

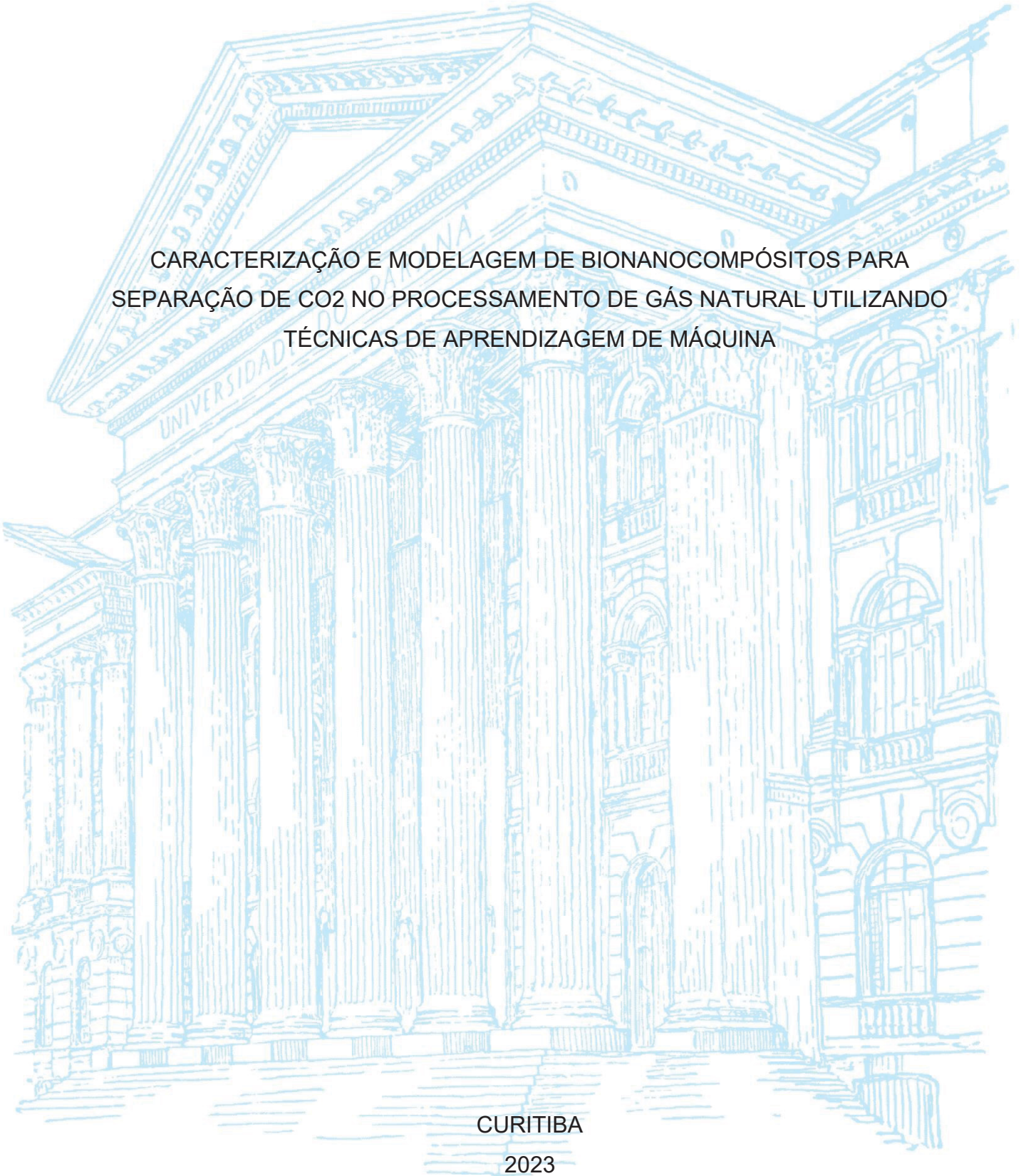
UNIVERSIDADE FEDERAL DO PARANÁ

GIOVANA SIGNORI IAMIN

CARACTERIZAÇÃO E MODELAGEM DE BIONANOCOMPÓSITOS PARA
SEPARAÇÃO DE CO₂ NO PROCESSAMENTO DE GÁS NATURAL UTILIZANDO
TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

CURITIBA

2023



GIOVANA SIGNORI IAMIN

CARACTERIZAÇÃO E MODELAGEM DE BIONANOCOMPÓSITOS PARA
SEPARAÇÃO DE CO₂ NO PROCESSAMENTO DE GÁS NATURAL UTILIZANDO
TÉCNICAS DE APRENDIZAGEM DE MÁQUINA

Dissertação apresentada ao curso de Pós-Graduação em Engenharia Química, Setor de Tecnologia, Universidade Federal do Paraná, como requisito parcial à obtenção do título de Mestre em Engenharia Química.

Orientador: Prof. Dr. Alexandre Ferreira Santos

Coorientador: Prof. Dr. Marcos Lúcio Corazza

CURITIBA

2023

DADOS INTERNACIONAIS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)
UNIVERSIDADE FEDERAL DO PARANÁ
SISTEMA DE BIBLIOTECAS – BIBLIOTECA CIÊNCIA E TECNOLOGIA

Iamin, Giovana Signori

Caracterização e modelagem de bionanocompósitos para separação de CO₂ no processamento de gás natural utilizando técnicas de aprendizagem de máquina. / Giovana Signori Iamin. – Curitiba, 2023.

1 recurso on-line : PDF.

Dissertação (Mestrado) - Universidade Federal do Paraná, Setor de Tecnologia, Pós-Graduação em Engenharia Química.

Orientador: Prof. Dr. Alexandre Ferreira Santos

Coorientador: Prof. Dr. Marcos Lúcio Corazza

1. Gás natural. 2. Aprendizado do computador. 3. Dióxido de carbono. 4 Modelagem. I. Santos, Alexandre Ferreira. II. Corazza, Marcos Lúcio. III. Universidade Federal do Paraná. Programa de Pós-Graduação em Engenharia Química. IV. Título.

Bibliotecária: Roseny Rivelini Morciani CRB-9/1585

ATA DE SESSÃO PÚBLICA DE DEFESA DE MESTRADO PARA A OBTENÇÃO DO GRAU DE MESTRA EM ENGENHARIA QUÍMICA

No dia vinte e sete de julho de dois mil e vinte e três às 09:00 horas, na sala Microsoft Teams, Plataforma online, foram instaladas as atividades pertinentes ao rito de defesa de dissertação da mestranda **GIOVANA SIGNORI IAMIN**, intitulada: **CARACTERIZAÇÃO E MODELAGEM DE BIONANOCOMPÓSITOS PARA SEPARAÇÃO DE CO2 NO PROCESSAMENTO DE GÁS NATURAL UTILIZANDO TÉCNICAS DE APRENDIZAGEM DE MÁQUINA**, sob orientação do Prof. Dr. ALEXANDRE FERREIRA SANTOS. A Banca Examinadora, designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA QUÍMICA da Universidade Federal do Paraná, foi constituída pelos seguintes Membros: ALEXANDRE FERREIRA SANTOS (UNIVERSIDADE FEDERAL DO PARANÁ), DANIEL EIRAS (UNIVERSIDADE FEDERAL DO PARANÁ), RAFAEL BRUNO VIEIRA (UNIVERSIDADE FEDERAL DE UBERLÂNDIA). A presidência iniciou os ritos definidos pelo Colegiado do Programa e, após exarados os pareceres dos membros do comitê examinador e da respectiva contra argumentação, ocorreu a leitura do parecer final da banca examinadora, que decidiu pela APROVAÇÃO. Este resultado deverá ser homologado pelo Colegiado do programa, mediante o atendimento de todas as indicações e correções solicitadas pela banca dentro dos prazos regimentais definidos pelo programa. A outorga de título de mestra está condicionada ao atendimento de todos os requisitos e prazos determinados no regimento do Programa de Pós-Graduação. Nada mais havendo a tratar a presidência deu por encerrada a sessão, da qual eu, ALEXANDRE FERREIRA SANTOS, lavrei a presente ata, que vai assinada por mim e pelos demais membros da Comissão Examinadora.

CURITIBA, 27 de Julho de 2023.

Assinatura Eletrônica

27/07/2023 15:42:36.0

ALEXANDRE FERREIRA SANTOS

Presidente da Banca Examinadora

Assinatura Eletrônica

27/07/2023 13:29:58.0

DANIEL EIRAS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

27/07/2023 13:22:01.0

RAFAEL BRUNO VIEIRA

Avaliador Externo (UNIVERSIDADE FEDERAL DE UBERLÂNDIA)



MINISTÉRIO DA EDUCAÇÃO
SETOR DE TECNOLOGIA
UNIVERSIDADE FEDERAL DO PARANÁ
PRÓ-REITORIA DE PESQUISA E PÓS-GRADUAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO ENGENHARIA QUÍMICA
- 40001016056P9

TERMO DE APROVAÇÃO

Os membros da Banca Examinadora designada pelo Colegiado do Programa de Pós-Graduação ENGENHARIA QUÍMICA da Universidade Federal do Paraná foram convocados para realizar a arguição da dissertação de Mestrado de **GIOVANA SIGNORI IAMIN** intitulada: **CARACTERIZAÇÃO E MODELAGEM DE BIONANOCOMPÓSITOS PARA SEPARAÇÃO DE CO2 NO PROCESSAMENTO DE GÁS NATURAL UTILIZANDO TÉCNICAS DE APRENDIZAGEM DE MÁQUINA**, sob orientação do Prof. Dr. ALEXANDRE FERREIRA SANTOS, que após terem inquirido a aluna e realizada a avaliação do trabalho, são de parecer pela sua APROVAÇÃO no rito de defesa.

A outorga do título de mestra está sujeita à homologação pelo colegiado, ao atendimento de todas as indicações e correções solicitadas pela banca e ao pleno atendimento das demandas regimentais do Programa de Pós-Graduação.

CURITIBA, 27 de Julho de 2023.

Assinatura Eletrônica

27/07/2023 15:42:36.0

ALEXANDRE FERREIRA SANTOS

Presidente da Banca Examinadora

Assinatura Eletrônica

27/07/2023 13:29:58.0

DANIEL EIRAS

Avaliador Interno (UNIVERSIDADE FEDERAL DO PARANÁ)

Assinatura Eletrônica

27/07/2023 13:22:01.0

RAFAEL BRUNO VIEIRA

Avaliador Externo (UNIVERSIDADE FEDERAL DE UBERLÂNDIA)

Rua Cel. Francisco Heráclito dos Santos, s/nº - CURITIBA - Paraná - Brasil

CEP 81531-980 - Tel: (41) 3361-3590 - E-mail: ppgeq@ufpr.br

Documento assinado eletronicamente de acordo com o disposto na legislação federal Decreto 8539 de 08 de outubro de 2015.

Gerado e autenticado pelo SIGA-UFPR, com a seguinte identificação única: 300297

Para autenticar este documento/assinatura, acesse <https://siga.ufpr.br/siga/visitante/autenticacaoassinaturas.jsp> e insira o código 300297

Dedico este trabalho à minha família querida (meu pai Gustavo, minha mãe Solange e minha irmã Melissa), e ao meu companheiro de vida, Gabriel. Vocês são minha inspiração, obrigada por todo o amor e por sempre estarem ao meu lado.

AGRADECIMENTOS

Primeiramente, agradeço pela sorte de ter uma família tão amável como a minha. Pai, mãe e Mel, obrigada por sempre estarem ao meu lado, tornando as coisas mais fáceis. Gostaria de agradecer ao meu amor e companheiro, Gabriel. Eu agradeço imensamente por todo o apoio que você me dá para seguir a minha carreira.

Minha imensa gratidão ao meu professor orientador Prof. Dr. Alexandre Ferreira Santos, que além de um pesquisador docente incrível, é uma pessoa maravilhosa. Prof, obrigada pela orientação e me proporcionar aprendizados e oportunidades únicas para minha carreira. Agradeço também ao meu coorientador, Prof. Dr. Marcos Corazza, por também ser um profissional inspirador e por todo o auxílio proporcionado nesses dois anos de mestrado.

Agradeço à UFPR, ao DEQ e a PPGEQ. Aqui, pude conhecer pessoas e profissionais incríveis que são importantes para mim e são minha inspiração.

Agradeço à banca da qualificação e da defesa, Prof. Dr. Daniel Eiras, Prof. Dr. Maurício Bezerra e Prof. Dr. Rafael Vieira. Muito obrigada por contribuírem para que este trabalho fosse ainda melhor. Gostaria de agradecer ao Prof. Dr. Daniel Eiras e o seu grupo de membranas também por me auxiliarem na fase experimental e me ensinarem tanto sobre um assunto que era totalmente novo para mim.

Agradeço aos meus colegas do LACTA, que me ajudaram nas mais diferentes atividades e situações. Sei que fiz amigos que se manterão para sempre.

Agradeço ao Prof. Dr. Luís Ricardo Kanda, que foi essencial para a criação do equipamento de umidificação de gases. Obrigada por facilitar este processo!

Agradeço ao grupo LEPAMAP-PRODIS, da Universitat de Girona. Conheci profissionais que fizeram toda a diferença no meu aprendizado. Ademais, também foram um dos pilares mais importantes para este trabalho acontecer, disponibilizando os dados utilizados para realizar os modelos, além das nanoceluloses para fazer a etapa experimental.

Agradeço o apoio financeiro do Programa de Recursos Humanos da Agência Nacional do Petróleo, Gás Natural e Biocombustíveis – PRH-ANP, suportado com recursos provenientes do investimento de empresas petrolíferas qualificadas na Cláusula de P, D&I da Resolução ANP nº 50/2015.

Por fim, agradeço a todos aqueles que fizeram parte desta trajetória de alguma forma!

RESUMO

O presente estudo teve como objetivo aplicar a aprendizagem de máquina na predição de propriedades relevantes de bionanocompósitos (álcool polivinílico e celulose nanofibriladas) e da nanocelulose. Esses bionanocompósitos podem ser utilizados na separação de gases através de membranas, logo, o foco da aplicação da pesquisa está relacionado ao uso destes materiais na separação de CO₂ no processamento de gás natural, relevante operação unitária para evitar emissões de gases de efeito estufa na atmosfera, manter boas condições dos equipamentos e tubulações das plataformas e gasodutos e promover a qualidade do gás natural produzido. Foram aplicados dois tipos de algoritmos para as predições: florestas randômicas e redes neurais. Das propriedades das nanoceluloses, predisseram-se a razão de aspecto e o rendimento de nanofibrilação. O primeiro é uma importante característica morfológica e que pode ter efeitos na determinação de propriedades mecânicas e de permeação dos bionanocompósitos, enquanto o segundo é um parâmetro que relaciona o grau de nanofibrilação das fibras da matéria-prima inicial. Os modelos gerados foram muito satisfatórios em termos de desempenho e utilizaram como variáveis de entrada algumas características químicas da matéria-prima, parâmetros do processo de pré-tratamento e nanofibrilação ou propriedades finais da nanocelulose. Para a razão de aspecto, destacou-se o algoritmo de floresta randômica, que alcançou um R² de 0,919 e um erro absoluto médio de 9,05 para predição dos dados de teste. Já para o modelo de rendimento de nanofibrilação, o modelo que obteve o melhor desempenho foi a rede neural, com um R² de 0,943 e um erro absoluto médio de 0,009 para o mesmo tipo de subconjunto. Das propriedades envolvendo a separação de gases, geraram-se modelos que predisseram a permeância de CO₂ e seletividade CO₂/CH₄ a partir de propriedades relacionadas à membrana e ao processo, como razão de aspecto e concentração de nanocelulose no compósito e pressão e umidade relativa dos gases de entrada. Para as duas propriedades, o melhor modelo foi a rede neural. No caso da permeância, obtiveram-se um R² de 0,859 e um erro absoluto médio de 0,0179 e, no caso da seletividade, um R² de 0,907 e um erro absoluto médio de 2,096, todos para o caso de predição de dados não vistos durante o treinamento dos modelos. Desta maneira, constatou-se a viabilidade de aplicação de técnicas de inteligência artificial na predição de propriedades relevantes de materiais e membranas, gerando modelos com capacidades preditivas muito boas mesmo na presença de uma quantidade limitada de dados experimentais. Ademais, foi possível compreender a influência, a nível preditivo, das variáveis de entrada estudadas na predição da variável de saída de interesse.

Palavras-chave: aprendizagem de máquina; nanocelulose; processamento de gás natural; separação de CO₂; modelagem.

ABSTRACT

The present study aimed at the application of machine learning in the prediction of relevant bionanocomposites (specifically polyvinyl alcohol and nanofibrillated cellulose composites) and nanocellulose properties. These bionanocomposites could be used in membrane gas separation operations. Thus, the focus here is related to the use of these materials in the separation of CO₂ from methane in natural gas processing. Separating CO₂ from natural gas is an extremely important unit operation in the oil and gas industry, in order to avoid greenhouse gas emissions, maintain equipment and lines in good conditions and improve the natural gas quality. Two types of machine learning algorithms were investigated: random forests and artificial neural networks. The nanocellulose properties predicted were aspect ratio and yield of nanofibrillation. The former is an important morphological characteristic that can affect the mechanical and permeation performances of the bionanocomposites, while the latter is a parameter that relates the degree of nanofibrillation in comparison to the original raw-material's fibers. The generated models for these outputs were very satisfactory in terms of prediction capacity. The input variables considered englobed some raw material chemical characteristics, pre-treatment and nanofibrillation step process parameters, as well as final nanocellulose properties. For the aspect ratio, the random forest model got the best results, with a R² score of 0,919 and a mean absolute error of 9,05, both metrics for the prediction of test data. As for the yield of nanofibrillation, the neural network was the top model, presenting a R² of 0,943 and a mean absolute error of 0,9 percentual points for the test subset as well. In relation to the properties involving the separation of gases, CO₂ permeance and CO₂/CH₄ selectivity were the outputs of interest evaluated. The inputs considered were related to membrane/material features as well as process parameters, such as nanocellulose aspect ratio and concentration in the composite, in addition to pressure and relative humidity of the gases. For both properties, the best model was the neural network. In the case of permeance, a R² score of 0,859 and a mean absolute error of 0,0179 were achieved, while for the selectivity a R² of 0,907 and a mean absolute error of 2,096 were obtained. These metrics are all related to the prediction of data unseen by the model during the training step. Therefore, the application of artificial intelligence techniques was found to be viable for the prediction of relevant material and membrane properties. Models with very good prediction abilities were generated, regardless the limitation in respect to the number of experimental data available for use. Furthermore, it was possible to comprehend the influence, in a predictive level, that the input variables investigated in this study had over the outputs of interest.

Key-words: machine learning; nanocellulose; natural gas processing; CO₂ separation; modelling

LISTA DE FIGURAS

FIGURA 1 - MONÔMERO DO PVA.	38
FIGURA 2 - ESTRUTURA HIERÁRQUICA DA CELULOSE EM PLANTAS.	39
FIGURA 3 - REPRESENTAÇÃO DE UM EXEMPLO DE REDE NEURAL COM DUAS CAMADAS OCULTAS E DOIS OUTPUTS DE SAÍDA.....	43
FIGURA 4 - REPRESENTAÇÃO ESQUEMÁTICA DE UMA FLORESTA RANDÔMICA PARA PROBLEMA DE REGRESSÃO.	46
FIGURA 5 - FLUXOGRAMA DO PROCESSO DE UMIDIFICAÇÃO E PERMEAÇÃO DOS GASES.....	60
FIGURA 6 - GRÁFICOS DE DISPERSÃO PARA A RAZÃO DE ASPECTO DE CNFS	72
FIGURA 7 - GRÁFICOS DE DISPERSÃO PARA O RENDIMENTO DE NANOFIBRILAÇÃO DE CNFS.....	73
FIGURA 8 – IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA.....	108
FIGURA 9 - IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA.....	109
FIGURA 10 - IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA DA SUPERFÍCIE DAS MEMBRANAS COMPÓSITAS DE PVA E CNF.	109
FIGURA 11 - GRÁFICOS DE DISPERSÃO DA PERMEANÇA EM FUNÇÃO DAS VARIÁVEIS DE ENTRADA	117
FIGURA 12 - GRÁFICOS DE DISPERSÃO DA SELETIVIDADE EM FUNÇÃO DAS VARIÁVEIS DE ENTRADA	119

LISTA DE GRÁFICOS

GRÁFICO 1 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA UTILIZANDO OS HIPERPARÂMETROS DO MODELO PRELIMINAR PARA PREDIÇÃO DA RAZÃO DE ASPECTO.....	78
GRÁFICO 2 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA INTERMEDIÁRIO PARA PREDIÇÃO DA RAZÃO DE ASPECTO.....	81
GRÁFICO 3 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA RAZÃO DE ASPECTO UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.....	82
GRÁFICO 4 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DA RAZÃO DE ASPECTO.....	84
GRÁFICO 5 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL UTILIZANDO OS HIPERPARÂMETROS PRELIMINARES PARA PREDIÇÃO DA RAZÃO DE ASPECTO.....	86
GRÁFICO 6 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO/TESTE (B) DO MODELO DE REDE NEURAL INTERMEDIÁRIO PARA PREDIÇÃO DE RAZÃO DE ASPECTO.....	88
GRÁFICO 7 - FORÇA RELATIVA ABSOLUTA DAS VARIÁVEIS DE ENTRADA PARA PREDIÇÃO DA RAZÃO DE ASPECTO UTILIZANDO O ALGORITMO DE REDE NEURAL.....	89
GRÁFICO 8 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO/TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DE RAZÃO DE ASPECTO.....	91

GRÁFICO 9 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA COM OS HIPERPARÂMETROS PRELIMINARES PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	93
GRÁFICO 10 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA INTERMEDIÁRIO PARA PREDIÇÃO DE RENDIMENTO DE NANOFIBRILAÇÃO.	95
GRÁFICO 11 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.	96
GRÁFICO 12 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE RENDIMENTO DE NANOFIBRILAÇÃO.	98
GRÁFICO 13 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL UTILIZANDO OS HIPERPARÂMETROS ORIGINAIS PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	100
GRÁFICO 14 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL INTERMEDIÁRIO PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	102
GRÁFICO 15 - FORÇA RELATIVA ABSOLUTA DAS VARIÁVEIS DE ENTRADA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO UTILIZANDO O ALGORITMO DE REDE NEURAL.	103
GRÁFICO 16 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	104
GRÁFICO 17 - PERDA DE MASSA (%) E DERIVADA DA MASSA (DTG) EM FUNÇÃO DA TEMPERATURA.	112

GRÁFICO 18 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE PERMEANÇA DE CO ₂	122
GRÁFICO 19 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA PERMEANÇA DE CO ₂ UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.	124
GRÁFICO 20 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DE PERMEANÇA DE CO ₂	126
GRÁFICO 21 - FORÇA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA PERMEANÇA DE CO ₂ UTILIZANDO O ALGORITMO DE REDES NEURAS.	128
GRÁFICO 22 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE SELETIVIDADE DE CO ₂	130
GRÁFICO 23 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA SELETIVIDADE DE CO ₂ UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.	132
GRÁFICO 24 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE REDE NEURAL PARA PREDIÇÃO DE SELETIVIDADE DE CO ₂	134
GRÁFICO 25 - FORÇA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA SELETIVIDADE DE CO ₂ UTILIZANDO O ALGORITMO DE REDES NEURAS.	136
GRÁFICO 26 - RELAÇÃO DE IMPORTÂNCIA RELATIVA DAS VARIÁVEIS DE ENTRADA PARA O MODELO DE PREDIÇÃO DA RAZÃO DE ASPECTO DA FLORESTA RANDÔMICA.	166
GRÁFICO 27 - VALORES PREDITOS VERSUS VALORES REAIS DE RAZÃO DE ASPECTO PARA A FLORESTA RANDÔMICA.	170
GRÁFICO 28 - VALORES PREDITOS VERSUS VALORES REAIS DE RAZÃO DE ASPECTO PARA A REDE NEURAL 2.	171

GRÁFICO 29 - RELAÇÃO DE IMPORTÂNCIA RELATIVA DAS VARIÁVEIS DE ENTRADA PARA O MODELO DE PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DA FLORESTA RANDÔMICA.....	175
GRÁFICO 30 - VALORES PREDITOS VERSUS VALORES REAIS DE RENDIMENTO DE NANOFIBRILAÇÃO PARA A FLORESTA RANDÔMICA.	179
GRÁFICO 31 - VALORES PREDITOS VERSUS VALORES REAIS DE RENDIMENTO DE NANOFIBRILAÇÃO PARA A REDE NEURAL 3.	180

LISTA DE QUADROS

QUADRO 1 - CARACTERIZAÇÃO DA MATÉRIA-PRIMA DE ORIGEM E CNF MECÂNICA	58
QUADRO 2 - RELAÇÃO DE BIBLIOTECAS E RESPECTIVAS FUNÇÕES E/OU CLASSES UTILIZADAS PARA APLICAÇÃO DAS TÉCNICAS DE APRENDIZAGEM DE MÁQUINA.....	62
QUADRO 3 - COMPOSIÇÃO QUÍMICA E ÍNDICE DE CRISTALINIDADE DAS POLPAS LIGNOCELULÓSICAS UTILIZADAS PARA GERAÇÃO DAS CNFs MECÂNICA E ENZIMÁTICA.	149
QUADRO 4 - CARACTERÍSTICAS DAS NANOFIBRAS DE CELULOSE PRODUZIDAS DE ACORDO COM DOSAGEM ENZIMÁTICA E INTENSIDADE DO PROCESSO DE NANOFIBRILAÇÃO.....	151
QUADRO 5 - BASE DE DADOS PARA MEMBRANAS DE PVA PURA, PVA+CNC E PVA+CNF.	158

LISTA DE TABELAS

TABELA 1 - COMPOSIÇÃO GENÉRICA DO GÁS NATURAL.....	28
TABELA 2 - DADOS DE PERMEABILIDADE E SELETIVIDADE AO CO ₂ PARA DIFERENTES BIOPOLÍMEROS	36
TABELA 3 - DADOS DE PERMEABILIDADE E SELETIVIDADE AO CO ₂ PARA DIFERENTES POLÍMEROS TRADICIONAIS.....	37
TABELA 4 - PRINCIPAIS CARACTERÍSTICAS MORFOLÓGICAS DE NCB, CNF E CNC.	38
TABELA 5 - FAIXA DE VALORES DAS VARIÁVEIS RELEVANTES PARA PERMEAÇÃO EM MEMBRANAS DE PVA E CNC.....	55
TABELA 6 - FAIXA DE VALORES DAS VARIÁVEIS RELEVANTES PARA PERMEAÇÃO EM MEMBRANAS DE PVA E CNF.....	56
TABELA 7 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE FLORESTA RANDÔMICA PARA OS MODELOS PRELIMINARES DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS	65
TABELA 8 – HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE REDE NEURAL PARA OS MODELOS PRELIMINARES DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.	67
TABELA 9 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE FLORESTA RANDÔMICA PARA OS MODELOS COMPLETOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.....	68
TABELA 10 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE REDE NEURAL PARA OS MODELOS COMPLETOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.	69
TABELA 11 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E VARIÁVEIS DE SAÍDA PARA MODELOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE FIBRILADA.....	75

TABELA 12 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL RAZÃO DE ASPECTO.....	76
TABELA 13 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL RENDIMENTO DE NANOFIBRILAÇÃO.	76
TABELA 14 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DOS MODELOS DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA RAZÃO DE ASPECTO DE NANOCELULOSES FIBRILADAS.....	79
TABELA 15 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E VALIDAÇÃO/TESTE DOS MODELOS DE REDE NEURAL PARA PREDIÇÃO DA RAZÃO DE ASPECTO DE NANOCELULOSES FIBRILADAS.....	87
TABELA 16 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DOS MODELOS DE FLORESTA RANDÔMICA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DE NANOCELULOSES FIBRILADAS.	94
TABELA 17 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E VALIDAÇÃO/TESTE DOS MODELOS DE REDE NEURAL PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	101
TABELA 18 - RESULTADOS DE PERMEANÇA PARA AS MEMBRANAS DE PVA PURO TESTADAS EM LABORATÓRIO.....	105
TABELA 19 - RESULTADOS DE SELETIVIDADE CO ₂ /CH ₄ PARA AS MEMBRANAS DE PVA PURO TESTADAS EM LABORATÓRIO....	106
TABELA 20 - RESULTADOS DE PERMEANÇA E SELETIVIDADE PARA AS MEMBRANAS DE PVA E CNF (4%) TESTADAS EM LABORATÓRIO.	107
TABELA 21 - RESISTÊNCIA À TRAÇÃO E MÓDULO DE ELASTICIDADE MEDIDOS PARA AS MEMBRANAS DE PVA PURO E COMPÓSITAS.	113
TABELA 22 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E VARIÁVEIS DE SAÍDA PARA MODELOS DE PREDIÇÃO DE PROPRIEDADES DE PERMEAÇÃO DE GASES.....	116

TABELA 23 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL PERMEANÇA.	120
TABELA 24 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL SELETIVIDADE.	120
TABELA 25 - MÉTRICAS DE DESEMPENHO PARA OS SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA PERMEANÇA DE CO ₂	123
TABELA 26 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE REDE NEURAL PARA PREDIÇÃO DA PERMEANÇA DE CO ₂	127
TABELA 27 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA SELETIVIDADE DE CO ₂	131
TABELA 28 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE REDE NEURAL PARA PREDIÇÃO DA SELETIVIDADE DE CO ₂	134
TABELA 29 - VALORES ABSOLUTOS DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON PARA AS VARIÁVEIS DE ENTRADA EM RELAÇÃO A RAZÃO DE ASPECTO.	163
TABELA 30 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E ÚNICA VARIÁVEL DE SAÍDA CONSIDERADAS PARA A MODELAGEM. .	164
TABELA 31 - RESULTADOS APRESENTADOS PARA MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA RAZÃO DE ASPECTO.	167
TABELA 32 - RESULTADOS DAS REDES NEURAS GERADAS COM DIFERENTES COMBINAÇÕES DE HIPERPARÂMETROS PARA A RAZÃO DE ASPECTO.	168
TABELA 33 - MÉTRICAS DE DESEMPENHO PARA OS MODELOS CONSIDERADOS DE CADA ALGORITMO TESTADO PARA PREDIÇÃO DA RAZÃO DE ASPECTO.	172
TABELA 34 - VALORES ABSOLUTOS DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON PARA AS VARIÁVEIS DE ENTRADA EM RELAÇÃO AO RENDIMENTO DE NANOFIBRILAÇÃO.	173

TABELA 35 - DESCRIÇÃO ESTATÍSTICA DA VARIÁVEL RENDIMENTO DE NANOFIBRILAÇÃO.	174
TABELA 36 - RESULTADOS APRESENTADOS PARA MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	176
TABELA 37 – PRINCIPAIS ARQUITETURAS DE REDES NEURAS GERADAS COM DIFERENTES COMBINAÇÕES DE HIPERPARÂMETROS PARA O RENDIMENTO DE NANOFIBRILAÇÃO.	177
TABELA 38 - MÉTRICAS DE DESEMPENHO PARA OS MODELOS CONSIDERADOS DE CADA ALGORITMO TESTADO PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.	181

LISTA DE ABREVIATURAS OU SIGLAS

ABS	- Absoluto
CNC	- Celulose nanocristalina
CNF	- Celulose nanofibrilada
DEA	- Dietanolamina
EAM	- Erro absoluto médio
EPAM	- Erros percentual absoluto médio
FPSO	- Unidades Flutuantes de Produção, Armazenamento e Descarga de Petróleo
GEE	- Gases do efeito estufa
HPH	- Homogeneizador a alta pressão
IPCC	- Painel Intergovernamental sobre Mudanças Climáticas
MEA	- Monoetanolamina
MEV	- Microscopia eletrônica de varredura
MOF	- <i>Metallic-Organic-Frameworks</i>
NCB	- Nanocelulose bacteriana
PDADMAC	- Poli cloreto de dialildimetillamônio
PHA	- Polihidroxialcanoato
PHB	- Polihidroxibutirato
PLA	- Ácido polilático
PVA	- Álcool polivinílico
PVAm	- Polivinil amina
ReLU	- Rectified Linear Unit
RQEM	- Raiz quadrada do erro médio
RS	- Força relativa
TAPPI	- Technical Association of the Pulp and Paper Industry
TEMPO	- N-OXIL-2,2,6,6- TETRAMETILPIPERIDINA
ATG	- Análise termogravimétrica
UFPR	- Universidade Federal do Paraná

SUMÁRIO

1 INTRODUÇÃO	23
1.1 JUSTIFICATIVA	24
1.2 OBJETIVOS	26
1.2.1 Objetivos gerais.....	26
1.2.2 Objetivos específicos.....	26
2 REVISÃO DA LITERATURA	27
2.1 PROCESSAMENTO DO GÁS NATURAL	27
2.2 TECNOLOGIAS DE SEPARAÇÃO DE CO ₂	29
2.2.1 Membranas	30
2.2.1.1 Método de produção de membranas.....	31
2.2.1.2 Mecanismos de transporte	32
2.2.1.3 Permeança, permeabilidade e seletividade	33
2.3 BIOPOLÍMEROS E BIONANOCOMPÓSITOS	34
2.3.1 Álcool polivinílico (PVA).....	37
2.3.2 Nanocelulose.....	38
2.3.2.1 Obtenção de nanocelulose	39
2.3.2.2 Propriedades da nanocelulose	41
2.4 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA E SUAS APLICAÇÕES	42
2.4.1 Algoritmos de aprendizagem de máquina	42
2.4.1.1 Redes neurais artificiais.....	43
2.4.1.2 Florestas randômicas	46
2.4.2 Algoritmos de otimização de hiperparâmetros	48
2.4.2.1 Procura manual	48
2.4.2.2 Otimização bayesiana	48
2.4.3 Aplicação de aprendizagem de máquina em sistemas com nanocelulose.....	49
2.4.4 Aplicação de aprendizagem de máquina na predição de propriedades de permeação de gases através de membranas.	51
3 MATERIAIS E MÉTODOS	53
3.1 OBTENÇÃO DOS DADOS.....	53
3.1.1 Predição das propriedades da nanocelulose.....	53
3.1.2 Permeança e seletividade de CO ₂ /CH ₄ em bionanocompósitos de PVA e nanocelulose.	54

3.1.2.1	Dados extraídos da literatura.....	55
3.1.2.2	Dados obtidos experimentalmente	57
3.1.3	Propriedades dos bionanocompósitos	60
3.1.3.1	Microscopia eletrônica de varredura (MEV).....	60
3.1.3.2	Análise termogravimétrica (ATG).....	60
3.1.3.3	Resistência à tração e módulo de elasticidade.....	61
3.2	MODELAGEM DOS DADOS.....	61
3.2.1	Linguagem e interface de programação	61
3.2.2	Avaliação preliminar e pré-processamento dos dados	62
3.2.3	Seleção dos algoritmos e variáveis de entrada e saída	63
3.2.4	Ajuste de hiperparâmetros	64
3.2.4.1	Modelos preliminares de predição das propriedades de nanocelulose	64
3.2.4.2	Modelos completos de predição das propriedades da nanocelulose e modelos de predição das propriedades de membrana	67
3.3	AVALIAÇÃO DOS MODELOS	70
4	RESULTADOS E DISCUSSÃO	71
4.1	AVALIAÇÃO DOS DADOS DE POLPAS LIGNOCELULÓSICAS E DE NANOFIBRAS DE CELULOSE PARA PREDIÇÃO DE PROPRIEDADES DE MATERIAL	71
4.2	MODELOS DE PREDIÇÃO: RAZÃO DE ASPECTO DE NANOFIBRAS DE CELULOSE	77
4.2.1	Florestas randômicas	77
4.2.2	Redes neurais	85
4.3	MODELOS DE PREDIÇÃO: RENDIMENTO DE NANOFIBRILAÇÃO DE NANOFIBRAS DE CELULOSE	92
4.3.1	Florestas randômicas	92
4.3.2	Redes neurais	99
4.4	RESULTADOS EXPERIMENTAIS DE PERMEAÇÃO COM MEMBRANAS DE PVA E CNF MECÂNICA	105
4.4.1	Resultados de permeação e seletividade.....	105
4.4.2	Microscopia eletrônica de varredura	108
4.4.3	Análise termogravimétrica	112
4.4.4	Determinação de propriedades mecânicas	113

4.5 AVALIAÇÃO DOS DADOS DE PERMEAÇÃO PARA PREDIÇÃO DE PERMEANÇA E SELETIVIDADE DE MEMBRANAS.....	115
4.6 MODELOS DE PREDIÇÃO: PERMEANÇA AO CO ₂	121
4.6.1 Florestas randômicas	121
4.6.2 Redes neurais	125
4.7 MODELOS DE PREDIÇÃO: SELETIVIDADE CO ₂ /CH ₄	129
4.7.1 Florestas randômicas	129
4.7.2 Redes neurais	133
5 CONCLUSÃO	137
6 SUGESTÕES PARA TRABALHOS FUTUROS	139
REFERÊNCIAS.....	140
APÊNDICE A – BASE DE DADOS PARA PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE: RAZÃO DE ASPECTO E RENDIMENTO DE NANOFIBRILAÇÃO.....	149
APÊNDICE B – BASE DE DADOS PARA MODELAGEM DAS PROPRIEDADES DE MEMBRANA	158
APÊNDICE C – RESULTADOS DA MODELAGEM PRELIMINAR DE PREDIÇÃO DA RAZÃO DE ASPECTO DE CNFS	163
APÊNDICE D – RESULTADOS DA MODELAGEM PRELIMINAR DE PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DE CNFS.....	173

1 INTRODUÇÃO

O aumento da temperatura média da Terra é uma das principais razões para a ocorrência das mudanças climáticas e de desastres naturais mais frequentes nas últimas décadas. Uma das principais causas para isto é a emissão de gases do efeito estufa (GEE) geradas principalmente pelo setor de transporte e energia, que utiliza primariamente combustíveis de origem fóssil, pela indústria de óleo e gás e pelos sistemas de gás natural (RUSSO *et al.*, 2021). Exemplos de componentes incluídos como GEE são o metano (CH₄), sulfeto de hidrogênio (H₂S) e dióxido de carbono (CO₂), sendo que este último corresponde a 76% do total de emissões.

Frente a estes problemas, diversos países e companhias estão se comprometendo a aplicar ou melhorar suas políticas de sustentabilidade, através da assinatura de tratados internacionais ou pela revisão de suas metas estratégicas de negócios, por exemplo. Contudo, apesar dos esforços e das necessidades urgentes, estudos demonstram que a tendência é de que as emissões continuem crescendo pelos próximos 30 anos, ao em vez de diminuírem (HO; LEO, 2021; RUSSO *et al.*, 2021).

Considerando-se este cenário, a separação e captura de carbono surgem como uma abordagem relevante para reduzir sua emissão na atmosfera e mitigar os efeitos decorrentes dessa prática (TAKHT RAVANCHI; SAHEBDEL FAR, 2014). Este processo já é aplicado na indústria de óleo e gás, e uma das operações unitárias aplicadas é a separação de gases por membranas. Este tipo de operação é capaz de obter bons desempenhos de purificação e, geralmente, não requer muita energia, não faz uso de solventes e não necessita de um investimento de capital muito alto (IULIANELLI; DRIOLI, 2020; OLAJIRE, 2010).

Destaca-se, entretanto, que, como a separação de CO₂ tem como objetivo reduzir os impactos ambientais associados ao setor de óleo e gás, é necessário desenvolver a sustentabilidade do processo de forma holística. Assim, deve-se também levar em consideração o material com o qual são produzidas as membranas, de forma que não só o processo, por si só, promova uma pegada de carbono/energia baixa, mas que seus materiais também a tenham. Desta forma, materiais biodegradáveis e preferencialmente obtidos a partir da biomassa são fortes candidatos (ANSALONI *et al.*, 2017; HO; LEO, 2021).

A nanocelulose é um potencial alvo de estudo, uma vez que a celulose é um dos polímeros naturais mais abundantes na Terra. Além disso, a nanocelulose possui excelentes propriedades mecânicas e que podem melhorar o desempenho de outros polímeros quando adicionados na forma de bionanocompósitos (DAI *et al.*, 2019; KLEMM *et al.*, 2018). Com relação a propriedades de permeação de gases, já existem pesquisas em que este material foi aplicado na fabricação de membranas sustentáveis para separação de CO₂ em correntes que continham CH₄, e os resultados até o momento são muito promissores, pois indicam uma melhora no desempenho de separação e seletividade ao CO₂ (JAHAN *et al.*, 2018a, 2018b; SAEED *et al.*, 2020; TORSTENSEN *et al.*, 2019).

Desta maneira, a investigação e a aceleração no desenvolvimento de bionanocompósitos feitos a partir da nanocelulose para aplicação na indústria petrolífera são muito importantes. Para isso, compreender e desenvolver modelos que possam relacionar características da nanocelulose, variáveis de processo e outras propriedades finais dos bionanocompósitos, como propriedades de separação, seriam muito úteis (ÖZKAN *et al.*, 2019). Com o intuito de facilitar a obtenção destes modelos, as técnicas de inteligência artificial e aprendizagem de máquina são fortes candidatas para isso, pois atuam bem na predição de variáveis mais complexas e não lineares, como as citadas anteriormente (GÉRON, 2019).

Este estudo propõe a produção de nanocelulose e sua aplicação em bionanocompósitos com o objetivo de caracterizar e modelar as propriedades destes materiais. O objetivo principal é desenvolver modelos que correlacionem propriedades de matéria-prima, nanocelulose, de processo e outras variáveis importantes com as características finais das nanoceluloses e dos bionanocompósitos, utilizando para isto técnicas de inteligência artificial.

1.1 JUSTIFICATIVA

É urgente a resolução dos problemas relacionados à mudança climática, que se tornaram mais evidentes nas últimas décadas devido à emissão de CO₂, principalmente advindos da indústria de óleo e gás e do uso de combustíveis fósseis no setor de transportes e energia (RUSSO *et al.*, 2021). Segundo o último relatório do Painel Intergovernamental em Mudanças Climáticas (IPCC), predições indicam uma alta probabilidade do aumento da temperatura média da Terra em pelo menos 1,5 °C

nos próximos vinte anos. Este aumento já é suficiente para aumentar o nível do mar e agravar secas e precipitações em diversas regiões do planeta (IPCC, 2012) .

Como mencionado anteriormente, o uso de combustíveis não renováveis no setor de transportes e o processamento industrial de petróleo e gás contabilizam uma grande porcentagem das emissões de CO₂ (24% para o primeiro e 16% para o segundo), geralmente emitido através dos gases de combustão e dos processos de refino (RUSSO *et al.*, 2021; TAKHT RAVANCHI; SAHEBDEL FAR, 2014). Uma operação chave que pode auxiliar a mitigar estas emissões de forma adequada é a separação de gases por membranas, mais alinhada aos princípios de sustentabilidade, uma vez que faz parte da Estratégia de Intensificação de Processos (IULIANELLI; DRIOLI, 2020). Ademais, o desenvolvimento e aperfeiçoamento das tecnologias de processo e de materiais para a redução dos impactos ambientais negativos é estratégico e de grande relevância para o Brasil e para as companhias petrolíferas nacionais, uma vez que fazem parte de acordos internacionais com metas ambientais desafiadoras.

Considerando que a celulose é um dos polímeros naturais mais abundantes na Terra, a nanocelulose pode ser usada para produzir bionanocompósitos para aplicação como membranas. A nanocelulose é biodegradável e de origem renovável e, por conta de suas propriedades, pode melhorar as características mecânicas, de permeação e seletividade dos filmes, além de aumentar a biodegradabilidade de polímeros atualmente empregados.

Além disso, a aprendizagem de máquina é uma ferramenta poderosa, fácil de aplicar e que pode auxiliar na resolução de problemas mais complexos (GÉRON, 2019). Visto a importância do assunto, desenvolver modelos que possam relacionar parâmetros de processo e da nanocelulose com propriedades dos bionanocompósitos pode auxiliar fortemente no avanço deste tipo de material, e assim, contribuir na obtenção de um método de separação de CO₂ mais sustentável.

1.2 OBJETIVOS

1.2.1 Objetivos gerais

O objetivo geral deste estudo é modelar as propriedades da nanocelulose e de bionanocompósitos utilizados como membranas para separação de gases no processamento de gás natural, aplicando para isso algoritmos de aprendizagem de máquina. Dados para a geração dos modelos estão relacionados à produção da nanocelulose fibrilada (CNF) e sua caracterização e à permeação e seletividade de CO₂/CH₄ utilizando membranas de álcool polivinílico (PVA) e PVA e nanocelulose, os quais foram obtidos a partir da literatura e experimentalmente.

1.2.2 Objetivos específicos

Os objetivos específicos desta pesquisa estão elencados abaixo:

- Construir uma base de dados com as propriedades de micro/nanoceluloses fibriladas a partir de diferentes fontes de biomassa e pré-tratamento;
- Obter micro/nanocelulose fibrilada a partir da polpa de eucalipto e aplicando pré-tratamento mecânico e homogeneizador a alta pressão;
- Sintetizar membranas compósitas de PVA com diferentes teores de micro/nanocelulose fibrilada mecânica da polpa de eucalipto;
- Caracterizar estas membranas a respeito da sua morfologia, propriedades mecânicas, químicas e de permeação ao CO₂ e CH₄;
- Construir uma base de dados com informações de permeação e seletividade ao CO₂ obtidas em laboratório e da literatura também;
- Avaliar os parâmetros que possam ter um efeito na determinação de propriedades importantes da nanocelulose, como razão de aspecto e rendimento de nanofibração, e construir modelos de aprendizagem de máquina para suas predições;
- Avaliar parâmetros que tenham um potencial efeito na permeação e seletividade das membranas e construir modelos de aprendizagem de máquina para suas predições.

2 REVISÃO DA LITERATURA

2.1 PROCESSAMENTO DO GÁS NATURAL

O gás natural é uma fonte importante de combustíveis e serve como matéria-prima para a produção de diversos derivados químicos. Além disso, é mencionado diversas vezes como uma fonte de energia de transição entre o óleo combustível e o carvão em direção ao uso de energias mais limpas e renováveis, devido ao menor potencial poluente e por poder ser utilizado em infraestruturas já existentes ou que requeiram adaptações de baixo custo (CHEN; LIU; JIN, 2021; EPE, 2018).

Conforme a lei 9.478/1997, conhecida como “Lei do Petróleo e Gás”, o gás natural pode ser definido como *“todo hidrocarboneto que permaneça em estado gasoso nas condições atmosféricas normais, extraído diretamente a partir de reservatórios petrolíferos ou gaseíferos, incluindo gases úmidos, secos, residuais e gases raros”*. Com relação à sua composição, é formado por uma mistura de hidrocarbonetos gasosos, principalmente da fração C1 até C4, sendo o metano (CH₄) seu componente principal. Além destes, podem estar presentes outras substâncias caracterizadas como impurezas como água, CO₂, H₂S e metais pesados (CNI, 2019).

A composição específica do gás natural bruto depende de alguns fatores naturais que influenciam em como as reservas de óleo e gás se formaram e se acumularam na bacia de origem. Uma composição genérica para o gás natural está apresentada na TABELA 1. Nesta tabela, observa-se que a quantidade de CO₂ presente pode ser de até 1%. Contudo, nas bacias do pré-sal, esta fração de CO₂ é maior ainda, podendo atingir valores de 10 até 20% (ALMEIDA, 2016).

TABELA 1 - COMPOSIÇÃO GENÉRICA DO GÁS NATURAL.

Componente	Variação na composição (% molar) (traços ≤ 0.1%)
Metano	70-98
Etano	1-10
Propano	traços-5
Butano	traços-2
Pentano	traços-1
Hexano	traços-0,5
Heptano e maiores	traços
Ciclopropano	traços
Ciclohexano	traços
Benzeno	traços
Nitrogênio	traços-15
Dióxido de carbono	traços-1
Gás sulfídrico	traços ocasionalmente
Hélio	traços-5
Água	traços-5
Outros compostos de nitrogênio e enxofre	traços ocasionalmente

Fonte: BISPO (2011) apud OBERLAENDER (2015)

De qualquer maneira, para fins de transporte, distribuição e utilização, é necessário que o gás natural esteja conforme padrões pré-determinados pela Resolução 16/2008 da Agência Nacional de Petróleo (AGÊNCIA NACIONAL DE PETROLÉO GÁS NATURAL E BIOCOMBUSTÍVEIS., 2008). Isto pode ser alcançado através do seu processamento, ou seja, uma sequência de operações unitárias a que se submete o gás natural (BRASIL, 1997).

O petróleo extraído dos poços passa por uma primeira etapa de tratamento conhecida como processamento primário. As operações unitárias requeridas dependerão da fase de desenvolvimento do campo de petróleo e podem abranger os seguintes estágios, dentre outros (OBERLAENDER, 2015):

- Separação do petróleo nas fases gasosas, líquidas e aquosas;
- Processar o óleo extraído, retirando-se água e sais dissolvidos;
- Processar o gás bruto, de forma a reduzir as frações de água, CO₂, H₂S e outras impurezas;

A etapa de desumidificação e redução de gases ácidos é um processo crítico, uma vez que evita a emissão de GEE e poluentes na atmosfera, corrige a qualidade do gás necessário para transporte e venda, evita a danificação de equipamentos e tubulações, o entupimento de linhas e a redução da eficiência de combustão do gás. Portanto, é uma fase do processamento que é de grande relevância e que muitas vezes acaba por torná-lo oneroso (FREITAS *et al.*, 2021). Em geral, inicialmente é realizada a retirada dos gases ácidos da corrente de gás natural (etapa de adoçamento/dessulfurização), e em seguida reduz-se o teor de água, em forma líquida ou de vapor, através do processo de desidratação. O adoçamento é comumente feito através de tecnologias como absorção, adsorção e separação por membranas (COELHO, 2007; THOMAS, 2004).

2.2 TECNOLOGIAS DE SEPARAÇÃO DE CO₂

Tecnologias geralmente empregadas para realizar a separação de CO₂ em correntes gasosas compreendem os processos de absorção química, de adsorção física e química, destilação criogênica e separação por membranas.

A absorção é uma operação unitária já estabelecida na indústria em geral para separação de CO₂ em correntes gasosas, principalmente em situações de pós-combustão. A rota por absorção química com uso de solventes à base de aminas, como MEA (monoetanolamina) e DEA (dietanolamina), é uma técnica economicamente viável e que apresenta ótimas taxas de recuperação do dióxido de carbono. Contudo, está associada a problemas de corrosão em equipamentos e tubulações e, devido à necessidade de regeneração dos solventes para que o CO₂ capturado seja liberado, a demanda energética desta operação unitária é alta, em torno da faixa de 4-6 MJ/ kg de CO₂ (IULIANELLI; DRIOLI, 2020). Além disso, requer um controle rigoroso da operação e, se tratando de Unidades Flutuantes de Produção, Armazenamento e Descarga de Petróleo (FPSO em inglês), não são muito adequadas, pois a instabilidade das plataformas leva a quedas de rendimento da operação (CHEN *et al.*, 2017).

A adsorção gás-sólido possui vantagens em relação à absorção química, como por exemplo, não fazer uso de solventes líquidos e, portanto, ter menos problemas de corrosão e ambientais. Contudo, para a regeneração do sólido, gastos energéticos relativamente consideráveis ainda são requeridos, na faixa de 2-3 MJ/kg

de CO₂. Ademais, os adsorventes comerciais disponíveis atualmente não possuem seletividades e capacidades altas, impedindo que esta tecnologia seja aplicada em escalas industriais de forma efetiva (IULIANELLI; DRIOLI, 2020).

A separação de gases por membranas é uma outra operação unitária que também pode ser aplicada para captura de CO₂ e que possui diversas vantagens em relação às explicitadas anteriormente. Por utilizar equipamentos mais compactos e que demandam um baixo consumo energético, está alinhada às diretrizes da estratégia de intensificação de processos. Para mais, possuem alta flexibilidade e facilidade de operação. Estas características a tornam a opção mais adequada para plantas *offshore* de processamento de gás natural (FREITAS *et al.*, 2021; RUSSO *et al.*, 2021). Porém, perduram algumas desvantagens ainda que se mantêm como desafios a serem superados, como problemas de plastificação e o contrabalanço entre permeabilidade e seletividade, que precisam ser melhorados.

Percebe-se, desta forma, que a tecnologia de separação por membranas é uma operação unitária chave em relação às tecnologias convencionais mencionadas anteriormente, pois possui um bom desempenho para separação de CO₂, requer baixos custos de investimento de capital, sua operação é fácil e requer pouca manutenção, demanda pouca energia e não faz uso de solventes, evitando problemas de corrosão. Ademais, considerando que é um processo de baixa pegada energética e sua aplicação está intimamente relacionada à redução da emissão de CO₂, é importante também que o material das membranas seja advindo de uma fonte renovável com baixa pegada de carbono. Desta forma, é imprescindível no contexto atual investigar novos materiais que estejam de acordo com uma economia circular e verde (HO; LEO, 2021).

2.2.1 Membranas

Uma membrana pode ser definida como uma interface através da qual componentes químicos são separados. Elas podem ser produzidas a partir de polímero, metais ou materiais cerâmicos. As membranas podem ter estruturas diferentes e são comumente divididas em dois grupos principais: membranas simétricas e anisotrópicas (HABERT; BORGES; NOBREGA, 2006).

Membranas simétricas são aquelas que apresentam uma única estrutura ao longo da sua espessura, podendo ser microporosas, densas ou eletricamente

carregadas. As primeiras possuem poros de magnitude de 0,01 a 10 μm . O fenômeno de separação ocorre, principalmente, pela diferença de tamanho das moléculas presentes no fluido a ser separado e na distribuição do tamanho dos poros. Em membranas densas não porosas, por sua vez, o regime de transporte das espécies químicas ocorre por difusão e solubilidade, e não pela diferença de tamanho das moléculas. Assim, em aplicações de separação de gases, por exemplo, componentes de tamanho molecular similar podem ser separados por conta da diferença de concentração de cada espécie química no material da membrana. Por fim, as membranas carregadas eletricamente possuem em sua estrutura íons carregados positiva ou negativamente e a separação ocorre, principalmente, devido à carga e concentração de cada espécie química presente no fluido (ABDULLAH *et al.*, 2018; BAKER, 2004).

Membranas anisotrópicas podem ser formadas pelo mesmo material ou não (neste último caso, são classificadas como compósitos) e apresentam mais de uma estrutura ao longo da matriz. Estas diferentes estruturas podem ser formadas ao mesmo tempo ou em etapas separadas. Estas membranas possuem a grande vantagem de possuírem um suporte mecânico, que não afeta a separação dos componentes, e uma camada seletiva mais fina, o que melhora o fluxo das espécies químicas através da sua espessura (BAKER, 2004).

2.2.1.1 Método de produção de membranas

A depender da metodologia aplicada, serão obtidas membranas com estruturas específicas. Membranas do tipo densas podem ser produzidas através dos métodos de evaporação de solvente (*solution casting*) e prensagem a quente (*thermal melt pressing*). A técnica por evaporação de solvente consiste na dissolução do polímero em um solvente de fácil evaporação e posterior espalhamento da mesma em uma superfície. Em um ambiente controlado, ocorre a evaporação do solvente, criando-se um filme denso do polímero. Esta técnica é comumente utilizada em laboratórios devido à sua facilidade de fabricação, contudo, por geralmente se obter membranas de espessuras maiores e ser do tipo densa e não porosa, o fluxo de permeação tende a ser mais baixo em razão da resistência apresentada pela espessura (HABERT; BORGES; NOBREGA, 2006). A prensagem a quente é uma técnica utilizada para polímeros que não dissolvem muito bem em solventes à

temperatura ambiente. Basicamente, a técnica consiste em comprimir o polímero entre duas placas aquecidas (BAKER, 2004).

Para membranas anisotrópicas, as técnicas frequentemente aplicadas são a inversão de fases e deposição de filme em suporte (*solution coating*). A técnica de inversão de fases consiste na separação de duas fases, uma sólida, rica em polímero, e outra líquida, pobre em polímero e que forma os poros (BAKER, 2004). A deposição de filme é outra técnica que consiste em criar um fino filme em cima de um suporte poroso. Nesse caso, a parte da membrana seletiva propriamente dita é o filme de polímero, que devido a sua espessura fina, permite um fluxo maior de permeação. O suporte poroso não interfere muito no transporte das moléculas gasosas (HABERT; BORGES; NOBREGA, 2006).

2.2.1.2 Mecanismos de transporte

Solução-difusão

Comumente, o mecanismo de transporte predominante em membranas densas é do tipo solução-difusão. Neste tipo de transporte, a transferência de massa ocorre devido a dois fatores: a interação química do componente gasoso em relação ao polímero (solubilidade) e a transferência devido ao tamanho da molécula (difusão). A transferência de massa através da membrana ocorre em razão do gradiente de potencial químico entre uma interface e outra (BAKER, 2004; RUSSO *et al.*, 2021).

O fluxo de uma espécie química através de uma membrana polimérica pode ser descrito pela EQUAÇÃO (1):

$$J = \frac{DS(p_1 - p_2)}{l} \quad (1)$$

Na EQUAÇÃO (1), D é o coeficiente de difusividade da molécula, S é o coeficiente de solubilidade, $(p_1 - p_2)$ representa a diferença de pressão entre os dois lados da membrana e l corresponde à espessura da mesma. Esta equação deriva da segunda lei de Fick e da lei de Henry, as quais descrevem um processo não estacionário de transferência de massa e da concentração de componentes, respectivamente, em função da pressão e de termos de solubilidade (BAKER, 2004; ISMAILA *et al.*, 2005).

Transporte facilitado

O transporte facilitado é um mecanismo onde um componente químico gasoso tem seu fluxo facilitado através de uma membrana devido a reações reversíveis entre ele e carreadores presentes no polímero, enquanto outros gases presentes na mistura permanecem inertes e são transportados somente pelo mecanismo de solução difusão (DENG; KIM; HÄGG, 2009). No caso do CO₂, o seu transporte pode ser facilitado através da sua reação com água, gerando o íon HCO₃⁻, o qual apresenta um coeficiente de difusão maior (DENG; HÄGG, 2010). Na outra interface da membrana, o dióxido de carbono é restituído (HONG et al., 2021; TORSTENSEN et al., 2019).

2.2.1.3 Permeança, permeabilidade e seletividade

A permeança é uma propriedade calculada através da divisão entre a vazão volumétrica de permeação (\dot{V}) pela respectiva área (A) e diferença de pressão (ΔP) aplicada à membrana, conforme a EQUAÇÃO (2). Geralmente é representado em GPU, o qual corresponde a $10^{-6} \frac{cm^3}{cm^2 \cdot s \cdot cmHg} = 2,7 \cdot 10^{-3} \frac{m^3}{m^2 \cdot h \cdot bar}$ (IULIANELLI; DRIOLI, 2020).

$$Pe = \frac{\dot{V}}{A \cdot \Delta P} \quad (2)$$

A permeabilidade de uma espécie química é outra propriedade e pode ser calculada multiplicando o coeficiente de solubilidade (S) e o coeficiente de difusão (D) ou multiplicando a permeança Pe pela espessura da matriz polimérica σ (HABERT; BORGES; NOBREGA, 2006), conforme a EQUAÇÃO (3).

$$P = Pe \cdot \sigma \quad (3)$$

A permeabilidade pode ser expressa em $\frac{cm^3 \cdot cm}{cm^2 \cdot s \cdot cmHg}$ ou, mais comumente, em Barrer, o qual equivale à $10^{-10} \frac{cm^3 \cdot cm}{cm^2 \cdot s \cdot cmHg}$.

A seletividade ideal de um gás é descrita pela razão entre as permeabilidades ou permeanças individuais de cada componente, conforme a EQUAÇÃO (4). Também

pode ser feita a relação entre as concentrações de cada espécie gasosa no permeado e no retentado provenientes de uma mistura dos componentes. Neste último caso chama-se de fator de separação, apresentado na EQUAÇÃO (5) (RUSSO et al., 2021).

$$\alpha_{i/j} = \frac{P_i}{P_j} \quad (4)$$

$$SF = \frac{X_{i,p}/X_{j,p}}{X_{i,f}/X_{j,f}} \quad (5)$$

2.3 BIOPOLÍMEROS E BIONANOCOMPÓSITOS

A produção de membranas a partir de materiais poliméricos é de grande interesse devido aos baixos custos e à facilidade no processamento (RUSSO et al., 2021). Contudo, pensando no contexto da aplicação em separação e captura de carbono, é pertinente reiterar a relevância de expandir o conceito da sustentabilidade associada a esse processo aos materiais com os quais são fabricadas as membranas. Desta forma, materiais poliméricos que sejam obtidos a partir de fontes renováveis, que sejam baratos e que estejam disponíveis de forma abundante, contribuindo para uma economia circular e menor pegada de carbono associado ao material, são características desejáveis (HO; LEO, 2021).

Existem algumas referências que definem os biopolímeros de uma forma mais abrangente e outras de forma mais restritiva. Por exemplo, segundo Christian (2016) e Smith, Moxon e Morris (2016), biopolímeros são materiais produzidos a partir de organismos vivos, tanto a partir de vegetais quanto a partir de bactérias e fungos. Outra abordagem mais ampla inclui também na definição alguns polímeros de origem fóssil e que são biodegradáveis (IBRAHIM *et al.*, 2019; RUSSO *et al.*, 2021).

Alguns exemplos de biopolímeros são o amido termoplástico, celulose, quitosana, ácido polilático (PLA), polihidroxialcanoatos e butiratos (PHA e PHB) e álcool polivinílico (PVA). As características atrativas desses polímeros estão relacionadas, principalmente, à maior sustentabilidade, abundância e disponibilidade da matéria-prima quando produzidos a partir de fontes renováveis, biocompatibilidade e biodegradabilidade (IBRAHIM *et al.*, 2019; RUSSO *et al.*, 2021).

Bionanocompósitos são materiais obtidos através da incorporação de um nanomaterial, ou seja, que possui pelo menos uma das suas dimensões na escala nanométrica (1-100 nm), em uma matriz biopolimérica (FERNANDES *et al.*, 2013; REDDY *et al.*, 2013). Geralmente, o objetivo por trás destas misturas está relacionado ao melhoramento de propriedades da fase considerada contínua, como as propriedades mecânicas, térmicas e químicas. São inúmeros os tipos de nanomateriais que podem ser aproveitados para este fim, desde componentes químicos inorgânicos, como alguns metais e zeólitas, e orgânicos, como fibras naturais e nanocelulose, a exemplo da celulose nanocristalina (CNC) e celulose nanofibrilada (CNF) (FERNANDES *et al.*, 2013).

Tratando-se da aplicação no processamento de gás natural e separação de CO₂, o estudo e a aplicação de biopolímeros para este fim é relativamente recente e têm apresentado resultados promissores, com seletividades favorecendo o transporte de dióxido de carbono em detrimento ao metano e permeabilidades comparáveis ou próximas das obtidas para polímeros tradicionais. Uma exceção em relação a novos estudos é o uso de membranas de acetato de celulose, pois é amplamente empregado no processamento de gás natural desde a década de 80 (IULIANELLI; DRIOLI, 2020; RUSSO *et al.*, 2021). Valores para permeabilidade e seletividade relativos à separação de CO₂ através de membranas biopoliméricas constam na TABELA 2 abaixo.

TABELA 2 - DADOS DE PERMEABILIDADE E SELETIVIDADE AO CO₂ PARA DIFERENTES BIOPOLÍMEROS

Polímero	Gases	Permeabilidade ao CO ₂ (Barrer)	Seletividade	OBS	Referência
PVA puro	CO ₂ /CH ₄	20,74*	21,75	P = 5 bar, 100% UR	(JAHAN <i>et al.</i> , 2018a)
PVA (98,5%) /CNC (1,5%)	CO ₂ /CH ₄	70,20*	39,16	P = 5 bar, 100% UR	(JAHAN <i>et al.</i> , 2018a)
PVA (98,5%) /CNC (1,5%)	CO ₂ /CH ₄	51,84*	36	P = 10 bar, 100% UR	(JAHAN <i>et al.</i> , 2018a)
PVA (98,5%) /CNC (1,5%)	CO ₂ /CH ₄	29,16*	30	P = 15 bar, 100% UR	(JAHAN <i>et al.</i> , 2018a)
CNF	CO ₂ /N ₂	17,66	515,79	P = 1 bar, T = 35°C 85% UR	(ANSALONI <i>et al.</i> , 2017)
	CO ₂ /CH ₄		363,7		
Acetato de celulose	CO ₂ /N ₂	13,8	39,5	P = 1 bar, T = 25°C	(KIM; LEE, 2015)
Acetato de celulose semicristalino	CO ₂ /CH ₄	1,8-6	32	T = 35°C	(NGUYEN <i>et al.</i> , 2019)
PLA	CO ₂ /CH ₄	65	250		(IULIANELLI; RUSSO; GALIANO, 2019)

Legenda: P – Pressão, T – Temperatura, UR – Umidade relativa

*: permeações medidas originalmente em m³(STP)/m² bar h e transformadas em barrer para fins de comparação

Fonte: Adaptado de RUSSO *et al.* (2021)

Dos polímeros mais tradicionais utilizados na indústria de gás natural tem-se as polímidas, polímidas fluoradas, poliamidas e polisulfonas. Dados de permeabilidade e seletividade referentes a esses polímeros estão apresentados na TABELA 3.

TABELA 3 - DADOS DE PERMEABILIDADE E SELETIVIDADE AO CO₂ PARA DIFERENTES POLÍMEROS TRADICIONAIS

Polímero	Gases	Permeabilidade ao CO ₂ (barrer)	Seletividade	OBS	Referência
Pebax 1657	CO ₂ /CH ₄	98	16	T = 35°C, P = 4 bar	(REIJERKERK <i>et al.</i> , 2010)
Matrimid 5218	CO ₂ /CH ₄	10	36	T = 35°C, P = 35 bar	(VU; KOROS; MILLER, 2003)
6FDA-durene	CO ₂ /CH ₄	458	16	T = 35°C, P = 10 bar	(LIU <i>et al.</i> , 2003)
6FDA-durene /mPDA (50:50)	CO ₂ /CH ₄	85	30	T = 35°C, P = 10 bar	(LIU <i>et al.</i> , 2003)
PIM-1	CO ₂ /CH ₄	2300	18	T = 25°C, P = 1 bar	(BUDD <i>et al.</i> , 2005)

Legenda: P – Pressão, T – Temperatura, UR – Umidade relativa

Fonte: Adaptado de IULIANELLI e DRIOLI (2020)

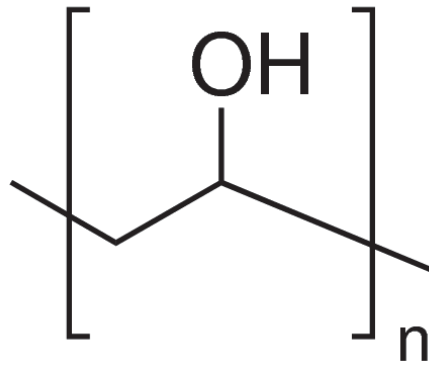
É possível observar a partir da TABELA 2 e TABELA 3 que os dados de permeação e seletividade das membranas tradicionais comerciais são, na sua maioria, superiores. Contudo, biopolímeros que contenham nanocelulose na matriz ou alguma funcionalização com aminas, por exemplo, apresentam maiores permeabilidades e seletividades. Estes dados demonstram que a utilização de biopolímeros para a separação de CO₂ de correntes que contenham metano tem potencial de aplicação, como por exemplo as membranas produzidas a partir de PVA. Este fato é reforçado também ao levar-se em conta que os testes experimentais foram realizados em condições úmidas, situação que ocorre no processamento de gás natural.

2.3.1 Álcool polivinílico (PVA)

O álcool polivinílico (PVA) é um polímero obtido a partir do acetato de vinila, através da sua polimerização e posterior hidrólise. O monômero correspondente ao álcool polivinílico está representado na FIGURA 1. O PVA é hidrofílico, não é tóxico, é biodegradável, possui boas propriedades mecânicas e apresentou resultados de permeação e seletividade ao CO₂ que indicam grande potencial para sua aplicação no processamento de gás natural através da separação de gases por membranas,

conforme mostrado na TABELA 2 (DOBLE; KUMAR, 2005; HAVSTAD, 2020; SAEED *et al.*, 2020).

FIGURA 1 - MONÔMERO DO PVA.



2.3.2 Nanocelulose

Nanocelulose é um material natural, hidrossolúvel, celulósico, que possui pelo menos uma de suas dimensões na escala nanométrica (1-100 nm). Abrange três principais tipos, a depender das características morfológicas e do processo de obtenção: celulose nanocristalina ou nanocristais de celulose (CNC), celulose nanofibriladas, também chamada de nanofibras de celulose, (CNF) e nanocelulose bacteriana (NCB) (KLEMM *et al.*, 2018). As principais diferenças estão elencadas na TABELA 4.

TABELA 4 - PRINCIPAIS CARACTERÍSTICAS MORFOLÓGICAS DE NCB, CNF E CNC.

Tipo de nanocelulose	Comprimento	Seção transversal	Grau de polimerização	Grau de cristalinidade
NCB	Diferentes tipos de rede de nanofibras	20 - 100 nm	4.000 – 10.000	Alto grau de cristalinidade
CNF	0,1 – 2 µm	5 – 60 nm	≥ 500	Menor grau de cristalinidade
CNC	100 – 250 nm (celulose vegetal) 100 nm – vários µm (a partir celulose bacteriana, algica)	5 – 70 nm	500 – 15.000	Médio grau de cristalinidade

Fonte: Adaptado de KLEMM *et al.* (2018)

Como se pode observar pela TABELA 4, as CNFs possuem um grau de cristalinidade menor em relação às outras formas de nanocelulose. Isto é devido à presença de regiões tanto cristalinas quanto amorfas na cadeia da celulose. Já as CNCs e NCBs possuem graus de cristalinidade maior. Além disso, as nanofibras possuem comprimentos maiores, na ordem de micrômetros, enquanto os comprimentos de CNCs são menores, de até 250 nanômetros (0,25 microns). Estas diferenças morfológicas refletem no seu aspecto. Por exemplo, as CNCs têm um formato parecido com grãos de arroz seco (chamados de *whiskers*), enquanto que as CNFs são finas e compridas, semelhantes a fios de espaguete. A nanocelulose bacteriana, por sua vez, possui altos graus de cristalinidade, contudo, possuem uma morfologia mais parecida com a formação de fibrilas, e não tanto como *whiskers*.

2.3.2.1 Obtenção de nanocelulose

A obtenção de cada tipo de nanocelulose, seja CNF ou CNC especificamente, dependerá fortemente do processo de produção. Fatores que influenciam nisso são a quantidade de energia aplicada no processo, a dosagem de reagentes e a composição química da matéria prima (KLEMM *et al.*, 2018). Essencialmente, a nanocelulose é gerada a partir da desintegração da fonte lignocelulósica utilizada. A estrutura da celulose contida nas plantas se organiza de forma hierárquica. Isto quer dizer que as paredes celulares das plantas são formadas por macrofibras de celulose, as quais são constituídas por microfibras de celulose e que, por sua vez, são constituídas pelas fibrilas elementares de celulose (TRACHE *et al.*, 2020). Esta estruturação pode ser observada na FIGURA 2.

FIGURA 2 - ESTRUTURA HIERÁRQUICA DA CELULOSE EM PLANTAS.

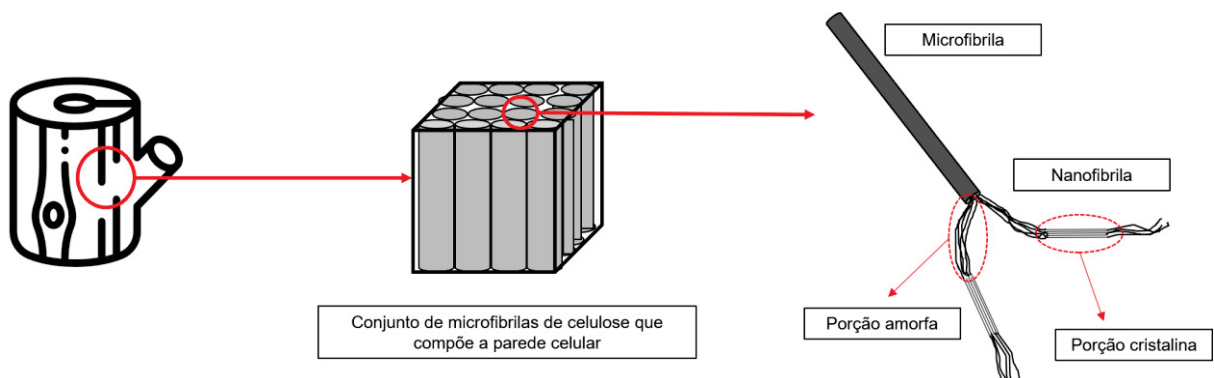


Imagem criada com recursos gratuitos do Flaticon.com

Fonte: Adaptado de NASIR *et al.* (2017)

As fibras elementares de celulose são formadas por regiões amorfas e cristalinas. A região cristalina é caracterizada por um alto grau de organização e fortes ligações de hidrogênio entre as cadeias de celulose, o que confere alta resistência e maior insolubilidade a diferentes solventes. Já a região amorfa possui ligações mais fracas, por isso é mais acessível a diferentes compostos e solventes químicos. Portanto, é mais propensa à quebra das ligações intermoleculares (VISAKH; THOMAS, 2010).

Desta maneira, as nanofibras de celulose, que são constituídas tanto por regiões amorfas quanto cristalinas, são geralmente produzidas a partir da delaminação mecânica das microfibrilas de celulose presentes nas polpas lignocelulósicas até se chegar nas estruturas elementares. Alguns processos comumente aplicados para isto são a homogeneização a alta pressão, microfluidização e moagem (SERRA-PARAREDA *et al.*, 2021a). Estas operações mecânicas requerem um alto valor energético para operar e para obter bons rendimentos de fibras na escala nanométrica. Com o intuito de reduzir a demanda energética e gerar materiais mais homogêneos em termos de diâmetro e comprimento, geralmente são empregados pré-tratamentos químicos juntamente com a desfibrilação mecânica, como o pré-tratamento através de oxidação TEMPO (N-OXIL-2,2,6,6- TETRAMETILPIPERIDINA), enzimáticos ou até mesmo mecânicos (DAI *et al.*, 2019). Estes pré-tratamentos facilitam a geração das fibras pois permitem que elas inchem e adquiram grupos funcionais em suas cadeias, promovendo a repulsão e conseqüente individualização das mesmas (SERRA-PARAREDA *et al.*, 2021d).

Já os nanocristais de celulose, como são caracterizados por serem praticamente a região cristalina das fibrilas elementares, necessitam de um processo mais agressivo para sua obtenção. As regiões amorfas, como explicado anteriormente, são mais suscetíveis à hidrólise devido ao impedimento estérico reduzido e a fatores cinéticos. Desta forma, geralmente as CNCs são produzidas através de hidrólise ácida, principalmente utilizando-se ácido sulfúrico (TRACHE *et al.*, 2020). Outros ácidos foram estudados recentemente em novos métodos de produção de nanocristais, dentre eles ácido maleico, oxálico e cítrico. O tipo de ácido influencia na carga superficial que será atribuída à nanocelulose gerada, nas suas

características morfológicas e na sua estabilidade coloidal e térmica (KLEMM *et al.*, 2018).

2.3.2.2 Propriedades da nanocelulose

A aplicação de cada tipo de nanocelulose em determinados ramos industriais depende diretamente das suas características morfológicas, cristalinidade, carga superficial e propriedades reológicas, térmicas e mecânicas. Na produção de bionanocompósitos, como na produção de membranas, a nanocelulose desempenha um papel de agente de reforço e como auxiliar no transporte de CO₂ através do filme (KLEMM *et al.*, 2018; TORSTENSEN *et al.*, 2019).

A respeito da sua capacidade de melhorar propriedades mecânicas, a nanocelulose pode incrementar consideravelmente propriedades como módulo de Young e resistência à tração. Por sua vez, o grau de melhoria depende das suas características morfológicas, como diâmetro e comprimento. Portanto, a razão de aspecto, correspondente à razão entre comprimento e diâmetro das fibras ou cristais, é uma propriedade muito relevante na caracterização das nanoceluloses (LEONG *et al.*, 2022).

Ademais, o tamanho da nanocelulose, e conseqüentemente a sua razão de aspecto, foi apontado como um fator relevante na permeação de CO₂ em membranas de PVA e polivinilamina (PVAm) para separação do carbono em correntes de gás natural e de combustão, de acordo com o relatado por alguns pesquisadores (JAHAN *et al.*, 2018a, 2018b; SAEED *et al.*, 2020; TORSTENSEN *et al.*, 2019). Evidências apontam que um menor tamanho de partícula promove uma melhor dispersão na matriz polimérica. Como as nanoceluloses possuem grupos hidroxilas ao longo das suas cadeias, que podem se ligar às moléculas de água, uma melhor dispersão significa uma rede mais coesa e melhor distribuída de água através da membrana. Como a água dissolve o CO₂ e gera íons bicarbonatos, que por sua vez possuem uma constante de difusão melhor do que espécies como o CH₄, são criados “caminhos preferenciais” nas membranas através da inserção de nanocelulose, o que facilita o transporte do gás de interesse. Além disso, a carga superficial também é relevante no processo de permeação de CO₂, pois também influencia na distribuição de água através da membrana e pode afetar a dispersão da nanocelulose na matriz.

O rendimento de nanofibrilação é um parâmetro importante e que também exerce uma contribuição positiva na resistência à tração de filmes poliméricos à base de nanocelulose (ESPINOSA *et al.*, 2020). Esta propriedade, apesar de não ser uma característica intrínseca a este nanomaterial, relaciona a massa obtida de fibras nanométricas ao final do processo de desfibrilação em comparação com a massa inicial. Assim, percebe-se que é relevante medi-la, justamente pela influência que exerce nas propriedades mecânicas e também, como parâmetro de processo, para se alcançar sua produção em escala industrial (SERRA-PARAREDA *et al.*, 2021b).

2.4 TÉCNICAS DE APRENDIZAGEM DE MÁQUINA E SUAS APLICAÇÕES

Aprendizagem de máquina é uma subárea da inteligência artificial. Como o próprio nome remete, é a área da programação em que a própria máquina consegue aprender e extrair informações a partir de dados que o usuário provê. Com este aprendizado, a máquina consegue gerar modelos de predição ou classificação, por exemplo, e melhorar o desempenho desses modelos ao longo do tempo, sem que o usuário precise programar este aperfeiçoamento de forma explícita (GÉRON, 2019). Aprendizagem de máquina é uma ferramenta que tem ganhado grande destaque e tem aplicações em diversas áreas, a exemplo dos ramos financeiro, comercial, econômico, educacional, da engenharia e ciência dos materiais.

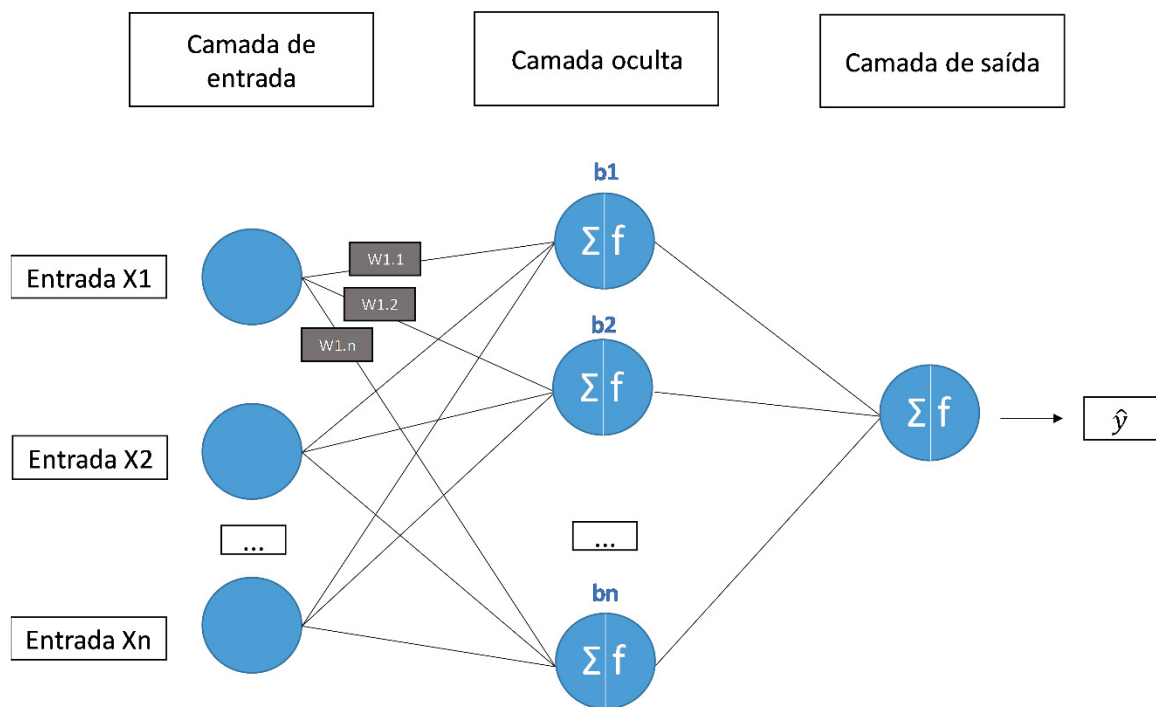
2.4.1 Algoritmos de aprendizagem de máquina

Existem diversos algoritmos que podem ser utilizados. Alguns dos mais conhecidos são as redes neurais artificiais, florestas randômicas, árvores de decisão, regressão linear múltipla, k-vizinhos mais próximos e máquina de vetores de suporte (TORRENTS-BARRENA; PELLICER, 2021). Por sua vez, a escolha dos algoritmos irá depender de alguns fatores, a exemplo do seu desempenho para o conjunto de dados utilizados, a complexidade do problema, o objetivo de interesse em relação à modelagem, entre outros.

2.4.1.1 Redes neurais artificiais

Rede neural artificial é um algoritmo que, inspirada na rede de neurônios humana, tenta criar conexões entre variáveis de interesse a partir de uma estrutura baseada em unidades processadoras que estão interligadas. É formada por camadas de neurônios, que estão conectadas entre si através de pesos associados aos nós. São essencialmente de três tipos: a camada de entrada, a(s) camada(s) oculta(s) e a camada de saída (HAYKIN, 1999). A FIGURA 3 representa uma rede neural genérica com uma única camada oculta e um único neurônio de saída.

FIGURA 3 - REPRESENTAÇÃO DE UM EXEMPLO DE REDE NEURAL COM DUAS CAMADAS OCULTAS E DOIS OUTPUTS DE SAÍDA.



Fonte: A autora (2023)

Observando a FIGURA 3, as variáveis de entrada da rede neural estão representadas como X1, X2 e Xn. Percebe-se também que cada neurônio de entrada está conectado por uma linha a cada neurônio da camada oculta. O mesmo ocorre entre os neurônios da camada oculta e da camada de saída. Essas interligações estão associadas a um peso, que cada neurônio precedente recebe. Aqui, estão

representados os pesos do neurônio X1 com cada neurônio da camada intermediária: W1.1, W1.2 e W1.n. Estes pesos representam a força de conexão entre os neurônios. Além dos pesos, outro parâmetro importante é o *bias*. Pela figura, percebe-se que existe um valor desse associado a cada neurônio da camada oculta. Ele tem como propósito permitir a ativação do neurônio mesmo quando a variável de entrada é nula. O *bias* é adicionado ao somatório ponderado realizado entre os pesos e os respectivos neurônios precedentes e que estão conectados ao neurônio do *bias* em questão.

Também pode-se verificar, na figura, os símbolos Σf nos neurônios da camada oculta e de saída, representando as funções de ativação. Funções de ativação são responsáveis por introduzir não-linearidade às redes, auxiliando-as a captar relações complexas entre os dados de entrada e a variável de interesse de saída. Realiza-se o somatório ponderado dos neurônios da camada precedente por seus respectivos pesos mais o *bias*, como mencionado no parágrafo anterior e, a depender do tipo de função, é definido como esta soma é transformada em um valor de saída do nó. Desta maneira, se o resultado desta função for um valor maior que zero ou um valor nulo, podem “ativar” ou não o respectivo neurônio.

As funções de ativação podem ser lineares ou não lineares e alguns exemplos são a *Rectified Linear Unit* (ReLU), sigmoidal e linear. A do tipo ReLU é uma das funções mais empregadas atualmente pois geralmente opera de forma mais rápida e produz melhores resultados e melhor generalização que outras funções. Ela é caracterizada por atribuir o valor nulo para entradas negativas, enquanto que entradas positivas são mantidas no mesmo valor, como apresentado na EQUAÇÃO (6) (NWANKPA *et al.*, 2018).

$$f(x) = \max(0, x) = \begin{cases} x_i, & \text{if } x_i \geq 0 \\ 0, & \text{if } x_i < 0 \end{cases} \quad (6)$$

A função sigmoidal também não é linear e foi muito utilizada no início da pesquisa com aprendizagem profunda (*deep learning*). É de fácil compreensão, é muito usada em redes mais rasas e como função de ativação entre a última camada escondida e a camada de saída em problemas de classificação binária. A função matemática, apresentada na EQUAÇÃO (7), gera respostas entre 0 e 1 (DUBEY; SINGH; CHAUDHURI, 2022).

$$f(x) = \left(\frac{1}{(1+exp^{-x})} \right) \quad (7)$$

A função de ativação linear não é empregada entre a camada de entrada e as camadas ocultas e entre as camadas ocultas pois não fornece nenhum tipo de não linearidade ao sistema, o que é uma característica geralmente desejada ao se aplicar redes neurais na modelagem de problemas (DUBEY; SINGH; CHAUDHURI, 2022). Contudo é comumente utilizada entre a última camada escondida e a camada de saída em problemas de regressão. Como a função linear, também chamada de função identidade, não modifica o valor recebido como entrada, ela deve ser aplicada em modelos de predição de quantidades, pois não se deseja modificar a resposta de saída final calculada pela rede (JIANG; JIANG; CHEN, 2022). A EQUAÇÃO (8) mostra a função identidade.

$$f(x) = x \quad (8)$$

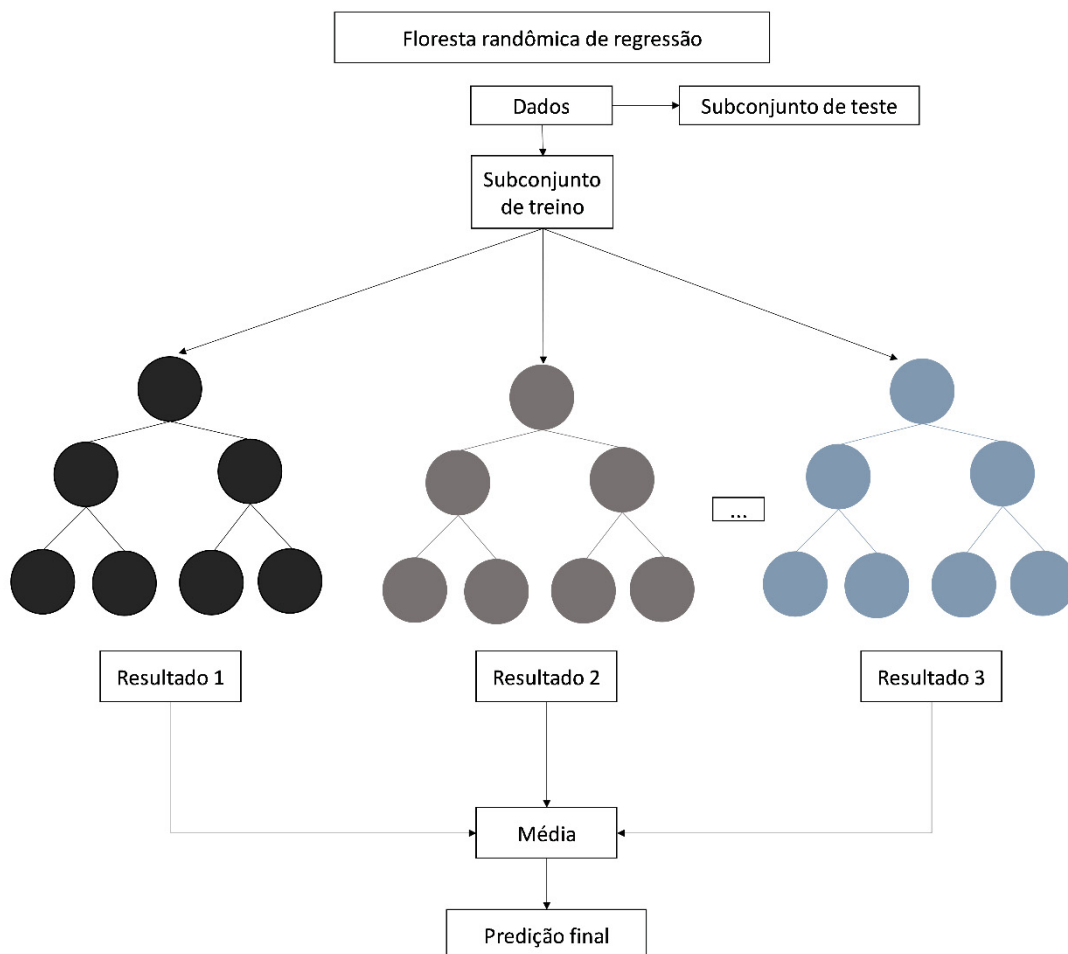
Os pesos mencionados até o momento estão diretamente relacionados na qualidade de predição do modelo de rede neural e são ajustados pela própria máquina por meio de funções de otimização interna. Para isso, é definida também uma “função de perda”, ou seja, uma função que irá calcular o erro entre a predição realizada pela rede e a resposta real. Esta comparação entre valor real e predito é feita até se atingir um erro mínimo e faz parte da etapa de treinamento da rede (CHRISTOPHER M BISHOP, 1996; DUDA; HART, 1973) .

O número de neurônios nas camadas, a quantidade de camadas ocultas e o tipo de função de ativação utilizado são chamados de hiperparâmetros do modelo. Estes hiperparâmetros não conseguem ser otimizados pelo próprio algoritmo e necessitam que o usuário faça o ajuste para que bons resultados de predição sejam gerados. Além destes, outros hiperparâmetros importantes são a taxa de aprendizagem, a função de otimização interna utilizada para o treinamento da rede, a quantidade de dados utilizada para treinar a rede por iteração – também chamado de tamanho do lote - , e o número de iterações total, também chamado de número de épocas (GÉRON, 2019).

2.4.1.2 Florestas randômicas

A floresta randômica é um algoritmo de combinação, ou *ensemble* em inglês. Isto quer dizer que este algoritmo gera previsões otimizadas ao se basear não somente em um único modelo de previsão, mas em um conjunto destes. Desta forma, corresponde a um agrupamento de árvores de decisão, o qual por si só já é um algoritmo de aprendizagem de máquina. A resposta de saída de um modelo do tipo floresta randômica é dada pela resposta mais votada entre as árvores de decisão que fazem parte do modelo, no caso de um problema de classificação, ou na média do valor gerado por cada árvore, no caso de problemas de regressão (BREIMAN, 2001). Esta estrutura está representada esquematicamente na FIGURA 4.

FIGURA 4 - REPRESENTAÇÃO ESQUEMÁTICA DE UMA FLORESTA RANDÔMICA PARA PROBLEMA DE REGRESSÃO.



Fonte: adaptado de SIGNORI-IAMIN *et al.* (2022)

Na FIGURA 4, é possível observar a estrutura de cada árvore de decisão que formam uma floresta randômica. A árvore inicia pela raiz (círculo mais acima de cada árvore) e, a partir deste ponto, o subconjunto de dados utilizados para treinar as árvores é sequencialmente dividido. Esta divisão ocorre de acordo com o que é decidido pelos nós de decisão, que são baseados nas variáveis de entrada do modelo.

As florestas randômicas possuem melhores desempenhos do que as árvores de decisão individuais pois “injetam” aleatoriedade conforme geram e ajustam um certo número de árvores independentes. Isto quer dizer que, conforme gera-se cada árvore, é selecionada uma certa quantidade aleatória de linhas de dados para ser o subconjunto de treinamento de cada árvore. No caso dos nós de decisão, é selecionado um subconjunto aleatório de variáveis de entrada para realizar as divisões. Por fim, os nós de resposta são chamados de nós do tipo folha, os quais não podem mais ser subdivididos (BREIMAN, 2001).

No caso de florestas randômicas para problemas de classificação, a divisão dos nós ocorre de forma a reduzir a entropia e obter o maior ganho de informação possível. Aqui, entropia representa uma métrica que indica a impureza de uma divisão em uma única árvore de decisão. Uma entropia de 0 corresponde a um nó totalmente puro, ou seja, em que se utilizando uma certa quantidade de variáveis de entrada divide-se 100% de dados em uma única classe. Uma entropia de valor 1 corresponde a um nó totalmente impuro, ou seja, onde a classe de resposta para os dados possui probabilidades uniformes de serem classificadas. Já o ganho de informação tende a ser maximizado. Este ganho é uma métrica que representa a redução da entropia antes e após a divisão do nó (GÉRON, 2019).

No caso de florestas randômicas para problemas de regressão, a função comumente utilizada para medir a qualidade da divisão dos nós é o erro entre valor predito e valor real do output de interesse. O algoritmo, portanto, busca sempre reduzir esse erro ao construir cada árvore. Esse erro pode ser representado pelo erro médio absoluto, erro médio quadrático ou desvio de Poisson (BREIMAN, 2001; SCIKIT-LEARN, 2022).

Os hiperparâmetros relativos a este algoritmo incluem o número de árvores, a máxima profundidade de cada árvore, o número máximo de variáveis de entrada levado em consideração na hora de fazer a divisão, o número mínimo de dados que devem estar presentes em cada nó e folha, dentre outros (PEDREGOSA; WEISS; BRUCHER, 2011).

2.4.2 Algoritmos de otimização de hiperparâmetros

Como mencionado previamente, os hiperparâmetros dos algoritmos de aprendizagem de máquina devem ser ajustados pelo próprio usuário. Estes hiperparâmetros são muito relevantes pois influenciam diretamente na qualidade de predição dos modelos. Algumas técnicas para ajustar esses valores estão descritas nas seções a seguir e são baseadas, em última análise, na combinação de valores de hiperparâmetros que proporcionam o modelo com melhor desempenho. Ou seja, são técnicas que não requerem o gradiente da função para realizar a otimização.

2.4.2.1 Procura manual

A procura manual é uma das formas mais simples para se realizar a otimização de hiperparâmetros. Nela, o próprio usuário modifica os valores dos parâmetros, roda os modelos e avalia o desempenho dos mesmos para escolher qual o melhor conjunto ajustado. É uma técnica simples e fácil de aplicar, contudo é demorada e limitada, a depender de quantos hiperparâmetros o usuário deseja avaliar. Uma estratégia plausível a ser aplicada é estabelecer alguns valores para os hiperparâmetros e modificar um parâmetro por vez, sempre a partir da última melhor configuração encontrada (BENGIO, 2012).

2.4.2.2 Otimização bayesiana

A otimização bayesiana é uma técnica que têm ganhando grande destaque entre a comunidade de ciência de dados, pois é uma forma de busca de hiperparâmetros considerada mais inteligente. Uma de suas características principais é que leva em conta o desempenho do modelo obtido na última iteração de busca para encontrar um novo ponto de teste. Isto ocorre porque é baseado na regra de probabilidade condicional de Bayes, representada pela EQUAÇÃO (9) (SHAHRIARI *et al.*, 2016).

$$p(w|D) = \frac{p(D|w)p(w)}{p(D)} \quad (9)$$

Na EQUAÇÃO (9), w corresponde a uma quantidade não observada e D é uma informação. Assim, $p(w|D)$ corresponde à distribuição posterior, ou seja, a probabilidade de w ocorrer sabendo-se da informação D . $P(D|w)$ representa a probabilidade de D considerando que w é verdadeira, $p(w)$ corresponde à distribuição anterior, ou seja, a probabilidade de w ocorrer antes de se considerar a informação D , e $p(D)$ corresponde à probabilidade da evidência ocorrer (SHAHRIARI *et al.*, 2016). Portanto, por conta do seu princípio de funcionamento, a otimização bayesiana tende a ser mais eficiente que outras técnicas para encontrar pontos ótimos, como em comparação a procura manual, mencionada anteriormente, ou a procura do tipo aleatória (SHIN *et al.*, 2020).

Ademais, o método baseia-se em um modelo de substituição para simular a superfície de resposta da função objetivo real (por exemplo, o coeficiente de determinação R^2 ou o erro médio absoluto do modelo) e uma função de aquisição. A função de aquisição serve para indicar qual o próximo ponto a ser testado pelo modelo, no caso, onde a função de aquisição é máxima (BROCHU; CORA; DE FREITAS, 2010; TURNER *et al.*, 2021). Conforme explicado anteriormente, a função de Bayes leva em consideração o resultado da iteração anterior para escolher o próximo ponto de teste. Assim, à medida que o número de iterações cresce, a função de substituição é ajustada de acordo com as novas observações que são feitas, ficando mais próxima da função objetivo real. Estas novas observações são baseadas nos pontos máximos da função de aquisição, que também é ajustada a cada nova iteração.

2.4.3 Aplicação de aprendizagem de máquina em sistemas com nanocelulose

Técnicas de aprendizagem de máquina já foram exploradas para modelar e estudar processos envolvendo a produção de celulose, a predição de propriedades mecânicas de bionanocompósitos derivados destes, e algumas propriedades da nanocelulose pura. Contudo, estas aplicações são recentes e, portanto, há muitas oportunidades de investigação na área.

Ozkan *et al.* (2019) utilizaram três algoritmos para prever propriedades mecânicas de bionanocompósitos de PVA e CNFs do tipo TEMPO reticuladas tanto com carbonato de amônio e zircônio quanto com glioxal, um tipo de aldeído. Os pesquisadores utilizaram regressão linear múltipla, redes neurais e florestas

randômicas para prever o módulo de Young, a resistência à tração e o alongamento de 6 filmes, produzidos a partir de três componentes com composições diferentes. As entradas dos modelos foram as propriedades mecânicas de filmes de CNF pura e de 9 tipos de filmes de dois componentes (CNF e PVA, CNF e carbonato de amônio e zircônio e CNF e glioxal), os quais tiveram concentrações da nanocelulose variando de 70-90%. Para o módulo de Young, a rede neural gerou as melhores previsões, enquanto que, para a resistência à tração, o desempenho das redes e das florestas randômicas foram melhores que a regressão linear. A regressão linear se mostrou mais acurada para a previsão do alongamento, seguida pela rede neural e floresta randômica.

Redes neurais foram utilizadas para prever o comprimento das fibras de celulose em um processo de produção de papel, considerando principalmente os efeitos da etapa de refino da polpa sob as propriedades morfológicas das fibras resultantes (ALMONTI *et al.*, 2019). Para tal, foram criados dois modelos, um a partir do conjunto total de 72 dados experimentais e um segundo a partir da redução para 52 pontos após tratamento dos dados. O modelo resultante com mais dados experimentais contou com uma arquitetura do tipo 13-25-1, ou seja, 13 neurônios na entrada, 25 neurônios na única camada escondida e 1 neurônio de saída, e o último modelo com uma arquitetura do tipo 11-7-1. O modelo do tipo 11-7-1 obteve um R^2 de 0,98 e um erro médio percentual de 1,07% somente.

Florestas randômicas, *gradient boosting* e máquina de vetores de suporte foram algoritmos empregados para prever o fator de enriquecimento celulósico e a recuperação de sólidos do processo de pré-tratamento de 45 tipos de biomassa diferentes. Para os pré-tratamentos, foram empregadas 80 variedades de solventes de líquidos iônicos, tipo de solvente utilizado para reduzir a recalcitrância da matéria-prima e ser melhor processada (PHROMPHITHAK; ONSREE; TIPPAYAWONG, 2021). O melhor desempenho foi obtido a partir das florestas randômicas, que obteve uma raiz quadrada do erro médio de 0,22 e um R^2 de 0,94 para o fator de enriquecimento e um erro de 0,07 e R^2 de 0,84 para a recuperação de sólidos.

Redes neurais com arquitetura 4-4-1 foram utilizadas também para prever a razão de aspecto de micro/nanofibras de celulose treinadas com 15 dados advindos de polpa de Pinus e validados com dados de outras fontes lignocelulósicas (abeto, álamo e eucalipto), utilizando-se como variáveis de entrada a quantidade de celulose, a energia aplicada no processo de fibrilação e o diâmetro e comprimento das fibras

da polpa pré-tratada (SANTOS *et al.*, 2022). O desempenho do modelo criado foi excelente, atingindo um valor de $R = 0,99$, e a validação gerou correlações muito satisfatórias ($R = 0,84$), considerando a diversidade das matérias-primas.

2.4.4 Aplicação de aprendizagem de máquina na predição de propriedades de permeação de gases através de membranas.

Recentemente, foram reportados alguns trabalhos acerca da predição da permeabilidade e seletividade de componentes gasosos em separação por membranas. Por exemplo, EBRAHIMI *et al.*, (2016) produziram membranas de matriz mista de polietersulfona com partículas de óxido de grafeno funcionalizados com amina e tiveram com objetivo prever a seletividade CO_2/CH_4 . Foram geradas membranas com diferentes teores de partícula (0 até 10%) e os gases foram empregados a diferentes pressões (0,5, 1, 2, 3 e 4 bar) e temperaturas (303, 323 e 343 K). 504 pontos experimentais (70% para treino e 30% para teste) foram obtidos para a modelagem utilizando um tipo de rede neural, chamado rede neural com função de base radial. As variáveis de entrada foram o teor de óxido de grafeno funcionalizado, a pressão e a temperatura. O modelo criado pelos pesquisadores teve uma capacidade preditiva muito satisfatória. Para o treino, foi obtido um R de 0,9033 e um erro quadrado médio de 0,0399, enquanto para o teste foi obtido um R de 0,9220 e um erro quadrado médio de 0,0345.

Outro exemplo foi a utilização do algoritmo de floresta randômica para prever a capacidade de membranas de matriz mistas baseadas em *metallic-organic-frameworks* (MOFs) na separação de CO_2 em correntes com CH_4 , para assim otimizar o desenvolvimento de novas membranas (GUAN *et al.*, 2022). Para isso, utilizaram 648 pontos experimentais obtidos da literatura acerca da permeação com MOFs na separação de CO_2 e CH_4 (como gases puros ou em mistura). Nesta base de dados, constaram 36 tipos de MOFs e 41 tipos diferentes de polímeros. As variáveis de saída foram a permeabilidade de CO_2 e a seletividade CO_2/CH_4 . As variáveis de entrada consideradas envolveram características das MOFs, características dos polímeros, características das membranas de matriz mista e condições de processo. São eles: tamanho de poro e área superficial das MOFs, tipo de MOF, teor de MOF na membrana, tamanho da partícula, tamanho da abertura dos poros, tipo de polímero,

permeabilidade do polímero puro, seletividade do polímero puro, espessura da membrana, pressão e temperatura.

O modelo gerado para a permeabilidade obteve um R^2 de 0,91 para o subconjunto de treino e 0,77 para o de teste, enquanto para a predição de seletividade obteve-se um R^2 de 0,89 para o treino e 0,70 para o teste. De acordo com os autores, a diferença entre as métricas de desempenho de treino e teste pode ter ocorrido devido à complexidade do processo estudado, além da coleção de informações baseados em artigos de grupos diferentes e com certa limitação de dados. Ainda assim, as métricas obtidas são satisfatórias. Ademais, a partir deste modelo foram desenvolvidos dois tipos de membranas de matriz mista que alcançaram um comportamento de permeabilidade e seletividade acima do limite superior de Robeson, (que correlaciona o contrabalanço entre permeabilidade e seletividade de membranas) e com boas propriedades mecânicas e melhor resiliência à plastificação e envelhecimento.

Outro trabalho que vale a pena mencionar é o desenvolvido por ZHANG *et al.*, (2022), que também utilizaram aprendizagem de máquina para prever o desempenho de permeabilidade e seletividade de CO_2 em comparação com N_2 utilizando compósitos de MOF com líquido iônico, e assim acelerar o desenvolvimento destes materiais. Para isso, iniciaram com 8167 dados a respeito da mistura de moléculas de $[\text{NH}_2\text{-Pmim}][\text{Tf}_2\text{N}]$ em MOF gerados com um software de simulação. Com estes pontos, treinou-se um modelo de floresta randômica para a predição da permeabilidade de CO_2 baseado em fatores estruturais e químicos das membranas, como grau de insaturação e percentual de metais na estrutura, densidade, fração de vazios, volume livre, volume acessível, dentre outros. A partir deste modelo, foram identificadas as variáveis principais que auxiliariam na obtenção de membranas com melhor desempenho. O resultado final obtido envolveu uma floresta randômica com R^2 de 0,728. As variáveis mais importantes, de acordo com o modelo, foram o volume acessível e área superficial gravimétrica, para a seletividade e permeabilidade respectivamente.

3 MATERIAIS E MÉTODOS

3.1 OBTENÇÃO DOS DADOS

3.1.1 Predição das propriedades da nanocelulose

Os dados para a predição de propriedades da nanocelulose foram obtidos em trabalho parceiro com o LEPAMAP-PRODIS da Universitat de Girona, os quais foram publicados recentemente em outras revistas (AGUADO *et al.*, 2022; BALEA *et al.*, 2021; SANCHEZ-SALVADOR *et al.*, 2022; SERRA-PARAREDA *et al.*, 2021b). Estes dados estão apresentados nos Quadros 3 e 4, no APÊNDICE A. As polpas lignocelulósicas utilizadas foram de 10 tipos no total, as quais, na sua maioria, foram submetidas a dois tipos de pré-tratamento separadamente: mecânico e enzimático. As polpas foram: polpa de Pinus (*Pinus radiata*) tratada termo mecanicamente (TMP), polpa kraft de Pinus branqueada (BKSP), polpa kraft de Pinus não branqueada (UKSP), polpa de Abeto (*Picea abies*) branqueada e termo mecanicamente tratada (BTMP), polpa de cânhamo (*Cannabis sativa L.*), polpa de juta (*Corchorus capsularis*), polpa de sisal (*Agave Sisalana*), polpa de Álamo (*Populus tremuloides*), polpa de Eucalipto (*Eucalyptus Globulus*) e polpa de papel reciclado.

Todas as polpas citadas receberam dois tipos de pré-tratamento, separadamente: o mecânico, o qual consistiu em um refinamento em um moinho PFI conforme os princípios de design da TAPPI T248 a 20.000 revoluções, e o enzimático (com exceção da polpa BTMP). O pré-tratamento enzimático foi realizado em duas condições de dosagem diferentes: 80 e 240 g/ton.

As polpas pré-tratadas, com uma consistência de 1 %p/p foram submetidas à etapa de desfibrilação em um homogeneizador a alta pressão (HPH) de modelo NS1001L PANDA 2 K-GEA (GEA Niro Soavy, Parma, Italy). Foram aplicados números crescentes de passes (de 3 a 9) e de pressões (300 a 900 bar), o que resultou em cinco níveis de energia aplicada nas polpas. A quantidade de energia consumida foi medida através de um medidor de energia Circutor (Barcelona, Spain), modelo CVM-C10. Desta maneira, 140 CNFs diferentes foram geradas e caracterizadas.

Acerca da caracterização das polpas pré-tratadas, as medidas de cinzas e extraíveis foram determinadas a partir de técnicas gravimétricas após combustão a 525°C e extração com etanol/benzeno, seguindo a metodologia TAPPI (TAPPI, 2020).

Carboidratos e lignina total foram determinados seguindo o procedimento analítico NREL/TP-510-42618 (SLUITER *et al.*, 2012).

A demanda catiônica foi determinada a partir de titulação potenciométrica com um polieletrólito catiônico, o poli cloreto de dialildimetillamônio (PDADMAC). Para isso, foi utilizado um detector de partícula carregada Mütek PCD 04 (BTG Instruments, Germany) (SERRA-PARAREDA *et al.*, 2021c).

Os parâmetros reológicos foram determinados a partir de suspensões de 1% de consistência com um viscosímetro rotacional do tipo Couette (cilindro co-axial), modelo PCE-RVI 2 V1L (PCE Instruments, Germany). A velocidade rotacional (N_i) foi variada de 0,5 a 200 rpm e a taxa de cisalhamento γ foi calculada para cada velocidade levando em conta as dimensões do vaso cilíndrico (R_o) e do eixo (R_i) (SERRA-PARAREDA *et al.*, 2021b).

As curvas logarítmicas de fluxo foram então ajustadas segundo a EQUAÇÃO (10) de Ostwald-de-Waele, onde k corresponde ao índice de consistência, n corresponde ao índice de fluxo e η representa a viscosidade aparente.

$$\eta = k \cdot \gamma^{-n} \quad (10)$$

Os valores de transmitância a 600 nm foram medidos em suspensões a 0,1 %p/p com um espectrofotômetro UV-Vis Shimadzu UV-160A.

O rendimento de nanofibrilação foi estimado a partir da fração não sedimentável após centrifugação de uma amostra de CNF a 0,2% de consistência submetido a 3400g por 20 minutos. A razão de aspecto foi estimada pela técnica do ponto de gelificação, calculado a partir da altura de sedimentos a várias concentrações (SANCHEZ-SALVADOR *et al.*, 2021; VARANASI; HE; BATCHELOR, 2013).

3.1.2 Permeança e seletividade de CO₂/CH₄ em bionanocompósitos de PVA e nanocelulose.

Os dados para a modelagem da permeação e seletividade de CO₂/CH₄ em membranas de PVA com diferentes teores de nanocelulose foram obtidos por duas maneiras: uma parte do banco de dados consistiu em resultados experimentais obtidos de trabalhos científicos de terceiros e outra parte foi resultante de atividades experimentais realizadas pela própria mestranda.

3.1.2.1 Dados extraídos da literatura

Os dados da literatura foram coletados de três artigos, especificamente de Jahan *et al.* (2018a; 2018b; 2020). No total, são 91 pontos de permeação de CO₂ em mistura com CH₄, utilizando membranas de PVA e celulose nanofibrilada (CNF) ou celulose nanocristalina (CNC) na composição. Deste total, 11 pontos se referem à permeação em membranas de PVA puro, 18 em membranas de PVA e CNF e 62 pontos são relativos ao comportamento de permeação para membranas de PVA e CNC.

Na TABELA 5 e TABELA 6 estão apresentadas as faixas de valores para as variáveis mais relevantes para a permeação descritas nos trabalhos, para cada tipo de membrana. Os testes que foram realizados na UFPR abrangeram os pontos experimentais dentro destas faixas de valores, na medida do possível de acordo com a estrutura existente.

TABELA 5 - FAIXA DE VALORES DAS VARIÁVEIS RELEVANTES PARA PERMEAÇÃO EM MEMBRANAS DE PVA E CNC.

Variável	Faixa de valores
Comprimento (nm) da nanocelulose	130
Diâmetro (nm) da nanocelulose	5,9
Razão de aspecto (calculada)	22,03
Concentração nanocelulose (%)	0 – 4
pH da solução mãe	6, 9 e 10
Temperatura (°C)	ambiente
Pressão (bar)	5, 10 e 15
Umidade relativa (%)	30 - 100

Fonte: A autora (2023)

TABELA 6 - FAIXA DE VALORES DAS VARIÁVEIS RELEVANTES PARA PERMEAÇÃO EM MEMBRANAS DE PVA E CNF.

Variável	Faixa de valores
Comprimento (nm) da nanocelulose	750
Diâmetro (nm) da nanocelulose	3
Razão de aspecto (calculada)	250
Concentração nanocelulose (%)	0 – 4
pH da solução mãe	5, 9 e 12
Temperatura (°C)	Ambiente
Pressão (bar)	5, 10 e 15
Umidade relativa (%)	95

Fonte: A autora (2023)

Nos artigos, todas as membranas foram produzidas com PVA de grau de hidrólise de 89% e massa molecular média de 89.000 a 90.000 do fornecedor Sigma Aldrich. A celulose nanocristalina foi comprada da Sigma Aldrich também, possuindo uma razão de aspecto de 22 nm aproximadamente. A celulose nanofibrilada foi produzida em laboratório, a partir da fosforilação com $(\text{NH}_4)_2\text{HPO}_4$ da polpa de madeira (60% de abeto norueguês e 40% de pinho escocês) dissolvida com sulfitos (GHANADPOUR *et al.*, 2015) e a razão de aspecto estimada foi de 250. As membranas foram produzidas pela técnica de deposição de filme em suporte utilizando soluções de 2% de PVA com diferentes concentrações de nanocelulose, variando de 0,5 a 4% em massa. O pH das soluções foi ajustado utilizando uma solução de 0,1 M de NaOH e variou de 5 a 12. A membrana seletiva foi suportada em folhas de polissulfona comercial para ultrafiltração.

Os ensaios de permeação foram realizados utilizando uma mistura de 40% molar de CO_2 e 60% molar de CH_4 , umidificada antes de entrar na célula de permeação. A umidade relativa dos gases variou de 30 a 100% e os ensaios foram realizados à temperatura ambiente. Os experimentos foram conduzidos em uma câmara onde a pressão de operação variou entre 5, 10 e 15 bar. A composição do permeado foi analisada utilizando cromatografia gasosa.

3.1.2.2 Dados obtidos experimentalmente

Para os pontos obtidos experimentalmente na UFPR, foram produzidos dois tipos de membranas diferentes: PVA puro e PVA com 4% de CNF mecânica de baixa carga (em relação a massa de PVA). O PVA (número CAS 9002-89-5) que foi utilizado possui um grau de hidrólise entre 87-89% e peso molecular médio na faixa de 85.000 – 124.000, do fornecedor Sigma-Aldrich.

A CNF mecânica foi fornecida pelo grupo LEPAMAP-PRODIS (Universitat de Girona) em suspensão de 1,54 %p/p em água. Para obtenção da nanocelulose, foi usado como matéria-prima polpa branqueada de eucalipto (*Eucalyptus Globulus*). A polpa foi primeiramente submetida ao pré-tratamento mecânico, que consistiu na aplicação de 20.000 rotações em um moinho PFI seguindo os princípios de desenho do equipamento de acordo com o procedimento padrão TAPPI T248 (TAPPI, 2020).

Sequencialmente à etapa de pré-tratamento mecânico, a polpa, a uma consistência de 1 %p/p, foi submetida à etapa de desfibrilação em um homogeneizador a alta pressão (HPH) de modelo NS1001L PANDA 2 K-GEA (GEA Niro Soavy, Parma, Italy). Foram aplicados 3 passes de 300 bar, 3 passes de 600 bar e 3 passes de 900 bar. Após a desfibrilação, obtiveram-se as celuloses nanofibriladas finais.

Todas as caracterizações da CNF utilizada na UFPR foram realizadas pelo laboratório parceiro. A polpa original de eucalipto foi caracterizada em termos da sua composição química, em especial os teores de celulose, hemicelulose, lignina total, cinzas e extraíveis. Cinzas e extraíveis foram determinadas a partir de técnicas gravimétricas após combustão a 525°C e extração com etanol/benzeno, seguindo a metodologia TAPPI (TAPPI, 2020). Carboidratos e lignina total foram determinados seguindo o procedimento analítico NREL/TP-510-42618 (SLUITER *et al.*, 2012).

As CNFs que foram utilizadas para produzir os bionanocompósitos foram caracterizadas em relação à razão de aspecto, carga superficial e viscosidade de uma suspensão a 1 %p/p. Na realidade, a primeira propriedade é a mais relevante para a modelagem, pois é a única característica da matéria-prima relatada nos artigos utilizados para construção do banco de dados de permeação e seletividade, além do tipo de nanocelulose (mecânica, fosforilada, cristalina etc).

A razão de aspecto foi calculada a partir da viscosidade intrínseca de uma suspensão a 1 %p/p de CNF em água deionizada. Para a determinação da

viscosidade, foi utilizado o viscosímetro Discovery HR10 da TA Instruments equipado com um sensor do tipo cone e placa (diâmetro de 40 mm e ajuste do cone de 2°). A viscosidade medida a uma taxa de cisalhamento de 1000 s⁻¹ foi utilizada para calcular as viscosidades específicas de acordo com a metodologia proposta por HUGGINS, (1942), KRAEMER (1938) e SIMHA (1940).

A carga superficial foi determinada a partir de titulação potenciométrica com um polieletrólito catiônico, o poli cloreto de dialildimetilamônio (PDADMAC). Para isso foi utilizado um detector de partícula carregada Mütek PCD 04 (BTG Instruments, Germany) (SERRA-PARAREDA *et al.*, 2021c).

No QUADRO 1 estão apresentadas a composição química da polpa de eucalipto utilizada para produção das nanoceluloses, bem como as características das CNFs.

QUADRO 1 - CARACTERIZAÇÃO DA MATÉRIA-PRIMA DE ORIGEM E CNF MECÂNICA

Celulose (%p/p)	Hemicelulose (%p/p)	Lignina total (%p/p)	Extraíveis (%p/p)	Cinzas (%p/p)
74,5	16,9	7,0	1,2	0,4
Razão de aspecto	Carga Superficial (µmol/g)	Viscosidade a 1 %p/p a 1000 s⁻¹ (Pa.s)		
64,3	229	0,00869		

Fonte: A autora (2023)

As membranas feitas na UFPR foram produzidas através da técnica de casting (LIU; CHAKMA; FENG, 2008; VENTURI; ANSALONI; BASCHETTI, 2016), por ser uma metodologia mais simples de aplicar em laboratório. A seguinte sequência de etapas foi realizada para produzir as membranas:

- Adicionar em um recipiente com tampa água deionizada e o PVA puro de acordo com a proporção necessária para adicionar a nanocelulose depois. Manter fixo sempre um teor de sólidos fixo de 2% em massa.
- Deixar a solução agitando em *roller* a frio por 48 horas
- Para completar a dissolução, chegando em uma solução límpida, aquecê-la a 75°C por 30 minutos mantendo agitação magnética constante.
- Para a produção de membranas com CNF, pesar a quantidade necessária, em balança analítica, em um béquer a parte.

- Adicionar a CNF ao recipiente, fazendo a transferência da solução de polímero ao béquer e retornando a suspensão ao recipiente por três vezes.

- Manter o recipiente fechado, aquecido a 75°C e agitando na chapa magnética intensamente para correta dispersão da suspensão. Realizar esta etapa por mais 30 minutos.

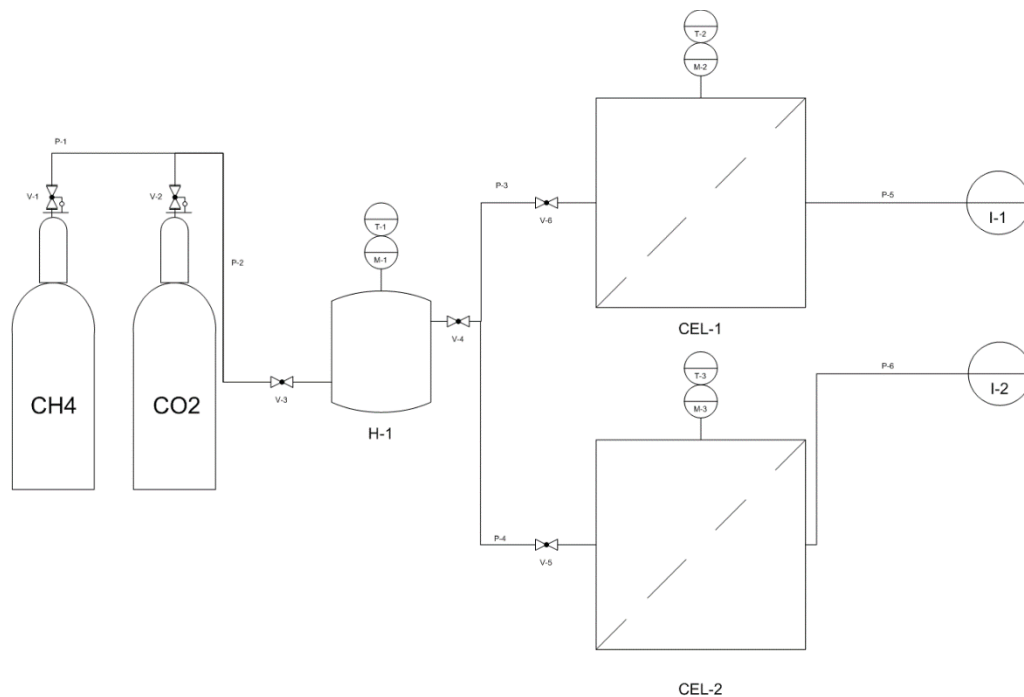
- Após a solução de PVA puro ou a suspensão com CNF estarem prontas, aplicar 20 ml da solução ou suspensão em uma placa de teflon de 10 cm de diâmetro, com o auxílio de uma pipeta volumétrica. No caso da suspensão de CNF, permanecer em agitação baixa neste momento para mantê-la sempre homogênea.

- Secar as membranas por 48 horas em estufa de secagem com circulação de ar (marca Nova Ética, modelo 400/4ND300°C, 320 W de potência) a 30°C.

- Posteriormente, submeter a tratamento térmico a 105°C em estufa de secagem (marca Ethik Technology , modelo 440-1D 27 litros, 2400 W de potência) por 1 hora para retirar qualquer água residual presente no sistema e permitir que ocorra a reticulação entre as cadeias de PVA e a hornificação das cadeias de celulose. A hornificação é o processo pelo qual as cadeias de celulose perdem a água adsorvida nelas, cristalizando-as e reduzindo seu tamanho. A consequência disto é que a nanocelulose não se dissolve novamente em água (VENTURI; ANSALONI; BASCHETTI, 2016).

Para a etapa da permeação de gases através das membranas, foi utilizado uma célula de permeação de volume variável/pressão constante. Os testes foram realizados com CH₄ e CO₂ separadamente. Estes gases foram umidificados previamente à sua entrada no equipamento. A umidificação foi realizada passando cada gás separadamente em um tanque de 800 ml contendo água deionizada à temperatura ambiente. O gás saturado em água adentrava então às duas células de permeação, organizadas em paralelo. Para medir a vazão de gás de permeado, foi utilizada uma bureta para aferir a vazão das bolhas de gás do permeado. Uma representação esquemática do sistema montado pode ser observada na FIGURA 5.

FIGURA 5 - FLUXOGRAMA DO PROCESSO DE UMIDIFICAÇÃO E PERMEAÇÃO DOS GASES.



Legenda: V - Válvulas globo. H – Umidificador. CEL – Célula de permeação. I – Instrumento de medição de vazão das bolhas. T – Medidor de temperatura. M – Manômetro.

Fonte: A autora (2023)

3.1.3 Propriedades dos bionanocompósitos

3.1.3.1 Microscopia eletrônica de varredura (MEV)

A morfologia da seção transversal das membranas foi inspecionada através da técnica de microscopia eletrônica de varredura. Para isso, as membranas foram fraturadas com N_2 líquido. A seção transversal foi metalizada com ouro e analisada em um microscópio modelo VEGA3 TESCAN a uma intensidade de feixe de elétrons de 15 kV.

3.1.3.2 Análise termogravimétrica (ATG)

As análises de ATG foram realizadas utilizando um equipamento de análise térmica modelo Perkin Elmer ATG 4000. As análises foram realizadas com um fluxo de 50 mL/min de N_2 , a uma taxa de aquecimento de 10K/min. A faixa de temperatura avaliada foi de 30 – 900 °C.

3.1.3.3 Resistência à tração e módulo de elasticidade

Para medição das propriedades de resistência à tração e módulo de elasticidade, foram produzidos corpos de prova à maneira da metodologia descrita nas seções anteriores, porém de tamanho 2 cm X 8 cm. O equipamento utilizado foi um texturômetro CT3 da Brookfield. A distância inicial entre as garras foi de 4,3 cm e a velocidade de separação das garras utilizada foi de 8,3 mm/s. Os corpos de prova foram deixados em estufa a 85°C durante à noite e mantidos em dessecador até o momento do teste.

A resistência à tração foi calculada dividindo-se a carga máxima pela área da seção transversal original. Conforme o mesmo método, o módulo de elasticidade foi calculado desenhando-se a tangente à porção linear da curva de tensão versus deformação e, a partir da seleção de qualquer ponto da tangente, foi dividida a tensão de tração pela deformação correspondente. A tensão de tração foi calculada, por sua vez, dividindo-se a carga pela área da seção transversal

3.2 MODELAGEM DOS DADOS

3.2.1 Linguagem e interface de programação

Toda a análise dos dados, modelagem e avaliação do desempenho dos modelos foi realizada utilizando a linguagem de programação Python e a plataforma Jupyter Notebook. Optou-se por utilizar estas ferramentas pois são facilmente acessíveis a quaisquer outros usuários, uma vez que são *open-source*, além de serem frequentemente implementadas na área de ciência de dados e aprendizagem de máquina. As bibliotecas, funções e classes utilizadas estão organizadas no QUADRO 2 abaixo.

QUADRO 2 - RELAÇÃO DE BIBLIOTECAS E RESPECTIVAS FUNÇÕES E/OU CLASSES UTILIZADAS PARA APLICAÇÃO DAS TÉCNICAS DE APRENDIZAGEM DE MÁQUINA.

Biblioteca	Funções/Classes	Objetivo de uso
Pandas	Read_excel; Describe; Info; Corr (correlation)	Manipulação, visualização e tratamento de dados.
Numpy	linspace	Processamento de matrizes; Funções matemáticas.
Matplotlib.pyplot	Subplots Savefig Xticks; Ylabel; Legend; Errorbar;	Criação de gráficos e visualização de dados.
Seaborn	Set_style Pairplot; Barplot; Scatterplot;	Criação de gráficos e visualização de dados
Pickle	Dump; Load	Salvar modelos de regressão linear e florestas randômicas.
Tensor Flow	keras.Sequential; keras.layers.Dense; keras.optimizers.Adam; tf.keras.losses.MeanAbsoluteError; tf.keras.losses.MeanSquareError; tf.keras.metrics.MeanAbsolutePercentageError; keras.models.load_model;	Criação, treinamento, salvamento e descarregamento de redes neurais artificiais.
sklearn.ensemble	RandomForestRegressor	Criação e treinamento de florestas randômicas.
sklearn.model_selection	train_test_split;	Pré-processamento de dados.
sklearn.metrics	mean_absolute_error; mean_absolute_percentage_error; mean_squared_error; r2_score	Cálculo de métricas de desempenho de modelos de aprendizagem de máquina.
SkOpt	BayesSearchCV; Real; Categorical; Integer	Função de otimização bayesiana de hiperparâmetros

Fonte: A autora (2023)

3.2.2 Avaliação preliminar e pré-processamento dos dados

Antes de realizar a modelagem propriamente dita, os dados brutos foram avaliados e pré-processados. A avaliação preliminar foi realizada através da plotagem de gráficos de dispersão, do cálculo dos coeficientes absolutos de Pearson, e da descrição dos dados em relação à tendência central, dispersão e forma da distribuição do conjunto de dados. Desta forma, para cada variável de entrada, foi descrito a

contagem de pontos com valores não nulos, o valor médio, o desvio padrão, o mínimo valor observado, os percentis de 25, 50 e 75% e o valor máximo observado.

O pré-processamento envolveu o tratamento de variáveis categóricas, a retirada das linhas que continham valores nulos para as variáveis de entrada, a separação do conjunto de dados entre treino e teste (para os modelos de floresta randômica) ou treino, validação e teste (para as redes neurais). A divisão entre dados para treino e validação/teste seguiu a proporção 70:30. A exceção foi para os modelos preliminares de predição de propriedades da nanocelulose, que por terem sido gerados com somente 20 pontos disponíveis à época, foi aplicada uma proporção 80:20 para maximizar o número de dados de treino. Vale ressaltar que as duas proporções são comumente utilizadas no aprendizado de máquina (PHROMPHITHAK; ONSREE; TIPPAYAWONG, 2021; REVATHY *et al.*, 2022).

Para o algoritmo de rede neural, onde a magnitude dos dados influencia no ajuste dos pesos, os dados das variáveis de entrada foram normalizados entre 0 e 1. As variáveis de saída foram mantidas com seus valores originais.

3.2.3 Seleção dos algoritmos e variáveis de entrada e saída

Os algoritmos de aprendizagem de máquina aplicados nesta investigação foram as redes neurais artificiais e as florestas randômicas. Apesar da grande variedade de algoritmos existentes na área, estes foram escolhidos pois são robustos, além de funcionarem a partir de princípios diferentes, trazendo uma variedade de aplicações para o presente estudo. Além disso, já há relatos de suas aplicações na área de nanocelulose e separação de gases por membranas.

As variáveis de saída que serão preditas e estão relacionadas à nanocelulose, material de produção das membranas, são a razão de aspecto de nanofibras de celulose e o rendimento de nanofibrilação. Já para a modelagem da separação de gases por membranas serão preditas a permeância do CO₂ e a seletividade CO₂/CH₄.

As variáveis de entrada consideradas inicialmente para predizer a razão de aspecto e o rendimento de nanofibrilação foram porcentagem mássica de celulose, hemicelulose, cinzas, extraíveis e lignina total, dosagem de enzima no pré-tratamento (g/ton), consumo de energia na etapa de fibrilação (kWh/kg), e propriedades finais da nanocelulose, como transmitância a 600 nm, demanda catiônica e os parâmetros reológicos índice de consistência (k) e índice de fluxo (n).

As variáveis de entrada consideradas para prever a permeação e a seletividade foram a razão de aspecto da nanocelulose, presença de CNF, a concentração de nanopartícula na membrana, o pH da solução mãe de polímero, a pressão e a umidade relativa dos gases de entrada.

A redução do número de variáveis para geração dos modelos finais levou em consideração alguns critérios e fez uso de algumas ferramentas e funções, como a função das florestas randômicas que calcula a importância das variáveis de entrada. Desta forma, considerou-se um número de variáveis de entrada de até 5, com o objetivo de tornar os modelos mais simples e baratos – do ponto de vista de coleção de informações-, ao mesmo tempo que mantendo um conjunto de variáveis finais escolhidas que tivessem, na sua soma, uma importância relativa alta.

3.2.4 Ajuste de hiperparâmetros

3.2.4.1 Modelos preliminares de predição das propriedades de nanocelulose

Inicialmente, foram gerados modelos preliminares para predição das propriedades da nanocelulose utilizando-se 20 dados experimentais. Estes 20 dados experimentais foram os relacionados às CNFs obtidas das polpas BKSP, UKSP, BTMP e TMP que foram submetidas ao pré-tratamento mecânico.

Os hiperparâmetros, neste caso, foram ajustados pela técnica manual. A procura manual foi empregada uma vez que estes modelos eram mais simples, foram produzidos com muito menos dados do que os modelos completos e foram ajustados somente alguns poucos hiperparâmetros de cada algoritmo. Portanto, uma procura manual seria suficiente para os modelos preliminares e não foi necessário empregar um algoritmo de otimização mais sofisticado para encontrar modelos com bons resultados. Os hiperparâmetros avaliados para o algoritmo de floresta randômica foram o número de árvores e o número mínimo de pontos experimentais a permanecerem em nós do tipo folha. O restante dos hiperparâmetros dos algoritmos das bibliotecas Scikit Learn foi mantido nos valores padrões. Ajustaram-se estes dois hiperparâmetros somente pois o número de preditores é um dos parâmetros mais importantes para regular a acurácia do modelo e o número mínimo de amostras presentes nas folhas condiciona o crescimento de cada árvore individualmente, gerando modelos mais suaves e controlando possíveis ocorrências de *overfitting*

(BREIMAN, 2001; GÉRON, 2019; PROBST; BOULESTEIX, 2018). O número avaliado de árvores foi de 1 a 200 e o número mínimo de amostras de 1 a 5. Desta maneira foram avaliadas algumas combinações de valores destes dois hiperparâmetros, para um modelo com todas as 11 variáveis de entrada disponíveis para o este conjunto de 20 pontos, e os respectivos resultados de predição comparados. O modelo mais simples e que fornecesse as melhores métricas foi selecionado. Na TABELA 7 encontram-se os valores, tanto o padrão quanto a faixa estudada, dos hiperparâmetros do algoritmo.

TABELA 7 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE FLORESTA RANDÔMICA PARA OS MODELOS PRELIMINARES DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS

Hiperparâmetro	Valor padrão	Valor/faixa estudada
<i>n_estimators</i>	Não	1-200
<i>Min_samples_leaf</i>	Não	1-5
<i>Criterion</i>	<i>Sim</i>	<i>'squared_error'</i>
<i>Max_depth</i>	Sim	<i>'None'</i> (nós são expandidos até que todas as folhas contenham menos que o número mínimo de amostras requeridas para um <i>split</i> ou quando todas as folhas estão puras)
<i>Min_samples_split</i>	Sim	2
<i>Min_weight_fraction_leaf</i>	Sim	0,0
<i>Max_features</i>	Sim	100%
<i>Max_leaf_nodes</i>	Sim	<i>None</i> (número ilimitado de nós do tipo folha)
<i>Min_impurity_decrease</i>	Sim	0
<i>Bootstrap</i>	Sim	<i>True</i>
<i>Oob_Score</i>	Sim	<i>False</i>
<i>Verbose</i>	Sim	0
<i>Warm_start</i>	Sim	<i>False</i>
<i>Ccp_alpha</i>	Sim	0,0
<i>Max_samples</i>	Sim	<i>None</i>

Fonte: A autora (2023)

Para as redes neurais artificiais preliminares, foram fixados diversos hiperparâmetros e avaliados diferentes valores para somente três. Desta forma, foi

utilizada somente uma camada escondida, uma vez que não foi usado um conjunto de dados extenso, fixou-se o valor da taxa de aprendizagem como 0,01, um valor tradicionalmente aplicado nos treinos de redes neurais e que tende a gerar bons resultados (BENGIO, 2012; SHIN *et al.*, 2020; XIE *et al.*, 2021), utilizou-se o algoritmo de otimização Adam, uma vez que é um dos algoritmos mais eficientes computacionalmente (KINGMA; BA, 2015), fixou-se um número máximo de épocas em 700, valor alto o suficiente para se obter convergência neste caso, e estabeleceu-se a função de ativação linear para camada de saída, pois é a indicada para problemas de regressão já que não modifica o valor final fornecido pela última camada escondida (JIANG; JIANG; CHEN, 2022).

Os hiperparâmetros que foram modificados incluíram o tamanho do lote com valores entre 2 e 16 (número total de pontos experimentais de treino), o número de neurônios na camada oculta e as funções de ativação desta camada (escolhidas entre sigmoideal e ReLU). Ressalta-se que o número de neurônios na camada escondida foi alterado entre o mesmo número de neurônios na camada de entrada e o dobro desta quantidade. O número de neurônios na primeira camada da rede mudava de acordo com o subconjunto de variáveis de entrada testados. A TABELA 8 apresenta um resumo dos valores dos hiperparâmetros usados para a rede neural. Da mesma forma que realizado para as florestas randômicas, a procura manual foi aplicada para os modelos de rede neural. O modelo mais simples e que fornecesse as melhores métricas era selecionado.

TABELA 8 – HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE REDE NEURAL PARA OS MODELOS PRELIMINARES DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.

Hiperparâmetro	Valor fixo	Valor/faixa estudada
N° de neurônios	Não	Mesmo número até o dobro de variáveis de entrada
Tamanho do lote	Não	2-16
Função de ativação intermediária	Não	Sigmoidal e ReLU
Taxa de aprendizagem	Sim	0,01
N° de camadas ocultas	Sim	1
Função de ativação da última camada	Sim	Linear
Função de otimização interna	Sim	Adam
N° máximo de épocas	Sim	700

Fonte: A autora (2023)

3.2.4.2 Modelos completos de predição das propriedades da nanocelulose e modelos de predição das propriedades de membrana

Para os modelos completos de predição de propriedades da nanocelulose e os modelos de predição de permeança e seletividade de membranas foi utilizada a otimização bayesiana. Para isso, foi aplicada a função BayesSearchCV da biblioteca Scikit Opt. Foi escolhida esta técnica uma vez que é um método mais eficiente em comparação com outras para quando se analisa uma quantidade maior de hiperparâmetros em uma faixa de valores possíveis maior. A função objetivo a ser maximizada quando o BayesSearchCV foi aplicado dependeu da função *score* padrão do algoritmo, a não ser especificado de outra maneira. Assim, para as florestas randômicas, a função foi o coeficiente de determinação R^2 . Para as redes neurais, a função a ser maximizada foi o negativo do erro absoluto médio. O número de iterações aplicado foi de 200 no máximo.

Para os modelos completos de predição de propriedades da nanocelulose e de desempenho das membranas utilizando a floresta randômica foram avaliados os hiperparâmetros *max_depth* (profundidade máxima das árvores), *max_features* (máximo de variáveis de entrada considerada para fazer a divisão dos nós), *min_samples_leaf* (número mínimo de amostras nos nós tipo folha) e *n_estimators* (número de árvores). Estes foram os hiperparâmetros escolhidos pois influenciam

diretamente na acurácia de predição, na eficiência computacional e na ocorrência de *over/underfitting* (ALSHRAIDEH; CASTILLO; GIL DEL VAL, 2020; PHROMPHITHAK; ONSREE; TIPPAYAWONG, 2021). O restante dos hiperparâmetros do algoritmo foi mantido com os valores padrão já presentes na função da biblioteca Scikit Learn. A faixa de valores estudada para cada hiperparâmetro está resumida na TABELA 9.

TABELA 9 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE FLORESTA RANDÔMICA PARA OS MODELOS COMPLETOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.

Hiperparâmetro	Valor padrão	Valor/faixa estudada
<i>n_estimators</i>	Não	1-400
<i>Min_samples_leaf</i>	Não	1-4
<i>Max_depth</i>	Não	1-100
<i>Max_features</i>	Não	1 – n° total de variáveis
<i>Criterion</i>	Sim	'squared_error'
<i>Min_samples_split</i>	Sim	2
<i>Min_weight_fraction_leaf</i>	Sim	0,0
<i>Max_leaf_nodes</i>	Sim	None (número ilimitado de nós do tipo folha)
<i>Min_impurity_decrease</i>	Sim	0
<i>Bootstrap</i>	Sim	True
<i>Oob_Score</i>	Sim	False
<i>Verbose</i>	Sim	0
<i>Warm_start</i>	Sim	False
<i>Ccp_alpha</i>	Sim	0,0
<i>Max_samples</i>	Sim	None

Fonte: A autora (2023)

Para as redes neurais artificiais, foram avaliados diferentes valores para quatro hiperparâmetros. Similarmente aos modelos preliminares, também foi utilizada na rede somente uma camada escondida, utilizou-se o algoritmo de otimização Adam e estabeleceu-se a função de ativação linear para camada de saída. O número de épocas máximo foi configurado para 2000 desta vez, uma vez que foi aplicada a função *early_stopping* para parar o treino da rede quando a métrica da função de perda do subconjunto de validação não melhorasse após 5 épocas seguidas.

Os hiperparâmetros que foram modificados incluíram o tamanho do lote com valores entre 2 e 32 (GÉRON, 2019), o número de neurônios na camada oculta (de 1

ao dobro de neurônio de entrada), as funções de ativação desta camada (escolhidas entre sigmoideal e ReLU) e o valor da taxa de aprendizagem (de 0,0001 até 0,1). A TABELA 10 apresenta um resumo dos valores dos hiperparâmetros usados para a rede neural.

TABELA 10 - HIPERPARÂMETROS E RESPECTIVA FAIXA DE VALORES ESTUDADOS PARA O ALGORITMO DE REDE NEURAL PARA OS MODELOS COMPLETOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE E MEMBRANAS.

Hiperparâmetro	Valor fixo	Valor/faixa estudada
N° de neurônios	Não	Mesmo número até o dobro de variáveis de entrada
Tamanho do lote	Não	2-32
Função de ativação intermediária	Não	Sigmoideal e ReLU
Taxa de aprendizagem	Não	0,0001-0,1
N° de camadas ocultas	Sim	1
Função de ativação da última camada	Sim	Linear
Função de otimização interna	Sim	Adam
N° máximo de épocas	Sim	2000

Fonte: A autora (2023)

Para fins de análise dos modelos, foram calculadas as forças relativas das variáveis de entrada para as redes neurais geradas utilizando a fórmula da EQUAÇÃO (11) (YOON; SWALES; SILVIA, 1993). Nesta equação, RS_{ji} representa a força relativa que a entrada “i” exerce sobre a saída “j”. W_{ki} representa o peso entre o “k” neurônio da camada oculta e a entrada “i” e U_{jk} representa o peso entre o neurônio “k” da camada oculta e neurônio de saída “j”. Desta maneira, no numerador é feita uma soma ponderada levando em conta os pesos associados aos “n” neurônios da camada oculta que tenham relação com o neurônio de entrada “i”. No denominador, é feita a soma ponderada absoluta dos pesos envolvidos com todos os “m” neurônios de entrada.

$$RS_{ji} = \frac{\sum_{k=0}^n (W_{ki} * U_{jk})}{\sum_{i=0}^m ABS(\sum_{k=0}^n (W_{ki} * U_{jk}))} \quad (11)$$

3.3 AVALIAÇÃO DOS MODELOS

Os modelos foram avaliados por quatro métricas estatísticas principais: coeficiente de determinação R^2 , raiz quadrada do erro médio (RQEM), erro absoluto médio (EAM) e erro percentual absoluto médio (EPAM).

EAM, EPAM, RQEM e R^2 são calculados pelas EQUAÇÕES (12), (13), (14) e (15), onde n representa o número de pontos experimentais, y_i o valor real, \hat{y}_i a predição da i -ésima amostra e \bar{y} representa o valor médio de todas as observações, este último calculado pela EQUAÇÃO (16).

$$\text{EAM} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (12)$$

$$\text{EPAM} = \frac{100\%}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (13)$$

$$\text{RQEM} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (14)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad (16)$$

4 RESULTADOS E DISCUSSÃO

4.1 AVALIAÇÃO DOS DADOS DE POLPAS LIGNOCELULÓSICAS E DE NANOFIBRAS DE CELULOSE PARA PREDIÇÃO DE PROPRIEDADES DE MATERIAL

Da tabela de caracterização da polpa, apresentada no APÊNDICE A, observa-se a grande diversidade de matérias-primas que foram utilizadas na etapa de predição de propriedades da nanocelulose. Pinus e abeto são exemplos de madeiras do tipo *softwood* (coníferas, com madeiras de baixa densidade), álamo e eucalipto são madeiras do tipo *hardwood* (madeiras mais densas e que possuem paredes celulares mais grossas), cânhamo, juta e sisal são plantas não lenhosas e a polpa reciclada é uma mistura de polpa kraft e papel ondulado (ASIF, 2009; SYDUZZAMAN *et al.*, 2020).

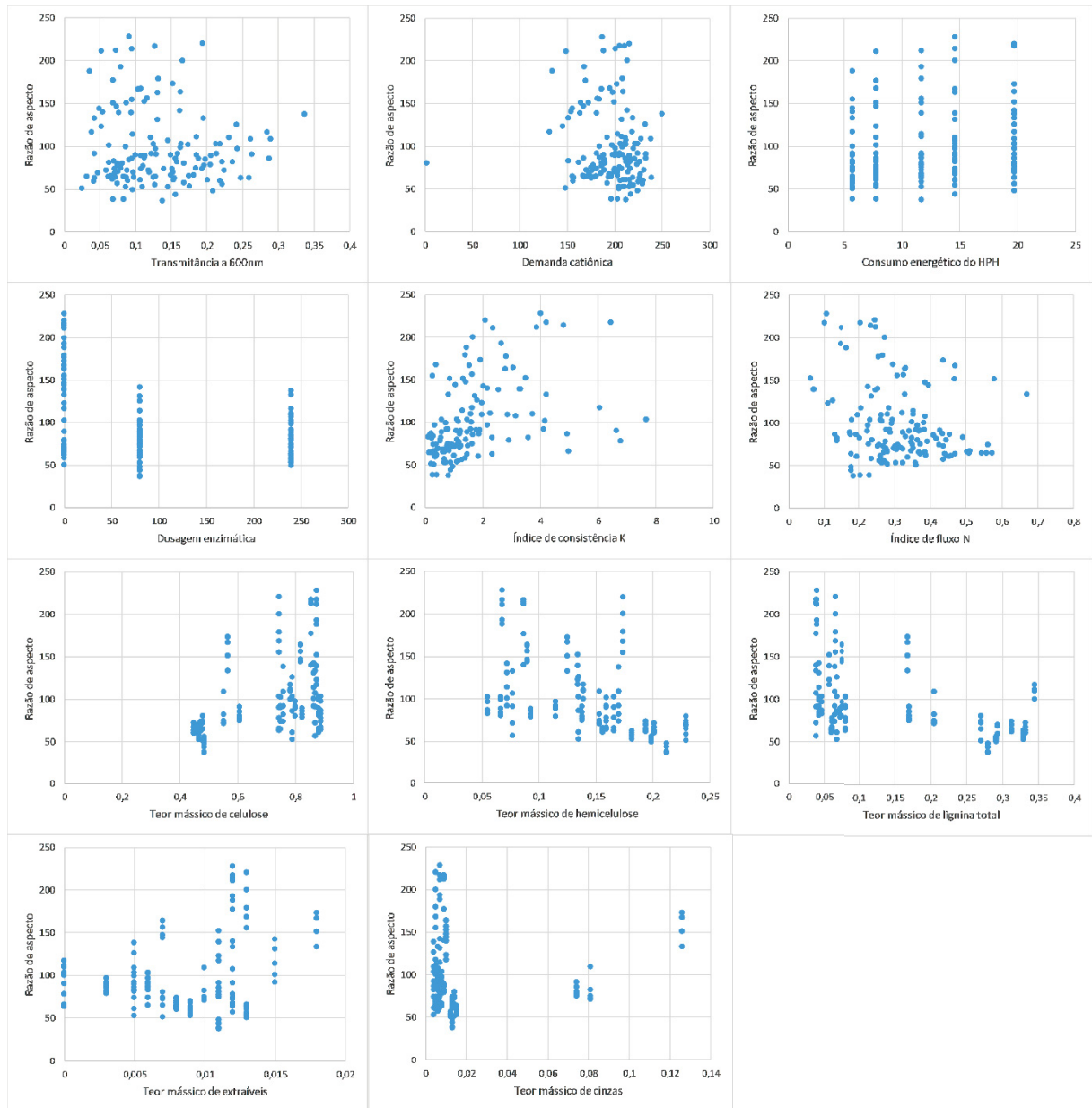
Esta diversidade é refletida na composição química. Por exemplo, verifica-se que as plantas não lenhosas possuem um teor de celulose maior do que as polpas de pinus sem branqueamento (TMP e UKSP), a de abeto (BTMP), álamo, eucalipto e de polpa reciclada. Já em relação ao teor de lignina é o contrário: sisal, juta e cânhamo são as polpas que possuem os menores valores para esta substância.

Analisando a tabela das propriedades da nanocelulose, apresentada no APÊNDICE A, onde estão elencados os valores das variáveis de processo e das nanoceluloses obtidas, percebe-se que há uma grande diversidade de razões de aspecto e rendimentos de nanofibrilação entre as 140 CNFs produzidas.

Das 12 variáveis no total, foram consideradas 11 para a geração dos modelos de predição de propriedade das nanoceluloses, pois excluiu-se o índice de cristalinidade, que havia sido medido somente para as polpas pré-tratadas mecanicamente. Não obstante, as variáveis consideradas abrangeram desde a composição química da matéria-prima pré-tratada, variáveis de processo como dosagem de enzima e intensidade de desfibrilação, e propriedades finais da nanocelulose, como parâmetros ópticos e reológicos.

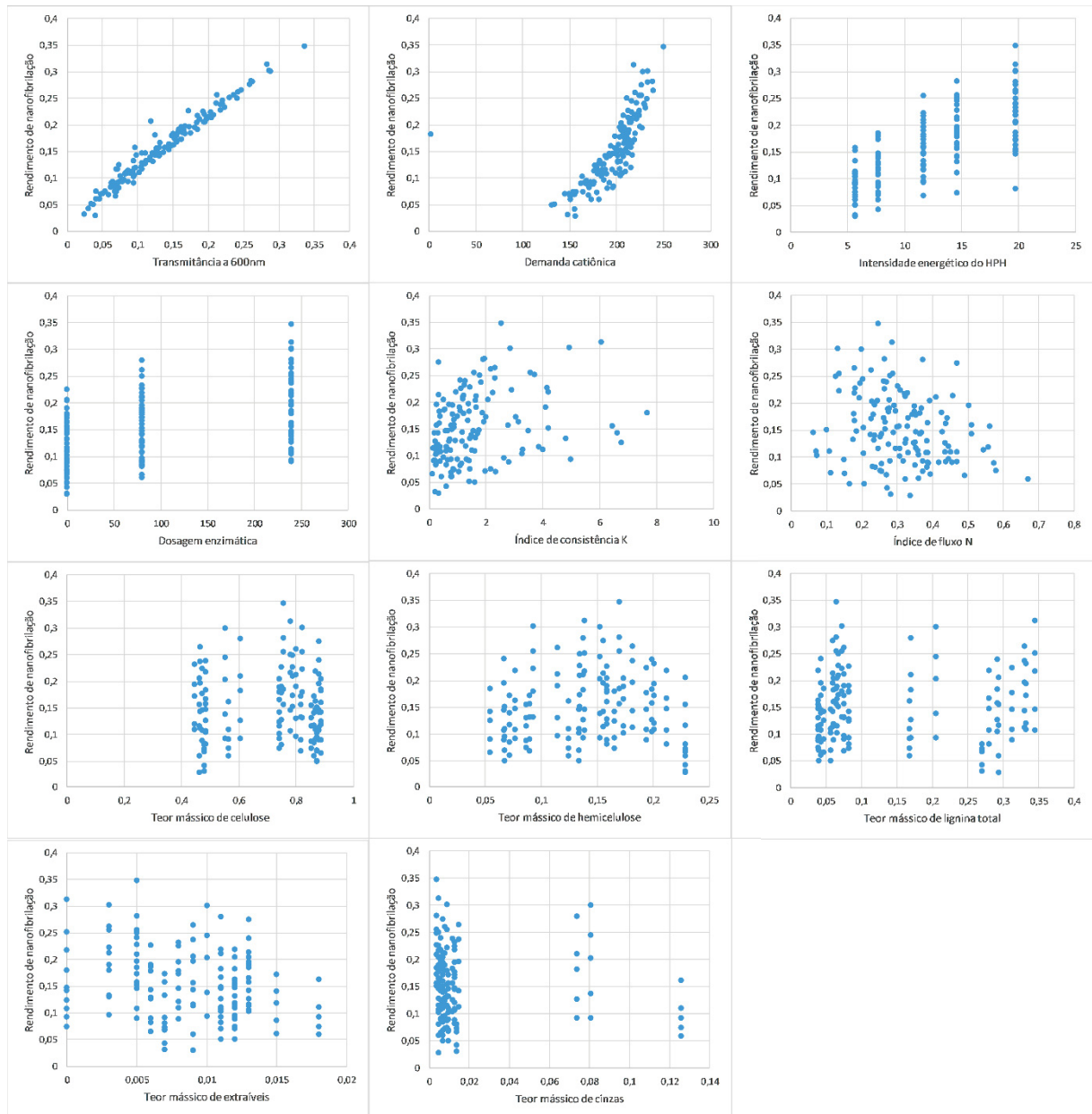
Nas FIGURA 6 e FIGURA 7, são exibidos os gráficos de dispersão das variáveis de entrada em função da razão de aspecto e rendimento de nanofibrilação, respectivamente.

FIGURA 6 - GRÁFICOS DE DISPERSÃO PARA A RAZÃO DE ASPECTO DE CNFS



Fonte: A autora (2023)

FIGURA 7 - GRÁFICOS DE DISPERSÃO PARA O RENDIMENTO DE NANOFIBRILAÇÃO DE CNFS



Fonte: A autora (2023)

Observando a FIGURA 6, percebe-se que não há claras tendências de correlação individualmente entre as variáveis de entrada e a razão de aspecto. Verifica-se, no gráfico da demanda catiônica versus razão de aspecto, a ocorrência de um valor discrepante, para o caso onde a demanda catiônica reportada foi de 2. Os únicos gráficos onde é possível visualizar algum tipo de tendência seriam para o de dosagem enzimática e composição química da matéria-prima pré-tratada, especialmente para os teores de celulose, hemicelulose e lignina. Para todos esses gráficos, há uma observação de uma ampla faixa de valores de razão de aspecto,

contudo, é possível perceber uma tendência de decréscimo da razão de aspecto com o aumento da dosagem enzimática, o aumento do teor de lignina total e o de hemicelulose, enquanto que percebe-se o oposto para o teor de celulose e extraíveis.

Observando a FIGURA 7, observa-se uma clara tendência linear entre a transmitância a 600 nm e o rendimento de nanofibrilação. É possível perceber uma tendência também de aumento entre a demanda catiônica e o rendimento, apesar de não ser linear. Com relação ao consumo energético, observa-se que existe uma ampla faixa de rendimentos reportados para cada nível de intensidade de energia aplicada. Contudo, percebe-se que ocorre um aumento do rendimento observado conforme se aumenta o consumo energético. Por fim, é possível observar uma tendência entre o aumento da dosagem enzimática e um aumento da faixa de valores do rendimento de nanofibrilação, apesar de ser mais sutil. Com relação às outras variáveis, não existem claras tendências que sejam significativas relatar.

Na TABELA 11 está a descrição estatística das variáveis de entrada e das variáveis de saída relacionados a esta etapa da modelagem.

TABELA 11 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E VARIÁVEIS DE SAÍDA PARA MODELOS DE PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE FIBRILADA.

Variável	Valores não nulos	Valor médio	Desvio padrão	Valor mínimo	25%	50%	75%	Valor máximo
Dosagem enzimática (g/ton)	140	102,86	100,32	0	0	80	240	240
Demanda catiônica (µeq/g)	140	195,76	28,86	2	181	202,5	214	250
Transmitância a 600 nm (%)	140	0,132	0,064	0,025	0,080	0,123	0,173	0,337
Consumo energético (kWh/kg)	140	11,87	5,01	5,67	7,7	11,67	14,59	19,72
Índice de consistência (k)	140	1,6345	1,4999	0,1308	0,6226	1,1943	1,9876	7,6787
Índice de fluxo (n)	140	0,312	0,115	0,062	0,237	0,303	0,383	0,670
Celulose (%p/p)	140	0,709	0,164	0,446	0,534	0,768	0,864	0,887
Hemicelulose (%p/p)	140	0,142	0,051	0,055	0,092	0,146	0,176	0,229
Lignina total (%p/p)	140	0,143	0,112	0,039	0,059	0,076	0,272	0,345
Extraíveis (%p/p)	140	0,0087	0,0043	0,0000	0,0057	0,0090	0,0120	0,0180
Cinzas (%p/p)	140	0,018	0,028	0,004	0,006	0,008	0,013	0,126
Razão de aspecto	140	98,81	44,97	37	65,75	85,5	117	228
Rendimento de nanofibração	140	0,1520	0,0654	0,0290	0,1025	0,1457	0,1957	0,3473

Fonte: A autora (2023)

Foram calculados os valores dos coeficientes de correlação de Pearson entre cada variável individual de entrada e as duas variáveis de saída de interesse. Nas

TABELA 12 e TABELA 13 estão apresentados os valores para razão de aspecto e rendimento de nanofibrilação, respectivamente.

TABELA 12 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL RAZÃO DE ASPECTO.

Variável de entrada	Coefficiente de correlação de Pearson
Hemicelulose (%p/p)	-0,5086
Lignina total (%p/p)	-0,4627
Índice de consistência (k)	0,4474
Celulose (%p/p)	0,4408
Dosagem de enzima (g/ton)	-0,4055
Extraíveis (%p/p)	0,2592
Índice de fluxo (n)	-0,2587
Consumo energético (kWh/kg)	0,2306
Demanda catiônica ($\mu\text{eq/g}$)	-0,1158
Cinzas (%p/p)	0,0902
Transmitância a 600 nm (%)	-0,0680

Fonte: A autora (2023)

TABELA 13 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL RENDIMENTO DE NANOFIBRILAÇÃO.

Variável de entrada	Coefficiente de correlação de Pearson
Transmitância a 600 nm (%)	0,9819
Consumo energético (kWh/kg)	0,7263
Demanda catiônica ($\mu\text{eq/g}$)	0,6925
Dosagem de enzima (g/ton)	0,4713
Índice de consistência (k)	0,2822
Extraíveis (%p/p)	-0,2422
Índice de fluxo (n)	-0,2261
Lignina total (%p/p)	0,0467
Cinzas (%p/p)	-0,0420
Hemicelulose (%p/p)	0,0411
Celulose (%p/p)	0,0165

Fonte: A autora (2023)

Observando os valores apresentados da TABELA 12, verifica-se que não há coeficientes maiores que 0,7 e próximos a 1. Contudo, dos elencados, a composição química da matéria-prima aparece como uma das variáveis com maior valor de

Pearson, em especial o teor de hemicelulose, lignina e celulose também. Os valores apresentados foram, respectivamente, -0,5086, -0,4627 e 0,4408. O índice de consistência e a dosagem enzimática também aparecem nas 5 variáveis com maior coeficiente, com os valores de 0,4474 e -0,4055. Ademais, observa-se que o teor de lignina, hemicelulose e dosagem de enzima possuem uma correlação negativa com a razão de aspecto. Estas particularidades estão em acordo com o que foi observado e descrito na avaliação dos gráficos de dispersão da FIGURA 6.

Já para os valores apresentados na TABELA 13, comprova-se que há uma correlação linear muito forte entre a transmitância a 600 nm e o rendimento de nanofibrilação, o qual apresenta um coeficiente de correlação de 0,9819. Em segundo lugar, tem-se o consumo energético, que apresenta um coeficiente de Pearson de 0,7263. A demanda catiônica apresentou um valor de 0,6925, muito próximo a 0,7. Já a dosagem enzimática não apresentou um coeficiente que denotasse uma correlação linear forte com o rendimento de nanofibrilação. Estas conclusões estão em consonância com o que é observado nos gráficos de dispersão da FIGURA 7.

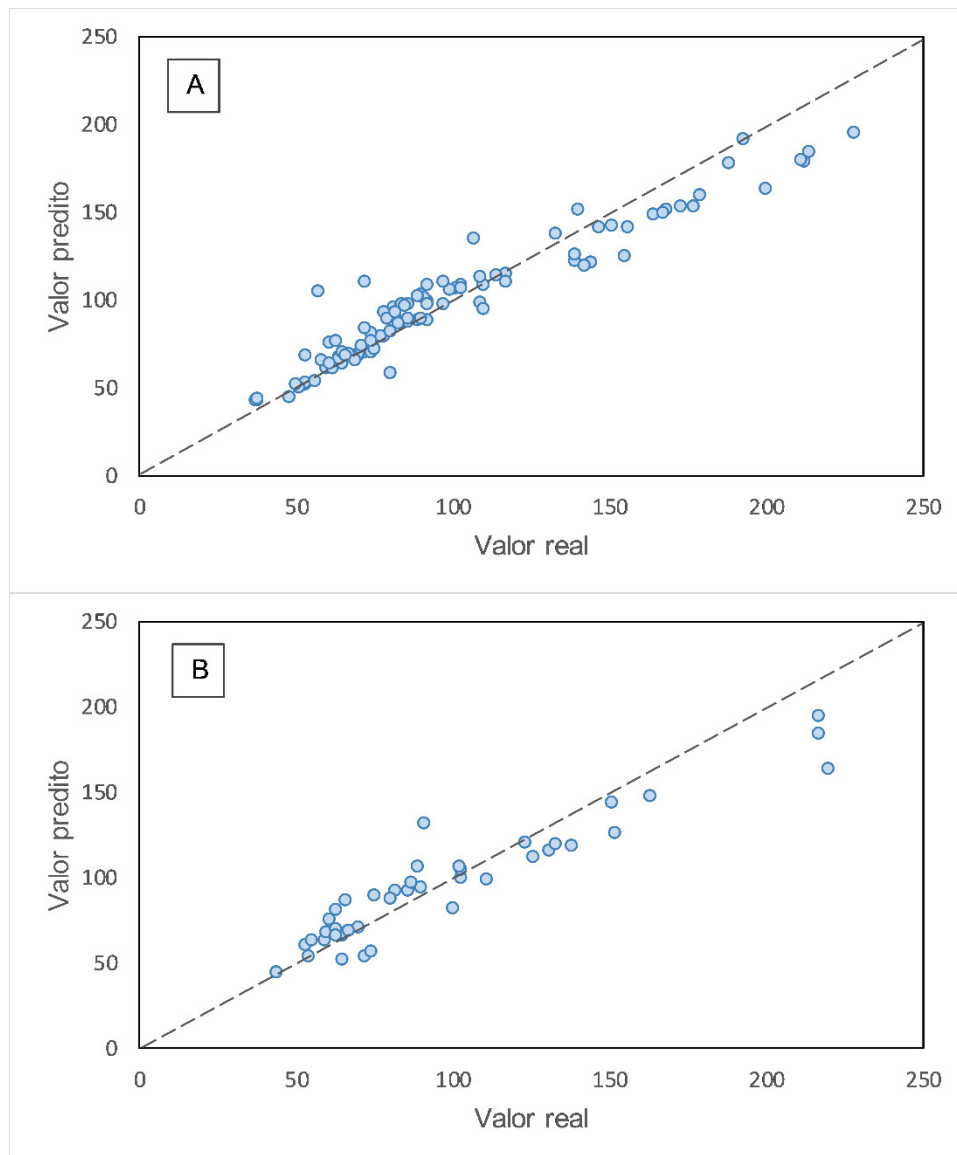
4.2 MODELOS DE PREDIÇÃO: RAZÃO DE ASPECTO DE NANOFIBRAS DE CELULOSE

4.2.1 Florestas randômicas

Para as florestas randômicas, foram gerados três modelos: um modelo utilizando a mesma arquitetura do modelo preliminar (modelo construído inicialmente com 20 dados experimentais de nanocelulose mecânica e que está descrito no APÊNDICE C), um modelo com todas as variáveis de entrada (ou seja, incluindo a dosagem enzimática, que é nova em relação ao modelo preliminar) chamado de intermediário, e um modelo final com as cinco variáveis mais importantes, de acordo com o resultado apresentado pela função de importância de variáveis embutida na biblioteca do Scikit Learn.

O modelo preliminar foi construído com 17 árvores e um número mínimo de 2 amostras por folha. As variáveis de entrada foram teor mássico de lignina total e celulose e consumo energético. O gráfico dos valores reais e valores preditos dos subconjuntos de treino e teste estão apresentados no GRÁFICO 1, enquanto as métricas estão elencadas na TABELA 14.

GRÁFICO 1 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA UTILIZANDO OS HIPERPARÂMETROS DO MODELO PRELIMINAR PARA PREDIÇÃO DA RAZÃO DE ASPECTO.



Fonte: A autora (2023)

TABELA 14 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DOS MODELOS DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA RAZÃO DE ASPECTO DE NANOCELULOSES FIBRILADAS.

Modelo	Subconjunto	EAM	EPAM	RQEM	R ²
Preliminar	Treino	9,95	9,92%	14,10	0,901
	Teste	12,35	12,47%	16,65	0,862
Intermediário	Treino	4,43	4,70%	5,63	0,984
	Teste	8,21	9,06%	11,40	0,935
Final	Treino	3,87	4,12 %	5,24	0,986
	Teste	9,05	9,94%	12,74	0,919

Fonte: A autora (2023)

O gráfico A do GRÁFICO 1 mostra os pontos relativamente próximos da reta de 45°. Quantitativamente, o R² obtido para o treino foi de 0,901. O erro percentual médio absoluto foi um pouco abaixo de 10%, precisamente 9,92%, enquanto o erro absoluto médio e a raiz do erro quadrado médio foi de 9,948 e 14,10, respectivamente. Percebe-se que a arquitetura preliminar foi um tanto quanto adequada para os novos dados, especialmente se considerado que nesta modelagem existem 85 pontos experimentais relacionados a nanoceluloses que provieram de um pré-tratamento enzimático, ou seja, um pré-tratamento que não foi contemplado no treino dos modelos preliminares. Como observado nos gráficos de dispersão da FIGURA 6, existem alguns indícios de correlação entre razão de aspecto e componentes químicos da matéria-prima, especialmente lignina, hemicelulose e celulose. Por conta da manutenção de algumas dessas variáveis tanto no conjunto de dados completo, quando no preliminar, manteve-se um grau de correlação satisfatório para os novos dados. Quando testado o modelo com pontos não vistos previamente pelo modelo, alcançou-se um R² um pouco abaixo de 0,90, no valor de 0,862 e erros próximos das métricas de treino. Observando o gráfico B do GRÁFICO 1 verificou-se que, para valores maiores de razão de aspecto, o desvio entre valor predito e valor real é maior.

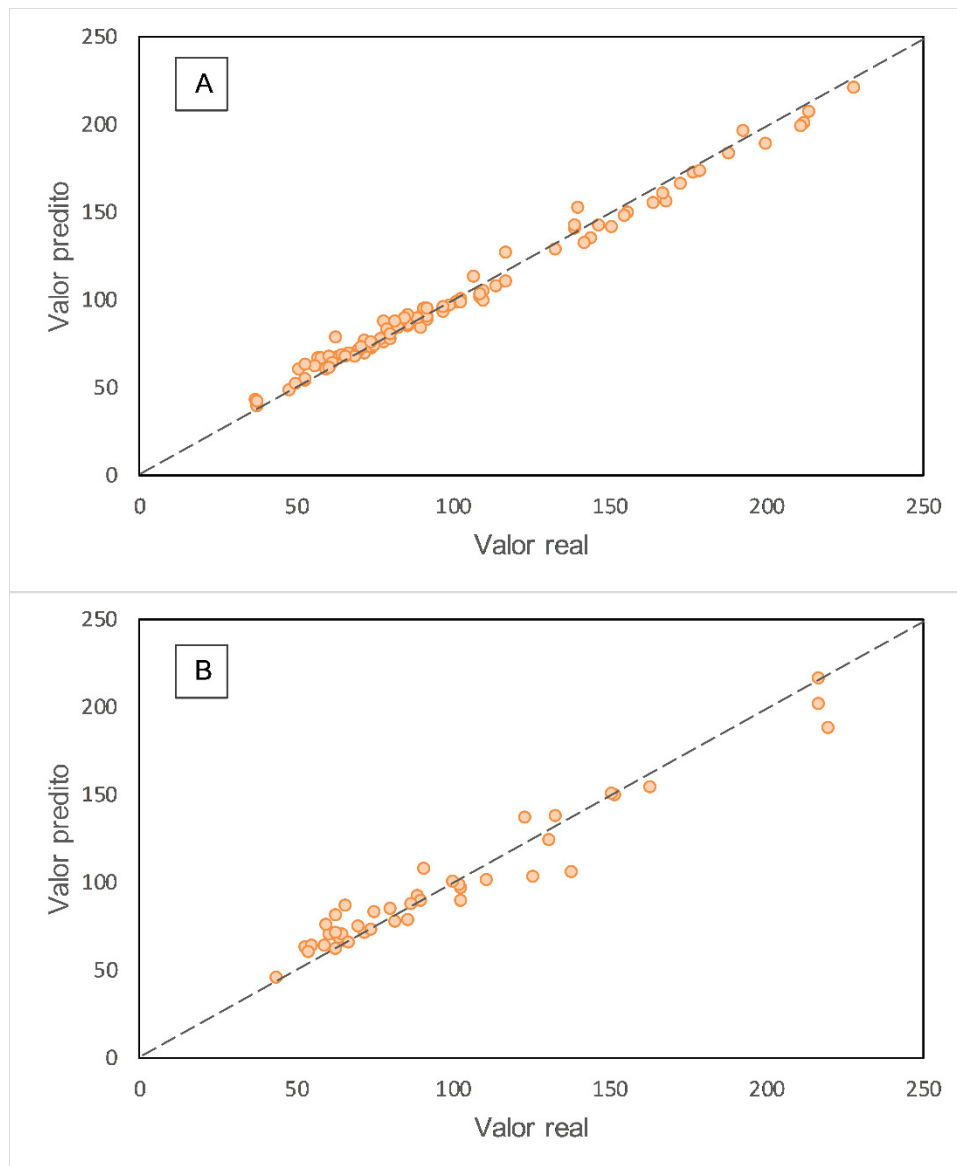
Desta maneira, constata-se que um modelo simples, com três variáveis de entrada e que contou somente com 20 pontos experimentais para chegar nesta arquitetura preliminar, teve um bom desempenho para predizer a razão de aspecto de nanoceluloses fibriladas advindas de matérias-primas e pré-tratamento que não integraram o conjunto de dados preliminar.

Um novo modelo de floresta randômica foi treinado com todos os dados novos e todas as variáveis de entrada, a fim de explorar outros hiperparâmetros do algoritmo

e em uma faixa de valores mais extensa. Este modelo foi chamado de modelo intermediário. Além disso, desejou-se entender se a importância das variáveis de entrada mudou frente a um novo e maior conjunto de dados, especialmente com a inserção de uma nova, a dosagem enzimática.

Aplicando a função de otimização bayesiana, a arquitetura que resultou nas melhores métricas de treino foi: profundidade máxima das árvores = 8, quantidade máxima de features para fazer a divisão dos nós = 8, número de árvores = 100 e número mínimo de pontos nas folhas = 1. Os gráficos dos valores preditos versus reais para o treino e teste estão apresentados no GRÁFICO 2. As métricas de desempenho para os subconjuntos mencionados estão apresentadas na TABELA 14.

GRÁFICO 2 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA INTERMEDIÁRIO PARA PREDIÇÃO DA RAZÃO DE ASPECTO.



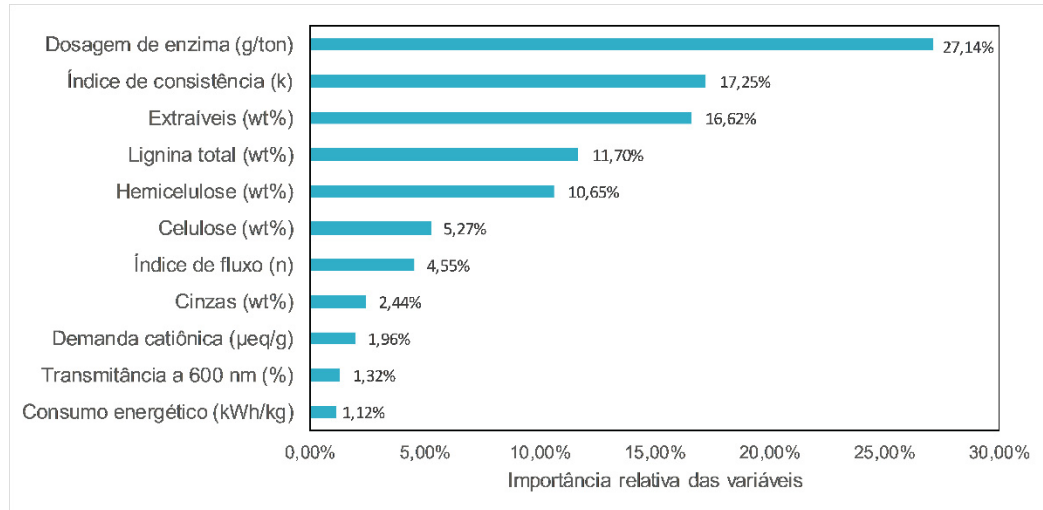
Fonte: A autora (2023)

A partir da análise do gráfico A do GRÁFICO 2, percebe-se que o treinamento dos novos dados utilizando a função de otimização bayesiana melhorou consideravelmente o desempenho preditivo com florestas randômicas. Percebe-se que os pontos experimentais estão bem mais próximos da reta de 45°. Pode-se concluir isto também a partir da análise dos valores das métricas de desempenho na TABELA 14. O coeficiente de determinação foi de 0,984 e o erro percentual médio absoluto foi menor que 5%, especificamente 4,70%. O erro médio absoluto decresceu para 4,432 e a raiz do erro quadrado médio para 5,63. A predição para os pontos de

teste também melhorou em comparação com o modelo com a arquitetura preliminar. O R^2 subiu para 0,935, e os erros EPAM, EAM e RQEM reduziram para 9,06%, 8,21 e 11,40, respectivamente.

A importância das variáveis de entrada foi calculada e elencada no GRÁFICO 3.

GRÁFICO 3 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA RAZÃO DE ASPECTO UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.



Fonte: A autora (2023)

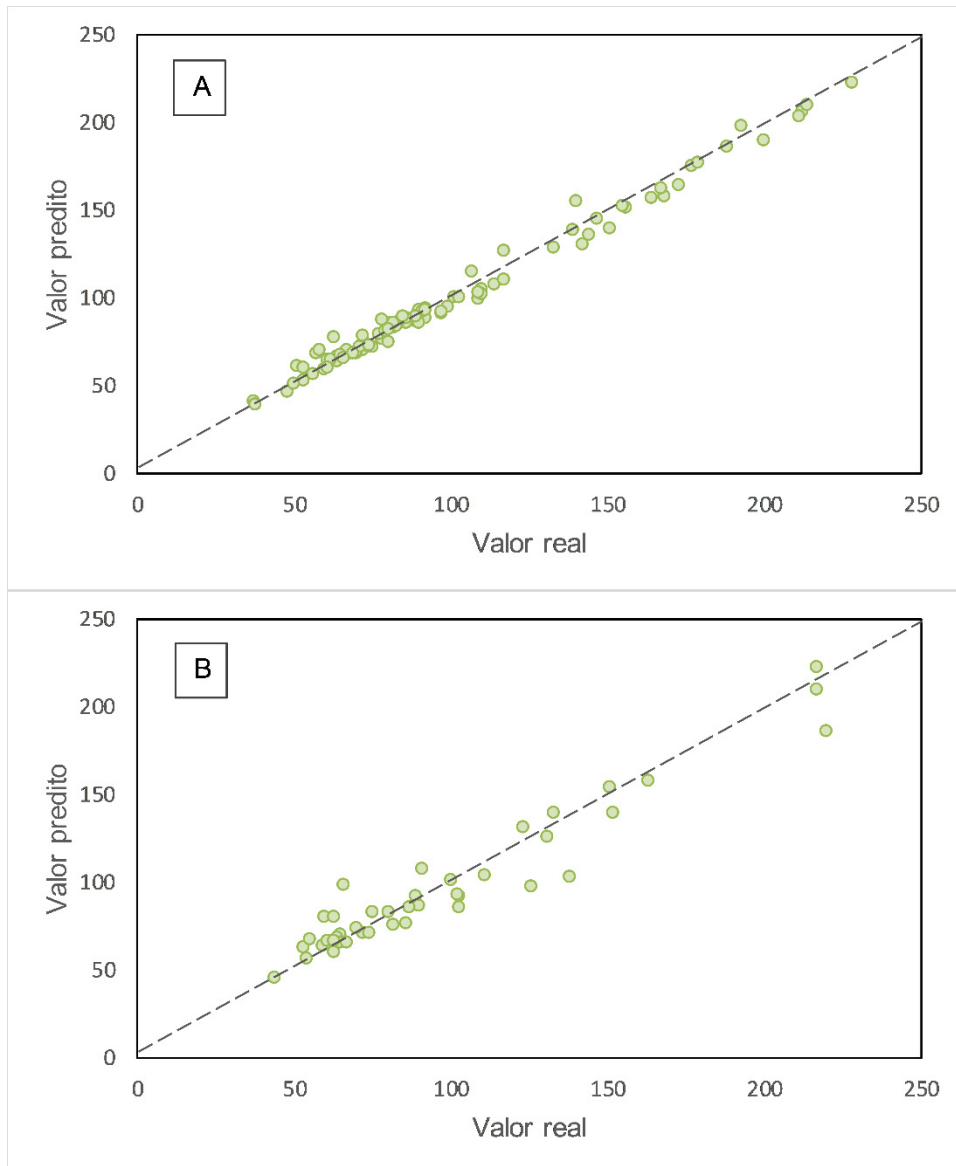
A variável que demonstrou mais importância dentre as estudadas foi a dosagem enzimática. Este parâmetro representou o tipo e nível de pré-tratamento aplicado, diferenciando entre enzimático e mecânico, este último representado pelo valor 0. Em consonância com o que foi descrito na avaliação dos dados, o tipo de pré-tratamento aplicado se mostrou como uma variável que se correlaciona bem à razão de aspecto. Em segundo lugar no nível relativo de importância, está o índice de consistência k , que é um parâmetro reológico. Com uma porcentagem próxima ao índice k , tem-se o teor de extraíveis, seguido do teor de lignina total e de hemicelulose. Desta forma, verifica-se que entre as cinco variáveis mais importantes para a floresta randômica, que juntas contabilizam quase 85% do total, três são relativas à composição da matéria-prima pré-tratada, bem como analisado nos gráficos de dispersão da FIGURA 6 e nos valores dos coeficientes de correlação de Pearson apresentados na TABELA 12.

Essas variáveis elencadas como mais importantes estão coerentes com o que é descrito na literatura acerca das influências na determinação da razão de aspecto

de nanoceluloses fibriladas. Por exemplo, o teor de lignina fornece maior rigidez e menor deformabilidade para as paredes celulares, o que dificulta a delaminação da celulose. Logo, possui um efeito negativo sobre a razão de aspecto, que aumenta conforme a desfibrilação é mais intensa (LI *et al.*, 2016; SERRA-PARAREDA *et al.*, 2022). Os efeitos sobre parâmetros reológicos de suspensões que contém nanocelulose fibrilada também é reportado na literatura. A redução do tamanho das partículas e a aplicação de taxas de cisalhamento sob este tipo de suspensão influenciam na reologia desses fluidos, especialmente porque a interação água e nanofibras tende a aumentar (AGUADO *et al.*, 2022; HUBBE *et al.*, 2017; WANG *et al.*, 2011). Já o tipo de pré-tratamento também possui influência sobre a razão de aspecto das nanoceluloses geradas. Por exemplo, o pré-tratamento mecânico atua principalmente de forma a delaminar as cadeias de celulose com aplicação de intensa energia, podendo haver uma presença maior de microfibras (ANG; HARITOS; BATCHELOR, 2019). Já o pré-tratamento enzimático se vale da hidrólise enzimática das ligações β 1-4 entre monômeros de glucose, reduzindo as cadeias em oligômeros menores (SINGH *et al.*, 2019).

Utilizando as cinco variáveis mais importantes para a floresta randômica, foi criado um último modelo. Também foi aplicada a função de otimização bayesiana para este caso. A arquitetura final obtida foi profundidade máxima das árvores = 38, quantidade máxima de features para fazer a divisão dos nós = 5, número de árvores = 51 e número mínimo de pontos nas folhas = 1. Os gráficos de valores preditos versus reais e as métricas de desempenho para os subconjuntos de treino e testes podem ser visualizados no GRÁFICO 4 e na TABELA 14.

GRÁFICO 4 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DA RAZÃO DE ASPECTO.



Fonte: A autora (2023)

Comparando os GRÁFICO 2 e GRÁFICO 4, percebe-se que a redução do total de variáveis para as cinco mais importantes não modificou consideravelmente o desempenho preditivo do modelo. Isto também é corroborado ao se analisarem as métricas de predição. Para o subconjunto de treino, o R^2 obtido foi de 0,986, enquanto as métricas de EAM, EPAM e RQEM foram 3,87, 4,12% e 5,24, respectivamente. Estas métricas foram levemente melhoradas, se contrapostas às de treino para o modelo intermediário. Ao testar o modelo final em novos dados, o R^2 obtido foi de 0,919 e o EAM, EPAM e RQEM foram de 9,05, 9,94% e 12,74, respectivamente. Neste

caso, houve uma leve piora nas métricas. De qualquer maneira, tanto para o treino quanto para o teste, a simplificação do modelo de floresta randômica utilizando as cinco variáveis mais importantes não afetou de maneira negativa na predição da razão de aspecto.

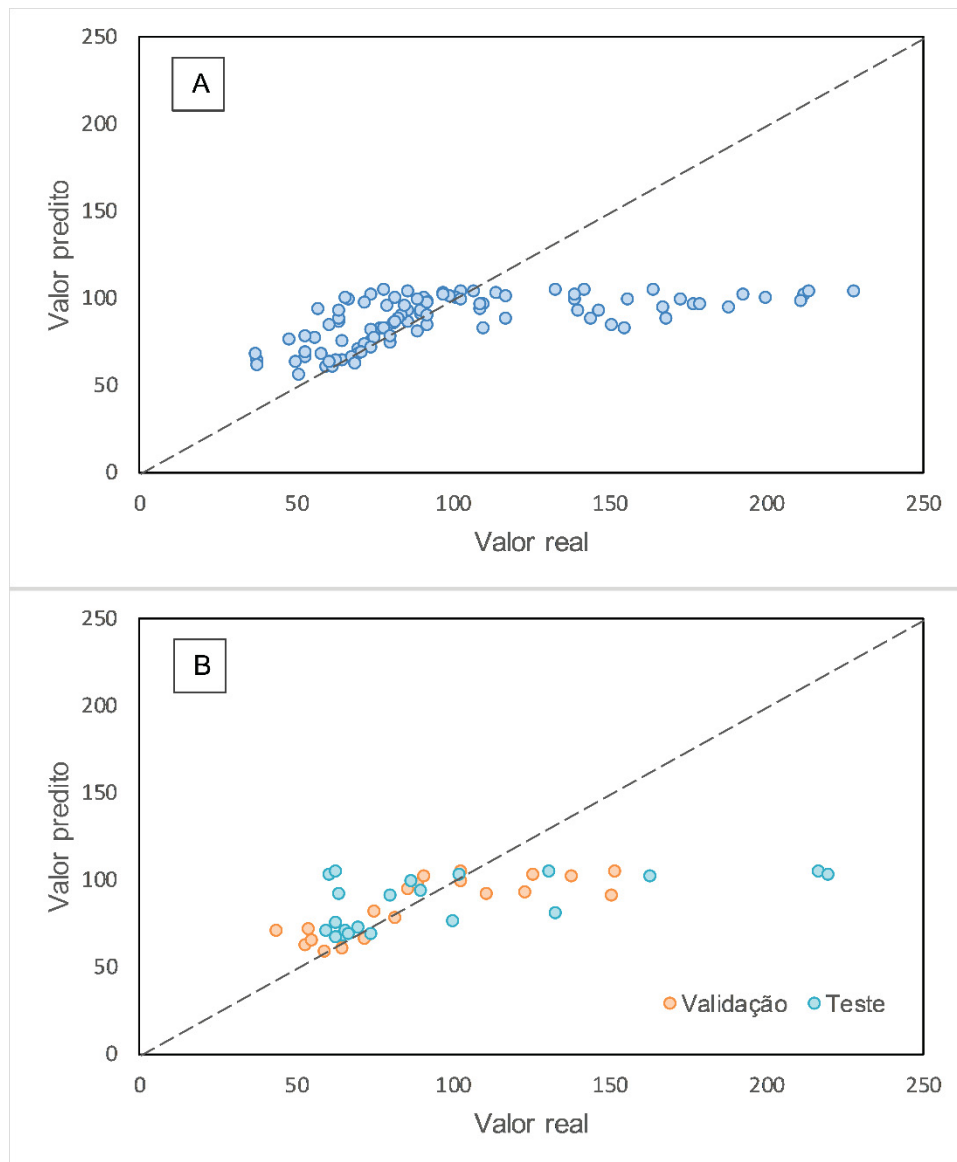
Desta forma, verifica-se que foi possível criar modelos simples utilizando o algoritmo de floresta randômica e que apresentaram métricas muito boas para a predição da razão de aspecto, especialmente considerando que as celuloses nanofibriladas produzidas advieram de diversas fontes de matéria-prima (madeiras moles, madeiras duras, plantas anuais) e com pré-tratamentos distintos e sob intensidades diferentes. Outro ponto relevante a ser levantado é a facilidade de medição das variáveis selecionadas para o modelo final. As propriedades químicas das polpas pré-tratadas podem ser medidas utilizando metodologias gravimétricas que fazem uso de equipamentos simples e reagentes baratos. Ademais, a dosagem enzimática é um parâmetro definido por quem está produzindo as nanoceluloses, enquanto que o índice de consistência k é um parâmetro que pode ser obtido a partir de análises reológicas feitas com um reômetro.

4.2.2 Redes neurais

Foram gerados três modelos de redes neurais: um modelo utilizando a mesma arquitetura da rede neural desenvolvida preliminarmente, um modelo com todas as 11 variáveis de entrada (chamado de intermediário aqui também) e um modelo final com as mesmas 5 variáveis escolhidas para o modelo de floresta randômica, para possibilitar a comparação entre os dois algoritmos. Não obstante, foi analisada a força relativa das variáveis de entrada para a predição da variável de saída no algoritmo de rede neural.

O modelo preliminar foi construído com uma camada oculta, 5 neurônios de entrada, 10 neurônios na camada oculta, um tamanho de lote de 2, uma taxa de aprendizagem de 0,01 e função de ativação sigmoideal. As variáveis de entrada foram teor mássico percentual de lignina total, celulose, hemicelulose e extraíveis e consumo energético (kWh/kg). O gráfico dos valores reais e valores preditos dos subconjuntos de treino e teste estão apresentados no GRÁFICO 5, enquanto as métricas estão elencadas na TABELA 15.

GRÁFICO 5 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL UTILIZANDO OS HIPERPARÂMETROS PRELIMINARES PARA PREDIÇÃO DA RAZÃO DE ASPECTO.



Fonte: A autora (2023)

TABELA 15 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E VALIDAÇÃO/TESTE DOS MODELOS DE REDE NEURAL PARA PREDIÇÃO DA RAZÃO DE ASPECTO DE NANOCELULOSES FIBRILADAS.

Modelo	Subconjunto	EAM	EPAM	RQEM	R ²
Preliminar	Treino	25,80	22,78%	40,09	0,1990
	Validação+teste	24,28	21,51%	38,47	0,2631
Intermediário	Treino	6,15	7,07%	8,26	0,9660
	Validação+teste	10,30	11,43%	14,06	0,9015
Final	Treino	12,54	13,70%	17,75	0,8429
	Validação+teste	13,57	13,85%	21,37	0,7725

Fonte: A autora (2023)

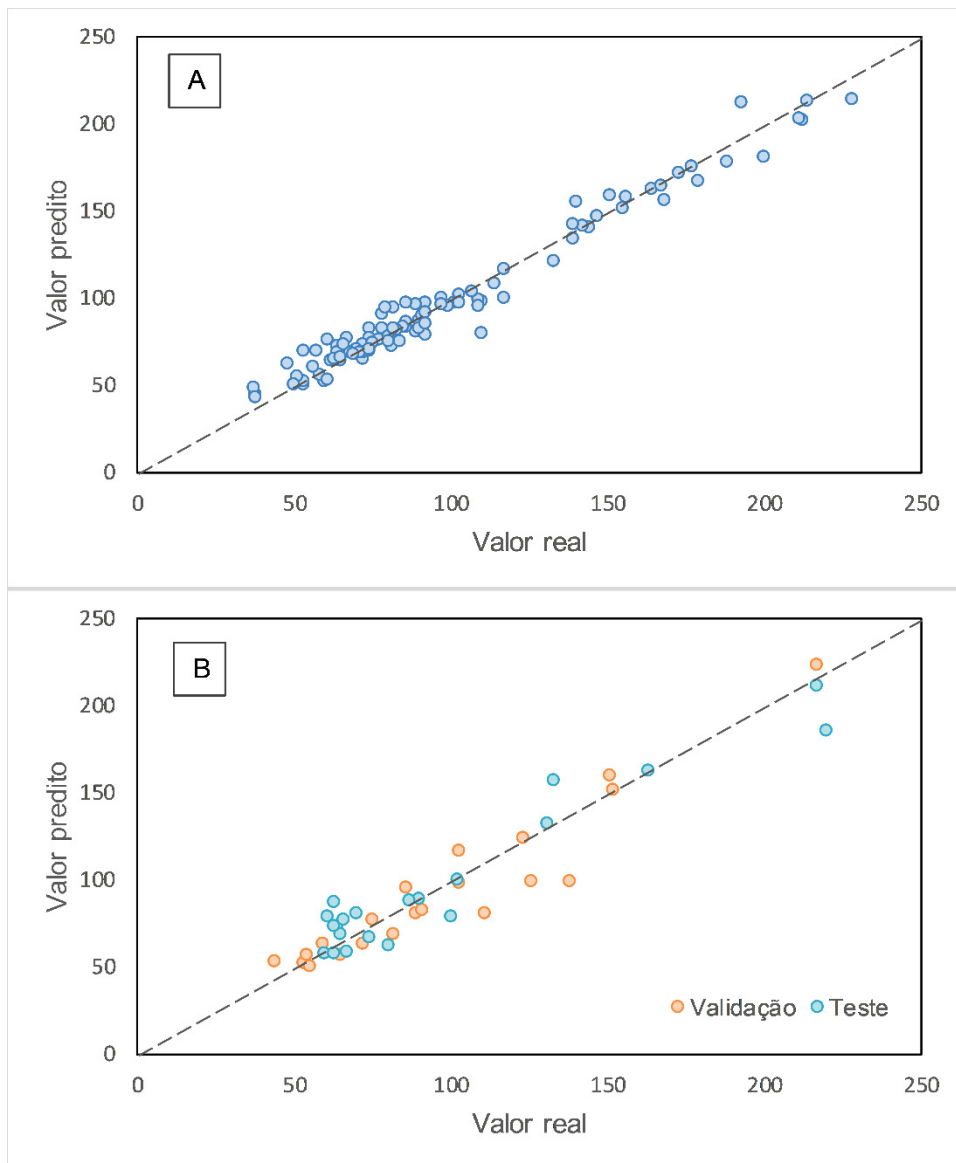
Ao se analisarem as métricas de desempenho na TABELA 15, as quais foram muito parecidas entre treino e validação/teste, pode-se verificar que, no geral, não houve uma boa adequação dos novos dados à arquitetura preliminar. Por exemplo, o coeficiente de determinação foi muito baixo, tanto para o subconjunto de treino quanto para o de validação e teste, apresentando, nessa ordem, os valores 0,1990 e 0,2631. Os erros percentuais médios foram um pouco altos, na ordem de 21 a 22%. O EAM do treino e validação/teste foram um tanto altos também, nos valores de 25,80 e 24,28, assim como o RQEM, de 40,09 e 38,47 respectivamente.

A partir da observação do GRÁFICO 5, constata-se que, majoritariamente, os valores preditos não ficaram próximos dos reais. Entretanto, é possível perceber que há duas regiões com comportamentos distintos: uma com razão de aspecto de 50 a 100 e outra entre 100 e 250. No caso da região de menor razão de aspecto, os valores preditos estão mais próximos da reta de 45°, enquanto que para as razões de aspectos maiores que 100, os valores preditos ficam praticamente em um platô de 100 e se distanciaram muito da reta. Provavelmente, isto está relacionado com os dados que foram utilizados para desenvolver esta arquitetura preliminar, que consistiu em 75% dos pontos com razões de aspecto entre 50 e 100 e o restante com razões de aspecto de 150 para cima. Naquele caso, os pontos originais de 50 a 100 foram muito bem preditos, enquanto aqueles com razões de aspecto maiores não. Para a modelagem com os novos dados, mas utilizando a arquitetura preliminar, percebe-se que esta tendência é mantida, apesar de, ainda assim, os pontos de 50 a 100 não terem sido perfeitamente preditos.

O segundo modelo de rede neural gerado (modelo intermediário) teve uma única camada oculta também, 11 neurônios de entrada (ou seja, todas as variáveis de

entrada, incluindo dosagem de enzima), 22 neurônios na camada oculta, um tamanho de lote de 3, uma taxa de aprendizagem de 0,1 e função de ativação sigmoideal. Esses parâmetros foram obtidos a partir da otimização Bayesiana aplicada. Os resultados gráficos do Pareto entre valores reais e valores preditos dos subconjuntos de treino e validação/teste estão apresentados no GRÁFICO 6, enquanto as métricas estão elencadas na TABELA 15, apresentada anteriormente.

GRÁFICO 6 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO/TESTE (B) DO MODELO DE REDE NEURAL INTERMEDIÁRIO PARA PREDIÇÃO DE RAZÃO DE ASPECTO.

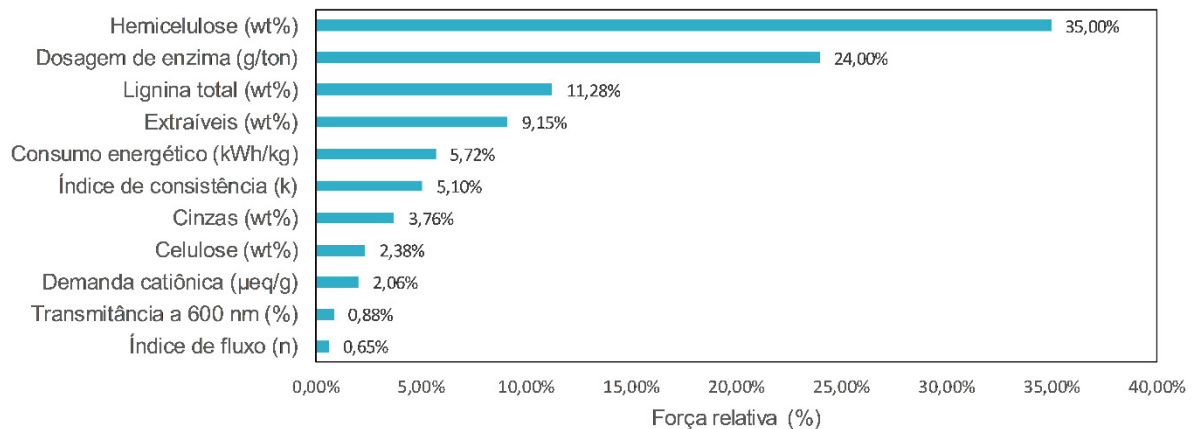


Fonte: A autora (2023)

Constata-se que a utilização de todas as variáveis e a aplicação da otimização Bayesiana contribuíram para o melhoramento da capacidade preditiva da rede neural. Neste caso, a otimização Bayesiana realizada obteve alguns parâmetros parecidos com a arquitetura preliminar: o uso da função de ativação sigmoidal na camada oculta e um tamanho de lote muito similar, além do número de neurônios na camada oculta ter sido também o dobro do número de neurônios na camada de entrada. Tanto os pontos reais de treino quanto de validação e teste foram bem preditos, uma vez que os pontos estão próximos da reta de 45°, além das métricas de desempenho indicarem um R^2 alto e erros baixos. Por exemplo, foi obtido para o treino um R^2 de 0,97, aproximadamente, e para a validação/teste de 0,9015. Além disso, o EAM do treino e validação/teste foram de 6,15 e 10,30, enquanto que o EPAM foi de 7,07% e 11,43%, respectivamente. O RQEM foi de 8,26 e 14,06, respectivamente também.

A partir deste modelo de rede neural criado com todas as variáveis de entrada, foram obtidas as forças relativas destas variáveis na predição da razão de aspecto. Os valores estão plotados no GRÁFICO 7.

GRÁFICO 7 - FORÇA RELATIVA ABSOLUTA DAS VARIÁVEIS DE ENTRADA PARA PREDIÇÃO DA RAZÃO DE ASPECTO UTILIZANDO O ALGORITMO DE REDE NEURAL.



Fonte: A autora (2023)

Para este algoritmo estudado, percebe-se que a hemicelulose desempenhou o papel de variável mais importante, com uma força relativa de 35% para a predição da razão de aspecto. A dosagem enzimática também demonstrou uma força relevante, similarmente ao ocorrido no estudo do algoritmo de floresta randômica, com uma força de 24%. As três seguintes variáveis mais importantes foram o teor de lignina total, extraíveis e consumo energético, apresentando os valores de 11,28%, 9,15% e

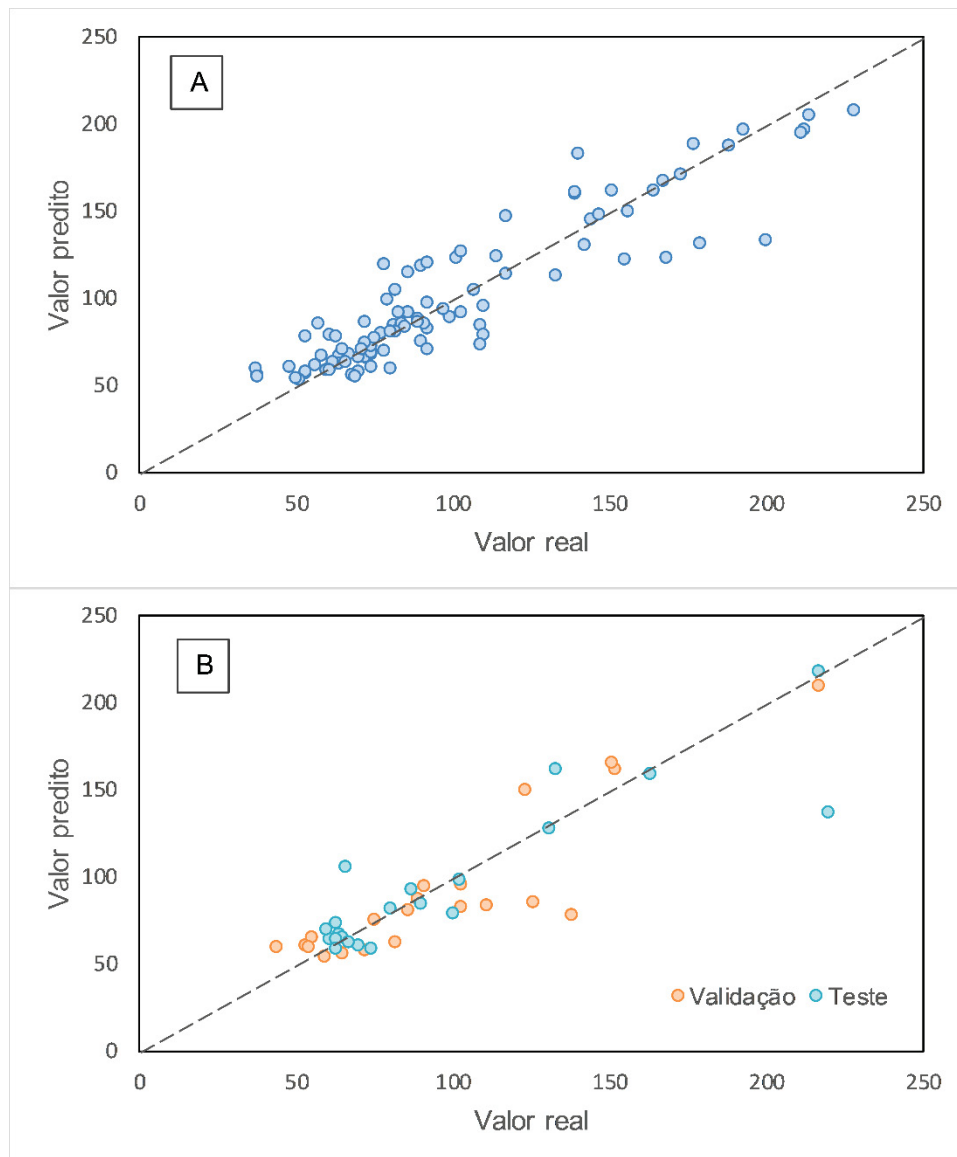
5,72%, respectivamente. Estas 5 variáveis somam, juntas, um total de 85,15% de força relativa.

Comparando-se com a importância relativa das mesmas variáveis, mas aplicadas no algoritmo de floresta randômica, observa-se que há uma certa similaridade entre elas. Apesar de não necessariamente as forças/importâncias relativas das variáveis terem que ser iguais entre os algoritmos, uma vez que o princípio de funcionamento de cada um é distinto, é interessante observar que a composição química se manteve como fator importante na predição da razão de aspecto, bem como a dosagem enzimática, que representa o tipo de pré-tratamento.

Para realizar uma comparação entre os desempenhos da rede neural e da floresta randômica, o terceiro modelo foi criado com as mesmas cinco variáveis mais importantes detalhadas pela importância relativa na floresta. Comparativamente com as cinco variáveis mais fortes para a rede neural, a única diferença está no uso da variável índice de consistência ao em vez do consumo energético, que apareceu com uma força um pouco maior que o índice de consistência (K) para a rede neural (5,72% contra 5,10%, respectivamente). Assim, as variáveis utilizadas para a predição foram: teor mássico de lignina total (%), hemicelulose (%), extraíveis (%), dosagem de enzima (g/ton) e o índice de consistência (K). O modelo teve uma única camada oculta, 5 neurônios de entrada, 10 neurônios na camada oculta, um tamanho de lote de 3, uma taxa de aprendizagem de 0,1 e função de ativação ReLU. Esses parâmetros, obtidos a partir da otimização Bayesiana, foram todos praticamente iguais à arquitetura obtida para as onze variáveis, com exceção da função de ativação, que desta vez foi a ReLU. Apesar do número de neurônios ser diferente, a proporção 1:2 entre neurônios de entrada: neurônios da camada oculta se manteve.

Os resultados gráficos do Pareto entre os valores reais e valores preditos dos subconjuntos de treino e validação/teste estão apresentados no GRÁFICO 8, enquanto as métricas estão na TABELA 15.

GRÁFICO 8 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO/TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DE RAZÃO DE ASPECTO.



Fonte: A autora (2023)

Percebe-se que a redução das variáveis não contribuiu para melhora da métrica ou mesmo para manutenção do bom desempenho obtido quando as onze variáveis foram aplicadas. Ao contrário, foram reportados maiores erros e R^2 menores. Por exemplo, o R^2 para o treino e validação/teste foram, respectivamente, de 0,8429 e 0,7725. Já o EAM, EPAM e RQEM foram, para o treino e validação/teste, de: 12,54 e 13,57; 13,70% e 13,85%; 17,75 e 21,37, respectivamente. Visualmente, esta redução no desempenho preditivo pode ser constatada ao se observar o GRÁFICO 8. Percebe-se que os valores seguem um pouco a tendência da reta de 45°, mas estão mais espalhados ao redor dela.

Comparando-se o modelo de floresta randômica e o obtido com a rede neural para predição da razão de aspecto, nota-se que o primeiro foi mais eficiente e teve melhores capacidades preditivas que o segundo. Por exemplo, foi possível obter bons resultados já com a arquitetura de floresta randômica preliminar, também com o modelo intermediário e principalmente com os resultados do modelo final. O modelo final de floresta randômica apresentou excelentes métricas, com R^2 altos, acima de 0,90, tanto para o subconjunto de treino quanto para o de teste, e erros de EAM, EPAM e RQEM mais baixos.

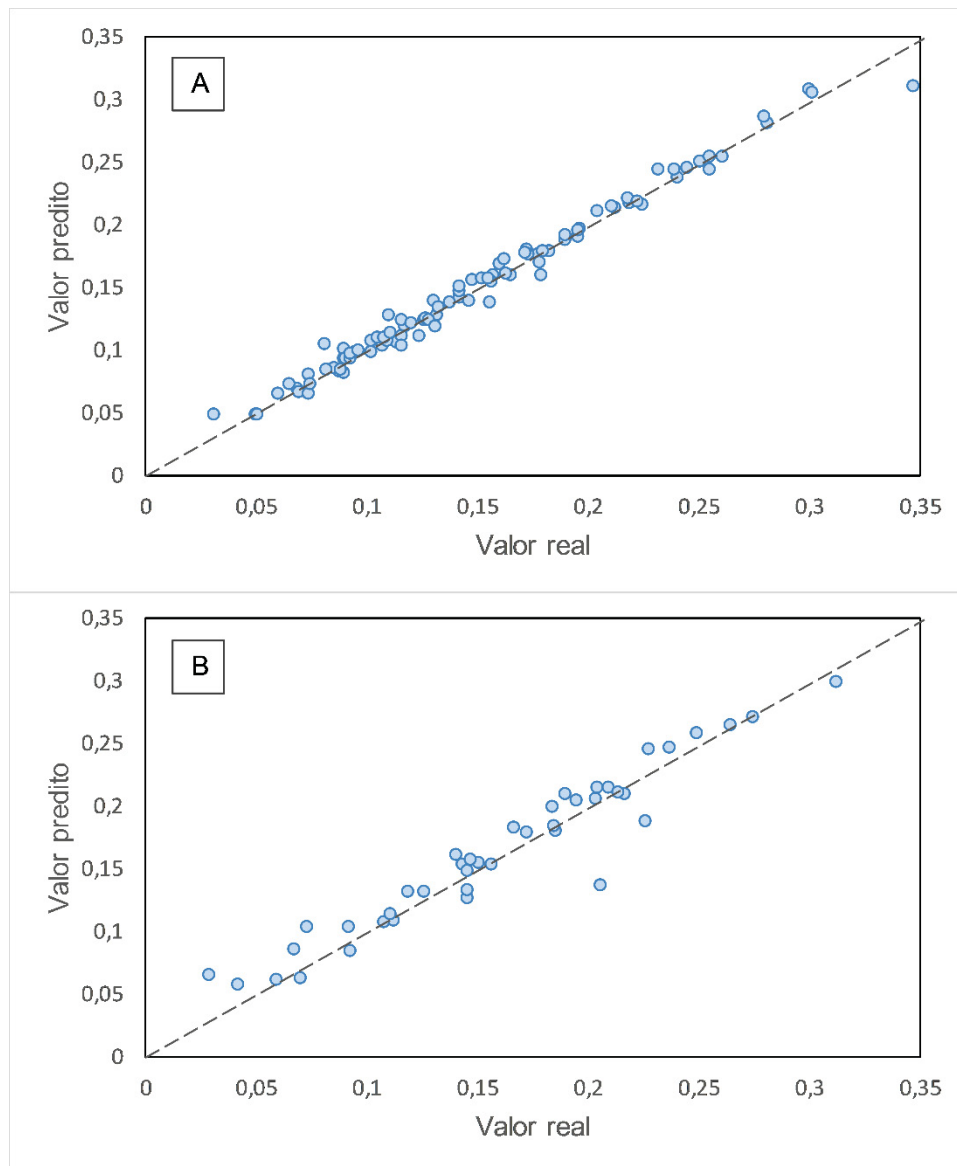
4.3 MODELOS DE PREDIÇÃO: RENDIMENTO DE NANOFIBRILAÇÃO DE NANOFIBRAS DE CELULOSE

4.3.1 Florestas randômicas

Foram gerados três modelos para predição do rendimento de nanofibrilação utilizando o algoritmo de floresta randômica: um modelo utilizando a mesma arquitetura do modelo preliminar (modelo construído inicialmente com 20 dados experimentais de nanocelulose mecânica e que está descrito no APÊNDICE D), um modelo com todas as variáveis de entrada (ou seja, incluindo a dosagem enzimática, que é nova em relação ao modelo preliminar) chamado de intermediário, e um modelo final com as duas variáveis mais importantes, de acordo com o resultado apresentado pela função da importância relativa das variáveis embutida na biblioteca do Scikit Learn.

O modelo preliminar foi construído com um número de árvores = 40 e um número mínimo de amostras nas folhas de 2. As variáveis de entrada foram demanda catiônica, transmitância a 600 nm e consumo energético. O gráfico dos valores reais e valores preditos dos subconjuntos de treino e teste estão apresentados no GRÁFICO 9, enquanto as métricas estão elencadas na TABELA 16.

GRÁFICO 9 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA COM OS HIPERPARÂMETROS PRELIMINARES PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.



Fonte: A autora (2023)

TABELA 16 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DOS MODELOS DE FLORESTA RANDÔMICA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DE NANOCELULOSES FIBRILADAS.

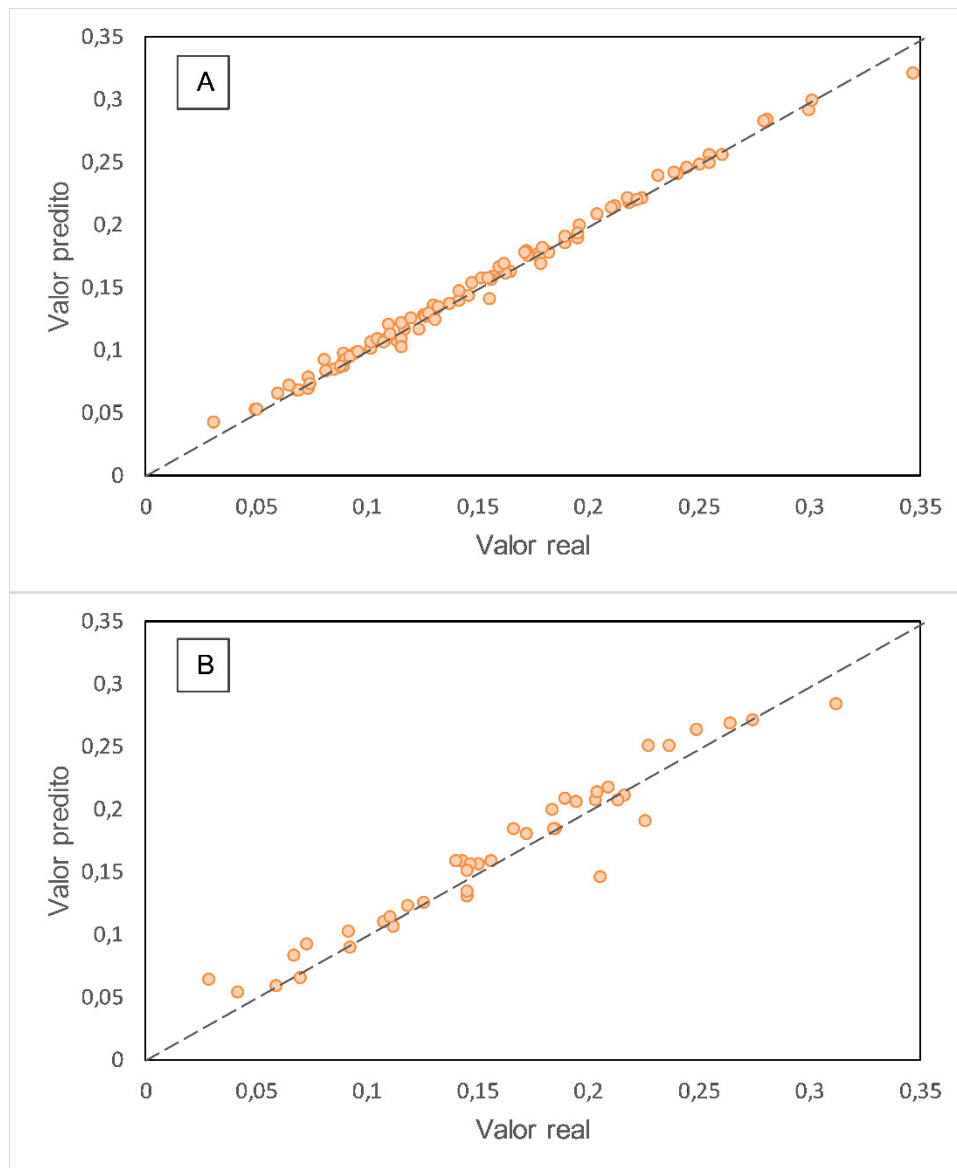
Modelo	Subconjunto	EAM	EPAM	RQEM	R²
Preliminar	Treino	0,005	4,65%	0,008	0,985
	Teste	0,012	11,38%	0,018	0,928
Intermediário	Treino	0,004	3,28%	0,005	0,993
	Teste	0,012	10,22%	0,016	0,937
Final	Treino	0,004	3,28 %	0,005	0,993
	Teste	0,013	12,31%	0,019	0,918

Fonte: A autora (2023)

Observando a figura A do GRÁFICO 9, percebe-se que a arquitetura preliminar se adequou muito bem ao novo subconjunto de treino. O valor do R² obtido de 0,985 foi alto e os erros gerados pequenos. Por exemplo, o EAM foi de 0,005, o EPAM de 4,65% e o RQEM de 0,008 somente. O subconjunto de teste também se adequou bem, com um R² acima de 0,9, especificamente no valor de 0,928. Os erros, apesarem de serem mais altos que os obtidos no treino, podem ser considerados baixos ainda. Observando a figura B do GRÁFICO 9, percebe-se que os dados ficaram próximos da reta de 45°. Desta maneira, a arquitetura preliminar e as variáveis consideradas como entrada do modelo foram já muito satisfatórias para predição de novos dados de rendimento de nanofibrilação.

O modelo gerado com todas as 11 variáveis de entrada possíveis e aplicando a função de otimização Bayesiana resultou numa arquitetura com as seguintes características: profundidade máxima das árvores = 38, quantidade máxima de features para fazer a divisão dos nós = 10, número de árvores = 150 e número mínimo de pontos nas folhas = 1. O GRÁFICO 10 apresenta os resultados plotados dos valores reais e preditos por esse segundo modelo. As métricas de desempenho estão apresentadas na TABELA 16.

GRÁFICO 10 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA INTERMEDIÁRIO PARA PREDIÇÃO DE RENDIMENTO DE NANOFIBRILAÇÃO.



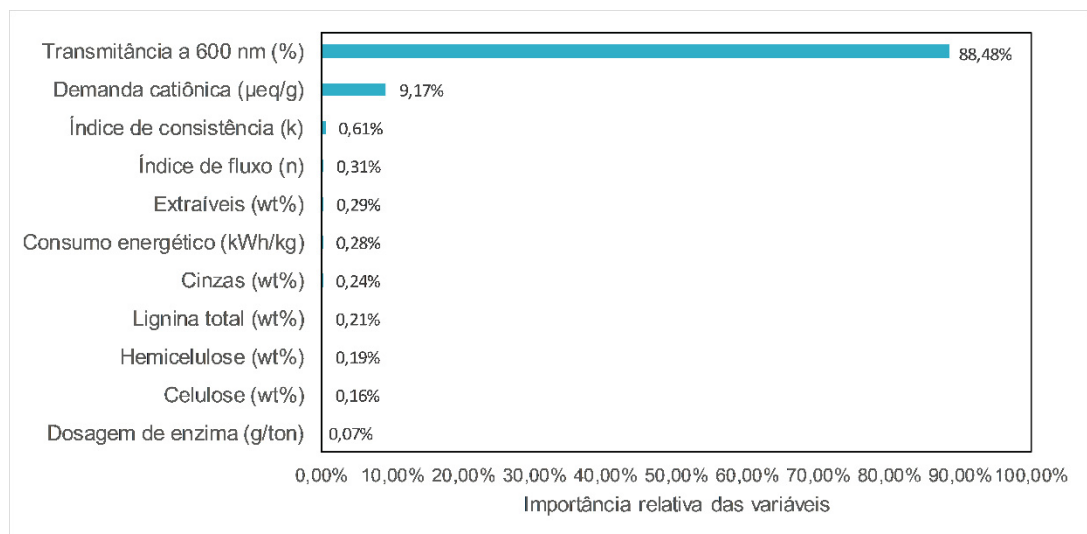
Fonte: A autora (2023)

Comparando o GRÁFICO 9 e GRÁFICO 10, verifica-se que o uso de todas as variáveis e uma nova arquitetura resultante da função de otimização permitiram um melhor desempenho tanto pelo subconjunto de treino quanto para o de teste. Quantitativamente, isto também pode ser verificado a partir da análise das métricas presentes na TABELA 16. Para o treino, o R^2 obtido foi muito próximo de 1, no valor de 0,993. O EAM, EPAM e RQEM foram levemente menores que o treino do modelo preliminar. Observando para o subconjunto de teste, o R^2 aumentou para 0,937 e os erros EPAM e RQEM foram um pouco menores também em relação ao teste do

modelo preliminar. Para o modelo intermediário, o EPAM foi de 10,22% e o RQEM foi de 0,016. O EAM se manteve o mesmo.

A partir deste modelo intermediário, foram avaliadas as importâncias relativas das variáveis para predição do rendimento de nanofibrilação. Os valores estão elencados no GRÁFICO 11.

GRÁFICO 11 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.



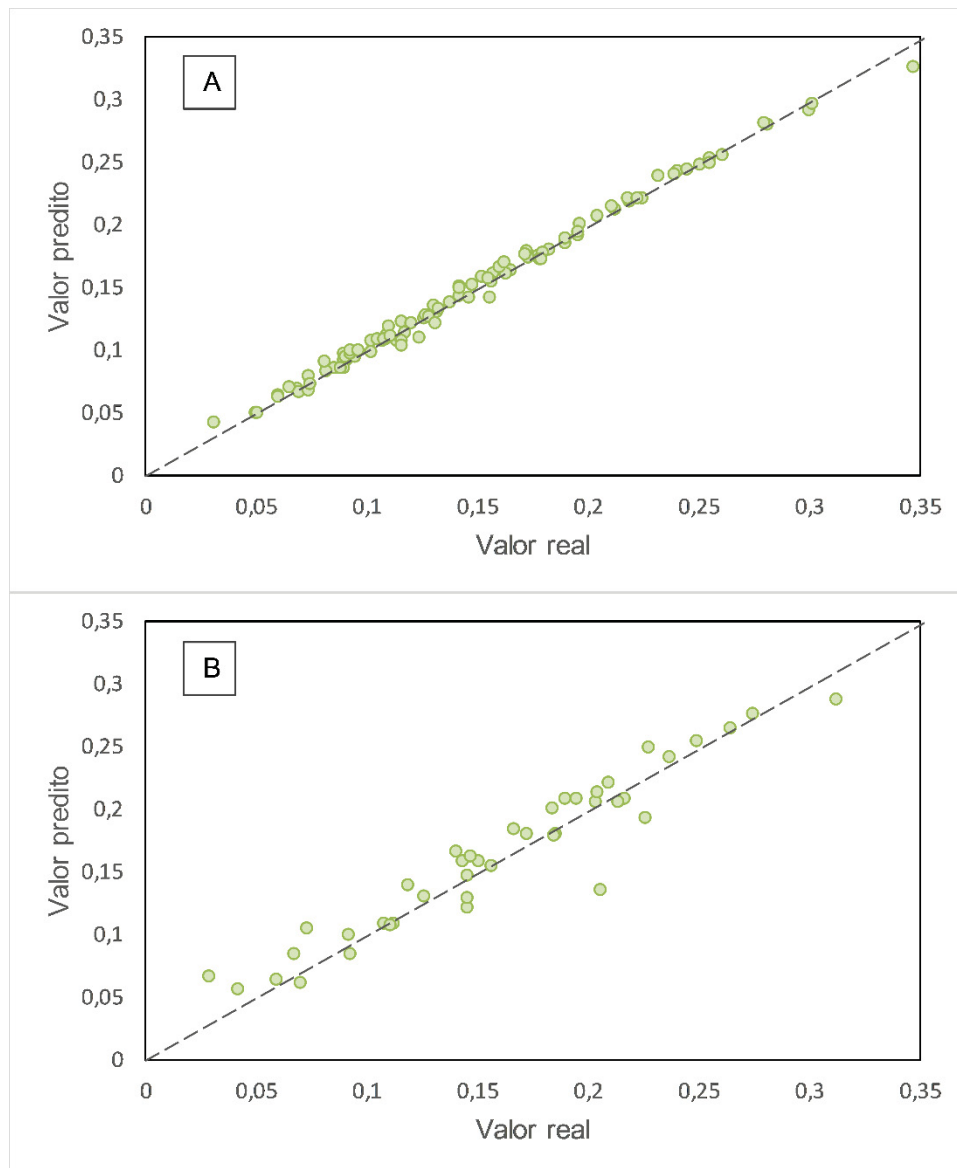
Fonte: A autora (2023)

Observa-se, pelo GRÁFICO 11, que a predição do rendimento de nanofibrilação está majoritariamente relacionada à variável “transmitância a 600 nm”, com uma contribuição de 88,48% do total. Em segundo lugar, tem-se uma atuação da demanda catiônica de 9,17%. O restante das variáveis possui importâncias ínfimas, com valores menores que 1%. Comparando com os valores dos coeficientes de Pearson, constata-se que há uma semelhança no que diz respeito à variável mais influente nos dois casos: a transmitância a 600 nm. Contudo, na TABELA 13, nota-se que o consumo energético possui uma relação um tanto significativa (linearmente) com o rendimento de nanofibrilação, enquanto que pelo GRÁFICO 11, este parâmetro praticamente não possui significância para o modelo de floresta randômica. A demanda catiônica tem uma importância relativamente forte tanto para o algoritmo de floresta quanto avaliando o coeficiente de Pearson. Estas observações também estão em consonância com o observado nos gráficos de dispersão da FIGURA 7.

Importante ressaltar também que, para o rendimento de nanofibrilação, o tipo de pré-tratamento não teve influência na predição do rendimento de nanofibrilação, como ocorreu para a razão de aspecto. Além disso, o tipo de matéria-prima também não parece desempenhar um papel relevante na predição, ou seja, o rendimento de nanofibrilação está diretamente relacionado a uma variável associada à nanocelulose final obtida. Na literatura, esta relação entre transmitância e rendimento de nanofibrilação já foi descrita. Conforme a proporção de fibras nanométricas em comparação com as micrométricas aumenta, constata-se um aumento na transparência da suspensão de CNFs. Ou seja, há uma transmitância mais alta da luz visível (MOVSESYAN *et al.*, 2022). Já a demanda catiônica está relacionada com os grupos funcionais negativos que remanescem das polpas lignocelulósicas (como carboxilatos ou grupos hidroxilas polarizados) e que, conforme ocorre a delaminação das cadeias de celulose, vão ficando mais expostas (SERRA-PARAREDA *et al.*, 2021c).

Como a importância das duas variáveis que estão no topo da lista do GRÁFICO 11 somam um total de 97,65% do total de importância, somente estas duas variáveis foram utilizadas para gerar o último modelo com o algoritmo de floresta randômica. Aplicando a função de otimização Bayesiana, a arquitetura final obtida contou com as seguintes características: profundidade máxima das árvores = 14, quantidade máxima de features para fazer a divisão dos nós = 2, número de árvores = 81 e número mínimo de pontos nas folhas = 1. O GRÁFICO 12 apresenta os valores preditos versus valores. As métricas de desempenho estão apresentadas também na TABELA 16.

GRÁFICO 12 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE RENDIMENTO DE NANOFIBRILAÇÃO.



Fonte: A autora (2023)

Comparando os GRÁFICOS GRÁFICO 10 e GRÁFICO 12, constata-se que, visualmente, os dois modelos estão com um desempenho muito similar. Avaliando e comparando as métricas presentes na TABELA 16, observa-se que as métricas de treino se mantiveram exatamente as mesmas em comparação com o modelo intermediário. Já para as métricas de teste, houve um pequeno decréscimo no valor do R^2 e um pequeno aumento nos erros, contudo, insignificantes e permanecendo ainda um desempenho de predição muito satisfatória para os dados não vistos pelo

modelo. Assim, para este modelo final, o R^2 para o subconjunto de teste foi de 0,918, enquanto o EAM, EPAM e RQEM foram 0,013, 12,31% e 0,019, respectivamente.

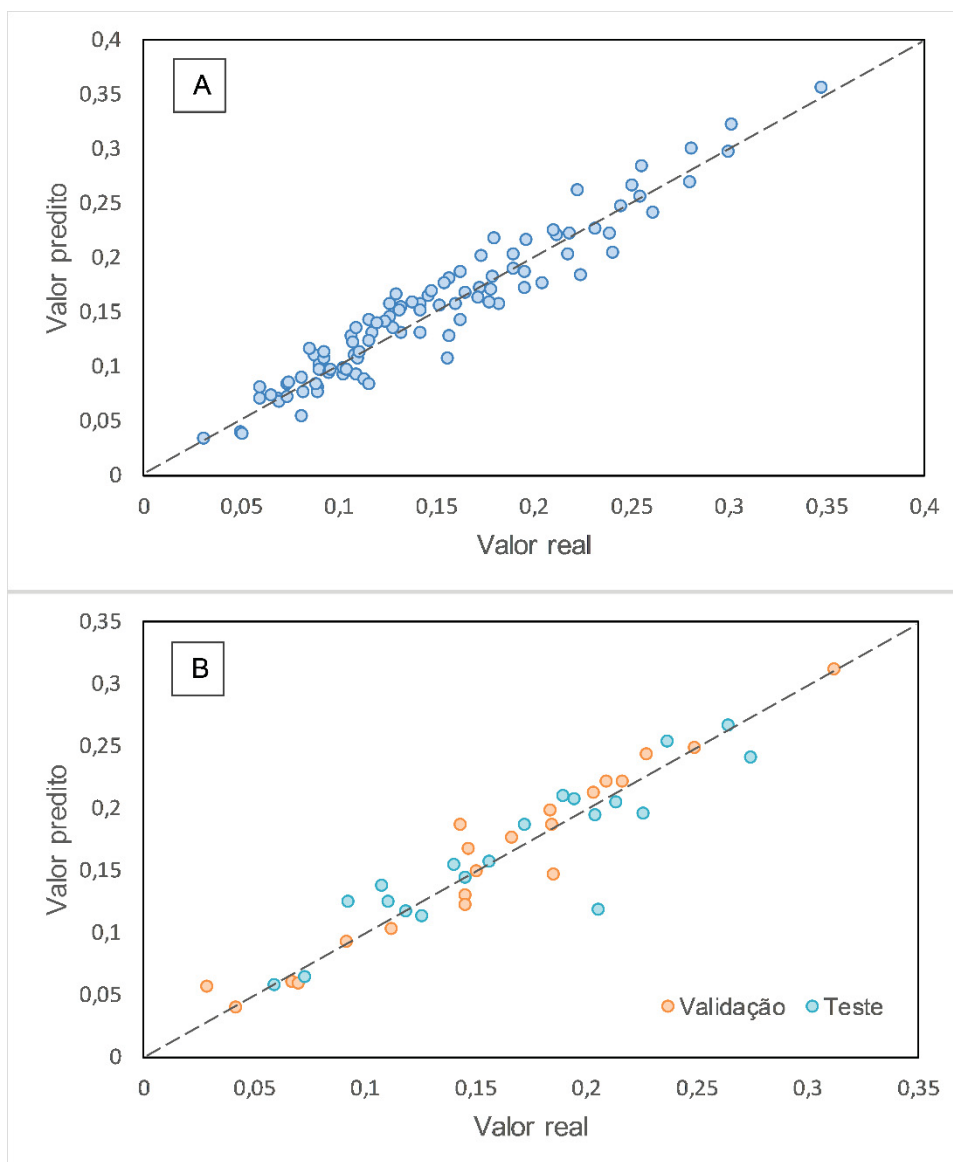
Todos estes indícios levam à conclusão de que a redução do total de variáveis para as duas mais importantes foi suficiente para manter um bom modelo de predição para o rendimento de nanofibrilação. Ademais, ressalta-se que a variável mais importante para a predição, a transmitância a 600 nm, é facilmente medível ao utilizar-se um espectrofotômetro, enquanto que a segunda variável mais importante, a demanda catiônica, poderia ser medida por um eletrodo. Por fim, observa-se que, mesmo utilizando-se um conjunto de dados preliminar pequeno de 20 pontos inicialmente, já havia sido possível determinar praticamente as mesmas variáveis mais importantes para descrição dos resultados de rendimento de nanofibrilação.

4.3.2 Redes neurais

Similarmente ao realizado para a predição da razão de aspecto, foram gerados três modelos de redes neurais: um modelo utilizando a mesma arquitetura da rede neural desenvolvida preliminarmente (modelo preliminar), um modelo com todas as 11 variáveis de entrada (modelo intermediário) e um modelo com as mesmas 2 variáveis escolhidas para o modelo de floresta randômica (modelo final), para possibilitar a comparação entre os dois algoritmos. A força relativa das variáveis de entrada para a predição da variável de saída no algoritmo de rede neural foi, ainda assim, realizada.

O modelo preliminar foi construído com uma camada oculta, 4 neurônios de entrada, 4 neurônios na camada oculta, um tamanho de lote de 16, uma taxa de aprendizagem de 0,01 e função de ativação ReLU. As variáveis de entrada foram demanda catiônica ($\mu\text{eq/g}$), transmitância a 600 nm, consumo energético (kWh/kg) e índice de consistência "k". Os resultados gráficos do Pareto entre valores preditos e reais estão apresentados no GRÁFICO 13, enquanto que as métricas estão na TABELA 17.

GRÁFICO 13 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL UTILIZANDO OS HIPERPARÂMETROS ORIGINAIS PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.



Fonte: A autora (2023)

TABELA 17 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E VALIDAÇÃO/TESTE DOS MODELOS DE REDE NEURAL PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.

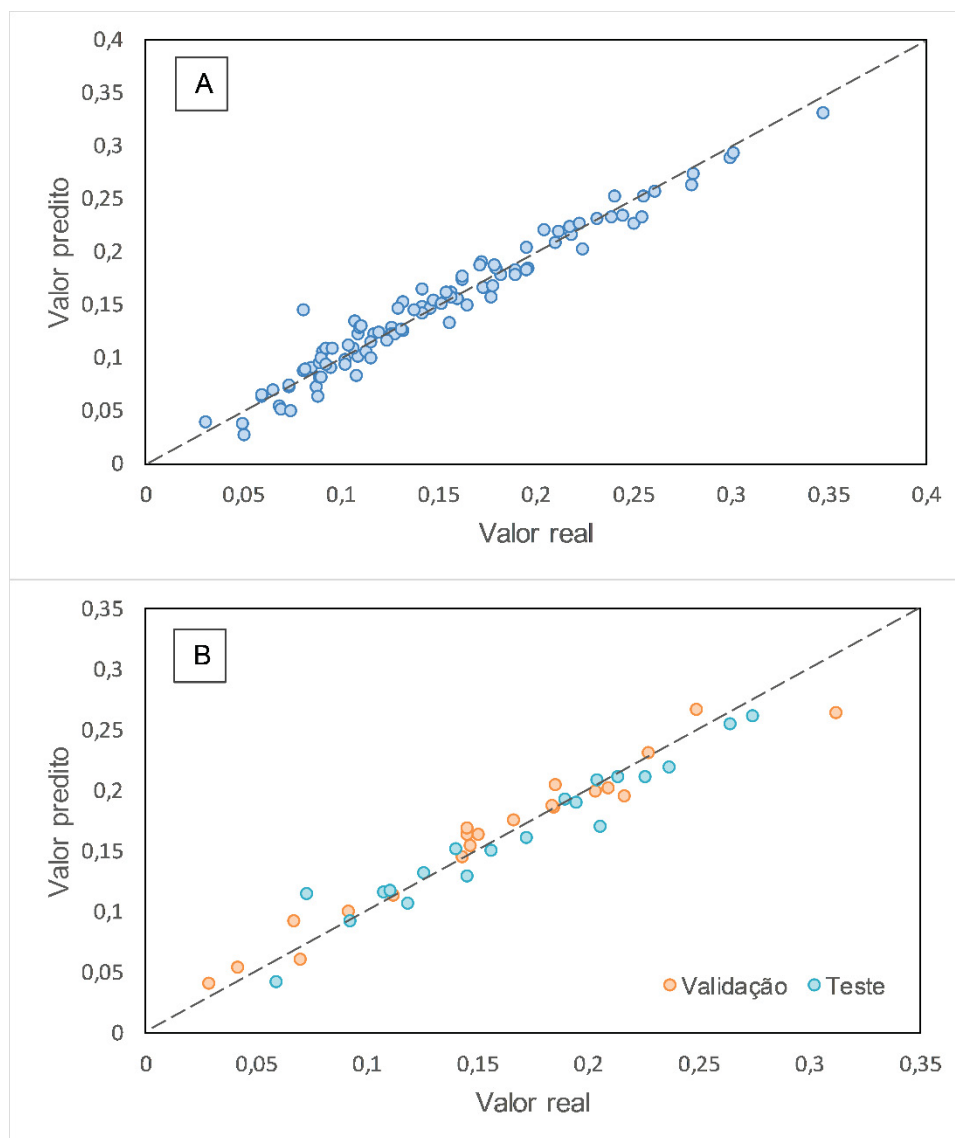
Modelo	Subconjunto	EAM	EPAM	RQEM	R ²
Preliminar	Treino	0,015	11,28%	0,018	0,920
	Validação+teste	0,015	11,92%	0,022	0,885
Intermediário	Treino	0,010	9,06%	0,014	0,955
	Validação+teste	0,013	10,46%	0,016	0,937
Final	Treino	0,008	6,63%	0,012	0,968
	Validação+teste	0,009	8,55%	0,016	0,943

Fonte: A autora (2023)

A partir da análise visual do gráfico de Pareto, verifica-se que a utilização dos hiperparâmetros originais, bem como as quatro variáveis de entrada citadas anteriormente, se adequou bem ao treinamento com novos dados de rendimento de nanofibrilação. Percebe-se que os valores reais e preditos estão bem próximos da reta de 45°. Esta constatação também é corroborada quando se analisam as métricas de desempenho. No caso do subconjunto de treino, o R² obtido foi de 0,920, enquanto que para a validação/teste foi de 0,885, dois valores altos e próximos. Ademais, os erros obtidos não foram altos. Por exemplo, o EAM foi de 0,015 tanto para os dados de treinamento quanto para os dados não vistos. O EPAM foi de 11,28% e 11,92% para o subconjunto de treino e validação/teste, respectivamente. Na mesma ordem citada, o RQEM foi de 0,018 e 0,022. Todos valores baixos. É interessante observar que esse modelo predisse bem valores de rendimento de nanofibrilação acima do valor máximo presente no subconjunto de treino preliminar, que foi de 20%, aproximadamente.

O modelo intermediário foi construído com uma camada oculta, 11 neurônios de entrada (ou seja, todas as variáveis de entrada), 12 neurônios na camada oculta, um tamanho de lote de 3, uma taxa de aprendizagem de 0,021, aproximadamente, e função de ativação sigmoidal. Os resultados gráficos do Pareto entre valores preditos e reais estão demonstrados no GRÁFICO 14, enquanto que as métricas estão na TABELA 17.

GRÁFICO 14 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL INTERMEDIÁRIO PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.



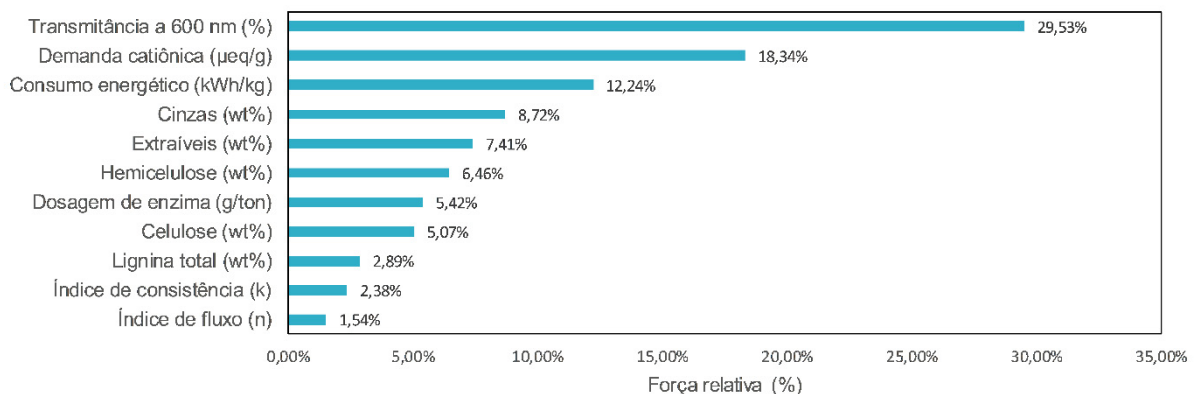
Fonte: A autora (2023)

Apesar do uso dos hiperparâmetros originais já ter providenciado um bom resultado de predição, o uso de todas as variáveis e a realização de uma busca Bayesiana para definição do valor dos hiperparâmetros da rede neural permitiu a melhora da capacidade preditiva do rendimento de nanofibrilação. Analisando-se o gráfico de Pareto, percebe-se que os dados estão muito bem alinhados com a reta de 45°, seguindo a sua tendência. Vale ressaltar que, tanto os dados de teste quanto de validação, ficaram bem próximos da reta. Além disso, as métricas apresentadas na TABELA 17 também demonstram essa melhora. O treino e o subconjunto de validação/teste apresentaram um valor de R^2 de 0,955 e 0,937. Os erros obtidos

também foram baixos. O EAM foi de 0,010 e 0,013, o EPAM de 9,06% e 10,46% e RQEM de 0,014 e 0,016, respectivamente.

A partir deste modelo intermediário, foram obtidas as forças relativas das variáveis na predição do rendimento de nanofibrilação. Os valores estão plotados no GRÁFICO 15.

GRÁFICO 15 - FORÇA RELATIVA ABSOLUTA DAS VARIÁVEIS DE ENTRADA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO UTILIZANDO O ALGORITMO DE REDE NEURAL.



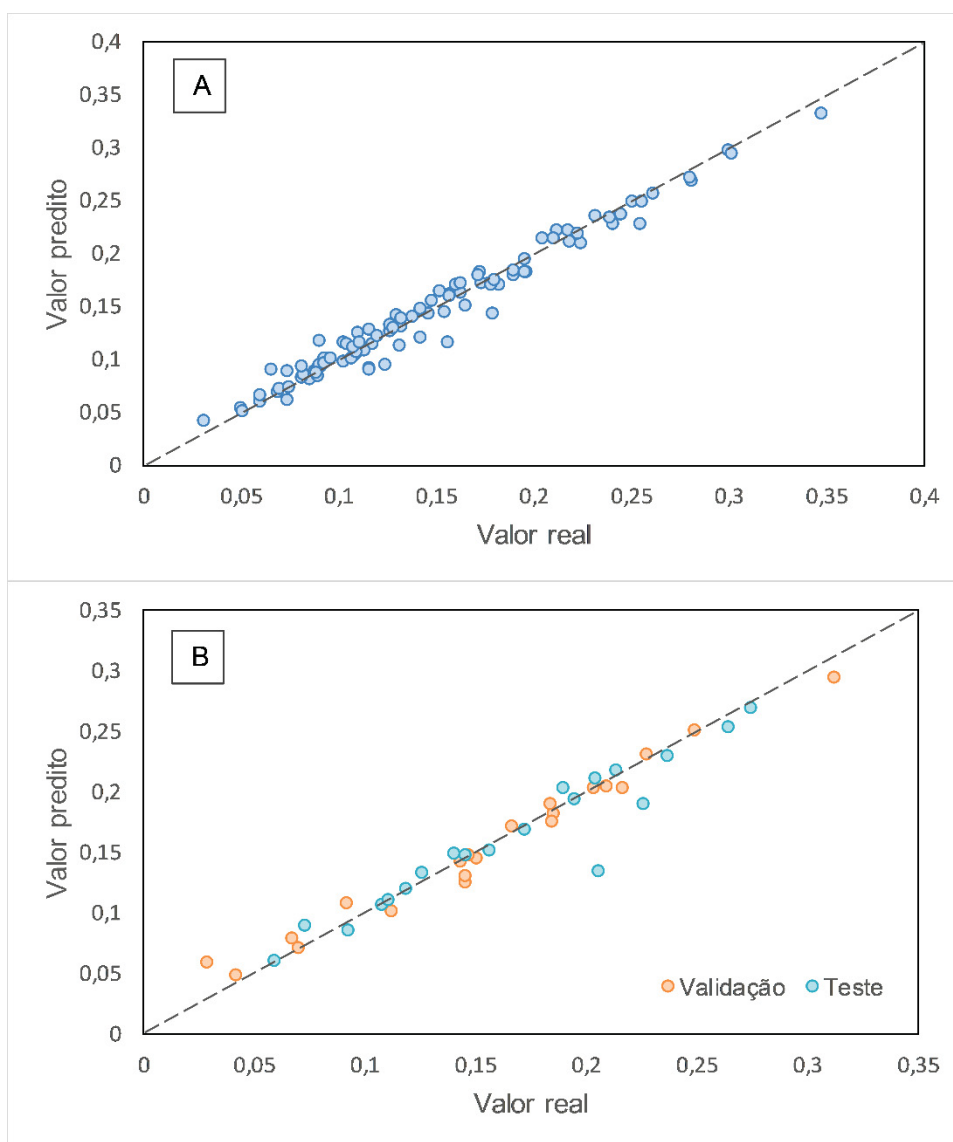
Fonte: A autora (2023)

Observando-se o GRÁFICO 15, percebe-se que a variável mais importante, para a rede neural, foi a transmitância a 600 nm, com uma força relativa de 29,53%. Em segundo lugar, a demanda catiônica, com uma força de 18,34%. Coincidentemente, estas duas variáveis também foram a mais e a segunda mais importante para o algoritmo de floresta randômica. Porém, diferentemente do modelo feito com árvores de decisão, a rede neural não atribuiu praticamente a totalidade das importâncias para estas duas variáveis. Em terceiro lugar, tem-se o consumo energético, com 12,24% de importância, em quarto, o teor de cinzas, de 8,72%, e em quinto o teor de extraíveis, com 7,41%. Juntas, estas cinco variáveis somam, ao todo, 76,24%. Desta maneira, percebe-se que, para se chegar a pelo menos 85% da importância total, seria necessário considerar ainda a força da hemicelulose e da dosagem enzimática, que respectivamente apresentaram as forças de 6,46% e 5,42%.

Para criação do terceiro modelo, também chamado de modelo final, foram utilizadas as mesmas duas variáveis utilizadas para o modelo final de floresta randômica. Ou seja, a transmitância a 600 nm e a demanda catiônica.

O modelo teve uma única camada oculta, 2 neurônios de entrada, 4 neurônios na camada oculta, um tamanho de lote de 3, uma taxa de aprendizagem de 0,04, aproximadamente, e função de ativação sigmoideal. Os resultados gráficos dos valores preditos versus reais estão apresentados no GRÁFICO 16, enquanto as métricas estão elencadas na TABELA 17.

GRÁFICO 16 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E VALIDAÇÃO E TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.



Fonte: A autora (2023)

Ao serem analisados os gráficos e as métricas, percebe-se que a redução para as duas variáveis mais fortes, apesar de estarem longe de somar, juntas, uma força relativa próxima de 100%, foi capaz de gerar um modelo ainda melhor que o

intermediário. Constatase que os dados ficaram mais próximos ainda da reta de 45°, além das métricas de desempenho terem melhorado. Por exemplo, o R² obtido para treino e validação/teste foram, respectivamente, 0,968 e 0,943. O EAM, de 0,008 e 0,009, o EPAM de 6,63% e 8,55% e o RQEM de 0,012 e 0,016, na mesma ordem.

Comparando o modelo gerado com o algoritmo de floresta e o gerado com rede neural, percebe-se que os dois lograram resultados muito bons de predição. No caso da floresta randômica, o subconjunto de treino teve um melhor desempenho que o da rede neural. Porém, no caso do subconjunto de validação/teste, foi o contrário. A rede neural performou um pouco melhor para dados não vistos para treino. Por conta disso, o modelo de rede neural final poderia ser considerado o melhor dos modelos obtidos.

4.4 RESULTADOS EXPERIMENTAIS DE PERMEAÇÃO COM MEMBRANAS DE PVA E CNF MECÂNICA

4.4.1 Resultados de permeação e seletividade

Os resultados obtidos de permeação e seletividade coletados para a membrana de PVA puro a diferentes pressões de operação estão apresentados na TABELA 18 e TABELA 19.

TABELA 18 - RESULTADOS DE PERMEANÇA PARA AS MEMBRANAS DE PVA PURO TESTADAS EM LABORATÓRIO.

Pressão (bar)	Perm. * CO₂ (1)	Perm. * CO₂ (2)	Perm. * CO₂ (3)	Perm. * CH₄ (1)	Perm. * CH₄ (2)	Perm. * CH₄ (3)
2	0,0107	0,0005	NP**	0,0626	0,0079	NP**
5	0,0528	0,0023	NP**	0,1390	0,0051	NP**
10	0,0501	0,0022	NP**	0,1010	0,0044	NP**

*Permeação: (m³ (STP).m².h⁻¹.bar⁻¹)

**NP: Não permeou

Fonte: A autora (2023)

TABELA 19 - RESULTADOS DE SELETIVIDADE CO₂/CH₄ PARA AS MEMBRANAS DE PVA PURO TESTADAS EM LABORATÓRIO.

Pressão (bar)	Seletividade (1)	Seletividade (2)	Seletividade (3)
2	0,17	0,07	N/A
5	0,38	0,46	N/A
10	0,50	0,50	N/A

Fonte: A autora (2023)

Observa-se que os resultados obtidos foram muito heterogêneos e não houve reprodutibilidade. Por exemplo, percebe-se que a membrana (1) apresentou maiores permeações, tanto de CO₂ quanto de CH₄, a membrana (2) apresentou uma permeação menor, para os dois componentes gasosos também, enquanto que a membrana (3) não permitiu o transporte de massa em valores medíveis de nenhuma das espécies em nenhuma das pressões testadas. Outra tendência observada foi de que, a 2 bar, obtiveram-se os menores valores de permeações, a 5 bar houve um aumento e a 10 bar houve uma redução. Ademais, apesar de ter sido utilizado um sistema de umidificação de gases, a permeação do CH₄, em todos os casos, com exceção da membrana (3) que não registrou permeação, foi maior que o de CO₂. Desta maneira, a seletividade CO₂/CH₄ não foi favorável para o componente de interesse a ser separado, apresentando um valor sempre abaixo de 1.

Foi realizada uma duplicata de testes com membranas de PVA e CNF (4%) e tendências similares foram constatadas. Estes valores estão apresentados na TABELA 20.

TABELA 20 - RESULTADOS DE PERMEANÇA E SELETIVIDADE PARA AS MEMBRANAS DE PVA E CNF (4%) TESTADAS EM LABORATÓRIO.

Pressão (bar)	Perm. * CO ₂ (1)	Perm. * CO ₂ (2)	Perm. * CH ₄ (1)	Perm. * CH ₄ (2)	Seletividade CO ₂ /CH ₄ (1)	Seletividade CO ₂ /CH ₄ (2)
2	Não foi possível medir	0,00006	Não foi possível medir	0,00038	N/A	0,16
5	Não foi possível medir	0,00043	Não foi possível medir	0,00038	N/A	1,14
10	Não foi possível medir	0,00053	Não foi possível medir	0,00108	N/A	0,49

Permeança medida em m³ (STP).m².h⁻¹.bar⁻¹

Fonte: A autora (2023)

Da TABELA 20, observa-se que não foi possível medir a permeação pela membrana (1) pois houve a presença de alguma falha que levou a um fluxo de gás de permeado muito alto, não sendo possível medir com os medidores de bolha disponíveis no laboratório. A membrana (2), por sua vez, apresentou um comportamento oposto: valores de permeação muito baixos. Para o CO₂ e CH₄, neste caso, a tendência de permeação foi de aumento conforme a pressão de operação aumentou. É interessante observar que o teste a 5 bar foi o único caso em que a seletividade CO₂/CH₄ foi um pouco maior que 1. Entretanto, é importante salientar que não é possível tirar conclusões muito certas com relação a tendências das permeações pelo fato de a permeação ter sido muito baixa. Ou seja, no momento de realizar a medição, o tempo de contagem para a bolha de gás avançar poucos mL é longo, o que leva a imprecisões na medida.

Algumas razões que podem ter contribuído para o comportamento observado ser diferente do relatado em literatura podem estar atreladas ao sistema de umidificação. No caso do sistema atual, não foi possível criar uma adaptação que gerasse vapor de água a pressão suficiente para passar pelo sistema, tampouco adaptar a célula com uma entrada para um gás de arraste com umidade, além de a umidificação ter tido que ser realizada nos gases puros, separadamente. No primeiro caso, alguns artigos consultados como referência executavam uma etapa de pré-umidificação das membranas com vapor de água. Após atingido o equilíbrio, passavam-se os gases úmidos a serem separados (VENTURI; ANSALONI;

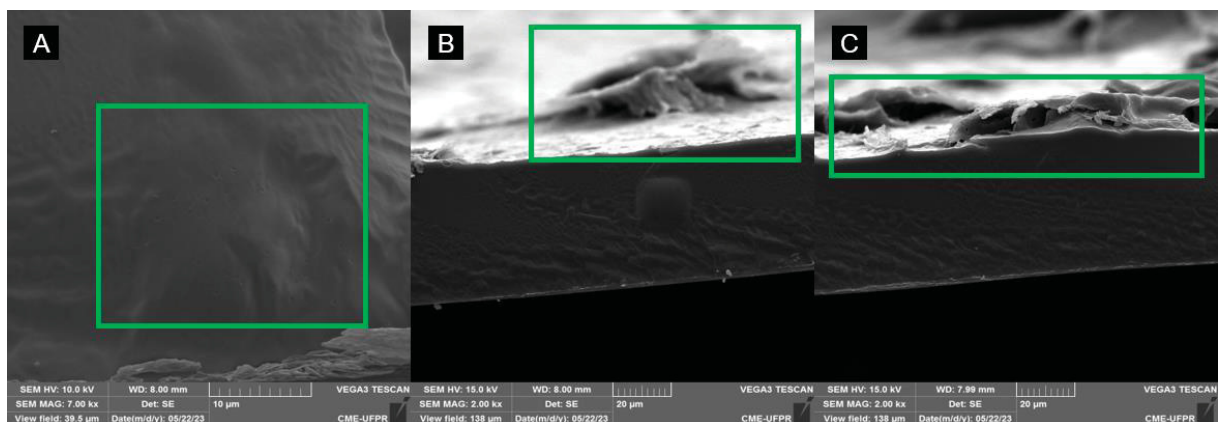
BASCETTI, 2016). No segundo caso, alguns sistemas de permeação contavam com gás de arraste úmido, o que pode ajudar a manter a mesma umidade nas duas interfaces da membrana (do lado do retentado e do lado do permeado) (JAHAN *et al.*, 2018a, 2018b, 2020; SAEED *et al.*, 2020). Por fim, alguns destes artigos também trabalharam utilizando uma mistura de CH₄ e CO₂, ao em vez dos gases puros (JAHAN *et al.*, 2018a, 2018b, 2020). Como o metano e o dióxido de carbono interagem com a água de formas distintas, ao ser feito o teste com os gases puros pode ser que a umidade da membrana não tenha sido mantida a mesma durante todo o teste.

Ademais, foi realizada a investigação das características das membranas que também possam ter influenciado neste comportamento constatado, especialmente seu aspecto visual obtido por microscopia eletrônica de varredura.

4.4.2 Microscopia eletrônica de varredura

As imagens obtidas da seção transversal das membranas de PVA puro e compósitas com CNF estão apresentadas nas FIGURA 8 e FIGURA 9. A FIGURA 10 corresponde à superfície da membrana compósita. Apesar de inicialmente não ter sido um objetivo a avaliação da superfície em si, uma das amostras deste filme se dobrou no momento de realizar a microscopia, e foi possível enxergar sua superfície de forma clara. Assim, optou-se por apresentar as imagens obtidas.

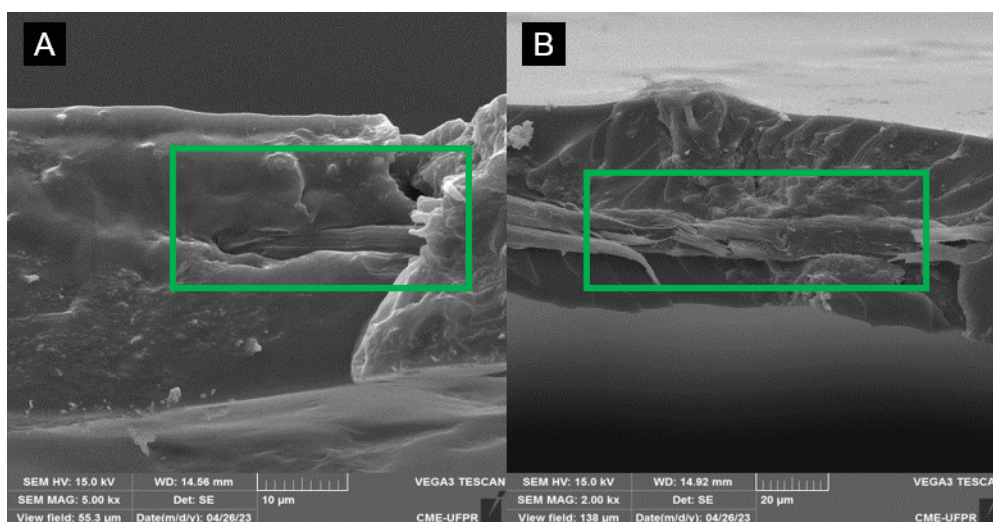
FIGURA 8 – IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA DA SEÇÃO TRANSVERSAL DAS MEMBRANAS DE PVA PURO



Fonte: A autora (2023)

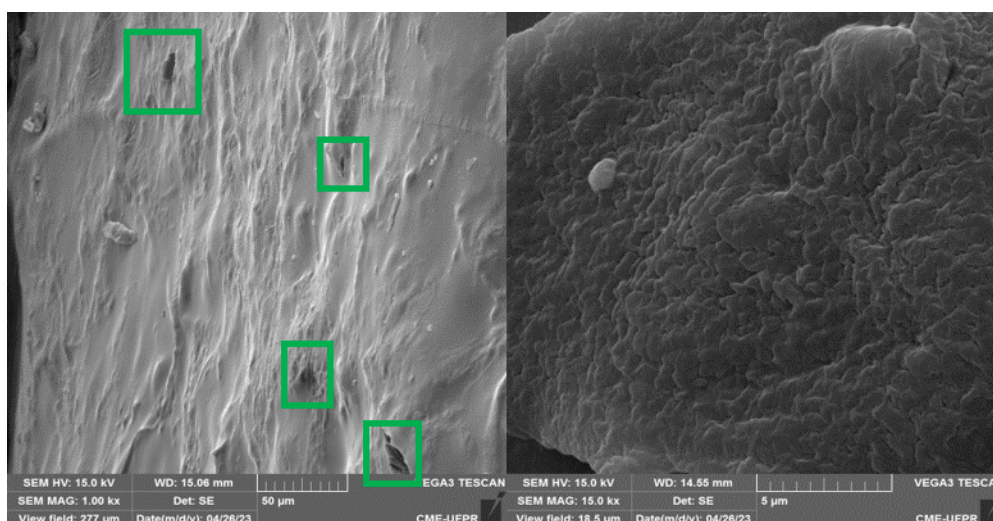
Analisando as imagens obtidas para a seção transversal das membranas de PVA puro, percebem-se duas características principais. Uma característica que se constata é que a membrana é densa, estando de acordo com a metodologia empregada para sua produção. Porém, observam-se algumas falhas na sua estrutura. Por exemplo, dentro da área verde demarcada na parte (A) da FIGURA 8, constatam-se pequenas concavidades. Ademais, apesar de não ser o foco da imagem, também é possível avaliar a superfície da membrana, na parte (B) e (C), dentro da área verde. Observam-se deformidades que deixam a superfície com um aspecto heterogêneo.

FIGURA 9 - IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA DA SEÇÃO TRANSVERSAL DAS MEMBRANAS COMPÓSITAS DE PVA E CNF.



Fonte: A autora (2023)

FIGURA 10 - IMAGENS DE MICROSCOPIA ELETRÔNICA DE VARREDURA DA SUPERFÍCIE DAS MEMBRANAS COMPÓSITAS DE PVA E CNF.



Fonte: A autora (2023)

A inserção de CNFs mantém a estrutura da membrana densa, porém com um aspecto mais desuniforme, menos liso, do que no caso do PVA puro, como pode se observar na FIGURA 9. Ainda assim, os indícios são de que ocorreu uma boa interação entre as CNFs e a matriz de PVA. Porém, é possível observar também a presença de macrofibras de celulose, que são formadas pelas micro e nanofibrilas. Estas estruturas maiores podem não ter interagido muito bem com o polímero, já que se percebe que a fratura deixou justamente esta parte exposta. Isto pode ocorrer, provavelmente, pela menor disponibilidade dos grupamentos hidroxila, os quais interagem fortemente com o PVA. É comum que uma suspensão de celulose que foi defibrilada mecanicamente contenha pedaços de celulose que não foram muito bem desagregados, em tamanhos micrométricos ou maiores, uma vez que o processo não conta com etapas de hidrólise ácida ou enzimática.

Por sua vez, visualizando a FIGURA 10, observa-se a superfície das membranas compósitas muito heterogêneas também. Dentro das áreas demarcadas em verde na parte (A) da figura, é possível ver orifícios e falhas. Na parte (B), visualiza-se uma estrutura extremamente rugosa. Estas duas texturas diferentes de superfície foram observadas para a mesma amostra, demonstrando a inhomogeneidade do filme.

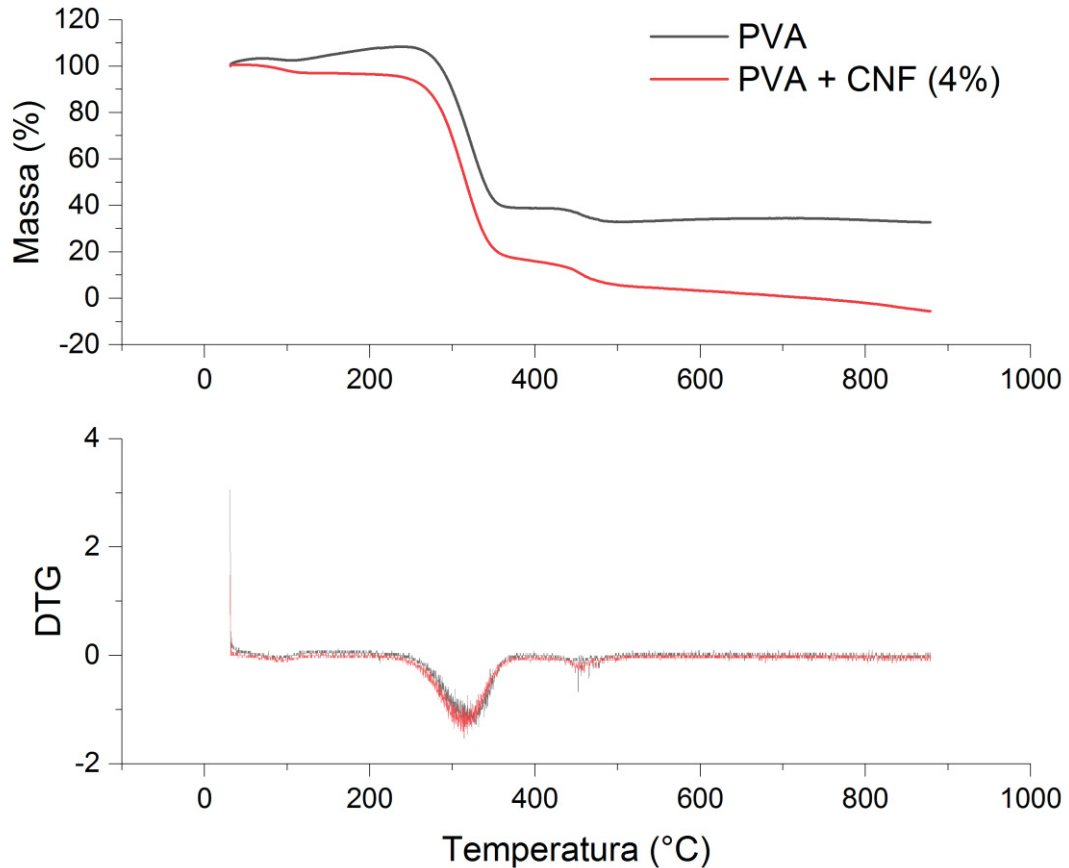
A partir da análise das imagens obtidas, constata-se que, de forma geral, as membranas não possuem uma característica homogênea, necessária para obter-se resultados reprodutíveis de permeação de gases. As moléculas gasosas de CO_2 e CH_4 são muito pequenas, com dimensões na ordem de poucos nanômetros, portanto, qualquer falha na estrutura facilita a passagem das moléculas por caminhos preferenciais. Desta maneira, o que se observa na etapa de permeação são resultados diversos, ora com característica de barreira, ora com características de alta permeabilidade. Curiosamente, percebeu-se uma tendência de a molécula de metano permear mais que o CO_2 , em vez de se manterem permeações parecidas. A presença destas falhas pode ter afetado o mecanismo de transporte principal, que pode ter sido o da difusão de Knudsen em vez do transporte facilitado. Este tipo de mecanismo é caracterizado pela presença de poros com raios menores que o caminho médio livre para a passagem das moléculas gasosas, e conseqüentemente estas moléculas se chocam contra as paredes dos poros com uma frequência maior do que com outras moléculas (BAKER, 2004). Desta maneira, aquelas com massa molecular menor permeiam a uma taxa maior. Como a massa molecular de CH_4 é menor que a do CO_2 , o hidrocarboneto pode ter sido favorecido no transporte.

As causas dessas falhas podem ser diversas. Por exemplo, foram utilizadas placas de Petri de teflon, uma vez que a nanocelulose e o PVA se aderem ao vidro. Entretanto, pode ser que este tipo de material ainda não seja o ideal para este compósito, pois ainda foi verificado uma certa aderência do filme no molde. Outra possível razão para a ocorrência de imperfeições está no método de secagem. Apesar de ter sido utilizada uma temperatura considerada baixa, de 30°C, foi empregada uma estufa de circulação de ar. O ar, com uma certa turbulência e em contato com o filme, pode ter contribuído para a aparição das imperfeições. Ademais, as membranas foram postas na estufa já previamente aquecida, além de que cada placa estava em uma posição diferente (algumas mais próximas do fundo do equipamento, outras mais próximas da porta de abertura). Estes fatores também podem contribuir para o aparecimento de falhas por conta da ocorrência de pontos quentes na matriz ou diferença de temperatura a que cada membrana estava submetida.

4.4.3 Análise termogravimétrica

Os resultados da análise termogravimétrica estão apresentados no GRÁFICO 17.

GRÁFICO 17 - PERDA DE MASSA (%) E DERIVADA DA MASSA (DTG) EM FUNÇÃO DA TEMPERATURA



Fonte: A autora (2023)

A análise termogravimétrica do filme de PVA puro e a do compósito com CNF (4%) demonstra a ocorrência de três fases de degradação principais. No caso do PVA puro, apesar de ter sido feita três vezes a análise, percebe-se que houve um aumento da massa de até uns 10%, aproximadamente. Este aumento de massa provavelmente ocorreu por erro de equipamento ou operação, já que não foi observado o mesmo fenômeno para o compósito.

O primeiro evento térmico, que ocorre a partir de aproximadamente 80°C até em torno de uns 120°C, pode ser atribuído à evaporação da água adsorvida no filme. É válido lembrar que os materiais trabalhados possuem afinidade pela água,

facilitando a adsorção de umidade presente no ambiente. O segundo evento térmico ocorre em temperaturas próximas tanto para o PVA quanto para o compósito. O início da degradação se dá em torno de 240 °C e termina próximo de 380 °C. As reações de degradação que ocorrem nesta etapa são principalmente de desidratação do PVA e geração de componentes voláteis e a despolimerização e desidratação da celulose (e outros componentes que possam estar presentes também, como hemicelulose) (GAN *et al.*, 2019; MAHARDIKA *et al.*, 2018). No caso da desidratação do PVA há formação de polienos, ou seja, hidrocarbonetos poliméricos onde estão presentes ligações duplas com e sem presença de hidroxilas. Como o grau de hidrólise do PVA utilizado é de 87-89%, há também a eliminação de grupos acetatos na forma de ácido acético (PENG; KONG, 2007). O último evento térmico ocorre entre 410 a 450°C. Nesta etapa, reações de ciclização, quebra de cadeia e decomposição molecular tomam parte, levando à formação de *char*, acetaldeído e outros (PENG; KONG, 2007). Nesta temperatura, componentes carbônicos também são degradados e transformados em *char*. As temperaturas observadas para estes eventos térmicos estão de acordo com o que é reportado em literatura, além da pouca diferença na temperatura de início de degradação percebida entre filmes de PVA puro e compósitos de PVA com CNF ou CNC a 4% de concentração (LEE *et al.*, 2020; VORONOVA *et al.*, 2015; YUWAWECH; WOOTHIKANOKKHAN; TANPICHAI, 2015). Desta maneira, para esta quantidade adicionada de nanopartícula, não foi possível constatar uma melhora significativa de estabilidade térmica.

4.4.4 Determinação de propriedades mecânicas

Os resultados medidos de resistência à tração e módulo de elasticidade para as membranas de PVA puro e PVA+CNF (4%) estão apresentados na TABELA 21.

TABELA 21 - RESISTÊNCIA À TRAÇÃO E MÓDULO DE ELASTICIDADE MEDIDOS PARA AS MEMBRANAS DE PVA PURO E COMPÓSITAS.

Membrana	Resistência à tração (MPa)	Módulo de elasticidade (GPa)
PVA puro	47,7 ± 17,6	1,6 ± 0,7
PVA + CNF (4%)	128,3 ± 33,1	4,7 ± 1,3

Fonte: A autora (2023)

Os valores de resistência à tração e módulo de elasticidade coletados foram similares aos encontrados em literatura. XING *et al.*, (2020) produziram CNFs a partir da hidrólise ácida da polpa de eucalipto (razão de aspecto de 13,4) e a misturaram com PVA (98-99% de grau de hidrólise) para produzir filmes biodegradáveis. Para um teor de 4% em massa de CNF, a resistência à tração reportada a uma velocidade de separação de 10 mm/min foi de 109 MPa, em comparação ao valor de 76 MPa obtido para o filme de PVA puro. Já para o módulo de elasticidade, o valor obtido para o mesmo compósito foi de 6,836 GPa, enquanto que para o PVA puro foi de 4,220 GPa. As amostras haviam sido secas previamente ao experimento. UEDA *et al.*, (2022) produziram filmes com 10% em volume de CNF (tratado com 20 passes em um homogeneizador a ultra alta pressão) e PVA. A resistência à tração obtida a uma velocidade de 5 mm/min foi de 115 MPa e o módulo de elasticidade de 6 GPa. Em contrapartida, para o PVA puro, a resistência foi de 30 MPa e o módulo de elasticidade foi de 1,6 GPa.

Apesar da velocidade de separação das garras ter sido diferente entre os experimentos citados e o realizado pela mestrandia, além de outras condições de teste poderem ter sido variadas também, percebe-se que o comportamento mecânico do compósito e do filme de PVA puro produzidos nas dependências da UFPR foi de acordo com o esperado. Percebe-se que a inserção de nanocelulose na matriz polimérica tornou os filmes mais duros, porém auxiliou na melhora do desempenho mecânico. Por exemplo, a resistência à tração apresentou um aumento de 169%, enquanto que o módulo de elasticidade apresentou uma melhora de 206%, quando se comparam os valores do filme de PVA puro e PVA com CNF (4%). Este efeito de reforço ocasionado pelas nanoceluloses já é largamente descrito por outros pesquisadores e é causado pelas interações entre os grupos hidroxilas presentes nas cadeias dos dois materiais (NIAZI *et al.*, 2020; PERESIN *et al.*, 2010; YUWAWECH; WOOTHIKANOKKHAN; TANPICHAI, 2015). Este é um dos indícios que pode evidenciar a boa dispersão e interação existente entre estes dois materiais. Contudo, apesar da evidente melhora das propriedades mecânicas, também se observa um desvio padrão acentuado nas medições. Isto pode ser mais um sinal da presença de falhas e da heterogeneidade das membranas produzidas.

4.5 AVALIAÇÃO DOS DADOS DE PERMEAÇÃO PARA PREDIÇÃO DE PERMEANÇA E SELETIVIDADE DE MEMBRANAS

Como não foi possível obter com sucesso os dados experimentais no laboratório, foram utilizados somente dados extraídos da literatura para a geração de modelos de aprendizagem de máquina para predição de propriedades de permeação de CO₂ para o processamento de gás natural.

Os dados de permeação de gases extraídos da literatura consistiram basicamente em três tipos de membrana: PVA puro, PVA com CNF e PVA com CNC. Avaliando a tabela dos dados no APÊNDICE B, observa-se que foram consideradas 5 variáveis iniciais para a modelagem: a razão de aspecto das nanoceluloses, o pH da solução mãe de membrana, a concentração de nanocelulose na membrana, a pressão de operação e a umidade relativa dos gases a serem permeados. Desta maneira, as variáveis consideradas abrangeram parâmetros de processo (pressão e umidade) e de material (razão de aspecto, pH da solução mãe e concentração de nanocelulose).

Como os dados advieram de artigos publicados, houve uma limitação na quantidade e nos tipos de variáveis que foram possíveis de serem utilizadas para a modelagem. Neste caso, a razão de aspecto contou com somente três valores diferentes, sendo 0 para o caso de membranas de PVA puro; o pH da solução mãe variou de 5 a 12, sendo a maioria das membranas produzidas a partir de uma solução de pH 9; a concentração de nanocelulose variou de 0% (no caso de PVA puro) até 4%; as pressões de operação variaram entre 5, 10 e 15 bar, sendo praticamente metade dos pontos coletados a 5 bar; e a umidade relativa variou de 30 a 100%, sendo a maioria dos pontos coletados a 90% de umidade para cima.

Já em relação às variáveis de saída, especificamente permeança e seletividade, observa-se que o menor valor registrado de permeança foi de 0,0125 m³ (STP).m⁻².h⁻¹.bar⁻¹ e o maior de 0,29 m³ (STP).m⁻².h⁻¹.bar⁻¹. A seletividade possui uma ordem de magnitude maior, registrando-se valores de 2,5 até 45. Todas estas observações podem ser visualizadas facilmente na TABELA 22, onde está a descrição estatística das variáveis de entrada e saída para os modelos de predição das propriedades de permeação de gases.

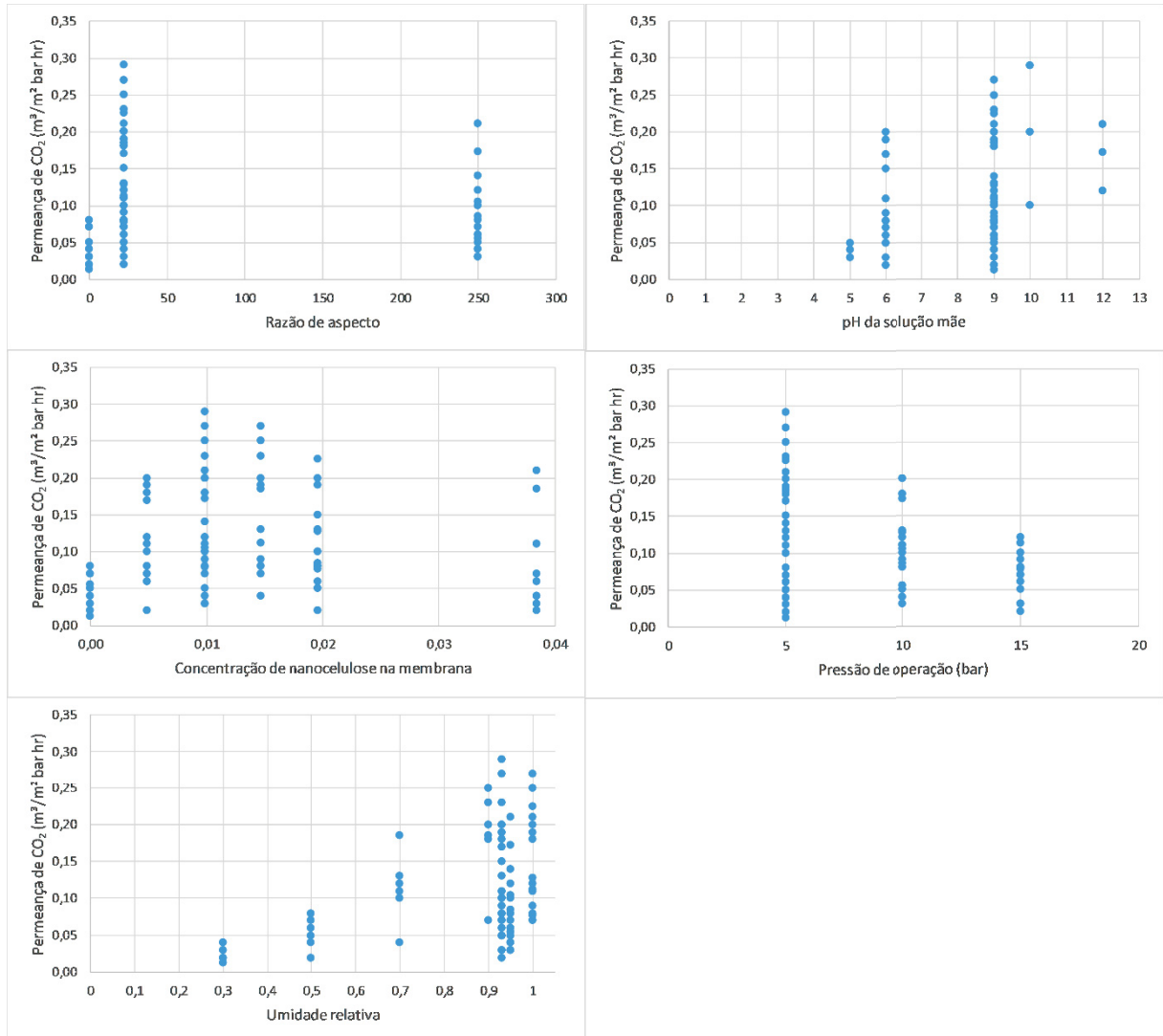
TABELA 22 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E VARIÁVEIS DE SAÍDA PARA MODELOS DE PREDIÇÃO DE PROPRIEDADES DE PERMEAÇÃO DE GASES.

Variável	Valores não nulos	Valor médio	Desvio padrão	Valor mínimo	25%	50%	75%	Valor máximo
Razão de aspecto (L/D)	91	64,46	92,91	0	22,03	22,03	22,03	250,00
pH da solução mãe	91	8,5	1,5	5,0	9,0	9,0	9,0	12,0
Concentração de nanocelulose	91	0,0132	0,0109	0,0000	0,0049	0,0099	0,0196	0,0384
Pressão de operação (bar)	91	8,6	4,1	5,0	5,0	5,0	10,0	15,0
Umidade relativa	91	0,86	0,20	0,30	0,90	0,93	0,95	1,00
Permeança (m ³ /m ² bar hr)	91	0,1102	0,0706	0,0125	0,0575	0,0900	0,1762	0,2900
Seletividade	91	23,4	10,1	2,5	17,2	24,0	30,5	45,0

Fonte: A autora (2023)

A FIGURA 11 mostra os gráficos de dispersão dos valores de permeança do CO₂ em função das variáveis de entrada consideradas.

FIGURA 11 - GRÁFICOS DE DISPERSÃO DA PERMEANÇA EM FUNÇÃO DAS VARIÁVEIS DE ENTRADA



Fonte: A autora (2023)

Analisando os gráficos, constata-se que, apesar de não haverem claras tendências lineares ou de alguma função específica que relacione cada variável de entrada com a de saída, é possível ainda perceber tendências de aumento ou diminuição da permeância do CO₂, especialmente com as variáveis de processo. Por exemplo, verifica-se que com o aumento da umidade relativa, há um aumento na faixa de valores de permeância observada. Esta observação é descrita na literatura e atribuída à afinidade tanto do PVA quanto das nanoceluloses pela água. Esses materiais se ligam facilmente a ela, que por sua vez, interage com a molécula de CO₂, transformando-a em um íon bicarbonato. Este íon aparentemente é transportado mais

facilmente do que o gás metano, assim, aumentando a transferência de massa de CO₂ através da membrana.

Com relação à pressão de operação, observa-se o contrário: quanto maior, menor fica a faixa de permeação observada. A literatura traz indícios de que este fenômeno pode ocorrer devido a uma maior compressão da membrana, resultando em perda de água e adensamento das cadeias poliméricas, o que dificulta a difusão dos gases pela membrana. Ademais, este fenômeno também pode ocorrer por efeitos de plastificação e saturação dos carreadores a pressões maiores (JAHAN *et al.*, 2018b).

Avaliando a concentração de nanocelulose e sua possível influência na permeação, percebe-se que as maiores permeações foram observadas para concentrações de nanocelulose de 1%. De fato, esta influência foi investigada nos artigos de Jahan *et al.* (2018a; 2018b; 2020), e observou-se que a concentração de 1% de nanocelulose era a ideal, dentre as avaliadas. Isto porque, nesta condição, há uma quantidade suficiente de nanopartículas que auxiliam na distribuição das moléculas de água na matriz, sem que ocorra um aumento significativo da cristalinidade da matriz ou que o reforço mecânico exercido pela nanocelulose atrapalhem a absorção de água pela membrana. Como a quantidade de água está diretamente ligada a um maior transporte de CO₂, o seu transporte é dificultado.

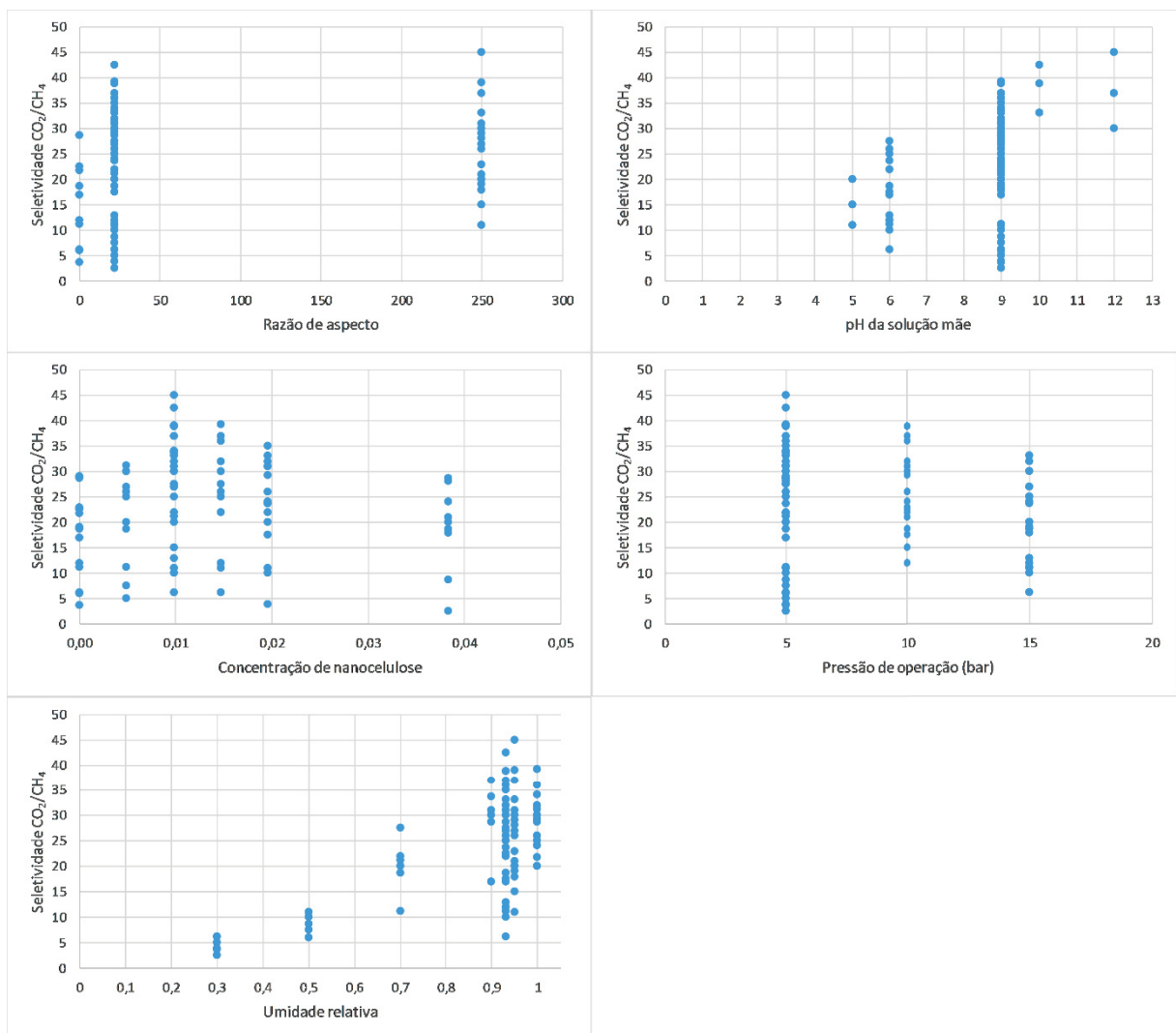
Sobre a influência do pH da solução mãe de membrana, há indícios de que a pHs maiores ocorre um favorecimento da permeação e seletividade. Isto, em realidade, parece estar atrelado a uma maior capacidade de absorção de água para membranas que foram produzidas com soluções a pHs maiores. A hipótese é de que isso ocorre por conta da diferença do conteúdo iônico (proveniente do NaOH) entre os filmes. Ademais, a pHs maiores, há um maior número de prótons livres disponíveis que podem reagir com o dióxido de carbono, facilitando a ligação com a água e a geração do íon bicarbonato (JAHAN *et al.*, 2020).

Por fim, como a razão de aspecto foi uma variável que só teve três valores diferentes reportados, não é possível estabelecer claras tendências com a permeação de CO₂. Contudo, é possível constatar que, para a razão de aspecto de 22 associado aos nanocristais de celulose, foi possível observar permeações maiores. Torstensen *et al.*, (2019) estudaram a permeação de CO₂ em membranas de PVA, CNF de baixa carga, CNF de alta carga e CNC, e observaram justamente um aumento na permeabilidade na presença de partículas menores ou com carga. A hipótese é de

que partículas com estas características conseguem se dispersar melhor pela matriz polimérica e distribuir melhor a água através do filme, facilitando o transporte de massa do CO_2 .

A FIGURA 12 mostra os gráficos de dispersão dos valores de seletividade em função das mesmas variáveis de entrada descritas nos parágrafos anteriores.

FIGURA 12 - GRÁFICOS DE DISPERSÃO DA SELETIVIDADE EM FUNÇÃO DAS VARIÁVEIS DE ENTRADA



Fonte: A autora (2023)

Da mesma forma comentada para a permeança, neste caso também não existem claras relações funcionais entre as variáveis e a seletividade. Contudo, é possível observar algumas tendências de aumento ou decréscimo para alguns casos específicos. Por exemplo, aqui também se vê uma influência da umidade

relativa. Conforme a umidade relativa do gás aumenta, valores cada vez maiores de seletividade são observados. No caso da pressão, as maiores seletividades são observadas a 5 bar. Parece haver uma redução da seletividade com o aumento da pressão, mas não tão clara quanto constatado para a permeança.

A respeito da concentração de nanocelulose, não há uma clara tendência de aumento ou redução da seletividade conforme a concentração de nanopartícula aumentou ou diminuiu na membrana. No entanto, é possível visualizar que a maior seletividade é observada para 1% de nanocelulose. Assim, nesta condição alcançou-se não só a maior permeança registrada para o CO₂, como também o seu maior favorecimento, o que aumentou a sua seletividade. Similarmente, observam-se as maiores seletividades para o caso de pHs maiores, pelas razões expostas previamente. A razão de aspecto não demonstra nenhuma relação significativa de influência na determinação da seletividade.

Os valores dos coeficientes de correlação de Pearson entre as variáveis de entrada e a permeança e seletividade estão elencados, respectivamente na TABELA 23 e TABELA 24.

TABELA 23 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL PERMEANÇA.

Variável de entrada	Coefficiente de correlação de Pearson
Umidade relativa	0,3923
Pressão de operação (bar)	-0,3528
pH da solução mãe	0,2476
Razão de aspecto (L/D)	-0,1680
Concentração de nanocelulose	0,0665

Fonte: A autora (2023)

TABELA 24 - VALORES DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON ENTRE AS VARIÁVEIS DE ENTRADA E A VARIÁVEL SELETIVIDADE.

Variável de entrada	Coefficiente de correlação de Pearson
Umidade relativa	0,6605
pH da solução mãe	0,3971
Razão de aspecto (L/D)	0,1608
Pressão de operação (bar)	-0,0666
Concentração de nanocelulose	0,0170

Fonte: A autora (2023)

Pela observação da TABELA 23, percebe-se que não há correlações lineares entre as variáveis de entrada e a variável de predição “permeança”. Isto já havia sido constatado ao se avaliarem os gráficos de dispersão na FIGURA 11. Ainda assim, os maiores valores de coeficiente estão atribuídos à umidade relativa e à pressão de operação, respectivamente. Isto está em consonância com o que havia sido verificado na análise dos gráficos de dispersão para esta variável. Além disso, para a umidade relativa tem-se uma influência positiva, enquanto que a pressão de operação exerce uma influência negativa. Isto também está refletido no sinal dos coeficientes, que foram positivo e negativo, respectivamente.

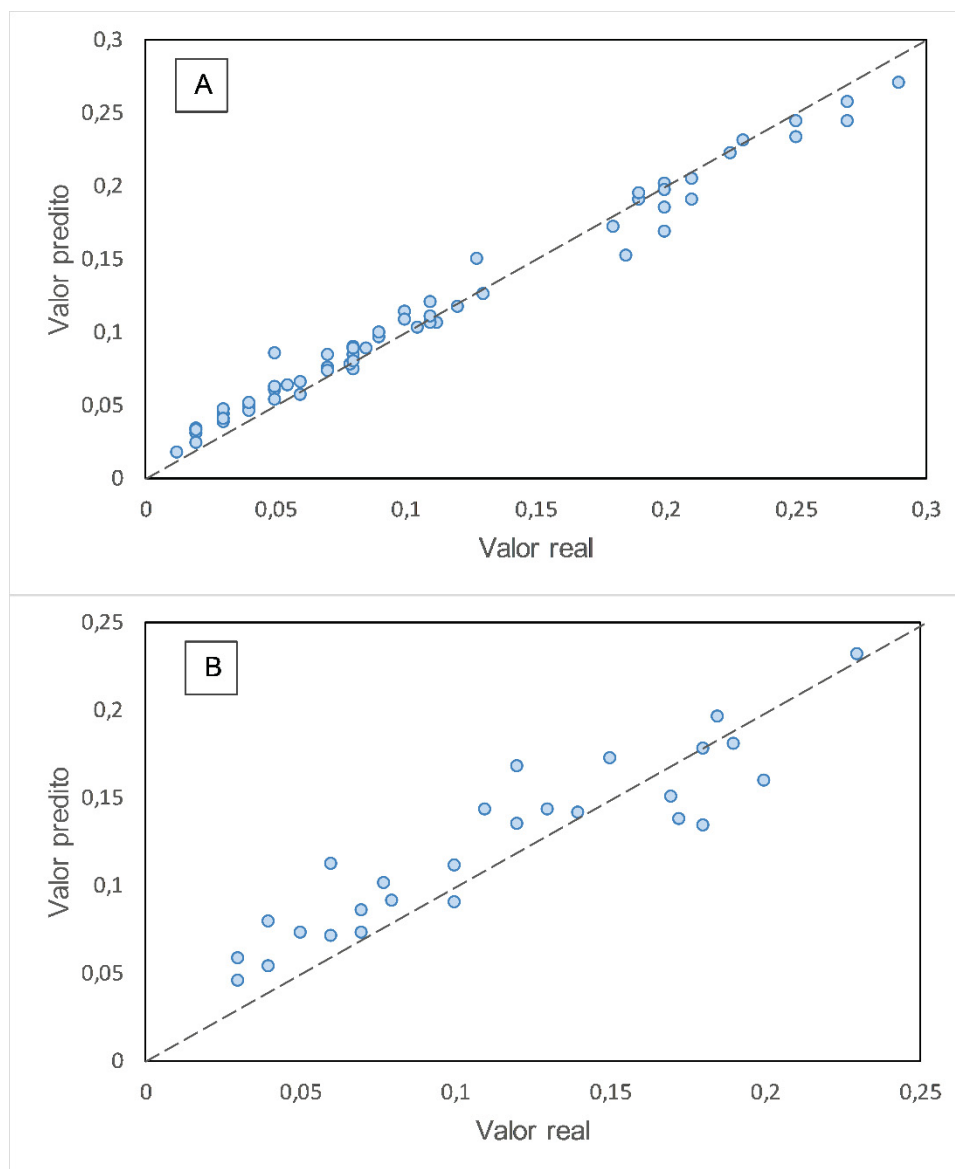
A respeito da TABELA 24, a umidade relativa também foi a variável com maior valor de coeficiente de Pearson, e pela análise dos gráficos da FIGURA 12, verifica-se que esta variável apresenta maior relação com a descrição da seletividade, sendo uma influência positiva. O restante das variáveis não apresentou coeficientes de Pearson altos.

4.6 MODELOS DE PREDIÇÃO: PERMEANÇA AO CO₂

4.6.1 Florestas randômicas

A partir dos dados apresentados e aplicando a função de otimização Bayesiana, o modelo final gerado fez uso dos seguintes valores para os hiperparâmetros otimizados: profundidade máxima das árvores = 25, quantidade máxima de features para fazer a divisão dos nós = 2, número de árvores = 117 e número mínimo de pontos nas folhas = 1. Os valores preditos foram plotados contra os valores reais e estão apresentados no GRÁFICO 18. As métricas de desempenho do modelo estão dispostas na TABELA 25.

GRÁFICO 18 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE PERMEANÇA DE CO₂.



Fonte: A autora (2023)

TABELA 25 - MÉTRICAS DE DESEMPENHO PARA OS SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA PERMEANÇA DE CO₂.

Métrica	Valor para treino	Valor para teste
EAM	0,0092	0,02049
EPAM	15,05%	26,57%
RQEM	0,0120	0,0252
R ²	0,9742	0,8120

Fonte: A autora (2023)

Avaliando o gráfico A do GRÁFICO 18 e as métricas apresentadas na TABELA 25, constata-se que foi possível obter um resultado excelente para o subconjunto de treino. A maioria dos pontos está bem próxima da reta de 45°, evidenciando a proximidade dos valores reais e preditos. Alcançou-se um R² muito próximo de 1, no valor de 0,9742, e erros médios relativamente baixos no geral, especialmente se consideramos a faixa de valores das permeanças, que variou de 0,0125 m³(STP).m⁻².bar⁻¹.h⁻¹ até 0,29 m³(STP).m⁻².bar⁻¹.h⁻¹. O EPAM, de aproximadamente 15%, foi um pouco elevado, provavelmente sendo afetado pelo fato de a ordem de grandeza da propriedade ser pequena. Quer-se dizer que pequenas variações absolutas no valor podem ocasionar variações grandes de erro percentual, principalmente para aqueles próximos a zero.

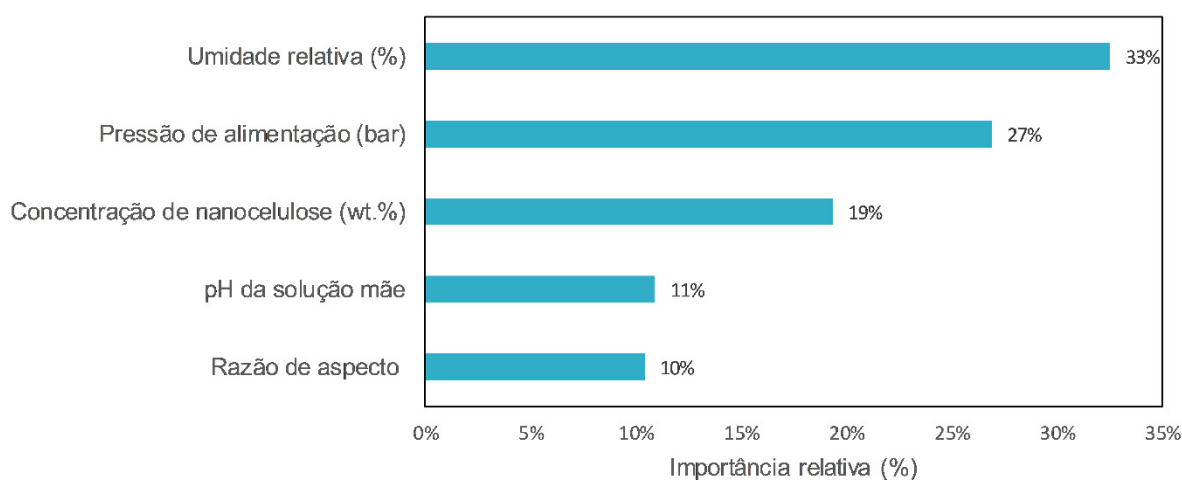
Quando avaliado o modelo para dados não vistos pelo algoritmo, ou seja, para o subconjunto de teste, observa-se que houve capacidade de predição satisfatória, apesar de não estar perfeita. O gráfico B do GRÁFICO 18 mostra que pontos no Pareto estão próximos da reta de 45°, porém um tanto esparsos ainda. O coeficiente de determinação obtido foi de 0,8120, um valor razoável. Os erros médios alcançados neste caso foram maiores em relação às métricas de treino, porém o modelo ainda é válido pois há poder de predição, apesar de não ser totalmente acurado. Pode-se dizer que há um grau de *overfitting*.

Neste caso, o desempenho razoável do modelo pode ser atribuído ao subconjunto de pontos utilizado e ao tipo de algoritmo empregado. É importante lembrar que há uma limitação na quantidade de dados utilizada para gerar este modelo, além das variáveis utilizadas também serem mais restritas, uma vez que foram utilizadas as informações que estavam disponíveis nos artigos. Provavelmente,

as métricas de teste possam ser melhoradas caso se obtenham mais valores experimentais e outras informações acerca do material, que possam descrever melhor a permeança de CO₂. Não obstante, apesar das limitações, o algoritmo conseguiu manter uma boa capacidade preditiva.

A importância relativa das variáveis utilizadas foi analisada para o modelo construído. Os valores correspondentes foram plotados no GRÁFICO 19.

GRÁFICO 19 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA PERMEANÇA DE CO₂ UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.



Fonte: A autora (2023)

Constata-se que umidade relativa foi a variável com maior importância na descrição da propriedade de interesse, correspondendo a 33% do total. Em segundo lugar aparece outra variável de processo, a pressão de alimentação. A razão de aspecto teve a menor importância, de 10%, mas ainda teve uma participação um tanto expressiva. Isto está em consonância com o que foi observado previamente na avaliação preliminar dos dados, na seção 4.5.

A umidade relativa é um parâmetro descrito na literatura bem relevante para membranas que possuem grande afinidade por água e em que o transporte de massa e separação de gases ocorre por transporte facilitado. A água atua como carreador da molécula de CO₂, favorecendo seu transporte em comparação com a molécula de metano. Quanto maior o teor de água na mistura gasosa a ser separada, mais água a membrana irá absorver e mais pontos de carreação existem para facilitar o transporte do CO₂ (DAI *et al.*, 2019). A pressão de alimentação também é uma variável de

processo muito relevante na separação de gases por membranas, uma vez que, levando em conta o mecanismo de solução-difusão, o fluxo de uma espécie química através de uma membrana polimérica é regido positivamente, dentre outros fatores, pela diferença de pressão entre os lados da membrana (ISMAILA *et al.*, 2005). Olhando-se pela perspectiva do transporte facilitado, há uma influência da pressão também, porém negativa. Primeiramente, os carreadores vão ficando saturados conforme aumenta-se a pressão, o que faz com que, a partir deste momento, somente o transporte por solução-difusão aumente com o aumento da pressão. Ademais, o aumento da pressão favorece a compressão destas membranas inchadas com água, resultando em perda de água e adensamento das cadeias poliméricas, o que dificulta a difusão dos gases (JAHAN *et al.*, 2018b).

À razão de aspecto da nanocelulose não foi atribuída a maior importância relativa. Apesar de ter influência sobre a permeação do dióxido de carbono, como discutido anteriormente, provavelmente aqui não esteve tão evidente esta contribuição por conta da limitação da quantidade de dados. Esta variável de entrada contou somente com três valores de razão de aspecto para treinamento do modelo: 0 para membranas de PVA puro, 22 para membranas com CNC e 250 para membranas com CNF. Como a maioria dos pontos correspondeu a CNC, ou seja, a maioria dos pontos continha valores de razão de aspecto de 22, não teve muita variação nos valores a ponto de ser um aspecto tão relevante para descrever a variável de saída de interesse. Por outro lado, para a umidade relativa foram reportados sete níveis diferentes, e assim, provavelmente o modelo utilizou mais esta variável para fazer a separação das amostras entre os diferentes valores de permeação.

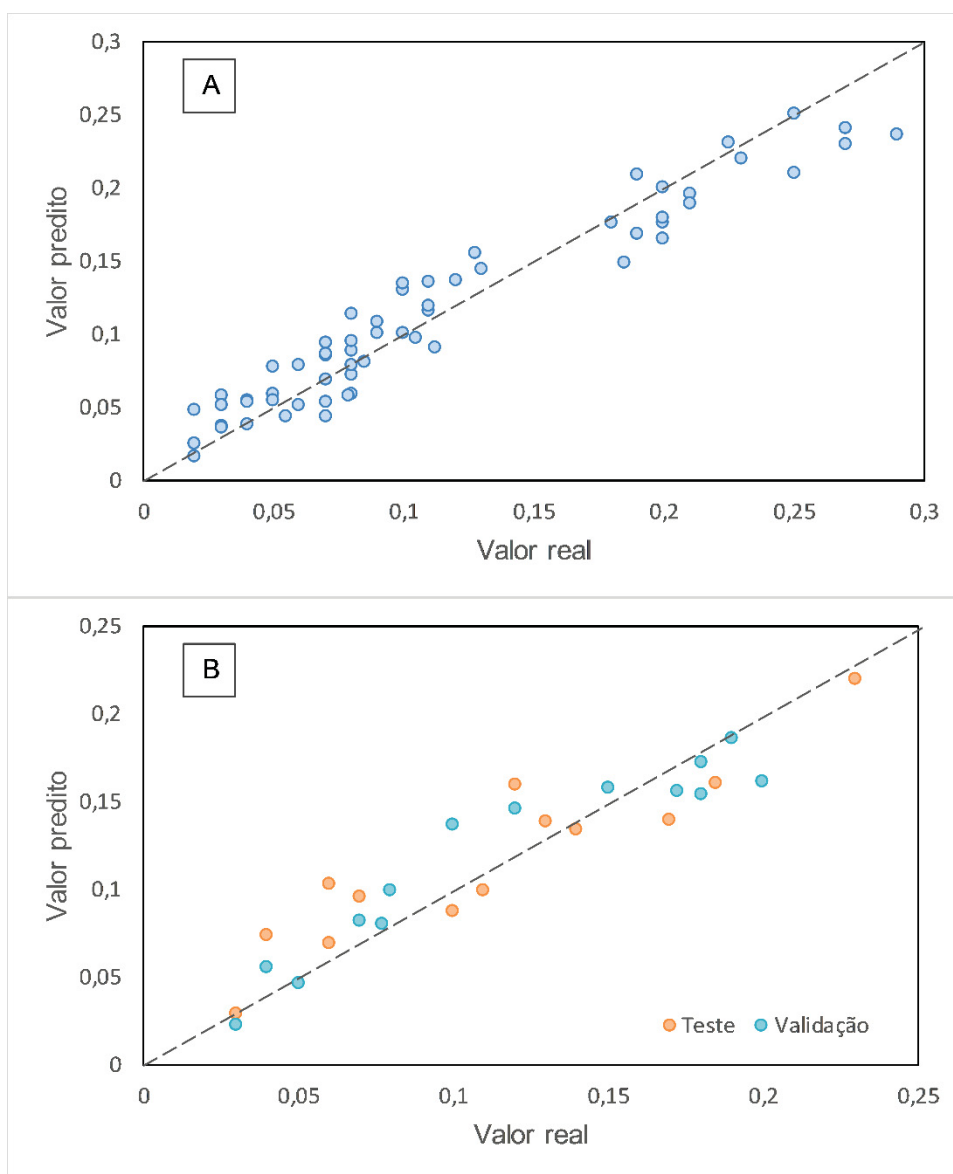
Portanto, seria interessante, para fins de melhoramento deste tipo de modelo preditivo, avaliar uma faixa maior de razão de aspecto das nanopartículas, além de estudar outras variáveis que possam interferir na propriedade de permeação de um gás, como por exemplo, a porosidade da membrana, carga superficial das nanopartículas, dentre outros.

4.6.2 Redes neurais

O modelo final gerado possuiu a seguinte arquitetura otimizada: uma camada oculta, 5 neurônios de entrada, 10 neurônios na camada oculta, um tamanho de lote de 6, uma taxa de aprendizagem de 0,013, aproximadamente, e função de ativação

ReLU. Os valores preditos foram plotados contra os valores reais e estão apresentados no GRÁFICO 20, tanto para o subconjunto de dados de treino quanto para o de teste. As métricas de desempenho do modelo estão dispostas na TABELA 26.

GRÁFICO 20 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE REDE NEURAL FINAL PARA PREDIÇÃO DE PERMEANÇA DE CO₂.



Fonte: A autora (2023)

TABELA 26 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE REDE NEURAL PARA PREDIÇÃO DA PERMEANÇA DE CO₂.

Métrica	Valor para treino	Valor para teste
EAM	0,0166	0,0179
EPAM	23,87	20,24
RQEM	0,0202	0,0218
R ²	0,9266	0,8591

Fonte: A autora (2023)

Da arquitetura obtida, constata-se que, dentro das iterações realizadas, foram necessários o máximo de neurônios contido na faixa de estudo. O tamanho do lote não foi alto e o valor da taxa de aprendizagem se manteve próxima de 0,01, que é um valor bastante usual de se utilizar na construção de redes neurais. A função de ativação que gerou o melhor resultado foi o ReLU.

O GRÁFICO 20 mostra que os pontos de treino não ficaram muito distantes da reta de 45°. As métricas de desempenho na TABELA 26 apresentam um R² de 0,9266, um EAM de 0,0166, EPAM de 23,87% e RQEM de 0,0202. Portanto, verifica-se que a rede neural se adequou bem aos dados de treino de permeança. Contudo, comparando-se com o modelo de floresta randômica, a capacidade preditiva foi um pouco pior.

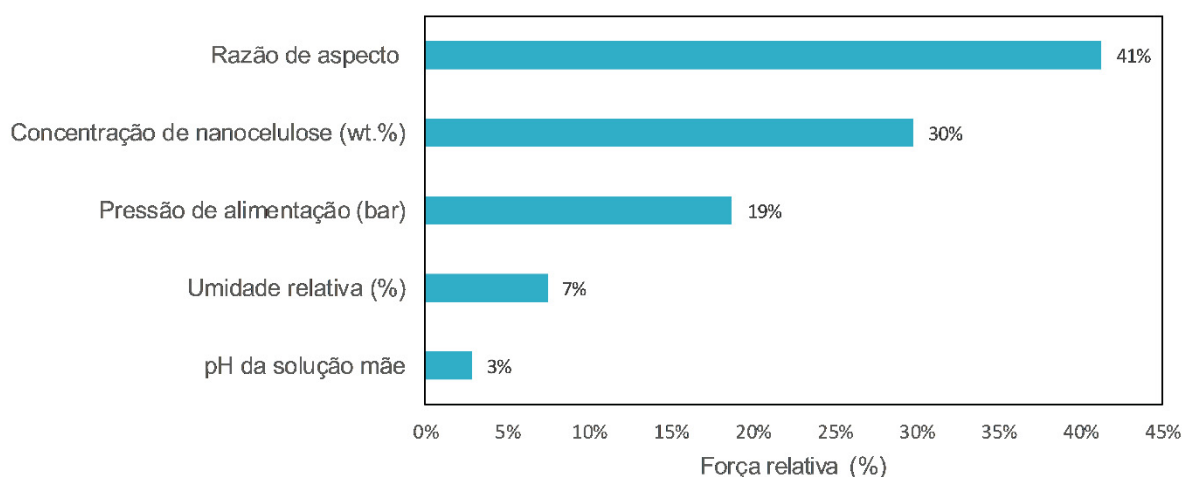
No entanto, ao se analisarem as métricas de desempenho para o subconjunto de teste, percebe-se que a rede neural conseguiu predizer melhor os dados não vistos pelo modelo. Neste caso, obteve-se um R² de 0,8591, valor maior do que o obtido para a floresta randômica. Os erros médios foram menores também. Para a rede neural, alcançou-se um EAM de 0,0179, um EPAM de 20,24% e um RQEM de 0,0218. Visualmente, isto também pode ser percebido pela observação da parte B do GRÁFICO 20, em que se percebe que os pontos estão um pouco mais próximos à reta de 45° em comparação com a parte B do GRÁFICO 18 para as florestas randômicas. Assim, aparentemente, no caso da rede neural ocorre uma menor tendência ao *overfitting* e uma melhor predição para novos dados.

É interessante perceber que, mesmo a rede neural tendo um melhor desempenho para o subconjunto de teste e as métricas de treino e teste estarem mais próximas, o EPAM ainda foi relativamente alto. Desta maneira, percebe-se que, para

os algoritmos estudados na faixa de hiperparâmetros avaliados, pode-se chegar até certo ponto em relação ao alcance de um EPAM mais baixo. Como mencionado na seção anterior, o fato de a ordem de grandeza da propriedade ser pequena deve estar afetando fortemente a variação percentual para os pontos preditos que ficaram mais distantes dos seus valores reais.

A força relativa das variáveis de entrada em relação à variável de saída foi calculada e plotada no GRÁFICO 21.

GRÁFICO 21 - FORÇA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA PERMEANÇA DE CO₂ UTILIZANDO O ALGORITMO DE REDES NEURAIAS.



Fonte: A autora (2023)

A força relativa das variáveis de entrada mudou consideravelmente em comparação com a importância relativa das mesmas variáveis usando o algoritmo de floresta randômica. Agora, a razão de aspecto das nanoceluloses desempenhou o papel de variável mais forte, seguida da concentração de nanocelulose nas membranas. Respectivamente, apresentaram uma força relativa de 41 e 30%. Já a pressão de alimentação ficou em terceiro lugar, com 19%. Curiosamente, a umidade relativa não desempenhou um papel significativo, como o encontrado para a floresta randômica. Sua força relativa foi de 7% somente.

É importante ressaltar que essa diferença de importância relativa das variáveis entre os algoritmos estudados pode ser esperada. Isto porque tratam-se de métodos de cálculo diferentes e de algoritmos diferentes. A floresta randômica é um algoritmo baseado na geração de árvores de decisão. Neste tipo de modelo, a

importância da variável está atrelada em quanto ela consegue reduzir o erro durante as divisões de nós. Neste caso, em geral, variáveis que possuem valores mais diferentes entre si, são preferidas pelo algoritmo para fazer a divisão pois dessa maneira conseguem separar melhor as amostras. Já em modelos de regressão como a rede neural, em que as variáveis são multiplicadas por pesos, as variáveis em que uma pequena mudança no seu valor gera grandes mudanças na predição são favorecidas.

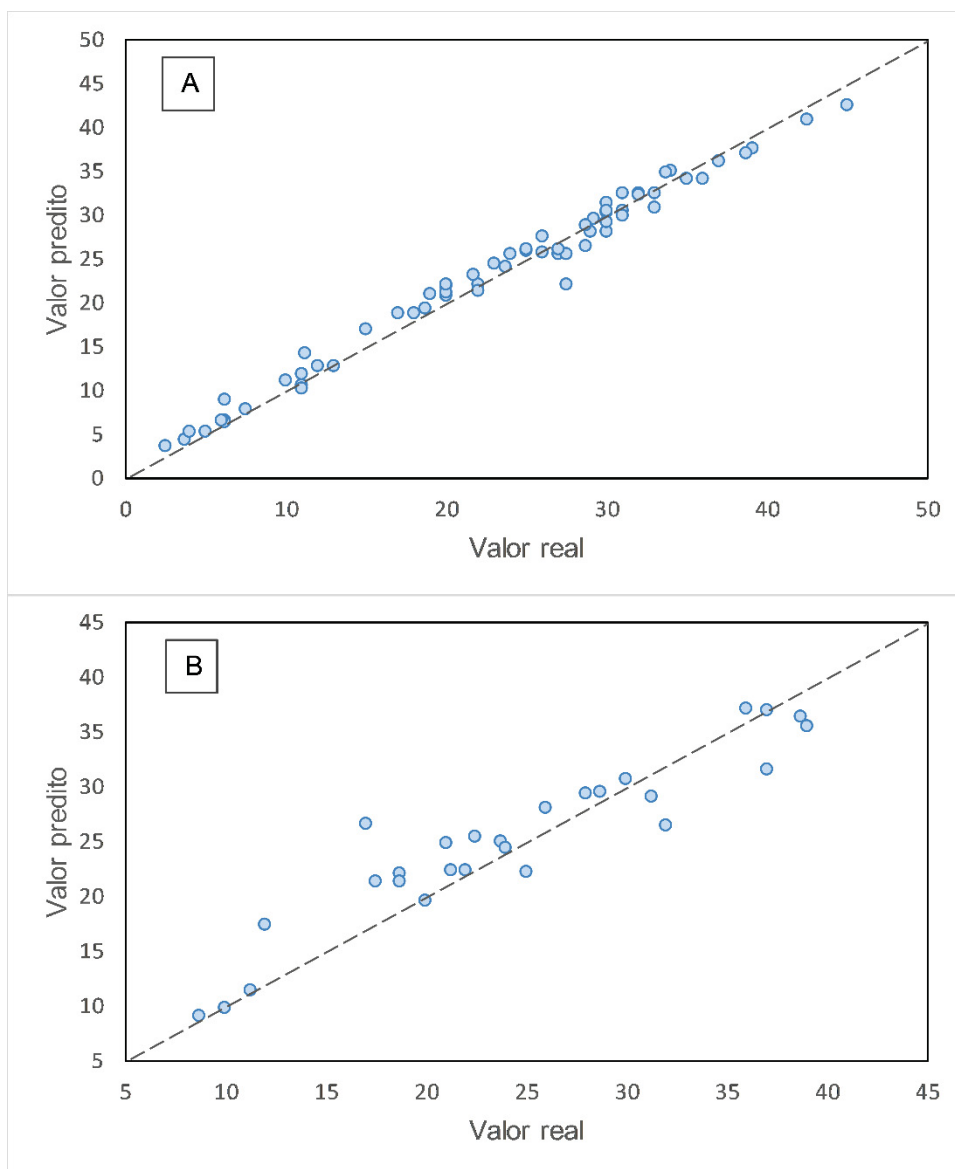
Desta maneira, para a rede neural, a razão de aspecto deve ter sido responsável por causar maiores mudanças na predição, do ponto de vista do modelo. Já para o modelo de floresta, a maior importância foi atribuída à variável com maior número de valores distintos, a umidade relativa.

4.7 MODELOS DE PREDIÇÃO: SELETIVIDADE CO₂/CH₄

4.7.1 Florestas randômicas

A partir dos dados apresentados e aplicando a função de otimização Bayesiana, o modelo final gerado fez uso dos seguintes valores para os hiperparâmetros otimizados: profundidade máxima das árvores = 100, quantidade máxima de features para fazer a divisão dos nós = 4, número de árvores = 400 e número mínimo de pontos nas folhas = 1. Os valores preditos foram plotados contra os valores reais e estão apresentados no GRÁFICO 22. As métricas de desempenho do modelo estão dispostas na TABELA 27.

GRÁFICO 22 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE FLORESTA RANDÔMICA FINAL PARA PREDIÇÃO DE SELETIVIDADE DE CO₂.



Fonte: A autora (2023)

TABELA 27 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA SELETIVIDADE DE CO₂.

Métrica	Valor para treino	Valor para teste
EAM	1,16	2,36
EPAM	7,08%	10,98%
RQEM	1,46	3,23
R ²	0,9809	0,8621

Fonte: A autora (2023)

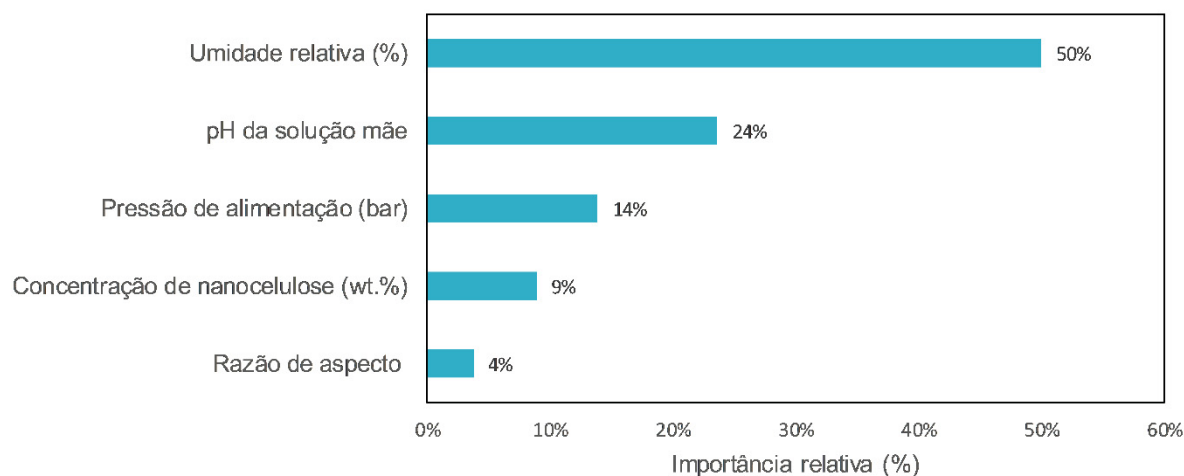
Observando o GRÁFICO 22, constata-se que o algoritmo de floresta randômica se adequou perfeitamente aos dados de treino do modelo, considerando que os pontos estão muito próximos à reta de 45°. Ademais, ao se avaliar as métricas de desempenho, também se percebe a excelente predição deste subconjunto: foi obtido um R² de 0,9809, um EAM de 1,16, EPAM menor que 10%, especificamente no valor de 7,08% e um RQEM de 1,46.

Já para os dados de teste, avaliando-se o gráfico B do GRÁFICO 22, percebe-se que, similarmente ao que ocorreu com os dados de permeância, também não houve a mesma capacidade excelente de predição constatada para o treino. Ademais, apesar do valor do R² para o subconjunto de teste ter sido 0,8621, ou seja, não tão próximo do R² obtido para o treino, o restante das métricas alcançadas foram muito próximas. O EAM foi de 2,36 unidades, o EPAM foi um pouco acima de 10%, no valor de 10,98% e o RQEM foi de 3,23. Estes erros não foram muito altos.

Evidencia-se então, que para os dados de seletividade, a predição utilizando os dados de literatura e o algoritmo de floresta randômica foi mais fácil. Isto também pode ser atribuído à ordem de grandeza dessa variável, que não é tão pequena quanto a permeância. Desta maneira, não sofre tanto, em termos de erros percentuais, quando os valores preditos ficam um pouco distantes dos reais.

A importância relativa das variáveis utilizadas foi analisada para o modelo construído. Os valores correspondentes foram plotados no GRÁFICO 23.

GRÁFICO 23 - IMPORTÂNCIA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA SELETIVIDADE DE CO₂ UTILIZANDO O ALGORITMO DE FLORESTA RANDÔMICA.



Fonte: A autora (2023)

Similarmente ao que ocorre para o modelo de floresta randômica da permeança, a umidade relativa aparece como a variável mais importante para a previsão da seletividade. Neste caso particular, o valor da importância é ainda mais alto, de 50%. O restante das variáveis não se mantém na mesma ordem, entretanto. O pH da solução mãe apresenta uma influência maior, até mais significativa que a pressão de alimentação, que aparece em terceiro lugar. Para a seletividade, a razão de aspecto não apresentou uma importância muito relevante, neste caso de somente 4%.

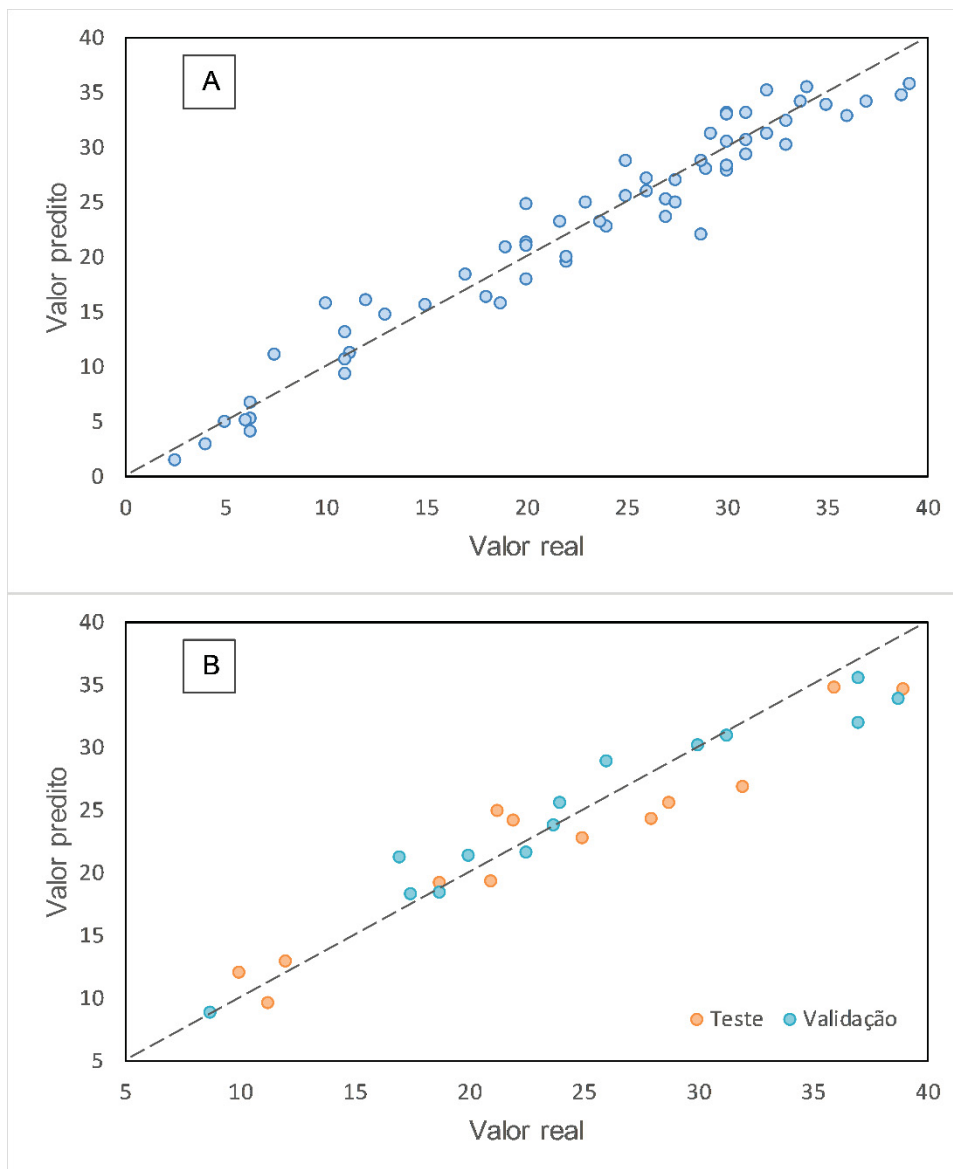
Ao analisar as duas variáveis mais importantes, o resultado apresentado é coerente com o que é descrito na literatura. Como discutido anteriormente, a umidade relativa do gás está diretamente atrelada à quantidade de água disponível para o filme absorver, enquanto o pH da solução mãe influencia em quanto de água o mesmo consegue absorver. Assim, é de se esperar que estes dois parâmetros tenham uma relevância grande na seletividade, mais até do que parâmetros como pressão de alimentação, que vai exercer uma influência mais ou menos parecida para a transferência de massa dos dois componentes gasosos, já que está relacionado com o transporte por solução-difusão. Curiosamente, a razão de aspecto não apresentou uma importância significativa, apesar de estar atrelada à questão da absorção e dispersão de água pela matriz. Talvez, isto esteja relacionado ao fato desta variável

somente apresentar três níveis diferentes de valores, como mencionado na seção anterior.

4.7.2 Redes neurais

O modelo final gerado possuiu a seguinte arquitetura otimizada: uma camada oculta, 5 neurônios de entrada, 10 neurônios na camada oculta, um tamanho de lote de 3, uma taxa de aprendizagem de 0,1 e função de ativação sigmoideal. Os valores preditos foram plotados contra os valores reais e estão apresentados no GRÁFICO 24, tanto para o subconjunto de dados de treino quanto para o de teste. As métricas de desempenho do modelo estão dispostas na TABELA 28.

GRÁFICO 24 - VALOR PREDITO VERSUS VALOR REAL PARA OS SUBCONJUNTOS DE TREINO (A) E TESTE (B) DO MODELO DE REDE NEURAL PARA PREDIÇÃO DE SELETIVIDADE DE CO₂.



Fonte: A autora (2023)

TABELA 28 - MÉTRICAS DE DESEMPENHO PARA SUBCONJUNTOS DE TREINO E TESTE DO MODELO DE REDE NEURAL PARA PREDIÇÃO DA SELETIVIDADE DE CO₂.

Métrica	Valor para treino	Valor para teste
EAM	2,100	2,096
EPAM	12,68%	8,78%
RQEM	2,704	2,651
R ²	0,9347	0,9069

Fonte: A autora (2023)

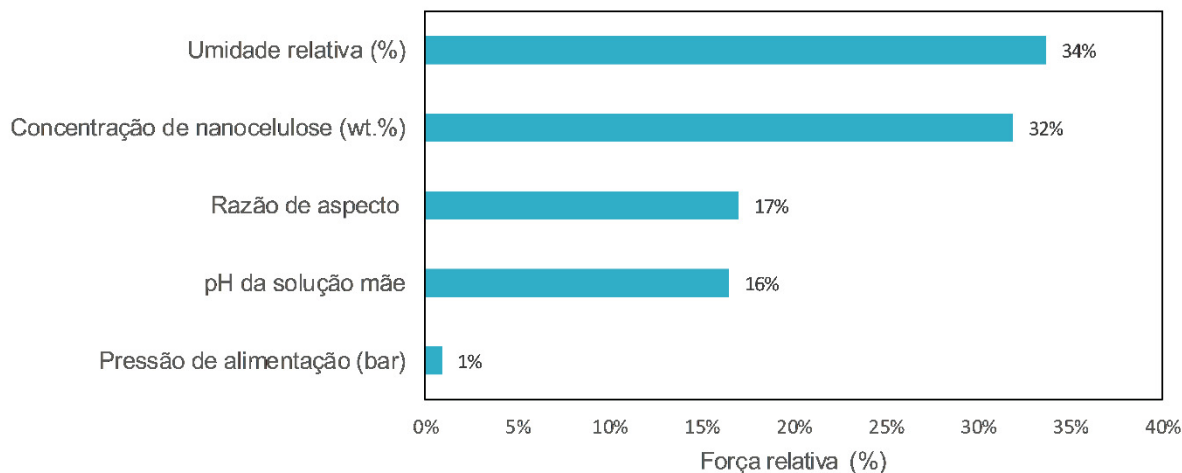
Analisando primeiramente a arquitetura da rede neural gerada, percebe-se que foram requeridos o número máximo de neurônios na camada oculta, além de ter sido necessário uma taxa de aprendizagem de 0,1. A função de ativação que melhor se adequou para este caso foi a sigmoidal. O tamanho do lote foi pequeno, similarmente ao caso da permeança, no valor de 3.

Pelo GRÁFICO 24, observa-se que tanto os pontos de treino quanto os de teste ficaram bem próximos da reta de 45°. Contudo, comparativamente com o algoritmo de floresta randômica, o subconjunto de treino não se ajustou tão perfeitamente, apesar de ainda ter uma capacidade preditiva muito boa. Isto pode ser verificado pelas métricas de desempenho alcançadas para este subconjunto. Por exemplo, o R^2 obtido foi de 0,9347, um valor muito satisfatório, enquanto que o EAM foi de 2,100, EPAM de 12,68% e o RQEM de 2,704, ou seja, erros relativamente baixos.

No entanto, foi obtido um desempenho melhor para predição de dados não vistos quando se compara a rede neural com a floresta randômica previamente obtida. Por exemplo, o R^2 alcançado foi de 0,9069, enquanto o EAM foi de 2,096, o EPAM de 8,78% e o RQEM de 2,651, todos valores melhores do que as métricas para o teste presentes na TABELA 27. Ademais, também se observa que as métricas de treino e teste para a rede neural ficaram ainda mais próximas, incluindo o R^2 dessa vez, aparentando um risco de *overfitting* menor ainda do que o apresentado pela floresta randômica.

A força relativa das variáveis de entrada em relação à variável de saída foi calculada e plotada no GRÁFICO 25.

GRÁFICO 25 - FORÇA RELATIVA DAS VARIÁVEIS PARA PREDIÇÃO DA SELETIVIDADE DE CO₂ UTILIZANDO O ALGORITMO DE REDES NEURAIS.



Fonte: A autora (2023)

Percebe-se que a umidade relativa também é a variável com maior força para a rede neural. Entretanto, o restante das variáveis está elencado em uma ordem diferente da apresentada para a floresta. Por exemplo, agora a concentração de nanocelulose está em segundo lugar como variável mais importante, possuindo uma força relativa de 32%. Em seguida, tem-se a razão de aspecto, com 17%, o pH da solução mãe, com 16% e, por fim, a pressão de alimentação com somente 1 % de força relativa.

Neste caso, ainda se observa uma prevalência das variáveis que possuem um impacto na quantidade de água absorvida pelas membranas, enquanto que a pressão de alimentação não teria tanto impacto na seletividade, uma vez que modificaria a princípio em mesmo grau na permeância das duas espécies químicas estudadas.

5 CONCLUSÃO

O presente trabalho teve como objetivos caracterizar membranas de PVA e nanocelulose fibrilada para separação de CO₂ no processamento de gás natural e a modelagem de propriedades relevantes do material, especificamente da nanocelulose, e da permeação do dióxido de carbono.

A respeito das propriedades da nanocelulose, foram gerados modelos para predição da razão de aspecto, importante característica morfológica e que pode ter efeitos na determinação de propriedades mecânicas e de permeação, e do rendimento de nanofibrilação, um parâmetro que relaciona a porcentagem de fibras nanométricas obtidas pós etapa de fibrilação. Para isso, foram usados os algoritmos de floresta randômica e rede neural e aplicada a função Bayesiana para procurar as melhores combinações de hiperparâmetros dentro do número de iterações estudado. Nos dois casos de propriedades e dos dois algoritmos, foram obtidos resultados muito satisfatórios, com boas capacidades preditivas. Contudo, o rendimento de nanofibrilação se destacou quanto às métricas de desempenho e R² alcançados, especialmente para o modelo de rede neural, que desempenhou melhor que a floresta randômica respectiva. No caso, o modelo neural alcançou um R² para os dados de validação/teste de 0,943 e um erro absoluto médio de 0,9 pontos percentuais (para uma faixa de 2,9 a 31,27% dos dados de teste) e foram empregadas como variáveis de entrada a transmitância a 600 nm e demanda catiônica. Para a razão de aspecto, o melhor resultado obtido foi com o algoritmo de floresta randômica, em que o modelo final alcançou um R² de 0,919 e um erro absoluto médio de 9,05 para o subconjunto de teste (para uma faixa validada/testada de 44 a 220). As variáveis de entrada do modelo envolveram componentes químicos da matéria-prima, como teor de lignina e hemicelulose, bem como parâmetros de processo, por exemplo o tipo de pré-tratamento representado pela dosagem enzimática.

Com relação aos experimentos realizados nas dependências da UFPR, apesar de não se ter logrado resultados reprodutíveis que pudessem ser usados na complementação do banco de dados de permeação e seletividade do CO₂, foi possível tirar algumas conclusões, como a influência da metodologia escolhida para produção das membranas na propriedade de permeação e a não efetividade do uso da técnica de evaporação de solvente, sendo provavelmente melhor o uso da técnica deposição

de filme em suporte, além de se verificar experimentalmente a melhoria das propriedades mecânicas dos filmes quando da incorporação de nanocelulose.

A modelagem das propriedades de permeança e seletividade do CO₂ em relação ao CH₄ para membranas de PVA e CNF e CNC foi realizada com dados da literatura. Também foram estudados os algoritmos de floresta randômica e redes neurais para as duas propriedades e aplicada a função Bayesiana. No caso da permeança, o melhor modelo foi a rede neural. O modelo gerado apresentou um R² de 0,859 e um erro absoluto médio de 0,0179 para os dados de teste (para uma faixa validada/testada de 0,03 a 0,23 m³(STP).m⁻². bar⁻¹.h⁻¹). Para a seletividade ocorreu de forma similar, sendo o melhor modelo o de rede neural também, que apresentou um R² de 0,907 e um erro absoluto médio de 2,096 para o mesmo subconjunto de dados (para uma faixa validada/testada de 8,75 a 39). A seletividade foi a propriedade relacionada à membrana mais fácil de modelar, possivelmente por conta de a ordem de grandeza ser maior que a permeança.

Constatou-se de que é viável aplicar técnicas de inteligência artificial na predição de propriedades relevantes de materiais e membranas, gerando modelos com capacidades preditivas muito boas mesmo na presença de uma quantidade limitada de dados experimentais. A aplicação da função Bayesiana permitiu a obtenção de bons modelos de predição. Além disso, proporcionou a avaliação de um número maior de hiperparâmetros e em uma faixa mais extensa de valores, em comparação com o uso da procura manual empregada na geração dos modelos preliminares, não dispendendo muito trabalho para isso. Ou seja, é uma forma mais produtiva de buscar novas combinações de hiperparâmetros.

Ademais, foi possível também compreender a influência, a nível preditivo, das variáveis de entrada estudadas na predição da variável de saída de interesse, além de ter sido viável utilizar a função embutida das florestas randômicas sobre a importância relativa das variáveis para guiar a redução do número de entradas e simplificação dos modelos. Estes modelos gerados bem como o procedimento de otimização adotado constituem uma importante contribuição deste estudo na direção da utilização da aprendizagem de máquina no desenvolvimento e aplicação de materiais mais sustentáveis, especificamente da nanocelulose, em processos industriais críticos, como o processamento de gás natural.

6 SUGESTÕES PARA TRABALHOS FUTUROS

As sugestões para trabalhos futuros estão elencadas abaixo:

- Produção de membranas de PVA e nanocelulose através da técnica de deposição de filme em suporte;
- Utilização de técnicas de geração de dados sintéticos para complementação de banco de dados;
- Criação de algoritmos que conectem e relacionem modelos de predição de propriedades de material com as propriedades de membrana;
- Realizar experimentos que investiguem melhor as possíveis interferências de outras propriedades da nanocelulose e de membranas com as propriedades de permeação e seletividade. Por exemplo, razão de aspecto, carga superficial, composição química da nanocelulose/polpa original, porosidade da membrana, dentre outros.

REFERÊNCIAS

ABDULLAH, N. *et al.* Membranes and Membrane Processes: Fundamentals. *In*: BASILE, A.; MOZIA, S.; MOLINARI, R. (org.). **Current Trends and Future Developments on (Bio-) Membranes: Photocatalytic Membranes and Photocatalytic Membrane Reactors**. [S. l.]: Elsevier, 2018. p. 45–70.

AGÊNCIA NACIONAL DE PETROLÉO GÁS NATURAL E BIOCOMBUSTÍVEIS. **RESOLUÇÃO ANP Nº 16, DE 17 DE JUNHO DE 2008**. Estabelece a especificação do gás natural, nacional ou importador, a ser comercializado em todo território nacional. Brasil: Diário Oficial da União: seção 1, Brasília, DF, ano 145, n. 115, p. 69-70, 18 jun., 2008.

AGUADO, R. *et al.* Micro-and Nanofibrillated Cellulose from Annual Plant-Sourced Fibers: Comparison between Enzymatic Hydrolysis and Mechanical Refining. **Nanomaterials**, [s. l.], v. 12, n. 9: 1612, 2022.

ALMEIDA, J. dos S. C. de. **Produção Offshore de Gás Natural Rico em CO₂: Cenário Dutos Subsea com MEG e Processamento Onshore**. 2016. 192 f. Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2016.

ALMONTI, D. *et al.* Artificial neural network in fibres length prediction for high precision control of cellulose refining. **Materials**, [s. l.], v. 12, n. 22: 3730, 2019.

ALSHRAIDEH, H.; CASTILLO, E. Del; GIL DEL VAL, A. Process control via random forest classification of profile signals: An application to a tapping process. **Journal of Manufacturing Processes**, [s. l.], v. 58, p. 736–748, 2020.

ANG, S.; HARITOS, V.; BATCHELOR, W. Effect of refining and homogenization on nanocellulose fiber development, sheet strength and energy consumption. **Cellulose**, [s. l.], v. 26, n. 8, p. 4767–4786, 2019.

ANSALONI, L. *et al.* Nanocellulose-based membranes for CO₂ capture. **Journal of Membrane Science**, [s. l.], v. 522, p. 216–225, 2017.

ASIF, M. Sustainability of timber, wood and bamboo in construction. *In*: KHATIB, J. M. (org.). **Sustainability of Construction Materials**. [S. l.]: Woodhead Publishing, 2009. p. 31–54.

BAKER, R. W. **Membrane technologies and applications**. 2nd. ed. West Sussex: John Wiley & Sons, Ltd, 2004.

BALEA, A. *et al.* Influence of pretreatment and mechanical nanofibrillation energy on properties of nanofibers from Aspen cellulose. **Cellulose**, [s. l.], v. 28, n. 14, p. 9187–9206, 2021.

BENGIO, Y. Practical recommendations for gradient-based training of deep architectures. **Lecture Notes in Computer Science**, [s. l.], v. 7700, p. 437–478, 2012.

BRASIL. **LEI Nº 9.478, DE 6 DE AGOSTO DE 1997**. Dispõe sobre a política energética nacional, as atividades relativas ao monopólio do petróleo, institui o Conselho Nacional de Política Energética e a Agência Nacional do Petróleo e dá outras providências. Diário Oficial da União: seção 1, Brasília, DF, ano 135, n. 150, p. 1-8, 07 ago, 1997.

BREIMAN, L. Random Forests. **Machine Learning**, [s. l.], v. 45, p. 5–32, 2001.

BROCHU, E.; CORA, V. M.; DE FREITAS, N. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. [s. l.], 2010. Disponível em: <http://arxiv.org/abs/1012.2599>. Acesso em 20 nov. 2022.

BUDD, P. M. *et al.* Gas separation membranes from polymers of intrinsic microporosity. **Journal of Membrane Science**, [s. l.], v. 251, n. 1–2, p. 263–269, 2005.

CHEN, C. *et al.* Amine – silica composites for CO₂ capture : A short review. **Journal of Energy Chemistry**, [s. l.], v. 26, n. 5, p. 868–880, 2017.

CHEN, X.; LIU, G.; JIN, W. Natural gas purification by asymmetric membranes: An overview. **Green Energy and Environment**, [s. l.], v. 6, n. 2, p. 176–192, 2021.

CHRISTIAN, S. J. Natural fibre-reinforced noncementitious composites (biocomposites). *In*: HARRIES, K. A.; SHARMA, B. (org.). **Nonconventional and Vernacular Construction Materials**. 1st. ed. [S. l.]: Woodhead Publishing, 2016. p. 111–126.

CHRISTOPHER M BISHOP. **Neural Networks for Pattern Recognition**. London: Oxford University Press, 1996.

CNI. **Especificação do gás natural: oportunidades e experiência internacional**. Brasília: [s. n.], 2019.111p

COELHO, A. de M. **Simulação e otimização dos processos de adoçamento e desidratação de gás natural e de refinaria**. 2007. 210 f. Dissertação (Mestrado em Engenharia Química) - Faculdade de Engenharia Química, Universidade Estadual de Campinas, Campinas, 2007.

DAI, Z. *et al.* A brief review of nanocellulose based hybrid membranes for CO₂ separation. **Fibers**, [s. l.], v. 7, n. 5: 40, 2019.

DENG, L.; HÄGG, M. B. Swelling behavior and gas permeation performance of PVAm/PVA blend FSC membrane. **Journal of Membrane Science**, [s. l.], v. 363, n. 1–2, p. 295–301, 2010.

DENG, L.; KIM, T. J.; HÄGG, M. B. Facilitated transport of CO₂ in novel PVAm/PVA blend membrane. **Journal of Membrane Science**, [s. l.], v. 340, n. 1–2, p. 154–163, 2009.

DOBLE, M.; KUMAR, A. Degradation of Polymers. *In*: DOBLE, M.; KUMAR, A. (org.). **Biotreatment of industrial effluents**. [S. l.]: Butterworth-Heinemann, 2005. p. 159–169.

DUBEY, S. R.; SINGH, S. K.; CHAUDHURI, B. B. Activation functions in deep learning: A comprehensive survey and benchmark. **Neurocomputing**, [s. l.], v. 503, p. 92–108, 2022.

DUDA, R. O.; HART, P. E. **Pattern Classification and Scene Analysis**. 1^oed. [S. l.]: Wiley–Blackwell, 1973.

EBRAHIMI, S. *et al.* PVA/PES-amine-functional graphene oxide mixed matrix membranes for CO₂/CH₄ separation: Experimental and modeling. **Chemical Engineering Research and Design**, [s. l.], v. 109, p. 647–656, 2016.

EPE. **Panorama do Refino e da Petroquímica no Brasil Nota Técnica Dpg-Spt Nº 04/2018**. [S. l.: s. n.], 2018. Disponível em: [https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-/topico-412/NT Refino e Petroquímica_2018.11.01.pdf](https://www.epe.gov.br/sites-pt/publicacoes-dados-abertos/publicacoes/PublicacoesArquivos/publicacao-/topico-412/NT%20Refino%20e%20Petroqu%C3%ADmica_2018.11.01.pdf). Acesso em 10 de out. 2022.

ESPINOSA, E. *et al.* Use of multi-factorial analysis to determine the quality of cellulose nanofibers: effect of nanofibrillation treatment and residual lignin content. **Cellulose**, [s. l.], v. 27, n. 18, p. 10689–10705, 2020.

FERNANDES, E. M. *et al.* Bionanocomposites from lignocellulosic resources: Properties, applications and future trends for their use in the biomedical field. **Progress in Polymer Science**, [s. l.], v. 38, n. 10–11, p. 1415–1441, 2013.

FREITAS, L. F. de L. e *et al.* Gás Natural: Cenário Da Produção E Processos De Tratamento. **Recursos Naturais: Energia de Biomassa Florestal**, [s. l.], p. 128–145, 2021.

GAN, P. G. *et al.* Thermal properties of nanocellulose-reinforced composites: A review. **Journal of Applied Polymer Science**, [s. l.], v. 137, n. 11: 48544, 2019.

GÉRON, A. **Hands-on Machine Learning whith Scikit-Learning, Keras and Tensorflow**. 2nd. ed. [S. l.]: O'Reilly Media, 2019.

GHANADPOUR, M. *et al.* Phosphorylated Cellulose Nano fi brils : A Renewable Nanomaterial for the Preparation of Intrinsically Flame-Retardant Materials. **Bio Macromolecules**, [s. l.], v. 16, p. 3399–3410, 2015.

GUAN, J. *et al.* Design and prediction of metal organic framework-based mixed matrix membranes for CO₂ capture via machine learning. **Cell Reports Physical Science**, [s. l.], v. 3, n. 5: 100864, 2022.

HABERT, A. C.; BORGES, C. P.; NOBREGA, R. **Processos de Separação por membranas**. Rio de Janeiro: e-papers, 2006.

HAVSTAD, M. R. Biodegradable Plastics. *In*: TREVOR M. LETCHER (org.). **Plastic Waste and Recycling**. [S. l.]: Academic Press, 2020. p. 97–129.

HAYKIN, S. **Neural Networks: a Comprehensive Foundation**. 2nd. ed. Singapore: Pearson Prentice Hall, 1999.

HO, N. A. D.; LEO, C. P. A review on the emerging applications of cellulose, cellulose derivatives and nanocellulose in carbon capture. **Environmental Research**, [s. l.], v. 197, 111100, 2021.

HONG, C. H. *et al.* Progress in polyvinyl alcohol membranes with facilitated transport properties for carbon capture. **Journal of Environmental Chemical Engineering**, [s. l.], v. 9, n. 6: 106783, 2021.

HUBBE, M. A. *et al.* Rheology of Nanocellulose-rich Aqueous Suspensions: A Review. **BioResources**, [s. l.], v. 12, n. 4, p. 9556–9661, 2017.

HUGGINS, M. L. The Viscosity of Dilute Solutions of Long-Chain Molecules. IV. Dependence on Concentration. **Journal of the American Chemical Society**, [s. l.], v. 64, n. 11, p. 2716–2718, 1942.

IBRAHIM, S. *et al.* Biopolymers From Crop Plants. **Reference Module in Materials Science and Materials Engineering**, [s. l.], p. 1–10, 2019.

IPCC. Summary for policymakers. *In*: FIELD, C. B. *et al.* (org.). **Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation: Special Report of the Intergovernmental Panel on Climate Change**. Cambridge and New York: Cambridge University Press, 2012. p. 1–19.

ISMAILA, A. F. *et al.* Understanding the Solution-Diffusion Mechanism in Gas Separation Membrane for Engineering Students. *In*: **Proceedings of the 2005 Regional Conference on Engineering Education**. Johor: [s. n.], 2005. p. 155–159.

IULIANELLI, A.; DRIOLI, E. Membrane engineering: Latest advancements in gas separation and pre-treatment processes, petrochemical industry and refinery, and future perspectives in emerging applications. **Fuel Processing Technology**, [s. l.], v. 206, 106464, 2020.

IULIANELLI, A.; RUSSO, F.; GALIANO, F. PLA Easy Fil – White-based membranes for CO₂ separation. **Greenhouse Gases Science and Technology**, [s. l.], v. 9, n. 2, p. 360-369, 2019.

JAHAN, Z. *et al.* Cellulose nanocrystal / PVA nanocomposite membranes for CO₂ / CH₄ separation at high pressure. **Journal of Membrane Science**, [s. l.], v. 554, p. 275–281, 2018a.

JAHAN, Z. *et al.* Decoupling the effect of membrane thickness and CNC concentration in PVA based nanocomposite membranes for CO₂/CH₄ separation. **Separation and Purification Technology**, [s. l.], v. 204, p. 220–225, 2018b.

JAHAN, Z. *et al.* Phosphorylated nanocellulose fibrils/PVA nanocomposite membranes for biogas upgrading at higher pressure. **Separation Science and Technology (Philadelphia)**, [s. l.], v. 55, n. 8, p. 1524–1534, 2020.

JIANG, Chao; JIANG, Canchen; CHEN, D. Densely Connected Neural Networks for Nonlinear Regression. **Entropy**, [s. l.], v. 24, n. 7: 876, 2022.

KIM, S.; LEE, Y. M. Rigid and microporous polymers for gas separation membranes. **Progress in Polymer Science**, [s. l.], v. 43, p. 1–32, 2015.

KINGMA, D. P.; BA, J. L. ADAM: A METHOD FOR STOCHASTIC OPTIMIZATION. **International Conference on Learning Representations**. [S. l.: s. n.], 2015. Disponível em: <https://doi.org/10.48550/arXiv.1412.6980>. Acesso em 15 fev. 2023.

KLEMM, D. *et al.* Nanocellulose as a natural source for groundbreaking applications in materials science: Today's state. **Materials Today**, [s. l.], v. 21, n. 7, p. 720–748, 2018.

KRAEMER, E. O. Molecular Weights of Celluloses and Cellulose Derivatives. **Industrial and Engineering Chemistry**, [s. l.], v. 30, n. 10, p. 1200–1203, 1938.

LEE, H. *et al.* Chemical and physical reinforcement behavior of dialdehyde nanocellulose in PVA composite film: A comparison of nanofiber and nanocrystal. **Carbohydrate Polymers**, [s. l.], v. 232, 115771, 2020.

LEONG, S. L. *et al.* Morphological Control of Cellulose Nanocrystals via Sulfuric Acid Hydrolysis based on Sustainability Considerations: An Overview of the Governing Factors and Potential Challenges. **Journal of Environmental Chemical Engineering**, [s. l.], v. 10, n. 4: 108145, 2022.

LI, Z. *et al.* Further understanding the response mechanism of lignin content to bonding properties of lignocellulosic fibers by their deformation behavior. **RSC Advances**, [s. l.], v. 6, n. 110, p. 109211–109217, 2016.

LIU, Y. *et al.* Effects of amidation on gas permeation properties of polyimide membranes. **Journal of Membrane Science**, [s. l.], v. 214, p. 83–92, 2003.

LIU, L.; CHAKMA, A.; FENG, X. Gas permeation through water-swollen hydrogel membranes. **Journal of Membrane Science**, [s. l.], v. 310, n. 1–2, p. 66–75, 2008.

MAHARDIKA, M. *et al.* Production of nanocellulose from pineapple leaf fibers via high-shear homogenization and ultrasonication. **Fibers**, [s. l.], v. 6, n.2: 28, 2018.

MOVSESYAN, A. *et al.* Mie Sensing with Neural Networks: Recognition of Nano-Object Parameters, the Invisibility Point, and Restricted Models. **Advanced Theory and Simulations**, [s. l.], v. 5, n. 2: 2100369, 2022.

NASIR, M. *et al.* Nanocellulose: Preparation methods and applications. In: MOHAMMAD JAWAID; BOUFI, S.; H.P.S., A. K. (org.). **Cellulose-Reinforced Nanofibre Composites: Production, Properties and Applications**. [S. l.]: Woodhead Publishing, 2017. p. 261–276. *E-book*. Disponível em: <http://dx.doi.org/10.1016/B978-0-08-100957-4.00011-5>.

NETTLETON, D. Selection of Variables and Factor Derivation. **Commercial Data Mining**, [s. l.], Morgan Kaufmann, 2014. p. 79-104.

NGUYEN, H. *et al.* Suppression of crystallization in thin films of cellulose diacetate and its effect on CO₂/CH₄ separation properties. **Journal of Membrane Science**, [s. l.], v. 586, p. 7–14, 2019.

NIAZI, M. B. K. *et al.* Mechanical and thermal properties of carboxymethyl fibers (CMF)/PVA based nanocomposite membranes. **Journal of Industrial and Engineering Chemistry**, [s. l.], v. 90, p. 122–131, 2020.

NWANKPA, C. *et al.* Activation Functions: Comparison of trends in Practice and Research for Deep Learning. [s. l.], p. 1–20, 2018. Disponível em <https://doi.org/10.48550/arXiv.1811.03378>. Acesso em 20 nov. 2022.

OBERLAENDER, D. T. **Simulação e otimização de uma unidade de desidratação de gás natural offshore: uma análise com aplicação operacional**. 2015.114 f. Dissertação (Mestrado em Ciências) - Programa de Pós-Graduação em Tecnologia de Processos Químicos e Bioquímicos Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2015.

OLAJIRE, A. A. CO₂ capture and separation technologies for end-of-pipe applications - A review. **Energy**, [s. l.], v. 35, n. 6, p. 2610–2628, 2010.

ÖZKAN, M. *et al.* Machine Learning assisted design of tailor-made nanocellulose films: A combination of experimental and computational studies. **Polymer Composites**, [s. l.], v. 40, n. 10, p. 4013–4022, 2019.

PEDREGOSA, F.; WEISS, R.; BRUCHER, M. Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, [s. l.], v. 12, p. 2825–2830, 2011.

PENG, Z.; KONG, L. X. A thermal degradation mechanism of polyvinyl alcohol/silica nanocomposites. **Polymer Degradation and Stability**, [s. l.], v. 92, n. 6, p. 1061–1071, 2007.

PERESIN, M. S. *et al.* Nanofiber Composites of Polyvinyl Alcohol and Cellulose Nanocrystals: Manufacture and Characterization. **Biomacromolecules**, [s. l.], v. 11, p. 674–681, 2010.

PHROMPHITHAK, S.; ONSREE, T.; TIPPAYAWONG, N. Machine learning prediction of cellulose-rich materials from biomass pretreatment with ionic liquid solvents. **Bioresource Technology**, [s. l.], v. 323, n. 124642, 2021.

PROBST, P.; BOULESTEIX, A. L. To tune or not to tune the number of trees in random forest. **Journal of Machine Learning Research**, [s. l.], v. 18, n. 2001, p. 1–8, 2018. Disponível em: <https://doi.org/10.48550/arXiv.1705.05654>. Acesso em 15 de out. 2022.

REDDY, M. M. *et al.* Biobased plastics and bionanocomposites: Current status and future opportunities. **Progress in Polymer Science**, [s. l.], v. 38, n. 10–11, p. 1653–1689, 2013.

REIJERKERK, S. R. *et al.* Poly(ethylene glycol) and poly(dimethyl siloxane): Combining their advantages into efficient CO₂ gas separation membranes. **Journal of Membrane Science**, [s. l.], v. 352, n. 1–2, p. 126–135, 2010.

REVATHY, G. *et al.* Random Forest Regressor based superconductivity materials investigation for critical temperature prediction. **Materials Today: Proceedings**, [s. l.], v. 66, p. 648–652, 2022.

RUSSO, F. *et al.* Biopolymers for sustainable membranes in CO₂ separation: a review. **Fuel Processing Technology**, [s. l.], v. 213, n. 106643, 2021.

SAEED, U. *et al.* Biogas upgrading with novel cellulose nano-crystals and polyvinyl amine nanocomposite membranes. **Polymer Testing**, [s. l.], v. 91, n. 106867, 2020.

SANCHEZ-SALVADOR, J. L. *et al.* Critical comparison of the properties of cellulose nanofibers produced from softwood and hardwood through enzymatic, chemical and mechanical processes. **International Journal of Biological Macromolecules**, [s. l.], v. 205, p. 220–230, 2022.

SANCHEZ-SALVADOR, J. L. *et al.* Simplification of gel point characterization of cellulose nano and microfiber suspensions. **Cellulose**, [s. l.], v. 28, n. 11, p. 6995–7006, 2021.

SANTOS, A. F. *et al.* Artificial neural network for aspect ratio prediction of lignocellulosic micro/nanofibers. **Cellulose**, [s. l.], v. 29, n. 10, p. 5609–5622, 2022.

SCIKIT-LEARN. **Random Forest Regressor model**. [S. l.], 2022. Disponível em: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestRegressor.html>. Acesso em: 11 jan. 2022.

SEGAL, L. *et al.* An Empirical Method for Estimating the Degree of Crystallinity of Native Cellulose Using the X-Ray Diffractometer. **Textile Research Journal**, [s. l.], v. 29, n. 10, p. 786–794, 1959.

SERRA-PARAREDA, F. *et al.* Chemical-free production of lignocellulosic micro- and nanofibers from high-yield pulps: Synergies, performance, and feasibility. **Journal of Cleaner Production**, [s. l.], v. 313, n. 127914, p. 1–10, 2021a.

SERRA-PARAREDA, F. *et al.* Correlation between rheological measurements and morphological features of lignocellulosic micro/nanofibers from different softwood sources. **International Journal of Biological Macromolecules**, [s. l.], v. 187, p. 789–799, 2021b.

SERRA-PARAREDA, F. *et al.* Potentiometric back titration as a robust and simple method for specific surface area estimation of lignocellulosic fibers. **Cellulose**, [s. l.], v. 28, n. 17, p. 10815–10825, 2021c.

SERRA-PARAREDA, F. *et al.* Techno-economic and environmental evaluation of a market pulp reinforced with micro-/nanofibers as a strengthening agent in packaging paper. **Journal of Cleaner Production**, [s. l.], v. 347, n. 131265, 2022.

SERRA-PARAREDA, F. *et al.* Tuning morphology and structure of non-woody nanocellulose: Ranging between nanofibers and nanocrystals. **Industrial Crops and Products**, [s. l.], v. 171, n. 113877, 2021d.

SHAHRIARI, B. *et al.* Taking the human out of the loop: A review of Bayesian optimization. **Proceedings of the IEEE**, [s. l.], v. 104, n. 1, p. 148–175, 2016.

SHIN, S. *et al.* Deep neural network model with Bayesian hyperparameter optimization for prediction of NO_x at transient conditions in a diesel engine. **Engineering Applications of Artificial Intelligence**, [s. l.], v. 94, n. 103761, 2020.

SIGNORI-IAMIN, G. *et al.* Prediction of cellulose micro/nanofiber aspect ratio and yield of nanofibrillation using machine learning techniques. **Cellulose**, [s. l.], v. 29, p. 9143–9162, 2022.

SIMHA, R. The Influence of Brownian Movement on the Viscosity of Solutions. **The Journal of Physical Chemistry**, [s. l.], v. 44, n. 1, p. 25–34, 1940.

SINGH, J. *et al.* Enzymatic Processing of Juice From Fruits / Vegetables : An Emerging Trend and Cutting Edge Research in Food. *In*: MOHAMMED KUDDUS (org.). **Enzymes in Food Biotechnology**. [S. l.]: Academic Press, 2019. p. 419–432.

SLUITER, A. *et al.* **Determination of Structural Carbohydrates and Lignin in Biomass**. [S. l.: s. n.], 2012.17p.

SMITH, A. M.; MOXON, S.; MORRIS, G. A. Biopolymers as wound healing materials. *In*: ÅGREN, M. S. (org.). **Wound Healing Biomaterials**. [S. l.]: Woodhead Publishing, 2016. v. 2, p. 261–287.

SYDUZZAMAN, M. *et al.* Plant-based natural fibre reinforced composites: A review on fabrication, properties and applications. **Coatings**, [s. l.], v. 10, n. 10, p. 1–34, 2020.

TAKHT RAVANCHI, M.; SAHEBDELFAH, S. Carbon dioxide capture and utilization in petrochemical industry: potentials and challenges. **Applied Petrochemical Research**, [s. l.], v. 4, n. 1, p. 63–77, 2014.

TAPPI. **TAPPI Standards, Technical Information Papers, and Useful Methods**. New York: Technical Association of the Pulp & Paper Industry, 2020.

THOMAS, J. E. **Fundamentos de engenharia de petróleo**. 2nd. ed. Rio de Janeiro: Editora Interciência, 2004.

TORRENTS-BARRENA, J.; PELLICER, E. Smart Cellulose Composites: Advanced Applications and Properties Prediction Using Machine Learning. **Encyclopedia of Materials: Composites**, [s. l.], v. 2, p. 527–538, 2021.

TORSTENSEN, J. Ø. *et al.* PVA / nanocellulose nanocomposite membranes for CO₂ separation from flue gas. **International Journal of Greenhouse Gas Control**, [s. l.], v. 81, p. 93–102, 2019.

TRACHE, D. *et al.* Nanocellulose: From Fundamentals to Advanced Applications. **Frontiers in Chemistry**, [s. l.], v. 8, n. 392, 2020.

TURNER, R. *et al.* Bayesian Optimization is Superior to Random Search for Machine Learning Hyperparameter Tuning: Analysis of the Black-Box Optimization Challenge 2020. [s. l.], 2021. Disponível em: <http://arxiv.org/abs/2104.10201>. Acesso em 02 de nov. 2022.

UEDA, T. *et al.* Structural, rheological, and mechanical properties of polyvinyl alcohol composites reinforced with cellulose nanofiber treated by ultrahigh-pressure homogenizer. **Materials Today Communications**, [s. l.], v. 33, n. 104316, 2022.

VARANASI, S.; HE, R.; BATCHELOR, W. Estimation of cellulose nanofibre aspect ratio from measurements of fibre suspension gel point. **Cellulose**, [s. l.], v. 20, n. 4, p. 1885–1896, 2013.

VENTURI, D.; ANSALONI, L.; BASCHETTI, M. G. Nanocellulose based facilitated transport membranes for CO₂ separation. **Chemical Engineering Transactions**, [s. l.], v. 47, p. 349–354, 2016.

VISAKH, P. M.; THOMAS, S. Preparation of bionanomaterials and their polymer nanocomposites from waste and biomass. **Waste and Biomass Valorization**, [s. l.], v. 1, n. 1, p. 121–134, 2010.

VORONOVA, M. I. *et al.* Thermal stability of polyvinyl alcohol/nanocrystalline cellulose composites. **Carbohydrate Polymers**, [s. l.], v. 130, p. 440–447, 2015.

VU, D. Q.; KOROS, W. J.; MILLER, S. J. Effect of condensable impurity in CO₂/CH₄ gas feeds on performance of mixed matrix membranes using carbon molecular sieves. **Journal of Membrane Science**, [s. l.], v. 221, n. 1–2, p. 233–239, 2003.

WANG, T. *et al.* Mechanisms of viscosity increase for nanocolloidal dispersions. **Journal of Nanoscience and Nanotechnology**, [s. l.], v. 11, n. 4, p. 3141–3150, 2011.

XIE, Y. *et al.* Stroke prediction from electrocardiograms by deep neural network. **Multimedia Tools and Applications**, [s. l.], v. 80, p. 17291–17297, 2021.

XING, L. *et al.* Biodegradable cellulose I (II) nanofibrils/poly(vinyl alcohol) composite films with high mechanical properties, improved thermal stability and excellent transparency. **International Journal of Biological Macromolecules**, [s. l.], v. 164, p. 1766–1775, 2020.

XU, Y. *et al.* The influence of nano-fibrillated cellulose as a coating component in paper coating. **BioResources**, [s. l.], v. 11, n. 2, p. 4342–4352, 2016.

YOON, Y.; SWALES, G.; SILVIA, E. M. A comparison of discriminant analysis versus artificial neural networks. **Journal of the Operational Research Society**, [s. l.], v. 44, n. 1, p. 51–60, 1993.

YUWAWECH, K.; WOOTHIKANOKKHAN, J.; TANPICHAI, S. Effects of two different cellulose nanofiber types on properties of poly(vinyl alcohol) composite films. **Journal of Nanomaterials**, [s. l.], v. 2015, n. 908689, 2015.

ZHANG, Z. *et al.* Machine learning aided high-throughput prediction of ionic liquid@MOF composites for membrane-based CO₂ capture. **Journal of Membrane Science**, [s. l.], v. 650, n. 120399, 2022.

APÊNDICE A – BASE DE DADOS PARA PREDIÇÃO DE PROPRIEDADES DA NANOCELULOSE: RAZÃO DE ASPECTO E RENDIMENTO DE NANOFIBRILAÇÃO.

QUADRO 3 - COMPOSIÇÃO QUÍMICA E ÍNDICE DE CRISTALINIDADE DAS POLPAS LIGNOCELULÓSICAS UTILIZADAS PARA GERAÇÃO DAS CNFs MECÂNICA E ENZIMÁTICA.

Pulp	Celulose (%p/p)	Hemicelulose (%p/p)	Lignina total (%p/p)	Extrativos (%p/p)	Cinzas (%p/p)	Índice de cristalinidade (%)
TMP_mec	47,9	22,9	27	0,7	1,4	73,3
TMP_80Henz	48,4	21,2	28	1,1	1,3	-
TMP_240Henz	48,4	19,9	29,2	1,3	1,2	-
UKSP_mec	74,2	16,6	8,2	0	0,8	81,1
UKSP_80Henz	74,8	15,9	8,2	0,6	0,5	-
UKSP_240Henz	78,0	13,9	34,5	0	0,5	-
BKSP_mec	85,3	8,7	3,9	1,2	0,9	87
BKSP_80Henz	86,3	7,2	4,3	1,5	0,7	-
BKSP_240Henz	86,8	7,7	3,9	1,2	0,6	-
BTMP_mec	46,2	22,9	29,4	0,9	0,5	76,1
Euca_mec	74,2	17,4	6,6	1,3	0,5	-
Euca_80Henz	78,8	13,5	6,8	0,5	0,4	-
Euca_240Henz	75,6	17,0	6,5	0,5	0,4	-
Álamo_mec	47,1	19,4	31,3	0,8	1,3	-
Álamo_80Henz	44,6	20,1	33,2	0,8	1,3	-
Álamo_240Henz	46,4	18,2	33,0	0,9	1,5	-
Cânhamo_mec	87,2	6,8	4,0	1,2	0,7	-

Pulp	Celulose (%p/p)	Hemicelulose (%p/p)	Lignina total (%p/p)	Extraíveis (%p/p)	Cinzas (%p/p)	Índice de cristalinidade (%)
Cânhamo_80Henz	88,7	5,5	4,7	0,6	0,6	-
Cânhamo_240Henz	87,8	6,7	4,3	0,5	0,6	-
Sisal_mec	87,2	13,4	5,7	1,1	1,0	-
Sisal_80Henz	88,7	15,9	6,1	1,2	0,7	-
Sisal_240Henz	87,8	15,6	6,0	1,3	0,7	-
Juta_mec	81,7	9,0	7,6	0,7	1,0	-
Juta_80Henz	79,8	11,5	7,6	0,3	0,8	-
Juta_240Henz	82,2	9,3	7,3	0,3	0,9	-
Rec_mec	56,4	12,5	16,8	1,8	12,6	-
Rec_80Henz	60,6	13,8	17	1,1	7,4	-
Rec_240Henz	55,1	15,3	20,5	1,0	8,1	-

Fonte: LEPAMAP-PRODIS (AGUADO *et al.*, 2022; BALEA *et al.*, 2021; SANCHEZ-SALVADOR *et al.*, 2022; SERRA-PARAREDA *et al.*, 2021b)

Os valores de consumo energético, dosagem de enzima, demanda catiônica (DC), transmitância a 600 nm (trans), parâmetros reológicos “k” e “n”, a razão de aspecto e o rendimento de nanofibrilação obtidos estão compilados abaixo.

QUADRO 4 - CARACTERÍSTICAS DAS NANOFIBRAS DE CELULOSE PRODUZIDAS DE ACORDO COM DOSAGEM ENZIMÁTICA E INTENSIDADE DO PROCESSO DE NANOFIBRILAÇÃO.

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
TMP_mec	0	5,67	148	2,5	0,222	0,281	51	3,12
	0	7,73	155	3,1	0,625	0,272	65	4,22
	0	11,67	170	5,9	0,789	0,270	72	6,74
	0	14,59	179	7,0	0,889	0,257	74	7,31
	0	19,72	190	7,4	1,107	0,239	80	8,11
TMP_80Henz	80	5,67	196	6,8	0,261	0,231	38	8,20
	80	7,73	203	8,3	0,419	0,205	38	10,70
	80	11,67	212	13,7	0,809	0,186	37	14,7
	80	14,59	217	15,6	0,900	0,180	44	16,69
	80	19,72	224	20,8	0,963	0,180	48	21,82
TMP_240Henz	240	5,67	204	9,6	0,308	0,360	50	10,47
	240	7,73	210	10,9	0,722	0,323	53	12,63
	240	11,67	215	14,5	0,878	0,305	53	15,73
	240	14,59	220	17,5	1,103	0,275	54	18,41
	240	19,72	229	22,1	1,274	0,270	56	23,92
UKSP_mec	0	5,67	157	4,2	1,461	0,373	63	7,40
	0	7,73	166	6,5	4,987	0,280	66	9,26
	0	11,67	177	7,4	6,780	0,276	78	12,41
	0	14,59	188	9,9	6,638	0,224	90	14,20
	0	19,72	206	12,5	7,679	0,233	103	17,94

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
UKSP_80Henz	80	5,67	171	6,2	0,625	0,384	65	8,16
	80	7,73	181	10,7	0,946	0,351	74	12,81
	80	11,67	198	14,9	1,232	0,342	90	17,85
	80	14,59	205	15,9	4,114	0,278	92	19,01
	80	19,72	212	17,3	4,161	0,265	102	22,63
UKSP_240Henz	240	5,67	184	8,6	1,623	0,383	100	10,77
	240	7,73	192	12,1	1,580	0,353	110	14,64
	240	11,67	207	18,5	2,254	0,322	111	21,7
	240	14,59	211	23,1	3,737	0,280	110	25,07
	240	19,72	218	28,4	6,066	0,285	117	31,27
BKSP_mec	0	5,67	154	5,4	2,178	0,253	140	7,46
	0	7,73	169	6,8	2,824	0,256	177	8,80
	0	11,67	188	7,2	3,873	0,247	212	11,57
	0	14,59	201	9,5	4,816	0,234	214	13,12
	0	19,72	210	12,7	6,549	0,204	217	15,48
BKSP_80Henz	80	5,67	181	4,2	0,802	0,360	92	6,042
	80	7,73	197	6,3	1,163	0,350	101	8,53
	80	11,67	203	9,6	1,299	0,353	114	11,77
	80	14,59	207	13,1	1,726	0,237	131	14,04
	80	19,72	213	16,2	2,014	0,227	142	17,16
BKSP_240Henz	240	5,67	192	7,4	0,665	0,437	57	9,13
	240	7,73	199	8,8	0,761	0,436	72	10,93
	240	11,67	206	12,9	1,781	0,427	91	14,72

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
	240	14,59	212	14,5	3,154	0,387	107	16,32
	240	19,72	219	19,5	4,217	0,329	133	21,87
BTMP_mec	0	5,67	156	4,1	0,349	0,338	59	2,90
	0	7,73	173	4,7	0,456	0,324	69	6,00
	0	11,67	185	7,0	0,616	0,307	68	11,60
	0	14,59	199	9,7	0,781	0,299	70	15,60
	0	19,72	214	11,9	1,201	0,278	70	20,60
	0	5,67	184	9,5	0,265	0,308	155	10,24
Euca_mec	0	7,73	195	10,7	0,394	0,296	168	11,57
	0	11,67	208	13,2	1,386	0,267	179	16,55
	0	14,59	213	16,6	1,672	0,273	200	18,97
	0	19,72	215	19,4	2,097	0,245	220	20,43
Euca_80Henz	80	5,67	204	14,5	0,885	0,266	53	15,21
	80	7,73	209	15,3	1,006	0,262	61	17,31
	80	11,67	216	18,8	1,199	0,193	86	20,95
	80	14,59	224	21,8	1,457	0,182	103	22,74
	80	19,72	232	24,12	1,798	0,125	126	24,94
	240	5,67	200	13,9	0,285	0,562	74	15,69
Euca_240Henz	240	7,73	212	15,8	0,548	0,349	82	18,47
	240	11,67	224	21,3	1,598	0,287	92	25,51
	240	14,59	238	26,1	1,965	0,264	109	28,12
	240	19,72	250	33,7	2,562	0,247	138	34,73
Álamo_mec	0	5,67	175	6,7	0,400	0,389	62	8,87

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
	0	7,73	180	8,3	0,617	0,386	65	10,86
	0	11,67	196	12,8	0,667	0,369	65	14,57
	0	14,59	203	15,2	0,832	0,353	70	17,77
	0	19,72	216	19,3	1,122	0,310	74	22,46
Álamo_80Henz	80	5,67	206	8,9	0,326	0,452	60	10,91
	80	7,73	215	10,6	0,334	0,446	61	12,04
	80	11,67	219	15,4	0,389	0,441	63	17,27
	80	14,59	227	18,1	0,840	0,348	67	19,46
	80	19,72	231	22,4	1,256	0,302	72	23,21
	240	5,67	209	8,6	0,664	0,359	53	11,26
	240	7,73	217	12,7	1,172	0,257	55	14,33
	240	11,67	225	16,8	1,403	0,237	58	19,67
	240	14,59	230	21,8	1,855	0,195	60	23,72
	240	19,72	239	24,7	2,340	0,180	63	26,47
	0	5,67	134	3,5	1,437	0,166	188	5,09
	0	7,73	149	5,2	2,362	0,151	211	6,98
Cânhamo_mec	0	11,67	168	7,9	2,656	0,149	193	9,52
	0	14,59	187	9,1	4,017	0,108	228	11,06
	0	19,72	205	12,7	4,221	0,101	217	15,10
	80	5,67	151	6,9	0,131	0,491	83	6,54
Cânhamo_80Henz	80	7,73	167	9,5	0,188	0,455	86	9,02
	80	11,67	184	11,19	0,221	0,435	87	12,59
	80	14,59	201	12,8	0,332	0,364	97	14,24

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
	80	19,72	210	16,4	0,570	0,295	103	18,56
	240	5,67	162	6,3	0,235	0,417	81	8,99
	240	7,73	189	9,1	0,281	0,332	84	10,80
Cânhamo_240Henz	240	11,67	200	10,7	0,588	0,289	89	14,56
	240	14,59	208	16,3	0,721	0,292	99	19,59
	240	19,72	221	21,2	1,117	0,262	103	24,05
	0	5,67	131	3,8	1,632	0,206	117	5,01
	0	7,73	145	5,2	1,979	0,112	123	7,05
Sisal_mec	0	11,67	163	7,6	3,278	0,072	139	10,26
	0	14,59	181	9,4	3,313	0,070	139	11,07
	0	19,72	199	11,2	3,494	0,062	152	14,59
	80	5,67	168	7,1	0,205	0,573	64	8,96
	80	7,73	179	9,8	0,271	0,558	64	11,87
Sisal_80Henz	80	11,67	194	14,9	0,366	0,511	67	16,02
	80	14,59	207	15,1	0,379	0,425	74	18,26
	80	19,72	214	19,6	0,517	0,392	78	20,46
	240	5,67	176	8,7	0,146	0,545	64	11,36
	240	7,73	193	12,5	0,203	0,510	64	14,24
Sisal_240Henz	240	11,67	203	17,4	0,286	0,503	66	19,57
	240	14,59	212	20,1	0,358	0,458	61	21,37
	240	19,72	226	25,9	0,355	0,470	63	27,46
	0	5,67	155	4,9	1,057	0,395	144	6,87
Juta_mec	0	7,73	167	7,3	1,425	0,386	147	9,07

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC (µeq/g)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
	0	11,67	182	11,6	1,641	0,325	156	13,27
	0	14,59	198	13,1	2,795	0,330	163	15,69
	0	19,72	209	16,3	3,59	0,332	164	17,24
Juta_80Henz	80	5,67	186	8,1	0,526	0,445	80	9,66
	80	7,73	201	12,2	0,747	0,384	92	13,02
	80	11,67	208	18,5	1,092	0,368	90	19,02
	80	14,59	215	20,5	1,226	0,324	89	21,20
	80	19,72	220	24,2	2,176	0,227	97	26,12
Juta_240Henz	240	5,67	204	11,3	1,461	0,176	89	13,22
	240	7,73	211	15,6	1,884	0,178	86	17,97
	240	11,67	223	20,4	2,917	0,136	79	22,26
	240	14,59	226	23,6	3,580	0,136	82	25,53
	240	19,72	233	28,7	4,942	0,131	86	30,17
Rec_mec	0	5,67	151	4,2	0,817	0,670	133	5,97
	0	7,73	164	6,8	0,870	0,579	151	7,43
	0	11,67	172	8,7	1,318	0,469	151	9,22
	0	14,59	186	10,3	1,533	0,470	167	11,01
	0	19,72	202	15,2	1,934	0,437	173	16,23
Rec_80Henz	80	5,67	175	7,9	0,788	0,372	75	9,27
	80	7,73	186	11,0	1,189	0,347	77	12,67
	80	11,67	2	16,8	1,421	0,364	80	18,21
	80	14,59	221	19,8	1,666	0,410	85	21,06
	80	19,72	233	26,3	1,911	0,373	91	28,00

Pulp	Dosagem de enzima (g/ton)	Consumo energético(kWh/kg)	DC ($\mu\text{eq/g}$)	Trans. (%)	k	n	Razão de aspecto	Rendimento (%)
Rec_240Henz	240	5,67	176	7,5	1,015	0,318	71	9,29
	240	7,73	189	11,8	1,501	0,253	72	13,81
	240	11,67	205	18,4	1,647	0,229	75	20,34
	240	14,59	216	22,1	2,334	0,203	82	24,49
	240	19,72	228	28,9	2,872	0,198	109	30,01

Fonte: LEPAMAP-PRODIS (AGUADO *et al.*, 2022; BALEA *et al.*, 2021; SANCHEZ-SALVADOR *et al.*, 2022; SERRA-PARAREDA *et al.*, 2021b)

APÊNDICE B – BASE DE DADOS PARA MODELAGEM DAS PROPRIEDADES DE MEMBRANA

QUADRO 5 - BASE DE DADOS PARA MEMBRANAS DE PVA PURA, PVA+CNC E PVA+CNF.

Material	Razão de aspecto	pH da solução mãe	Concentração de nanocelulose (%)	Pressão de alimentação (bar)	Umidade relativa (%)	Permeância de CO ₂ (m ³ (STP).m ⁻² .bar ⁻¹ .h ⁻¹)	Seletividade CO ₂ /CH ₄
PVA + CNC	22,03	9	3,8%	5	30%	0,020	3
PVA + CNC	22,03	9	2,0%	5	30%	0,020	4
PVA + CNC	22,03	9	0,5%	5	30%	0,020	5
PVA + CNC	22,03	9	1,0%	5	30%	0,030	6
PVA + CNC	22,03	9	1,5%	5	30%	0,040	6
PVA + CNC	22,03	9	3,8%	5	50%	0,040	9
PVA + CNC	22,03	9	2,0%	5	50%	0,050	11
PVA + CNC	22,03	9	0,5%	5	50%	0,060	8
PVA + CNC	22,03	9	1,0%	5	50%	0,070	10
PVA + CNC	22,03	9	1,5%	5	50%	0,080	11
PVA + CNC	22,03	9	0,5%	5	70%	0,100	20
PVA + CNC	22,03	9	3,8%	5	70%	0,110	19
PVA + CNC	22,03	9	1,0%	5	70%	0,120	21
PVA + CNC	22,03	9	2,0%	5	70%	0,130	22
PVA + CNC	22,03	9	1,5%	5	70%	0,185	28
PVA + CNC	22,03	9	0,5%	5	90%	0,180	30
PVA + CNC	22,03	9	3,8%	5	90%	0,185	29
PVA + CNC	22,03	9	2,0%	5	90%	0,200	31
PVA + CNC	22,03	9	1,0%	5	90%	0,230	34

Material	Razão de aspecto	pH da solução mãe	Concentração de nanocelulose (%)	Pressão de alimentação (bar)	Umidade relativa (%)	Permeança de CO ₂ (m ³ (STP).m ⁻² .bar ⁻¹ .h ⁻¹)	Seletividade CO ₂ /CH ₄
PVA + CNC	22,03	9	1,5%	5	90%	0,250	37
PVA + CNC	22,03	6	2,0%	15	93%	0,050	10
PVA + CNC	22,03	6	0,5%	15	93%	0,060	11
PVA + CNC	22,03	6	1,5%	15	93%	0,070	12
PVA + CNC	22,03	9	0,5%	15	93%	0,070	27
PVA + CNC	22,03	6	1,0%	15	93%	0,079	13
PVA + CNC	22,03	6	0,5%	10	93%	0,080	19
PVA + CNC	22,03	6	2,0%	10	93%	0,080	18
PVA + CNC	22,03	9	1,5%	15	93%	0,080	25
PVA + CNC	22,03	9	2,0%	15	93%	0,080	24
PVA + CNC	22,03	6	1,5%	10	93%	0,090	22
PVA + CNC	22,03	9	1,0%	15	93%	0,100	32
PVA + CNC	22,03	9	2,0%	10	93%	0,100	31
PVA + CNC	22,03	10	1,0%	15	93%	0,100	33
PVA + CNC	22,03	6	1,0%	10	93%	0,110	22
PVA + CNC	22,03	9	0,5%	10	93%	0,110	30
PVA + CNC	22,03	9	1,5%	10	93%	0,130	32
PVA + CNC	22,03	6	2,0%	5	93%	0,150	24
PVA + CNC	22,03	6	0,5%	5	93%	0,170	25
PVA + CNC	22,03	9	1,0%	10	93%	0,180	37
PVA + CNC	22,03	6	1,5%	5	93%	0,190	26
PVA + CNC	22,03	9	2,0%	5	93%	0,190	35

Material	Razão de aspecto	pH da solução mãe	Concentração de nanocelulose (%)	Pressão de alimentação (bar)	Umidade relativa (%)	Permeação de CO ₂ (m ³ (STP).m ⁻² .bar ⁻¹ .h ⁻¹)	Seletividade CO ₂ /CH ₄
PVA + CNC	22,03	6	1,0%	5	93%	0,200	28
PVA + CNC	22,03	9	0,5%	5	93%	0,200	30
PVA + CNC	22,03	10	1,0%	10	93%	0,200	39
PVA + CNC	22,03	9	1,5%	5	93%	0,230	36
PVA + CNC	22,03	9	1,0%	5	93%	0,270	39
PVA + CNC	22,03	10	1,0%	5	93%	0,290	43
PVA + CNC	22,03	9	0,5%	15	100%	0,070	20
PVA + CNC	22,03	9	3,8%	15	100%	0,070	20
PVA + CNC	22,03	9	2,0%	15	100%	0,077	24
PVA + CNC	22,03	9	1,0%	15	100%	0,090	25
PVA + CNC	22,03	9	3,8%	10	100%	0,110	24
PVA + CNC	22,03	9	1,5%	15	100%	0,113	30
PVA + CNC	22,03	9	0,5%	10	100%	0,120	26
PVA + CNC	22,03	9	2,0%	10	100%	0,128	29
PVA + CNC	22,03	9	1,0%	10	100%	0,180	30
PVA + CNC	22,03	9	0,5%	5	100%	0,190	31
PVA + CNC	22,03	9	1,5%	10	100%	0,200	36
PVA + CNC	22,03	9	3,8%	5	100%	0,210	29
PVA + CNC	22,03	9	2,0%	5	100%	0,225	32
PVA + CNC	22,03	9	1,0%	5	100%	0,250	34
PVA + CNC	22,03	9	1,5%	5	100%	0,270	39
PVA + CNF	250	9	0,0%	15	95%	0,030	19

Material	Razão de aspecto	pH da solução mãe	Concentração de nanocelulose (%)	Pressão de alimentação (bar)	Umidade relativa (%)	Permeança de CO ₂ (m ³ (STP).m ⁻² .bar ⁻¹ .h ⁻¹)	Seletividade CO ₂ /CH ₄
PVA + CNF	250	9	3,8%	15	95%	0,030	18
PVA + CNF	250	5	1,0%	15	95%	0,030	11
PVA + CNF	250	9	3,8%	10	95%	0,040	21
PVA + CNF	250	5	1,0%	10	95%	0,040	15
PVA + CNF	250	5	1,0%	5	95%	0,050	20
PVA + CNF	250	9	0,0%	10	95%	0,055	23
PVA + CNF	250	9	3,8%	5	95%	0,060	28
PVA + CNF	250	9	2,0%	15	95%	0,060	20
PVA + CNF	250	9	0,0%	5	95%	0,070	29
PVA + CNF	250	9	1,0%	15	95%	0,080	27
PVA + CNF	250	9	2,0%	10	95%	0,085	26
PVA + CNF	250	9	2,0%	5	95%	0,100	33
PVA + CNF	250	9	1,0%	10	95%	0,105	31
PVA + CNF	250	12	1,0%	15	95%	0,120	30
PVA + CNF	250	9	1,0%	5	95%	0,140	39
PVA + CNF	250	12	1,0%	10	95%	0,173	37
PVA + CNF	250	12	1,0%	5	95%	0,210	45
PVA PURO	0	9	0,0%	5	30%	0,013	4
PVA PURO	0	9	0,0%	5	50%	0,020	6
PVA PURO	0	9	0,0%	5	70%	0,040	11
PVA PURO	0	9	0,0%	5	90%	0,070	17
PVA PURO	0	6	0,0%	15	93%	0,020	6

Material	Razão de aspecto	pH da solução mãe	Concentração de nanocelulose (%)	Pressão de alimentação (bar)	Umidade relativa (%)	Permeança de CO ₂ (m ³ (STP).m ⁻² .bar ⁻¹ .h ⁻¹)	Seletividade CO ₂ /CH ₄
PVA PURO	0	6	0,0%	10	93%	0,030	12
PVA PURO	0	9	0,0%	15	93%	0,030	19
PVA PURO	0	6	0,0%	5	93%	0,050	17
PVA PURO	0	9	0,0%	10	93%	0,050	23
PVA PURO	0	9	0,0%	5	93%	0,070	29
PVA PURO	0	9	0,0%	5	100%	0,080	22

Fonte: Adaptado de (JAHAN *et al.*, 2018a, 2018b, 2020)

APÊNDICE C – RESULTADOS DA MODELAGEM PRELIMINAR DE PREDIÇÃO DA RAZÃO DE ASPECTO DE CNFS

Foram gerados modelos preliminares com os dados de CNF que estavam disponibilizados à época em que se iniciou a modelagem. Estas amostras foram aquelas produzidas a partir das polpas BKSP, UKSP, BTMP e TMP que foram submetidas ao pré-tratamento mecânico. Este conjunto de dados preliminar contou com 20 pontos experimentais.

As variáveis de entrada foram as mesmas apresentadas na seção 3.1.1, com exceção da dosagem enzimática e com adição do índice de cristalinidade. O índice de cristalinidade foi estimado a partir do pico (002) da celulose tipo I presente nos padrões de difração de raio X, seguindo a metodologia empírica de Segal (SEGAL *et al.*, 1959).

Os valores dos coeficientes de correlação de Pearson de cada variável de entrada em relação a razão de aspecto estão apresentados com seu valor absoluto na TABELA 29 abaixo.

TABELA 29 - VALORES ABSOLUTOS DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON PARA AS VARIÁVEIS DE ENTRADA EM RELAÇÃO A RAZÃO DE ASPECTO.

Variável de entrada	Coeficiente de correlação de Pearson
Hemicelulose (%p/p)	0,8870
Índice de cristalinidade (%)	0,8407
Celulose (%p/p)	0,7690
Lignina total (%p/p)	0,7092
Índice de fluxo (n)	0,6042
Extraíveis (%p/p)	0,5489
Índice de consistência (k)	0,4778
Transmitância a 600 nm (%)	0,4142
Demanda catiônica ($\mu\text{eq/g}$)	0,3857
Consumo energético (kWh/kg)	0,2377
Cinzas (%p/p)	0,0095

Fonte: A autora (2023)

As variáveis que indicam uma correlação mais significativa com a variável de interesse são aquelas com valores iguais ou maiores que 0,7 (NETTLETON, 2014). Portanto, o teor de hemicelulose, índice de cristalinidade, celulose e lignina total possuem uma relação mais importante com a razão de aspecto.

Os tipos de variáveis e a descrição estatística deste subconjunto de dados preliminar estão apresentados na TABELA 30.

TABELA 30 - DESCRIÇÃO ESTATÍSTICA DAS VARIÁVEIS DE ENTRADA E ÚNICA VARIÁVEL DE SAÍDA CONSIDERADAS PARA A MODELAGEM.

Variável	Contagem de valores não nulos	Valor médio	Desvio padrão	Valor mínimo	25%	50%	75%	Valor máximo
Demanda catiônica ($\mu\text{eq/g}$)	20	179,2	20,1	148	163,8	178	192,2	214
Transmitância a 600 nm (%)	20	0,0727	0,0298	0,0250	0,0522	0,07	0,0955	0,1270
Consumo energético (kWh/kg)	20	11,88	5,12	5,67	7,73	11,60	14,59	19,72
Índice de consistência (k)	20	2,7364	2,5620	0,2216	0,7419	1,3311	4,8584	7,6787
Índice de fluxo (n)	20	0,272	0,041	0,204	0,245	0,271	0,285	0,373
Celulose (%p/p)	20	0,634	0,726	0,462	0,475	0,611	0,770	0,853
Hemicelulose (%p/p)	20	0,178	0,060	0,087	0,146	0,198	0,229	0,229
Lignina total (%p/p)	20	0,171	0,115	0,039	0,071	0,176	0,276	0,294
Extraíveis (%p/p)	20	0,007	0,005	0,000	0,005	0,008	0,010	0,012
Cinzas (%p/p)	20	0,009	0,003	0,000	0,007	0,009	0,010	0,014
Índice de cristalinidade (%)	20	0,7938	0,0535	0,7330	0,7540	0,7860	0,8258	0,8700
Razão de aspecto	20	101,9	56,6	51	67,5	73	112,2	217

Fonte: A autora (2023)

Desempenho dos modelos

A modelagem foi realizada com florestas randômicas e redes neurais. O conjunto total de dados foi dividido de forma aleatória em uma razão 80:20 entre subconjuntos de treino e teste, não havendo um de validação para estes modelos preliminares por conta do número limitado de pontos. Os resultados para cada algoritmo estão apresentados nas subseções abaixo.

Florestas Randômicas

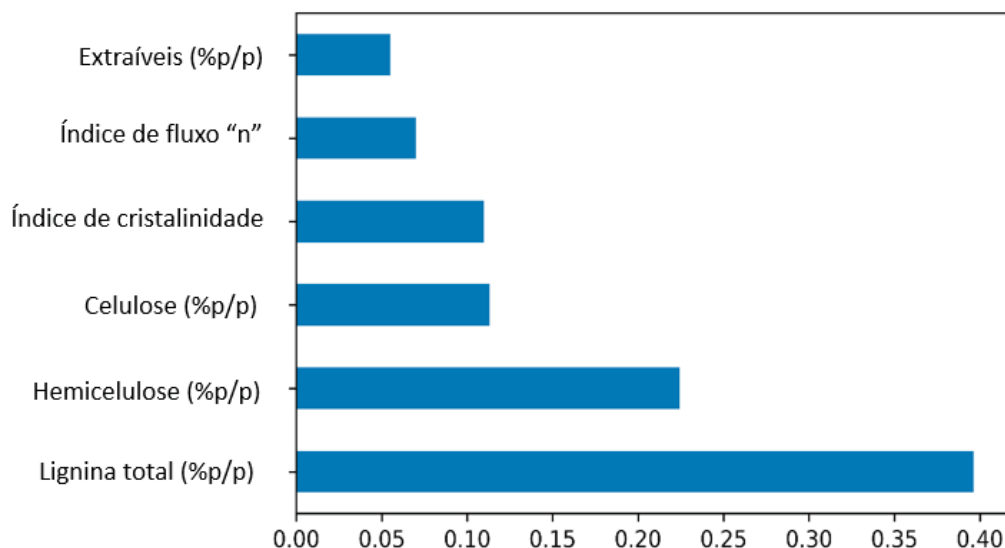
Para este tipo de modelo, mantiveram-se os valores pré-estabelecidos pela função `RandomForestRegressor` do Scikit Learn de todos os hiperparâmetros com exceção do número de árvores e do número mínimo de pontos experimentais a permanecerem em nós do tipo folha. Ajustaram-se estes dois hiperparâmetros somente pois o número de preditores é um dos parâmetros mais importantes para regular a acurácia do modelo e o número mínimo de amostras presentes nas folhas condiciona o crescimento de cada árvore individualmente, gerando modelos mais suaves e controlando possíveis ocorrências de *overfitting* (BREIMAN, 2001; GÉRON, 2019; PROBST; BOULESTEIX, 2018). O número avaliado de árvores foi de 1 a 200 e o número mínimo de amostras de 1 a 5. Na TABELA 7 da seção 3.2.4.1 encontram-se os valores padrão dos hiperparâmetros que não foram ajustados.

Para determinação do número de árvores, foram avaliadas os desempenhos do subconjunto de treino utilizando todas as 11 variáveis de entrada possíveis (descritas previamente na TABELA 30), um número de árvores de 1 a 200 e um número mínimo de pontos em nós do tipo folha de 1 a 5. A partir deste teste foi determinado que 17 estimadores era o número ideal e um mínimo de 2 amostras nas folhas. Estes valores geraram os melhores resultados sem demandar uma estrutura muito grande e computacionalmente dispendiosa, além de reduzir as chances de *overfitting*.

A partir dessas definições, avaliou-se o uso de conjuntos menores de variáveis de entrada para gerar modelos mais simples e baratos. Para isso, levou-se em consideração o resultado da importância relativa das variáveis para um modelo

completo, os coeficientes de Pearson e a facilidade de medição das variáveis. O gráfico que relaciona as importâncias relativas para as 6 variáveis mais relevantes está apresentado no GRÁFICO 26 abaixo.

GRÁFICO 26 - RELAÇÃO DE IMPORTÂNCIA RELATIVA DAS VARIÁVEIS DE ENTRADA PARA O MODELO DE PREDIÇÃO DA RAZÃO DE ASPECTO DA FLORESTA RANDÔMICA.



Fonte: A autora (2023)

Pelo GRÁFICO 26 pode-se observar que as variáveis mais importantes são geralmente componentes químicos das polpas pré-tratadas. É interessante observar que os coeficientes de correlação de Pearson também apresentaram como maiores valores os componentes químicos das polpas, a exemplo do conteúdo de celulose, hemicelulose e lignina total, além do índice de cristalinidade. Portanto, levando-se em consideração este resultado, os valores de coeficientes de correlação de Pearson e buscando-se, para esta etapa preliminar, gerar modelos mais simples com variáveis mais fáceis de medir e mais reprodutíveis, o conjunto de 11 variáveis foi afinado para subconjuntos de 3 a 5.

Foram testados modelos com diversas combinações de variáveis de entrada e seus resultados computados. O modelo que melhor se adequou contou com três variáveis de entrada, sendo (lignina total (%p/p), celulose (%p/p) e consumo energético (kWh/kg)). Os resultados para os subconjuntos de treino e teste estão apresentados na TABELA 31.

TABELA 31 - RESULTADOS APRESENTADOS PARA MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DA RAZÃO DE ASPECTO.

Nº de variáveis de entrada utilizadas	Tipo de dado	EAM	EPAM	RQEM	R ²
3	Treino	10,019	9,02	14,237	0,928
	Teste	8,723	5,37	15,265	0,941

Fonte: A autora (2023)

Observa-se que o modelo gerado obteve valores de R² bons, acima de 0,9 e erros absolutos relativamente baixos. Cabe ressaltar que estes subconjuntos de variáveis compreendeu combinações daquelas mais importantes indicadas no GRÁFICO 26 e na TABELA 29, mas também com outras variáveis que não estavam presentes, por exemplo, o consumo energético na etapa de fibrilação. Este parâmetro, apesar de não configurar como um dos mais importantes ou possuir correlação de Pearson maior que 0,7, adentrou no conjunto de variáveis de entrada testados pois é um parâmetro que representa o processo, é de fácil medição e gerou bom desempenho.

Redes Neurais Artificiais

Para as redes neurais artificiais preliminares foi seguida a mesma lógica de gerar modelos menos complexos. Assim, fixaram-se a maioria dos hiperparâmetros e modificaram-se três. Os hiperparâmetros fixados foram a taxa de aprendizagem igual a 0,01, o uso do algoritmo de otimização Adam, a aplicação de uma única camada oculta, o número de épocas estabelecido no máximo em 700 e o uso da função de ativação linear para a camada de saída.

Os hiperparâmetros que foram modificados incluíram o tamanho da batelada, que foi avaliado com valores entre 2 a 16, o número de neurônios na camada oculta e as funções de ativação desta camada (escolhidas entre sigmoideal e ReLU). Ressalta-se que o número de neurônios na camada escondida foi alterado entre o mesmo número de neurônios na camada de entrada e o dobro desta quantidade. O número de neurônios na primeira camada da rede mudava de acordo com o subconjunto de variáveis de entrada testados. Para testar estes hiperparâmetros foi

aplicada uma procura manual, uma vez que não eram muitos hiperparâmetros a serem investigados, e desta maneira foram avaliadas algumas combinações deles.

Na TABELA 32 estão apresentados os resultados obtidos para algumas das principais arquiteturas de redes neurais testadas a fim de comparação.

TABELA 32 - RESULTADOS DAS REDES NEURAIS GERADAS COM DIFERENTES COMBINAÇÕES DE HIPERPARÂMETROS PARA A RAZÃO DE ASPECTO.

Modelo N°	N° de variáveis de entrada utilizadas	N° de neurônios na camada oculta	Lote	Função de ativação	Tipo de dado	EAM	EPAM	RQEM	R ²
1	11	22	2	Sigmoidal	Treino	1,576	1,23	3,887	0,995
					Teste	6,488	6,64	7,283	0,987
2	5	10	2	Sigmoidal	Treino	5,760	4,10	11,020	0,957
					Teste	8,544	6,29	12,610	0,960
3	5	10	2	ReLU	Treino	7,696	8,31	12,242	0,947
					Teste	11,010	10,22	13,472	0,954
4	5	10	10	Sigmoidal	Treino	27,215	16,35	51,07	0,072
					Teste	34,519	20,51	61,00	0,057
5	4	8	2	Sigmoidal	Treino	10,209	6,65	19,733	0,861
					Teste	16,431	11,27	26,374	0,824

Fonte: A autora (2023)

Os cinco modelos apresentados diferem em relação ao número de variáveis de entrada, número de neurônios na camada oculta, ao tamanho do lote e à função de ativação intermediária. O modelo 1 corresponde a um modelo gerado com todas as variáveis de entrada possíveis para as polpas BTMP_mec, BKSP_mec, UKSP_mec e TMP_mec. O valor de tamanho do lote foi mantido baixo e a função utilizada foi sigmoidal. O desempenho deste modelo foi excepcional, com erros baixos e valores de R² altos. Contudo, pode-se observar uma certa discrepância entre os valores de EAM, EPAM e RQEM dos subconjuntos de treino e teste, o que pode estar indicando um certo nível de *overfitting* para este modelo.

O modelo 2 também utiliza o mesmo número de tamanho do lote e função de ativação intermediária que o modelo 1, contudo, utiliza 5 variáveis de entrada e 10 neurônios na camada oculta. Estas 5 variáveis testadas foram hemicelulose (%p/p),

celulose (%p/p), extraíveis (%p/p), lignina total (%p/p) e consumo energético (kWh/kg). Este modelo gerou resultados muito satisfatórios também. Apesar do desempenho ter sido um pouco pior que o modelo com 11 variáveis de entrada, os erros não foram excessivamente altos e os valores de R^2 estão acima de 0,9. Ademais, as métricas de treino e teste ficaram mais parecidas, não indicando possíveis *overfitting* nem *underfitting*.

O modelo 3 possui quase toda a mesma arquitetura que o modelo 2, com exceção da função de ativação, onde foi testada a ReLU. Neste caso, os valores de R^2 de treino e teste foram bem próximos que os do modelo 2, porém, as métricas de erro foram um pouco maiores, principalmente para EAM e EPAM.

No modelo 4 foi testada uma arquitetura igual ao modelo 2, porém com um tamanho de lote maior, no valor de 10. O desempenho do modelo não foi satisfatório, e por tanto, mantiveram-se valores de tamanho do lote menores para as arquiteturas.

Por fim, o modelo 5 apresentado contou com 4 variáveis de entrada - hemicelulose (%p/p), celulose (%p/p), lignina total (%p/p) e consumo energético (kWh/kg) -, oito neurônios na camada oculta, tamanho de lote de 2 e função de ativação sigmoideal. Apresentou valores de R^2 mais inferiores em relação aos modelos apresentados e erros muito maiores, principalmente para o subconjunto de teste. Além disso, pode-se observar uma discrepância grande entre as métricas de treino e teste, o que pode estar indicando *overfitting*.

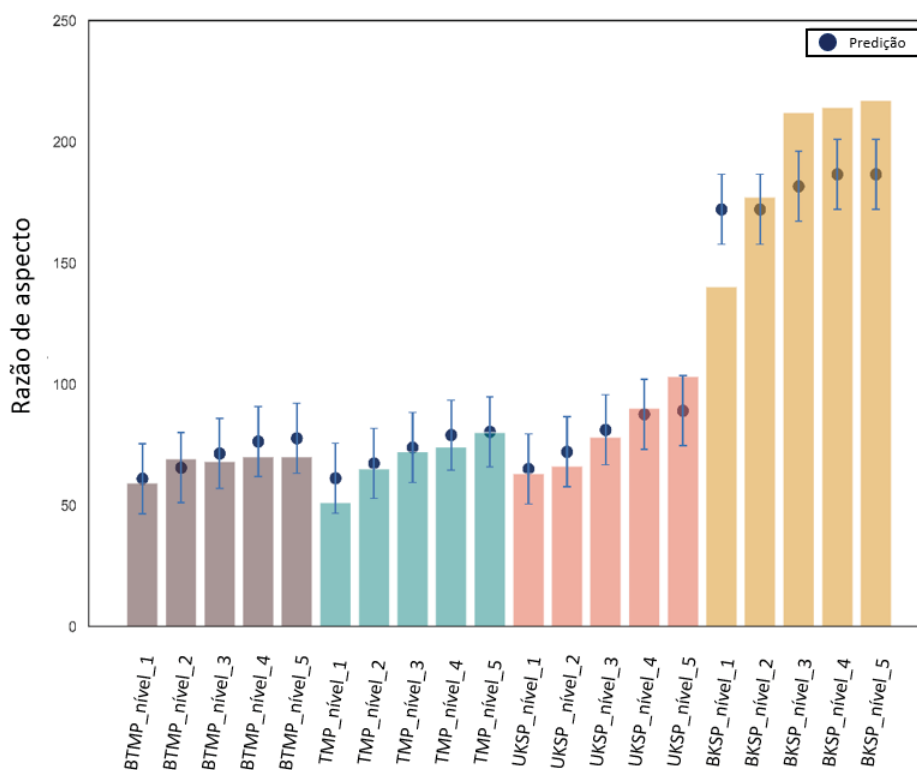
Desempenho comparativo entre os diferentes algoritmos de aprendizagem de máquina

A partir dos modelos considerados melhores para cada algoritmo, foram gerados gráficos comparativos dos resultados preditos de razão de aspecto (representados por círculos) e os valores reais da propriedade (representados pelas barras). No eixo das abscissas estão apresentadas as amostras derivadas de cada grupo de polpa e ordenadas de acordo com o nível crescente de energia aplicado ao sistema de fibrilação, sendo a notação “nível_1” correspondente ao de menor energia e “nível_5” ao de maior energia, conforme os valores apresentados no QUADRO 5.

O GRÁFICO 27 mostra os valores preditos pelo modelo de floresta randômica. Pode-se observar que os valores de predição e os reais estão próximos para todos os

grupos de amostras, com exceção do BKSP. Isto pode ter ocorrido devido ao fato de existir neste conjunto de dados preliminar somente 5 pontos com razões de aspecto mais altos, no intervalo de 140 a 217. Ademais, percebe-se que a razão de aspecto predita acompanha a tendência de crescimento conforme o nível de energia aplicado para gerar cada amostra aumentou.

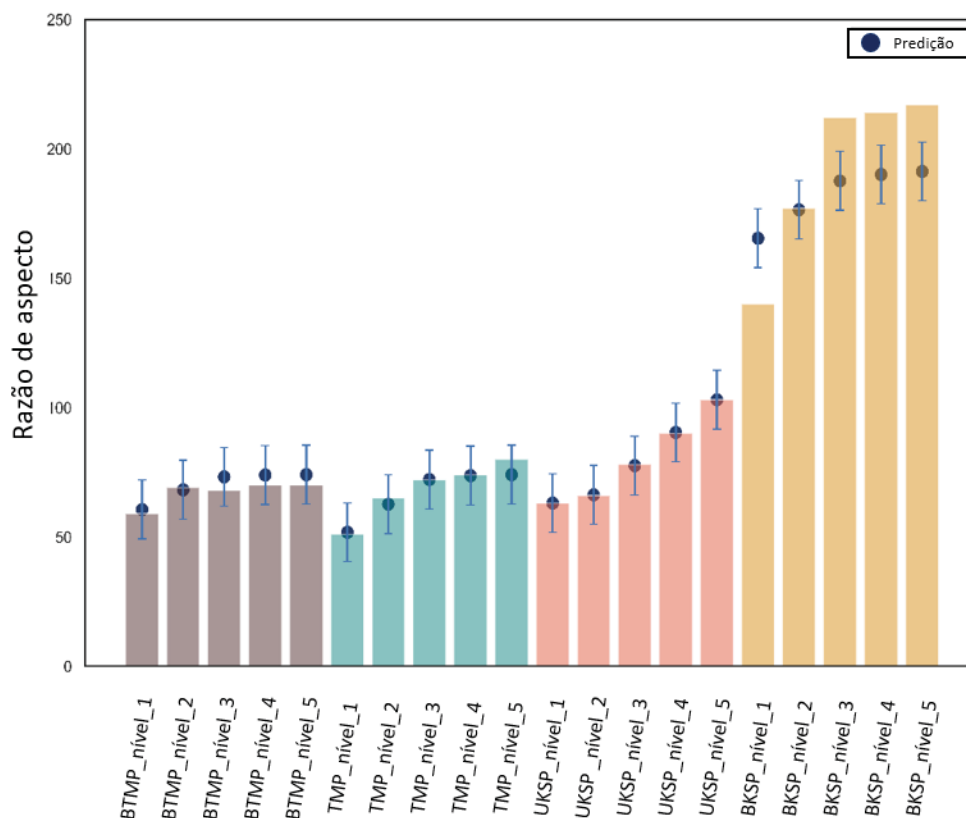
GRÁFICO 27 - VALORES PREDITOS VERSUS VALORES REAIS DE RAZÃO DE ASPECTO PARA A FLORESTA RANDÔMICA.



Fonte: A autora (2023)

O GRÁFICO 28 mostra os valores preditos pelo modelo 2 das redes neurais, ou seja, o que obteve melhor desempenho com um número de variáveis de entrada mais otimizados. Pode-se observar que os valores de predição e reais estão muito próximos para todos os grupos de amostras, com exceção do BKSP. A explicação para isto é a mesma apresentada para o modelo de floresta randômica, uma vez que a quantidade de pontos com razões de aspecto maiores é limitada nesta primeira etapa de estudo.

GRÁFICO 28 - VALORES PREDITOS VERSUS VALORES REAIS DE RAZÃO DE ASPECTO PARA A REDE NEURAL 2.



Fonte: A autora (2023)

A observação dos gráficos acima permite constatar que os modelos gerados conseguiram captar a tendência de aumento da razão de aspecto com o aumento da aplicação de energia na etapa de fibrilação, apesar de alguns modelos terem conseguido modelar melhor essa tendência que outros.

Ademais os modelos gerados foram simples, com poucas variáveis, obtendo bons desempenhos a partir de variáveis de entrada de fácil aquisição relacionadas à composição química das polpas pré-tratadas, além do consumo energético na etapa de fibrilação. A determinação dos teores de celulose, hemicelulose, lignina total e extraíveis não requer o uso de técnicas cromatográficas muito complexas, sendo possível determiná-las a partir das metodologias gravimétricas descritas pela *Technical Association of the Pulp and Paper Industry* (TAPPI). Além disso, o consumo energético do equipamento de fibrilação pode ser facilmente lido em linha.

Foram geradas métricas para o conjunto total de dados (ou seja, incluindo o treino e teste) a fim de facilitar a comparação entre os algoritmos testados. Estes

resultados estão reunidos na TABELA 33 abaixo. Comparando-se os valores, destaca-se que todos os modelos obtiveram desempenhos muito satisfatórios, com coeficientes de correlação de R^2 acima de 0,9 e erros médios percentuais abaixo de 10%. Apesar disso, comparativamente, o modelo que gerou os melhores resultados foram as redes neurais, com erros médios percentuais abaixo de 5% e R^2 de aproximadamente 0,96.

TABELA 33 - MÉTRICAS DE DESEMPENHO PARA OS MODELOS CONSIDERADOS DE CADA ALGORITMO TESTADO PARA PREDIÇÃO DA RAZÃO DE ASPECTO.

Modelo	EAM	EPAM	RQEM	R^2
Floresta randômica	9.760	8.18%	14.448	0.9313
Rede neural	6.317	4.54%	11.356	0.9575

Fonte: A autora (2023)

APÊNDICE D – RESULTADOS DA MODELAGEM PRELIMINAR DE PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DE CNFS

Avaliação preliminar dos dados

As variáveis de entrada consideradas para os modelos preliminares de predição da razão de aspecto elencadas no APÊNDICE C foram utilizadas para geração dos modelos de floresta randômica e rede neural para predição do rendimento de nanofibrilação. Os valores absolutos dos coeficientes de correlação de Pearson estão apresentados na TABELA 34. A descrição estatística da variável de saída deste caso está apresentada na TABELA 35. As variáveis de entrada foram omitidas pois suas descrições já estão presentes na TABELA 30 do APÊNDICE C.

TABELA 34 - VALORES ABSOLUTOS DOS COEFICIENTES DE CORRELAÇÃO DE PEARSON PARA AS VARIÁVEIS DE ENTRADA EM RELAÇÃO AO RENDIMENTO DE NANOFIBRILAÇÃO.

Variável de entrada	Coefficiente de correlação de Pearson
Transmitância a 600 nm (%)	0,9363
Demanda catiônica ($\mu\text{eq/g}$)	0,8966
Consumo energético (kWh/kg)	0,7703
Índice de consistência (k)	0,5706
Cinzas (%p/p)	0,4530
Índice de fluxo (n)	0,4168
Índice de cristalinidade	0,3425
Celulose (%p/p)	0,3105
Lignina total (%p/p)	0,2845
Hemicelulose (%p/p)	0,2546
Extraíveis (%p/p)	0,0779

Fonte: A autora (2023)

TABELA 35 - DESCRIÇÃO ESTATÍSTICA DA VARIÁVEL RENDIMENTO DE NANOFIBRILAÇÃO.

Variável	Contagem de valores não nulos	Valor médio	Desvio padrão	Valor mínimo	25%	50%	75%	Valor máximo
Rendimento de nanofibrilação (variável de saída)	20	0,102	0,049	0,029	0,072	0,090	0,134	0,206

Fonte: A autora (2023)

Desempenho dos modelos

O mesmo subconjunto de dados de treino e teste utilizado da seção anterior foi utilizado para modelar o rendimento de nanofibrilação. Os resultados para o cada algoritmo estão apresentados nas subseções abaixo.

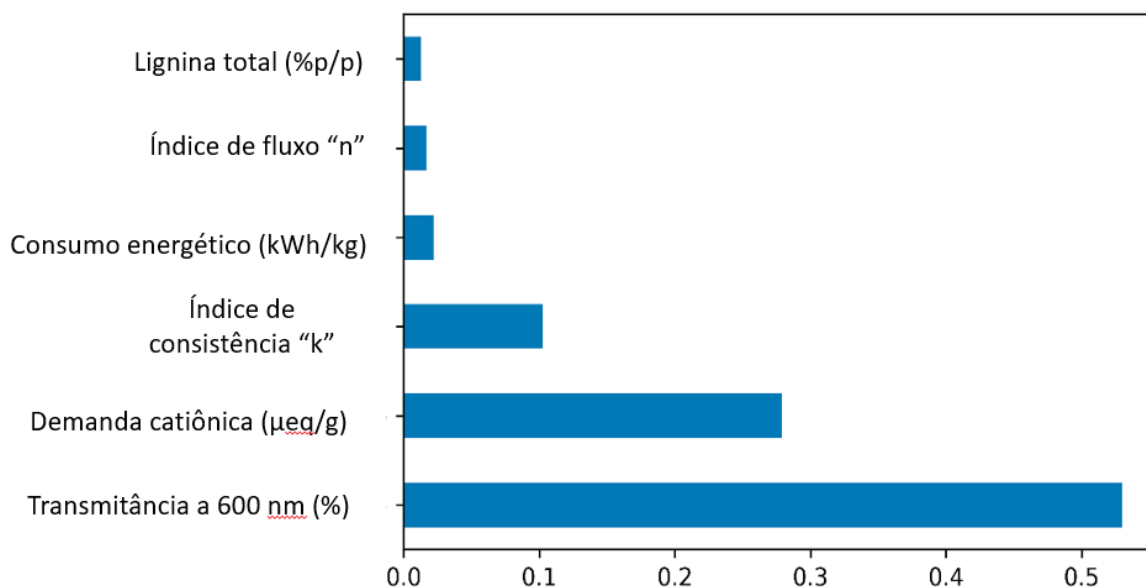
Ademais, as mesmas considerações realizadas para a modelagem da razão de aspecto utilizando os dois algoritmos mencionados são feitas para a modelagem do rendimento de nanofibrilação. Eventuais diferenças nas decisões tomadas acerca da definição de hiperparâmetros ou outro aspecto relacionado aos algoritmos serão explicitados.

Florestas Randômicas

Seguindo a mesma lógica para a definição do número de árvores e pontos em nós do tipo folha para a predição da razão de aspecto, foram determinados 40 estimadores e um número mínimo de 2 amostras nos nós do tipo folha. Estes valores de hiperparâmetros geraram os melhores resultados sem demandar uma estrutura muito grande e computacionalmente dispendiosa e com menores riscos de *overfitting*.

O gráfico que relaciona as importâncias relativas para 6 variáveis mais relevantes está apresentado no GRÁFICO 29.

GRÁFICO 29 - RELAÇÃO DE IMPORTÂNCIA RELATIVA DAS VARIÁVEIS DE ENTRADA PARA O MODELO DE PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO DA FLORESTA RANDÔMICA.



Fonte: A autora (2023)

Para o rendimento de nanofibrilação, as variáveis mais importantes estão relacionadas principalmente às características finais da nanocelulose e percebe-se que a importância se concentra principalmente em três variáveis: transmitância a 600 nm, demanda catiônica e índice de consistência "k". Aqui também se percebe que os dois maiores coeficientes de correlação de Pearson coincidem com os apresentados no GRÁFICO 29. Portanto, levando-se em consideração este resultado, os valores de coeficientes de correlação de Pearson e geração de modelos mais simples, o conjunto de 11 variáveis foi afinado para subconjuntos de 3 a 5 variáveis de entrada, seguindo a mesma lógica aplicada para os modelos de razão de aspecto.

A TABELA 36 apresenta os resultados gerados para o modelo de predição de rendimento de nanofibrilação com florestas randômicas escolhido.

TABELA 36 - RESULTADOS APRESENTADOS PARA MODELO DE FLORESTA RANDÔMICA PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.

Tipo de dado	EAM	EPAM	RQEM	R ²
Treino	0,0136	17,44	0,0172	0,858
Teste	0,0135	34,19	0,0167	0,884

Fonte: A autora (2023)

O desempenho da floresta randômica para o rendimento de nanofibrilação não gerou métricas tão boas quanto as que foram possíveis de obter para a razão de aspecto. Ainda assim, os valores de R² dos modelos em geral foram satisfatórios. Os erros absolutos foram em torno de 1,4 pontos percentuais, tanto para o treino quanto para o teste. Percebe-se que o valor do erro percentual médio para o subconjunto de teste foi muito alto, na ordem de 30%. Estes tipos de valores não são desejáveis, contudo, vale ressaltar o fato de que 20% dos dados totais que fizeram parte do subconjunto de teste correspondem a somente 4 pontos experimentais na prática. Desta maneira, as métricas são mais afetadas quando algum dos pontos não é predito de forma acurada.

O modelo requereu a demanda catiônica, a transmitância e o consumo de energia como variáveis de entrada para gerar os melhores resultados de desempenho. Novamente, para os modelos de predição do rendimento de nanofibrilação manteve-se este último parâmetro nas simulações pois é uma variável de processo de fácil obtenção.

Redes Neurais Artificiais

Para as redes neurais artificiais preliminares foi seguida a mesma lógica de gerar modelos menos complexos, da mesma forma que foi aplicado para a razão de aspecto. Os hiperparâmetros fixados foram os mesmos, com os mesmos valores. Os hiperparâmetros modificados também foram iguais e incluíram o tamanho da batelada, o número de neurônios na camada oculta e as funções de ativação desta camada (escolhidas entre sigmoidal e ReLU). Para testar estes hiperparâmetros foi aplicada também uma procura manual.

Devido ao número extenso de combinações realizadas, estão apresentadas na TABELA 37 os resultados obtidos para algumas das principais arquiteturas de redes neurais geradas a fim de comparação.

TABELA 37 – PRINCIPAIS ARQUITETURAS DE REDES NEURAIS GERADAS COM DIFERENTES COMBINAÇÕES DE HIPERPARÂMETROS PARA O RENDIMENTO DE NANOFIBRILAÇÃO.

Modelo N°	N° de variáveis de entrada utilizadas	N° de neurônios na camada oculta	Lote	Função de ativação	Tipo de dado	EAM	EPAM	RQEM	R ²
1	11	11	16	ReLU	Treino	0,0056	6,939	0,0068	0,978
					Teste	0,0095	16,481	0,0098	0,960
2	5	5	16	ReLU	Treino	0,0041	5,412	0,0062	0,981
					Teste	0,0047	9,190	0,0048	0,990
3	4	4	16	ReLU	Treino	0,0057	6,035	0,0076	0,972
					Teste	0,0059	9,544	0,0073	0,978
4	4	4	2	ReLU	Treino	0,0151	17,852	0,0190	0,830
					Teste	0,0230	51,385	0,0263	0,713
5	4	4	16	Sigmoid	Treino	0,0112	13,165	0,0148	0,895
					Teste	0,0164	36,422	0,0179	0,867
6	4	4	2	Sigmoid	Treino	0,0157	21,017	0,0187	0,833
					Teste	0,0126	20,044	0,0179	0,867

Fonte: A autora (2023)

Os seis modelos apresentados diferem em relação ao número de variáveis de entrada, número de neurônios na camada oculta, ao tamanho de lote e a função de ativação intermediária. A menor razão neurônios camada entrada: neurônios camada oculta que gerou os melhores resultados foi a de 1:1. Partindo de uma análise holística, percebe-se que neste caso a função de ativação ReLU se mostrou mais adequada para esta situação e um *batch size* maior gerou melhores resultados.

O modelo 1 corresponde a um modelo gerado com todas as variáveis de entrada possíveis, tamanho de lote de 16 e função de ativação ReLU. O desempenho deste modelo foi muito bom, com erros baixos e valores de R² altos, a exceção do erro percentual médio do subconjunto de teste. Contudo, percebe-se uma certa

discrepância entre as métricas de erro de treino e teste, apesar de terem sido boas no geral para os dois subconjuntos de dados.

O modelo 2 também utiliza o mesmo número de tamanho de lote e função de ativação intermediária que o modelo 1, contudo, utiliza 5 variáveis de entrada e 5 neurônios na camada oculta. Estas 5 variáveis testadas foram: demanda catiônica, transmitância, consumo de energia, índice de consistência (k) e extraíveis. Este modelo gerou resultados excelentes também. Os valores de R^2 de treino e teste são muito próximos de 1 e as métricas de treino e teste são mais baixas que o do modelo 1. Porém, há uma certa diferença principalmente entre o EPAM e o RQEM.

No modelo 3 foram avaliadas quase todas as mesmas cinco variáveis citadas anteriormente, com exceção do teor de extraíveis. Neste caso, os valores de R^2 , apesar de um pouco menores, foram bem próximos ao do modelo de 5 variáveis. Além disso, as outras métricas de treino e teste foram satisfatórias e ficaram mais similares do que no caso do modelo 4.

Com a arquitetura do modelo 4 foi avaliado o uso de um tamanho de lote pequeno, no valor de 2, mantendo-se a mesma estrutura do modelo 3. Observa-se que o tamanho de lote menor prejudicou o desempenho do modelo, principalmente em relação às métricas do subconjunto de teste.

No modelo 5 foi testada a mesma arquitetura do modelo 3, mas com outra função de ativação (a sigmoide) e no caso do modelo 6, também foi testado um valor de tamanho de lote menor. Os dois modelos apresentaram desempenhos piores que os modelos que utilizaram função de ativação ReLU. O modelo 6 teve um desempenho pior em relação ao 5 também, principalmente para as métricas de treino, o que demonstra novamente o fato de um tamanho de lote menor não ter sido benéfico neste caso.

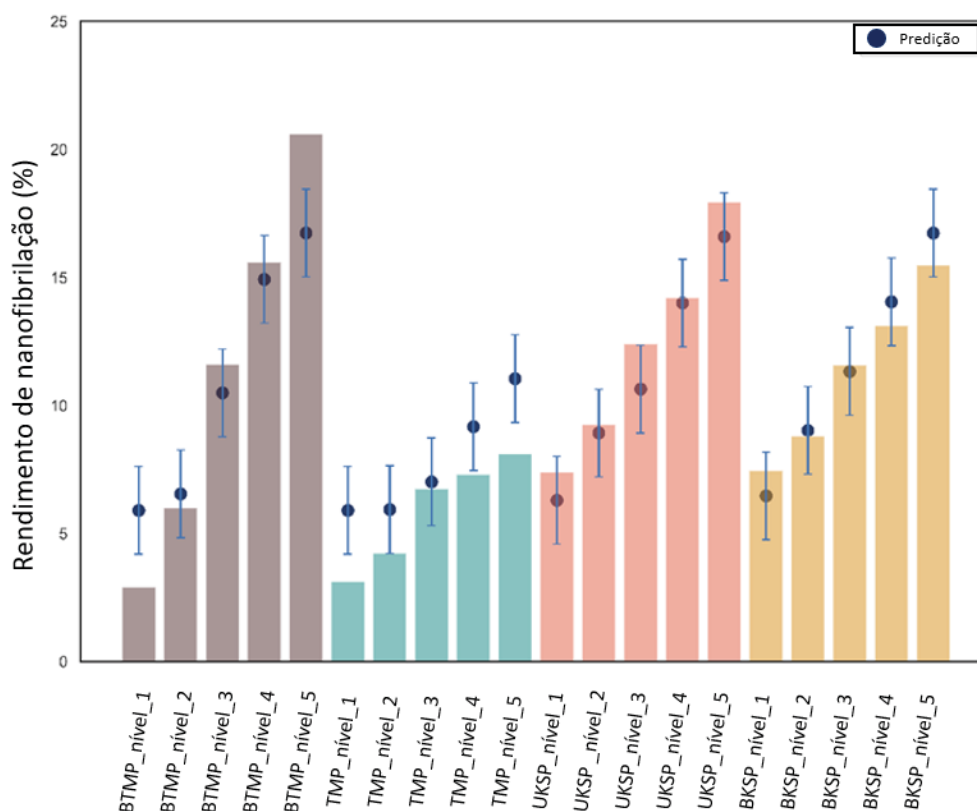
Desempenho comparativo entre os diferentes algoritmos de aprendizagem de máquina

A partir dos modelos considerados melhores para cada algoritmo, foram gerados gráficos comparativos dos resultados preditos de rendimento de nanofibrilação (representados por círculos) e os valores reais da propriedade (representados pelas barras). No eixo das abscissas estão apresentadas as amostras

derivadas de cada grupo de polpa e ordenadas de acordo com o nível crescente de energia aplicado ao sistema de fibrilação, sendo a notação “nível_1” correspondente ao de menor energia e “nível_5” ao de maior energia, da mesma forma explicada para a predição da razão de aspecto.

O GRÁFICO 30 mostra os valores preditos pelo modelo de floresta randômica. O rendimento de nanofibrilação também é afetado pela quantidade de energia aplicada no sistema, e apesar da falta de precisão na predição para algumas amostras, o modelo consegue captar as tendências de aumento do rendimento com o aumento do consumo energético. Dentre os algoritmos avaliados, a floresta randômica foi a que performou pior para predizer esta propriedade. Um dos grupos de amostras menos favorecidos foi o TMP, onde o modelo superestimou todos os pontos. Isto pode ter ocorrido pois é o único grupo que não obtém o mesmo padrão e valores de rendimento de nanofibrilação que os outros três.

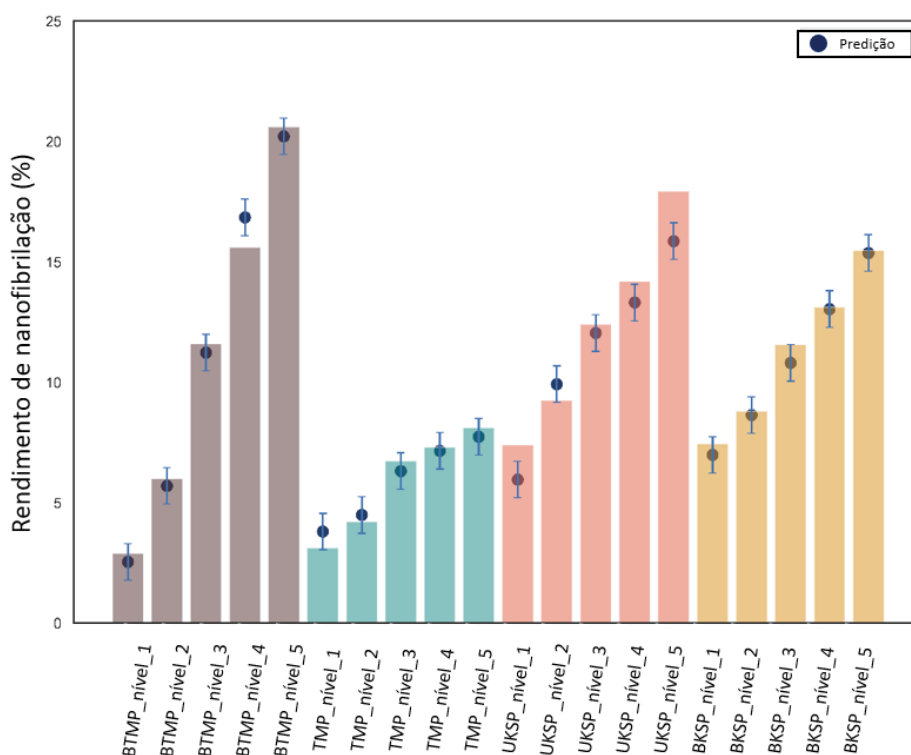
GRÁFICO 30 - VALORES PREDITOS VERSUS VALORES REAIS DE RENDIMENTO DE NANOFIBRILAÇÃO PARA A FLORESTA RANDÔMICA.



Fonte: A autora (2023)

O GRÁFICO 31 mostra os valores preditos pelo modelo 3 das redes neurais, o que obteve melhor desempenho ao avaliar-se as métricas de treino e teste. Novamente, neste caso a rede neural foi o algoritmo que obteve as melhores métricas em relação ao conjunto de algoritmos testados. Os erros apresentados são menores que os da floresta randômica. Ademais, conseguiu prever muito bem o grupo TMP, diferentemente dos outros modelos.

GRÁFICO 31 - VALORES PREDITOS VERSUS VALORES REAIS DE RENDIMENTO DE NANOFIBRILAÇÃO PARA A REDE NEURAL 3.



Fonte: A autora (2023)

De forma geral, os modelos gerados para o rendimento de nanofibrilação foram simples, obtendo bons desempenhos a partir de variáveis de entrada de fácil aquisição. Neste caso, as variáveis estão mais relacionadas às propriedades finais da nanocelulose. O consumo energético do equipamento de fibrilação pode ser facilmente lido em linha. Outros parâmetros baratos e facilmente medidos são a demanda catiônica, que pode ser medido com o uso de potenciômetros. A transmitância pode ser facilmente medido com o uso de fotômetros para detectar a perda na intensidade da luz incidente (SERRA-PARAREDA *et al.*, 2021b; XU *et al.*,

2016). Por sua vez, a medida da viscosidade requer o uso de um reômetro e o teor de extraíveis pode ser medido através de técnicas gravimétricas da TAPPI.

Para o rendimento de nanofibrilação também foram geradas métricas para o conjunto total de dados (ou seja, incluindo o treino e teste) a fim de facilitar a comparação entre os algoritmos testados. Estes resultados estão reunidos na TABELA 38 abaixo. Avaliando os algoritmos, neste caso a floresta randômica teve o pior desempenho entre os modelos, com um erro médio percentual alto principalmente. A rede neural mostrou ser o melhor modelo para este caso também, com um EPAM abaixo de 10% e um R^2 de aproximadamente 0,98. Destaca-se para este algoritmo que o EAM e RQEM ficaram abaixo de 1 ponto percentual.

TABELA 38 - MÉTRICAS DE DESEMPENHO PARA OS MODELOS CONSIDERADOS DE CADA ALGORITMO TESTADO PARA PREDIÇÃO DO RENDIMENTO DE NANOFIBRILAÇÃO.

Modelo	EAM	EPAM	RQEM	R²
Floresta randômica	1,358%	20,79%	1,712%	0,8720
Rede neural	0,577%	6,74%	0,758%	0,9749

Fonte: A autora (2023)