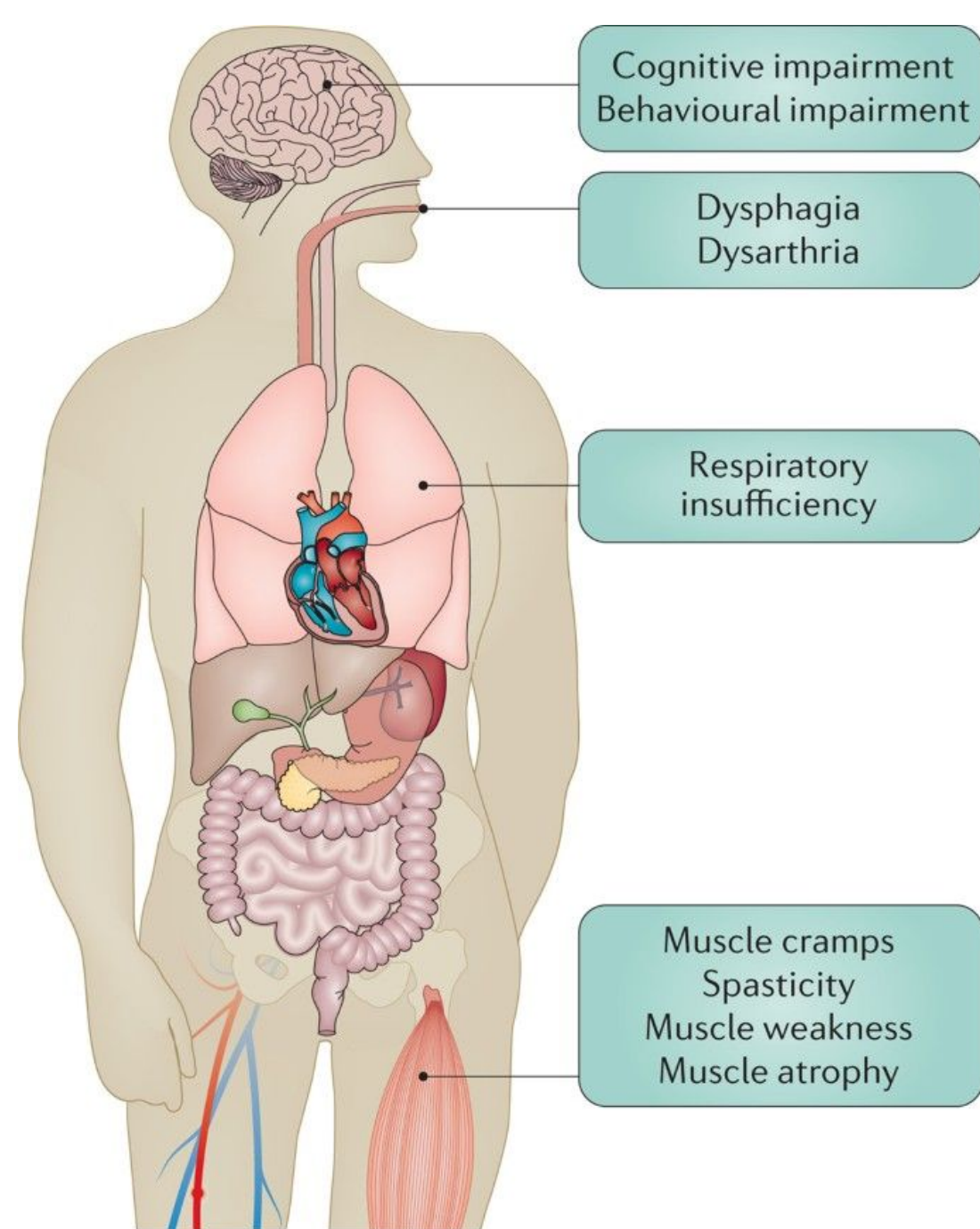


Tema:

Modelo automatizado para mapeamento de bibliografia existente sobre bases de dados de Esclerose Lateral Amiotrófica

Introdução



Nature Reviews | Disease Primers

A Esclerose Lateral Amiotrófica (ELA) é uma doença neurodegenerativa rara que afeta os neurônios motores, com origem genética em cerca de 15% dos casos, enquanto a maioria permanece de causa desconhecida. Diversos estudos atualmente buscam realizar o prognóstico de pacientes, no entanto, os avanços na área são restringidos por pequenas quantidades de dados

Fonte: Hardiman et al. (2017)

relativos à doença e pela falta de caracterizações mais completas. Por essa razão, existem argumentações na comunidade científica à favor da criação de uma base de dados global, centralizando diferentes modalidades de dados obtidos ao longo de anos de pesquisa.

Motivação e Objetivo

A partir de trabalhos prévios utilizando imagens em ressonância magnética de pacientes, foram detectadas dificuldades como a escassez de dados e a falta de informações centralizadas sobre bases existentes, resultando em limitações significativas. Diante desse cenário, identificou-se a importância de criar uma plataforma que reúna informações sobre os datasets disponíveis para ELA, especificando o número de participantes e os tipos de dados fornecidos (clínicos, imagens médicas e/ou coletas genéticas). Assim, o objetivo desta nova etapa do estudo é automatizar a coleta de informações em artigos acadêmicos relacionados a bases de dados de ELA, utilizando técnicas de Processamento de Linguagem Natural (NLP).

Integrantes: Fernanda Namie Takemoto Furukita

Professor(a) Orientador(a): Prof. Dr. Edson Satoshi Gomi

Metodologia

O trabalho foi estruturado em cinco etapas principais:

- 1. Coleta de textos acadêmicos:** Reunir artigos relacionados a bases de dados de Esclerose Lateral Amiotrófica.
- 2. Pré-processamento dos textos:** Foi realizada a extração dos resumos dos artigos e a simplificação das informações para adequação aos modelos.
- 3. Aplicação de modelos de NLP:** Três modelos foram utilizados para identificar os tipos de dados abordados nos textos. Para validação das informações coletadas, foi implementada a técnica de Triple Modular Redundancy.
- 4. Disponibilização dos resultados por API:** Os resultados processados pelos modelos foram integrados em uma API.
- 5. Desenvolvimento do Front-End:** Para facilitar o acesso aos resultados, foi desenvolvido um front-end utilizando a framework React. As etapas de análise de texto e integração com a API foram realizadas em Python, enquanto a interface foi construída com React.



Resultados

Como resultado da coleta automatizada, 980 artigos dentro da temática foram encontrados, dos quais 666 foram baixados. Dentre estes, cujos textos passaram pelo pré-processamento e pela análise por modelos de NLP, obteve-se aprovação do votador para datasets coletados, tipos de dados e quantidade de dados respectivamente para 412, 405 e 479 artigos.

