como o *deep learning*, para realizar estas tarefas.

Será também de referir que o PLN, quando auxiliado por outros métodos, pode ter diversas tarefas, desde facilitação da escrita de relatórios, comunicação rápida com especialistas quando determinados termos chave são detetados ou uniformização da escrita, levando assim a uma standardização dos relatórios e registos clínicos, o que, por sua vez, irá facilitar as futuras análises do PLN na área.

## 2 Métodos

Recorreu-se a diversas bases de dados académicas para a procura de estudos e artigos científicos relacionados com o PLN em Imagiologia. As aplicações encontradas incidiram sobre os mais variados espectros da área, sendo que se apresentam aqueles que se achou terem maior relevância e aplicabilidade noutros ambientes além dos do estudo.

Procurou-se ao máximo evitar a parcialidade de alguns dos artigos investigados, pois alguns foram publicados em revistas e jornais científicos e procuravam o patrocínio de determinadas ferramentas ou *softwares*, sendo que alguns foram até publicados por empresas que pretendiam patrocinar os seus próprios produtos e técnicas.

Teve-se em consideração os resultados demonstrados em cada artigo para categorizar a relevância científica de cada um e os métodos utilizados nos estudos, de forma a perceber se teriam capacidade de replicação em ambientes diferentes daqueles em que foram desenvolvidos.

## 3 Resultados

Apresentam-se em seguida os resultados recolhidos, organizados pelas diferentes aplicabilidades do PLN em Imagiologia.

### 3.1 Detecção de erros

Erros de gramática, ortografia e uso de palavras incorretas são erros comuns em relatórios de imagens médicas, com 23–36% de relatórios finais contendo erros.

Para combater este fenómeno, um grupo do criou um modelo de PLN que deteta automaticamente interseções, exclusões e substituições inadequadas de palavras em relatórios de imagiologia. Para tal, utilizaram 1.192.098 relatórios de TCs cerebrais e radiografias torácicas de três instituições distintas.

No final, o modelo foi capaz de detetar cerca de 90% das frases erradas.[2]

### 3.2 Detecção de quedas

Um estudo realizado nesta categoria foi a identificação de pacientes com HIV que sofreram quedas, através da análise de relatórios de radiologia. Deu-se início a este estudo devido a idosos com HIV terem uma menor densidade óssea que um idoso normal, o que significa que estão em maior risco de desenvolver fraturas em episódios de quedas.

Foram acedidos processos eletrónicos (VACS-VC) e através de uma ferramenta PLN chamada YTEX foram extraídos conceitos UMLS (Unified Medical Language System) com relevância dos relatórios e desenvolveram-se máquinas de vetores de suporte para classificar os relatórios.

Este foi um estudo com bastante sucesso, sendo que para um PPV (Positive Predictive Value) de 92.57% obteve uma sensibilidade de 94.48%. [3]

### 3.3 Acompanhamento da evolução de tomografias computadorizadas

Os relatórios das tomografias computadorizadas (TC), apesar de seguirem orientações, são escritos em texto livre, o que dificulta muito a automação.

Deste modo, um estudo tentou criar um pipeline de PLN para emparelhar medições de relatórios de TCs, de forma a automatizar não só o processo de acompanhamento de um determinado paciente, mas também o *data mining* destes relatórios.

Para tal, utilizaram 283 relatórios de TCs abdominais e 311 TCs torácicas de 50 pacientes.

Os seus resultados foram relativamente bem sucedidos, obtendo uma precisão de 90% e uma sensibilidade de 78%. No entanto, apenas obtiveram 57% de desempenho sem erros. [4]

### 3.4 Correlação entre mamografias e patologias

Este estudo visou investigar a eficiência de PLN em encontrar subtipos de cancro através da procura de características específicas em relatórios de mamografia BI-RADS (Breast Imaging Reporting and Data System).

Um grande problema na recolha de relatórios eletrónicos de saúde é a existência de texto não estruturado o que limita a interpretação e a recolha de dados de exames de mamografia.

Este grupo desenvolveu uma ferramenta baseada em PLN chamada MOTTE, que é capaz de extrair e processar milhares de dados de relatórios de mamografia. Através da ferramenta MOTTE foram extraídos parâmetros, e palavras-chave de 543 relatórios de pacientes BI-RADS categoria 5 com cancro de um total de 717, que foram utilizados no algoritmo.

Em termos de resultados, o algoritmo conseguiu extrair palavras-chave com correlações a vários subtipos de cancro. Os dados obtidos foram bastante próximos daqueles obtidos através da recolha manual. Estes são aspetos positivos sendo que através de PLN o processo de extração de dados pode ser bastante mais rápido e em maior escala. [5]

### 3.5 Identificação de fraturas em ossos longos

A *Pediatric Emergency Care Applied Research Network (PECARN)* desenvolveu um modelo de PLN capaz de identificar fraturas nos ossos longos em emergências pediátricas através de relatórios livremente escritos (não estruturados).

Utilizou-se um conjunto de treino de 1000 relatórios radiológicos e um conjunto de teste de 500, em conjunto com uma *cross-validation* de 10-*folds*, de modo a encontrar a melhor configuração do modelo.

No final, o modelo criado obteve uma exatidão de 95.8%, uma sensibilidade de 96.9% e precisão de 94%, sendo assim possível identificar fraturas em ossos longos a partir de relatórios médicos.[6]

### 3.6 Classificação automática de cancro pulmonar

De forma a automatizar a deteção e classificação de cancro pulmonar, duas instituições americanas desenvolveram um modelo de PLN que capturasse características visuais de nódulos pulmonares em relatórios estruturados de TCs pulmonares.

Para tal, utilizaram exames de pacientes com alto risco de cancro pulmonar, interpretados por radiologistas que produziram os relatórios estruturados. De seguida, utilizaram 3 conjuntos de treino, de 500, 120 e 383 relatórios, seguido de um conjunto de teste de 198 relatórios (total de 1201 relatórios).

No final, o modelo criado obteve valores de sensibilidade de 75% e especificidade de 98%, e a maioria dos erros ocorreram em casos em que existia de facto nódulos pulmonares, mas ainda não eram considerados no modelo de classificação utilizado.

Assim, de uma forma geral, foi possível a classificação automática de cancro pulmonar, quando os relatórios eram produzidos de uma forma estruturada e padronizada.[7]

### 3.7 Simplificação de relatórios de radiologia para o consumidor

Muitas instituições médicas começam, hoje em dia, a utilizar portais para que pacientes possam aceder aos seus relatórios de radiologia mais facilmente. Este estudo teve como objetivo testar o uso de uma terminologia chamada CHV (Consumer Health Vocabulary) para a estandardização e simplificação de relatórios de modo a torná-los acessíveis aos consumidores.

O processo da estandardização dos relatórios é feito usando uma ferramenta chamada MetaMap que ajuda a ligar e a substituir certos termos nos relatórios com aqueles preferidos pela terminologia CHV. Se existirem termos médicos que a terminologia CHV não conseguir traduzir, são então dadas explicações baseadas no UMLS (Unified Medical Language System).

Para avaliar o desempenho deste processo, foi analisada a cobertura geral, e a semelhança entre os termos traduzidos e aqueles que estavam no relatório de radiologia.

Com esses parâmetros em mente, a terminologia CHV teve uma cobertura de 88.5% sendo que dos 792 termos selecionados, 695 foram traduzidos. Para o segundo parâmetro, dos termos traduzidos através do CHV, houve uma semelhança de aproximadamente 72.6% com os termos dos relatórios de radiologia.[8]

### 3.8 Identificação de anomalias acidentais em pacientes que sofreram um trauma

As descobertas de anomalias em pacientes são, muitas das vezes, feitas acidentalmente. Quando um paciente faz um exame de radiologia de rotina, ou quando sofreu um trauma, é possível encontrar anomalias (congénitas ou não) reportadas nos relatórios acerca de outro assunto, que não está relacionado com o motivo que levou o paciente a realizar o exame em primeiro lugar. Estas poderiam ficar escondidas e, eventualmente, afetar negativamente o paciente, se não fosse a realização do exame após o trauma.

Para este estudo foram obtidos 170 052 relatórios de radiologia. Através do uso de PLN foi possível simplificar os relatórios de modo a obter classificadores de diferentes modalidades.

Com o uso de métodos *deep learning* e classificadores mais tradicionais foi criado um algoritmo que consegue avaliar os relatórios de radiologia. Coleções de terminologia médica, como a SNOMED, foram usadas para poder encontrar possíveis identificadores que auxiliassem nesta avaliação.

Dos relatórios originais, foram obtidos 4,181 relatórios que melhor se adequavam ao que se estava à procura. Desses 4,181 relatórios, apenas 10.5% acabaram por revelar uma anomalia acidental.

Neste estudo, foi concluído que o uso da terminologia SNOMED não teve um efeito significativo no desempenho dos algoritmos de *machine learning*. Pelo contrário, o uso de CNNs (Convolutional Neural Networks) melhorou o sistema de avaliação, com valores de precisão e sensibilidade mais elevados.[9]

### 3.9 Detecção da presença de pneumonia em crianças através de relatórios de radiologia

A pneumonia é bastante prevalente na população infantil, sendo a maior causa de hospitalização de crianças nos Estados Unidos. Sistemas de avaliação de dados e classificadores de doenças ineficientes foram a principal motivação por detrás do seguinte estudo.

Foram usados algoritmos de *machine learning* e ferramentas PLN como MEDLee, na extração de conceitos e na identificação de relatórios de exames efetuados à região do peito. Em auxílio ao estudo foi também usado um repositório de relatórios de crianças, o PHIS+.

No desenvolvimento da aplicação PLN, foram usados 282 relatórios para treinar e testar o modelo. Para o treino do modelo foram criados relatórios com os conceitos e identificadores relevantes ao estudo. Para testar a capacidade do modelo, e ver se era capaz de classificar relatórios corretamente, foram extraídos

alguns dos identificadores mais comuns na pneumonia dos relatórios a ser testados.

Para concluir, este modelo de classificação teve resultados positivos, com uma sensibilidade de 71% e uma especificidade de 96.2%. Este teve um melhor desempenho do que avaliação efetuada manualmente por especialistas do domínio, que obtiveram uma sensibilidade de apenas 52.7%.[10]

### 3.10 Desambiguação de abreviaturas em relatórios médicos com recurso a CNN

O recurso a abreviaturas e siglas no preenchimento de relatórios médicos é uma prática extensiva. De forma a que os algoritmos de *machine learning* consigam ser eficazes, é necessário decifrar aquilo a que cada abreviatura faz referência, já que conceitos diferentes podem ter a mesma abreviação (e.g. *myasthenia gravis* e miligrama). O processo de decifração passa por conseguir interpretar o sentido da palavra, baseado no contexto envolvente. Isto implica, como é óbvio, a existência de um inventário dos possíveis sentidos de cada palavra.

Tradicionalmente, algoritmos de Support Vector Machine (SVM) ou de *machine learning* eram usados para esta tarefa. Contudo, este tipo de abordagem envolve a necessidade de realizar *feature engineering* (divisão dos dados em grupos que possam ser aplicados aos algoritmos de aprendizagem com supervisão - *supervised learning*). Este é um processo dispendioso, o que levou à procura de soluções que não necessitem deste passo intermédio.

O estudo defende assim a utilização de *deep learning* Convolutional Neural Networks (CNN) para realizar o processo de desambiguação. Esta estratégia revelou ainda ser sempre mais eficaz do que os algoritmos de SVM, rondando os 1 a 4 pontos percentuais de melhoria.

Contudo, estes modelos de aprendizagem sem supervisão exigem uma grande quantidade de dados para o seu treino. Assim, um dos problemas enfrentados no teste destes novos modelos é o baixo número de dados disponíveis. Surge então a necessidade de criar novos dados, pela geração automática de abreviaturas ou siglas em relatórios médicos reais. A estratégia passa por um processo de engenharia inversa, no qual os documentos são percorridos em busca de possíveis sentidos de uma abreviatura, previamente agrupados num repositório de abreviaturas e os seus possíveis sentidos. As formas extensas são então substituídas pela respetivas siglas e o contexto em que surgem é associado àquele sentido. Os novos documentos são então usados para testar e treinar o modelo. Um dos problemas reportados com esta estratégia é o não equilíbrio na ocorrência dos diferentes possíveis sentidos nos documentos. Por exemplo, novamente para a sigla "mg", o seu surgimento nos novos documentos deve-se maioritariamente à presença da palavra "miligrama", e não da palavra *myasthenia gravis*. O estudo concluiu ainda que este problema pode ser reduzido recorrendo a técnicas de *clustering* aplicados aos possíveis sentidos e à anotação manual de outros possíveis sentidos quando a abreviatura apenas tem um possível sentido associado no repositório.[10]

### 3.11 *De-identification* e proteção de privacidade

Quando se trabalha com relatórios médicos para a extração de dados, tem de se ter sempre muita atenção para não extrair informação pessoal ou sensível dos pacientes. O processo de remoção destas secções de informação sensível é chamado de *de-identification*, ou anonimização em português.

Os métodos existentes de *machine learning*, expressões regulares e baseados em dicionários não são aplicáveis para os relatórios médicos, pois os dados surgem, na maior parte das vezes, de forma desestruturada e as palavras não estão todas catalogadas.

Um estudo realizado sobre 113 relatórios cardiológicos escritos em alemão refere que a utilização de um algoritmo de *deep learning* baseado em *Long Short-Term Networks* (LSTM) bidirecional conseguiu melhorar o processo de anonimização de 70% para 93% na escala de pontuação F2.[11]

## 4 Discussão

### 4.1 Aplicações clínicas de PLN em imagiologia

Pode-se assim observar a importância de PLN na área da Imagiologia médica, sendo que com o auxílio da inteligência artificial esta torna-se uma ferramenta altamente eficaz. O poder de processamento das máquinas atuais é capaz de auxiliar o trabalho de profissionais da área, sistematizando tarefas que de outro modo demorariam muito tempo a ser concluídas.

Como foi verificado, as aplicações de PLN acabam espalhadas por diversas áreas devido à sua flexibilidade, sendo que nomeadamente na área clínica é comum que estas tomem funções mais informativas, ou classificação de doenças do que a funções clínicas. Isto pode ser devido ao facto de que modelos que não usam *deep learning*, que são os mais utilizados devido à sua transparência, acabam por ser menos eficazes e, como tal, têm funções mais restritas.

Outro problema são os *datasets* utilizados, nomeadamente o facto de serem por vezes pequenos ou de poderem conter informação que não esteja equilibrada. Isto, principalmente nos modelos de *deep learning*, complica a sua utilização. Se se pensar nos comuns relatórios médicos, quer sejam de consultas ou exames realizados, muitos costumam conter linguagem corrente ou até com incoerências gramaticais, pois os profissionais de saúde trabalham geralmente sobre pressão. Além disso, cada profissional desenvolve geralmente o seu próprio método de escrita de relatório, não havendo portanto um formato ou estilo comum que facilite a aplicação dos algoritmos tradicionais.

### 4.2 Utilização e conhecimento do léxico médico

Também deve ser referido que descobertas que envolvam estes modelos e que utilizem o léxico médico ainda têm números reduzidos, sendo que será importante alinhar as arquiteturas descritas com este léxico para ajudar os modelos a ter um melhor desempenho mesmo com *datasets* mais reduzidos, sendo de especial

importância o conhecimento de como se expressa a doença, a área afetada e os aspetos da patologia em questão.

Outra barreira presente nesta área é a variabilidade dos termos aplicados, incluindo sintaxe, posicionamento de observações e por fim, a utilização de diversas formas diferentes de fazer a mesma observação. Isto obriga a um maior treino dos modelos forçando-os possivelmente assim a necessitar de um treino muito mais detalhado.

### 4.3 A questão do idioma

Como o PLN trabalha sobre o texto escrito, tem de ter em conta o idioma sobre o qual está a trabalhar, ao contrário de outros métodos de *machine learning* que trabalham sobre dados numéricos ou gráficos. Assim, todas as nuances abordadas até agora são posteriormente afetadas e limitadas pela barreira linguística, colocando entraves à escalabilidade das suas aplicações. Esta questão leva geralmente à necessidade da intervenção de especialistas na área, ou mesmo à total transformação dos modelos criados.

### 4.4 Parâmetros de avaliação e sua definição

Outro fator importante a referir, é a heterogeneidade dos estudos em questão. Quando avaliamos humanos, é geralmente fácil fazê-lo através da análise de anotações. No entanto, quando tentamos aplicar os mesmos parâmetros a estes tipos de modelos depara-mo-nos com outros tipos de problemas. Nomeadamente, a necessidade de ser capaz de explicar os resultados obtidos por modelos, evitando assim o fenómeno de *black boxes*. A necessidade de assegurar a transparência dos modelos é assim um dos grandes desafios nesta área. A falta de validação externa e a falta de *datasets*, por muitos destes serem de instituições, e como tal, privados, constitui outro problema. Por um lado o treino destes modelos é extremamente complexo e precisa de grandes quantidades de dados. Por outro, por razões éticas, morais e de segurança, muitos destes *datasets* não podem ser utilizados pelo público, podendo levar a cenários de *overfitting* do modelo, ou simplesmente a uma dificuldade elevada de conseguir replicar os resultados dos mesmos.

### 4.5 Possíveis limitações da *review*

Como muitos outros estudos semelhantes, onde o objetivo é concatenar os estudos existentes na área, existirão problemas na seleção de quais os que deverão ser abordados. Em primeiro lugar, será altamente possível que as *key words* utilizadas na pesquisa não tenham encontrado estudos que poderiam ser relevantes. Por outro lado, também será possível que alguns estudos tenham sido descartados na fase de síntese devido às parcialidades inerentes a todos os processos de pesquisa humanos.



#### 4.6 Considerações finais e progresso de PLN na área médica

De uma forma geral, pode-se concluir que o PLN quando aplicado à imagiologia pode suportar áreas diversas, desde a diminuição da carga horária de médicos, ajudar na tomada de decisões, adicionar valor a relatórios com a sintetização da informação relevante, assim como ajudar a encontrar diagnósticos perdidos durante a triagem. No entanto, pelo motivos referidos anteriormente, estes modelos são extremamente ineficientes no que toca a doenças raras. Poderá ser por estes motivos que nos últimos anos se tem verificado um aumento do uso deste tipo de modelos para extrair informação de relatórios ou classificar doenças, mas não necessariamente em aplicações clínicas.

Como identificado por Casey et al., alguns pontos a ter em conta para estudos futuros seriam um aumento da transparência do que é reportado e das descobertas em questão, incluindo informação relevante ao *dataset* utilizado, como o seu tipo, os dados que foram usados para teste e para treino, as características das anotações assim como a metodologia usada para o seu desenvolvimento e as métricas de desempenho dos modelos, que deveriam ser tão abrangentes quanto possível. Assegurar uma comparação significativa entre modelos tendo em conta diversos métodos diferentes, o uso de *datasets* padrão para permitir a comparação de modelos com os de outros estudos e, por fim, os *datasets* utilizados, assim como o código em questão, deveriam ser disponíveis publicamente. No entanto, será de referir que por motivos de segurança e privacidade, a publicação dos *datasets* utilizados poderá ser inviável. [13]

## 5 Bibliografia

### References

1. Kunio, D. (2006). Diagnostic imaging over the last 50 years: research and development in medical imaging science and technology. *Physics in Medicine and Biology*, 51(R5).
2. Zech, J. R., Forde, J. O., Titano, J. J., Kaji, D., Costa, A., Oermann, E. K. (2019). Detecting insertion, substitution, and deletion errors in radiology reports using neural sequence-to-sequence models. *Annals of Translational Medicine*, 7(11), 233. <https://doi.org/10.21037/atm.2018.08.11>
3. Bates, J., Fodeh, S. J., Brandt, C. A., Womack, J. A. (2016). Classification of radiology reports for falls in an HIV study cohort. *Journal of the American Medical Informatics Association*, 23(e1), e113-e117. <https://doi.org/10.1093/jamia/ocv155>
4. Sevenster, M., Bozeman, J., Cowhy, A., Trost, W. (2015). A natural language processing pipeline for pairing measurements uniquely across free-text CT reports. *Journal of Biomedical Informatics*, 53, 36–48. <https://doi.org/10.1016/j.jbi.2014.08.015>
5. Patel, T., Puppala, M., Ogunti, R., Ensor, J., He, T., Shewale, J. B., Ankerst, D. P., Kaklamani, V. G., Rodriguez, A., Wong, S. T. C., Chang, J. C. (2017). Correlating mammographic and pathologic findings in clinical decision support using natural language processing and data mining methods. *Cancer*, 123(1), 114–121. <https://doi.org/10.1002/cncr.30245>

6. Grundmeier, R. W., Masino, A. J., Casper, T. C., Dean, J. M., Bell, J. F., Enríquez, R. G., Deakynne, S. J., Chamberlain, J., Alpern, E. R. (2016). Identification of Long Bone Fractures in Radiology Reports Using Natural Language Processing to support Healthcare Quality Improvement. *Applied Clinical Informatics*, 07(04), 1051–1068. <https://doi.org/10.4338/aci-2016-08-ra-0129>
7. Beyer, S., McKee, B., Regis, S., McKee, A. B., Flacke, S., Saadawi, G. M. E., Wald, C. (2017). Automatic Lung-RADSTM classification with a natural language processing system. *Journal of Thoracic Disease*, 9(9), 3114–3122. <https://doi.org/10.21037/jtd.2017.08.13>
8. Qenam, B., Kim, T. Y., Carroll, M. J., Hogarth, M. (2017). Text Simplification Using Consumer Health Vocabulary to Generate Patient-Centered Radiology Reporting: Translation and Evaluation. *Journal of Medical Internet Research*, 19(12), e417. <https://doi.org/10.2196/jmir.8536>
9. Trivedi, G., Hong, C., Dadashzadeh, E. R., Handzel, R., Visweswaran, S. (2019). Identifying incidental findings from radiology reports of trauma patients: An evaluation of automated feature representation methods. *International Journal of Medical Informatics*, 129, 81–87. <https://doi.org/10.1016/j.ijmedinf.2019.05.021>
10. Meystre, S. M., Gouripeddi, R., Tieder, J. S., Simmons, J. M., Srivastava, R., Shah, S. S. (2017). Enhancing Comparative Effectiveness Research With Automated Pediatric Pneumonia Detection in a Multi-Institutional Clinical Repository: A PHIS+ Pilot Study. *Journal of Medical Internet Research*, 19(5), e162. <https://doi.org/10.2196/jmir.6887>
11. Richter-Pechanski P, Amr A, Katus HA, Dieterich C. Deep Learning Approaches Outperform Conventional Strategies in De-Identification of German Medical Reports. *Stud Health Technol Inform*. 2019 Sep 3;267:101-109. doi: 10.3233/SHTI190813. PMID: 31483261.
12. Joopudi, V., Dandala, B., Devarakonda, M. V. (2018). A convolutional route to abbreviation disambiguation in clinical text. *Journal of Biomedical Informatics*, 86, 71–78. <https://doi.org/10.1016/j.jbi.2018.07.025>
13. Casey, A., Davidson, E., Poon, M. T. C., Dong, H., Duma, D., Grivas, A., Grover, C., Suárez-Paniagua, V., Tobin, R., Whiteley, W., Wu, H., Alex, B. (2021). A systematic review of natural language processing applied to radiology reports. *BMC Medical Informatics and Decision Making*, 21(1). <https://doi.org/10.1186/s12911-021-01533-7>