

TRABALHO DE CONCLUSÃO DE CURSO

ENGENHARIA FLORESTAL

**TAMANHO DA POPULAÇÃO NO TREINAMENTO DE RNA PARA
ESTIMAR O VOLUME COMERCIAL DA CAATINGA ARBÓREA
COM O NEAT**

LOIC AYMARD MFOUMBYT MFOUMBYT



Loic Aymard Mfoumbyt Mfoumbyt

**TAMANHO DA POPULAÇÃO NO TREINAMENTO DE RNA PARA ESTIMAR
O VOLUME COMERCIAL DA CAATINGA ARBÓREA COM O NEAT**

Trabalho de Conclusão de Curso apresentado ao Instituto de Ciências Agrárias da Universidade Federal de Minas Gerais, como requisito parcial, para a obtenção de título de Bacharel em Engenharia Florestal.

Orientador: Christian Dias Cabacinha



UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS AGRÁRIAS
COLEGIADO DO CURSO DE GRADUAÇÃO EM ENGENHARIA FLORESTAL

ATA DE DEFESA DE MONOGRAFIA / TRABALHO DE CONCLUSÃO DE CURSO (TCC)

Aos 19 dias do mês de março de 2021, às 14h00min, o estudante Loic Aymard Mfoumyby Mfoumyby, matrícula 2016500039, defendeu o Trabalho intitulado “**TAMANHO DA POPULAÇÃO NO TREINAMENTO DE RNA PARA ESTIMAR O VOLUME COMERCIAL DA CAATINGA ARBÓREA COM O NEAT**” tendo obtido a média (92,0) .

Participaram da banca examinadora os abaixo indicados, que, por nada mais terem a declarar; assinam eletronicamente a presente ata.

Nota: 92,0 (noventa e dois)

Orientador(a): Christian Dias Cabacinha

Nota: 92,0 (noventa e dois)

Examinador(a): Renato Dourado Maia

Nota: 92,0 (noventa e dois)

Examinador(a): Iandra Alexia Rodrigues Silva



Documento assinado eletronicamente por **Christian Dias Cabacinha, Professor do Magistério Superior**, em 19/03/2021, às 16:08, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **Renato Dourado Maia, Professor do Magistério Superior**, em 19/03/2021, às 16:14, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



Documento assinado eletronicamente por **landra Aléxia Rodrigues Silva, Usuário Externo**, em 19/03/2021, às 16:18, conforme horário oficial de Brasília, com fundamento no art. 5º do [Decreto nº 10.543, de 13 de novembro de 2020](#).



A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **0632196** e o código CRC **8F3CB236**.

Este documento deve ser editado apenas pelo Orientador e deve ser assinado eletronicamente por todos os membros da banca.

Referência: Processo nº 23072.213256/2021-70

SEI nº 0632196

AGRADECIMENTOS

Gostaria de agradecer a Deus pelos acontecimentos na minha vida no Brasil. Graças a Ele, consegui viver momentos incríveis, conhecer pessoas diferentes na minha vida. Agradeço a Deus também por sempre estar comigo em todos os momentos que eu precisava.

Agradeço a minha mãe Marie Madeleine, pelas orações e apoio em todos os momentos bons e ruins, durante esses seis anos distante dela e por acreditar sempre em mim. Ao meu pai Siméon pela força e pelos conselhos nos momentos importantes. A todos os membros da minha família, pela força psicológica, em especial ao meu tio Zouba Bouka, e minhas irmãs, Sanie e Stela.

Aos meus colegas de turma, pelas conversas, por nossas “farras” e por terem compartilhado comigo momentos incríveis durante nossas viagens técnicas. A meu amigo e irmão que Deus me deu, Luiz Otávio, obrigado pela presença e por sempre me ouvir quando precisei.

Obrigado a todos os professores do curso de Engenharia Florestal, que participaram de forma direta ou indireta dessa jornada, em especial, ao meu orientador, Professor Christian, por confiar em mim e pelo conhecimento adquirido no Laboratório de Pesquisa em Mensuração Florestal (LAMEF). Agradeço também aos membros do Núcleo de Estudos em Manejo Florestal (NEMAF), pelo crescimento pessoal e acadêmico que a convivência me proporcionou. Foram momentos ricos de troca de experiências e de aprendizado. Obrigado à Professora Adriana pela inspiração e pelos conselhos.

Obrigado à UFMG pela oportunidade, pelo acolhimento desde as aulas de língua portuguesa no *campus* Pampulha em Belo Horizonte até a conclusão do curso de Engenharia Florestal no Instituto de Ciências Agrárias em Montes Claros.

Por fim, gostaria de agradecer a cada pessoa que esteve comigo e compartilhou seu tempo e momentos nestes seis anos no Brasil. Serei sempre grato por tudo.

Obrigado a todos!!!

RESUMO

A Caatinga Arbórea é uma tipologia vegetacional da Caatinga, bioma exclusivamente brasileiro, composta por diversas espécies de interesse medicinal, alimentício (humano e animal) e econômico. O conhecimento do seu estoque volumétrico é fundamental na elaboração de planos de manejo e o uso das Redes Neurais Artificiais é um método que se destaca por sua capacidade de produzir excelentes resultados. Assim, o objetivo deste trabalho foi avaliar o treinamento de RNA com diferentes números de indivíduos para o algoritmo genético NEAT, na estimativa do volume comercial de árvores de Caatinga arbórea para identificação da melhor configuração para o treinamento com menor custo computacional. Dados de volume comercial com casca de 300 indivíduos foram obtidos para compor a base de dados para o processamento com as Redes Neurais Artificiais usando o algoritmo genético NEAT. O processamento foi composto por duas fases, na qual a primeira continha redes com 10 indivíduos até 100 em intervalos de 10 indivíduos e a segunda, redes de 100 indivíduos até 1000 em intervalos de 100 indivíduos. Realizou-se uma análise das estatísticas correlação e raiz quadrada do erro médio para cada fase para comparar as diferentes configurações que geraram resultados satisfatórios. Após a análise, foi constatado que os resultados da validação foram diferentes para as fases 1 e 2. As redes com número de indivíduos inferior a 300 no treinamento foram descartadas e a partir de 400 indivíduos, com densidade populacional de 0,1, apresentaram resultados mais satisfatórios. Além disso, foram comparados esses resultados a partir do teste de Sinal com nível de significância de 5%. Assim, concluiu-se que a melhor configuração para o treinamento de redes neurais artificiais para estimar o volume comercial com casca, utilizando como variáveis de entrada o DAP e HT com o algoritmo NEAT com menor custo computacional é considera 500 indivíduos para o tamanho da população com densidade populacional de 0,1.

Palavras-chave: Caatinga. Algoritmo Genético. Produção Volumétrica. Tamanho Populacional.

LISTA DE ILUSTRAÇÕES

Figura 1 – Estrutura de uma rede de funções de base radial	13
Figura 2 – Arquitetura de uma rede Multilayer Perceptron (MLP) do tipo feedforward...	14
Figura 3 – Localização da FLONA Contendas do Sincorá com a respectiva alocação dos três blocos onde foram aplicados os tratamentos e as atividades de cubagem rigorosa.....	16
Figura 4- Boxplot para a correlação (a) e RQME% (b) para a fase 1 de processamento dos dados e para a correlação (c) e RQME% (d) para a fase 2 de processamento dos dados..	21
Figura 5 - <i>Boxplot</i> para a correlação do treinamento (a), correlação da validação (b), RQME do treinamento (c) e RQME da validação (d) para a fase 2 de processamento dos dados.....	22

LISTA DE TABELAS

Tabela 1 – Frequência absoluta por classe diamétrica dos indivíduos mensurados na Floresta Nacional Contendas do Sincorá (FLONA), de espécies da caatinga arbórea.	17
Tabela 2 – Valores globais das variáveis quantitativas mensuradas na cubagem.....	18
Tabela 3 – Características estatísticas do comportamento das melhores redes neurais artificiais com uso do algoritmo NEAT para estimativa do volume comercial da casca (VCcc), considerando populações de 10 até 100 indivíduos.....	20
Tabela 4 – Características estatísticas do comportamento das melhores redes neurais artificiais com uso do algoritmo NEAT para estimativa do volume comercial da casca (VCcc), considerando populações de 100 a 1000 indivíduos.....	20
Tabela 5 – Resultados das comparações de redes na fase de treinamento para RQME do teste de sinais.....	24
Tabela 6 – Resultados das comparações de redes na fase de validação para RQME do teste de sinais da estimativa.....	25

SUMÁRIO

1. INTRODUÇÃO	9
2. REFERENCIAL TEÓRICO	11
2.1 Volumetria de Espécies Nativas.....	11
2.2 Redes Neurais Artificiais (RNA)	12
2.3 Tipos de RNAs	13
2.3.1 <i>Radial Basis Function</i>	13
2.3.2 <i>Multilayer Perceptron</i>	14
2.4 Algoritmos	14
2.4.1. <i>Backpropagation</i>	14
2.4.2. Algoritmo Genético: <i>Neuroevolution of Augmenting Topologies</i> (NEAT).....	15
3. MATERIAL E MÉTODOS	16
3.1. Caracterização da Área de Estudo.....	16
3.2. Área Amostral e Critérios de Inclusão	16
3.3. Cubagem Rigorosa	17
3.4. Redes Neurais Artificiais	18
3.5. Critérios Estatísticos para Avaliação do Desempenho da RNA	19
4. RESULTADOS E DISCUSSÃO	20
5. CONCLUSÃO	26
REFERÊNCIAS	27

1. INTRODUÇÃO

A Caatinga é um bioma exclusivamente brasileiro que ocupa uma área de 844.453 km², 10% do território do país, distribuída nos estados do Ceará, Alagoas, Minas Gerais, Bahia, Maranhão, Paraíba, Pernambuco, Piauí, Rio Grande do Norte e Sergipe. Subdivisão deste bioma, a Caatinga Arbórea é composta por diversas espécies de interesse medicinal, alimentício (humano e animal) e econômico (RIZZINI, 1997), além da utilização da sua madeira para serraria e carvão, como exemplo da *Jatropha molíssima* (Pohl) Baill. (Pinhão), a *Ziziphus joazeiro* Mart (Juazeiro) e *Myracrodruon urundeuva* Allemão (Aroeira), comumente destacadas em planos de manejo no bioma.

Diante da importância e uso da Caatinga, em especial da Caatinga Arbórea, tipologia vegetal bastante manejada, faz-se necessário, realizar estudos para obtenção do estoque volumétrico presente com maior precisão, para adequada supressão dos indivíduos, valoração e quantificação do que será explorado (SANQUETTA et al., 2014).

Comumente, para conhecimento do estoque de madeira, usa-se equações volumétricas ajustadas para estimar o volume da vegetação, porém os modelos matemáticos encontrados na literatura nem sempre conseguem explicar o comportamento dessa variável de forma satisfatória devido à alta variabilidade que envolve algumas situações, como solo, espécie e bioma interferindo diretamente no comportamento de forma das árvores e consequentemente em seu volume (REZENDE et al., 2006).

Como saída, pode-se fazer uso de outras ferramentas de aprendizagem computacional, como as Redes Neurais Artificiais (RNA) por exemplo.

As Redes Neurais Artificiais (RNAs) são sistemas complexos que, através da entrada de dados, detectam padrões e funcionam similarmente a um cérebro humano, capazes de fazer interpolações e extrapolações de forma linear e geram cenários distintos em diferentes arquiteturas que são algoritmos ligadas aos neurônios (HAYKIN, 2001; BRAGA et al., 2014). Essa metodologia tem se destacado pela sua capacidade de produzir excelentes resultados e pela tendência em gerar erros mínimos. Alguns trabalhos foram realizados para estimar o volume de árvores da Caatinga, a partir de RNA, com resultados promissores (ABREU, 2019; GOMES, 2017; ARAÚJO, 2015).

Um dos algoritmos que pode se destaca para o treinamento de RNA para árvores nativas é o Neuroevolution of Augmenting Topologies (NEAT). O NEAT, é um algoritmo genético de otimização e busca, baseado na teoria da seleção natural e genética de Darwin, que a partir de uma técnica de um algoritmo evolutivo ajusta simultaneamente os pesos sinápticos e a topologia de uma Rede Neural Artificial.

Ao mesmo tempo que o NEAT se configura como um excelente algoritmo para o treinamento de redes para estimar volume de espécies nativas, um alto tempo computacional é exigido pelo o algoritmo, quando em sua parametrização um grande número de indivíduos. No algoritmo NEAT, há a possibilidade de alterar configurações de população e suas densidades iniciais.

Nesse sentido, Gomes (2017), realizou testes para as populações de 300, 600, 800, 1000 e densidades de 0,1 e 0,5 em referência aos parâmetros avaliados por Araújo (2015). O autor verificou que o número de populações iniciais necessárias, as redes NEAT apresentaram estabilidade com populações acima de 800 no momento de treinamento, e seu comportamento na validação é influenciado pela quantidade de neurônios de entrada. Não é recomendado o uso de densidade de 0,5, em que os melhores desempenhos se mostraram a partir de 800 populações e densidade de 0,1.

Contudo, não se verificou nas abordagens com uso do NEAT, o uso de populações inferiores a 300 indivíduos. Valores de ciclos inferiores são interessantes, por se tratar de um algoritmo com alto consumo de tempo e processamento se comparado às outras topologias e tipos de redes (STANLEY; MUKKULAIEN, 2002; CAAMAÑOS et al., 2014).

Diante disso, o presente trabalho teve por objetivo avaliar o treinamento de RNA com diferentes números de indivíduos para o algoritmo genético NEAT, na estimativa do volume comercial de árvores de Caatinga arbórea, na busca da melhor configuração para a geração de redes satisfatórias com o menor custo computacional.

2. REFERENCIAL TEÓRICO

2.1 Volumetria de Espécies Nativas

No Brasil, o uso da lenha no Nordeste é uma prática recorrente que acontece desde o início da colonização e perdura de maneira intensa até os dias atuais, aumentando constantemente a pressão sobre os recursos florestais naturais (FIGUEIRÔA et al., 2005).

O estudo do volume representa uma das melhores formas de conhecer o potencial produtivo em florestas e o conhecimento do seu estoque é essencial para todas as atividades de planejamento (MACHADO et al., 2000; THOMAS et al., 2006; FRANCO et al., 1998). Contudo, ainda são poucos estudos que analisam a volumetria para espécies nativas brasileiras, considerando as distinções existentes no país, tanto no que tange a variabilidade genética da flora brasileira, quanto variações topográficas em algumas regiões do país. A existência de tal informação não se apresenta de forma oficial em diversos estados brasileiros (COLPINI et al., 2009; SANTOS et al., 2012).

Existem várias formas de se obter o volume da vegetação. Uma delas, é a partir da cubagem, que consiste em coletar dados para obter os volumes, a partir da medição de diâmetros em diferentes posições dos fustes das árvores e da soma dos volumes calculados para as seções definidas pelas alturas de medição. Por efeito de apresentar fustes com características diferentes para florestas tais como plantadas ou nativas, aplicação do método de cubagem necessita escolha de uma das fórmulas recomendadas. As fórmulas de Huber, Newton e Smalian para determinação são as mais usuais dentro do processo de mensuração florestal.

Pode-se obter o volume de árvores também pelo método de deslocamento de água. Isso pode ser operacionalizado por um xilômetro, que consiste em um tanque cheio de água associado a um circuito eletrônico que mede o volume de água deslocado quando uma tora é submersa no tanque. Apesar de ser um método que proporciona a medida mais próxima do valor exato do volume de fustes de árvores (SOUZA et al., 2017), este procedimento não é muito usado no setor florestal e não apresenta muitos estudos.

Os dados obtidos por esses métodos servem para ajustar equações matemáticas por regressão, para estimar o volume das árvores as equações de volume são divididas em equações de simples entradas, em que o volume é estimado somente a partir dos diâmetros a 1,30 m do solo (DAP) e para sua eficiência, exigem fortes correlações entre o volume e altura das árvores. Já as equações de dupla entrada, estabelecem uma relação entre o

volume e o DAP e altura das árvores. Esse segundo grupo, por usar duas variáveis (DAP e altura) para estimar o volume, tendem gerar estimativas mais precisas.

As equações podem ser ajustadas por espécie, por grupo de espécies, o que é convencional, ou para algumas espécies importantes, a depender do interesse em termos de planos de manejo, baseado nos objetivos do inventário florestal (SCOLFORO et al., 2008; SANQUETTA et al., 2011; SOARES et al., 2011a; SANTOS et al., 2012). Entretanto na prática de medições, dentro de uma rotina clássica de inventário florestal, raramente se ajusta equações de volume por espécie. Isso se deve ao fato de que existem espécies que são raras em certas comunidades vegetais, o que impede a composição de uma boa base de informações de cubagem rigorosa representativa (em número de observações) para tais espécies. Um outro aspecto é que não raro, alguns indivíduos não são identificados, durante o levantamento de dados. Esses aspectos impedem a estratificação da base de dados para ajustes de equações por espécie.

Nas últimas décadas, com foco na melhoria de estimativas de volume de árvores em florestas plantas e nativas e redução do número de indivíduos amostrados em inventários florestais, técnicas de aprendizagem de máquina passaram a ser avaliadas, com destaque para as redes neurais artificiais.

2.2 Redes Neurais Artificiais (RNA)

As Redes Neurais Artificiais (RNA) são modelos computacionais inspirados no sistema nervoso dos seres vivos. São sistemas que atuam de forma semelhante sendo constituídos por unidades de processamento simples, também denominadas neurônios artificiais, conectadas entre si de maneira particular com propósito efetuar determinada tarefa (BINOTI, 2010).

Durante o processamento, as redes têm tendência a demonstrar capacidades de aprender com o modelo padrão, dando resultados com informações relevantes baseado no que foi aprendido. Deste modo, os neurônios artificiais são apresentados como modelos matemáticos similares aos neurônios biológicos e processam as informações recebidas e ponderadas por pesos sinápticos com resposta única (HAYKIN, 2001; BRAGA et al., 2007).

A aplicação de Redes Neurais Artificiais na área florestal tem demonstrado alta relevância por apresentar bons resultados, por exemplo, na classificação de estágios sucessionais na Amazônia (KUPLICH, 2009), estimação do volume de árvores (BINOTI

et al., 2014), modelagem da distribuição diamétrica (DIAMANTOPOULOU et al., 2015) e modelagem do crescimento em diâmetro (REIS et al., 2016).

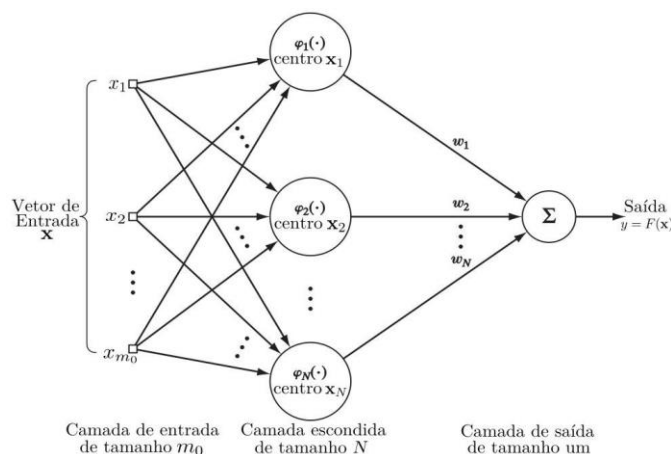
2.3 Tipos de RNAs

Na área florestal, diversos tipos de arquiteturas de Redes Neurais Artificiais foram avaliadas. As *Radial Basis Function* (RBF) e redes *Multilayer Perceptron* (MLP) são as mais usadas e ambas são classificadas como RNA do tipo *feedforward* (BRAGA et al., 2014).

2.3.1 Radial Basis Function

A *Radial Basis Function* (RBF) (Figura 1), constitui-se em uma rede de função de base radial, é composta por três camadas que possuem funções diferentes. A camada de entrada é formada por unidades sensoriais que conectam a rede ao ambiente em que ela está inserida; a segunda camada, única camada oculta existente na rede, executa uma transformação não linear do espaço de entrada para um espaço oculto; a camada de saída que se apresenta de forma linear projetada para dar resposta da rede a um padrão de ativação na camada de saída (HAYKIN, 2001; BRAGA et al., 2014).

Figura 1 – Estrutura de uma rede de funções de base radial.

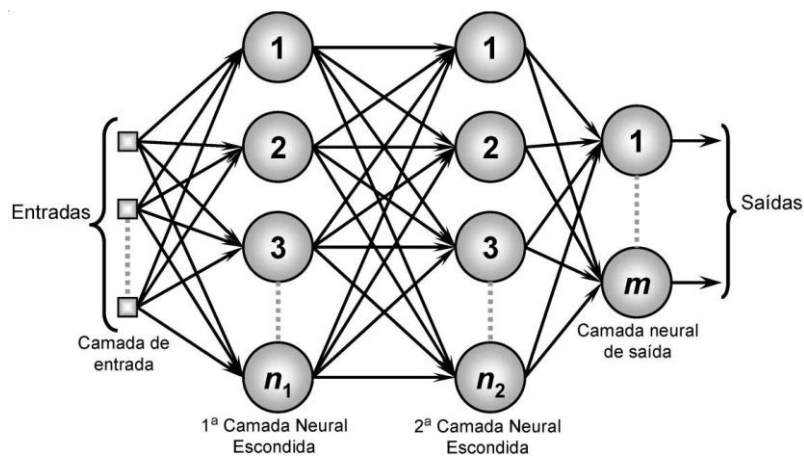


Fonte: HAYKIN et al., 2009.

2.3.2 Multilayer Perceptron

As redes *Multilayer Perceptron* (MLP) (Figura 2), são uma das mais importantes classes de redes neurais e são mais aplicadas para resolver diversos problemas. São redes que possuem uma camada de entrada que recebe os valores de entrada e as transfere ponderadas por pesos sinápticos à camada intermediária ou oculta que aplica transformações matemáticas e as transfere para a camada de saída que, por sua vez, fornece a resposta da rede. Durante este processo, o treinamento é efetuado de forma supervisionada através de um algoritmo muito popular conhecido como algoritmo de *backpropagation* (GOMES et al., 2017).

Figura 2 – Arquitetura de uma rede Multilayer Perceptron (MLP) do tipo feedforward.



Fonte: Gomes et al. (2017)

2.4 Algoritmos

2.4.1. Backpropagation

No sistema computacional cada rede é adaptada a um algoritmo que consegue proporcionar ou executar um treinamento. Existem diversos algoritmos reconhecidos como determinante para resolver problemas complexos. No caso do *backpropagation*, é algoritmo baseado na regra de aprendizagem apresentando várias formas de se obter um resultado satisfatório, com modelo da correção de erro (SILVA et al., 2017).

O algoritmo *backpropagation* contém duas etapas através das diferentes camadas da rede: uma etapa para frente, a propagação (*forward*), e outra para trás, vista como retropropagação (*backward*). No caso da propagação, existe um elemento chamado de

vetor de entrada, aplicado às unidades sensoriais da rede, onde informações se propagam ao longo da rede, camada por camada. Para outra situação, existe um sinal de erro na saída da rede subtraída da resposta desejada. Esse sinal de erro é então retropropagado, camada por camada, da camada de saída para a camada de entrada (BRAGA et al., 2014; HAYKIN, 2009; SILVA et al., 2010).

2.4.2. Algoritmo Genético: *Neuroevolution of Augmenting Topologies* (NEAT)

Além do *backpropagation*, existem também algoritmos denominados como genéticos e capazes de solucionar problemas mais complexos para dentro de estudos de Inteligência Artificial (AI). Um algoritmo genético tem como finalidade a otimização e a procura da melhor solução para diversos problemas. Seu funcionamento é baseado no aspecto de evolução natural, ou melhor, pratica da teoria da seleção natural, favorecendo o início de grande variedade de seres vivos com padrão bem definido. É a chave no processo de avaliação de elementos ou espécies a se adaptar em meio (KRAMER, 2017).

Neste caso, o Algoritmo mais usado é o *Neuroevolution of Augmenting Topologies* (NEAT). O NEAT é um algoritmo genético que objetiva o seu aproveitamento de uma rede pelo meio de sua estrutura, diminuindo o tamanho da pesquisa dos pesos das conexões de uma rede neural. É um algoritmo que apresenta princípios da computação evolutiva e dos algoritmos genéticos, trabalhando uso de genes, genótipos, fenótipos e seu processo evolutivo. O NEAT funciona de forma que, os genes possam ser dispostos de maneira ordenada quando existe cruzamento. Os genomas são representações de conectividade da rede. Existe lista de genes de conexão distribuída em diferente genoma, cada um referenciando a dois genes de nó conectados, ou seja, gene de conexão possui um nó de entrada, um nó de saída, seus respectivos pesos, a habilidade de estar ou não ativado e um número de inovação, que permite encontrar os genes correspondentes (STANLEY; MIIKKULAINEN, 2002).

Foram medidos todos os indivíduos dentro da parcela com diâmetro a 1,30 m do solo (DAP) superior ou igual a 5 cm, usando a suta para medição e estimativa visual para determinar a variável altura em todas as parcelas.

3.3. Cubagem Rigorosa

Os dados de cubagem usados neste estudo foram coletados por Gomes (2017) e cedidos para este estudo. Detalhes da coleta são descritos a seguir.

Os diâmetros foram agrupados em classes diamétricas para obter-se amostragem significativa do volume comercial com casca (VC_{cc}, em m³), considerando amplitude de 5 cm por classes.

A cubagem foi realizada utilizando o método de Smalian, com medições das seções ao longo do fuste, nas posições 0,1; 0,3; 0,5; 0,7; 0,9; 1,1; 1,3; 2,0; metros e assim sucessivamente, até a altura total (CAMPOS; LEITE, 2013; MACHADO; FIGUEIREDO FILHO, 2009; SOARES et al., 2006). Desse modo, foi estabelecido um limite mínimo de 3 cm de diâmetro ao medir cada seção, bem como realizadas avaliações de galhos com circunferência superior à 3 cm de diâmetro (Tabela 1).

Tabela 1 – Frequência absoluta por classe diamétrica dos indivíduos mensurados na Floresta Nacional Contendas do Sincorá (FLONA), de espécies da caatinga arbórea. Fonte: Gomes, 2017.

Classes	Intervalos	Frequência cubagem	Vol. (m ³) ¹
1	4,9 — 9,9	191	2,98483
2	9,9 — 14,9	63	3,26639
3	14,9 — 19,9	23	2,58930
4	19,9 — 24,9	12	2,76315
5	24,9 — 29,9	3	0,94196
6	29,9 — 34,9	1	0,21927
7	34,9 — 39,9	3	1,37436
8	39,9 — 44,9	2	1,28857
9	44,9 — 49,9	1	0,72836
10	49,9 — 54,9	1	1,07307
TOTAL		300	17,22926

¹ volume comercial da cubagem rigorosa referente ao somatório por classe diamétrica.

Ao considerar os dados de volume resultantes da cubagem rigorosa com 300 observações de 24 espécies identificadas e 5 espécies a nível de gênero, observou-se também, 63 espécies não identificadas devido a perda de folhas decorrente à época de precipitação muito baixa. As seguintes variáveis quantitativas foram consideradas: diâmetro com casca a 1,3 m de altura (DAP), altura total – em metros (HT) e o volume comercial com casca (VCcc, em m³).

A partir dos dados de cubagem rigorosa, obteve-se os volumes comerciais (até a altura de fuste) de cada indivíduo mensurado e gerou-se a base de dados para o treinamento das RNA. Essa base constituiu-se de dados para cada indivíduo mensurado de DAP, altura total (HT) e volume comercial com casca (VCcc), além da identificação a nível de espécie.

Tabela 2 – Valores globais das variáveis quantitativas mensuradas na cubagem.

Variável	Mínimo	Máximo	Médio	Desvio Padrão
DAP (cm)	5,00	52,60	10,50	6,96
HT (m)	3,50	11,50	6,74	1,41
VCcc (m ³)	0,00354	1,07307	0,05743	0,11392

Fonte: Gomes, 2017.

3.4. Redes Neurais Artificiais

Neste estudo, a arquitetura de rede usada foi a MLP. Foram testadas 30 redes para cada estratégia avaliada com o algoritmo *Neuroevolution of Augmenting Topologies* (NEAT), utilizando o software NeuroForest 4.0.

As RNA treinadas, apresentaram uma configuração em relação às variáveis utilizadas, compatíveis às equações de volume de dupla entrada. Na camada de entrada, utilizou-se dois neurônios, para as variáveis DAP e altura total (HT), oito neurônios na camada oculta, e um neurônio na camada de saída para o volume comercial com casca (VCcc)

A base de dados foi dividida em 70% para treinamento e 30% para validação. No treinamento das redes foi considerado apenas a densidade populacional de 0,1, conforme ARAÚJO (2015), 1.000 ciclos, conforme Gomes (2017), com um número de indivíduos variando de 10 até 100 em um intervalo de 10 indivíduos para a primeira fase do processamento e de 100 a 1000 com intervalo de 100 para a segunda fase do processamento.

3.5. Critérios estatísticos para avaliação do desempenho das RNA

Para estudo do comportamento das redes, foram avaliados o coeficiente de correlação entre os valores observados e estimados do volume ($r_{Y\hat{Y}}$) (Eq. 1) e a raiz do quadrado médio do erro absoluto (RQME) (Eq. 2) e relativo (RQME%) (Eq. 3).

$$r_{Y\hat{Y}} = \frac{\text{cov}(Y, \hat{Y})}{\sqrt{s^2(Y)s^2(\hat{Y})}} \quad (\text{Eq. 1})$$

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (\text{Eq. 2})$$

$$\text{RMSE}_{\%} = \frac{100}{\bar{Y}} * \sqrt{\frac{\sum_{i=1}^n (Y_i - \hat{Y}_i)^2}{n}} \quad (\text{Eq. 3})$$

Onde: Y = volume comercial real; \hat{Y} = volume comercial estimado; \bar{Y} = média aritmética do volume comercial; n = número de observações.

Utilizou-se gráficos do tipo *boxplot* para comparação dos resultados das 30 redes para uma análise complementar.

As redes que apresentaram os melhores resultados de treinamento foram comparadas, a partir da raiz do quadrado médio do erro absoluto (RQME), com o teste de sinais com um nível de significância estatística de 5% (SILVA JR; MAIA; CABACINHA, 2018). Esta análise foi realizada no software BioEstat 5.0.

4. RESULTADOS E DISCUSSÃO

As redes neurais artificiais foram treinadas e validadas em duas fases, a primeira com 10 até 100 indivíduos e a segunda de 100 a 1000, os resultados observados para os melhores resultados de treinamento e validação são apresentados na Tabela 3 e 4 respectivamente.

Tabela 3 – Características estatísticas do comportamento das melhores redes neurais artificiais com uso do algoritmo NEAT para estimativa do volume comercial da casca (VC_{cc}), considerando populações de 10 até 100 indivíduos.

Tamanho da população	RNA (Id)	Treinamento		Validação	
		$r_{\hat{Y}Y}\%$	RQME%	$r_{\hat{Y}Y}\%$	RQME%
10	24	96,80	47,59	99,02	39,06
20	14	96,80	47,58	99,03	39,20
30	8	96,80	47,59	99,04	38,99
40	14	96,80	47,59	99,02	39,13
50	30	96,79	47,57	99,01	39,20
60	18	96,80	47,58	99,03	39,05
70	10	96,80	47,59	99,04	38,88
80	10	96,80	47,59	99,03	39,07
90	24	96,80	47,59	99,10	37,98
100	10	96,79	47,58	99,05	38,63

Onde: RNA (Id)= Identificador da rede; $r_{\hat{Y}Y}\%$ =correlação entre volume comercial real e estimado e RQME%= raiz do quadrado médio do erro.

Fonte: Do autor, 2021.

Tabela 4 – Características estatísticas do comportamento das melhores redes neurais artificiais com uso do algoritmo NEAT para estimativa do volume comercial da casca (VC_{cc}), considerando populações de 100 a 1000 indivíduos.

Tamanho da população	RNA (Id)	Treinamento		Validação	
		$r_{\hat{Y}Y}\%$	RQME%	$r_{\hat{Y}Y}\%$	RQME%
100	10	96,79	47,58	99,05	38,63
200	3	96,80	47,59	99,02	39,15
300	14	97,90	38,84	97,88	45,10
400	17	97,35	39,72	96,86	47,33
500	25	98,19	35,40	97,60	41,95
600	21	98,32	34,86	96,31	44,59
700	28	98,22	35,77	98,40	46,19
800	11	98,90	28,28	96,95	48,92
900	12	98,42	33,73	96,25	63,27
1000	26	98,22	35,70	98,09	42,80

Onde: RNA (Id)= identificador rede; $r_{\hat{Y}Y}\%$ =correlação entre volume comercial real e estimado e RQME%= raiz do quadrado médio do erro.

Fonte: Do autor, 2021.

Observa-se pelos os resultados da Tabela 3, que para a fase 1 de processamento, tanto para o treinamento como para a validação, os valores de correlação e da raiz do quadrado médio do erro não sofreram grandes mudanças quando o tamanho da população aumentou de 10 para 100 indivíduos.

Destaque para a população com 90 indivíduos, que retornou a maior correlação entre volume real e estimado (99,10%) e a menor raiz do quadrado do médio do erro para a validação $\pm 0,0218 \text{ m}^3$ (37,98%).

Já na fase 2 de processamento no treinamento, observa-se que ao aumentar o número de indivíduos de 100 até 800 as estatísticas analisadas tenderam a valores melhores, isto é, as correlações aumentaram e a RQME diminuíram, inclusive a valores melhores que os observados na fase 1. Para as populações de 900 e 100 a correlação voltou a diminuir enquanto a RQME aumentou.

Destaque para a população com 900 indivíduos, que embora tenha gerado resultados muito bons na fase de treinamento, retornou valores de correlação (96,25%) e RQME igual a $\pm 0,0363 \text{ m}^3$ (63,27%) para a validação.

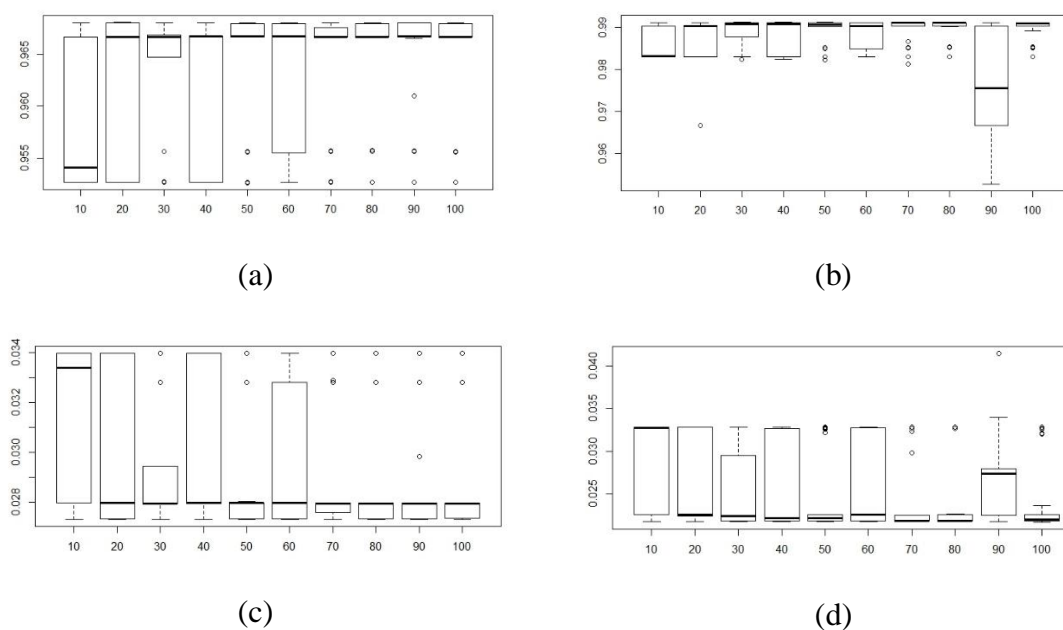
Contudo, o que chamou bastante atenção na fase 2, foi o comportamento verificado para as estatísticas na validação. Embora as redes tenham registrado excelentes estatísticas para o treinamento, na validação as estatísticas foram piores que as observadas na fase 1 de processamento.

Na fase 1 de processamento (Figura 4), observou-se tanto no treinamento como na validação, uma grande variação no comportamento das redes neurais para alguns tamanhos de população.

A mediana para a correlação aumentou a partir da rede com 20 indivíduos no treinamento (Figura 4a) e praticamente não se alterou até a rede com 100 indivíduos. Já para o RQME a rede com 10 indivíduos apresentou a maior mediana, a partir de 20 indivíduos a mediana diminuiu e praticamente não se alterou até a rede com 100 indivíduos.

Com relação a validação verificou-se uma grande dispersão e uma forte assimetria nos resultados de correlação e RQME (Figura 4b, 4d) nas redes com 10, 20, 40 e 60 indivíduos. Para a rede 10 do RQME no treinamento, a assimetria foi positiva, para as demais a assimetria foi negativa. Embora as demais configurações tenham apresentado uma dispersão menor, verificou-se ainda uma assimetria negativa para a rede com 30 indivíduos e as demais configurações uma assimetria positiva, sendo que para as redes 50, 70, 80, 90 e 100 a mediana foi igual ao primeiro quartil, resultado que pode estar associado ao grande número de valores discrepantes identificados na análise do *boxplot*.

Figura 4 - *Boxplot* para a correlação do treinamento (a), correlação da validação (b), RQME do treinamento (c) e RQME da validação (d) para a fase 1 de processamento dos dados.



Fonte: Do autor, 2021.

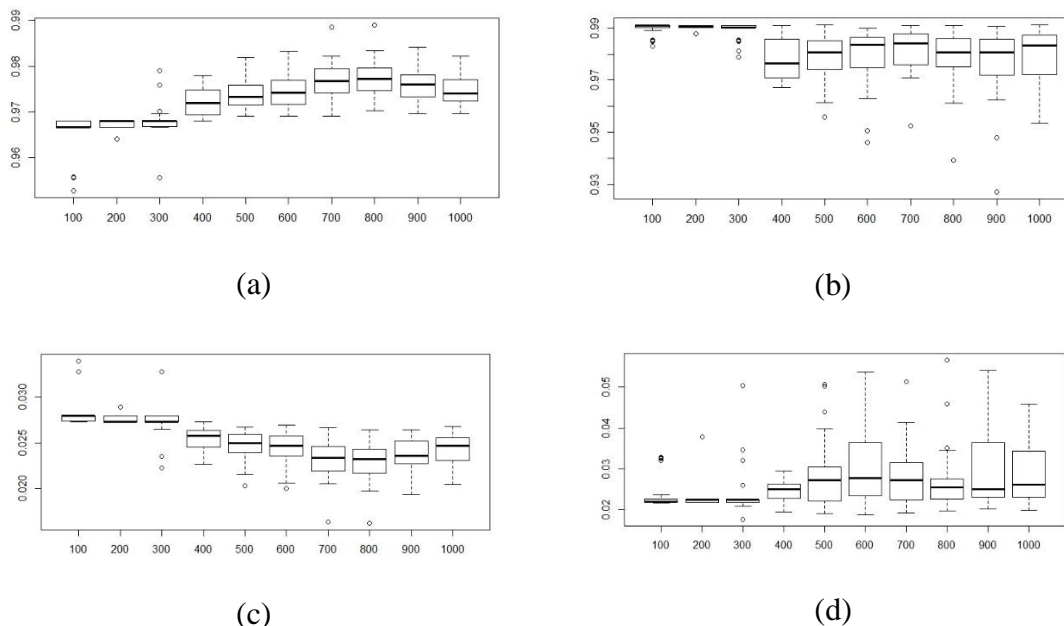
Diante dos resultados verificados na fase 1 de processamento, as redes com tamanho para a população inferiores a 70 indivíduos foram consideradas insatisfatórias e descartadas para a estimativa de volume comercial para Caatinga arbórea.

Na fase 2 de processamento (Figura 5), observou-se resultados mais satisfatórios, exceto para as configurações considerando 100, 200 e 300 indivíduos que apresentaram ainda comportamento semelhante ao descrito para as redes da fase 1 de processamento.

Esse comportamento era esperado, embora as redes com o maior número de indivíduos o custo computacional na etapa de treinamento seja maior, populações maiores toleram maior número de espécies potenciais, ou seja, maior quantidade de RNA testadas com alta diversidade topológica, aumentando as chances de obter bons resultados (STANLEY; MIIKKULAIEN, 2002). Comportamento semelhante foi também verificado por Araújo (2015) e Gomes (2017).

Para o treinamento, a partir de 400 indivíduos a mediana da correlação aumentou até a rede com 800 e voltou a diminuir para as redes 900 e 1.000 indivíduos. Para o RQME, como esperado, verificou-se o contrário, a partir de 400 indivíduos os valores das medianas decresceram até o mínimo, que ocorreu na rede com 800, voltando a subir para as redes 900 e 1.000 indivíduos.

Figura 5 - *Boxplot* para a correlação do treinamento (a), correlação da validação (b), RQME do treinamento (c) e RQME da validação (d) para a fase 2 de processamento dos dados.



Fonte: Do autor, 2021.

Na validação, de maneira geral, observa-se correlações mais altas para as populações iniciais inferiores a 400 indivíduos, diminuição e uma maior variação nos valores de correlação até 1.000 indivíduos. Para a RQME, observa-se valores mais baixos para populações iniciais inferiores a 400 indivíduos, aumento e uma maior variação nos valores de RQME até 1.000 indivíduos.

O ótimo desempenho do NEAT para populações com números de indivíduos acima de 300 no treinamento, pode estar associado ao algoritmo que busca diferentes topologias de rede a partir de uma estrutura básica e evolui a depender da complexidade do problema. Estas topologias podem envolver variadas combinações entre as ligações das variáveis e os neurônios, flexibilização dos pesos, ampliando assim as chances da rede em encontrar o erro mínimo global (STANLEY e MIIKKULAINEN, 2002).

Araújo (2015) quando avaliou redes NEAT para estimar o volume de formações florestais no estado de Minas Gérias, testou as seguintes configurações para o treinamento: número de indivíduos da população variando em 100, 300, 500, 600, 800 e 1.000 e densidades de conexões de 0,1 e 0,5. A autora constatou que somente a partir de populações maiores que 600 foram geradas redes com bom desempenho e concluiu que a partir das duas densidades de conexões analisadas (0,1 e 0,5), é possível obter estimativas volumétricas com bom grau de exatidão para as onze tipologias florestais estudadas e diferentemente ao verificado neste estudo, recomendou populações com número de

indivíduos maiores que 600, já que com populações inferiores não foram encontradas redes com bom desempenho para essa finalidade.

Gomes (2017), avaliando o desempenho de redes treinadas com o algoritmo NEAT para caatinga arbórea, verificou que o melhor resultado foi obtido para a densidade de conexão 0,1 e população de 800 indivíduos.

Após a análise dos resultados, percebe-se que as redes a partir de 400 indivíduos geram resultados mais satisfatórios, mesmo com estatísticas inferiores para a validação quando comparados as redes com populações menores.

Os resultados sugerem que o algoritmo NEAT, quando se considera populações iniciais com muitos indivíduos, tendem a gerar redes com resultados mais confiáveis (redes com menor variação nos resultados), sobretudo quando são treinadas poucas redes.

Neste estudo foram geradas, para cada estratégia estudada, 30 redes, com razoável variação nos resultados para populações inferiores a 70 indivíduos.

Entre 70 e 100, as redes geram resultados bastante similares, portanto sem aparente ganho em retornar o erro mínimo global. Quando se aumenta os intervalos entre os tamanhos das populações de 10 para 100 indivíduos, verificou-se que até 300 indivíduos não há melhoras e a partir de 400 até 1.000, há um aparente ganho no retorno do mínimo global, em redes com menores erros.

De maneira complementar, comparou-se as redes a partir de 400 indivíduos pelo teste de Sinal com o nível de significância de 5% para a RQME do treinamento (Tabela 5) e da validação (Tabela 6).

Tabela 5 – Resultados das comparações de redes na fase de treinamento para RQME do teste de sinais, com um nível de significância de 5%. considerando o desempenho das redes com população a partir de 400 até 1000 indivíduos.

	400	500	600	700	800	900	1000
400		•	•	•	•	•	•
500			ns	•	•	•	ns
600				•	•	•	ns
700					ns	ns	ns
800						ns	•
900							ns
1000							

Onde: “•” significativo para $\alpha=0,05$; ns = não significativo.

Fonte: Do autor, 2021.

Os resultados evidenciaram que no treinamento, a rede de 800 indivíduos (que apresentou menor mediana para a RQME) não se diferenciou estatisticamente apenas das redes com 700 e 900 indivíduos.

Já para a validação a menor mediana para o RQME foi gerada para 400 indivíduos e apenas a rede de 600 indivíduos se diferenciou estatisticamente desta.

Tabela 6 – Resultados das comparações de redes na fase de validação para RQME do teste de sinais, com um nível de significância de 5%, considerando o desempenho das redes com população a partir de 400 até 1000 indivíduos.

	400	500	600	700	800	900	1000
400		ns	•	ns	ns	ns	ns
500			ns	ns	ns	ns	ns
600				ns	ns	ns	ns
700					ns	ns	ns
800						•	ns
900							ns
1000							

Onde: “•” significativo para $\alpha=0,05$; ns = não significativo.

Fonte: Do autor, 2021.

Como a raiz do quadrado médio do erro (RQME), é obtida a partir da diferença entre valor observado e valor estimado, ao comparar os valores obtidos para o RQME para as diferentes estratégias avaliadas para a validação, ou seja, quando a rede estima o volume comercial para dados que não participaram do treinamento da rede, constatou-se que todas as redes treinadas foram estatisticamente semelhantes à rede com 500 indivíduos, sendo esse portanto, o menor tamanho para a população recomendado para o treinamento de redes com o algoritmo NEAT, para volume comercial de Caatinga arbórea com menor custo computacional em relação aos valores ora apontados pela literatura.

5. CONCLUSÃO

Os resultados desse trabalho evidenciam a potencialidade do uso de rede neurais treinadas com o algoritmo *Neuroevolution of Augmenting Topologies* (NEAT) para estimar o volume na caatinga arbórea. Assim como, permitem concluir que durante o treinamento de redes para estimar o volume comercial com casca utilizando como variáveis de entrada o DAP e a HT, o algoritmo NEAT apresentou resultados mais robustos com populações a partir de 400 indivíduos e recomenda-se utilizar pelo menos 500 indivíduos com densidade populacional de 0,1 e 1.000 ciclos durante o treinamento.

REFERÊNCIAS

- ABREU, J. C. de. **Alternativas para estimar o volume de árvores individuais em formações florestais no estado de Minas Gerais**. 2019. 82 f. Tese (Doutorado em Ciência Florestal) - Universidade Federal de Viçosa, Viçosa. 2019.
- ARAÚJO, A. C. S. C. **Redes neurais artificiais para quantificação volumétrica de florestas nativas no estado de Minas Gerais**. 2015. 73 f. Dissertação (Mestrado em Ciência Florestal) – Universidade dos Vales do Jequitinhonha e Mucuri, Diamantina. 2015.
- ARAÚJO, B. H. P. de. **Modelagem da altura, volume e afilamento de *Calycophyllum spruceanum* Benth. empregando regressão e redes neurais artificiais**. Manaus-AM: INPA, 2015b, 53 p. Dissertação (Mestrado em Ciências de Florestas Tropicais) – Instituto Nacional de Pesquisas da Amazônia.
- BINOTI, D. H. B.; BINOTI, M. L. M. da S.; LEITE, H. G. Configuração de redes neurais artificiais para estimação do volume de árvores. **Ciência da Madeira**, v. 5, n. 1, p. 58-67, 2014.
- BINOTI, M. L. M. da S.; BINOTI, D. H. B.; LEITE, H. G. Aplicação de redes neurais artificiais para estimação da altura de povoamentos equiâneos de eucalipto. **Revista Árvore**, v. 37, n. 4, p. 639-645, 2013
- BINOTI, M. L. M. **Redes neurais artificiais para prognose da produção de povoamentos não desbastados de eucalipto**. Viçosa-MG: UFV, 2010, 54 p. Dissertação (Mestrado em Ciências Florestais) – Universidade Federal de Viçosa
- BRAGA, A. P.; CARVALHO, A. P. L. F.; LUDERMIR, T. B. **Redes neurais artificiais: teoria e aplicações**. 2. ed.- [Reimpr.]. LTC, Rio de Janeiro, 2014, 26 p.
- CAMPOS, J. C. C.; LEITE, H. G. **Mensuração florestal: perguntas e respostas**. 4.ed., Viçosa: Editora UFV, 2013. 605p.
- COULSON, R.N.; FOLSE, L.J.; LOH, D. K. Artificial intelligence and natural resource management. **Science**, v. 237, n. 4812, p. 262-267, 1987.
- FERNANDES, Moabe Ferreira; QUEIROZ, Luciano Paganucci de. Vegetação e flora da Caatinga. **Ciência e cultura**, v.70, n.4, p.51-56, 2018.
- FIGUEIREDO, E. O.; SCHROEDER, R.; PAPA, D. de A. Fatores de Forma para 20 Espécies Florestais comerciais da Amazônia. **Comunicado Técnico**, n. 173. Embrapa, Rio Branco, AC, 2009.
- FIGUEIRÔA, J. M. et al., Madeireiras, In: SAMPAIO, E. V. S. B.; et al. **Espécies da Flora Nordestina de Importância Econômica Potencial**. Recife: APNE, 2005, p. 101-133.
- GOMES, S. H. M. **Redes neurais artificiais para estimativa de volume de espécies da caatinga arbórea**. 2017. 125 f. Dissertação (Mestrado em Ciências Florestais) - Universidade Estadual do Sudoeste da Bahia, Vitória da Conquista. 2017.

GORGENS, E. B.; LEITE, H. G.; SANTOS, H. do N.; GLERIANI, J. M. Estimação do volume de árvores utilizando redes neurais artificiais. **Revista Árvore**, v. 33, n. 6, p.1141-1147, 2009.

HAYKIN, S. **Neural networks and learning machines**. 3. ed. Prentice Hall, New York, 2009, 906 p.

HAYKIN, S. **Redes neurais: princípios e prática**. 2. ed. Bookman, Porto Alegre, 2001, 900 p.

IBAMA – Instituto Brasileiro de Meio Ambiente e dos Recursos Naturais Renováveis. **Plano de Manejo: Floresta Nacional Contendas do Sincorá, Volume I**. MMA, Brasília, 2006.

KRAMER, O. **Genetic Algorithms Essentials**. International publishing, 2017

LEITE, H. G. et al. Modelo de afilamento de cerne de *Tectona grandis* L.f. Revista Scientia Forestalis, **Piracicaba**, v. 39, n. 89, p.53-59, mar. 2011.

MACHADO, S.A.; FIGUEIREDO FILHO, A. **Dendrometria**, 2ª.ed., 1.ª reimpressão, Guarapuava:UNICENTRO, 2009. 309p.

REZENDE, A. V.; VALE, A. T.; SANQUETTA, C. R.; FIGUEIREDO FILHO, A.; FELFILI, J. M. Comparação de modelos matemáticos para estimativa do volume, biomassa e estoque de carbono da vegetação lenhosa de um cerrado sensu stricto em Brasília, DF. **Scientia Forestalis**, n. 71, p. 65-76, 2006.

RIZZINI, C.T. **Tratado de Fitogeografia do Brasil: Aspectos ecológicos, sociológicos e florísticos**. 2. ed. Âmbito Cultural, Rio de Janeiro, Volume único ,1997, 747 p.

SANQUETTA, C. R.; DALLA CORTE, A. P.; RODRIGUES, A. L.; WATZLAWICK, L. F. **Inventários Florestais: planejamento e execução**. 3. ed. Multi-graphic gráfica e editora, Curitiba, 2014, 406 p.

SANQUETTA, C. R.; DALLA CORTE, A. P.; JACON, A. D. Crown area and trunk diameter relationship for tree species at a mixed-araucaria natural forest in the midsouthern Parana state, Brazil. **Floresta**, v. 41, n. 1, p. 63-72, 2011.

SANTOS, A. T. dos; MATTOS, P. P. de; BRAZ, E. M.; ROSOT, N. C. Equação de volume e relação hipsométrica em plantio de *Ocotea porosa*. **Pesquisa florestal brasileira**, v. 32, n. 69, p. 13-21, 2012.

SCOLFORO, J. R.; RUFINI, A. L.; MELLO, J. M. de; OLIVEIRA, A. D. de; CASTRO SILVA, C. P. de; TRUGILHO; P. F. Equações de volume, peso de matéria de seca e produção de tanino do angico-vermelho (*Anadenanthera colubrina* (Benth) Brenan), em Minas Gerais. In: SCOLFORO, J. R.; OLIVEIRA, A. D.; ACERBI JÚNIOR, F. W. (Ed.). **Inventário Florestal de Minas Gerais - Equações de Volume, Peso de Matéria Seca e Carbono para Diferentes Fisionomias da Flora Nativa**. Lavras: UFLA, 2008, p. 117-128.

SILVA, I. N. d. et al. Artificial Neural Networks : A Practical Course. 1. ed. [S.l.]: Springer International Publishing, 2017. ISBN 978-3-319-43162-8,978-3-319-43161-1. Citado 3 vezes nas páginas 29, 30 e 36.

Silva Júnior, E. M. da, Maia, R. D., & Cabacinha, C. D. (2018). Bee-inspired RBF network for volume estimation of individual trees. *Computers and Electronics in Agriculture*, 152, 401–408. <https://doi.org/10.1016/j.compag.2018.07.036>.

SOARES, C. P. B.; NETO, F. de P.; SOUZA, A. L. de (Ed.). **Dendrometria e Inventário Florestal**. 2. ed. UFV, Viçosa, 2011a, 272 p.

SOUZA, A. L. de; SOARES, C. P. B. **Florestas nativas**: estrutura dinâmica e manejo. UFV, Viçosa, 2013, 322 p.

SOUZA, H. S. et al. Comparação de métodos de cubagem para eucalipto. *Rev. Bras. Biom.*, Lavras, v.35, n.1, p.17-26, 2017.

STANLEY, K. O.; MIIKKULAINEN, R. Evolving Neural Networks through augmenting Topologies. **Evolutionary Computation**, v. 10, n. 2, p. 99-127, 2002.