

Classificação Foliar de Ervas Medicinais Utilizando Algoritmos de Inteligência Artificial

Kenedy F. S. Silva

Agenda

- Introdução
- Trabalhos Relacionados
- Metodologia
- Base de Dados
- Pré-processamento das Imagens
- Classificadores
- Análise Experimental
- Conclusão
- Referências

Introdução

- Reconhecimento de imagens de plantas
- Identificação de Recursos
- Valor Medicinal
- Utilização de recursos
- Eficiência
- Machine Learning

Trabalhos Relacionados

- Literature review of image features and classifiers used in leaf based plant recognition through image analysis approach
 - Amala Sabu; K. Sreekumar
- Plant Classification Using Artificial Neural Networks
 - Luciano D. S. Pacifico; Valmir Macario; Joao F. L. Oliveira
- Identification of selected medicinal plant leaves using image features and ANN
 - R. Janani; A. Gopal

Metodologia

- Coleta de dados
- Criação da base de dados
- Pré-processamento das imagens
- Scikit-learn
- Validação cruzada
- Classificação das imagens
- Teste de hipótese

Base de Dados

- Coleta de imagens da internet
 - Google Imagens
 - Imagens de ambientes reais
- Total de 540 imagens
 - Recursos que poderão ser aumentados com inclusão de mais plantas.
 - Atualmente temos apenas 13 espécies.

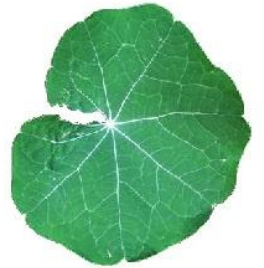
Exemplo de Imagens



Pré-processamento das Imagens

- Separação das folhas
- Binarização
- Remoção do background
- Extração das características
 - Textura
 - Cores
- Criação da base de classificação / pré-processada
 - 13 atributos
 - 7 de textura
 - 3 de cores

Exemplo de imagens



Extração de Características

- Textura (Método GLCM)
 - Contraste
 - Dissimilaridade
 - Homogeneidade
 - Energia
 - Correlação
 - Segundo Momento Angular
 - Entropia
- Cores (Dados calculados para RGB)
 - Média
 - Desvio Padrão

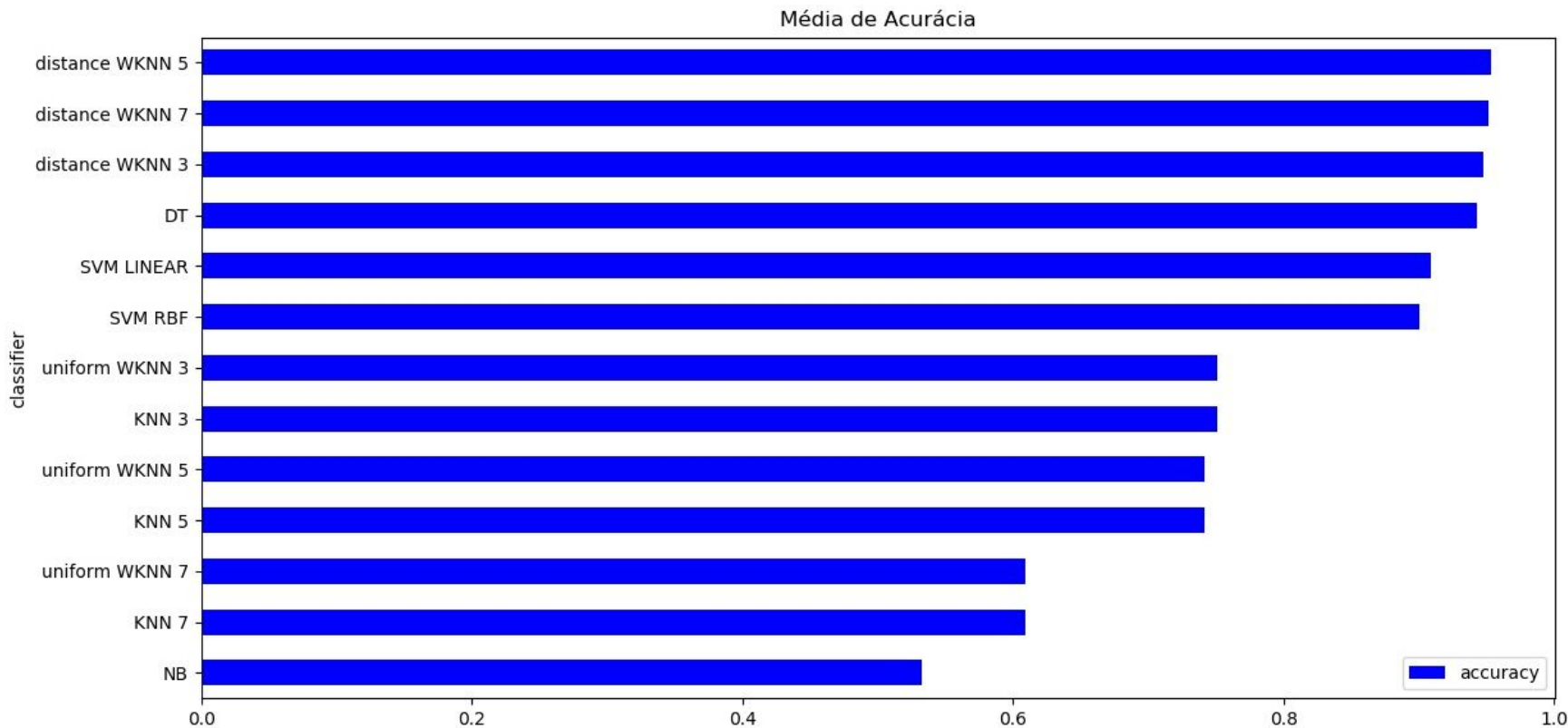
Classificadores

- Árvore Decisão (DT)
- Naive Bayes (NB)
- KNN
 - Variações de 3, 5 e 7
 - Distância Euclidiana
- W-KNN
 - Pesos (Weights) - Uniforme e Distância
 - Variações de 3, 5 e 7
 - Distância Euclidiana
- SVM
 - Linear
 - RBF
- Logistic Regression

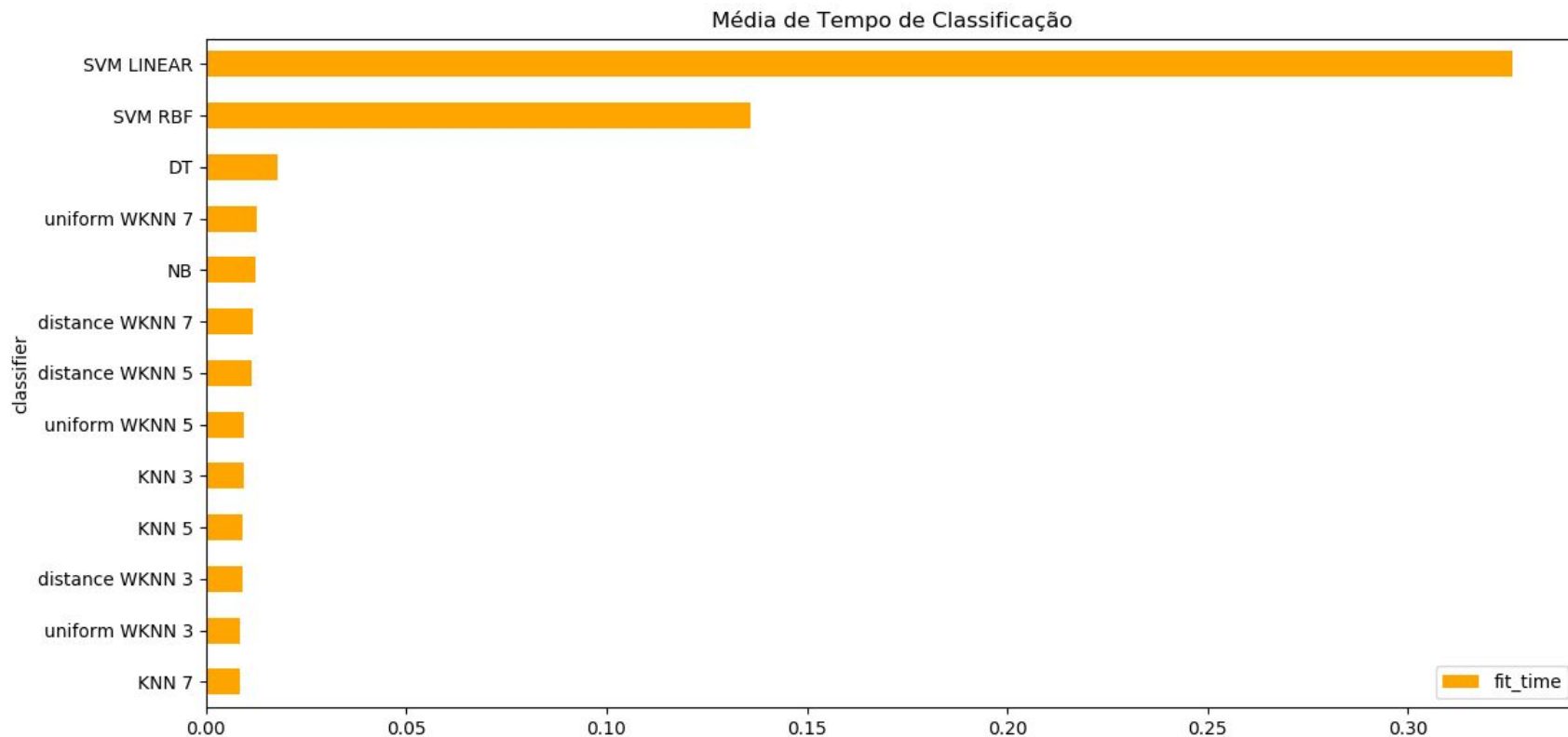
Análise Experimental

- Teste de Hipótese de Correção de Holm-Sidak (Teste post hoc)
 - Método utilizado para neutralizar o problema das comparações múltiplas
 - É um método simples de controlar a taxa de erro da família
 - Sem rejeição
 - Valores $p < 0,01$
- Análise das médias
 - Acurácia
 - Desvio padrão
 - Mediana

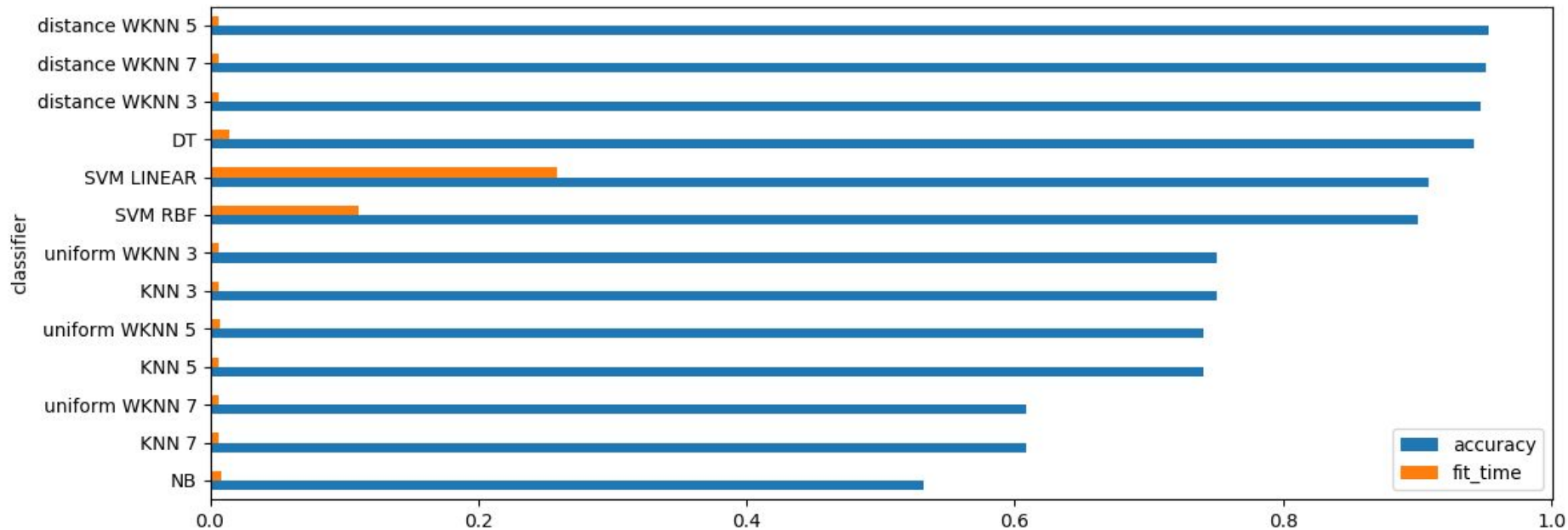
Análise das médias



Tempo de Execução



Médias / Tempo de Execução



Médias Obtidas

Classificador	Acurácia	Desvio Padrão	Mediana
NB	53,40%	0,05586	53,70%
KNN 7	60,70%	0,05461	61,11%
U WKNN 7	60,70%	0,05461	61,11%
KNN 3	74,91%	0,06009	75,70%
U WKNN 3	74,91%	0,06009	75,70%
KNN 5	75,25%	0,05447	75,70%
U WKNN 5	75,25%	0,05447	75,70%
SVM LINEAR	91,25%	0,04536	92,52%
SVM RBF	91,25%	0,04251	91,60%
DT	94,14%	0,03728	94,44%
D WKNN 3	94,36%	0,03334	94,44%
D WKNN 7	94,70%	0,03162	94,44%
D WKNN 5	94,77%	0,03506	94,44%

Conclusão

- Total de testes realizados = 650
- Base nova
- Índice de acerto alto
- Acurácia acima de 60% para todos os algoritmos
- Destaque para o W-KNN ($k = 5$) - 94,44%
- Pouco tempo de processamento

Referências

- [1] A. Sabu and K. Sreekumar, "Literature review of image features and classifiers used in leaf based plant recognition through image analysis approach," in Inventive Communication and Computational Technologies (ICICCT), 2017 International Conference on. IEEE, 2017, pp. 145–149
- [2] Pacifico, Luciano DS, Valmir Macario, and Joao FL Oliveira. "Plant Classification Using Artificial Neural Networks." 2018 International Joint Conference on Neural Networks (IJCNN). IEEE, 2018.
- [3] Cheirsilp, Ronnarit. 3D multimodal image analysis for lung-cancer assessment. The Pennsylvania State University, 2016.
- [4] Hall-Beyer, Mryka. (2017). GLCM Texture: A Tutorial v. 3.0 March 2017. 10.13140/RG.2.2.12424.21767.
- [5] R. Janani and A. Gopal, "Identification of selected medicinal plant leaves using image features and ANN," 2013 International Conference on Advanced Electronic Systems (ICAES), Pilani, 2013, pp. 238-242. doi: 10.1109/ICAES.2013.6659400
- [6] R.M. Haralick, K. Shanmugam, and I. Dinstein, "Textural Features for Image Classification" , IEEE Trans. On Systems, Man and Cybernetics, 1973, pp.610-621.
- [7] I. Rodríguez-Fdez, A. Canosa, M. Mucientes, A. Bugarín, STAC: a web platform for the comparison of algorithms using statistical tests, in: Proceedings of the 2015 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 2015.

Referências

- [8] SHANNON Entropy. 1. 2017. Disponível em: <https://en.wiktionary.org/wiki/Shannon_entropy>. Acesso em: 06 nov. 2018.
- [9] 1.9 Naive Bayes: Gaussian Naive Bayes. 2017. Disponível em: <https://scikit-learn.org/stable/modules/naive_bayes.html#gaussian-naive-bayes>. Acesso em: 07 nov. 2018.
- [10] Lorena, Ana Carolina, and André CPLF de Carvalho. "As Máquinas de Vetores Suporte." (2002).
- [11] Pedregosa, Fabian, et al. "Scikit-learn: Machine learning in Python." Journal of machine learning research 12.Oct (2011): 2825-2830.b
- [12] Kadir, Abdul, et al. "Leaf classification using shape, color, and texture features." arXiv preprint arXiv:1401.4447 (2013).
- [13] REGRESSÃO Logística. 2018. Disponível em: <https://pt.wikipedia.org/wiki/Regress%C3%A3o_log%C3%ADstica>. Acesso em: 09 nov. 2018.
- [14] Perktold, Josef, Skipper Seabold, and Jonathan Taylor. "Statsmodels: Statistics in python." (2017).
- [15] OPÇÕES ANOVA: Testes post hoc. Disponível em: <https://www.ibm.com/support/knowledgecenter/pt-br/SSEP7J_10.1.1/com.ibm.swg.ba.cognos.ug_cr_rptstd.10.1.1.doc/c_id_obj_anova.html>. Acesso em: 06 nov. 2018.

Referências

[16] Mordvintsev, Alexander, and K. Abid. "Opencv-python tutorials documentation." Obtenido de <https://media.readthedocs.org/pdf/opencv-python-tutroals/latest/opencv-python-tutroals.pdf> (2014).