

REDE NEURAL ARTIFICIAL PARA O EQUILÍBRIO LÍQUIDO-LÍQUIDO DA MISTURA GLICEROL-METANOL-BIODIESEL DE ÓLEO DE CÁRTAMO¹

ARTIFICIAL NEURAL NETWORK FOR LIQUID-LIQUID EQUILIBRIUM OF GLYCEROL-METHANOL-SAFFLOWER BIODIESEL MIXTURE

Richard da Silva Cunha², Gisele Atsuko Medeiros Hirata³ e Tiago Dias Martins⁴

RESUMO

A determinação do Equilíbrio Líquido-Líquido (ELL) na produção de Biodiesel é essencial para o projeto e monitoramento do processo de produção. Assim, a procura por ferramentas de modelagem mais eficientes para esse fim é crucial, já que os modelos termodinâmicos clássicos possuem limitações intrínsecas. Com isso, as redes neurais artificiais (RNAs) vêm ganhando espaço por sua facilidade, capacidade de generalização e abrangência, precisão nos resultados, etc. Tendo isso em vista, o objetivo desse trabalho foi obter duas RNAs para tratar do ELL da mistura glicerol-metanol-biodiesel de óleo de cártamo à 298,15 K e 101,325 kPa. Uma capaz de classificar se a mistura está na região bifásica de ELL ou de fase líquida homogênea e outra capaz de calcular a composição das fases líquidas em equilíbrio (rica em biodiesel e rica em glicerol). Ambos os modelos consideraram a composição global da mistura como entrada. Os resultados obtidos se mostraram eficazes, sendo que o primeiro modelo foi capaz de classificar corretamente as fases da mistura com 98 % de precisão e o segundo calculou as composições em equilíbrio com 4,98 % de erro (um valor inferior ao obtido utilizando-se o método UNIQUAC).

Palavras-chave: Equilíbrio de Fases; Inteligência Artificial; Estabilidade de Fases; Biocombustíveis.

ABSTRACT

The determination of the Liquid-Liquid Equilibrium (LLE) in biodiesel production is essential for the design and monitoring of the production process. Therefore, the search for efficient modeling tools is crucial, since classical thermodynamic models have intrinsic limitations. Artificial neural networks (ANNs) are gaining ground due to their ease of use, generalization capacity, accuracy of results, etc. Within this context, the objective of this work was to obtain two ANNs to deal with the LLE of the glycerol-methanol-safflower oil biodiesel mixture at 298.15 K and 101.325 kPa. The aim of the first model was to classify whether the mixture is in the biphasic LLE region or in a homogeneous liquid phase, and the other one to calculating the composition of the liquid phases in equilibrium (rich in biodiesel and rich in glycerol). Both models considered the overall composition of the mixture as input. The obtained results showed that, with the first model was able to correctly classify the phases of the mixture with 98% accuracy, and the second one predicted the equilibrium compositions with a 4.98% error (lower than the one obtained using the UNIQUAC method).

Keywords: Phase Equilibrium; Artificial Intelligence; Phase stability; Biofuels.

¹ Parte de trabalho de mestrado

² Universidade Federal de São Paulo - UNIFESP. E-mail: rcunha@unifesp.br. ORCID: <https://orcid.org/0009-0000-1720-0170>

³ Universidade Federal de São Paulo - UNIFESP. E-mail: gisele.hirata@unifesp.br. ORCID: <https://orcid.org/0000-0002-7652-8667>

⁴ Universidade Federal de São Paulo - UNIFESP. Email: tdmartins@unifesp.br. ORCID: <https://orcid.org/0000-0002-3452-7033>

1 INTRODUÇÃO

A produção de biodiesel envolve reações de transesterificação entre um álcool de cadeia curta e uma fonte de moléculas de triglicerídeos, podendo ser óleo de origem vegetal, gordura animal etc. (Meneghetti *et al.*, 2013).

A obtenção do biodiesel em escala industrial, por sua vez, compreende operações unitárias importantes, como a separação dos produtos da reação: como o glicerol e o biodiesel (McCormick; Moriarty, 2023). Essa etapa normalmente é realizada empregando a extração líquido-líquido, que explora a afinidade de uma substância com um solvente a fim de separá-la da substância de interesse. Por isso, há a necessidade de prever se o sistema formado entre os produtos e o solvente apresenta separação de fases, a fim de verificar a viabilidade de sua separação, bem como as composições obtidas no equilíbrio.

Em geral, o sistema é considerado como uma mistura ternária. Assim, o equilíbrio entre fases líquida pode ser descrita pela igualdade de fugacidades ($\hat{f}$) de cada substância entre as fases (Sandler, 2017):

$$\hat{f}_{\text{glicerol}}^I = \hat{f}_{\text{glicerol}}^{II} \quad (1)$$

$$\hat{f}_{\text{alcohol}}^I = \hat{f}_{\text{alcohol}}^{II} \quad (2)$$

$$\hat{f}_{\text{biodiesel}}^I = \hat{f}_{\text{biodiesel}}^{II} \quad (3)$$

Para uma fase líquida, a fugacidade de cada substância pode ser obtida a partir de um modelo de energia livre de Gibbs de excesso. Usualmente, utilizam-se os modelos UNIFAC, UNIQUAC e NRTL, e Wilson para essa tarefa (Sandler, 2017). Porém, elas possuem limitações intrínsecas que diminuem a abrangência e aplicabilidade, tais como: necessidade de grande espaço computacional, cálculos mais demorados, aplicabilidade limitada devido à simplificações etc. De acordo com Misikov *et al.* (2022), para sistemas complexos é difícil encontrar uma solução analítica que os descrevam, já que é necessário levar em conta inúmeros fatores internos e externos, além de interações específicas entre os componentes que são difíceis de descrever.

Bazooyar *et al.* (2021), em seu estudo na obtenção do ELL contendo Biodiesel de diferentes origens (óleo de Canola, Milho, Girassol etc.), Glicerol e Etanol, citaram que os modelos termodinâmicos UNIFAC, NRTL e UNIQUAC são mais comuns para esses sistemas e geram bons resultados, entretanto, são limitados às condições específicas de temperatura e pressão, não predizendo bem dados fora dessas condições. Es' Haghi *et al.* (2024) relataram que o modelo NRTL é uma ferramenta que pode ser utilizada para modelagem de sistemas não-ideais, mas enfrenta desafios, pois é necessário determinação de dados experimentais para a obtenção de seus parâmetros e isso envolve a resolução de equações altamente não-lineares, com isso tendo dificuldades para prever situações em

que o sistema esteja num estado de temperatura (T) e pressão (P) críticos, por exemplo. Já Padilha *et al.* (2016) mostra a complexidade em prever sistemas contendo biomoléculas, pois envolve interações moleculares (forças de Van de Waals, forças de hidrogênio etc.) e os modelos termodinâmicos possuem limitações na determinação de sistemas envolvendo proteínas.

Com isso, as Redes Neurais Artificiais (RNAs) têm ganhado relevância, pois possuem um grande potencial de aplicação e agilidade na determinação de resultados. Segundo Beigzadeh *et al.* (2012), redes neurais artificiais são mais adequadas para obtenção de funções mais complexas. Já, Yatim *et al.* (2024) mencionaram os benefícios da utilização de RNAs, que inclui o baixo custo, rapidez nos cálculos e alta acurácia nos resultados. Gao e Loney (2002) destacaram que para aplicação dos modelos termodinâmicos, deve existir uma grande biblioteca de dados, pois os mesmos parâmetros não podem ser aplicados para sistemas/misturas diferentes. E as redes neurais são aplicadas justamente para driblar esses problemas, pois possuem uma estrutura adaptável para diferentes sistemas.

RNAs são sistemas compostos por um conjunto de unidades básicas denominadas neurônios artificiais, organizados em camadas. Esses neurônios tem a função de receber informações em suas entradas, ponderá-las, processá-las e gerar um valor de saída que é transmitido para os neurônios que estão na camada seguinte até a camada de saída, cujo valor esperado é a resposta desejada (Haykin, 2009).

As RNAs já foram aplicadas com sucesso para o cálculo do ELL de diversos sistemas, não se limitando apenas à determinação de parâmetros como composição no equilíbrio, mas também, parâmetros como viscosidade, densidade, condutividade térmica, coeficiente de partição etc. (Melo *et al.*, 2020; Ottaiano *et al.*, 2021; Kim; Moon, 2023; Pierantozzi; Hosseini, 2024). Entretanto, uma aplicação interessante que foi encontrada, foi a resolução do problema de estabilidade da fase, realizada por Schmitz *et al.* (2006) e Pandharipande e Moharkar (2012).

O estudo do equilíbrio de fases em termodinâmica se concentra em duas abordagens: o cálculo das composições em equilíbrio e a determinação da estabilidade de fases, que consiste em determinar se o sistema em questão apresenta uma ou mais fases estáveis (Michelsen, 1982; Sandler, 2017). A determinação da estabilidade de fases é de suma importância na indústria, visto que há oscilações nas composições das misturas, e, com isso, torna-se importante se conhecer quais as composições que geram duas fases, propiciando assim, a separação da substância de interesse (Michelsen, 1982).

Tendo isso em vista, este trabalho teve como objetivo obter duas arquiteturas de RNA. A primeira visando a determinação da estabilidade de fases, ou seja, determinar se, a dada temperatura, pressão e composição global, a mistura existirá em uma fase homogênea ou se haverá diferentes fases líquidas em equilíbrio. A segunda, para o cálculo das composições de cada substância no ELL da mesma mistura, nas mesmas condições de T e P.

O objeto de estudo deste trabalho é o sistema glicerol-metanol-biodiesel de óleo de cártamo. O motivo desta escolha foi o interesse pela produção de biodiesel utilizando essa matéria-prima, largamente disponível no Brasil e que apresenta ótimo custo-benefício (Rios *et al.*, 2023).

2 MÉTODOS

2.1 COLETA DE DADOS

Para a obtenção das RNAs foram coletados os valores de ELL da mistura sistema glicerol (1)-metanol (2)-biodiesel de óleo de cártamo (3), provenientes do trabalho realizado por Rios *et al.* (2023), que obtiveram dados experimentais sob pressão de 101,325 kPa, e temperatura de 298,15 K.

2.2 DESENVOLVIMENTO DA RNA PARA DETERMINAÇÃO DA ESTABILIDADE DE FASES

A determinação da estabilidade de fases do sistema consiste em verificar se uma mistura com composição determinada está dentro do envelope de fases do diagrama ternário, o que indica que o sistema possui duas fases líquidas em equilíbrio; caso contrário, é indicativo de uma única fase homogênea.

Uma vez que Rios *et al.* (2023) reportou a curva binodal do ELL da mistura ternária em questão, é possível gerar um banco de dados robusto para a obtenção e validação do modelo de RNA.

Para gerar o banco de dados de modo a apresentar uma quantidade de observações que fosse suficiente para a RNA aprender tal comportamento, foram geradas 1100 observações para toda a faixa de composições.

A geração de dados foi realizada da seguinte forma: para cada valor de fração mássica do glicerol (w_1), entre 0 e 1, gerou-se aleatoriamente valores de fração mássica de metanol (w_2), também entre 0 e 1.

Comparando-se o valor de w_2 gerado com o valor experimental da curva binodal era possível se determinar se a mistura apresentava ELL ou não: se o valor gerado estivesse acima do valor conhecido, sabia-se que a mistura se apresentava em uma fase homogênea, caso contrário, em duas fases líquidas coexistindo em equilíbrio.

Vale salientar que para se calcular o valor da fração mássica do biodiesel (w_3), basta aplica a equação:

$$w_3 = 1 - w_2 - w_1 \quad (4)$$

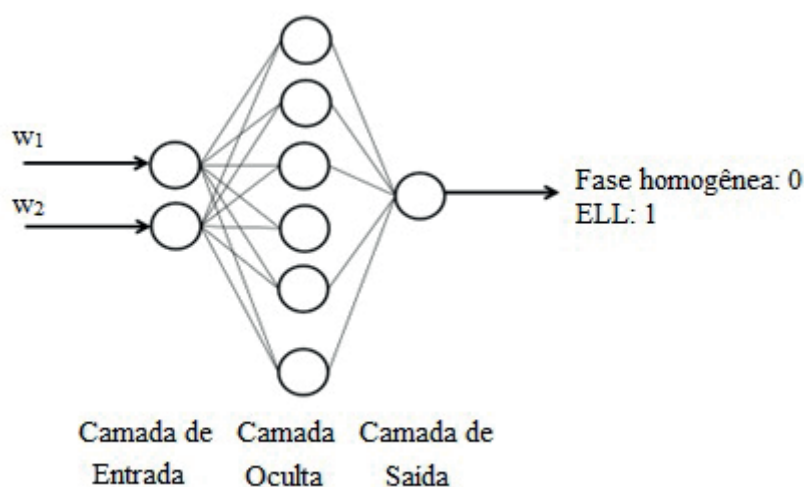
Para iniciar a obtenção da rede neural, foram consideradas como entradas as composições globais da mistura (fração molar de glicerol (w_1) e de metanol (w_2)). Como saída, se a mistura se apresentava em uma fase homogênea, o valor atribuído foi igual a 0, caso contrário, o valor era igual a 1.

Diferentes estruturas de RNAs com uma camada oculta foram testadas, variando-se o número de neurônios nessa camada, combinações de funções de ativação na camada oculta e na camada de saída, e algoritmos de otimização. As configurações testadas foram:

- Algoritmo de otimização: Levenberg-Marquardt (trainlm) e Levenberg Marquardt com regularização bayesiana (trainbr);
- Funções de ativação na camada oculta: tangente hiperbólica (tansig) e logarítmica sigmoideal (logsig);
- Funções de ativação na camada de saída: tansig, linear (purelin) e saturada linear simétrica (satlin);
- Número de neurônios na camada oculta: 2 a 9.

Dos dados coletados, os 660 pontos (60 % dos dados gerados) foram utilizados para a obtenção da RNA. Essa etapa consistia em três etapas: treinamento, teste e validação, em que 70 %, 15 %, e 15 %, respectivamente, das observações disponíveis foram utilizadas. O segundo conjunto de 440 pontos foi utilizado em uma etapa de simulação da RNA (validação externa), que foi realizada após a etapa de obtenção. Um exemplo de como a rede foi configurada para essas simulações está ilustrada na Figura 1.

Figura 1 - Estrutura geral da RNA para a determinação da estabilidade de fases.



Fonte: os autores.

Para a determinação do melhor modelo, todas as combinações possíveis levando em conta as opções apresentadas na Tabela 1, foram realizadas gerando resultados independentes. Esse tipo de abordagem permite verificar os resultados para cada arquitetura de RNA configurada no software computacional e garante um espaço de procura bastante abrangente. Além disso, cada configuração de RNA foi treinada 5 vezes.

Para estimar os pesos e bias das RNAs, utilizou-se a Equação como função objetivo (F_{ob}) a ser minimizada na etapa de treinamento. Vale ressaltar que quanto mais próximo de zero é o valor obtido, em tese, melhor é a rede.

$$F_{ob} = \frac{1}{nm} \sum_{i=1}^n \sum_{j=1}^m (y_{i,j}^{\text{exp}} - y_{i,j}^{\text{RNA}})^2 \quad (5)$$

em que n é o número de variáveis de saída e m é o número de pontos experimentais, y^{exp} e y^{RNA} representam, respectivamente, o valor real e o valor calculado pela RNA.

Para problemas de classificação como este, a escolha do melhor modelo de RNA recai em uma determinação estatística mais detalhada. Para isso, após o treinamento, foi calculado o valor do AUC (do inglês, *Area Under Curve*), que é obtido através da área sob a curva ROC, ou *Receiver Operating Characteristic Curve*. A curva ROC é utilizada para determinação do desempenho de um modelo de classificação binária e para determinar o ponto de corte abaixo do qual a resposta é igual a 0, ou acima do qual a resposta é igual a 1.

A curva ROC é formada pelos valores de taxa de verdadeiros positivos (TPR - Eq. (6)), *versus* a taxa de falsos positivos (FPR - Eq.), calculados após o treinamento da RNA, considerando os dados da validação e diferentes pontos de corte.

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$FPR = \frac{FP}{FP + TN} \quad (7)$$

em que: TP, FN, FP e TN são os números de verdadeiros positivos, falsos negativos, falsos positivos e verdadeiros positivos, respectivamente.

O ponto de corte ótimo, correspondente ao modelo mais eficiente é determinado analisando-se a curva ROC de modo a se maximizar o valor de TPR, minimizando-se o valor de FPR.

Além desses parâmetros, também foi calculada a razão de verossimilhança negativa (LR- - Eq. (8)) e positiva (LR+ - Eq.(9)). Quanto mais próximo de zero o valor de LR- e quanto mais positivo for o valor de LR+, mais confiável é o modelo estudado.

$$LR- = \frac{TPR}{FPR} \quad (8)$$

$$LR+ = \frac{FNR}{TNR} \quad (9)$$

em que: FNR e TNR são as taxas de falsos negativos e positivos, respectivamente.

Além disso, outra métrica calculada foi a acurácia (Acc). A acurácia foi calculada para todas as etapas e é dada pela Eq. :

$$Acc = \frac{TP + TN}{P + N} \quad (10)$$

A saturação dos neurônios (φ) da camada oculta foi outro parâmetro calculado. A saturação dos neurônios é uma consequência do comportamento das funções de ativação empregadas, especialmente quando se usa funções do tipo sigmoidais. Essas funções possuem a sua imagem limitada assintoticamente. Assim, caso as saídas do neurônio estejam concentradas apenas nos valores extremos, o que se tem é uma RNA cuja etapa de aprendizado não foi bem-sucedida (Glorot; Bengio, 2010). O parâmetro φ pode variar de 0 a 1, e é calculado conforme as Eqs. (11)-(13).

$$\varphi = \frac{\sum_{b=1}^B \left| \bar{g}_b \right| f_b}{\sum_{b=1}^B f_b} \quad (11)$$

em que:

$$\bar{g}_b = \frac{2(\bar{g}_b - g_L)}{g_U - g_L} - 1 \quad (12)$$

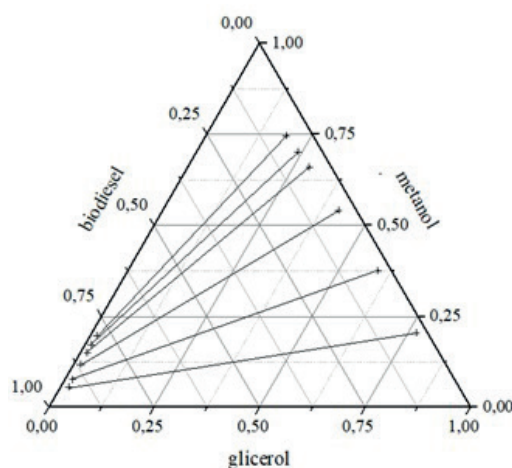
$$\bar{g}_b = \begin{cases} \sum_{k=1}^{f_b} g_k / f_b & \text{se } f_b > 0 \\ 0 & \text{se } f_b < 0 \end{cases} \quad (13)$$

em que: B é o número total de intervalos, b é o índice do intervalo, f é o número de saídas do intervalo b, g é o valor da saída do neurônio, U e L são o menor e maior valor de g, respectivamente. Considera-se que um modelo de RNA é confiável para valores de φ em torno de 0,5.

Por fim, a fim de se verificar a precisão do modelo a partir de dados experimentais, a RNA foi simulada considerando os dados experimentais de equilíbrio reportados em Rios *et al.* (2023).

2.3 DESENVOLVIMENTO DA RNA PARA O CÁLCULO DO ELL

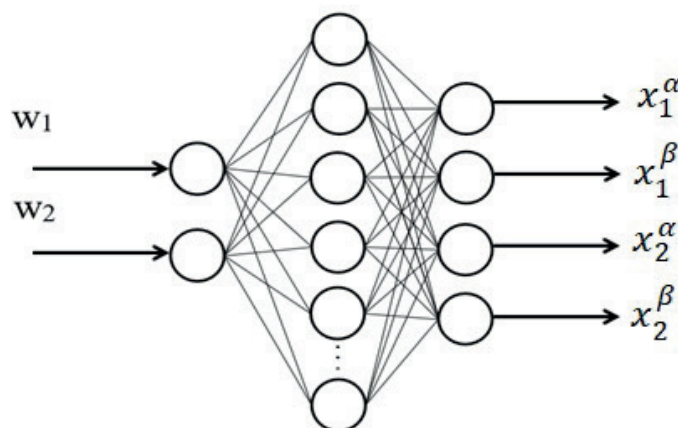
O cálculo do ELL consiste em se calcular as composições de equilíbrio da mistura ternária a partir da composição global dessa mistura, em determinada T e P. Para isso, utilizaram-se os dados de ELL obtidos no trabalho de Rios *et al.* (2023), conforme apresentado na Figura 2.

Figura 2 - Representação das linhas de amarração de ELL dos dados utilizados neste trabalho.

Fonte: os autores

Na Figura 2, qualquer ponto sobre a linha que liga os valores de composição que estão em equilíbrio corresponde a uma composição global da mistura, que, naquela condição de T e P, gera um sistema em ELL (Sandler, 2017). Assim, uma vez que a quantidade de dados experimentais é insuficiente, para a obtenção da RNA, foram gerados dados de composição global sobre cada linha de amarração. Foram gerados, aleatoriamente uma quantidade de 100 pontos de composição geral sobre estas linhas.

Como entrada da RNA foram consideradas a composição global do glicerol (w_1) e metanol (w_2) e como saída as composições destas substâncias na fase rica em biodiesel (fase α) e na fase rica em glicerol (fase β), respectivamente. Assim, as RNAs continham duas variáveis de entrada e quatro variáveis de saída. A estrutura geral de uma RNA com duas camadas ocultas está ilustrada na Figura 3. Nela, x_1^α , x_1^β , x_2^α , x_2^β são as frações molares do glicerol e do metanol na fase rica em biodiesel e na fase rica em glicerol, respectivamente.

Figura 3 - Estrutura geral da RNA para o cálculo do equilíbrio de fases.

Fonte: os autores

Diferentes estruturas de RNAs com uma e duas camadas ocultas foram testadas, variando o número de neurônios, funções de ativação na(s) camada(s) oculta(s) e na camada de saída, e algoritmos de otimização. Dos dados gerados, houve uma separação para treinamento, teste e validação de 70 %, 15 %, e 15 %, respectivamente. As opções de cada configuração testada foram:

- Algoritmo de otimização: trainlm, trainbr, Powell-Beale (traincgb), Resilient-Backpropagation (trainrp) e gradiente conjugado (trainscg).
- Funções de ativação na camada oculta: tansig, logsig, purelin, satlins e linear positiva saturada (poslin);
- Funções de ativação na camada de saída: tansig, logsig, purelin e poslin;
- Número de neurônios nas camadas ocultas: 4 a 40.

Cada configuração de RNA foi treinada 5 vezes. A Eq. (5) também foi usada como função objetivo a ser minimizada. Nesse caso, também foi realizada uma etapa de validação externa da RNA utilizando os pontos experimentais reportados no estudo de Rios *et al.* (2023).

Para a análise dos resultados obtidos, considerou-se dois parâmetros estatísticos. Nesse caso, uma vez que as saídas são numéricas, foi calculado o Erro Percentual Absoluto Médio (MAPE - Eq.(14)) e o coeficiente de correlação de Pearson.

$$MAPE = 100 \frac{1}{n} \sum_{i=1}^n \frac{|y_i^{\text{exp}} - y_i^{\text{RNA}}|}{y_i^{\text{exp}}} \quad (14)$$

Para a determinação das melhores RNAs, a média do MAPE de cada variável de saída da etapa de validação externa, foi considerada. Para todos os treinamentos das RNAs foram utilizadas rotinas computacionais em Matlab®, versão R2025a.

3 RESULTADOS E DISCUSSÃO

3.1 DESENVOLVIMENTO DA RNA PARA DETERMINAÇÃO DA ESTABILIDADE DE FASES

No que diz respeito à obtenção do modelo neural para a determinação do estado termodinâmico da mistura glicerol-metanol-biodiesel de óleo de cártamo, um total de 360 configurações de RNAs foram treinadas. Os dez melhores resultados obtidos estão dispostos na Tabela 1, em ordem decrescente de Acc da etapa de validação externa.

Tabela 1 - Resumo dos resultados obtidos para o modelo de estabilidade de fases.

Algoritmo	Função	Ativação	# Neurônios	F_{ob} Treino	F_{ob} Teste	F_{ob} Valid.	Acc. Simula	AUC	ϕ
	Oculto	Saída							
trainbr	tansig	purelin	9	$8,8.10^{-5}$	$4,7.10^{-6}$	$5,1.10^{-3}$	0,986	0,056	0,924
trainbr	logsig	purelin	9	$1,6.10^{-1}$	$1,6.10^{-1}$	$1,6.10^{-1}$	0,982	1,000	0,856
trainbr	logsig	logsig	9	$1,6.10^{-1}$	$1,4.10^{-1}$	$1,4.10^{-1}$	0,980	1,000	0,939
trainbr	logsig	logsig	7	$1,6.10^{-1}$	$1,7.10^{-1}$	$1,7.10^{-1}$	0,980	1,000	0,531
trainbr	tansig	satlins	9	$2,2.10^{-3}$	$1,0.10^{-5}$	$1,0.10^{-5}$	0,977	0,000	0,954
trainbr	logsig	satlins	4	$8,4.10^{-3}$	$3,9.10^{-4}$	$4,9.10^{-3}$	0,976	0,133	0,914
trainbr	tansig	satlins	5	$6,7.10^{-3}$	$1,1.10^{-2}$	$2,6.10^{-2}$	0,976	0,118	0,757
trainbr	logsig	satlins	9	$1,1.10^{-3}$	$1,4.10^{-2}$	$8,7.10^{-3}$	0,976	0,067	0,922
trainbr	logsig	satlins	7	$1,6.10^{-2}$	$3,5.10^{-2}$	$2,5.10^{-2}$	0,974	0,944	0,837

Ao se observar os resultados apresentados na Tabela 1, nota-se que não há muita variação nos valores de Acc. Desse modo, a priori, qualquer RNA poderia ser utilizada. No entanto, é importante combinar o efeito dos valores de AUC e ϕ para a determinação do melhor modelo.

Outra observação é que a quantidade de neurônios na camada intermediária não apresenta correlação direta com o F_{obj} do treinamento. Ambos os algoritmos de otimização apresentaram bons resultados e se mostram equilibrados entre as dez melhores RNAs. O mesmo pode ser observado com relação às funções de ativação da camada oculta. Por outro lado, com relação às funções de ativação da camada de saída, nota-se que a função satlins gerou altos valores de acurácia. No entanto, os valores de AUC são bastante baixos, o que demonstra que esses modelos não são confiáveis e podem ser desprezados.

A determinação do melhor modelo recaiu na combinação de maior AUC e menor valor de ϕ , garantindo, assim, o modelo mais confiável. Assim, levando em conta todos os parâmetros para a correta determinação de uma boa acurácia, dentre os resultados obtidos, a melhor configuração de rede neural é mostrada na Tabela 3 em negrito, com 7 neurônios na camada oculta, tendo sido treinada com o algoritmo de otimização trainbr e tendo como função de ativação logsig nas duas camadas.

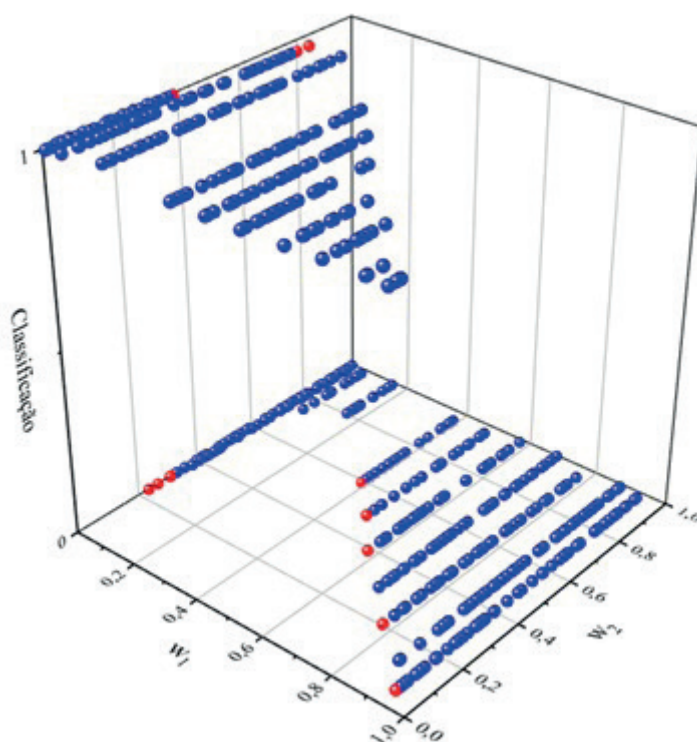
A Figura 4 mostra os resultados obtidos de forma mais visual, evidenciando os pontos em que a rede não conseguiu prever corretamente (pontos em vermelho), e os pontos em que a rede calculou corretamente (pontos em azul). Pode-se observar que a rede neural conseguiu bons resultados na predição da fase em que se encontra o sistema em equilíbrio líquido-líquido envolvendo a mistura considerada. Os poucos erros gerados concentram-se na região de composição de glicerol entre 0,3 e 0,4, aproximadamente.

Os dados de composição global dos experimentos de ELL, reportados por Rios *et al.* (2023), foram empregados em uma simulação do melhor modelo encontrado. A acurácia da classificação realizada pela melhor RNA foi de 100 %, ou seja, a RNA classificou corretamente o ELL.

Uma questão importante a ser endereçada é a possibilidade de sobreajuste (overfitting). Para evitar esse problema, a saturação dos neurônios, ϕ , foi calculada e a RNA foi escolhida considerando

o menor valor possível, aliado ao maior AUC. Neste trabalho o valor de ϕ encontrado foi de 0,531, valor muito próximo ao ideal, de 0,5, conforme indicado por Glorot; Bengio (2010). Por fim, é importante ressaltar que o melhor modelo apresenta 29 parâmetros ajustáveis. Para o método trainbr, o número efetivo de parâmetros, ou seja, o número de graus de liberdade da RNA pode ser calculado. Para a melhor RNA esse número foi igual a 6, que é bastante inferior ao número total de parâmetros. Isto, em conjunto com os resultados de validação interna e externa indicam que o modelo não possui risco de sobreajuste.

Figura 4 - Resultados da classificação da estabilidade de fases. No eixo z, o valor igual a 1 se refere à região de ELL e o valor igual a 0 se refere à região correspondente à fase líquida homogênea.



Fonte: os autores

3.2 DESENVOLVIMENTO DA RNA PARA O CÁLCULO DO ELL

No que diz respeito ao cálculo das composições das fases em equilíbrio da mistura glicerol-metanol-biodiesel de óleo de cártamo, no total, foram treinadas 4700 RNAs com uma camada oculta, e 7565 RNAs com duas camadas ocultas. Para compreender melhor o desempenho de cada conjunto de RNAs, a Tabela 2 apresenta a contagem de RNAs, por faixa da média do MAPE das variáveis de saída, e por algoritmo de otimização, para as RNAs com uma camada oculta.

Da Tabela 2, é possível notar que a maioria dos erros se concentrou na faixa de 20 % a 30 % e nenhuma RNA apresentou média do MAPE abaixo de 5 %, o que compromete o desempenho da predição. Além disso, os algoritmos trainlm e trainbr foram os que apresentaram a maior quantidade de erros baixos, entre 5,0 % e 10,0 %. A exceção foi o algoritmo trainscg, que somente apresentou

erros acima de 10 %. Por esse motivo, RNAs com duas camadas foram necessárias. Na Tabela 3 é apresentada a contagem de RNAs, por faixa da média do MAPE das variáveis de saída, e por algoritmo de otimização, para as RNAs com duas camadas ocultas.

Tabela 2 - Distribuição da média de MAPE e algoritmo de otimização.

Média do MAPE (%)	Trainbr	Traincgb	Trainlm	Trainrp	Trainscg	Total
0,0 - 1,0	0	0	0	0	0	0
1,0 - 5,0	0	0	0	0	0	0
5,0 - 10,0	43	1	27	17	0	88
10,0 - 15,0	137	53	198	221	76	685
15,0 - 20,0	133	122	88	134	148	625
20,0 - 30,0	283	231	316	299	250	1379
30,0 - 40,0	227	271	226	206	269	1199
40,0 - 50,0	117	197	67	51	158	590
> 50,0	0	65	17	13	39	134

Da Tabela 3, pode-se notar que a eficiência das RNAs aumentou substancialmente com a adição de uma camada oculta, uma vez que mais de 6 % das RNAs treinadas apresentaram erros abaixo de 10 % (contra 1,8 % das RNAs com uma camada oculta). Além disso, foi possível obter erros abaixo de 5 %, mesmo que para apenas uma configuração de RNA. Na Tabela 4 são apresentados os resultados das dez melhores RNAs obtidas, em ordem da média do MAPE das variáveis de saída.

Tabela 3 - Distribuição da média de MAPE por algoritmo de otimização.

Média do MAPE (%)	Trainbr	Trainlm	Total
0,0 - 1,0	0	0	0
1,0 - 5,0	0	1	1
5,0 - 10,0	344	167	511
10,0 - 15,0	975	855	1830
15,0 - 20,0	405	581	986
20,0 - 30,0	423	1154	1577
30,0 - 40,0	1127	881	2008
40,0 - 50,0	505	110	615
> 50,0	6	31	37

Tabela 4 - Resumo dos resultados obtidos para o modelo de cálculo do equilíbrio.

Algoritmo	Função Ativação			#Neurônios		MAPE Simulação (%)				Média do MAPE (%)
	Camada oculta		Saída	Camada oculta		x_1^α	x_2^α	x_1^β	x_2^β	
	1ª	2ª		1ª	2ª					
trainlm	purelin	logsig	tansig	5	10	4,26	4,44	8,53	2,68	4,98
trainbr	purelin	tansig	purelin	15	10	3,10	6,13	9,26	5,97	6,11
trainbr	purelin	tansig	purelin	30	5	1,66	5,57	9,19	9,07	6,37
trainbr	logsig	tansig	purelin	5	15	3,04	7,79	11,28	4,25	6,59
trainbr	tansig	logsig	purelin	10	10	4,00	8,20	8,99	5,24	6,61

Algoritmo	Função Ativação			#Neurônios		MAPE Simulação (%)				Média do MAPE (%)
	Camada oculta		Saída	Camada oculta		x_1^α	x_2^α	x_1^β	x_2^β	
	1ª	2ª		1ª	2ª					
trainlm	logsig	purelin	purelin	5	20	2,09	10,04	8,57	6,14	6,71
trainlm	tansig	purelin	purelin	5	5	5,02	5,39	5,47	11,41	6,82
trainbr	logsig	logsig	purelin	5	5	1,92	6,55	9,66	9,84	6,99
trainbr	logsig	logsig	purelin	30	5	1,29	5,68	10,77	10,61	7,09
trainbr	tansig	logsig	purelin	5	15	3,86	8,04	10,60	5,91	7,10

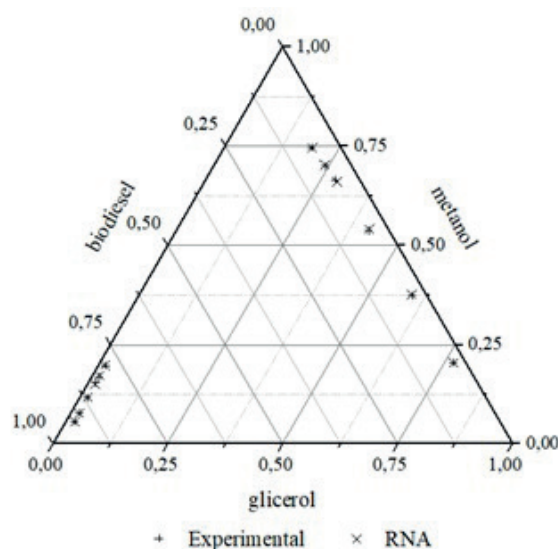
Nessa tabela, pode-se observar que só são reportadas RNAs com duas camadas ocultas. A menor média do MAPE obtida para uma RNA com uma camada oculta foi de 7,61 %, que apresentou MAPE de 1,97 %, 7,71 %, 10,53 %, e 10,23 % para x_1^α , x_1^β , x_2^α , x_2^β , respectivamente.

Além disso, pode-se notar que o algoritmo trainbr figura repetidas vezes dentre os melhores resultados, porém não foi com este algoritmo que o melhor resultado foi obtido. Com relação às funções de ativação, pode-se notar que as funções não-lineares aparecem mais vezes nas camadas ocultas, e que a função linear aparece majoritariamente na camada de saída, exceto na RNA que apresentou o melhor resultado.

Para esse caso, a melhor RNA foi treinada com o método trainlm, e apresenta 5 neurônios na primeira camada oculta e 10 neurônios na segunda. Para sistemas com componentes quimicamente semelhantes aos analisados neste trabalho, caso dos estudos de Moghadam e Asgharzadeh (2016) e Misikov *et al.* (2022), foram utilizadas duas camadas intermediárias para encontrar melhores resultados. Moghadam e Asgharzadeh (2016) utilizaram duas camadas com 3 e 4 neurônios, já Misikov *et al.* (2022) utilizaram 5 e 3. Ambos os autores reportam erros abaixo de 5 %. Ambos os estudos obtiveram RNAs específicas por mistura. Embora alguns dos 10 melhores resultados encontrados nas simulações tenham tido muitos neurônios na camada intermediária, a melhor rede neural encontrada aqui apresenta valores mais próximos aos encontrados na literatura.

Rios *et al.* (2023) apresentaram um valor de RMSE (a raiz quadrada da Eq.) igual a 3,20 ao realizar a modelagem termodinâmica empregando o modelo UNQUAC. Para fins de comparação, neste trabalho o RMSE dos dados de validação externa (que são os mesmos empregados pelos autores ao ajustar o modelo UNQUAC) foi calculado e foi igual a 1,83, o que ressalta uma maior precisão em relação ao obtido com a modelagem clássica. A eficiência do modelo proposto neste trabalho fica evidente no diagrama ternário apresentado na Figura 5, que compara o valor da composição experimental e os valores calculados pela RNA.

Figura 5 - Comparação entre os valores de composição experimental e calculado.



4 CONCLUSÃO

Este trabalho teve como objetivo a obtenção de RNAs para determinação da estabilidade de fases e do ELL do sistema envolvendo a mistura de Glicerol (1) + Metanol (2) + Biodiesel de Óleo de Cártamo (3) de sistemas sob pressão de 101,325 kPa, e na temperatura de 298,15 K.

Para a determinação da estabilidade da mistura, foram testadas diversas configurações de RNAs, variando-se o número de camadas ocultas, funções de otimização e algoritmos de otimização. Com isso, pôde-se observar que houve modelos de RNAs que apresentaram melhores resultados do que outros. O melhor resultado apresentou valores de acurácia acima de 98 % nas etapas de treinamento, teste e validação. A RNA tinha uma estrutura contendo 7 neurônios em uma única camada oculta.

Por sua vez, a determinação do ELL da mistura envolveu uma análise muito mais ampla, devido à complexidade do problema, foi necessário variar o número de camadas ocultas, funções de ativação, algoritmos de otimização e neurônios, gerando milhares de estruturas de RNAs, a fim de encontrar uma arquitetura que promovesse maior acurácia possível. A estrutura que apresentou o melhor resultado consistiu em uma RNA com duas camadas ocultas, contendo 5 e 10 neurônios cada. Quando comparado com o método UNIQUAC, os resultados obtidos pela RNA apresentou menor RMSE, evidenciando que as RNAs podem ser usadas para determinação de ELL de forma satisfatória e mais eficiente em relação à modelagem termodinâmica tradicional.

Por fim, este trabalho restringiu a modelagem do ELL à uma única temperatura e pressão do sistema, e considerando uma mistura visando demonstrar o potencial da ferramenta na resolução destes problemas. Trabalhos futuros podem considerar a expansão do conjunto de dados considerando outras misturas, condições experimentais de interesse.

REFERÊNCIAS

- BAZOOYAR, B.; SHAAHMADI, F.; ANBAZ, M. A.; JOMEKIAN, A. Intelligent modelling and analysis of biodiesel/alcohol/glycerol liquid-liquid equilibria. **Journal of Molecular Liquids**, v. 322, p. 114972, 2021.
- BEIGZADEH, R.; RAHIMI, M.; SHABANIAN, S. R. Developing a feed forward neural network multilayer model for prediction of binary diffusion coefficient in liquids. **Fluid phase equilibria**, v. 331, p. 48-57, 2012.
- ES' HAGHI, P.; MOHAMMADI, A.; SHAYESTEH, K.; SEDDIGHI, H. Comparative Analysis of Computational Prediction Methods for Liquid-Liquid Equilibria in Ternary Water-Ethanol-Benzene Mixtures. **Iranian Journal of Analytical Chemistry**, v. 11, n. 1, p. 1-11, 2024.
- GAO, L.; LONEY, N. W. New hybrid neural network model for prediction of phase equilibrium in a two-phase extraction system. **Industrial & engineering chemistry research**, v. 41, n. 1, p. 112-119, 2002.
- GLOROT, X.; BENGIO, Y. **Understanding the difficulty of training deep feedforward neural networks**. Proceedings of the thirteenth international conference on artificial intelligence and statistics, 2010.
- HAYKIN, S. **Neural Networks and Learning Machines**. 3rd ed. New York: Prentice Hall, 2009.
- KIM, M.; MOON, J. H. Deep neural network prediction for effective thermal conductivity and spreading thermal resistance for flat heat pipe. **International Journal of Numerical Methods for Heat & Fluid Flow**, v. 33, n. 2, p. 437-455, 2023.
- MCCORMICK, R.; MORIARTY, K. **Biodiesel Handling and Use Guide**. National Renewable Energy Laboratory (NREL), Golden, CO (United States), 2023.
- MELO, E.; OLIVEIRA, E.; MARTINS, T. A neural network correlation for molar density and specific heat of water: Predictions at pressures up to 100 MPa. **Fluid Phase Equilibria**, v. 506, p. 112411, 2020.
- MENEGHETTI, S. P.; MENEGHETTI, M. R.; BRITO, Y. C. A reação de transesterificação, algumas aplicações e obtenção de biodiesel. **Revista Virtual de Química**, v. 5, n. 1, p. 63-73, 2013.
- MICHELSSEN, M. L. The isothermal flash problem. Part I. Stability. **Fluid Phase Equilibria**, v. 9, n. 1, p. 1-19, 1982.
- MISIKOV, G. K.; PETROV, A.; TOIKKA, A. Application of artificial neural networks for the analysis of data on liquid-liquid equilibrium in three-component systems. **Theoretical Foundations of Chemical Engineering**, v. 56, n. 2, p. 200-207, 2022.

MOGHADAM, M.; ASGHARZADEH, S. On the application of artificial neural network for modeling liquid-liquid equilibrium. **Journal of Molecular Liquids**, v. 220, p. 339-345, 2016.

OTTAIANO, G. Y.; DA CRUZ, I. N.; DA CRUZ, H. S.; MARTINS, T. D. Estimation of vaporization properties of pure substances using artificial neural networks. **Chemical Engineering Science**, v. 231, p. 116324, 2021.

PANDHARIPANDE, S.; MOHARKAR, Y. Artificial neural network modeling of equilibrium relationship for partially miscible liquid-liquid ternary system. **International Journal of Computer Applications**, v. 52, n. 8, 2012.

PIERANTOZZI, M.; HOSSEINI, S. Density and viscosity modeling of liquid adipates using neural network approaches. **Journal of Molecular Liquids**, v. 397, p. 124134, 2024.

RIOS, R. B.; EVANGELISTA NETO, A. A.; MEDEIROS, H. A. D.; SILVA, F. F. M.; OLIVEIRA, H. N. M.; BARROS NETO, E. L.; CHIAVONE-FILHO, O. Liquid-Liquid Equilibria of Glycerol+ Alcohol+ Safflower Biodiesel Systems: Measurement and Modeling. **Journal of Chemical & Engineering Data**, v. 68, n. 7, p. 1716-1727, 2023.

SANDLER, S. I. **Chemical, biochemical, and engineering thermodynamics**. John Wiley & Sons, 2017.

SCHMITZ, J. E.; ZEMP, R. J.; MENDES, M. J. Artificial neural networks for the solution of the phase stability problem. **Fluid Phase Equilibria**, v. 245, n. 1, p. 83-87, 2006.

YATIM, F. E.; AIT BELALE, R.; ABALA, I.; CHHITI, Y.; RUJAS, N. M.; AGUILAR, F.; ALAOUI, F. E. M. H. Blending biomass-based liquid biofuels for a circular economy: Measuring and predicting density for biodiesel and hydrocarbon mixtures at high pressures and temperatures by machine learning approach. **Renewable Energy**, v. 234, p. 121146, 2024.