



# Simpósio de Integração Acadêmica

“Bicentenário da Independência: 200 anos de ciência, tecnologia e inovação no Brasil e 96 anos de contribuição da UFV”

SIA UFV 2022



## PREDIÇÃO GENÔMICA DA RESISTÊNCIA A CERCOSPORIOSE EM *COFFEA ARÁBICA* VIA REDES NEURAIS ARTIFICIAIS

Victória Manhago Salvador<sup>1</sup>, Moysés Nascimento<sup>1</sup>, Noé Mitterhofer Eiterer Ponce de Leon da Costa<sup>1</sup>, Eveline Teixeira Caixeta<sup>1</sup>, Antonio Carlos Baião de Oliveira<sup>1</sup>

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

**Palavras-Chaves:** café, melhoramento genético, inteligência artificial

Área temática: Inteligência Artificial; Grande Área: Ciências Agrárias; Categoria do Trabalho: Pesquisa

### 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 prever 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.

### Objetivos

O objetivo desse trabalho foi comparar a capacidade preditiva de diferentes arquiteturas de RNA na predição do mérito de genótipos quanto a resistência a cercosporiose em *Coffea arábica*.

### Material e Métodos

Foram utilizados dados de resistência a cercosporiose de 195 genótipos de *Coffea arábica* genotipados para 3824 marcadores. 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. Foram consideradas para o ajuste diferentes números de iterações, sendo eles 10, 20, 30, 40, 50, 60, 100, 300 e 500. Para a comparação foram utilizadas a capacidade preditiva e erro quadrático médio com métricas.

### Resultados e Discussão

**Tabela 1.** Resumo das melhores RNAs de cada iteração com base na capacidade preditiva do conjunto teste (CP teste), onde o número de neurônios na primeira camada (NC1) e na segunda camada (NC2), capacidade preditiva do conjunto de treino (CP treino) e erro quadrático médio (EQM).

Iterações	NC1	NC2	CP treino	CP teste	EQM
50	2	12	0,40305	0,60040	0,81248
30	4	10	0,35326	0,57928	0,57928
40	2	6	0,29164	0,55954	0,97501
20	8	3	0,37817	0,54714	0,87770
60	8	3	0,49642	0,52955	0,95210
100	3	16	0,39504	0,52123	0,91156
300	8	18	0,78288	0,49045	0,98498
500	2	16	0,37279	0,46369	0,99977
1000	4	19	0,53088	0,44114	1,04575

### Conclusões

Esse resultado pode ser explicado devido ao fenômeno denotado por overfitting, ou seja, um modelo obtém ótimos resultados em sua população de treinamento, mas possui pobre performance na realização de predições.

### Apoio Financeiro



### Agradecimentos

