



Simpósio de Integração Acadêmica

“Ciências Básicas para o Desenvolvimento Sustentável”

SIA UFV 2023



Redução de marcadores para predição genômica de produtividade em *Coffea arabica* via Redes Neurais Artificiais

Victória Manhago Salvador¹, Moysés Nascimento, Weverton Gomes da Costa, Noé Eiter, Eveline Teixeira Caixeta

¹Universidade Federal de Viçosa, Centro de Ciências Exatas, Departamento de Estatística, Viçosa, Minas Gerais.
Autor correspondente: victoria.salvador@ufv.br

Palavras-chave: café, melhoramento genético, inteligência artificial, GWS

Introdução

O melhoramento genético do cafeeiro é um processo lento que demora anos. Para acelerar a identificação de variedades superiores é de suma importância que métodos de seleção mais eficientes e rápidos sejam desenvolvidos. Nesse contexto, a Seleção Genômica (SG), a qual permite predizer o mérito genético de plantas com base de informações moleculares com grande acurácia, se apresenta como uma metodologia interessante. Dentre os métodos utilizados para realizar a predição genômica, as Redes Neurais Artificiais (RNA) capturam naturalmente as relações não-lineares entre os marcadores a partir dos dados de entrada. Desta forma, o uso de RNA em SG vem aumentando. Porém, o grande número de marcas apresenta-se como um problema para o ajuste de uma rede neural considerando um tempo razoável. Desta forma, faz-se necessário a utilização de técnicas para reduzir a dimensionalidade dos dados a fim de ajustar modelos de Inteligência Computacional

Objetivos

Assim, o objetivo desse trabalho foi comparar a capacidade preditiva de diferentes arquiteturas de RNA considerando diferentes números de marcadores como entrada da rede para predição do mérito de genótipos quanto a produtividade em *Coffea arabica*.

Material e Método

Para isso, foram utilizados dados de 195 genótipos de *C. arabica*. Inicialmente, foram considerados 3824 marcadores, que após redução das marcas (usando algoritmos de aprendizado de máquina pelo método Random Forest) foram diminuídos para 1077 marcas de importância. Os dados foram divididos em treinamento, 80% dos genótipos (156 indivíduos), e teste, 20% dos genótipos (39 indivíduos). As RNA utilizadas são do tipo Perceptron de Múltiplas Camadas, com duas camadas ocultas, cada uma variando 1 a 20 neurônios.

Resultados e Discussão

Quando foram utilizados todos os 3824 marcadores, a RNA, composta de 8 neurônios na primeira e também na segunda camada, apresentou uma capacidade preditiva (CP) igual à 0,54 e o tempo de espera foi de 45 segundos. Após a redução do número de marcas considerado para o ajuste da RNA, o valor de CP foi igual à 0,64 e o tempo de espera foi reduzido para 3 segundos. O melhor resultado foi obtido quando se considerou uma RNA composta de 19 neurônios na primeira camada e 2 neurônios na segunda camada

Conclusões

De maneira geral, foi possível observar que a redução do número de marcas, por meio de algoritmos de aprendizado de máquina, acarretou no aumento da capacidade preditiva evidenciando que modelos de inteligência computacional podem ser ajustados com menos recursos computacionais e menos tempo.

Agradecimentos



Apoio financeiro

