9ª Jornada Científica e Tecnológica do IFSULDEMINAS
6º Simpósio da Pós-Graduação
ISSN 2319-0124

MONTAGEM E TREINAMENTO DE REDES NEURAIS PERCEPTRON PARA IDENTIFICAÇÃO DE PROTEÍNAS EFETORAS

[1]Gabriel S. OLIVEIRA; [2]Leonardo F. MOREIRA; [3]Claudinei O DOTTO; [4]Gustavo J. SILVA

RESUMO

Biotecnologia é o ramo que integra várias áreas da biologia com tecnologias. Com essa integração, diversos cursos direcionados às áreas computacionais tiveram a oportunidade de utilizar os conhecimentos fornecidos por essas formações para aplicar em dados biológicos. Esta pesquisa mostra a classificação de proteínas efetoras utilizando redes neurais perceptron com base em sete características biológicas. Foram avaliados os resultados obtidos por um treinamento realizado em uma rede neural perceptron para verificar o percentual de acerto que a rede realizava. Com o decorrer da pesquisa foi constatado que a rede neural treinada conseguiu alcançar uma taxa aceitável de acerto de 91.5%.

Palavras-chave: Bioinformática; Proteoma; Análise de Sequência

1. INTRODUÇÃO

A biotecnologia atua em diversas áreas, desde a produção de fármacos até trabalhar em conjunto com outras ciências, buscando métodos que resolvam problemas e mecanismos de inovação para determinado fim. Assim, pode-se estudar sequências genéticas de DNA e proteínas.

As proteínas exercem um papel muito importante dentro do organismo humano, pois fornecem o material tanto para construção como para a manutenção de todos os nossos órgãos e tecidos, além de outras funções químicas. Segundo Araújo (2008), mais de mil projetos já foram desenvolvidos no mundo sobre proteínas. Sendo assim, existe uma vasta quantidade de dados sobre esse assunto. Existem bactérias que podem secretar proteínas durante a interação com a célula dentro do organismo humano, podendo alterar o processo celular e causar algum tipo de doença. Essas proteínas que alteram os processos no hospedeiro denominam-se proteínas efetoras (ALVAREZ-MARTINEZ; CHRISTIE, 2009).

Esta pesquisa justifica-se pela importância na identificação de proteínas efetoras e tem como objetivo geral desenvolver uma rede neural capaz de ajudar na identificação de proteínas efetoras de acordo com suas características de hidropatia, sinal de localização nuclear e sinal de localização mitocondrial, utilizando redes neurais perceptron.

[1] IFSULDEMINAS - Campus Muzambinho. Muzambinho/MG - Email: gsantannamb@hotmail.com
[2] IFSULDEMINAS - Campus Muzambinho. Muzambinho/MG - Email: leonardomoreiramg@gmail.com
[3] IFSULDEMINAS - Campus Muzambinho. Muzambinho/MG - Email: claudinei.dotto@gmail.com
[4] IFSULDEMINAS - Campus Muzambinho. Muzambinho/MG - Email: gustavo.jose@muz.ifsuldeminas.edu.br

ISSN 2319-0124

2. MATERIAL E MÉTODOS

No desenvolvimento, foi adquirida uma base de dados com proteínas efetoras e não efetoras de bactérias no endereço ftp.ncbi.nlm.nih.gov/genomes/Bacteria/. Foram criados dois grupos, um de proteínas efetoras e outro de não efetoras. As proteínas declaradas como efetoras foram comprovadas em laboratório de acordo com Lockwood et al. (2011).

As ferramentas de bioinformática utilizadas nas análises dos grupos foram Hydrocalc Proteome (responsável por analisar características relacionadas à hidropatia), NLStradamus (localiza sinal de localização nuclear) e TargetP (encontra padrões relacionado ao sinal de localização mitocondrial) (EMANUELSSON et al., 2007). Com os resultados das características das proteínas, foram criados dois documentos, um para cada grupo de proteína.

Foram criados mais dois documentos, denominados treinamento e pós-treinamento. No documento de treinamento foram colocadas as 100 primeiras proteínas de cada um dos grupos um total de 200 proteínas, sendo 100 efetoras e 100 não efetoras para que fosse possível utilizar o software Weka (DAMASCENO, 2010). No outro documento (pós-treinamento) foram inseridas as demais proteínas, afirmando que 44 delas eram efetoras e 148 não efetoras.

Na fase de testes o documento de treinamento foi adicionado a uma função do software Weka, chamada Multilayer Perceptron, no qual foi fornecida a opção de visualizar de forma gráfica do perceptron, na qual pode-se alterar as variáveis de treinamento como taxa de erro, momentum e número de épocas. Foram realizados 2 tipos de testes com treinamento da rede, o primeiro foi com uma camada de 4 neurônios, o segundo com 2 camadas, sendo cada uma possuindo 4 neurônios, totalizando 8 neurônios no final.

3. RESULTADOS E DISCUSSÕES

Na análise do documento de treinamento da rede as proteínas se separam de acordo com características analisadas, conforme mostra a Figura 1, gerada na ferramenta Weka.

Figura 2 – Tela de Cadastro de CT-e

Figura 3 – Login ao sistema

ISSN 2319-0124

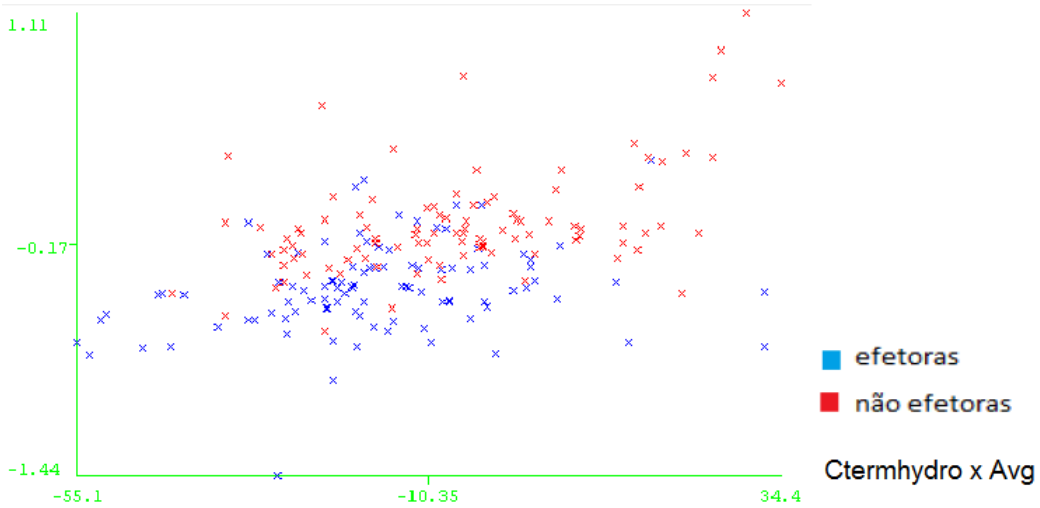

Figura 1 - Gráfico Ctermhydro x Avg.

Foi feito um treinamento com uma camada de 4 neurônios, que foi ajustado o limite de época, que se resume em quantidade de vezes que é feito o ajuste dos pesos em uma rede. A Tabela 1 mostra os resultados percentuais referentes à taxa de acertos.

Tabela 1 – Treinamento Perceptron de 4 neurônios com limite de épocas.

Resultado por épocas – Perceptron 4 neurônios		
Número de épocas	Erro por época	Taxa de acerto %
1000	0.0903257	89
5000	0.0740908	91
10000	0.0721023	91
100000	0.0692064	91
1000000	0.0674174	91
5000000	0.0621091	91.5

ISSN 2319-0124

Percebe-se que o melhor resultado foi o treinamento da rede neural com 4 neurônios e 5000000 de épocas, apresentando uma taxa de acerto de 91.5%,

4. CONCLUSÕES

A pesquisa em si, teve como objetivo principal, treinar uma rede neural capaz de classificar as proteínas efetoras, o que ajuda pesquisadores na cura de doenças. Conclui-se também que a rede neural alcançou uma taxa de 91.5% com as características apresentadas. Recomenda-se, para futuros trabalhos, aumentar o número de características analisadas, visando aumentar a taxa de acertos.

AGRADECIMENTOS

Ao Instituto Federal de Educação, Ciência e Tecnologia do Sul de MG pelo apoio recebido no projeto de iniciação científica.

REFERÊNCIAS

ALVAREZ-MARTINEZ, C. E.; CHRISTIE, P. J. Biological diversity of prokaryotic type IV secretion systems. Microbiology and molecular biology reviews. Washington, p. 775-808. dez. 2009.

ARAÚJO, Nilberto Dias de et al. A era da bioinformática: Seu potencial e suas implicações para as ciências da saúde , Estud Biol. P. 775-808. Jan/dez 2008.

EMANUELSSON, O et al. Locating proteins in the cell using TargetP, SignalP, and related tools. **Nature**. London, p. 953-971. abr. 2007.

LOCKWOOD, S. et al. Identification of Anaplasma marginale Type IV Secretion System Effector Proteins. Plos One. San Francisco, 6, e0027724. nov. 2011.

MELO, MARCELO DAMASCENO DE () INSTITUTO FEDERAL DE EDUCAÇÃO, Ciência E Tecnologia Do Rio Grande Do Norte/Campus Macau). Introdução a mineração de dados utilizando o weka. *V Connepi*, n. 1, 2010.