

UNIVERSIDADE FEDERAL DE SANTA MARIA  
CENTRO DE TECNOLOGIA  
CURSO DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

Júlia Gabriela Santi Acosta

**SISTEMA WEB COM APLICAÇÃO DE REDES NEURAIIS PARA  
ESTIMAR PALEOTEMPERATURAS COM BASE EM ASSOCIAÇÕES  
DE FORAMINÍFEROS**

Santa Maria, RS  
2022

**Júlia Gabriela Santi Acosta**

**SISTEMA WEB COM APLICAÇÃO DE REDES NEURAIIS PARA ESTIMAR  
PALEOTEMPERATURAS COM BASE EM ASSOCIAÇÕES DE FORAMINÍFEROS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação, Área de Concentração em Ciências Exatas e da Terra, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Sistemas de Informação**. Defesa realizada por videoconferência.

ORIENTADOR: Prof. Joaquim Vinicius Carvalho Assunção

Santa Maria, RS  
2022

**Júlia Gabriela Santi Acosta**

**SISTEMA WEB COM APLICAÇÃO DE REDES NEURAIIS PARA ESTIMAR  
PALEOTEMPERATURAS COM BASE EM ASSOCIAÇÕES DE FORAMINÍFEROS**

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Sistemas de Informação, Área de Concentração em Ciências Exatas e da Terra, da Universidade Federal de Santa Maria (UFSM, RS), como requisito parcial para obtenção do grau de **Bacharel em Sistemas de Informação**.

**Aprovado em 15 de fevereiro de 2022:**

  
\_\_\_\_\_  
**Joaquim Vinicius Carvalho Assunção, Dr. (UFSM)**  
(Presidente/Orientador)

  
\_\_\_\_\_  
**Maria Alejandra Gomez Pivel, Dra. (UFRGS) (videoconferência)**

  
\_\_\_\_\_  
**Jonas Bulegon Gassen, Dr. (UFSM) (videoconferência)**

Santa Maria, RS  
2022

## RESUMO

# SISTEMA WEB COM APLICAÇÃO DE REDES NEURAIIS PARA ESTIMAR PALEOTEMPERATURAS COM BASE EM ASSOCIAÇÕES DE FORAMINÍFEROS

AUTORA: Júlia Gabriela Santi Acosta

ORIENTADOR: Joaquim Vinicius Carvalho Assunção

Esta monografia refere-se ao desenvolvimento de um sistema relacionado à predição de paleotemperaturas a partir da análise de foraminíferos, que são organismos unicelulares envoltos em uma concha encontrados em todos os ambientes marinhos. Quando o ciclo do foraminífero acaba, essas conchas são depositadas no fundo do oceano, transformando-se em fósseis que têm sido usados para investigar as mudanças ambientais ocorridas ao longo de milhões de anos. Para desenvolver estudos sobre estas conchas podemos usar técnicas de Redes Neurais Artificiais (RNA). Muitos trabalhos recentes que fazem a utilização de RNA envolvendo foraminíferos, se centram no reconhecimento de imagens, porém, nesta monografia o objeto de treinamento da rede foi diferente, pois foi utilizado um conjunto de dados com as informações de abundâncias relativas de foraminíferos juntamente com a respectiva temperatura de seu *habitat*. Para permitir a utilização destas técnicas por pesquisadores, foi criado um sistema com a linguagem R e o pacote *shiny*, que utiliza técnicas de RNA para prever a temperatura dos oceanos com base na abundância relativa das espécies de foraminíferos, utilizando-se de configurações diferentes de RNA dependendo da profundidade e período do ano. Para avaliar os modelos, rodadas de testes com diversas configurações de RNA foram feitas, com objetivo de buscar o modelo mais adequado para cada característica, foram salvos 21 modelos de RNA no sistema para o usuário poder utilizar na predição das paleotemperaturas dependendo da configuração desejada. Também é possível utilizar uma segunda opção, a predição baseada no método de árvore ANOVA. Esta monografia atingiu seu objetivo já que visava encontrar modelos iniciais com resultados satisfatórios para estimar paleotemperaturas através de um sistema web.

**Palavras-chave:** Predição, Aprendizagem de Máquina, Inteligência Artificial, Árvore ANOVA.

## ABSTRACT

# WEB SYSTEM WITH APPLICATION OF NEURAL NETWORKS TO ESTIMATE PALEOTEMPERATURES BASED ON FORAMINIFERA ASSOCIATIONS

AUTHOR: Júlia Gabriela Santi Acosta

ADVISOR: Joaquim Vinicius Carvalho Assunção

This monograph is centered on the development of a system related to the prediction of paleotemperatures from the analysis of foraminifera, which are single-celled shell-encased organisms found in all marine environments. When the foraminifera cycle ends, these shells are deposited in the deep of the ocean, turning into fossils that have been used to investigate environmental changes over millions of years. To develop studies on these shells we can use Artificial Neural Network (ANN) techniques. Many recent works that use ANN involving foraminifera focus on image recognition, however, in this monograph, the network training object was a dataset with the relative abundances information of foraminifera that was used along with the respective temperature of its habitat. To allow the use of these techniques by researchers, a system with the R language and the *shiny* package was created, which uses ANN techniques to predict the temperature of the oceans based on the relative abundance of foraminifera species, using different configurations of ANN depending on depth and period of the year. To evaluate the models, rounds of tests with different ANN configurations were carried out, in order to find the most suitable model for each characteristic, 21 ANN models were saved in the system for the researcher to use in the prediction of paleotemperatures depending on the desired configuration. It is also possible to use a second option, the prediction based on the ANOVA tree method. This monograph achieved its objective as it aimed to find initial models with satisfactory results to estimate paleotemperatures through a web system.

**Keywords:** Prediction, Machine Learning, Artificial Intelligence, ANOVA Tree.

## LISTA DE FIGURAS

|  |    |
|--|----|
| Figura 2.1 – Foraminíferos Planctônicos .....  | 12 |
| Figura 2.2 – Províncias de foraminíferos planctônicos no oceano moderno e temperatura superficial do mar (TSM) ..... | 13 |
| Figura 2.3 – Coleta de testemunhos .....   | 14 |
| Figura 2.4 – Rede Neural Biológica .....   | 16 |
| Figura 2.5 – Arquitetura básica de RNA .....   | 17 |
| Figura 2.6 – Objetos linearmente separáveis .....  | 18 |
| Figura 2.7 – Redes Neurais Multicamadas .....  | 19 |
| Figura 2.8 – Objetos não-linearmente separáveis .....  | 19 |
| Figura 2.9 – Outliers e Inliers .....  | 25 |
| Figura 4.1 – Treinamento da Rede Neural inicial .....  | 29 |
| Figura 4.2 – Caso 1: Primeira parte .....  | 29 |
| Figura 4.3 – Caso 1: Segunda parte .....   | 30 |
| Figura 4.4 – Caso 1: Terceira parte .....  | 30 |
| Figura 4.5 – Caso 1: Quarta parte .....  | 31 |
| Figura 4.6 – Caso 2: Primeira parte .....  | 32 |
| Figura 4.7 – Caso 2: Segunda parte .....   | 32 |
| Figura 4.8 – Caso 2: Terceira parte .....  | 33 |
| Figura 4.9 – <i>Anova Tree</i> .....   | 34 |
| Figura 4.10 – <i>Help</i> .....  | 35 |
| Figura 4.11 – <i>Neural Network Tab</i> .....  | 36 |
| Figura 4.12 – <i>Graphs Tab</i> : Comparação entre valores preditos e reais .....                                    | 36 |
| Figura 4.13 – <i>Graphs Tab</i> : Rede Neural criada pela aplicação .....  | 37 |
| Figura 5.1 – Diagrama de Atividades .....  | 39 |
| Figura 5.2 – Código em R para gerar a árvore .....   | 41 |
| Figura 5.3 – Árvore ANOVA gerada no sistema .....  | 41 |
| Figura 6.1 – Diagrama do principal fluxo do sistema (Caso 2 do tópico 4.3) .....                                     | 43 |

## LISTA DE TABELAS

|   |    |
|---|----|
| Tabela 2.1 – Exemplo com as classes de foraminíferos e diferentes temperaturas por parâmetro..... | 15 |
| Tabela 5.1 – Exemplo da planilha gerada para testar diferentes configurações .....                | 40 |
| Tabela 5.2 – Exemplo da planilha gerada com os menores erros para cada configuração ..            | 40 |

## LISTA DE ABREVIATURAS E SIGLAS

|             |                            |
|-------------|----------------------------|
| <i>RNA</i>  | Rede Neural Artificial     |
| <i>Tanh</i> | Tangente Hiperbólica       |
| <i>MSE</i>  | <i>Mean Square Error</i>   |
| <i>MAE</i>  | <i>Mean absolute error</i> |
| <i>TAM</i>  | Técnica do Análogo Moderno |
| <i>RNM</i>  | Rede Neural Multicamada    |
| <i>MLN</i>  | <i>Multilayer Network</i>  |
| <i>RNB</i>  | Redes Neurais Biológicas   |
| <i>ML</i>   | <i>Machine Learning</i>    |

## SUMÁRIO

|              |  |           |
|--------------|--|-----------|
| <b>1</b>     | <b>INTRODUÇÃO</b> .....  | <b>9</b>  |
| 1.1          | OBJETIVOS .....  | 10        |
| <b>1.1.1</b> | <b>Objetivos específicos</b> .....   | <b>10</b> |
| <b>2</b>     | <b>REFERENCIAL TEÓRICO</b> .....   | <b>11</b> |
| 2.1          | FORAMINÍFEROS .....  | 11        |
| <b>2.1.1</b> | <b>Coleta de amostras</b> .....  | <b>14</b> |
| <b>2.1.2</b> | <b>Características que influenciam na predição</b> .....                     | <b>15</b> |
| <b>2.1.3</b> | <b>Técnicas utilizadas para predições no que tange a foraminíferos</b> ..... | <b>15</b> |
| 2.2          | REDES NEURAS ARTIFICIAIS .....   | 15        |
| <b>2.2.1</b> | <b>Perceptron</b> .....  | <b>17</b> |
| <b>2.2.2</b> | <b>Rede Neural Multicamadas</b> .....  | <b>18</b> |
| <b>2.2.3</b> | <b>Funções de ativação</b> .....   | <b>20</b> |
| 2.2.3.1      | <i>Sigmóide</i> .....  | 20        |
| 2.2.3.2      | <i>Tanh</i> .....  | 20        |
| 2.2.3.3      | <i>ReLU</i> .....  | 21        |
| 2.2.3.4      | <i>Softmax</i> .....   | 21        |
| <b>2.2.4</b> | <b>Treino e aprendizagem</b> .....   | <b>21</b> |
| <b>2.2.5</b> | <b>Métricas de desempenho</b> .....  | <b>22</b> |
| 2.2.5.1      | <i>Acurácia</i> .....  | 23        |
| 2.2.5.2      | <i>Precisão</i> .....  | 23        |
| 2.2.5.3      | <i>Mean Squared Error (MSE)</i> .....  | 23        |
| 2.2.5.4      | <i>Mean Absolute Error (MAE)</i> .....                                       | 23        |
| 2.2.5.5      | <i>Root Mean Squared Error (RMSE)</i> .....                                  | 24        |
| 2.2.5.6      | <i>Relative Absolute Error (RAE)</i> .....                                   | 24        |
| <b>2.2.6</b> | <b>Árvore ANOVA</b> .....  | <b>25</b> |
| 2.3          | TRABALHOS RELACIONADOS E CONSIDERAÇÕES .....                                 | 25        |
| <b>3</b>     | <b>METODOLOGIA</b> .....   | <b>27</b> |
| <b>4</b>     | <b>DESENVOLVIMENTO</b> .....   | <b>28</b> |
| 4.1          | ESTRUTURA DA REDE NEURAL .....   | 28        |
| 4.2          | CASO 1: USUÁRIO CRIANDO SEU PRÓPRIO MODELO .....                             | 29        |
| 4.3          | CASO 2: USUÁRIO UTILIZANDO MODELOS PREVIAMENTE SALVOS ....                   | 31        |
| 4.4          | <i>MENU ANOVA TREE</i> .....   | 33        |
| 4.5          | <i>MENU HELP</i> .....   | 34        |
| 4.6          | MENUS AUXILIARES DO SISTEMA .....  | 35        |
| <b>5</b>     | <b>RESULTADOS E CONSIDERAÇÕES</b> .....                                      | <b>38</b> |
| 5.1          | REDES NEURAS: MODIFICAÇÃO DE PARÂMETROS .....                                | 38        |
| 5.2          | <i>ANALYSIS OF VARIANCE (ANOVA)</i> .....                                    | 41        |
| <b>6</b>     | <b>CONCLUSÃO</b> .....   | <b>42</b> |
| 6.1          | TRABALHOS FUTUROS .....  | 44        |
|              | <b>REFERÊNCIAS BIBLIOGRÁFICAS</b> .....                                      | <b>45</b> |

# 1 INTRODUÇÃO

Foraminíferos são organismos marinhos unicelulares que são envoltos em uma concha. Tais conchas demoram para se degradar a ponto de servir como registros fósseis (HAYWARD et al., 2021). Estes fósseis compõem registros sedimentares, que permitem, a partir de sua análise, construir ecossistemas passados (CHAPORI et al., 2015). A reconstrução destes ecossistemas permite conhecer a sequência exata de eventos paleoceanográficos, e isto é vital para compreender os mecanismos por trás das mudanças climáticas (PETRÓ et al., 2016), podemos verificar esta premissa no trabalho de Petró et al. (2016) e também em Sijinkumar, Nath e Clemens (2016).

Malmgren et al. (2001) citam técnicas para fazer a investigação de paleotemperaturas do mar utilizando foraminíferos. E este estudo traz um comparativo entre as técnicas, sugerindo que a técnica de Redes Neurais Artificiais (RNA<sup>1</sup>) teve melhor desempenho em relação às outras. Utilizamos RNA para fazer previsões de paleotemperaturas a partir da análise de foraminíferos, fornecendo um sistema que auxilia os pesquisadores a investigar a correlação e possíveis padrões entre diferentes espécies de foraminíferos, temperatura em determinados períodos do ano e profundidade do oceano.

Petró et al. (2016) salientam em seu trabalho que a temperatura da superfície do mar constitui o principal fator ambiental que controla a distribuição e abundância dos foraminíferos. O objetivo principal do sistema será prever a temperatura média para diferentes épocas do ano, e profundidade do oceano com base em uma determinada amostra representativa da associação de foraminíferos planctônicos. Também serão exploradas algumas questões que podem influenciar nestes dados, como foi feito no trabalho de Petró et al. (2016) que utilizaram um banco de dados com uma calibração específica para cada bacia oceânica com intuito de minimizar o erro introduzido por espécies crípticas.

O avanço tecnológico recente proporcionou o armazenamento de um grande volume de dados (*e.g.*, Big Data) e um poder computacional muito maior. Isto tornou possível a evolução dos modelos de RNA, não apenas permitindo a aplicação de conceitos que não eram viáveis com o poder computacional anterior, como também os tornando melhores (MIIKKULAINEN et al., 2019). Portanto, é esperada a criação de modelos de RNA capazes de obter resultados melhores para estimar as paleotemperaturas.

São poucos os trabalhos que envolvem RNA e foraminíferos, e utilizam *datasets* para fazer as previsões. A maioria deles estão na parte de classificação de imagens como foi feito por Johansen e Sørensen (2020), Johansen et al. (2021) e Mitra et al. (2019). Para a utilização da RNA, precisa-se de um entendimento de como ajustar sua arquitetura e parâmetros visando maximizar a taxa de acerto e generalização do modelo. Portanto, para obter um resultado satisfatório precisamos unir o entendimento sobre a utilização de RNA e foraminíferos. É nesta

interdisciplinaridade que a monografia estará centrada, provendo uma ferramenta que utilize as técnicas de RNA para ajudar os pesquisadores a realizar predições de paleotemperaturas de acordo com amostras de conchas foraminíferas.

## 1.1 OBJETIVOS

O trabalho objetiva o desenvolvimento de um sistema web para auxiliar os pesquisadores a obter informações a partir dos foraminíferos, testando novos modelos e configurações de RNA para estimar a paleotemperatura superficial do mar a partir de proporções de foraminíferos.

### 1.1.1 Objetivos específicos

- Desenvolver um sistema web que permita o pesquisador fazer a predição de uma determinada característica, focando principalmente na temperatura média anual.
- Considerar os hiperparâmetros de profundidade e período do ano para a construção da RNA.
- Testar diferentes configurações de redes conforme a profundidade e período do ano e utilizar a melhor configuração de cada hiperparâmetro.

---

<sup>1</sup>A sigla RNA será utilizado ao longo do texto, tanto para Redes Multicamadas, quanto para Perceptron, porém, quando nos referimos às Redes usadas nos modelos obtidos, no desenvolvimento e no resultado desta monografia designamos a sigla para Redes Neurais Multicamadas.

## 2 REFERENCIAL TEÓRICO

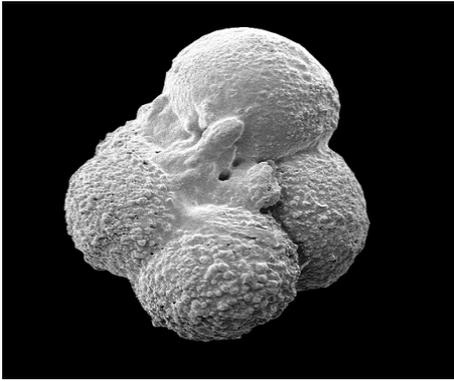
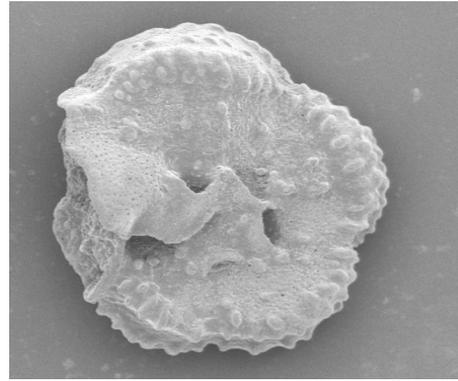
Esta seção contém uma explicação mais detalhada sobre assuntos necessários para o entendimento da monografia. O tópico 2.1 de foraminíferos visa a fundamentação teórica para entender mais sobre os foraminíferos e como funciona o estudo sobre eles, já no tópico 2.2 sobre RNA será o embasamento para criar os modelos de predição. Entendendo melhor sobre esses dois assuntos teremos a base teórica para, a partir de informações sobre as associações de foraminíferos, deduzir uma informação paleoambiental, por fim no tópico 2.3 são abordados alguns trabalhos que envolvem os temas desta monografia e também algumas considerações em relação aos tópicos anteriores.

### 2.1 FORAMINÍFEROS

Foraminíferos são microorganismos encontrados em todos os ambientes marinhos, eles possuem uma carapaça constituída principalmente de carbonato de cálcio ( $CaCO_3$ ), e também possuem diferentes tipos de estruturas que são formadas com a influência de um conjunto de fatores abióticos do ambiente. (MURRAY, 2006; HAYWARD et al., 2021). Quando o ciclo de vida dos foraminíferos acaba, essas carapaças são depositadas no fundo do oceano, e acabam por fazer parte do sedimento marinho. Estes fósseis de foraminíferos são estudados para reconstruir ecossistemas passados, eles têm sido usados para investigar as mudanças ambientais ocorridas ao longo de milhões de anos. (MURRAY, 2006).

Foraminíferos são organismos pertencentes ao Reino Chromista, Filo Retaria e Sub-filo Foraminifera, com indicações para seu surgimento desde o início do período Cambriano (BOUDAGHER-FADEL, 2018). Quanto ao seu modo de vida podem ser classificados em bentônicos - caracterizados por viverem no fundo do mar, e os planctônicos - que vivem na coluna d'água, geralmente até mil metros de profundidade. Eles são classificados principalmente em relação à composição e morfologia da carapaça (University College London, 2002). Nas imagens da figura 2.1 é possível verificar foraminíferos de modo de vida planctônicos, que são o foco do estudo nesta monografia, e algumas de suas classificações específicas.

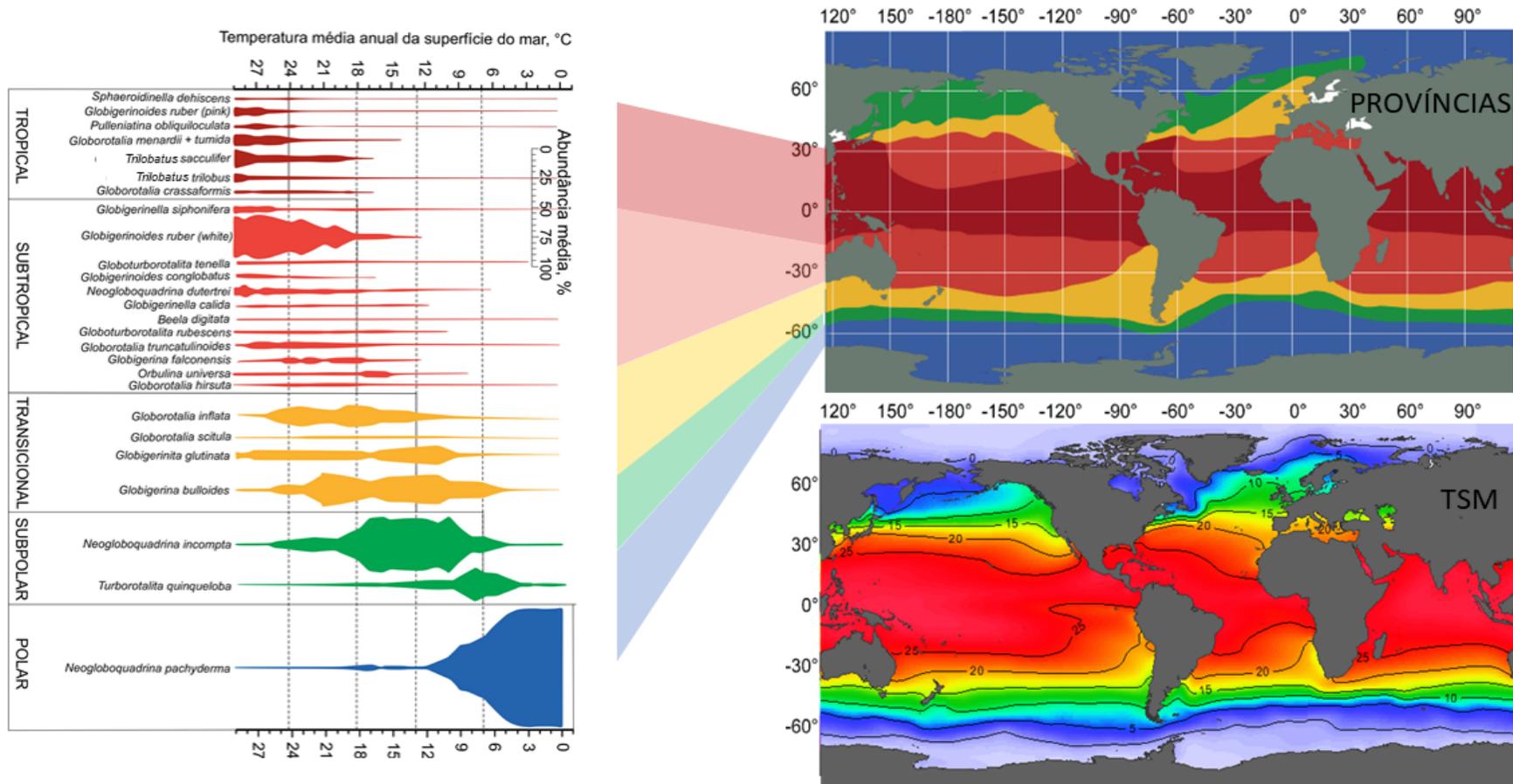
Figura 2.1 – Foraminíferos Planctônicos

(a) *Hedbergella sigali* Moullade(b) *Abathomphalus mayaroensis* (Bolli)

Fonte: University College London (2002)

As proporções de amostras de foraminíferos usadas para predição são de foraminíferos planctônicos, estas espécies têm sua distribuição controlada principalmente pela temperatura (Figura 2.2). Como os foraminíferos produzem carapaças, eles têm um alto potencial para se preservar nos sedimentos. Desta forma, coletando amostras antigas de sedimentos, podemos deduzir a temperatura superficial do oceano no passado, isto é, a paleotemperatura. Também vale a pena mencionar a importância de reconstruir essa variável, já que a temperatura é o principal parâmetro do sistema climático. Isto pode ser verificado no trabalho de Petró et al. (2016) que consiste em reconstruir mudanças paleoceanográficas, partindo da análise dos foraminíferos planctônicos obtidos. A técnica utilizada no trabalho baseia-se na premissa de que a temperatura é o principal fator que controla a distribuição da fauna.

Figura 2.2 – Províncias de foraminíferos planctônicos no oceano moderno e temperatura superficial do mar (TSM)

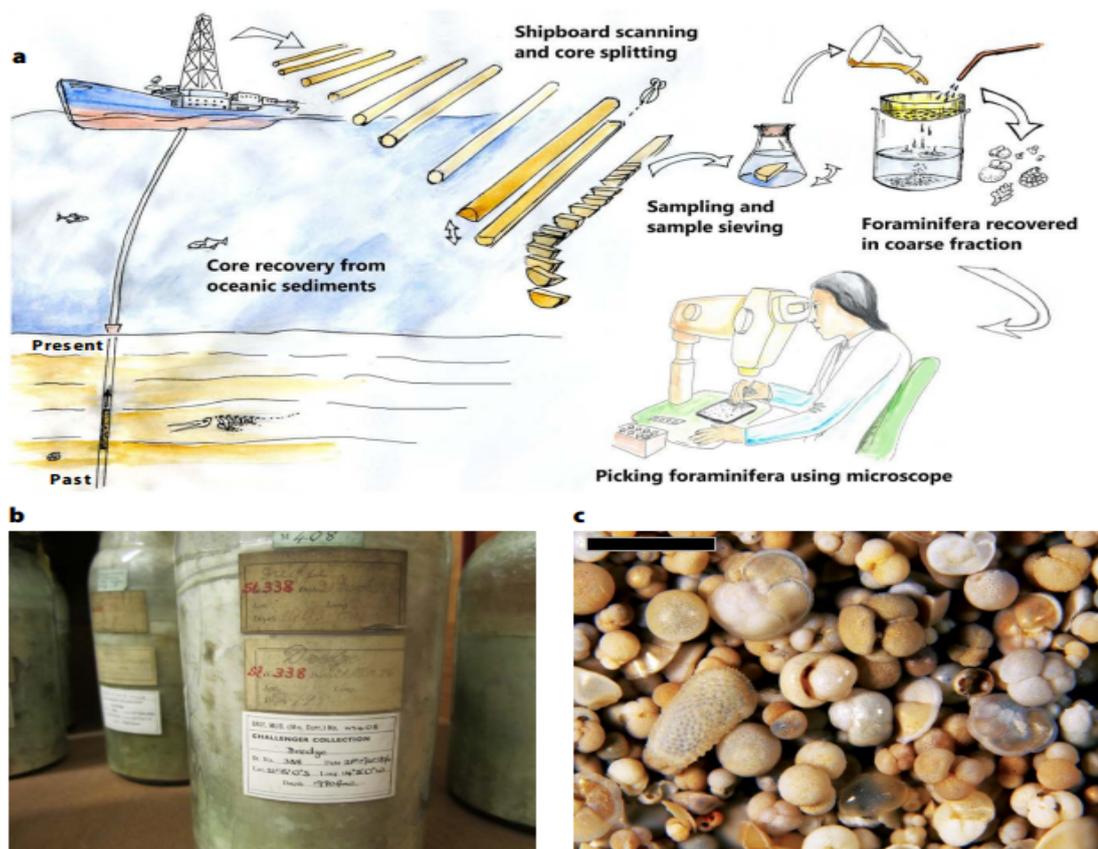


Fonte: Kucera (2007)

### 2.1.1 Coleta de amostras

As amostras de foraminíferos costumam ser obtidas a partir da extração de sedimentos do fundo do oceano em um determinado ponto de coleta que é definido pelos pesquisadores. Neste local definido, um testemunhador extrai um cilindro vertical de sedimentos conhecido como testemunho. Depois da extração, há uma série de etapas, o sedimento retirado é lavado, peneirado, seco e pesado e é feita a triagem, por fim, os microfósseis são colocados em lâminas e os foraminíferos são identificados. A Figura 2.3 feita por Yasuhara et al. (2020) exemplifica o funcionamento da coleta de amostras. O fluxo citado acima seria para uma coleta manual das amostras, no caso dos dados de foraminíferos planctônicos utilizados para o treinamento das RNA nesta monografia, foram obtidos do banco de dados ForCenS (SICCHA; KUCERA, 2017) e de García Chapori e Kucera (2019).

Figura 2.3 – Coleta de testemunhos



Fonte: Yasuhara et al. (2020)

### 2.1.2 Características que influenciam na predição

Petró et al. (2016) descrevem que utilizaram estimativas de inverno, verão e temperaturas médias anuais, pois as abundâncias de espécies de foraminíferos variam em função da temperatura, a qual, por sua vez, varia ao longo do ano. No trabalho de Hayes et al. (2005) também é feita essa comparação sobre períodos de inverno e verão, inferindo que a predição de temperatura com as técnicas utilizadas foi mais bem sucedida no inverno. No trabalho de Petró et al. (2016) foi apresentado um diagrama com a distribuição das amostras dos foraminíferos planctônicos com diferentes profundidades, Sijinkumar, Nath e Clemens (2016) também falam da questão de diferentes profundidades de coleta de sedimentos de foraminíferos. Essas são características que vão diferir na hora da coleta e geralmente são transcritas na planilha, podendo ter resultados de temperatura diferentes, como no exemplo ilustrativo da tabela 2.1 que foi baseado na planilha utilizada para treinamento das RNA.

Tabela 2.1 – Exemplo com as classes de foraminíferos e diferentes temperaturas por parâmetro.

| Globorotalia crassaformis | Neogloboquadrina incompta | annual-0m | jan-mar-75m |
|---------------------------|---------------------------|-----------|-------------|
| 0.4                       | 0.6                       | 14.56     | 12.34       |
| 0.2                       | 0.8                       | 10.50     | 14.44       |
| 0.7                       | 0.3                       | 11.56     | 13.37       |

Fonte: Autora

### 2.1.3 Técnicas utilizadas para predições no que tange a foraminíferos

Uma técnica muito utilizada na previsão das características dos foraminíferos é a Técnica do Análogo Moderno (TAM). O trabalho do Donato et al. (2018) explica que na TAM é feita uma comparação entre as amostras de fósseis de um banco de dados com amostras modernas usando um coeficiente de similaridade. Para tal, as estimativas paleoambientais são obtidas a partir dos parâmetros ambientais que foram medidos no local das amostras modernas.

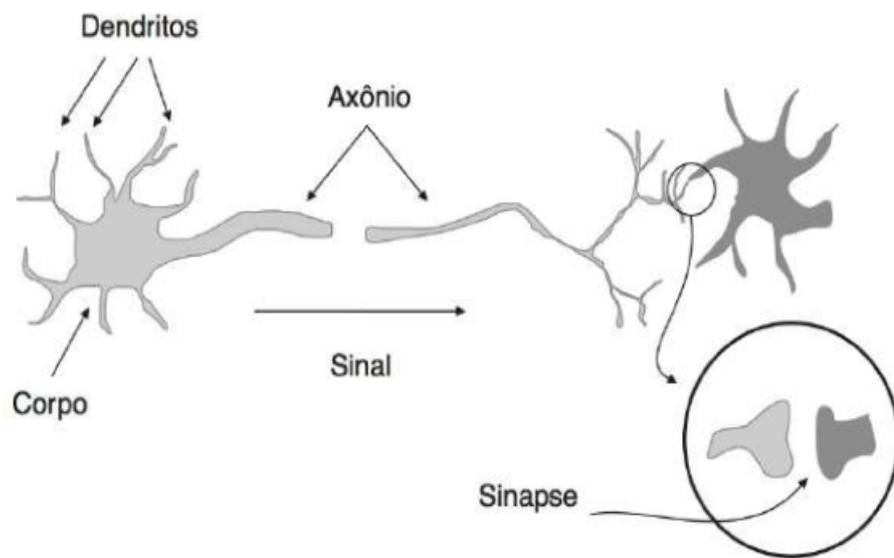
Malmgren et al. (2001), fizeram um comparativo entre técnicas utilizadas em previsões paleoambientais, além da TAM, foram referenciadas as técnicas: *Imbrie-Kipp transfer function*, *SIMMAX*, *Revised analog method* e também das RNA onde é sugerido que as RNA poderiam estimar melhor que as outras técnicas.

## 2.2 REDES NEURAIIS ARTIFICIAIS

Redes neurais artificiais são técnicas de aprendizado de máquina que simulam o mecanismo de aprendizado de organismos biológicos (AGGARWAL, 2018). O cérebro possui cerca

de 100 bilhões de neurônios, estes neurônios são considerados unidades básicas do sistema nervoso, e é através dos sinais elétricos que eles transmitem de uma extremidade a outra formando circuitos que podem processar informações recebidas e fornecer uma resposta. Estes sinais elétricos são transmitidos dos dendritos aos terminais do axônio através do corpo do axônio e continuam a ser transmitidos através da sinapse de um neurônio para outro (figura 2.4). O funcionamento do sistema nervoso depende deste grupos de neurônios que trabalham juntos. (HAN et al., 2018)

Figura 2.4 – Rede Neural Biológica



Fonte: Faceli et al. (2011)

Nas RNA, a saída não é fornecida constantemente, ela é gerada quando a entrada excede um certo limite, a função que recebe um sinal de entrada e produz um sinal de saída após um certo valor limite é chamada de função de ativação (2.2.3). Em síntese, os neurônios recebem sinais e geram outros sinais, assim como neurônios biológicos recebem vários inputs de outros neurônios do qual farão a sinapse, os neurônios das redes neurais, também recebem várias entradas, então eles as adicionam e processam a soma com uma função de ativação. Então poderíamos sintetizar dizendo que tanto as RNA quanto as Redes Neurais Biológicas (RNB) recebem dados de entrada, realizam um processamento e fornecem uma saída. Sendo assim, as respostas de neurônios biológicos e neurônios artificiais são semelhantes (HAN et al., 2018).

Portanto, é justo sugerir o sucesso de inspiração biológica para criação das redes neurais e, à medida que aprendemos mais sobre como o cérebro humano funciona isto pode ser replicado em outras configurações. (AGGARWAL, 2018).

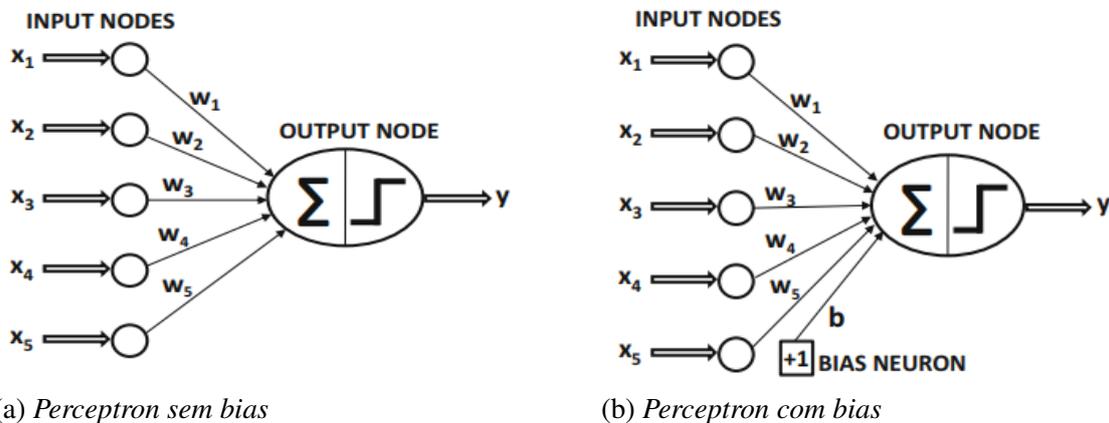
A RNA tem vantagens em relação ao aprendizado de máquina tradicional, além de que por permitir mais opções de arquitetura, ela proporciona uma abstração maior para expressar compreensão semântica. Ela também fornece uma maneira simples de ajustar a complexidade

de um modelo, adicionando ou removendo neurônios da arquitetura de acordo com a disponibilidade de dados de treinamento ou poder computacional (AGGARWAL, 2018).

### 2.2.1 Perceptron

A arquitetura do Perceptron proposta por Rosenblatt é mostrada na figura 2.5, na qual uma única camada de entrada transmite as informações para o nó de saída, os valores de entrada vão depender das informações relacionadas à predição que se deseja fazer, e a conexão entre a entrada e a saída contém os pesos  $w_1...w_d$  responsáveis por armazenar o conhecimento adquirido pela rede, que pode ser inicialmente definido com base em algum critério ou de forma aleatória, estes pesos serão multiplicados com os valores de entrada e adicionados no nó de saída (2.5a). Em muitas configurações, há uma parte da previsão, que é conhecida como *bias*, que é utilizada para casos em que a abordagem inicial não é suficiente para a previsão. Para isto precisamos incorporar uma variável de polarização adicional  $b$  que captura essa parte invariável da previsão (2.5b).

Figura 2.5 – Arquitetura básica de RNA



Fonte: Aggarwal (2018)

Desta forma, a função de entrada de um neurônio é definida pela soma ponderada das entradas, representada na equação:

$$\sum_{i=0}^n a_i w_i \quad (2.1)$$

Para casos que há a necessidade da utilização de *bias* a equação da função de entrada será:

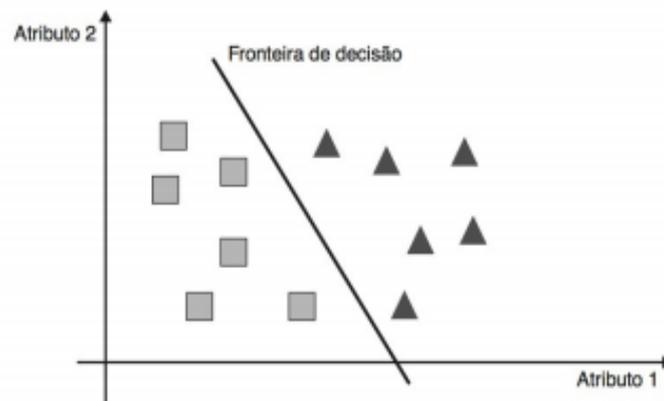
$$\sum_{i=0}^n a_i w_i + b \quad (2.2)$$

Depois de passar pela função de entrada, o resultado passa por uma função de ativação

$f$  cujo resultado define a ativação de uma determinada saída, mais detalhes sobre funções de ativação estão descritos no tópico 2.2.3.

Vale ressaltar que apesar da perceptron conter duas camadas, a camada de entrada não realiza nenhum cálculo, ela apenas transmite os valores. Então a camada de entrada não é incluída na contagem do número de camadas em uma rede neural, por isto a perceptron é considerada *single-layer network*. A maioria dos modelos básicos de aprendizado de máquina podem ser facilmente representados como arquiteturas de rede neural simples utilizando o modelo perceptron. No entanto, o modelo limita-se a problemas que são linearmente separáveis, ou seja, perceptrons só funcionam para problemas onde podemos traçar uma reta no plano cartesiano bidimensional que nos permite obter duas classes distintas, exemplificado na figura 2.6 (AGGARWAL, 2018).

Figura 2.6 – Objetos linearmente separáveis



Fonte: Faceli et al. (2011)

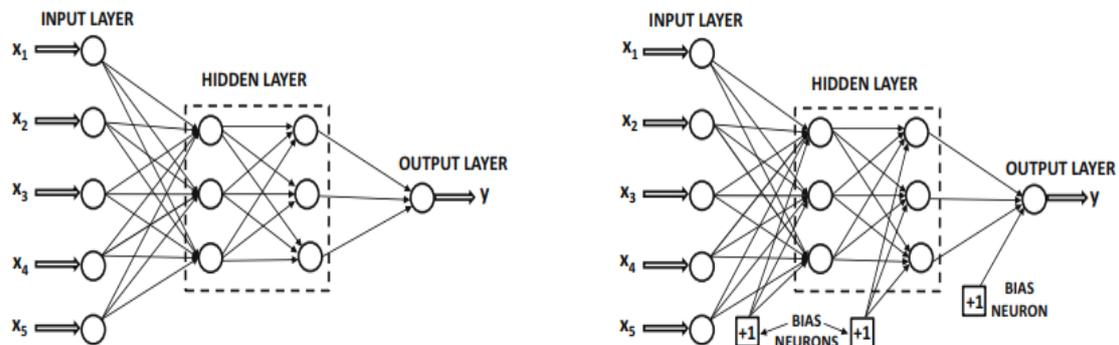
## 2.2.2 Rede Neural Multicamadas

Enquanto o modelo Perceptron contém uma camada de entrada e outra de saída, em que a camada de saída é a única camada que realiza cálculos completamente visíveis, as Redes Neurais Multicamadas (RNM) contêm várias camadas computacionais. As camadas intermediárias adicionais (entre entrada e saída) são conhecidas como camadas ocultas porque os cálculos realizados não são visíveis.

RNM são conhecidas como redes *feed-forward*, por ter camadas sucessivas que alimentam uma a outra, iniciando na entrada e indo para a saída. A arquitetura padrão de redes *feed-forward* assume que todos os nós em uma camada são conectados àquelas da próxima camada. Exemplos de redes multicamadas com ou sem *bias* são mostrados na Figura 2.7 (a) e (b), respectivamente. Em cada caso, a Rede Neural contém três camadas, como citado na

seção 2.2.1, a camada de entrada não é contada, porque ela simplesmente transmite os dados e nenhum cálculo é realizado nessa camada.

Figura 2.7 – Redes Neurais Multicamadas



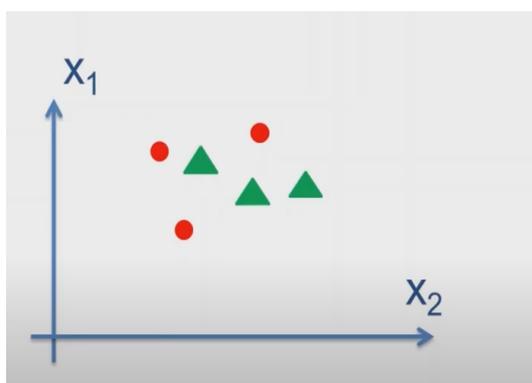
(a) RNM sem bias

(b) RNM com bias

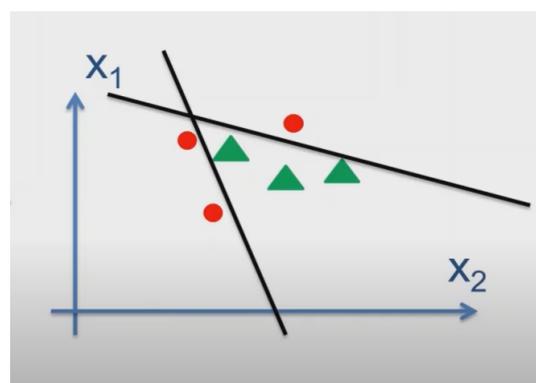
Fonte: Aggarwal (2018)

Diferente das perceptrons, podemos utilizar as RNM para resolver problemas não lineares. Em problemas não-lineares exemplificado na figura 2.8 vamos precisar de diversos neurônios para compor a fronteira de decisão entre as classes, enquanto nos problemas lineares apenas traçamos uma fronteira de decisão (Figura 2.6), para isso vamos ter que usar uma RNA que tenha mais neurônios, os neurônios das camadas intermediárias serão usados para compor a fronteira de decisão (Figura 2.8b), é neste caso que usamos a RNM, também chamada de *multilayer network* (MLN) (AGGARWAL, 2018).

Figura 2.8 – Objetos não-linearmente separáveis



(a) Sem fronteiras



(b) Com fronteiras

Fonte: Cérebro Eletrônico IFSP<sup>1</sup>

<sup>2</sup>Disponível em <<https://www.youtube.com/watch?v=ImB05A7Sdvc>> Acesso em 15 de novembro de 2021.

### 2.2.3 Funções de ativação

A escolha da função de ativação é uma parte importante do projeto da rede neural (AGGARWAL, 2018). As funções de ativação permitem que pequenas mudanças nos pesos e *bias* causem apenas uma pequena alteração no *output*, esse é o fato crucial que permitirá que uma RNA aprenda. Elas também decidem se um neurônio deve ser ativado ou não, ou seja, se a informação que o neurônio está recebendo é relevante. A função de ativação é a transformação não linear que fazemos ao longo do sinal de entrada, esta saída transformada é então enviada para a próxima camada de neurônios como entrada.

Quando não temos a função de ativação, os pesos e bias simplesmente fazem uma transformação linear, uma equação linear é simples de resolver, mas é limitada na sua capacidade de resolver problemas complexos. Uma rede neural sem função de ativação é essencialmente apenas um modelo de regressão linear, a função de ativação faz a transformação não-linear nos dados de entrada, tornando-o capaz de aprender e executar tarefas mais complexas. (DATA SCIENCE ACADEMY, 2020)

Existem vários tipos de função de ativação, cada qual a ser usado em diferentes situações. Os mais populares são: Sigmóide, Tanh, ReLU e Softmax, todos eles detalhados na continuação.

#### 2.2.3.1 Sigmóide

A função sigmóide é amplamente utilizada até hoje, ela varia de 0 a 1 tendo um formato de S, se centra essencialmente em tentar empurrar os valores de Y para os extremos, isto é algo muito desejável quando temos que classificar valores para uma classe específica. Outro aspecto interessante dela é que quando temos vários neurônios com a função sigmóide como função de ativação, a saída não é linear. Sua fórmula é representada por:

$$f(x) = \frac{1}{(1 + e^{-x})} \quad (2.3)$$

Podemos ter alguns problemas com a sigmóide quando os gradientes se tornam muito pequenos, isso significa que o gradiente está se aproximando de zero e a rede não está realmente aprendendo, outro problema que ela sofre é que os valores variam apenas de 0 a 1, mas nem sempre desejamos que os valores enviados ao próximo neurônio sejam todos do mesmo sinal.

#### 2.2.3.2 Tanh

Tanh funciona de forma semelhante à função sigmóide, é uma versão escalonada dela, a diferença é que ela varia de -1 a 1, e com isso acaba por solucionar o problema dos valores que

tem na sigmóide (2.2.3.1). Ela é representada pela equação:

$$f(x) = \frac{2}{(1 + e^{-2x}) - 1} \quad (2.4)$$

### 2.2.3.3 ReLU

ReLU é a função de ativação mais utilizada ao projetar RNA atualmente (DATA SCIENCE ACADEMY, 2020). Ela é não linear, o que significa que podemos facilmente copiar os erros para trás e ter várias camadas de neurônios ativados pela função.

A principal vantagem de usar a função ReLU sobre outras funções de ativação é que ela não ativa todos os neurônios ao mesmo tempo. Ou seja, para a função ReLU, se a entrada for negativa, ela será convertida em zero e o neurônio não será ativado. Isso significa que, apenas alguns neurônios são ativados, tornando a rede esparsa e eficiente. Mas, assim como a sigmóide (2.2.3.1), a ReLU também pode ter problemas com os gradientes que se deslocam em direção a zero. Sua equação é representada por:

$$f(x) = \max(0, x) \quad (2.5)$$

### 2.2.3.4 Softmax

A função softmax também é um tipo de função sigmóide, ela é útil quando tentamos lidar com problemas de classificação, enquanto a função sigmóide é capaz de lidar com apenas duas classes, a softmax pode lidar com várias classes. A função softmax transforma as saídas para cada classe para valores entre 0 e 1 e também divide pela soma das saídas, isso gera a probabilidade de a entrada estar em uma determinada classe (DATA SCIENCE ACADEMY, 2020). Pode ser definida como:

$$\frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \text{ para } j = 1, \dots, K \quad (2.6)$$

## 2.2.4 Treino e aprendizagem

Na fase de treino iniciamos com a seleção do conjunto de dados que será utilizado na RNA, isto é, um conjunto que contém informações dos neurônios de entrada e saídas esperadas para RNA. Esse conjunto será separado em dados de treino, compreendendo os dados usados

para que a rede aprenda as relações entre entrada e saídas esperadas, e os dados de teste, que serão os dados utilizados para validar o valor que a rede está predizendo, testando sua capacidade de generalização.

Na Rede Neural de camada única, o processo de treinamento é relativamente simples porque o erro pode ser calculado como uma função direta dos pesos, que permite fácil aplicação do gradiente. No caso de redes multicamadas, o problema é que a perda é uma função de composição complicada dos pesos nas camadas anteriores. O gradiente de uma função de composição é calculado usando o algoritmo *retropropagation* exemplificado nos próximos parágrafos (AGGARWAL, 2018).

No algoritmo de *retropropagation*, na etapa denominada *forward* os dados de entrada são inseridos na camada inicial e propagados de camada em camada, fazendo os cálculos usando o conjunto inicial de pesos, até chegarem na camada de saída, as saídas produzidas pela RNA são comparadas com as saídas esperadas e é realizado o cálculo do erro. Quanto menor for o erro, melhor será a predição da rede. O erro usado neste algoritmo é chamado de *Mean Squared Error* (MSE) e é representado pela equação:

$$\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (2.7)$$

Após o cálculo do erro, ele é propagado da camada de saída para a de entrada, numa etapa chamada de *backward*. É durante esta etapa que os pesos das conexões entre os neurônios de cada camada são ajustados, e é através do método de descida do gradiente que o ajuste com intuito de minimizar o erro é feito. Quanto menor ele for, mais próximo a saída da rede estará da resposta esperada (BRAGA; LUDERMIR; CARVALHO, 2000; AGGARWAL, 2018).

### 2.2.5 Métricas de desempenho

Escolher a métrica certa é muito importante para avaliar os modelos de *machine learning* (ML). Várias métricas são propostas para esta finalidade, em alguns modelos olhar para uma única métrica pode não fornecer a imagem completa do problema que está se tentando resolver, e pode ser necessário usar um subconjunto das métricas para ter uma avaliação concreta de seus modelos (Shervin Minaee, 2019).

As primeiras métricas citadas abaixo são mais utilizadas para casos em que o problema é de classificação, são exemplos desse problema: classificação de imagens, reconhecimento facial, restrição de conteúdo e classificação de texto.

### 2.2.5.1 Acurácia

É uma métrica muito utilizada e simples de calcular sendo definida como o número de previsões corretas dividido pelo número total de previsões, multiplicado por 100. É representada pela equação:

$$A = \left(\frac{Ta}{T}\right) * 100 \quad (2.8)$$

### 2.2.5.2 Precisão

Existem casos que a acurácia não é um bom indicador de desempenho, como quando é desequilibrada, ou seja, uma classe é mais frequente do que outras. Neste caso é possível ter uma acurácia alta, mas mesmo assim o resultado desejado sempre prever a mesma classe. Essa métrica é representada pela fórmula:

$$P = \frac{\textit{PositivoVerdadeiro}}{(\textit{PositivoVerdadeiro} + \textit{PositivoFalso})} \quad (2.9)$$

Já as métricas abaixo são utilizadas para casos relacionados a regressão, que seriam para prever valores alvo contínuos, tais como o problema proposto na monografia. Outros exemplos incluem: previsão de preços de residências, sistemas de preços de comércio eletrônico, previsão do tempo e previsão do mercado de ações.

### 2.2.5.3 Mean Squared Error (MSE)

MSE é talvez a métrica mais popular usada para problemas de regressão. Basicamente, usando essa fórmula encontramos o erro quadrático médio entre os valores previstos e reais. Essa métrica foi citada na seção 2.2.4 da monografia e é representada pela fórmula 2.7.

### 2.2.5.4 Mean Absolute Error (MAE)

MAE é uma métrica que encontra a distância absoluta média entre os valores previstos e desejados. O MAE é conhecido por ser mais robusto para os *outliers* (pontos que não se encaixam no modelo exemplificado na figura 2.9) do que o MSE. A principal razão é que no MSE, ao elevar os erros ao quadrado, os *outliers* (que geralmente apresentam erros maiores do que outras amostras) recebem mais atenção e dominância no erro final e isto impacta os

parâmetros do modelo (Shervin Minaee, 2019). É representada pela equação:

$$\frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2.10)$$

#### 2.2.5.5 Root Mean Squared Error (*RMSE*)

Como o próprio nome indica, é similar ao MSE, porém a raiz quadrada retorna a escala dos erros que foram elevados ao quadrado. É representado pela equação:

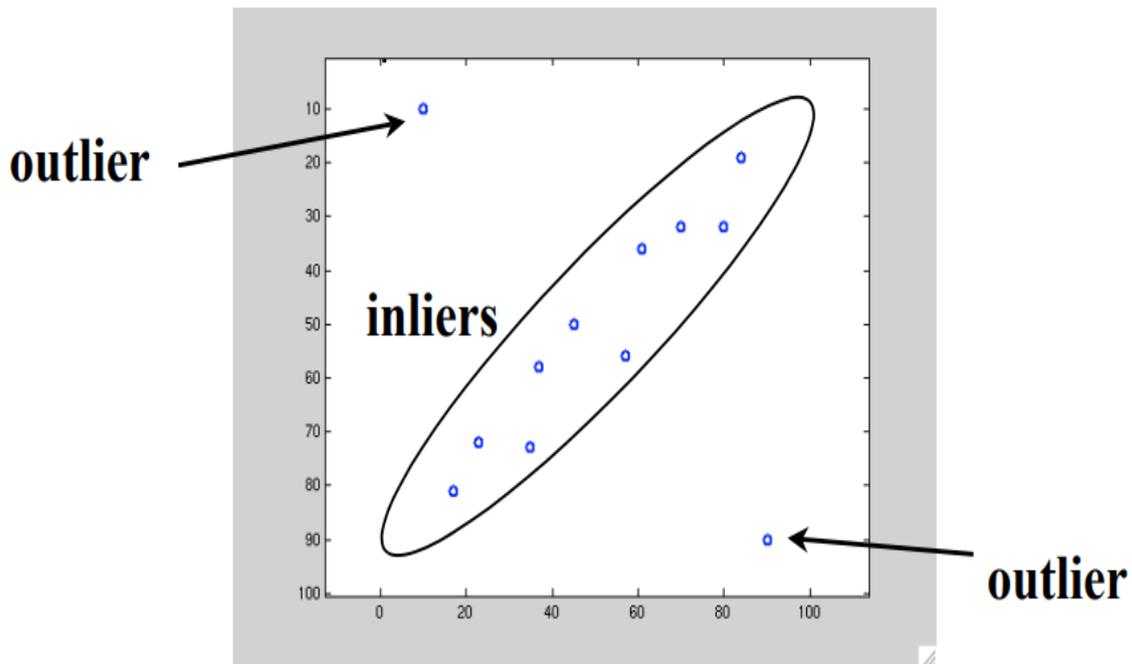
$$\sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2} \quad (2.11)$$

#### 2.2.5.6 Relative Absolute Error (*RAE*)

O RAE corresponde ao MAE relativizado pela divisão do MAE usando a média como valor subtraído no somatório. Pode ser calculado por:

$$\frac{\sum_{i=1}^n |Y_i - \hat{Y}_i|}{\sum_{i=1}^n |Y_i - \bar{Y}_i|} \quad (2.12)$$

Figura 2.9 – Outliers e Inliers



Fonte: Robert Collins<sup>2</sup>

### 2.2.6 Árvore ANOVA

ANOVA é um teste estatístico para detectar diferenças na média do grupo quando há uma variável dependente paramétrica ou uma ou mais variáveis independentes. Este teste é utilizado para verificar quando existe diferenças entre as médias de populações determinando quando a relação é significativa ou não.

Ela é usada para ajudar a entender como os diferentes grupos respondem, quando entendemos esta relação podemos começar a aprofundar o conhecimento nas conexões entre as médias e com isto verificar o que está impulsionando o comportamento.

## 2.3 TRABALHOS RELACIONADOS E CONSIDERAÇÕES

Muitos trabalhos envolvendo foraminíferos e RNA estão centrados no reconhecimento de imagens de conchas foraminíferas como em Johansen e Sørensen (2020), Johansen et al. (2021), Zhong et al. (2017), Marchant et al. (2020) e Mitra et al. (2019). Dentre os trabalhos que usam *dataset* e RNA estão Malmgren et al. (2001) e Hayes et al. (2005) ambos são uma

<sup>3</sup>Disponível em <<http://www.cse.psu.edu/rtc12/CSE486/lecture15.pdf>> Acesso em 30 de novembro de 2021.

comparação entre as técnicas para prever a temperatura a partir de dados de abundâncias relativas das diferentes espécies de foraminíferos. Hayes et al. (2005) inclusive cita o trabalho de Malmgren et al. (2001), pois é um dos trabalhos que envolvem RNA e foraminíferos que é referencia na área e tem resultados muito interessantes. Porém desde que estes trabalhos foram publicados, os modelos e resultados das redes neurais melhoraram muito, possibilitando criar uma RNA que consiga fazer previsões com maior taxa de acerto.

Diferente destes trabalhos citados anteriormente, em que o objetivo era a comparação entre as técnicas, nesta monografia o objetivo está focado em prover uma ferramenta para os pesquisadores, que facilite o uso de RNA em previsões relativas a foraminíferos.

Foi utilizada uma Rede Neural *feed-forward* Multicamada, pois redes *feed-forward* são usadas principalmente para aprendizagem supervisionada nos casos em que os dados a serem aprendidos não são sequenciais nem dependentes do tempo. No caso do modelo proposto na monografia além de não ter o aspecto temporal, ele não tem o intuito de continuar aprendendo (MCGONAGLE, et al., 2021). Pelo que foi dito em 2.1.2 e pela tabela com dados de foraminíferos no formato da tabela 2.1, a qual foi utilizada para o treinamento das RNA, que tem resultados de temperatura média anual para parâmetros diferentes de período do ano e profundidade, verificamos que esses parâmetros devem ser considerados na previsão da paleotemperatura. Então, por isso será feito o treinamento de vários modelos com as diferentes configurações de profundidade e período do ano escolhidas.

### 3 METODOLOGIA

O trabalho está dividido em três etapas, como especificado na subseção 3.0.2. A primeira será centrada no entendimento do problema e a criação do modelo de predição, sendo dividida em:

- 1. Revisão da literatura para entender melhor o problema;
- 2. Criação do protótipo do sistema.
- 3. Validação do protótipo com o pesquisador.
- 4. Criação do modelo de RNA que consiga prever a temperatura média anual para uma determinada associação de foraminíferos.

Na segunda parte o foco será no desenvolvimento do sistema e na criação e otimização da RNA, sendo dividido em:

- 1. Desenvolvimento dos módulos do sistema web propostos no protótipo utilizando a linguagem R<sup>4</sup> com a biblioteca *Shiny*<sup>5</sup>.
- 2. Criação e comparação dos modelos de RNA visando modelos com menor erro.
- 3. Escrita da monografia.

Por fim, na terceira etapa, o foco estará nos ajustes finais e validação com o pesquisador sobre ambos, sistema e RNA, e também na finalização da escrita da monografia.

---

<sup>4</sup>Disponível em <<https://www.r-project.org/>> Acesso em 16 de julho de 2021.

<sup>5</sup>Disponível em <<https://shiny.rstudio.com/>> Acesso em 16 de julho de 2021.

## 4 DESENVOLVIMENTO

Neste capítulo, estão descritos os parâmetros e a estrutura utilizados para o treinamento da RNA e o funcionamento do sistema<sup>6</sup>, como o usuário irá interagir com ele e qual são os principais aspectos relativos à sua utilização. O link com o código do sistema se encontra no GitHub<sup>7</sup>.

### 4.1 ESTRUTURA DA REDE NEURAL

Foi utilizada uma planilha baseada no banco de dados ForCenS (SICCHA; KUCERA, 2017) e em García Chaporí e Kucera (2019) com aproximadamente 1600 linhas, que contém abundância relativa de diversas espécies de foraminíferos e 21 resultados diferentes que variam conforme o período do ano e profundidade. Nesta planilha primeiramente foi feita uma limpeza nos dados, alguns valores de resultado que estavam afetando o cálculo da rede foram retirados, pois na planilha continha algumas *strings* e números muito grandes que eram diferentes do padrão observado nos outros resultados.

Em seguida, para o treinamento, reservamos as proporções das comunidades de foraminíferos e apenas um dos resultados, que foi escolhido com base no período do ano e profundidade selecionado nos filtros do sistema. Foi feita a normalização dos dados utilizando a técnica de *min-max* e separado 90% em dados de treinamento e 10% de teste.

Quanto aos parâmetros do treinamento, foi utilizado 0,01 para *threshold*, que seria o critério de parada da rede, o algoritmo *retropropagation* explicado no capítulo 2.2.4,  $1e+06$  para *stepmax*, que corresponde ao tempo que a rede ficará treinando, e *hidden* igual a  $c(5, 2)$ , representado por um vetor que serve para especificar quantos neurônios tem em cada camada escondida, porém este valor foi utilizado apenas para o Caso 1 da imagem 4.2 enquanto para o Caso 2 da imagem 4.3 é abordado no capítulo 5 contendo um valor diferente de camadas escondidas para cada configuração. Na Figura 4.1 está o código que corresponde à estes parâmetros da rede inicial utilizada no trabalho.

Após a criação da RNA inicial exemplificada no capítulo 4, foi criado um algoritmo para modificar os parâmetros de neurônios das camadas escondidas, este algoritmo é o treinamento exemplificado no capítulo 4 com dois laços para alterar os neurônios. Os laços foram de 5 a 15 testando as possibilidades, guardando ao final de cada execução o resultado dos erros e a média entre eles.

---

<sup>6</sup>Disponível em <<https://vfhnkg-julia0gabriela-santi0acosta.shinyapps.io/foraminifera-predictor/>>

<sup>7</sup>Disponível em <<https://github.com/gsjulia/TCC-foraminifera-predictor>>

Figura 4.1 – Treinamento da Rede Neural inicial

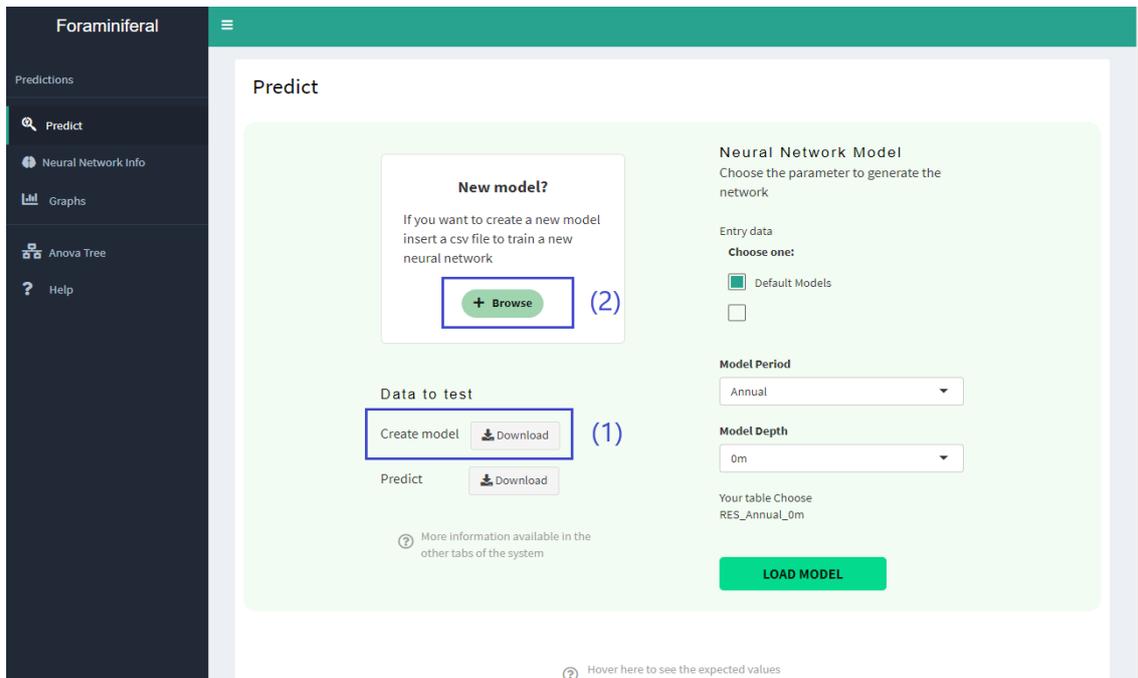
```
# Criação do treinamento da Rede Neural inicial
nn = neuralnet(str_c(outputExpected, " ~ ", formula), data=train,
               algorithm = "rprop+", startweights = NULL,
               hidden = c(5, 2), stepmax = 1e+06,
               lifesign = "none", threshold = 0.01,
               );
```

Fonte: Autora

## 4.2 CASO 1: USUÁRIO CRIANDO SEU PRÓPRIO MODELO

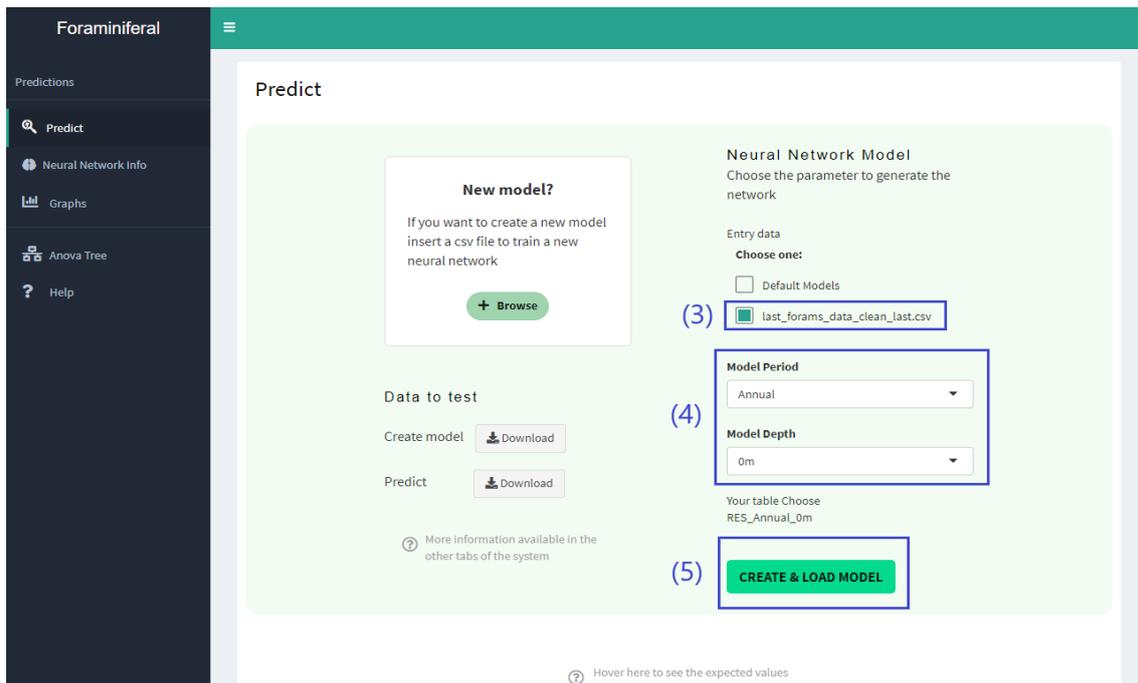
Neste primeiro caso, o usuário usa sua própria planilha para fazer o treinamento da rede. Na figura 4.2 (1) ele pode optar por baixar a planilha de teste para verificar a formatação, e em seguida no (2) fazer *upload* do modelo. Na figura 4.3, depois que tiver feito o *upload* no passo anterior, é possível selecionar a planilha importada (3), selecionar os parâmetros desejados (4) e criar o modelo (5). Após ser criado (imagem 4.4), é possível verificar se o modelo foi carregado com sucesso (6), suas métricas de desempenho podem ser vistas na seção 4.10. Na parte (7) poderá baixar a planilha que contém apenas as associações de foraminíferos, já no botão (8) será onde o usuário importará esse arquivo, que será usado para fazer a previsão da temperatura. Por fim, é possível visualizar as previsões no sistema (Figura 4.5).

Figura 4.2 – Caso 1: Primeira parte



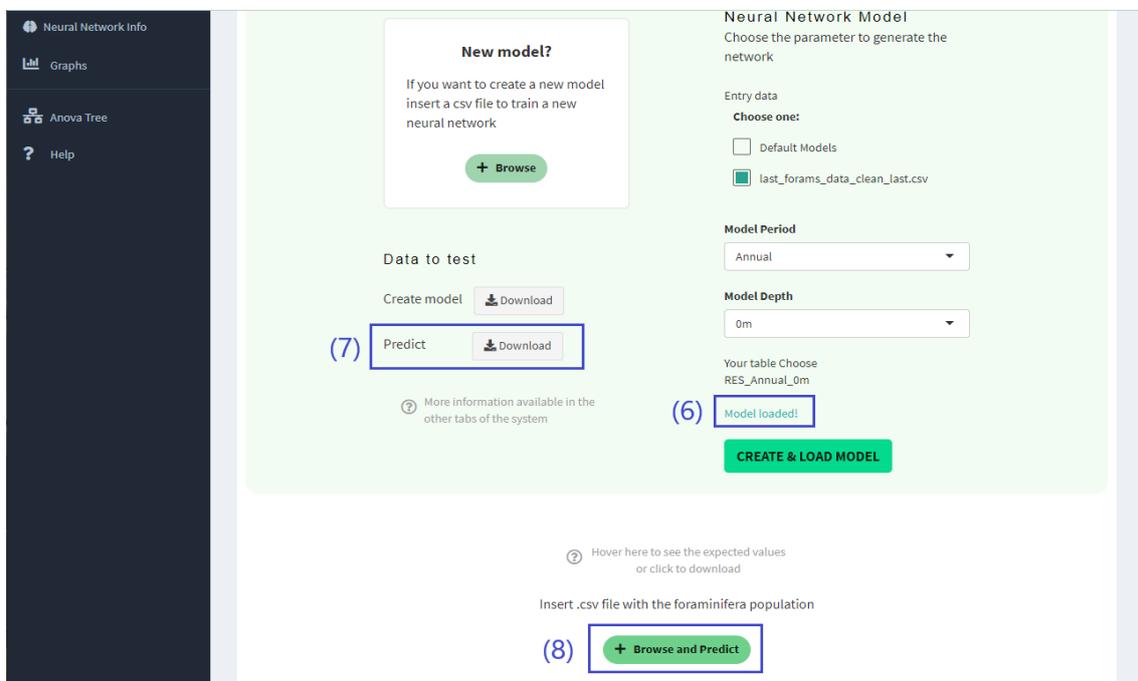
Fonte: Autora

Figura 4.3 – Caso 1: Segunda parte



Fonte: Autora

Figura 4.4 – Caso 1: Terceira parte



Fonte: Autora

Figura 4.5 – Caso 1: Quarta parte

Hover here to see the expected values or click to download

Insert .csv file with the foraminifera population

[+ Browse and Predict](#)

[Download](#)

Show **10** entries Search:

|    | Globoturborotalita_rubescens | Sphaeroidinella_dehiscens | Globigerinita_glutinata | unidentified.original...excluídos | Valor_predito    |
|----|------------------------------|---------------------------|-------------------------|-----------------------------------|------------------|
| 78 | 0.017276423                  | 0                         | 0.020325203             | 0                                 | 48.4141057337332 |
| 33 | 0.019447288                  | 0                         | 0.078812692             | 0                                 | 44.5730671648191 |
| 33 | 0.018672199                  | 0                         | 0.049792531             | 0                                 | 32.5382221109472 |
| 31 | 0.009424084                  | 0                         | 0.060732984             | 0.006282722                       | 60.1247167819073 |
| 34 | 0.006282723                  | 0                         | 0.080628272             | 0.003141361                       | 63.7944094355835 |
| 72 | 0.003131524                  | 0                         | 0.102296451             | 0                                 | 11.4977670700483 |
| 27 | 0.010427529                  | 0                         | 0.091762252             | 0                                 | 27.3738268622756 |
| 72 | 0.017489712                  | 0                         | 0.101851852             | 0                                 | 26.0403377836363 |
| 34 | 0.013104839                  | 0                         | 0.064516129             | 0                                 | 70.0247453274836 |
| 36 | 0.019547325                  | 0                         | 0.065843621             | 0                                 | 3.24828346119107 |

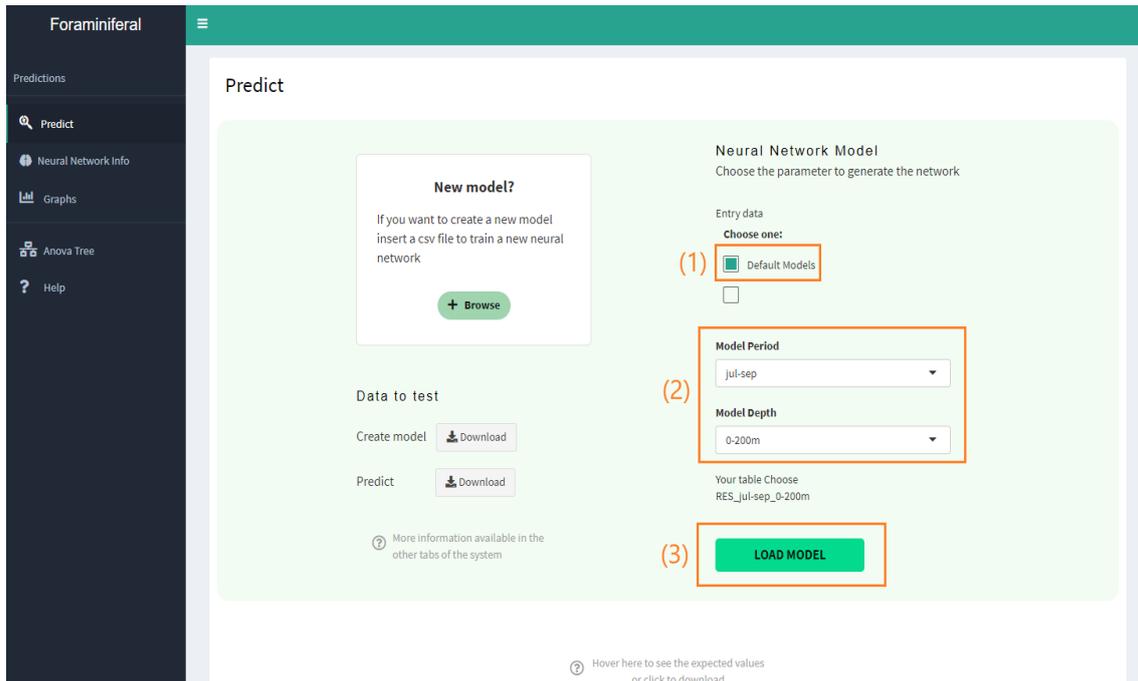
Showing 1,521 to 1,530 of 1,548 entries Previous 1 ... 151 152 **153** 154 155 Next

Fonte: Autora

#### 4.3 CASO 2: USUÁRIO UTILIZANDO MODELOS PREVIAMENTE SALVOS

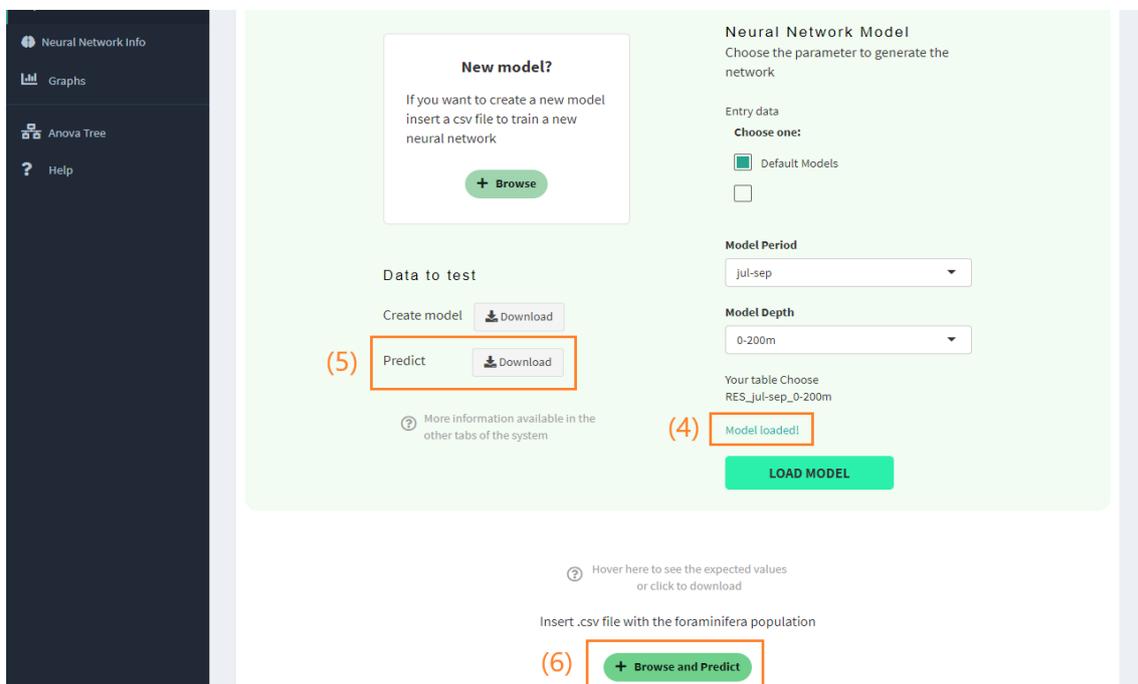
Para o segundo caso, o usuário utiliza os modelos salvos para fazer a predição. Na figura 4.6 é selecionada a opção *Default Models* (1), será selecionado o parâmetro desejado (2) que terá uma configuração de RNA diferente para cada parâmetro, e selecionará o botão de carregar o modelo (3). Após o modelo carregado, na figura 4.7 ficará visível se ele foi carregado com sucesso (4) e suas métricas de desempenho podem ser vistas na seção 4.10. Os valores teste de populações de foraminíferos estão disponíveis no (5) e o usuário poderá importar esses valores, ou usar a sua planilha, no botão (6). Por fim, a figura 4.8 representa a planilha que aparece no sistema dos valores preditos de temperatura média anual.

Figura 4.6 – Caso 2: Primeira parte



Fonte: Autora

Figura 4.7 – Caso 2: Segunda parte



Fonte: Autora

Figura 4.8 – Caso 2: Terceira parte

Insert .csv file with the foraminifera population

[+ Browse and Predict](#)

[Download](#)

Show  entries

Search:

| loboturborotalita_rubescens | Sphaeroidinella_dehiscens | Globigerinita_glutinata | unidentified.original...excluídos | Valor_predito    |
|-----------------------------|---------------------------|-------------------------|-----------------------------------|------------------|
| 0                           | 0                         | 0.128486056             | 0.00996016                        | 26.1154473258841 |
| 0                           | 0                         | 0.120758483             | 0                                 | 39.6174844345868 |
| 0                           | 0.003835091               | 0.04506232              | 0.04506232                        | 128.979134147051 |
| 0                           | 0                         | 0.041625372             | 0.008919722                       | 46.2484007984481 |
| 0                           | 0                         | 0.111111111             | 0.029239766                       | 25.5324726317576 |
| 0                           | 0                         | 0.097847358             | 0.05479452                        | 21.3341082427805 |
| 0                           | 0                         | 0.141865079             | 0.013888888                       | 35.9883515141117 |
| 0                           | 0                         | 0.131736527             | 0                                 | 26.0094364786554 |
| 0                           | 0                         | 0.14014014              | 0.006006006                       | 25.7898720893014 |
| 0                           | 0                         | 0.117882118             | 0.006993007                       | 26.0055223563359 |

Showing 1 to 10 of 1,548 entries

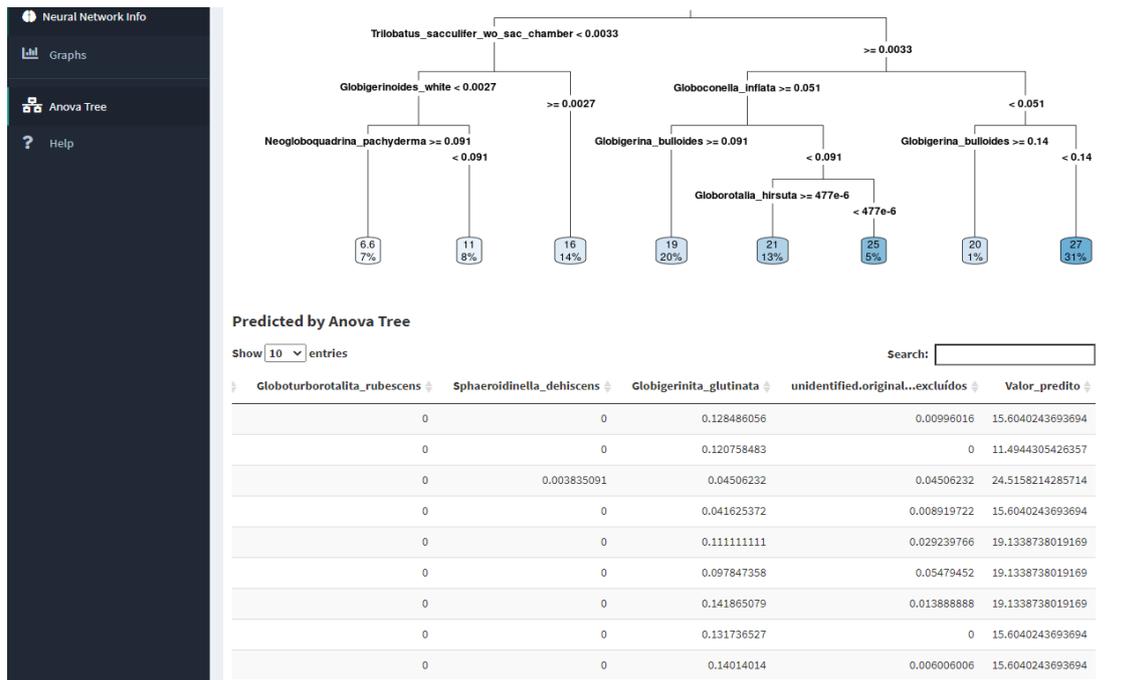
Previous  2 3 4 5 ... 155 Next

Fonte: Autora

#### 4.4 MENU ANOVA TREE

O menu da imagem 4.9 contém o gráfico que representa a árvore usando o método ANOVA, este gráfico é gerado a partir do treinamento armazenado no arquivo de dados para o caso 2 (*default models*) ou gerado durante o treinamento para o caso 1 em que o usuário cria seu próprio modelo. Neste menu, em um segundo momento, também é gerada a tabela com as paleotemperaturas estimadas pelo método ANOVA após o usuário clicar em *Browse and Predict* e inserir a planilha no *Home* exemplificado no caso 1 (8) da imagem 4.8 e no caso 2 (6) da imagem 4.7. Este menu pode ser utilizado como segunda opção para as previsões, visto que em alguns casos a rede está apresentando temperaturas negativas. Mais informações sobre a árvore ANOVA estão descritas na subseção 2.2.6 e no capítulo 5.

Figura 4.9 – Anova Tree



Fonte: Autora

#### 4.5 MENU HELP

No menu da imagem 4.10, há informações detalhadas sobre a árvore ANOVA, e sobre os erros utilizados como métrica de desempenho da RNA, também há os valores dos erros resultantes do treinamento feito no *Home* quando o usuário clica no botão de carregar o modelo destacado (3) da imagem 4.6 ou em (5) da imagem 4.3.

Figura 4.10 – Help

**Help**

See more info about the Neural Network metrics below, this is a Regression problem so the metrics will be different types of errors, the smaller the error, the better the model.

More info is available in: [Regression Errors & Anova Tree](#)

**Model Info**

**MEAN SQUARED ERROR: 1.37**

MSE is perhaps the most popular metric used for regression problems. Using this formula we find the mean square error between the predicted and actual values.

**ROOT MEAN SQUARED ERROR: 1.17**

It is similar to MSE, but the square root returns the scale of the squared errors.

**MEAN ABSOLUTE ERROR: 0.81**

MAE is a metric that finds the mean absolute distance between predicted and desired values.

**RELATIVE ABSOLUTE ERROR: 0.15**

It is the MAE relativized by dividing the MAE using the average as the value subtracted from the sum.

**Anova Tree**

Analysis of variance (ANOVA) is a statistical test for detecting differences in the group mean when there is one parametric dependent variable & one or more independent variable.

Fonte: Autora

#### 4.6 MENUS AUXILIARES DO SISTEMA

Dentre os outros *menus* do sistema há o *Neural Network* (Figura 4.11), que possui os valores de treino e teste utilizados para o treinamento do modelo, explicado no primeiro parágrafo da seção 2.2.4. O *menu Graphs* possui gráficos relativos ao treinamento da RNA, sendo o primeiro (Figura 4.12) sobre a comparação entre valores Reais e valores preditos pela RNA e o segundo (Figura 4.13) é a própria visualização da Rede Neural onde as entradas são as classes de foraminíferos e a saída é a temperatura média anual, esta saída contém os nomes das características usadas no treinamento. Por exemplo, a palavra *RES\_Annual\_0m* significa que as características utilizadas foram o período anual e uma profundidade de zero metros.

Figura 4.11 – *Neural Network Tab*

Foraminiferal

Predictions

- Predict
- Neural Network Info
- Graphs
- Anova Tree
- Help

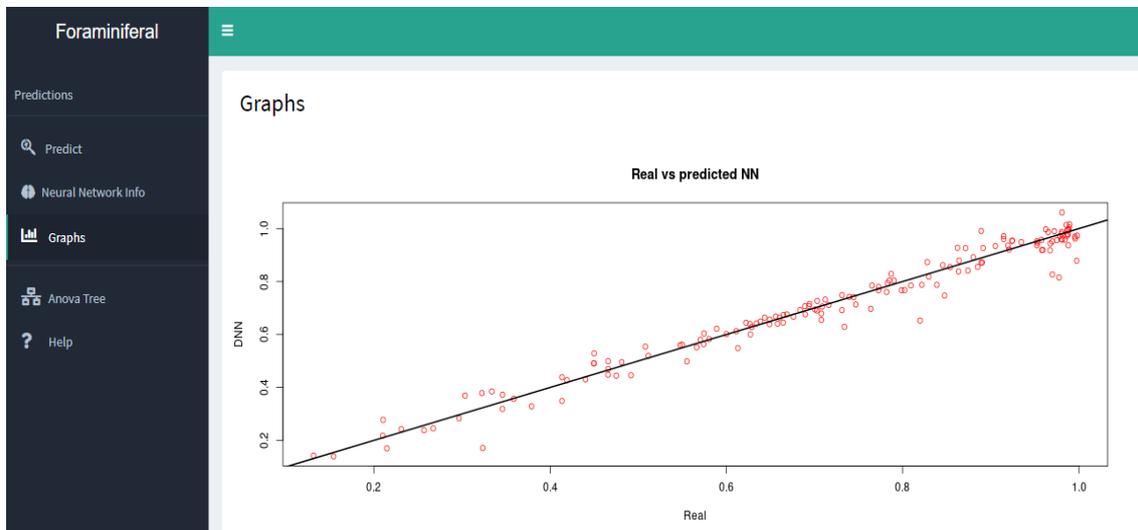
### Neural Network Information

Show  entries Search:

|     | Globorotalia_crassaformis | Globorotalia_hirsuta | Globorotalia_menardi_...<br>Globorotalia_tumida | Globorotalia_scitula | Globo |
|-----|---------------------------|----------------------|---|----------------------|-------|
| 36  | 0                         | 0                    | 0   | 0.013                |       |
| 39  | 0                         | 0.035035035          | 0.004004004                                     | 0.005005005          |       |
| 45  | 0                         | 0.003                | 0.05  | 0.002                |       |
| 46  | 0                         | 0.029                | 0.034   | 0                    |       |
| 48  | 0                         | 0.035                | 0.017   | 0.002                |       |
| 63  | 0                         | 0.008359457          | 0   | 0                    |       |
| 69  | 0                         | 0.001998002          | 0.002997003                                     | 0.015984016          |       |
| 73  | 0.014527845               | 0.004842615          | 0   | 0.00968523           |       |
| 100 | 0                         | 0                    | 0   | 0.002680965          |       |
| 109 | 0                         | 0.036101083          | 0.034296029                                     | 0                    |       |

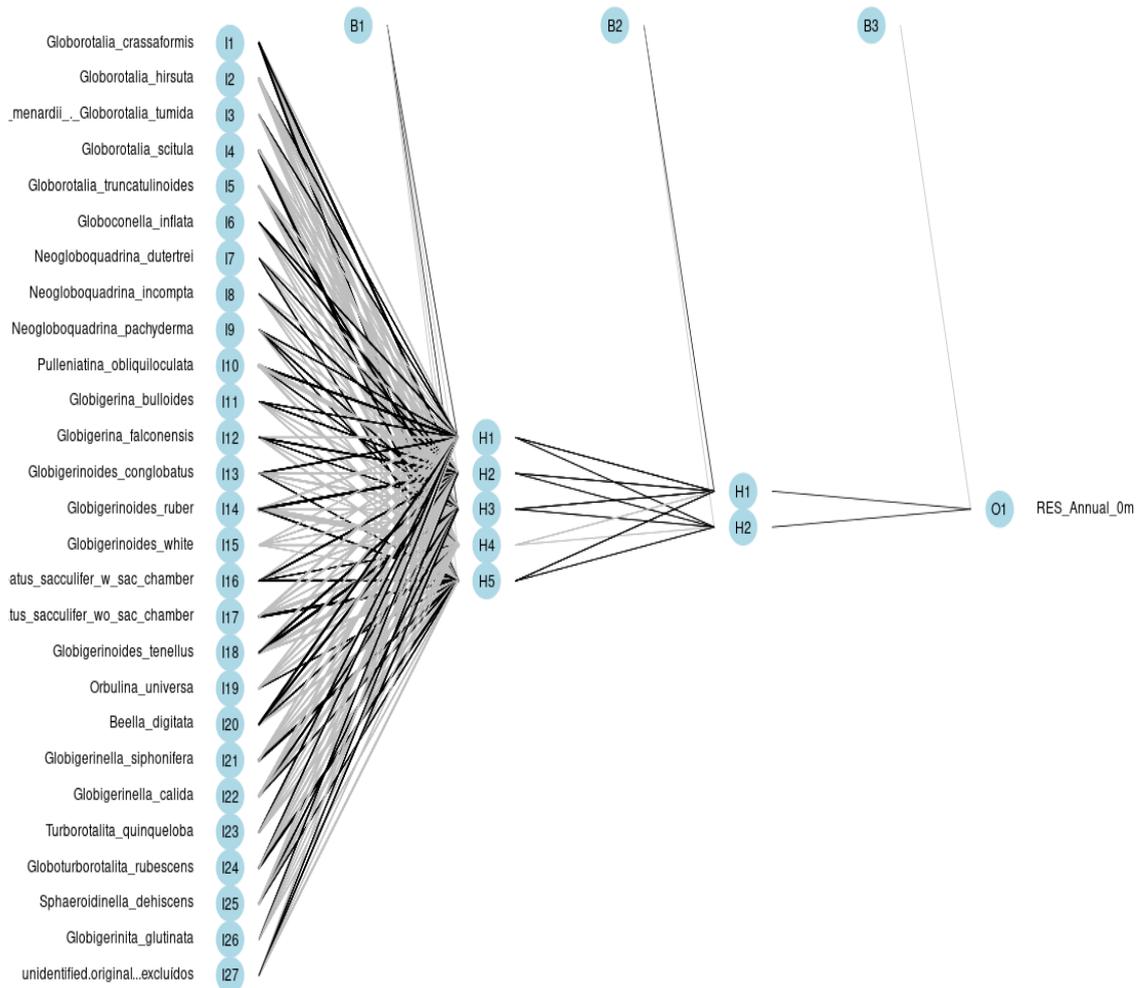
Showing 1 to 10 of 154 entries Previous      ...  Next

Fonte: Autora

Figura 4.12 – *Graphs Tab*: Comparação entre valores preditos e reais

Fonte: Autora

Figura 4.13 – *Graphs Tab*: Rede Neural criada pela aplicação



Fonte: Autora

## 5 RESULTADOS E CONSIDERAÇÕES

Como resultados da monografia, além do sistema e a Rede inicial que foram citados no capítulo 4, há os valores obtidos pelas técnicas de IA e árvore ANOVA. Neste capítulo, estão descritas observações sobre os diferentes tipos de treinamento feitos para descobrir os resultados médios de temperatura, destacando os que obtiveram melhores resultados. Também é possível verificar como funciona o fluxo do sistema na figura 5.1.

### 5.1 REDES NEURAIS: MODIFICAÇÃO DE PARÂMETROS

Considerando que o problema que esta monografia pretendia resolver é um problema de regressão, os erros do modelo foram usados como métrica de desempenho para escolher a melhor configuração, estes erros, descritos no subtópico 2.2.5 são: *Mean Absolute Error* (MAE) 2.10, *Mean Squared Error* (MSE) 2.7, *Root Mean Squared Error* (RMSE) 2.11, *Relative Absolute Error* (RAE) 2.12.

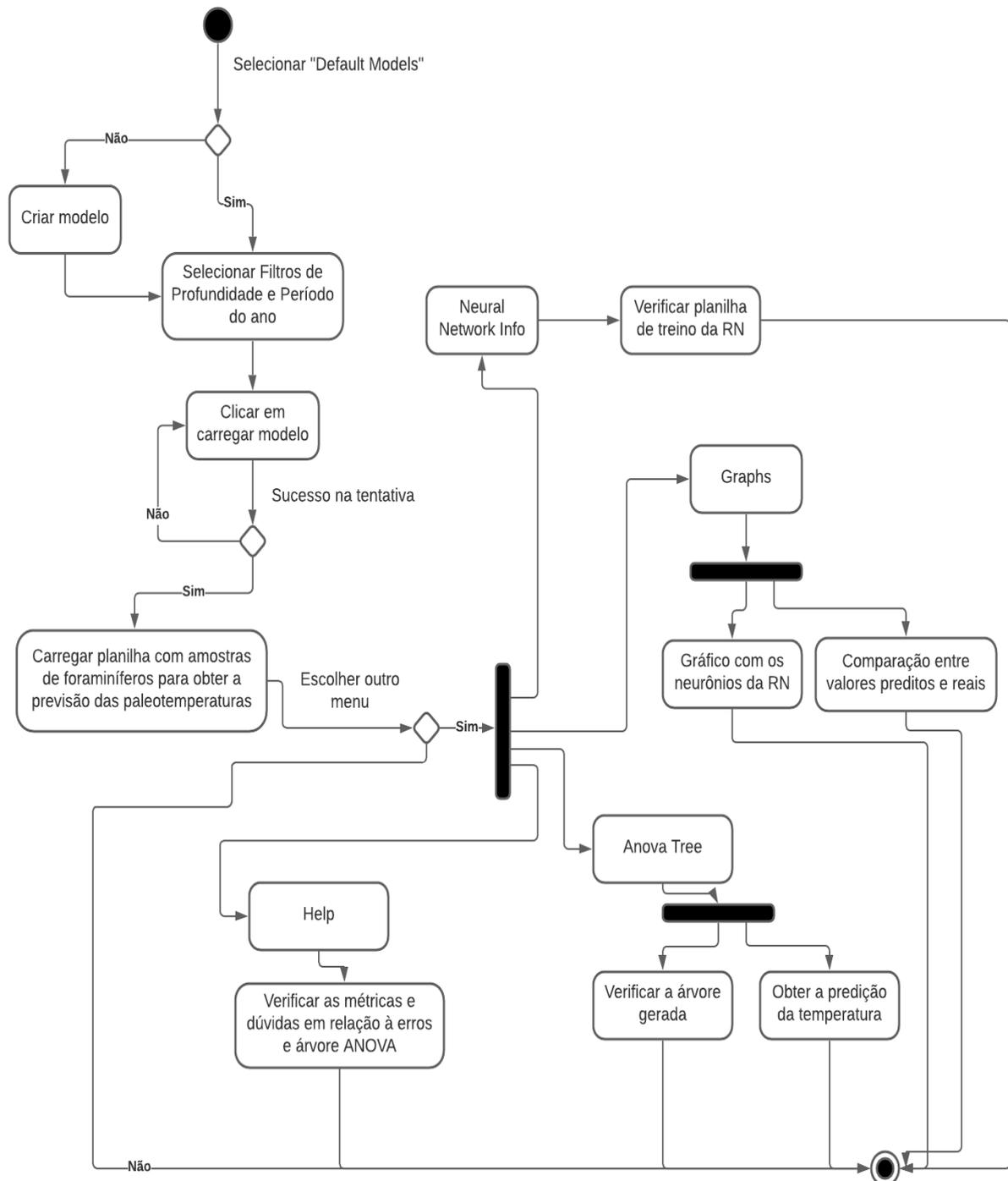
A bateria de testes feita para cada um dos 21 diferentes resultados da planilha com intuito de encontrar a quantidade de neurônios que gera o melhor resultado da rede, isso se aplica somente para o Caso 2: Usuário Utilizando Modelos Previamente Salvos, exemplificado na figura 4.3, pois neste caso, o treinamento já foi previamente realizado e está guardado em um arquivo de dados. Os modelos previamente salvos podem ter configurações diferentes uns dos outros, foi salvo o que tem o menor erro na verificação feita pelo algoritmo, todos os códigos e planilhas geradas encontram-se no GitHub<sup>8</sup> e podem ser reutilizados caso seja necessário fazer a alteração do arquivo de dados.

A Tabela 5.1 exemplifica alguns valores obtidos na planilha gerada pelo algoritmo em que o resultado era Anual com 0 metros. Para verificar o melhor modelo foram identificadas as duas linhas com a menor média entre os erros. No entanto, se esse valor fosse igual ou tivesse apenas 0,01 de diferença foi considerado a configuração com menor número de neurônios, isto pode ser verificado na tabela 5.2 que é a tabela gerada com as melhores configurações, a célula que contém os parâmetros escolhidos está destacada com um \* ao lado, essa verificação foi feita para cada 21 casos. É importante destacar que este foi apenas um estudo inicial, então podem ser feitos mais testes com intuito de gerar modelos mais precisos, envolvendo até situações diferentes das apresentadas nesta seção.

---

<sup>8</sup>Disponível em <<https://github.com/gsjulia/TCC-foraminifera-predictor>>

Figura 5.1 – Diagrama de Atividades



Fonte: Autora

Tabela 5.1 – Exemplo da planilha gerada para testar diferentes configurações

| Linha | Neurônios escondidos | MSE  | MAE  | RMSE | RAE  |
|-------|----------------------|------|------|------|------|
| 1     | 5, 5                 | 1,19 | 0,77 | 1,09 | 0,14 |
| 10    | 5, 14                | 0,79 | 0,68 | 0,89 | 0,12 |
| 15    | 6, 8                 | 1,08 | 0,75 | 1,04 | 0,13 |
| 30    | 7, 12                | 1,74 | 0,77 | 1,32 | 0,14 |
| 51    | 9, 11                | 1,04 | 0,72 | 1,02 | 0,13 |
| 70    | 11, 8                | 0,79 | 0,62 | 0,89 | 0,11 |
| 80    | 12, 7                | 0,85 | 0,66 | 0,92 | 0,12 |
| 82    | 12, 9                | 0,76 | 0,64 | 0,87 | 0,11 |
| 91    | 13, 7                | 0,86 | 0,69 | 0,93 | 0,12 |
| 105   | 14, 10               | 0,92 | 0,71 | 0,96 | 0,13 |

Fonte: Autora

Tabela 5.2 – Exemplo da planilha gerada com os menores erros para cada configuração

| Linha | Configuração       | Menor média | Neurônios do menor | Segunda menor média | Neurônios do segundo menor |
|-------|--------------------|-------------|--------------------|---------------------|----------------------------|
| 1     | RES-Annual-0m      | 0,58        | (10, 12)           | 0,59                | (6, 12)*                   |
| 2     | RES-Annual-50m     | 0,7         | (13, 12)           | 0,7                 | (13, 8)*                   |
| 3     | RES-Annual-75m     | 0,84        | (11, 9)*           | 1,11                | (14, 10)                   |
| 4     | RES-Annual-100m    | 0,66        | (14, 5)*           | 0,67                | (13,13)                    |
| 5     | RES-Annual-0-75m   | 0,68        | (15, 7)*           | 0,71                | (13, 14)                   |
| 6     | RES-Annual-0-100m  | 0,66        | (11, 10)*          | 0,84                | (15, 9)                    |
| 7     | RES-Annual-0-200m  | 0,56        | (14, 10)*          | 0,58                | (14, 6)                    |
| 8     | RES-jan-mar-0m     | 1,16        | (14, 11)*          | 1,18                | (10, 6)                    |
| 9     | RES-jan-mar-50m    | 0,78        | (15, 9)*           | 0,82                | (8, 9)                     |
| 10    | RES-jan-mar-75m    | 0,8         | (6, 14)*           | 1,16                | (8, 5)                     |
| 11    | RES-jan-mar-100m   | 0,74        | (7, 9)*            | 0,92                | (7, 11)                    |
| 12    | RES-jan-mar-0-75m  | 0,89        | (15, 14)*          | 0,96                | (10, 6)                    |
| 13    | RES-jan-mar-0-100m | 0,82        | (15, 9)*           | 0,84                | (14, 7)                    |
| 14    | RES-jan-mar-0-200m | 0,76        | (13, 14)           | 0,76                | (8, 8)*                    |
| 15    | RES-jul-sep-0m     | 0,96        | (15, 5)            | 0,97                | (10, 6)*                   |
| 16    | RES-jul-sep-50m    | 0,77        | (11, 9)*           | 0,9                 | (15, 6)                    |
| 17    | RES-jul-sep-75m    | 0,77        | (14, 5)*           | 0,99                | (14, 14)                   |
| 18    | RES-jul-sep-100m   | 0,76        | (9, 15)*           | 0,93                | (12, 8)                    |
| 19    | RES-jul-sep-0-75m  | 0,99        | (14, 12)*          | 1,26                | (14, 13)                   |
| 20    | RES-jul-sep-0-100m | 0,66        | (14, 11)*          | 0,88                | (14, 13)                   |
| 21    | RES-jul-sep-0-200m | 0,76        | (9, 9)*            | 0,78                | (5, 11)                    |

Fonte: Autora

## 5.2 ANALYSIS OF VARIANCE (ANOVA)

No caso do problema de predição de paleotemperaturas apresentado na monografia, a árvore ANOVA é utilizada para mostrar as proporções de comunidades de foraminíferos que foram os mais importantes para identificar a temperatura, que pode ser verificado na Figura 5.3, presente no menu Anova Tree do sistema. Neste mesmo menu, após a inserção dos dados das abundâncias relativas de espécies de foraminíferos foi gerada a tabela resultante com as paleotemperaturas no menu Home e também pode ser verificado, abaixo da imagem da árvore no menu ANOVA Tree, a tabela resultante do método ANOVA, que pode ser utilizada para casos em que a Rede apresente um resultado discrepante. Globalmente, no mar, a temperatura superficial varia aproximadamente entre -1.83 e 30.3, então resultados abaixo ou acima desses valores seriam considerados incomuns. A fórmula da Biblioteca *rpart* para a criação da árvore pode ser verificada na Figura 5.2.

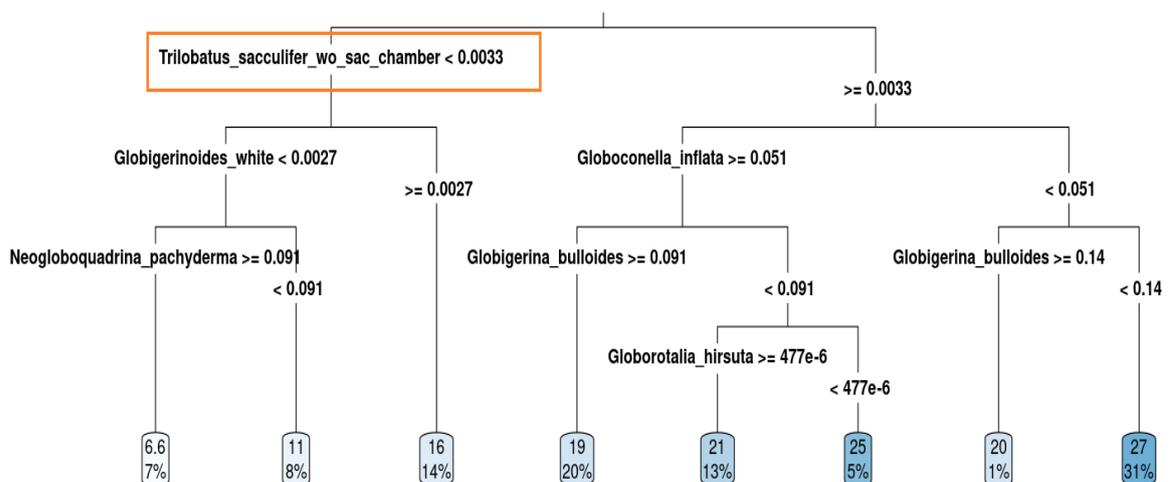
Figura 5.2 – Código em R para gerar a árvore

```
# Gerando a Árvore anova
formula = str_c(dfNames[!dfNames %in% outputExpected],
               collapse="+")

foramTree <- rpart(str_c(outputExpected, " ~ ", formula), data = df,
                  method='anova')

rpart.plot(foramTree, type = 3, digits = 2)
```

Figura 5.3 – Árvore ANOVA gerada no sistema



## 6 CONCLUSÃO

O objetivo deste trabalho foi desenvolver um sistema web onde pesquisadores pudessem, de maneira simplificada, estimar a temperatura média superficial do mar para um local específico com base em amostras de foraminíferos fósseis desse local. A ideia principal do sistema é possibilitar que apenas inserindo uma planilha com as abundâncias relativas de foraminíferos, e selecionando os dados de calibração incluindo a profundidade considerada para temperatura superficial e a época do ano de interesse obtém-se uma tabela com a predição das paleotemperaturas.

O caso 2 citado no parágrafo anterior utiliza os treinamentos salvos em um arquivo de dados do sistema e está descrito no Fluxograma 6.1 e no Diagrama de Atividades 5.1, o tópico 4.3 tem a explicação detalhada e as imagens sobre esta parte do sistema. Também há a possibilidade de gerar os parâmetros da Rede com uma planilha nova, como foi exemplificado no caso 1 do tópico 4.2. O caso 2 é o mais indicado, pois o sistema ainda não tem uma grande verificação de erros para novas planilhas então pode acontecer inconsistências que afetarão os resultados no caso 1.

Além disso, nos modelos salvos no arquivo de dados (caso 2) foram escolhidos a quantidade de neurônios que obtiveram os melhores resultados, enquanto no caso 1 em que o usuário insere a planilha para fazer o treino, a camada escondida será sempre fixa de (5, 2). A quantidade fixa de neurônios para camada escondida, foi escolhida baseada no que apresentava um erro pequeno mas que não fosse uma quantidade grande de neurônios porque isto faria com que o treinamento demorasse muito mais (Figura 4.1), futuramente pode ser adicionado um módulo no sistema que permita o usuário escolher este valor.

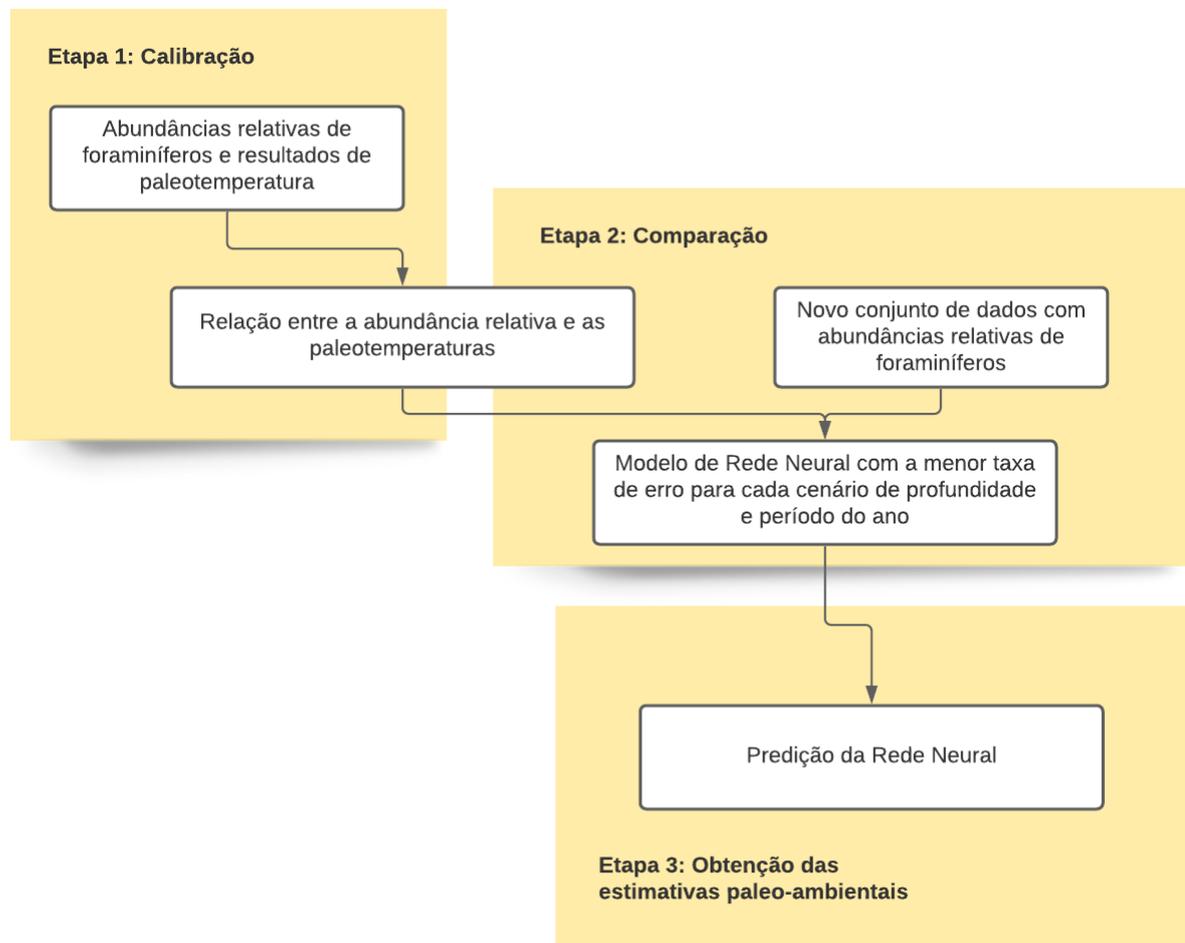
No capítulo 2, foram retratados alguns trabalhos envolvendo a utilização de RNA para predição de paleotemperaturas. O diferencial desta monografia foi proporcionar que o pesquisador não necessite de conhecimentos prévios de Redes Neurais e ainda assim possa obter resultados a partir da ferramenta. Vale ressaltar que há vários modelos que ainda não foram explorados, conforme descrito no capítulo 4, os testes foram realizados com apenas duas camadas escondidas e entre 5 e 15 neurônios, mesmo que em alguns casos o aumento da quantidade de neurônios não diminuía o erro, ou seja, não melhorava o resultado, poderia ser alterada a quantidade de camadas escondidas ou alterar outro parâmetro da Rede (Figura 4.1) com o objetivo de obter um resultado ainda melhor.

Como o escopo do trabalho era grande e envolvia tanto a parte do desenvolvimento do sistema como a parte de predição de paleotemperaturas, houveram algumas dificuldades. Apesar de as RNA treinadas apresentarem resultados interessantes, ainda há algumas melhorias que devem ser feitas futuramente, pois algumas amostras estão apresentando temperaturas negativas, o que é algo estranho, já que geralmente nesta temperatura a água salgada congelaria, até poderia ser algo razoável para uma estimativa de paleotemperatura para um local muito frio

localizado em altas latitudes, mas não foi o caso para os dados utilizados nesta monografia. Para resolver o problema nestes casos específicos onde a temperatura está negativa, poderia ser usado o resultado da árvore ANOVA ao invés do resultado da RNA, pois além da árvore ANOVA ter apresentado resultados interessantes ela tem uma vantagem em relação as RNA, ela permite verificar a importância das diferentes classes de foraminíferos mostrando qual teve maior impacto na estimativa.

Também vale destacar que nesta monografia não visamos encontrar o melhor modelo entre todos, mas sim procuramos por um modelo inicial com um resultado satisfatório, posteriormente podem ser explorados modelos híbridos e até mesmo técnicas de Rede Neurais mais avançadas.

Figura 6.1 – Diagrama do principal fluxo do sistema (Caso 2 do tópico 4.3)



Fonte: Autora

## 6.1 TRABALHOS FUTUROS

Para trabalhos futuros sugerimos a resolução do problema de temperaturas negativas citado no capítulo 6, podendo ser utilizado um modelo híbrido, bem como outra técnica de RNA ou até mesmo a alteração dos parâmetros utilizados na criação da RNA inicial.

É importante ressaltar que ainda existem vários pequenos ajustes que poderiam ser feitos no sistema em termos de experiência de usuário, começando pela questão de indicação de erros, caso o usuário insira alguma planilha com formatação errada.

Nesse sentido, poderia também ser feito um filtro mais detalhado em relação às planilhas inseridas, como na limpeza inicial da planilha utilizada que havia um número extremamente grande, já que os resultados de paleotemperaturas tendem a seguir um padrão de faixa de valores, pode-se desconfiar de valores muito altos ou muito baixos.

Como foi adicionada a árvore ANOVA, que estava fora do escopo inicial do trabalho, em um trabalho futuro poderia ser adicionados outros algoritmos que podem ter um bom resultado na predição de paleotemperaturas, como o algoritmo *Random Forest*. Ainda sobre a árvore ANOVA, já que ela mostra as classes mais importantes de foraminíferos, poderia ser verificado como o modelo reagiria sem a utilização das classes menos importantes, pois estas classes poderiam estar atuando como um *outlier* (Figura 2.9).

Outras questões relativas à entrada de dados podem ser melhor exploradas, como é o caso das bacias oceânicas, que podem ser analisadas da mesma forma que os parâmetros de profundidade e períodos do ano que foram usados com diferentes configurações de RNA. Outra sugestão, seria adicionar um local que permita o usuário escolher a quantidade de neurônios para o treinamento da Rede Neural a partir dos que geram um erro menor.

## REFERÊNCIAS BIBLIOGRÁFICAS

- AGGARWAL, C. C. **Neural Networks and Deep Learning: A Textbook**. [S.l.]: Springer International, 2018. 497 p.
- BOUDAGHER-FADEL, M. K. **Evolution and geological significance of larger benthic foraminifera**. [S.l.]: London: UCL, 2018.
- BRAGA, A. d. P.; LUDERMIR, T. B.; CARVALHO, A. P. d. L. F. **Redes neurais artificiais: teoria e aplicações**. [S.l.]: Livros Técnicos e Científicos, 2000. 248 p.
- CHAPORI, N. G. et al. Sea-surface temperature reconstruction of the quaternary western south atlantic: New planktonic foraminiferal correlation function. In: **Palaeogeography, Palaeoclimatology, Palaeoecology**. [S.l.: s.n.], 2015. v. 425, p. 67–75.
- DATA SCIENCE ACADEMY. **Deep Learning Book**. 2020. Acessado em 15 novembro 2021. Disponível em: <<https://www.deeplearningbook.com.br>>.
- DONATO, V. D. et al. Palaeoenvironmental reconstructions through compositional data analysis. In: **Alpine and Mediterranean Quaternary**. [S.l.: s.n.], 2018. v. 31, n. 1, p. 59–73.
- FACELI, K. et al. **Inteligência Artificial: Uma abordagem de aprendizado de máquina**. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda., 2011. 375 p.
- García Chapori, N. L.; Kucera, M. data set, **Planktonic foraminifera census counts from the western South Atlantic**. PANGAEA, 2019. Disponível em: <<https://doi.org/10.1594/PANGAEA.907931>>.
- HAN, S. et al. Artificial neural network: Understanding the basic concepts without mathematics. In: **Dement Neurocogn Disord**. [S.l.: s.n.], 2018. v. 17,3, p. 83–89.
- HAYES, A. et al. Glacial mediterranean sea surface temperatures based on planktonic foraminiferal assemblages. **Quaternary Science Reviews**, v. 24, n. 7, p. 999–1016, 2005. ISSN 0277-3791. Multiproxy Approach for the Reconstruction of the Glacial Ocean surface. Disponível em: <<https://www.sciencedirect.com/science/article/pii/S0277379104002124>>.
- HAYWARD et al. **World Foraminifera Database**. 2021. Acesso em: 16 julho 2021. Disponível em: <<http://www.marinespecies.org/foraminifera>>.
- JOHANSEN, T. H.; SØRENSEN, S. A. Towards detection and classification of microscopic foraminifera using transfer learning. **arXiv preprint arXiv:2001.04782**, 2020.
- JOHANSEN, T. H. et al. Instance segmentation of microscopic foraminifera. **arXiv preprint arXiv:2105.14191**, 2021.
- KUCERA, M. Chapter six planktonic foraminifera as tracers of past oceanic environments. **Developments in Marine Geology**, Volume 1, p. 213–262, 12 2007.
- MALMGREN, B. A. et al. Comparison of statistical and artificial neural network techniques for estimating past sea surface temperatures from planktonic foraminifer census data. In: **Paleoceanography**. [S.l.: s.n.], 2001. v. 16, n. 5, p. 520–530.
- MARCHANT, R. et al. Automated analysis of foraminifera fossil records by image classification using a convolutional neural network. **Journal of Micropalaeontology**, v. 39, n. 2, p. 183–202, 2020. Disponível em: <<https://jfm.copernicus.org/articles/39/183/2020/>>.

MCGONAGLE, et al. **Feedforward Neural Networks**. 2021. Acessado em 30 novembro 2021. Disponível em: <<https://brilliant.org/wiki/feedforward-neural-networks/>>.

MIKKULAINEN, R. et al. Artificial intelligence in the age of neural networks and brain computing. In: **Evolving deep neural networks**. [S.l.]: Elsevier, 2019. p. 293–312.

MITRA, R. et al. Automated species-level identification of planktic foraminifera using convolutional neural networks, with comparison to human performance. In: **Marine Micropaleontology**. [S.l.]: Elsevier, 2019. v. 147, p. 16–24.

MURRAY, J. W. **Ecology and applications of benthic foraminifera**. [S.l.]: Cambridge university press, 2006.

PETRÓ, S. M. et al. Paleooceanographic changes through the last 130 ka in the western south atlantic based on planktonic foraminifera. In: **Revista Brasileira de Paleontologia**. [S.l.: s.n.], 2016. v. 19, n. 1, p. 3–14.

Shervin Minaee. **Popular Machine Learning Metrics**. 2019. Acessado em 30 novembro 2021. Disponível em: <<https://towardsdatascience.com/>>.

SICCHA, M.; KUCERA, M. Forcens, a curated database of planktonic foraminifera census counts in marine surface sediment samples. **Scientific Data**, 2017.

SIJINKUMAR, A. V.; NATH, B. N.; CLEMENS, S. North atlantic climatic changes reflected in the late quaternary foraminiferal abundance record of the andaman sea, north-eastern indian ocean. In: **Palaeogeography, Palaeoclimatology, Palaeoecology**. [S.l.: s.n.], 2016. v. 446, p. 11–18.

University College London. **Foraminifera**. 2002. Acesso em: 08 novembro 2021. Disponível em: <<https://www.ucl.ac.uk/GeolSci/micropal/foram.html>>.

YASUHARA, M. et al. Instance segmentation of microscopic foraminifera. **Oceanography**, v. 33, n. 2, 2020.

ZHONG, B. et al. A comparative study of image classification algorithms for foraminifera identification. In: **2017 IEEE Symposium Series on Computational Intelligence (SSCI)**. [S.l.: s.n.], 2017. p. 1–8.