

VladmiR: Uma ferramenta para diagnóstico de Alzheimer a partir de miRNA circulante utilizando aprendizado de máquina.

Jean Rodrigues¹; Andrei Lucas Padilha Pereira², João Gabriel Moreira de Souza³ ; Frederico Schimit Kremer⁴

¹*Universidade Federal de Pelotas – eujean@live.com*

²*Universidade Federal de Pelotas – andreilsp2pp@gmail.com*

³*Universidade Federal de Pelotas – joaomoreirasouza99@gmail.com*

⁴*Universidade Federal de Pelotas – fred.s.kremer@gmail.com*

1. INTRODUÇÃO

A Doença do Alzheimer foi identificada pela primeira vez em 1906 pelo psiquiatra Alois Alzheimer. É uma doença neurodegenerativa crônica que ainda não tem tratamento preventivo ou para retardar os efeitos na neuroplasticidade do cérebro e no armazenamento da memória.

O Alzheimer (DA) leva a perda dos neurônios e das sinapses no córtex cerebral, provocando a degeneração no lobo temporal, no lobo parietal e em partes do lobo frontal (WENK, 2003). Isso acontece devido ao mau enovelamento das proteínas beta amiloide e tau.

Essas proteínas são usadas como um dos métodos de diagnóstico, mas não é um dos mais eficientes porque também precisa de uma avaliação cognitiva. Mas para tentar achar uma forma mais eficiente de análise e que pudesse usar sangue, foi analisado os micro-RNAs.

Micro-RNAs são estruturas monocatenárias com um tamanho de até 25 nucleotídeos. A sua função é atuar como um RNA não codificante e silenciador pós-transcricional, onde eles se ligam com mRNAs específicos e regulam sua estabilidade e tradução. Eles têm demonstrado seu potencial como biomarcadores não invasivos de sangue e soro para uma ampla variedade de patologias humanas. Uma vez que a desregulação da expressão de miRNA pode estar envolvida em disfunção neurológica ou processos neurodegenerativos [LAIDINGER P. et al, 2013].

A pesquisa promovida por LAIDINGER P. et al deixou as amostras de micro-RNA em aberto no site do NCBI. Então foi pensado num modelo de Machine Learning que pudesse analisar essas estruturas de forma mais precisa e sob diferentes dimensões para ter uma acurácia maior sobre a correlação de ocasionalidade entre a má formação de micro-RNAs e as ocorrências de Alzheimer.

2. METODOLOGIA

O projeto VladmiR teve base na linguagem de programação Python, escolhemos esta por ser o foco do laboratório e por sua complementaridade com modelos de Machine Learning. Iniciando o projeto, buscamos no banco de dados da NCBI por arquivos no formato SRA sobre miRNAs correspondentes a Alzheimer, após o download convertemos com a ferramenta “fastq_to_fasta”, do pacote “SRA-Tools”, para formato fasta.

Com os dados brutos prontos e convertidos, utilizando a biblioteca “jellyfish”, está especializada em leitura de arquivos fasta e multi-fasta contendo sequências



de DNA para organização de k-mers, assim organizamos em k-mers 7, 8 e 9. K-mers devidamente distribuídos, foi montado o dataframe com a biblioteca “pandas” para facilitar a visualização de atributos e amostras.

A parte de processamento dos dados, especialmente o treinamento, foi feita utilizando as bibliotecas “scikit-learn”, para montar os modelos ML, e “seaborn”, para visualização dos resultados. Com scikit-learn escolhemos os modelos RandomForest, ExtraTreesClassifier, K-nearest neighbors(KNN) e Naive Bayes para comparação da capacidade de discriminação entre pacientes e não pacientes.

3. RESULTADOS E DISCUSSÃO

Após o treinamento dos modelos, obtivemos diferentes performances para cada um deles. RandomForest e ExtraTreesClassifier obtiveram 0.94 e 0.88% de acurácia, respectivamente (Figura 1), sendo um valor consideravelmente bom levando em consideração que as métricas para medir a performance de um modelo variam de 0 a 1, assim, discriminando com excelência pacientes onde a doença foi observada dos pacientes que faziam parte do grupo controle. Atualmente o modelo está sendo aplicado utilizando dados com diferentes números de kmers, a fim de identificar se há melhora na performance.

Métricas de performance dos modelos aplicados

Algoritmo	Precisão	Recall	Acurácia
RandomForest	1 1.0	1 0.92	0.94
ExtraTreeClassifier	1 0.92	1 0.92	0.89
K-nearest neighbors	1 0.84	1 0.84	0.77
Naive Bayes	1 0.75	1 0.76	0.61

Figura 1: tabela contendo as notas para diferentes métricas, entre elas: precisão, revocação e acurácia de cada modelo aplicado neste estudo.

Atualmente a investigação e diagnóstico se dá por avaliação de um médico, exame de sangue, exames de imagem e avaliação neuropsicológica, contudo, em grande parte, tal procedimento é realizado quando o paciente em questão já apresenta determinado comprometimento cognitivo leve, um estágio que precede o desenvolvimento da doença [Gandy S. et al, 2011], o que pode dificultar ou até inviabilizar o tratamento uma vez que a doença já apresenta certo nível de severidade. É de suma relevância rever métodos já impostos, que utilizam a análise de biomarcadores como A β e Tau apenas para confirmar o diagnóstico e trazer a atenção para métodos atuais tal qual a abordagem trazida aqui, que visa o diagnóstico antes de surgir sintomas, assim, garantindo o tratamento precoce [FRANKFOR SV, TULNER LR, VAN CAMPEN JP, VERBEEK MM, JANSEN RW, BEIJNEN JH. et al, 2008.].

A problemática, visando modelos preditivos, que precisa ser pautada é a quantidade de dados disponíveis para treinamento e validação de modelos, que



não são suficientes de maneira que garanta eficiência no diagnóstico para pacientes brasileiros (*data drift*). Outro ponto, modelos treinados com dados de pessoas de determinada região do globo, poderão garantir acurácia no diagnóstico no país onde o mesmo foi desenvolvido, justamente por conta das peculiaridades genômicas da própria população. Sendo essas, as limitações desta abordagem.

4. CONCLUSÕES

A abordagem trazida aqui foi capaz de discriminar, a partir de miRNA circulante, pacientes que apresentaram a doença em questão de pacientes que não. Mas, para além disso, a necessidade de esforços em cima de novos métodos para diagnósticos de Alzheimer, e não apenas, a partir dessa abordagem, acredita-se que seja possível também discriminar outras patologias por meio da biologia molecular, assim, encontrando marcadores moleculares e desenvolvendo tratamentos menos invasivos.

5. REFERÊNCIAS BIBLIOGRÁFICAS

Wenk GL. **Neuropathologic changes in Alzheimer's disease.** J Clin Psychiatry. 2003;64 Suppl 9:7-10. PMID: 12934968.

Leidinger, P., Backes, C., Deutscher, S. *et al.* **A blood based 12-miRNA signature of Alzheimer disease patients.** Genome Biol 14, R78 (2013)

Gandy, S: **Perspective: prevention is better than cure.** Nature 2011, 475:S15

Frankfort SV, Tulner LR, van Campen JP, Verbeek MM, Jansen RW, Beijnen JH: **Amyloid beta protein and tau in cerebrospinal fluid and plasma as biomarkers for dementia:** a review of recent literature. Curr Clin Pharmacol 2008, 3:123-13

Albert Einstein. **Alzheimer.** Disponível em:
<https://www.einstein.br/doencas-sintomas/alzheimer>