

# Avaliação de indicadores sanguíneos para diagnósticos da doença de Parkinson e Parkinsonismos utilizando algoritmos de aprendizado de máquina

LUÍSA RIBEIRO BEZERRA, Universidade Federal de Minas Gerais

A Doença de Parkinson e os Parkinsonismos são síndromes importantes a serem estudadas para obter diagnóstico o mais rápido possível, retardando o aparecimento de mais sintomas, uma vez que não há cura, e sim tratamentos por meio de medicamentos para o alívio dos sintomas da doença. O uso do aprendizado de máquina, neste estudo, é para a identificação de indicadores sanguíneos que possivelmente têm alguma correlação para com estas síndromes.

*Parkinson's disease and Parkinsonisms are important enough to justify further study in order to secure faster diagnosis, delaying the onset of symptoms, since despite the lack of cure, treatments exist to alleviate such symptoms. Machine learning can be used in this context to identify blood indicators related to the disease, predicting its onset before it fully develops*

## KEYWORDS

Doença de Parkinson, Parkinsonismos, aprendizado de máquina, Árvore de Decisão, Naive Bayes Gauciano

## ACM Reference format:

Luísa Ribeiro Bezerra. 2018. Avaliação de indicadores sanguíneos para diagnósticos da doença de Parkinson e Parkinsonismos utilizando algoritmos de aprendizado de máquina. 1, 1, Article 1 (June, 2018), 7 pages.

## 1 INTRODUÇÃO

A doença de Parkinson é considerada clinicamente uma perturbação degenerativa crônica do sistema nervoso central, sendo uma característica bem comum o distúrbio da coordenação motora do paciente, dado pela perda de células cerebrais que produzem a dopamina. Lentamente e gradualmente, os sintomas da doença se manifestam e desenvolvem ao longo do tempo - dificuldade para caminhar, conversar ou realizar outras tarefas simples do dia-a-dia [1]. É uma doença que não há precedentes indicativos da sua causa, porém há uma frente que acredita na possível correlação dela com fatores genéticos e fatores ambientais [2]. Em adendo, não há uma cura para a doença de Parkinson, mas sim tratamentos por meio de medicamentos para o alívio dos sintomas da doença [1][3] - por isso é importante ter o diagnóstico antecipadamente, uma vez que facilita no tratamento da doença, retardando de forma positiva o aparecimento dos sintomas. Outro ponto importante de ser avaliado futuramente por meio do diagnóstico é a separação de Parkinson para Parkinsonismos, síndrome que tem sintomas parecidos com a doença de Parkinson, porém com tratamentos diferentes. Muitas vezes, os Parkinsonismos são tratados como Parkinson, o que acaba possivelmente agravando o estado do paciente, uma vez que o tratamento não é feito de forma adequada ao seu real diagnóstico [4]. Neste trabalho, a abordagem será em cima de indicadores que são visivelmente importantes para a identificação de Parkinson e Parkinsonismos no geral.

Visto a importância deste assunto para a medicina e para a área da saúde humana, é possível fazer um estudo por meio de aprendizado de máquina com objetivo de encontrar novas correlações entre sintomas e características (como variáveis de exames de sangue, sono e influência genética). Por meio dos algoritmos de aprendizado de máquina *Decision Tree Classifier* e *Gaussian Naive Bayes* do scikit-learn - biblioteca de aprendizado de máquina de código aberto para a linguagem de programação Python [5] -, treinamos os modelos juntamente dos dados amostrais obtidos pelo estudo longitudinal do PPMI [6], que avalia a progressão da desordem em um paciente. Neste primeiro momento, conseguir extrair as características que são possíveis indicadores da doença de Parkinson e dos Parkinsonismos baseado nos elementos sanguíneos é o objetivo principal.

---

## 2 FUNDAMENTAÇÃO TEÓRICA

### 2.1 Doença de Parkinson

A doença de Parkinson é uma progressiva desordem neurológica do sistema nervoso central, resultado da perda de células cerebrais que produzem a dopamina. O sintoma mais evidente são os tremores por regiões no corpo, dentre elas as mãos, braços, pernas, mandíbula e rosto, além da rigidez dos membros e tronco [1]. Outros possíveis sintomas são problemas sensoriais, urinários, emocionais, psiquiátricos e a perturbações do sono [2][3]. São mais comuns em pessoas com mais de 60 anos, sendo mais comum entre homens do que em mulheres [2]. Não há exames de sangue ou laboratoriais que comprovadamente auxiliem no diagnóstico da doença de Parkinson - o diagnóstico é dado baseado nos sintomas e em exames neuro imagiológicos [1].

### 2.2 Parkinsonismos

Os Parkinsonismos em geral são síndromes específica caracterizada por um conjunto de sinais e sintomas, como o tremor, hipocinesia, rigidez e instabilidade postural. Pode ser causado por diversas outras doenças, toxinas, doenças metabólicas e outras condições neurológicas. A causa mais comum de parkinsonismo é a condição neurodegenerativa conhecida como doença de Parkinson - porém, os tratamentos para cada tipo de Parkinsonismos são bem específicos [4].

### 2.3 Aprendizado de Máquina

O aprendizado de máquina é um subcampo da Ciência da Computação que aprofundou-se no estudo de reconhecimento de padrões para o aprendizado [7], com algoritmos que podem aprender de seus erros e fazer previsões sobre dados. O modelo construído é com base no tipo de algoritmo, nas amostras de entrada e as decisões para suas previsões são fundamentadas pelos dados amostrais - ou seja, generaliza a entrada a partir de suas experiências.

### 2.4 Decision Tree Classifier

O algoritmo de aprendizado de máquina de classificação *Decision Tree Classifier* utiliza uma árvore de decisão como um modelo preditivo, que classifica a “população” para vários conjuntos com base em algumas propriedades escolhidas desta população. Partindo da raiz da árvore, o algoritmo classifica as instâncias ao chegar em algum nodo folha que fornece a classe da instância [8][9].

### 2.5 Gaussian Naive Bayes

O classificador bayesiano *Gaussian Naive Bayes* é um classificador de natureza supervisionada e probabilístico que assume uma distribuição normal baseado no “Teorema de Bayes”, com uma suposição de independência entre os preditores (variáveis) [10].

---

## 3 TRABALHOS RELACIONADOS

A grande maioria dos trabalhos relacionados com o aprendizado de máquina e a doença de Parkinson e Parkinsonismos está voltada para estudos motores, como o tremor nos pacientes que têm os sintomas comuns nos Parkinsonismos [11][12]. Há também estudos relacionados com a frequência da voz do paciente [13], com análise de imagens de tomografia [14], dentre outros, todos focados em obter uma alta acurácia em relação ao diagnóstico da doença de Parkinson.

## 4 METODOLOGIA

Para esta pesquisa, foram selecionados dois algoritmos de classificação para receberem os dados amostrais e retornarem a acurácia para o diagnóstico da doença de Parkinson: *Decision Tree Classifier* e *Gaussian Naive Bayes*. Ambos os algoritmos foram fornecidos pela biblioteca scikit-learn [5].

Os dados amostrais obtidos pelo estudo longitudinal PPMI avaliam a progressão da desordem nos pacientes [6]. Dos dados coletados, 54 indicadores foram selecionados para a avaliação, com objetivo de extrair indicadores com maior influência no estudo do Parkinson correlacionado ao sangue - São eles:

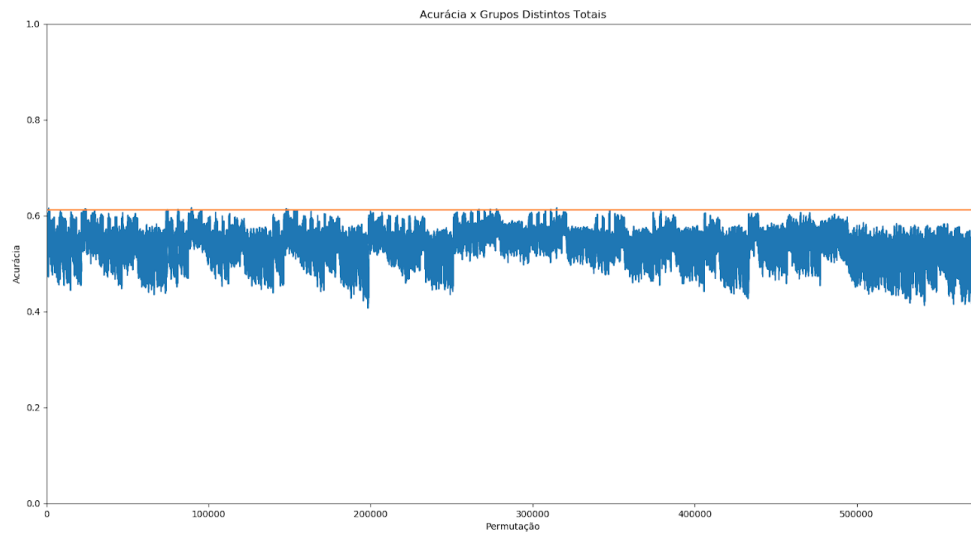
HMT10	HMT20
HMT11	HMT21
HMT12	RCT18
HMT13	HMT40
HMT15	HMT2
HMT16	HMT3
HMT17	HMT7
HMT18	HMT8
HMT19	HMT9
HMT92	HMT68
HMT93	HMT69
HMT95	CGT283
HMT96	AMT7453
RCT1	CGT284
RCT4	RCT392
RCT5	RCT183
RCT6	BAT318
RCT8	RCT1407
RCT11	BIOMOMPD
RCT12	BIODADPD
RCT3580	FULSIBPD/FULSIB
CGT626	HAFSIBPD/HAFSIB
RCT13	MAGPARPD/MAGPAR
RCT15	PAGPARPD/PAGPAR
RCT16	MATAUPD/MATAU
RCT3579	PATAUPD/PATA
RCT17	KIDSPD/KIDSNUM

Para a injeção de dados nos algoritmos, conjuntos de cinco indicadores foram feitos por meio de permutações desses 54 elementos sanguíneos. A ideia proposta, então, para uma boa escolha de indicadores, seria a combinação do resultado dos dois algoritmos, fazendo uma análise dos resultados para garantir possíveis indicadores com uma boa atuação na identificação de Parkinson por meio de exame de sangue - buscando uma minimização do viés dos algoritmos sob os dados, tornando o resultado o mais imparcial possível.

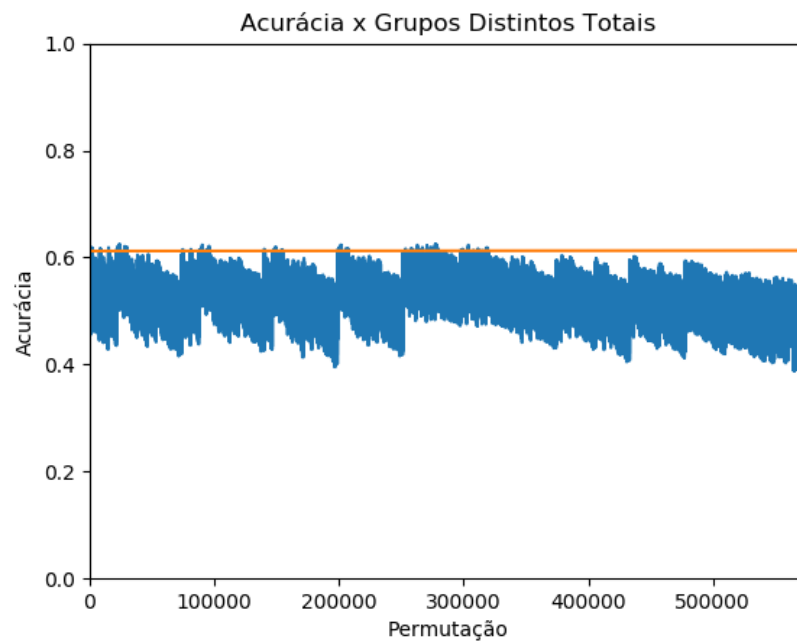
## 5 EXPERIMENTOS E RESULTADOS

O experimento se inicializa por meio da execução do classificador de *Decision Tree Classifier*, obtido pela biblioteca scikit-learn, e após isso do algoritmo de *Gaussian Naive Bayes*, juntamente dos dados obtidos pelo estudo PPMI. Após a execução do algoritmo, foi contabilizado, baseado em todos os conjuntos possíveis de cinco elementos dos 54 indicadores, as acurácias obtidas por meio de ambos os algoritmos. A ideia, aqui, é plotar um

gráfico que nos dê uma acurácia geral obtida por meio da utilização dos modelos. Os valores para cada execução estão ilustrados nos gráficos das Figuras 1 e 2:

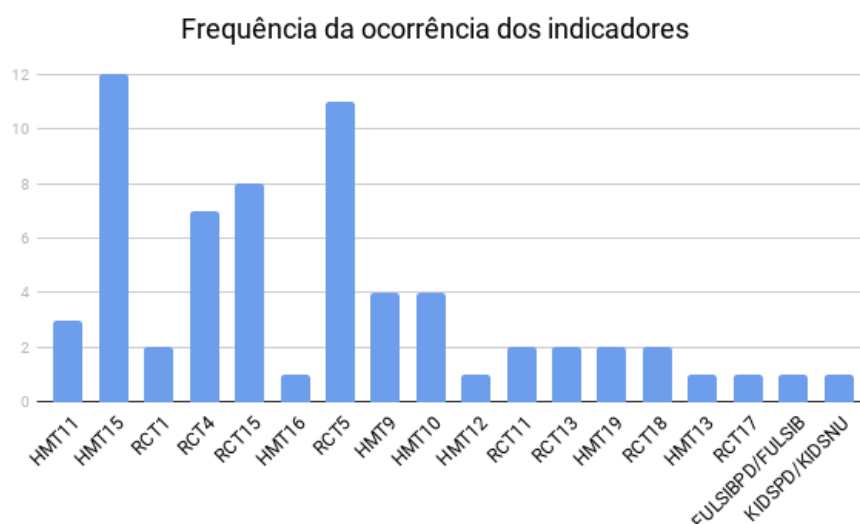


**Fig. 1. Acurácia obtida para todos os grupos distintos baseados no algoritmo Árvore de Decisão**

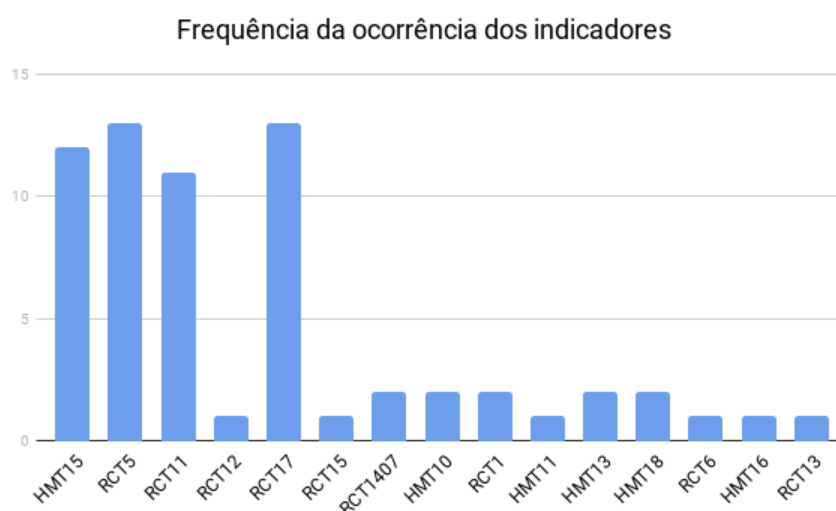


**Fig. 2. Acurácia obtida para todos os grupos distintos baseados no algoritmo *Gaussian Naive Bayes***

Foi então definido um corte para a análise dos melhores indicadores - 61.3% no algoritmo de Árvore de Decisão e 62.2% no algoritmo *Gaussian Naive Bayes* -, selecionando o “Top 13” conjuntos de melhor acurácia. Visto que o corte nos retorna conjunto de cinco elementos sanguíneos, foi extraída a frequência da aparição dos indicadores presentes em todos os conjuntos selecionados. Os dois gráficos abaixo – Figuras 3 e 4 - ilustram a frequência da aparição de cada indicador acima do corte:



**Fig. 3. Frequência da ocorrência dos indicadores baseado no algoritmo Árvore de Decisão**



**Fig. 4. Frequência da ocorrência dos indicadores baseado no algoritmo *Gaussian Naive Bayes***

## 6 CONCLUSÃO

Por meio desta pesquisa, que tende a avaliar indicadores de potenciais importâncias para a identificação da Doença de Parkinson e Parkinsonismos, é possível identificar, em uma base de dados pré-disponibilizada no estudo longitudinal do PPMI, alguns indicadores com maior relevância quando se fala em Parkinson e Parkinsonismos. Foi utilizado dois algoritmos para a validação e minimização do viés dos algoritmos sob os dados, tornando o resultado o mais imparcial possível - para tal, é comparado os indicadores mais frequentes para toda acurácia acima de 60% em ambos os modelos.

É conclusivo que há uma grande ocorrência, segundo o corte feito sob os conjuntos de maior acurácia obtidos através do algoritmo de *Decision Tree Classifier*, dos bioindicadores HMT15 (12 ocorrências), RCT5 (11 ocorrências), RCT4 (7 ocorrências), RCT15 (8 ocorrências), HMT9 (4 ocorrências) e HMT10 (4 ocorrências). Tais indicadores têm a correspondência segundo a Tabela 1.

Código	Nome	Ocorrência
HMT15	Neutrophils (%)	12
RCT5	AST (SGOT)	11
RCT4	ALT (SGPT)	7
RCT15	Serum Sodium	8
HMT9	Lymphocytes	4
HMT10	Monocytes	4

**Tabela 1. Indicadores mais frequentes baseado no algoritmo *Decision Tree Classifier***

Em relação ao algoritmo de *Gaussian Naive Bayes*, foi analisado uma ocorrência similar em alguns pontos, sendo mais expressivos nos indicadores HMT15 (12 ocorrências), RCT5 (13 ocorrências), RCT11 (11 ocorrências) e RCT17 (13 ocorrências). Tais indicadores têm a correspondência segundo a Tabela 2.

Código	Nome	Ocorrência
HMT15	Neutrophils (%)	12
RCT5	AST (SGOT)	13
RCT11	Serum Glucose	11
RCT17	Serum Bicarbonate	13

**Tabela 2. Indicadores mais frequentes baseado no algoritmo *Gaussian Naive Bayes***

Há, visivelmente, uma necessidade da melhoria do algoritmo, tendendo a otimizá-lo, e da escolha dos indicadores para uma melhoria no algoritmo e na sua acurácia, porém é possível captar alguns indicadores sanguíneos que podem ter relação com a Doença de Parkinson e Parkinsonismos.

A seleção de indicadores feita neste estudo é também importante no âmbito social, visto que exames sanguíneos são, normalmente, de alto custo, e reduzir o número de indicadores pode ser uma boa abordagem para a redução do valor - garantindo a oportunidade de exame para uma população maior pela melhoria da

acessibilidade monetária do exame. Além disso, a diminuição no gasto computacional para o estudo sanguíneo dos pacientes é expressiva, uma vez que, tendo um número seletivo de indicadores, é possível analisá-los profundamente com um tempo computacional muito menor, facilitando a acessibilidade por mais máquina e rapidez nos resultados de análise.

## 7 REFERÊNCIAS

- [1] <https://www.ninds.nih.gov/Disorders/All-Disorders/Parkinsons-Disease-Information-Page> NINDS. 30 de junho de 2016. Consultado em 18 de julho de 2016
- [2] [https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(14\)61393-3/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(14)61393-3/fulltext) Kalia, LV; Lang, AE (29 de agosto de 2015). Parkinson's disease. *Lancet*. 386 (9996): 896–912.
- [3] <https://onlinelibrary.wiley.com/doi/abs/10.1111/jnc.13691> Sveinbjornsdottir, S (11 de julho de 2016). The clinical symptoms of Parkinson's disease. *Journal of Neurochemistry*. 139: 318–324.
- [4] Christine, C. W., & Aminoff, M. J. (2004). Clinical differentiation of Parkinsonian syndromes: prognostic and therapeutic relevance. *The American journal of medicine*, 117(6), 412-419
- [5] <http://scikit-learn.org/stable/index.html>
- [6] PPMI: Site. 2018. Disponível em: <<http://www.ppmi-info.org/>>. Acessado em 23 mar. 2018
- [7] <https://global.britannica.com/technology/machine-learning>
- [8] Mitchell, Tom. “Machine Learning”, McGraw-Hill, 1997.
- [9] <http://scikit-learn.org/stable/modules/tree.html>
- [10] [http://scikit-learn.org/stable/modules/naive\\_bayes.html](http://scikit-learn.org/stable/modules/naive_bayes.html)
- [11] <http://publications.lib.chalmers.se/records/fulltext/250423/250423.pdf> Cecilia Hernqvist & Matilda Rosander. Machine learning for symptoms quantification of Parkinson's disease patients. (2017)
- [12] <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0188226> Warwick R. Adams. High-accuracy detection of early Parkinson's Disease using multiple characteristics of finger movement while typing. (30 de novembro de 2017)
- [13] [https://www.researchgate.net/publication/262817615\\_Intelligent\\_Parkinson\\_Disease\\_Prediction\\_Using\\_Machine\\_Learning\\_Algorithms](https://www.researchgate.net/publication/262817615_Intelligent_Parkinson_Disease_Prediction_Using_Machine_Learning_Algorithms) V.S Sriram, Tarigoppula & Venkateswara Rao, M & V Satya Narayana, G & Pandu Ranga Vital, T & Dowluru, Kaladhar SVGK. (2013). Intelligent Parkinson Disease Prediction Using Machine Learning Algorithms. *IJEIT*. 3. 212-215.
- [14] <http://iopscience.iop.org/article/10.1088/1741-2560/12/2/026008> Francisco P M Oliveira and Miguel Castelo-Branco 2015 *J. Neural Eng.* 12 026008. (2015). Computer-aided diagnosis of Parkinson's disease based on [123I]FP-CIT SPECT binding potential images, using the voxels-as-features approach and support vector machines.