UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TECNOLOGIA E GEOCIÊNCIAS
DEPARTAMENTO DE ENGENHARIA QUÍMICA
CURSO DE GRADUAÇÃO EM ENGENHARIA QUÍMICA

MANUELA LEAL NADER

DESENVOLVIMENTO E APLICAÇÃO DE MODELO DE APRENDIZADO DE MÁQUINA NA IDENTIFICAÇÃO DE PROBLEMAS NO PROCESSO DE FERMENTAÇÃO ALCOÓLICA INDUSTRIAL A PARTIR DE DADOS EXPERIMENTAIS E TEÓRICOS

Recife
2023

MANUELA LEAL NADER

DESENVOLVIMENTO E APLICAÇÃO DE MODELO DE APRENDIZADO DE MÁQUINA NA IDENTIFICAÇÃO DE PROBLEMAS NO PROCESSO DE FERMENTAÇÃO ALCOÓLICA INDUSTRIAL A PARTIR DE DADOS EXPERIMENTAIS E TEÓRICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Engenharia Química.

Orientadora: Prof.ª Dra. Glória Maria Vinhas

Coorientador: Prof. Dr. Denys Ewerton da Silva Santos

Recife

2023

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Nader, Manuela Leal.
Desenvolvimento e aplicação de modelo de aprendizado de máquina na identificação de problemas no processo de fermentação alcoólica industrial a partir de dados experimentais e teóricos / Manuela Leal Nader. - Recife, 2023.
84 p. : il., tab.

Orientador(a): Glória Maria Vinhas
Cooorientador(a): Denys Ewerton da Silva Santos
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Tecnologia e Geociências, Engenharia Química - Bacharelado, 2023.
Inclui referências, apêndices.

1. aprendizado de máquina. 2. cinética. 3. contaminação. 4. fermentação alcoólica. 5. simulação. I. Vinhas, Glória Maria. (Orientação). II. Santos, Denys Ewerton da Silva. (Coorientação). IV. Título.

660 CDD (22.ed.)

MANUELA LEAL NADER

DESENVOLVIMENTO E APLICAÇÃO DE MODELO DE APRENDIZADO DE MÁQUINA NA IDENTIFICAÇÃO DE PROBLEMAS NO PROCESSO DE FERMENTAÇÃO ALCOÓLICA INDUSTRIAL A PARTIR DE DADOS EXPERIMENTAIS E TEÓRICOS

Trabalho de Conclusão de Curso apresentado ao Curso de Graduação em Engenharia Química da Universidade Federal de Pernambuco, como requisito parcial para obtenção do título de Bacharel em Engenharia Química.

Aprovado em: 28/09/2023

BANCA EXAMINADORA

_____

Profa. Dra. Glória Maria Vinhas (Orientadora)

Universidade Federal de Pernambuco

_____

Prof. Dr. Denys Ewerton da Silva Santos (Coorientador)

Universidade Federal de Pernambuco

_____

Prof. Dr. Pedro Ferreira de Souza Filho (Examinador Interno)

Universidade Federal de Pernambuco

_____

Prof. Dr. Felipe Pedro da Costa Gomes (Examinador Interno)

Universidade Federal de Pernambuco

Aos meus pais, que sempre me apoiaram em todas as minhas escolhas.

AGRADECIMENTOS

Em primeiro lugar, gostaria de agradecer aos meus dois orientadores, a professora Dra. Glória Vinhas e o professor Dr. Denys Santos, por todo o suporte técnico e emocional ao longo do desenvolvimento desse trabalho. Aos meus pais, por todo o apoio e por serem meus exemplos de caráter e humanidade. À minha família, que sempre torceu por mim e comemorou minhas conquistas. À minha parceira de estudos Júlia Cabral, sem a qual não sei se teria tido forças para chegar até aqui. Aos meus amigos, que me trouxeram leveza e alegria nos momentos que mais precisei. A todos os membros do grupo de pesquisa BioMAT, em especial à professora Dra. Thereza Soares, por tudo que fizeram por mim ao longo dos meus quase seis anos de iniciação científica. A Gisely Alves, por toda a paciência e por todo o suporte no laboratório. À Universidade Federal de Pernambuco e em especial ao Departamento de Engenharia Química que forneceu os meios materiais para o meu desenvolvimento acadêmico e pessoal. A todos os professores e profissionais que contribuíram para minha formação, direta ou indiretamente.

Mas sei que uma dor assim pungente
Não há de ser inutilmente
A esperança dança
Na corda bamba de sombrinha
E em cada passo dessa linha
Pode se machucar
(O Bêbado e A Equilibrista, 1979).

RESUMO

Uma das vias industriais de produção do etanol combustível consiste na fermentação alcoólica do caldo de cana-de-açúcar. Infelizmente, a contaminação microbiana do caldo por fungos e bactérias pode interferir neste processo, acarretando em uma diminuição na produção de etanol e perda de eficiência. Sendo assim, estratégias para identificar e evitar a contaminação estão sendo bastante utilizadas nesse tipo de indústria. Nesse âmbito, o aprendizado de máquina surge como uma importante ferramenta, sendo um dos princípios da Indústria 4.0. O presente trabalho visa o estudo dos diferentes parâmetros que interferem na fermentação e a aplicação de um modelo de aprendizado de máquina na detecção de contaminação em processos fermentativos. Para isso, foram realizados experimentos variando a fonte de carbono, a concentração de substrato e a ausência ou presença de contaminação bacteriana. Verificou-se uma redução na conversão de substrato em etanol nos casos com contaminação. Então, foram realizadas simulações de um modelo cinético a fim de gerar um grande número de dados para a aplicação de modelos de aprendizado de máquina. Também foi proposta uma modificação nesse modelo para simular a contaminação no processo fermentativo. Por fim, o modelo de redes neurais do tipo perceptron multicamadas foi utilizado para o problema de classificação binária, com ou sem contaminação, dos dados provenientes das simulações. Observou-se uma alta acurácia do modelo. Aplicou-se a metodologia de validação cruzada e a sintonização de hiperparâmetros, obtendo resultados ainda melhores, os quais foram comparados com a performance de outros modelos de classificação. Com isso, foi possível perceber o grande potencial da aplicação de modelos de aprendizado de máquina em processos industriais.

Palavras-chave: aprendizado de máquina; cinética; contaminação; fermentação alcoólica; simulação.

ABSTRACT

One of the industrial routes for producing fuel ethanol consists of the alcoholic fermentation of sugarcane juice. Unfortunately, microbial contamination of the broth by fungi and bacteria can interfere in this process, resulting in a decrease in ethanol production and loss of efficiency. Therefore, strategies to identify and avoid contamination are widely used in this type of industry. In this context, machine learning emerges as an important tool, which is one of the principles of Industry 4.0. The present work aims to study different parameters that affect fermentation and the application of a machine learning model to detect contamination in fermentation processes. Hence, experiments were carried out varying the carbon source, substrate concentration and in absence or presence of bacterial contamination. There was a reduction in the conversion of substrate to ethanol in cases with contamination. Then, simulations of a kinetic model were carried out in order to generate a large number of data for the application of machine learning models. A modification of such model was also proposed to simulate contamination in the fermentation process. Finally, a multilayer perceptron neural network model was used for a binary classification problem, with or without contamination, of data from simulations. A high accuracy of the model was observed. The cross-validation methodology and hyperparameter tuning were applied, obtaining even better results, which were compared with the performance of other classification models. In this regard, it is clear the great potential of applying machine learning models in industrial processes.

Keywords: alcoholic fermentation; contamination; kinetics; machine learning; simulation.

LISTA DE ILUSTRAÇÕES

Figura 1 –	Via de Embden-Meyerhof-Parnas: descrição das etapas, intermediários e reação global	22
Quadro 1 –	Microrganismos e seus produtos finais da fermentação	22
Figura 2 –	Esquema de um modelo de redes neurais artificiais com várias camadas de unidades de processamento	29
Figura 3 –	Fluxograma da metodologia utilizada neste trabalho	33
Figura 4 –	Cinco sistemas referentes aos cinco experimentos realizados	35
Quadro 2 –	Variáveis do banco de dados, seus significados e intervalos de valores possíveis	39
Figura 5 –	Concentração de sacarose em função do tempo para os experimentos realizados com melaço (A) e sacarose (B) como fontes de carbono	41
Figura 6 –	Concentração de glicose em função do tempo para cada um dos experimentos realizados	41
Figura 7 –	Concentração de frutose em função do tempo para cada um dos experimentos realizados	42
Figura 8 –	Concentração de etanol em função do tempo para cada um dos experimentos realizados	43
Figura 9 –	Massa de CO2 produzida em função do tempo para cada um dos experimentos realizados	44
Figura 10 –	Concentração de biomassa em função do tempo para cada um dos experimentos realizados	45
Figura 11 –	Concentração de glicose em função do tempo para cada um dos experimentos realizados	46
Figura 12 –	Concentração de biomassa em função do tempo para cada um dos experimentos realizados	46
Figura 13 –	Concentração de etanol em função do tempo para cada um dos experimentos realizados	47
Figura 14 –	Gráfico da linearização de lnX pelo tempo para o experimento F	48

Figura 15 – Gráfico da linearização de lnX pelo tempo para o experimento G 48
Figura 16 – Gráfico da linearização de lnX pelo tempo para o experimento H 49
Figura 17 – Concentrações de biomassa, substrato e produto em função do tempo obtidas através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016), com S_0 = 111,5 g/L, X_0 = 25 g/L, P_0 = 0 g/L e μ_{max} = 0,5 h^{-1} 51
Figura 18 – Concentrações de biomassa, substrato e produto em função do tempo obtidas por Oliveira, Oliveira, Tacin e Gattás (2016), com S_0 = 111,5 g/L, X_0 = 25 g/L, P_0 = 0 g/L e μ_{max} = 0,5 h^{-1} 51
Figura 19 – Concentração de substrato em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato 52
Figura 20 – Concentração de produto em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato 52
Figura 21 – Concentração de biomassa em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato 53
Figura 22 – Concentração de substrato em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa 54
Figura 23 – Concentração de produto em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa 54
Figura 24 – Concentração de biomassa em função do tempo obtida 55

através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa
Figura 25 – Concentração de substrato em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação 55
Figura 26 – Concentração de produto em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação 56
Figura 27 – Concentração de biomassa em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação 56
Figura 28 – Concentração de substrato em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β 57
Figura 29 – Concentração de produto em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β 57
Figura 30 – Concentração de biomassa em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β 58
Figura 31 – Quantidade de pontos provenientes de simulações com (1) e sem (0) contaminação no conjunto de treino 59
Figura 32 – Quantidade de pontos provenientes de simulações com (1) 59

e sem (0) contaminação no conjunto de validação
Figura 33 – Matriz de confusão para o conjunto de validação com modelo de redes neurais MLP 60
Figura 34 – Comparação entre a acurácia obtida ao utilizar diferentes modelos de classificação 61

LISTA DE TABELAS

Tabela 1 –	Descrição dos cinco experimentos realizados	34
Tabela 2 –	Descrição dos três experimentos realizados	36
Tabela 3 –	Valores das constantes do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para simulação de um processo em batelada de fermentação alcoólica	38
Tabela 4 –	Concentração de sacarose, glicose e frutose em soluções preparadas com sacarose e melaço antes e após o processo de autoclavagem	42
Tabela 5 –	Valores dos fatores de conversão de substrato em produto e de substrato em células para três ensaios realizados	49
Tabela 6 –	Valores das produtividades em célula e em produto e do rendimento experimental em relação ao teórico para os três ensaios realizados	50
Tabela 7 –	Acurácias obtidas para os conjuntos de teste e de validação com o modelo de redes neurais MLP	60
Tabela 8 –	Acurácias obtidas para cada uma das iterações da metodologia de validação cruzada com o modelo de redes neurais MLP, a média e o desvio padrão desses valores	61
Tabela 9 –	Combinação dos hiperparâmetros do modelo MLP que apresentou melhor desempenho e a acurácia da validação cruzada	62

LISTA DE ABREVIATURAS E SIGLAS

ATP	Adenosina trifosfato
MLP	Perceptron multicamadas
EMP	Embden-Meyerhof-Parnas
NADH	Dinucleótido de nicotinamida e adenina
ABS	Absorbância

LISTA DE SÍMBOLOS

μ	Velocidade específica de crescimento microbiano
X	Concentração de biomassa
S	Concentração do substrato limitante
P	Concentração de produto
t	Instante de tempo
μ_{max}	Velocidade específica máxima de crescimento microbiano
K_m	Constante de Monod
$Y_{X/S}$	Fator de conversão de substrato em células
$Y_{P/S}$	Fator de conversão de substrato em produto
K_S	Constante de saturação
K_i	Parâmetro de inibição para açúcares
P_{max}	Parâmetro de inibição para o etanol
n	Poder tóxico do etanol
α	Parâmetro da equação de Luedeking and Piret
X0	Concentração inicial de biomassa
S0	Concentração inicial de substrato
P0	Concentração inicial de produto
Pr_X	Produtividade em células
Pr_P	Produtividade em produto

SUMÁRIO

1	**INTRODUÇÃO**	**19**
1.1	OBJETIVO GERAL	20
1.2	OBJETIVOS ESPECÍFICOS	20
2	**REVISÃO DA LITERATURA**	**21**
2.1	FERMENTAÇÃO	21
2.1.1	**Produção industrial de bioetanol**	**23**
2.1.1.1	Microrganismos envolvidos na produção industrial de bioetanol	24
2.1.1.2	Contaminação microbiana na produção industrial de bioetanol	24
2.1.2	**Cinética de um processo fermentativo**	**25**
2.2	INTELIGÊNCIA ARTIFICIAL	27
2.2.1	**Algoritmos de inteligência artificial**	**27**
2.2.1.1	Redes neurais artificiais	29
2.3	INDÚSTRIA 4.0	30
3	**METODOLOGIA**	**32**
3.1	MATERIAIS	33
3.2	MICRORGANISMOS	33
3.3	ETAPA 1: PROCEDIMENTO EXPERIMENTAL PARA AVALIAÇÃO DE DIFERENTES CONDIÇÕES DE FERMENTAÇÃO	34
3.4	ETAPA 2: PROCEDIMENTO EXPERIMENTAL PARA DETERMINAÇÃO DOS PARÂMETROS DA FERMENTAÇÃO VARIANDO A CONCENTRAÇÃO DE GLICOSE	36
3.5	ETAPA 3: METODOLOGIA PARA SIMULAÇÃO DE UM PROCESSO FERMENTATIVO EM BATELADA	37
3.6	ETAPA 4: METODOLOGIA PARA DETECÇÃO DE CONTAMINAÇÃO POR INTELIGÊNCIA ARTIFICIAL	39
4	**RESULTADOS E DISCUSSÃO**	**41**
4.1	ETAPA 1: AVALIAÇÃO DA FERMENTAÇÃO ALCOÓLICA EM DIFERENTES CONDIÇÕES	41
4.2	ETAPA 2: DETERMINAÇÃO DOS PAR METROS DA FERMENTAÇÃO VARIANDO A CONCENTRAÇÃO DE	45

	GLICOSE	
4.3	ETAPA 3: SIMULAÇÃO DE UM PROCESSO FERMENTATIVO EM BATELADA	50
4.4	ETAPA 4: IDENTIFICAÇÃO DE CONTAMINAÇÃO EM PROCESSOS FERMENTATIVOS POR INTELIGÊNCIA ARTIFICIAL	58
5	**CONCLUSÃO**	**63**
	REFERÊNCIAS	**65**
	APÊNDICE A – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE SACAROSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA	**70**
	APÊNDICE B – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE GLICOSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA	**71**
	APÊNDICE C – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE FRUTOSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA	**72**
	APÊNDICE D – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE ETANOL EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA	**73**
	APÊNDICE E – CÓDIGO EM FORTRAN 90 PARA SIMULAÇÃO DO PROCESSO FERMENTATIVO EM DIFERENTES CONDIÇÕES	**74**
	APÊNDICE F – CÓDIGO EM PYTHON PARA TREINO E VALIDAÇÃO DE MODELO DE APRENDIZADO DE MÁQUINA	**78**
	APÊNDICE G – VALORES DEFAULT DOS HIPERPARÂMETROS DO MODELO DE REDES NEURAIS MLP DA BIBLIOTECA SCIKIT-LEARN	**84**

1 INTRODUÇÃO

A fermentação é um processo amplamente utilizado na indústria, por exemplo, na área de alimentos temos a produção de vinhos, cervejas e pães, e na indústria de combustíveis, a geração de bioetanol. Referente a esse último processo, o Brasil é responsável por 27,5% da produção mundial de etanol, cuja matéria-prima principal é a cana-de-açúcar (Vidal, 2022).

Vários fatores podem interferir no rendimento de produção do etanol, como as concentrações de açúcares, a temperatura, o pH e a salinidade do meio (Americano da Costa *et al*., 2014). Além disso, a conversão de substrato em etanol pode ser comprometida, caso haja contaminação por outros microrganismos, gerando baixa produtividade. Ademais, para resolver esse último problema, haverá a necessidade de esterilização dos equipamentos, resultando em parada na produção (Seo *et al*., 2020). Sendo assim, estratégias de controle e identificação de interferências estão sendo cada vez mais utilizadas nesse tipo de indústria.

Aliada à eficiência produtiva, a simulação de processos industriais permite a otimização das condições de operação, maximizando a capacidade de produção. O uso dessa ferramenta implica no desenvolvimento de um modelo matemático que represente a dinâmica da operação. Vários modelos foram propostos para a fermentação alcoólica, utilizando diferentes fontes de carbono e em temperaturas distintas (Liu *et al*., 2014; Oliveira; Oliveira; Tacin; Gattás, 2016).

Um conceito inovador surgiu em 2011, com a definição da indústria 4.0. Segundo Cañas, Mula, Díaz-Madroñero e Campuzano-Bolarín (2021), ele se baseia na flexibilidade dos sistemas de produção, na sua capacidade de tomar decisões de maneira inteligente, de se auto configurar, manter e projetar de maneira sustentável, com ou sem auxílio humano, por meio de modelos, algoritmos e heurísticas.

Nesse cenário, a inteligência artificial surge como uma importante ferramenta. O aprendizado de máquina, uma das técnicas de inteligência artificial, vem sendo aplicado no desenvolvimento de sensores virtuais, no monitoramento de processos e na predição de parâmetros (Zhu; Rehman; Wang; Shahzad, 2020; Barchi *et al*., 2016; Roell *et al*., 2022). Ainda, a detecção de falhas no processo produtivo já é uma realidade (Lo; Flaus; Adrot, 2019).

Redes neurais artificiais consistem em um modelo de aprendizado de máquina que pode ser aplicado tanto em problemas de classificação quanto de

clusterização. Esse método pode ser utilizado na identificação de falhas em um processo industrial (Heo; Lee, 2018).

No entanto, muito ainda pode ser desenvolvido nessa área, tanto a nível dos modelos e da capacidade de processamento quanto a nível de aplicação na indústria. Dessa forma, esse trabalho de conclusão de curso apresenta os seguintes objetivos geral e específicos:

1.1 OBJETIVO GERAL

Desenvolver modelo de aprendizado de máquina a partir de amostras obtidas através de experimentos e de modelo para identificação de problemas no processo de fermentação alcoólica industrial.

1.2 OBJETIVOS ESPECÍFICOS

Dentre os objetivos específicos propostos neste trabalho podemos destacar:

- Realizar cinco fermentações alcoólicas em diferentes condições, variando a fonte de carbono e a existência de contaminação bacteriana;
- Determinar os parâmetros da fermentação alcoólica em diferentes concentrações de glicose;
- Simular um modelo cinético em diferentes condições para gerar vários perfis de consumo de substrato, formação de produto e crescimento microbiano;
- Propor um modelo de aprendizado de máquina para identificar possíveis problemas no processo industrial de fermentação alcoólica.

2 REVISÃO DA LITERATURA

2.1 FERMENTAÇÃO

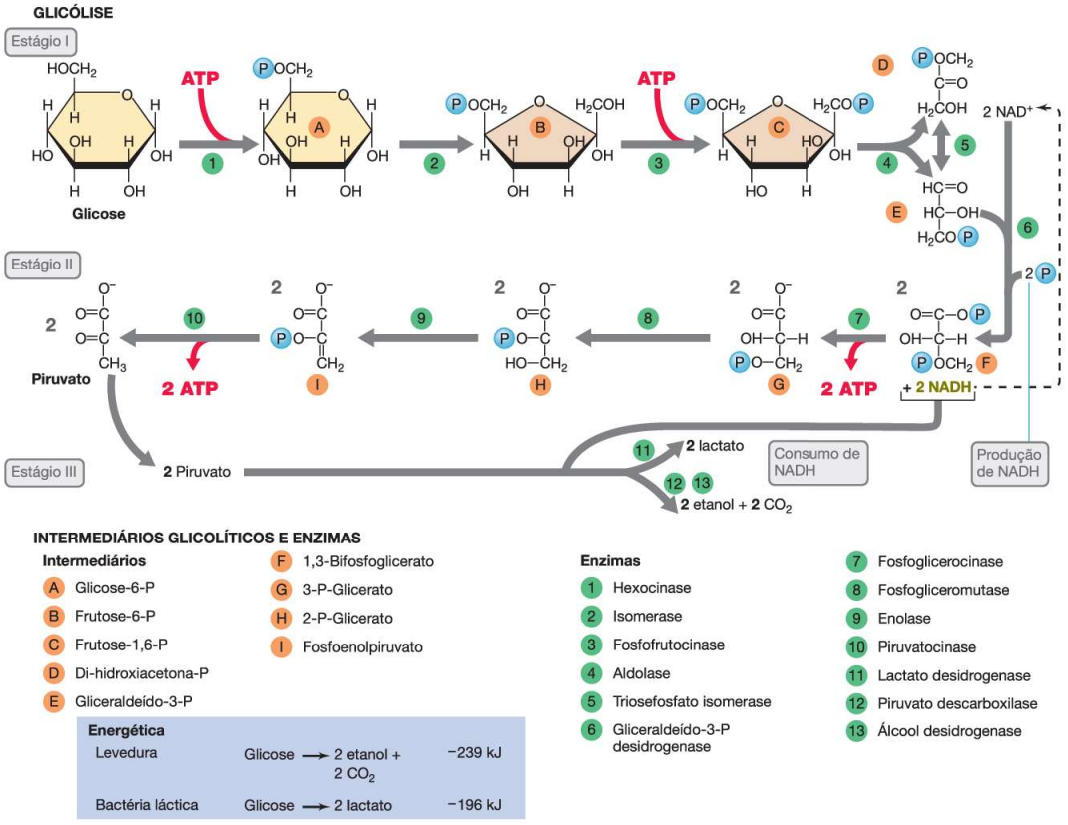
A fermentação consiste em reações catabólicas realizadas por microrganismos, ou seja, reações químicas catalisadas por enzimas que resultam em produção energética (Tortora; Funke; Case, 2017). A decomposição de carboidratos é a mais comum dessas reações, as quais envolvem a movimentação de elétrons ao longo de uma cadeia. Diferentemente da respiração celular, em que o oxigênio é a molécula receptora final desses elétrons, moléculas orgânicas como o piruvato e a acetilcoenzima A são os receptores finais durante a fermentação (Voidarou *et al*., 2020). Além disso, a respiração ocorre em condições aeróbias e a fermentação em condições anaeróbias.

A depender do substrato e da espécie de microrganismo, o processo de oxidação dos açúcares pode seguir diferentes vias. Para microrganismos homofermentativos, a via de Embden–Meyerhof–Parnas (EMP) é a mais comum. Já os microrganismos heterofermentativos utilizam a via da fosfocetolase (Voidarou *et al*., 2020).

A via EMP, também conhecida como glicólise, é a primeira etapa tanto da respiração celular quanto da fermentação. A partir dela, uma molécula de glicose é convertida em duas moléculas de piruvato. São duas as etapas que compõem essa via (Figura 1). A primeira consiste em uma fase preparatória, que ocorre com consumo de duas moléculas de ATP para produção de gliceraldeído-3-fosfato (Lee; Cho; Choi; Woo, 2021). Já a segunda compreende as reações redox com produção de piruvato e formação de quatro moléculas de ATP e duas de NADH. Dessa forma, a glicólise resulta em um saldo positivo de duas moléculas de ATP, geradas por meio da fosforilação a nível de substrato (Madigan *et al*., 2016).

Para que o processo de glicólise ocorra novamente, as duas moléculas de NADH produzidas devem ser novamente convertidas a NAD+. Em microrganismos fermentativos, isso ocorre em uma terceira etapa, que corresponde à conversão do piruvato em diferentes produtos a depender do microrganismo em questão (Madigan *et al*., 2016).

Figura 1 – Via de Embden-Meyerhof-Parnas: descrição das etapas, intermediários e reação global



Fonte: Madigan *et al*. (2016).

Os principais microrganismos fermentadores são as bactérias produtoras de ácido lático, as leveduras e os bolores (Sharma *et al*., 2020). Para as bactérias lácticas, o produto final da fermentação é o lactato, obtido através da redução do piruvato. Já as leveduras produzem etanol e gás carbônico (Madigan *et al*., 2016). O Quadro 1 traz os produtos finais da fermentação por diferentes microrganismos.

Quadro 1 – Microrganismos e seus produtos finais da fermentação

Microrganismos	**Produtos finais da fermentação**
Leveduras	Álcool
Bactérias láticas (*Lactobacillus* spp. etc.)	Ácido lático (homo e heterofermentação)
Clostridium spp., *Butyrivibrio* spp., *Bacillus* spp. e outros anaeróbios	Ácido butírico
Propionebacterium spp., *Veilonella* spp., *Bacteroides* spp., algumas *Clostridia* spp	Ácido propiônico
Acetobacter spp., *Gluconobacter* spp., *Bacillus subtilis*	Ácido acético
Enterobacteriaceae (*Escherichia* spp., *Enterobacter* spp., *Salmonella* spp.)	Ácidos mistos (ácido acético, fórmico, láctico, succínico)

Adaptado de: Voidarou *et al*. (2021).

Esses produtos são bastante explorados industrialmente. Na indústria alimentícia, estão envolvidos na produção de iogurtes, pães, salames e bebidas alcóolicas, entre outros (Sharma *et al*., 2020). Além disso, a indústria de combustíveis também utiliza a fermentação alcoólica para produção de bioetanol. Nesse aspecto, o Brasil encontra-se como o segundo maior produtor mundial de bioetanol, correspondendo a 27,5% do cenário global (Vidal, 2022).

2.1.1 Produção industrial de bioetanol

A obtenção de bioetanol pode ser dividida em três processos: o pré-tratamento da matéria-prima, a hidrólise dos açúcares e a fermentação (Tse; Wiens; Reaney, 2021). A depender do tipo da matéria-prima utilizada, o pré-tratamento é necessário para gerar açúcares que podem ser utilizados na fermentação, aumentando a velocidade do processo.

No Brasil, o etanol é majoritariamente produzido a partir da cana-de-açúcar. Através da aplicação de pressão ou do processo de difusão, o suco da cana é extraído e segue para a etapa de clarificação. Então, a formação dos cristais de sacarose ocorre através da evaporação desse suco clarificado. Após centrifugação, os cristais de sacarose são coletados e o melaço é formado, o qual possui uma alta concentração de sacarose, glicose e frutose (Basso; Basso; Rocha, 2011). Dessa forma, a etapa de hidrólise não é requerida. No processo de produção de etanol, podem ser utilizados o suco, o melaço ou uma combinação dos dois.

A fermentação industrial ocorre em equipamentos chamados de biorreatores. Esse processo pode ocorrer de forma contínua ou em bateladas, com ou sem alimentação. Ainda, o meio no qual os microrganismos crescem pode ser líquido, sólido ou um mosto com alta concentração de açúcar. A escolha do modo de operação e da característica do meio deve ser feita de acordo com as características da matéria-prima e do microrganismo utilizado, além de uma análise financeira apropriada (Tse; Wiens; Reaney, 2021). Na maior parte das destilarias brasileiras, a produção ocorre em batelada alimentada e em meio líquido, caracterizando uma fermentação submersa (Varize *et al*., 2022).

2.1.1.1 Microrganismos envolvidos na produção industrial de bioetanol

Grande parte das leveduras pode realizar a fermentação alcoólica, sendo a *Saccharomyces cerevisiae* a mais utilizada industrialmente. Isso ocorre devido à sua capacidade de se desenvolver na presença ou não de oxigênio e à sua alta tolerância ao etanol (Tse; Wiens; Reaney, 2021).

Ao longo desse processo, os microrganismos estão sujeitos a inúmeros fatores que podem levar a uma redução na produção de etanol, à produção de subprodutos indesejáveis, à contaminação microbiana, entre outros (Tse; Wiens; Reaney, 2021). Dentre os principais fatores, pode-se ressaltar as concentrações de açúcares, a temperatura, o pH e a salinidade do meio (Americano da Costa *et al*., 2014; Devos; Colla, 2022). As leveduras da espécie *Saccharomyces cerevisiae*, por exemplo, apresentam uma melhor eficiência na presença de glicose em comparação à frutose (Tse; Wiens; Reaney, 2021).

2.1.1.2 Contaminação microbiana na produção industrial de bioetanol

Segundo Seo *et al*. (2020), a falta ou a baixa eficiência da esterilização do meio contendo a fonte de carbono utilizada na fermentação industrial e a reutilização das leveduras após centrifugação tornam tais processos sujeitos à contaminação por outros microrganismos. Sabe-se que esses microrganismos podem ser leveduras, como as dos gêneros *Dekkera, Schizosaccharomyces* e *Candida*, ou bactérias, especialmente as produtoras de ácido lático.

A contaminação por bactérias produtoras de ácido lático gera uma competição por nutrientes com a levedura, além da produção de metabólitos que podem reprimir a fermentação (Tse; Wiens; Reaney, 2021). Essa repressão é observada na forma de redução da produção de etanol, formação de espuma no meio, floculação e diminuição na viabilidade das células de levedura. Esses microrganismos contaminantes suportam condições adversas como concentrações altas de etanol, açúcar e a acidez do meio. Dentre os microrganismos, as bactérias produtoras de ácido lático são as mais encontradas em casos de contaminação do processo de fermentação alcoólica (Branco; Diniz; Albergaria, 2022).

Várias estratégias foram desenvolvidas para minimizar os riscos de contaminação bacteriana no processo fermentativo. A lavagem das células a serem

reutilizadas com uma solução de ácido sulfúrico é bastante empregada para esse fim. No entanto, a viabilidade das leveduras pode ser bastante reduzida após vários ciclos. Antibióticos também são comumente utilizados, mas a questão da resistência a partir de mutações torna-se um problema recorrente. Uma alternativa é o uso de produtos naturais com ação antimicrobiana, como materiais lignocelulósicos provenientes de biomassa vegetal, ácidos extraídos de lúpulo e quitosana obtida da casca de camarões e caranguejos (Seo *et al*., 2020).

2.1.2 Cinética de um processo fermentativo

Um modelo cinético do processo de fermentação alcoólica constitui uma descrição matemática do crescimento celular das leveduras, do consumo dos açúcares e da produção de etanol ao longo do tempo. Esses modelos podem ser fenomenológicos, sendo obtidos a partir das equações diferenciais que descrevem o processo, ou empíricos, obtido por meio do ajuste de dados experimentais (Du *et al*., 2022).

Primeiramente, é necessária a descrição dos balanços materiais para cada componente. Em um processo de fermentação em batelada, tem-se que a taxa de acúmulo é igual a taxa de consumo ou de formação, uma vez que não há entrada e saída de material. Considerando o volume constante ao longo da fermentação, os balanços materiais para as células, o substrato e o produto equivalem às equações 1, 2 e 3, respectivamente (Zanardi; Carvalho; Costa Júnior, 2016). Nelas, a concentração de biomassa corresponde à variável X, a concentração de substrato à variável S e a concentração de produto à variável P. Ainda, as constantes $Y_{X/S}$ e $Y_{P/S}$ equivalem às conversões de substrato em células e em biomassa, respectivamente.

$$\frac{dX}{dt} = \mu X \tag{1}$$

$$\frac{dS}{dt} = -\frac{1}{Y_{X/S}}\frac{dX}{dt} \tag{2}$$

$$\frac{dP}{dt} = \frac{1}{Y_{P/S}}\frac{dX}{dt} \tag{3}$$

O modelo de Monod é bastante utilizado para descrever o crescimento microbiano (Monod, 1949). Segundo ele, a velocidade específica de crescimento (μ) é uma função da concentração do substrato limitante, a qual pode ser expressa pela equação 4.

$$\mu = \frac{\mu_{max} S}{K_m + S} \tag{4}$$

São duas as constantes presentes nesse modelo. A primeira consiste na velocidade específica máxima de crescimento microbiano (μ_{max}) e a segunda corresponde à concentração de substrato na qual a velocidade é igual à metade da velocidade máxima (K_m). Graficamente, essas constantes podem ser determinadas a partir da linearização de Lineweaver–Burk (Lineweaver; Burk, 1934).

Apesar do crescimento microbiano depender da ocorrência de diversos processos enzimáticos simultaneamente, é bastante comum simplificá-lo como um processo enzimático descrito pelo modelo de Michaelis-Menten e com um único substrato (Amirian; Irwin; Finkel, 2022). Ainda, tem-se que, em condições de saturação de substrato, o crescimento microbiano apresenta uma cinética de ordem zero e que, em condições de escassez de substrato, esse processo consiste em uma cinética de primeira ordem (Xu, 2020).

Desde que foi desenvolvido, foram propostas diversas modificações para o modelo de Monod. Zentou *et al.* (2021) propuseram uma modificação a fim de levar em consideração a inibição por subprodutos da fermentação. A limitação de substrato e a inibição por substrato e produto foram incorporadas ao modelo de Monod por Oliveira, Oliveira, Tacin e Gattás (2016), sendo validado com dados de fermentação alcoólica em batelada do caldo de cana-de-açúcar utilizando a *Saccharomyces cerevisiae*.

Também, outros modelos mais robustos foram desenvolvidos a fim de adequar melhor diferentes condições. Liu *et al*. (2014) propuseram um modelo para a produção de etanol por leveduras da espécie *Saccharomyces cerevisiae*, que leva em consideração o efeito da temperatura no processo de sacarificação e fermentação simultânea.

Sonego *et al.* (2016) estudaram o uso do modelo híbrido Andrews-Levenspiel para representar o processo de fermentação extrativo, operando em batelada

alimentada com extração de dióxido de carbono. Ainda, Rivera *et al.* (2013) propuseram um procedimento para otimização dos parâmetros cinéticos do processo fermentativo baseados em algoritmos genéticos, os quais podem ser considerados métodos de inteligência artificial.

A modelagem da cinética da reação é necessária para possibilitar a simulação desses processos, a qual permite uma análise de efeito das várias variáveis envolvidas. Sendo assim, é uma ferramenta bastante utilizada na otimização das condições de operação (Borges, 2008).

2.2 INTELIGÊNCIA ARTIFICIAL

Métodos de inteligência artificial consistem em algoritmos computacionais que, a partir da interpretação de um conjunto de dados complexos, e não necessariamente correlacionados, podem se adaptar de forma a realizar tarefas específicas ou atingir metas (Haenlein; Kaplan, 2019). Nesse ramo, o aprendizado de máquina, ou *machine learning*, é um dos métodos mais utilizados. A partir de modelos matemáticos construídos com base em dados de treinamento, algoritmos de aprendizado de máquina podem fazer predições ou ainda tomar decisões (Khanzode; Sarode, 2020).

O uso dessa técnica, se bem ajustada para o problema, traz inúmeros benefícios, como uma maior velocidade em realizar tarefas quando comparado a humanos, uma maior taxa de sucesso com menores erros e a possibilidade de resolver problemas complexos e de descobrir novas correlações (Khanzode; Sarode, 2020).

2.2.1 Algoritmos de inteligência artificial

Os algoritmos de aprendizado de máquina podem ser divididos em quatro categorias: aprendizagem supervisionada, não supervisionada, semi-supervisionada e por reforço. A aprendizagem supervisionada consiste na comparação entre o resultado esperado e o obtido através do modelo (Das; Dey; Pal; Roy, 2015). Sendo assim, o conjunto de dados de treinamento utilizado possui um rótulo, e a determinação desse rótulo é o objetivo do modelo (Sarker, 2021). Esse tipo de metodologia é bastante utilizado para problemas de classificação, nos quais se quer

separar os dados em diferentes classes, e de regressão, cujo rótulo a ser determinado é um valor real (Nasteski, 2017).

Já a aprendizagem não supervisionada se baseia na exploração de dados não rotulados, a fim de identificar padrões, separar em grupos e explorar o conjunto de dados. Dessa forma, esses modelos são empregados em problemas de clusterização, em que se quer agrupar dados com características semelhantes. Ainda, a aprendizagem semi-supervionada se trata de uma combinação entre os dois tipos anteriores, sendo utilizada em problemas de classificação e clusterização em que existem poucos dados rotulados (Sarker, 2021).

Por fim, a aprendizagem por reforço consiste na avaliação automática da ação que melhora a eficiência com base em um determinado contexto ou ambiente. A decisão é tomada de forma a aumentar uma função recompensa ou minimizar o risco. Além de problemas de classificação, pode ser utilizada no controle de variáveis, no âmbito da automação e otimização de processos (Sarker, 2021).

A validação desses modelos deve ocorrer com um conjunto de dados novos, ou seja, que não foram utilizados para treinar o modelo. No entanto, a indisponibilidade de conjuntos de dados muito grandes pode dificultar essa tarefa. Uma solução bastante utilizada consiste na validação cruzada, em que o treinamento ocorre iterativamente e, em cada iteração, uma parte diferente dos dados é utilizada para o treino e outra para validação. Dessa forma, todo o conjunto de dados é utilizado para ambas as tarefas (Vabalas; Gowen; Poliakoff; Casson, 2019).

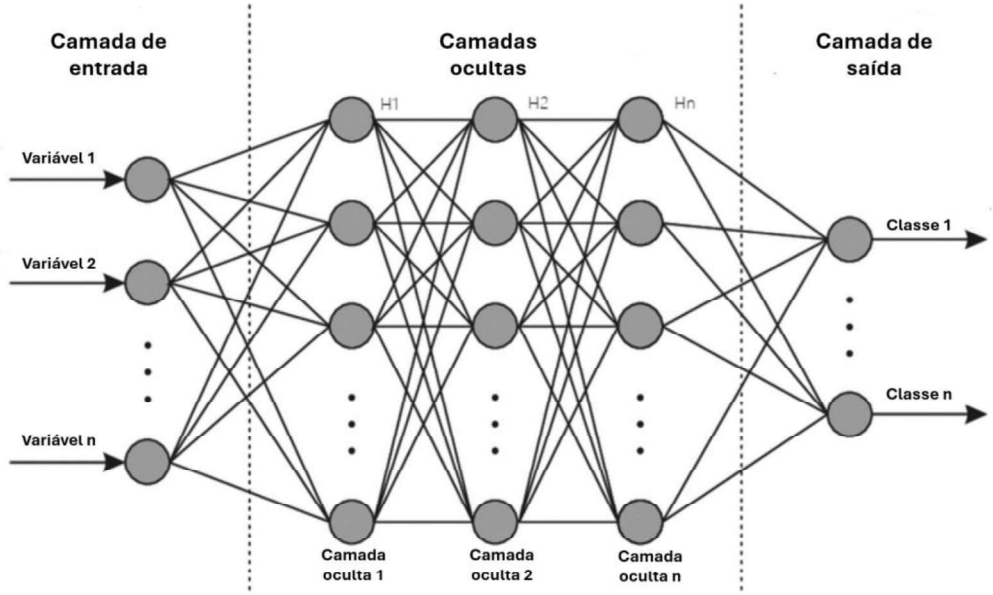
Muitos modelos já foram desenvolvidos e encontram-se disponíveis em diversas bibliotecas, como a *scikit-learn* da linguagem Python (Pedregosa *et al*., 2011). Para problemas de classificação, exemplos de modelos bastante utilizados são o classificador k-vizinhos mais próximos, a regressão logística, o classificador *Gradient Boosting*, o classificador árvore de decisão e o classificador florestas aleatórias (Khan *et al*., 2022; Phillips; Cripps; Lau; Hodkiewicz, 2015; Safari; Parvinnia; Haddad, 2021). Outro modelo muito comum consiste nas redes neurais artificiais.

2.2.1.1 Redes neurais artificiais

Redes neurais artificiais constituem um modelo de aprendizado de máquina que tenta mimetizar o comportamento do sistema nervoso biológico. Ele pode ser aplicado a problemas de classificação, regressão e clusterização. Nele, cada unidade de processamento local recebe uma entrada, a qual é multiplicada por um valor ou peso. Então, a soma ponderada das entradas recebidas passa por uma função de ativação que gera uma saída (Mohamed, 2017). A partir da combinação de várias dessas unidades de processamento, obtém-se uma rede conectada em que o aprendizado ocorre pelo ajuste desses pesos (Lo; Flaus; Adrot, 2019).

A partir desse princípio, foram desenvolvidos outros modelos mais robustos com a implementação de várias camadas de unidades de processamento. Um bastante popular é o perceptron multicamadas (MLP), formado por uma camada de entrada, uma ou mais camadas ocultas e uma camada de saída (Sarker, 2021). Nesse caso, a camada de saída pode ser entendida como a probabilidade de determinada entrada pertencer a cada uma das classes disponíveis, sendo selecionada a maior (Heo; Lee, 2018). A Figura 2 traz um esquema do funcionamento desse método, sendo cada um dos nós equivalente a uma unidade de processamento.

Figura 2 – Esquema de um modelo de redes neurais artificiais com várias camadas de unidades de processamento



Adaptado de: Sarker (2021).

Esse modelo pertence a uma nova categoria de algoritmos de aprendizado de máquina, conhecido com aprendizado profundo ou *deep learning*. Os métodos de aprendizado profundo apresentam um ótimo desempenho no contexto de dados cada vez mais numerosos, e um melhor detalhamento deste ponto pode ser encontrado no trabalho de Torres *et al.* (2021).

Um problema para o qual podemos aplicar os algoritmos descritos é o de classificação, comumente encontrado na indústria, na detecção de falhas de um processo produtivo. Nesse viés, Heo e Lee (2018) propuseram um modelo para o diagnóstico de falhas em um processo industrial, utilizando redes neurais artificiais, que obteve uma acurácia de 97,26%, demonstrando o poder dessa ferramenta para auxiliar no controle de falhas e na automatização para correção de processos.

A aplicação desse modelo de inteligência artificial na indústria é vantajosa devido à sua habilidade de se adaptar a novos cenários e de lidar com dados cujo ruído é alto. No entanto, a demora no tempo de treino desses algoritmos e a necessidade de grande número de dados podem ser empecilhos para certos casos (Ray, 2019). Dessa forma, pesquisas estão sendo elaboradas para facilitar a implementação de métodos de inteligência artificial na indústria.

2.3 INDÚSTRIA 4.0

A produção industrial evoluiu de acordo com o desenvolvimento de novas tecnologias. A Primeira Revolução Industrial permitiu o aumento produtivo a partir do uso de máquina movidas à agua ou vapor. A Segunda Revolução Industrial consistiu na utilização de energia elétrica, permitindo a produção em massa. Já a automação da produção levou à Terceira Revolução Industrial (Xu; Xu; Li, 2018). No entanto, as mudanças promovidas pela integração entre produção e conectividade em rede configuram o início de uma nova revolução industrial.

Nesse cenário, o conceito de indústria 4.0 surgiu em 2011 e pode ser definido como um sistema de produção flexível, capaz de utilizar modelos, algoritmos e heurística para tomar decisões, com ou sem ajuda humana, de maneira inteligente, e de se auto manter, projetar e configurar de forma sustentável (Cañas; Mula; Díaz-Madroñero; Campuzano-Bolarín, 2021). Sendo assim, métodos de inteligência artificial e aprendizado de máquina podem ser considerados imprescindíveis para a revolução das indústrias inteligentes (Cioffi *et al*., 2020).

As aplicações desses métodos são bastante variadas na indústria. Um excelente exemplo é o desenvolvimento de sensores virtuais, que permitem a determinação indireta, em tempo real, de variáveis de difícil medição. Isso é realizado pela aplicação de modelos matemáticos a dados de variáveis facilmente detectadas (Zhu; Rehman; Wang; Shahzad, 2020). Ainda, algoritmos de aprendizado de máquina podem ser utilizados na predição de variáveis de processo, como a velocidade de produção de determinada substância, quando a modelagem matemática do processo em si é um desafio (Roell *et al*., 2022).

Outra área em que o aprendizado de máquina traz inúmeros benefícios é a detecção de anomalias em processos industriais. A partir de dados obtidos através de sensores, modelos podem ser desenvolvidos para reconhecer comportamentos anormais, os quais podem ser causados por falhas nos equipamentos e nos sistemas de controle, problemas na matéria-prima, entre outros (Lo; Flaus; Adrot, 2019). Sendo assim, a rápida localização e identificação da causa dessas falhas permite uma resposta mais eficaz.

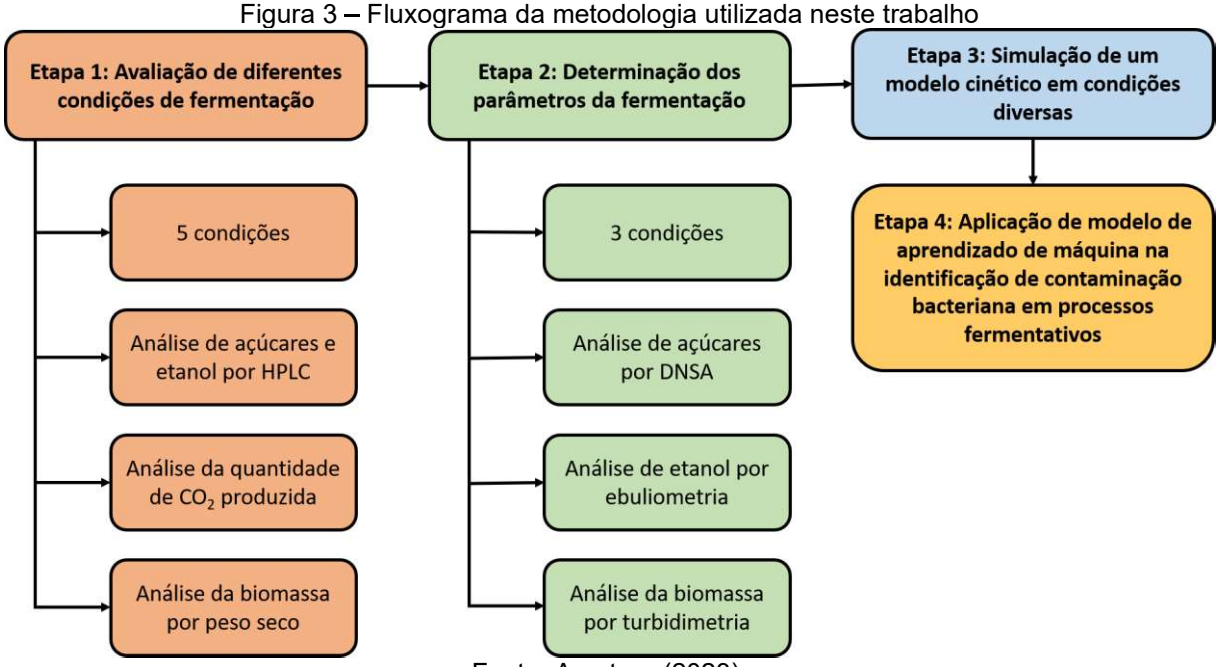
3 METODOLOGIA

Os experimentos de fermentação foram realizados no Laboratório de Microbiologia (DEQ-UFPE), as análises de etanol e açúcares foram realizadas no Laboratório de Cromatografia Instrumental (DEQ-UFPE) e a metodologia teórica foi desenvolvida no Laboratório de Modelagem Molecular e Métodos Computacionais avançados (DQF-UFPE).

Para melhor compreensão deste trabalho, foram realizados quatro estudos. No primeiro (Etapa 1), foram realizadas fermentações com diferentes fontes e concentração de carbono, bem como fermentação com levedura isolada e com cultura mista de levedura e bactéria. A adição da bactéria tem o propósito de simular uma fermentação com contaminação bacteriana. No segundo estudo (Etapa 2), foram realizadas fermentações variando a concentração de glicose, usando métodos de análises diferentes dos usados na Etapa 1 para determinar a biomassa e as concentrações de glicose e etanol. Aliado a essa variação de glicose, foram calculados os parâmetros da fermentação.

Já a Etapa 3 consistiu na simulação de um modelo cinético de fermentação alcoólica para obtenção de diversas curvas de crescimento microbiano, consumo de substrato e formação de produto, as quais foram utilizadas como dados para o modelo de aprendizado de máquina. Por último, a Etapa 4 correspondeu à aplicação de modelos de aprendizado de máquina na identificação de contaminação bacteriana em processos fermentativos. A Figura 3 traz um fluxograma da metodologia utilizada.

Figura 3 – Fluxograma da metodologia utilizada neste trabalho



Fonte: A autora (2023).

3.1 MATERIAIS

Para os experimentos de fermentação alcoólica, foram utilizados os seguintes equipamentos e vidrarias: 5 erlenmeyers de 2 L, 5 balões volumétricos de 1 L, 5 erlenmeyers de 250 mL, 5 erlenmeyers de 125 mL, 80 béqueres de 5 mL, 40 pipetas graduadas de 5 mL, um cromatógrafo líquido de alta eficiência com detector de índice de refração e coluna HPX-87H, uma balança analítica de 4 casas decimais, uma autoclave, uma estufa e uma balança de 1 casa decimal. A composição do meio de cultura foi a seguinte: 100 g/L de glicose, sacarose ou melaço equivalente a 100 g/L de açúcar, 2 g/L de extrato de levedura, 1 g/L de ureia, 1 g/L de KH_2PO_4, 0,5 g/L de $MgSO_4.7H_2O$ e água destilada.

3.2 MICRORGANISMOS

Os microrganismos utilizados foram a levedura *Saccharomyces cerevisiae* e a bactéria *Bacillus clausii*, presente no probiótico comercial Enterogermina®. Essa última foi usada para simular contaminação nos biorreatores.

3.3 ETAPA 1: PROCEDIMENTO EXPERIMENTAL PARA AVALIAÇÃO DE DIFERENTES CONDIÇÕES DE FERMENTAÇÃO

Nessa etapa do trabalho foram feitas cinco fermentações, variando as fontes de carbono e inoculando a suspensão de bactéria comercial (Enterogermina®) para propor dois sistemas de fermentação alcoólica com contaminação. Assim, foram preparados 10 L de meio de cultura contendo 10 L de água destilada, 5,0039 g de $MgSO_4.7H_2O$ (0,5 g/L), 10,0099 g de KH_2PO_4 (1 g/L), 10,0018 g de ureia (1 g/L) e 19,9916 g de extrato de levedura (2 g/L). Esse meio foi dividido em cinco amostras de 2 L cada, e foram adicionadas 200 g de glicose (100 g/L) em três, 200 g de sacarose (100 g/L) em um e 240,9 g de melaço cujo Brix era de 83% no último. Essa quantidade de melaço foi adicionada para que houvesse equivalência em quantidade de açúcar, mesmo que parte dos açúcares do melaço não fossem fermentáveis. Os cinco experimentos encontram-se descritos na Tabela 1.

Tabela 1 – Descrição dos cinco experimentos realizados

Experimento	Fonte de carbono	Percentual em massa de levedura	Percentual em massa de bactéria
A	Melaço	100%	0%
B	Sacarose	100%	0%
C	Glicose	100%	0%
D	Glicose	75%	25%
E	Glicose	50%	50%

Fonte: A autora (2023).

Dos 2 L de cada solução, 1350 mL foram adicionados aos erlenmeyers de 2 L, 450 mL foram adicionados aos balões de 1 L, 150 mL foram adicionados aos erlenmeyers de 250 mL e 50 mL foram adicionados aos erlenmeyers de 125 mL. Por fim, os meios foram esterilizados em autoclave a 121 °C e 1 atm por 15 minutos.

Então, foi realizada a preparação dos pré-inóculos nos erlenmeyers de 250 e 125 mL. Para isso, foi feita uma suspensão de levedura com turbidez equivalente à turbidez da suspensão de bactéria comercial (Enterogermina®). Visto que a embalagem da suspensão bacteriana indicava a presença de 4.109 células, utilizou-se o número 4 da escala McFarland para a suspensão de levedura, o qual equivale a 12.108 células. Após a inoculação, os erlenmeyers foram colocados na mesa agitadora por 20 horas e meia.

Passado esse tempo, foi realizada a transferência dos pré-inóculos para os erlenmeyers de 2 L e para os balões volumétricos de 1 L, nos quais foram realizados os experimentos de cinética microbiana (Figura 4). Três amostras de 5 mL cada foram coletadas de duas em duas horas nas primeiras 12 horas e após 24 horas do início da fermentação. A cada amostragem, foi medida a variação da massa dos balões após agitação para desprendimento do CO_2, a qual foi considerada igual a massa de CO_2 formada.

Figura 4 – Cinco sistemas referentes aos cinco experimentos realizados

Fonte: A autora (2023).

Para a medição do peso seco em cada amostra, oitenta béqueres foram previamente tarados após aquecimento em estufa a 100 °C por 24 horas. Então, as amostras de 5 mL foram colocadas nos béqueres, sendo posteriormente colocados na estufa a 100 °C até peso constante e a massa medida novamente.

Já as quantidades de açúcares e de etanol foram medidas através um cromatógrafo líquido de alta eficiência com detector de índice de refração e coluna HPX-87H. Também, foram feitas curvas de calibração com soluções padrões de sacarose, glicose, frutose e etanol para relacionar a área dos picos obtida através do cromatograma e a concentração de cada substância, as quais se encontram nos apêndices A, B, C e D. A fase móvel consistiu em uma solução de H_2SO_4 a 0,05 M e a vazão da fase móvel foi igual a 0,6 mL/min.

3.4 ETAPA 2: PROCEDIMENTO EXPERIMENTAL PARA DETERMINAÇÃO DOS PARÂMETROS DA FERMENTAÇÃO VARIANDO A CONCENTRAÇÃO DE GLICOSE

Para esta etapa, foram utilizadas a mesma composição do meio de cultura descrita anteriormente e a mesma espécie de levedura. Os três experimentos realizados encontram-se descritos na Tabela 2, os quais variaram em relação à concentração de glicose do meio.

Tabela 2 – Descrição dos três experimentos realizados

Experimento	Concentração de glicose
F	50 g/L
G	100 g/L
H	150 g/L

Fonte: A autora (2023).

Os biorreatores consistiram em três balões volumétricos de 3 L, os quais receberam 2 L de meio e 200 mL do pré-inóculo preparado nas mesmas condições. As amostras foram retiradas a cada 3 horas nas primeiras 12 horas e uma última amostra após 24 horas de fermentação.

A análise de glicose foi realizada por meio do método do DNSA. Para isso, 0,5 mL da amostra foram transferidos para um tubo de Folin Wu, assim como 1 mL do reagente DNS. Então, o tubo foi aquecido em banho maria a 100 °C por cinco minutos e, posteriormente, resfriados em banho de gelo por dois minutos. Acrescentou-se água destilada até a marca de 12,5 mL no tubo, homogeneizou-se a solução final e mediu-se a absorbância em um espectrofotômetro, com comprimento de onda igual a 540 nm. Foram realizadas diluições quando necessário para que a transmitância se mantivesse na faixa de confiabilidade do equipamento. Por fim, a concentração de glicose foi obtida a partir da equação 5.

$$ABS = 0{,}2983x - 0{,}027 \quad (5)$$

Já a concentração de etanol foi medida através do método da ebuliometria. Primeiramente, um ebuliômetro foi calibrado com água destilada para marcação do teor 0 °GL na régua. Em seguida, 50 mL das amostras após centrifugação foram

adicionados ao equipamento e o teor alcoólico foi medido por meio da régua uma vez que a solução entrou em ebulição.

Por fim, a concentração de biomassa foi determinada por turbidimetria. Para isso, foi realizada uma curva de calibração relacionando a concentração de biomassa com a absorbância em um espectrofotômetro (Equação 6).

$$ABS = 1{,}9519x - 0{,}0358 \quad (6)$$

Então, cada amostra foi centrifugada e lavada com água destilada. Transferiu-se a amostra para um balão volumétrico, o qual teve seu volume completado com água destilada. Mediu-se a absorbância da solução final no espectrofotômetro. Dessa forma, a concentração de biomassa foi determinada, levando em consideração a diluição realizada.

Também, os fatores de conversão foram calculados segundo as equações 7 e 8. As produtividades em célula e em produto foram obtidas por meio das equações 9 e 10 e a eficiência foi determinada pela razão entre o fator de conversão de substrato em produto pelo rendimento teórico (51,11%).

$$Y_{P/S} = \frac{P_{Final} - P_{Inicial}}{S_{Inicial} - S_{Final}} \quad (7)$$

$$Y_{X/S} = \frac{X_{Final} - X_{Inicial}}{S_{Inicial} - S_{Final}} \quad (8)$$

$$Pr_X = \frac{X_{Final} - X_{Inicial}}{\Delta t_{Total}} \quad (9)$$

$$Pr_P = \frac{P_{Final} - P_{Inicial}}{\Delta t_{Total}} \quad (10)$$

3.5 ETAPA 3: METODOLOGIA PARA SIMULAÇÃO DE UM PROCESSO FERMENTATIVO EM BATELADA

Para aplicação do modelo de aprendizado de máquina para detecção de problemas em um processo, é necessário um grande número de dados de forma a garantir a robustez do modelo proposto. Sendo assim, foi utilizado o modelo cinético descrito por Oliveira, Oliveira, Tacin e Gattás (2016) para simular o processo de fermentação alcoólica em batelada em diferentes condições. As equações 11, 12 e

13 descrevem os balanços de massa referente às células, ao produto e ao substrato, respectivamente.

$$\frac{dX}{dt} = \frac{\mu_{max} S}{K_S + S + {S^2}/{K_i}} \left(1 - \frac{P}{P_{max}}\right)^n X \tag{11}$$

$$\frac{dP}{dt} = \alpha \frac{\mu_{max} S}{K_S + S + {S^2}/{K_i}} \left(1 - \frac{P}{P_{max}}\right)^n X \tag{12}$$

$$\frac{dS}{dt} = -\frac{\alpha}{Y_{P/S}} \frac{\mu_{max} S}{K_S + S + {S^2}/{K_i}} \left(1 - \frac{P}{P_{max}}\right)^n X \tag{13}$$

Os valores utilizados para as constantes foram os propostos por Oliveira, Oliveira, Tacin e Gattás (2016). A Tabela 3 traz esses valores.

Tabela 3 – Valores das constantes do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para simulação de um processo em batelada de fermentação alcoólica

Constante	Valor
$\boldsymbol{\mu_{max}}$	0,25 h^{-1}
$\boldsymbol{K_S}$	0,0061 g/L
$\boldsymbol{K_i}$	139,7 g/L
$\boldsymbol{P_{max}}$	94,2 g/L
$\boldsymbol{n}$	4,12
$\boldsymbol{\alpha}$	4,87
$\boldsymbol{Y_{P/S}}$	0,4

Fonte: A autora (2023).

Para simular a contaminação bacteriana no processo, foi adicionado um novo termo de consumo de substrato. A alteração proposta encontra-se na equação 14, sendo o valor utilizado para a constante β igual a 1,2. Foi escolhido um valor acima de uma unidade de forma a simular a competição de substrato entre as leveduras e as bactérias, visto que há um maior consumo de substrato sem formação do produto desejado na mesma proporção em casos de contaminação.

$$\frac{dS}{dt} = -\beta \frac{\alpha}{Y_{P/S}} \frac{\mu_{max} S}{K_S + S + {S^2}/{K_i}} \left(1 - \frac{P}{P_{max}}\right)^n X \tag{14}$$

Para a simulação, foi escrito um código na linguagem de programação Fortran 90 de forma a resolver numericamente as equações de balanço, obtendo os

perfis de concentração de biomassa, de produto e de substrato em função do tempo. O código pode ser encontrado no apêndice E.

Os processos simulados distinguiram-se quanto às concentrações iniciais de substrato e de biomassa e quanto à presença ou não de contaminação. A concentração de substrato variou de 70 a 120 g/L. Já a concentração de biomassa variou de 15 a 35 g/L. A concentração inicial de produto foi igual a zero em todas as simulações.

3.6 ETAPA 4: METODOLOGIA PARA DETECÇÃO DE CONTAMINAÇÃO POR INTELIGÊNCIA ARTIFICIAL

O conjunto de dados consistiu em 87822 pontos de 2142 simulações, sendo metade com contaminação, cada um deles contendo 8 variáveis (Quadro 2). Esses dados foram provenientes da simulação do processo fermentativo em batelada, como descrito no tópico anterior, e correspondem aos diferentes ensaios simulados. Então, ele foi dividido aleatoriamente em dois subconjuntos, sendo 80% reservado para o treinamento do modelo e o restante para validação.

Quadro 2 – Variáveis do banco de dados, seus significados e intervalos de valores possíveis

Variáveis	**Significado**	**Intervalo de valores possíveis**
Rótulo	Classificação da simulação quanto à contaminação	0 (não contaminado) e 1 (contaminado)
X0	Concentração inicial de biomassa (g/L)	15 – 35
S0	Concentração inicial de substrato (g/L)	70 – 120
P0	Concentração inicial de produto (g/L)	0
t	Instante de tempo da simulação (h)	0,2 – 8
X	Concentração de biomassa (g/L)	Variável
S	Concentração de substrato (g/L)	Variável
P	Concentração de produto (g/L)	Variável

Fonte: A autora (2023).

Para aplicação de um método supervisionado de aprendizado de máquina, é necessário determinar qual a variável que se quer predizer. Nesse caso, o objetivo é classificar quanto à contaminação. Então, foi realizado o treinamento de um modelo de redes neurais do tipo MLP, por meio da biblioteca scikit-learn (Pedregosa *et al*., 2011). A performance do modelo foi avaliada através da acurácia e da matriz de confusão da predição do conjunto de validação. Uma vez que o conjunto de dados é

relativamente pequeno, no contexto da inteligência artificial, utilizou-se a validação cruzada para potencialmente melhorar o modelo proposto. Também foi realizado um estudo da melhor combinação de valores para três hiperparâmetros do modelo, o número de camadas ocultas, a taxa de aprendizagem e o tamanho dos mini lotes para otimizadores estocásticos.

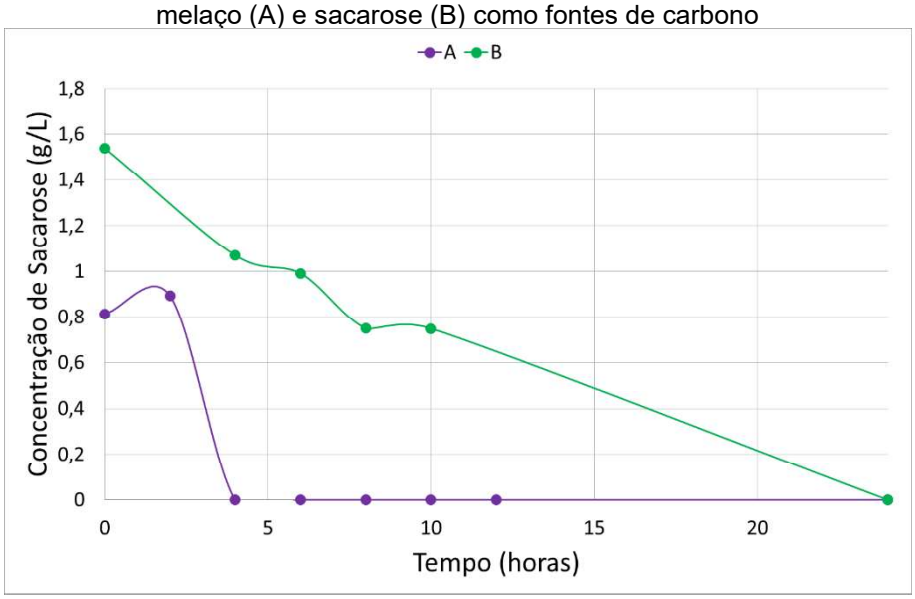
Ainda, foi realizada uma comparação entre o modelo de redes neurais MLP e outros cinco modelos de classificação com validação cruzada, verificando qual deles apresentou melhor desempenho frente ao problema proposto. Os modelos foram a regressão logística, o classificador árvore de decisão, o classificador florestas aleatórias, o classificador *Gradient Boosting* e o classificador k-vizinhos mais próximos, todos presentes da biblioteca scikit-learn (Pedregosa *et al*., 2011). Os códigos utilizados para treinamento, validação e avaliação da performance dos modelos foram implementados na linguagem de programação Python, na plataforma Kaggle, e encontram-se no apêndice F.

4 RESULTADOS E DISCUSSÃO

4.1 ETAPA 1: AVALIAÇÃO DA FERMENTAÇÃO ALCOÓLICA EM DIFERENTES CONDIÇÕES

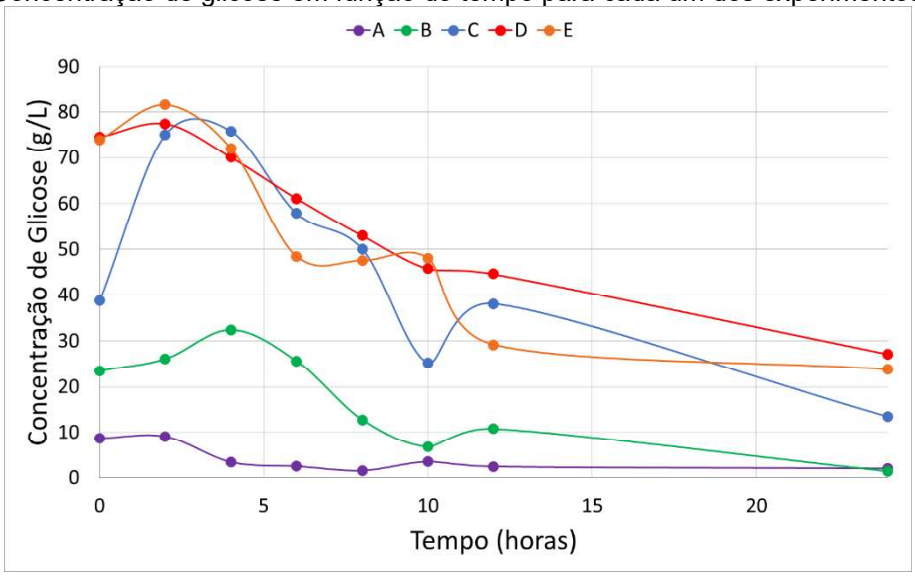
As Figuras 5, 6 e 7 trazem a variação das concentrações dos diferentes tipos de açúcares em função do tempo de fermentação para os cinco experimentos realizados.

Figura 5 – Concentração de sacarose em função do tempo para os experimentos realizados com melaço (A) e sacarose (B) como fontes de carbono



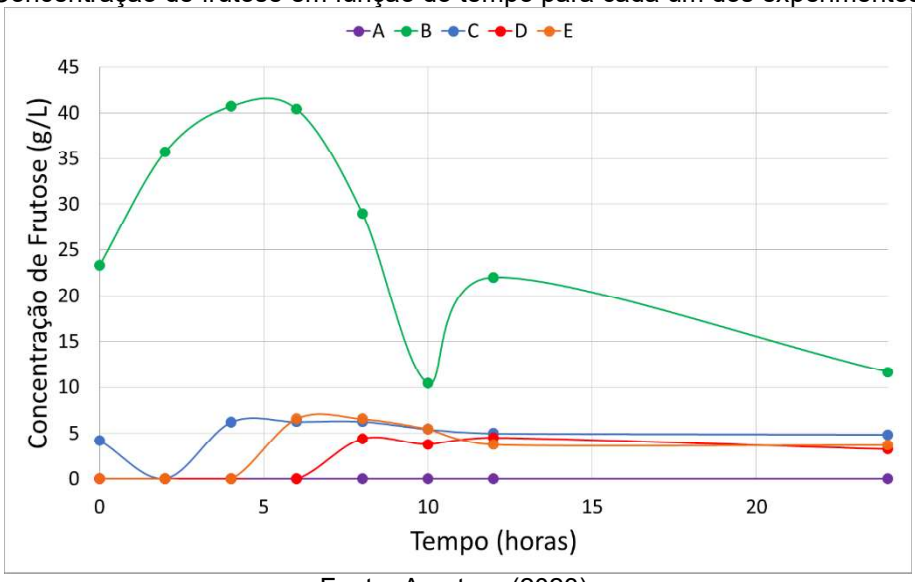
Fonte: A autora (2023).

Figura 6 – Concentração de glicose em função do tempo para cada um dos experimentos realizados



Fonte: A autora (2023).

Figura 7 – Concentração de frutose em função do tempo para cada um dos experimentos realizados



Fonte: A autora (2023).

Através da Figura 5, é possível perceber que a concentração de sacarose é desprezível nos experimentos A e B, cujos meios foram preparados com melaço e sacarose, respectivamente. Dessa forma, esperava-se uma maior concentração de sacarose. Uma hipótese para explicar esse resultado seria a hidrólise desses açúcares durante o processo de autoclavagem, visto que os meios foram preparados e esterilizados uma semana antes da realização dos experimentos.

Para testar essa hipótese, foi medida a concentração de sacarose, glicose e frutose em duas soluções, uma contendo sacarose e outra melaço, antes e após passarem pelo processo de autoclavagem. Os resultados encontram-se na Tabela 4.

Tabela 4 – Concentração de sacarose, glicose e frutose em soluções preparadas com sacarose e melaço antes e após o processo de autoclavagem

Concentração (g/L)	**Solução de sacarose**	**Solução de Melaço**
Sacarose antes	107,23	47,59
Sacarose após	117,75	52,44
Glicose antes	0	11,72
Glicose após	0,63	14,59
Frutose antes	0	10,31
Frutose após	1,96	10,91

Fonte: A autora (2023).

A partir da Tabela 4, observa-se que parte da sacarose foi convertida em glicose e frutose na solução de sacarose, sendo desprezível para a solução de melaço. Contudo, a conversão observada não foi significativa para justificar a

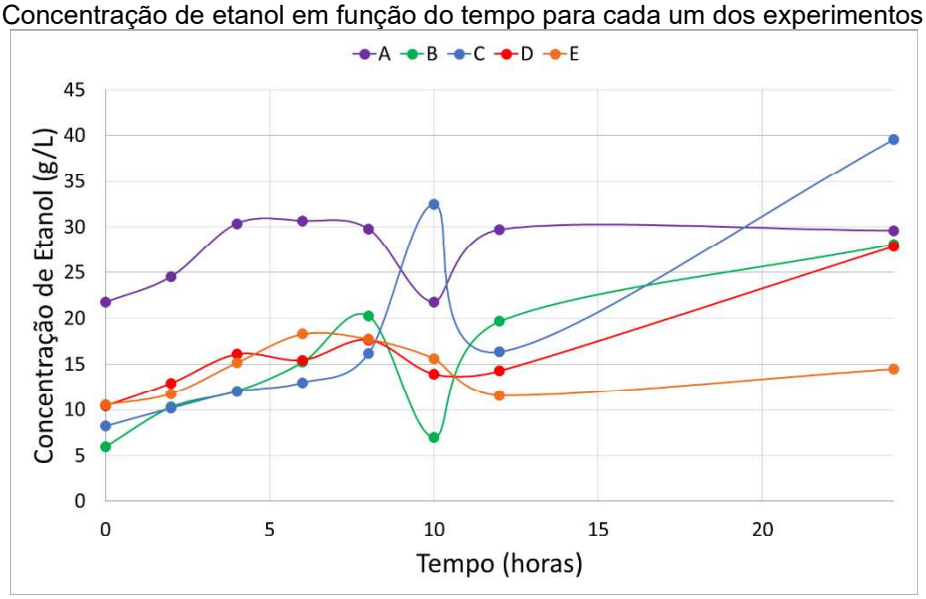
ausência de sacarose nos experimentos. Outra hipótese é que, uma vez que as amostras foram refrigeradas e não congeladas, o processo fermentativo tenha continuado. Então, foi realizada uma análise cromatográfica da amostra correspondente ao tempo t = 0 após 40 dias da realização dos experimentos, a qual apresentava produção de bolhas e presença de leveduras sedimentadas. Verificou-se então que a concentração de açúcares era próxima de zero, indicando que o processo fermentativo continuou nas amostras refrigeradas.

As análises cromatográficas foram realizadas em ordem inversa do tempo de amostragem e levaram duas semanas para serem finalizadas, devido a problemas nos equipamentos. Dessa forma, é possível que a quantificação das substâncias das amostras equivalentes aos tempos iniciais não seja representativa do início da fermentação.

A partir da análise das Figuras 6 e 7, observa-se que houve o consumo da glicose e da frutose ao longo do processo fermentativo. No entanto, percebe-se novamente uma concentração baixa no início do processo.

Além da quantificação dos substratos em cada amostra, foi verificada a produção de etanol em função do tempo para os cinco experimentos realizados (Figura 8).

Figura 8 – Concentração de etanol em função do tempo para cada um dos experimentos realizados



Fonte: A autora (2023).

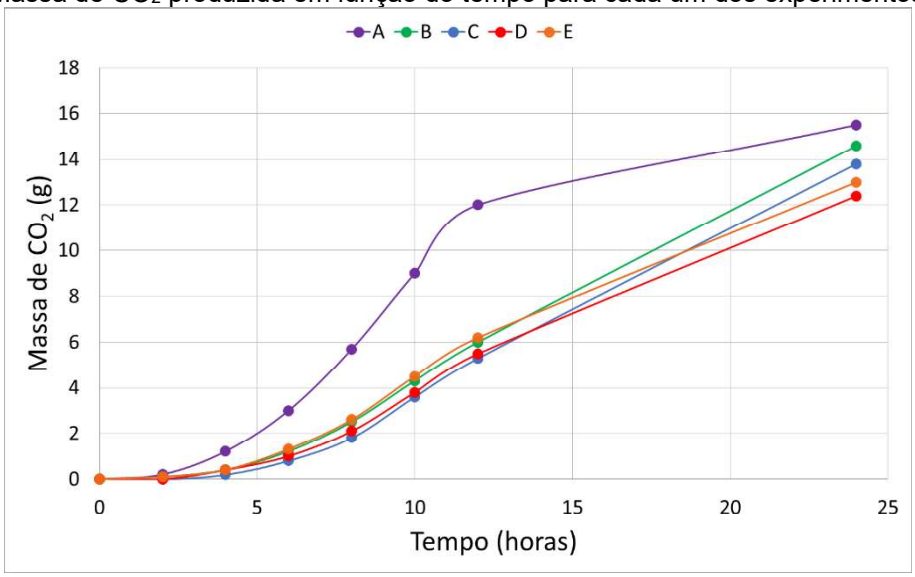
Pode-se observar que a concentração final de etanol para os experimentos realizados com contaminação (D e E) são menores quando comparados com os

sem contaminação (A, B e C), o que está em concordância com o reportado por Carvalho-Netto *et al*. (2015). Esse resultado pode ser explicado pela competição entre a levedura e a bactéria pela fonte de carbono, assim como a inibição do metabolismo da levedura pelos subprodutos da fermentação bacteriana (Tse; Wiens; Reaney, 2021).

Também, pode-se calcular a redução percentual da produção de etanol. Para os experimentos acima, tem-se que essa redução foi equivalente a 63,36% para 50% de contaminação e 29,67% para 25% de contaminação, ambos em relação ao experimento A, que utilizou a mesma fonte de carbono (Glicose). Esses valores são bem maiores do que o reportado por Silva *et al*. (2023), o qual foi igual a 18% de redução da produção de etanol devido à contaminação bacteriana por lactobacilos em um processo de fermentação alcoólica.

Também foi analisada a produção de dióxido de carbono em função do tempo para os cinco experimentos. Os resultados encontram-se na Figura 9.

Figura 9 – Massa de CO_2 produzida em função do tempo para cada um dos experimentos realizados

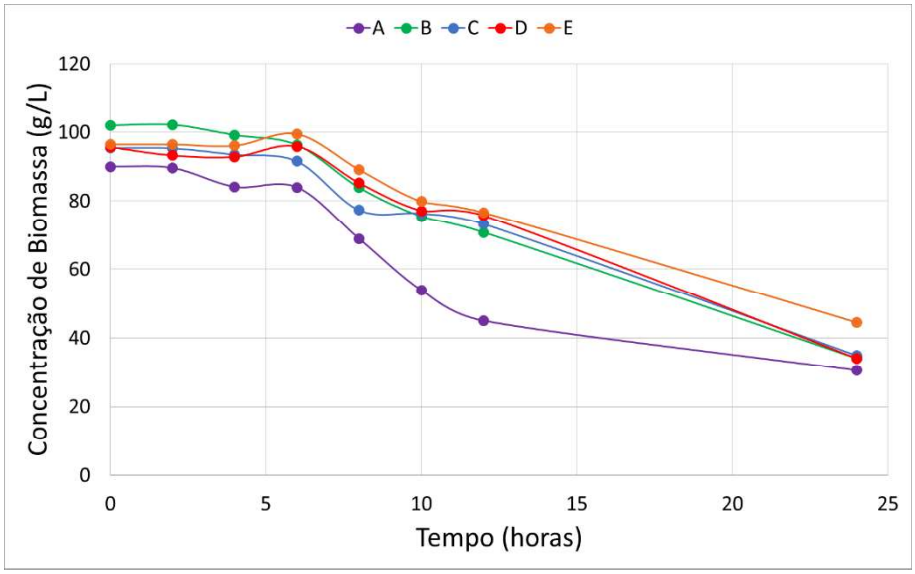


Fonte: A autora (2023).

Novamente, observa-se que houve a uma menor produção de dióxido de carbono ao longo do tempo nos experimentos realizados com contaminação (D e E). Esse resultado pode ser explicado ao analisar o metabolismo da espécie *Bacillus clausii*, utilizada como contaminante. Segundo Bhatawadekar *et al.* (2020), ela é capaz de fermentar glicose, mas sem produção de gás.

A Figura 10 traz a variação da concentração de biomassa em função do tempo para os cinco experimentos realizados. É importante ressaltar que, como essa análise foi realizada através do peso seco, não é possível diferenciar entre a massa de leveduras e a de bactérias.

Figura 10 – Concentração de biomassa em função do tempo para cada um dos experimentos realizados



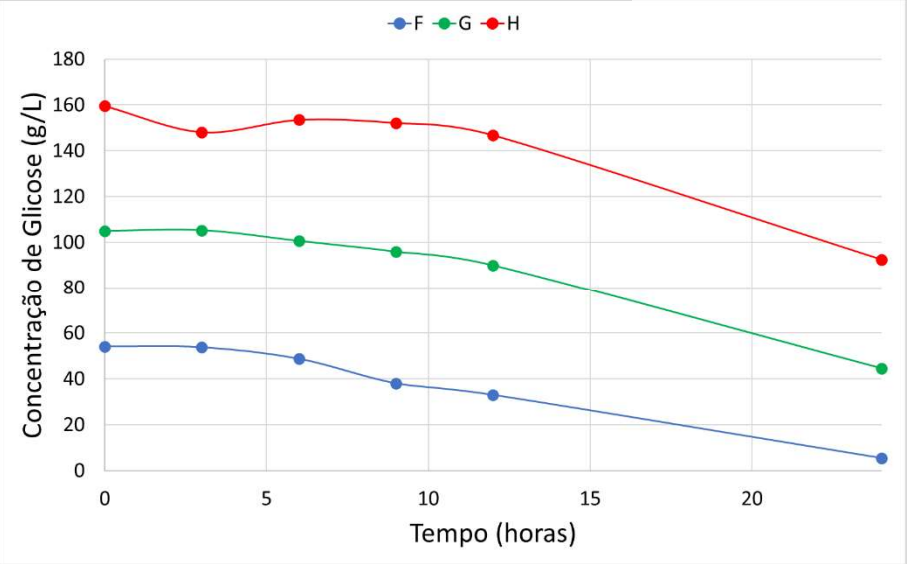
Fonte: A autora (2023).

Uma vez que os perfis de concentração de biomassa indicam a fase estacionária seguida da fase de declínio ou morte celular ao invés do crescimento microbiano, pode-se inferir que a metodologia da medição do peso seco das amostras não foi adequada para quantificação real da concentração de biomassa. Visto que a concentração de açúcar inicial do meio era de 100 g/L, tem-se que essa metodologia proporcionou a quantificação dos sólidos totais, não apenas da biomassa.

4.2 ETAPA 2: DETERMINAÇÃO DOS PARÂMETROS DA FERMENTAÇÃO VARIANDO A CONCENTRAÇÃO DE GLICOSE

A Figura 11 traz a variação da concentração de glicose em função do tempo para os três experimentos realizados.

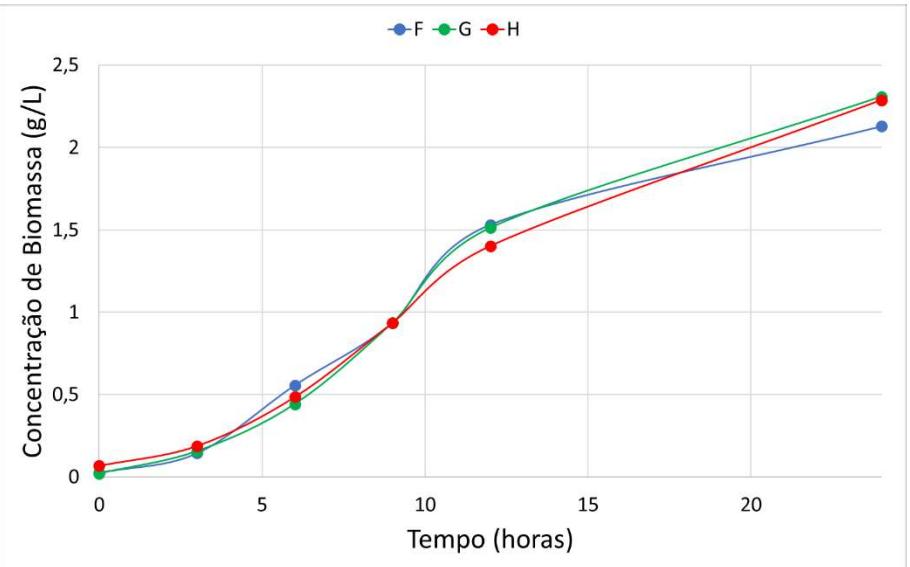
Figura 11 – Concentração de glicose em função do tempo para cada um dos experimentos realizados



Fonte: A autora (2023).

A partir da Figura 11, observa-se o decaimento da concentração de glicose em todos os experimentos devido ao consumo pela levedura dessa fonte de carbono. Também foi analisada a variação da concentração de biomassa a partir do método da turbidimetria (Figura 12).

Figura 12 – Concentração de biomassa em função do tempo para cada um dos experimentos realizados

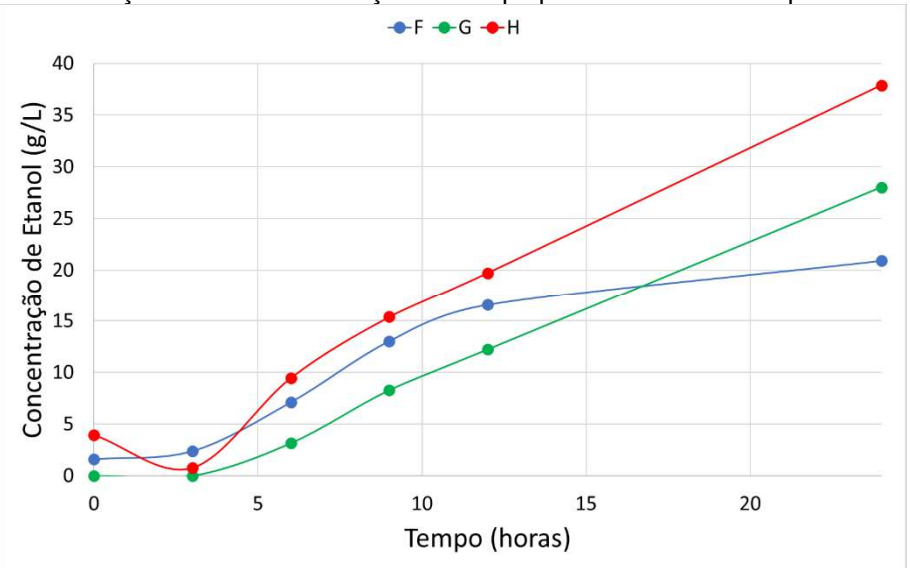


Fonte: A autora (2023).

É possível notar pela Figura 12 que os perfis de concentração de biomassa ao longo do tempo estão coerentes com o perfil esperado para a fase exponencial de crescimento microbiano. Dessa forma, verifica-se uma melhor adequação desse

método em relação à medição do peso seco. Ainda, foi verificada a produção de etanol ao longo do tempo para cada experimento, a qual foi quantificada pelo método da ebuliometria (Figura 13).

Figura 13 – Concentração de etanol em função do tempo para cada um dos experimentos realizados



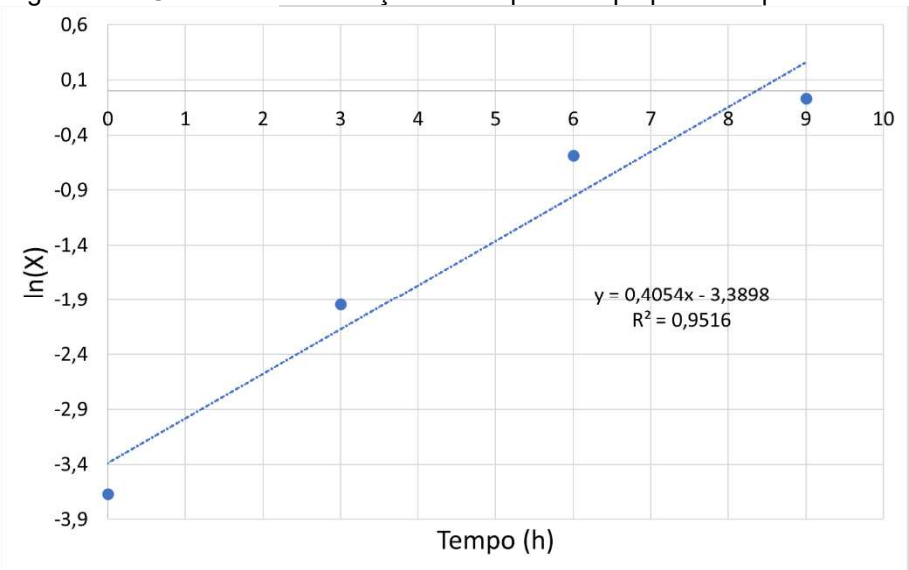
Fonte: A autora (2023).

Por meio da análise da Figura 13, pode-se perceber que o experimento realizado com a maior concentração inicial de substrato (H) levou a uma maior produção de etanol. Assim como o experimento com menor concentração inicial de glicose (F) apresentou a menor produção de etanol. Isso pode ser explicado pelo efeito *Crabtree*, segundo o qual altas concentrações de glicose favorecem o processo fermentativo com produção de etanol por *S. cerevisiae*, mesmo na presença de oxigênio (Zhang *et al.*, 2022).

A *Saccharomyces* é uma levedura que apresenta seu metabolismo em condições aeróbio facultativo, podendo sobreviver na ausência ou presença de O_2. Os produtos finais da metabolização da glicose vão depender das condições ambientais a que estas leveduras estarão submetidas. De uma forma geral, a levedura metaboliza anaerobicamente o açúcar para produzir energia na forma de ATP para manutenção de sua vida, crescimento e multiplicação. Neste processo fermentativo, o CO_2 e etanol são apenas produtos de excreção (Otterstedt *et al.*, 2004; Parapouli; Vasileiadis; Afendra; Hatziloukas, 2020). Assim, adequando as condições de processo é possível obter melhores fatores de conversão de substrato em produto e acompanhar quanto de substrato é convertido em biomassa celular.

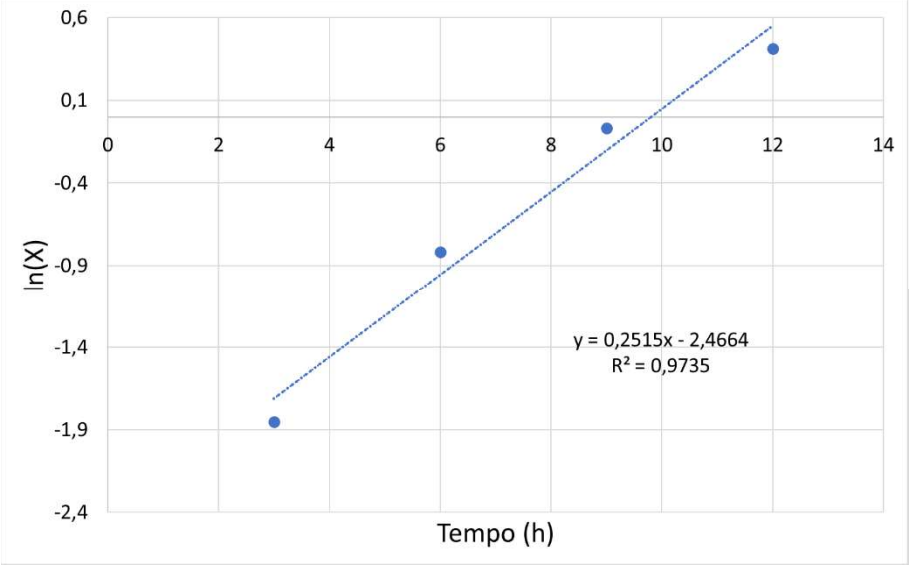
Então, a partir dos dados experimentais acima, foram calculados os parâmetros da fermentação alcoólica. Primeiramente, foi determinada a fase exponencial de crescimento para cada experimento por meio do melhor coeficiente de ajuste da reta para diferentes intervalos do gráfico lnX vs t. As Figuras 14, 15 e 16 trazem o melhor ajuste determinado para os experimentos F, G e H, respectivamente.

Figura 14 – Gráfico da linearização de lnX pelo tempo para o experimento F



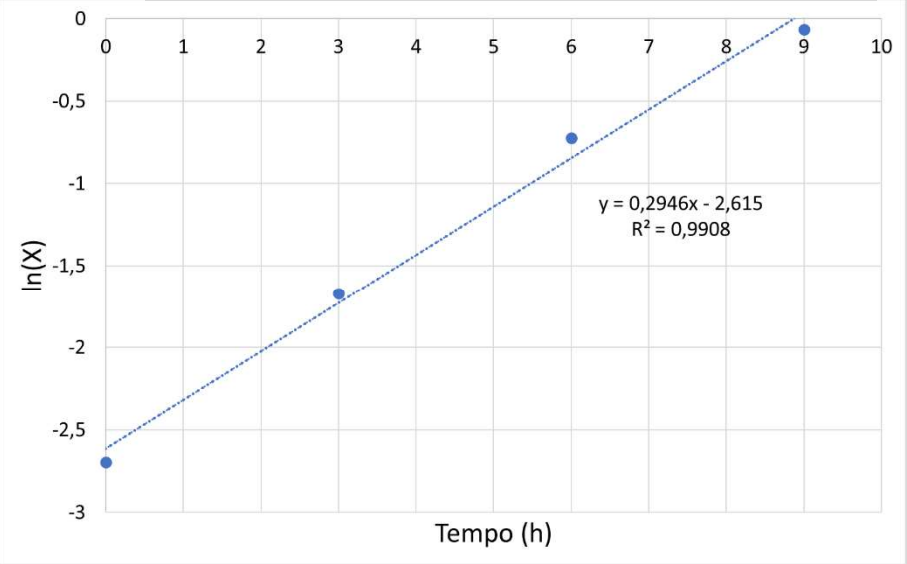
Fonte: A autora (2023).

Figura 15 – Gráfico da linearização de lnX pelo tempo para o experimento G



Fonte: A autora (2023).

Figura 16 – Gráfico da linearização de lnX pelo tempo para o experimento H



Fonte: A autora (2023).

A partir dessas equações, é possível obter o valor da velocidade máxima de crescimento celular, a qual é equivalente à inclinação da reta. Dessa forma, tem-se que essa constante é igual a 0,4054, 0,2515 e 0,2946 h^{-1} para os experimentos F, G e H, com concentrações iniciais de glicose de 50, 100 e 150 g/L, respectivamente.

Para processos fermentativos em batelada, utilizando *Saccharomyces cerevisiae* e cana-de-açúcar, foram reportados valores iguais a 0,5 h^{-1} para μ_{max} (Oliveira; Oliveira; Tacin; Gattás, 2016). Já para o mesmo processo utilizando glicose como substrato, Zentou *et al.* (2019) reportaram valores iguais a 0,179, 0,133, 0,65 e 0,084 h^{-1} para essa constante. Pode-se atribuir as variações observadas do valor de μ_{max} a diferenças nos meios de cultura, variações nas condições de operação, concentração e cepa do microrganismo utilizado.

Também foram calculados os fatores de conversão, os quais se encontram na Tabela 5.

Tabela 5 – Valores dos fatores de conversão de substrato em produto e de substrato em células para três ensaios realizados

Experimento	$Y_{P/S}$	$Y_{X/S}$
F	0,3974	0,0432
G	0,4642	0,0379
H	0,5060	0,0331

Fonte: A autora (2023).

Os valores do fator de conversão de substrato em produto encontram-se próximos 0,40, o qual foi reportado por Oliveira, Oliveira, Tacin e Gattás (2016).

Ainda, avaliou-se as produtividades em célula e em produto, assim como a eficiência. Os resultados encontram-se na Tabela 6.

Tabela 6 – Valores das produtividades em célula e em produto e do rendimento experimental em relação ao teórico para os três ensaios realizados

Experimento	Pr_X (g/L.h)	Pr_P (g/L.h)	Eficiência
F	0,0876	0,8054	77,75%
G	0,0952	1,1671	90,82%
H	0,0925	1,4136	99,00%

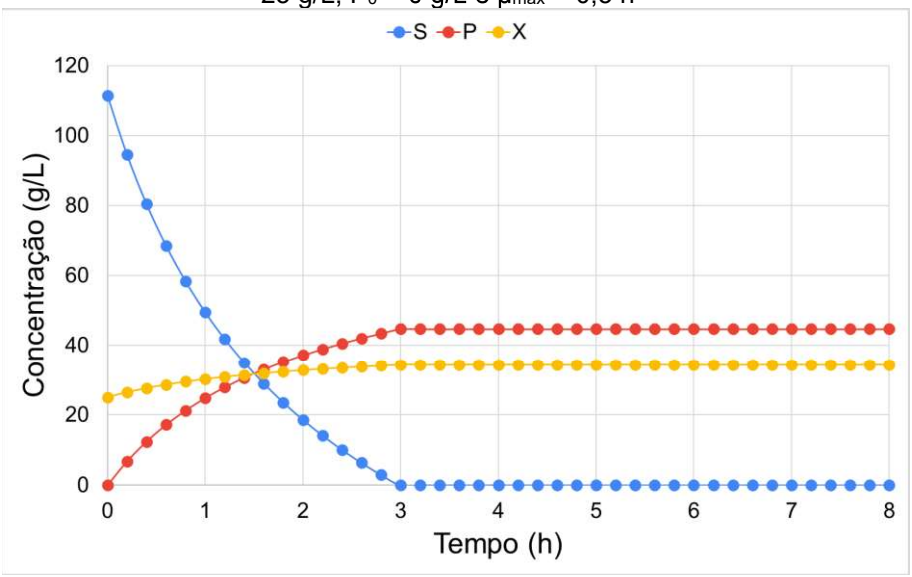
Fonte: A autora (2023).

Observa-se que o rendimento experimental foi muito alto, indicando que quase todo o substrato foi utilizado para produção de etanol. Segundo Silva *et al.* (2023), um rendimento de 90 a 92% pode ser considerado excepcional.

4.3 ETAPA 3: SIMULAÇÃO DE UM PROCESSO FERMENTATIVO EM BATELADA

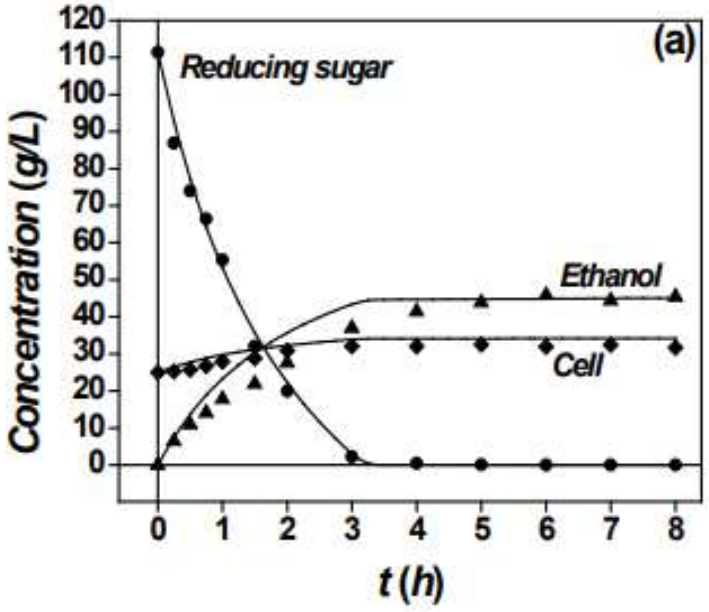
A Figura 17 traz os perfis de concentração de substrato, produto e biomassa obtidos através do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016), utilizando os mesmos valores iniciais que eles utilizaram, de forma a reproduzir o perfil encontrado por eles (Figura 18). Sendo assim, utilizou-se uma concentração inicial de biomassa igual a 25 g/L, uma concentração inicial de substrato igual a 111,5 g/L e uma concentração inicial de produto igual a zero. Também, o valor utilizado para a velocidade máxima de crescimento celular foi de 0,5 h^{-1}.

Figura 17 – Concentrações de biomassa, substrato e produto em função do tempo obtidas através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016), com S_0 = 111,5 g/L, X_0 = 25 g/L, P_0 = 0 g/L e μ_{max} = 0,5 h^{-1}



Fonte: A autora (2023).

Figura 18 – Concentrações de biomassa, substrato e produto em função do tempo obtidas por Oliveira, Oliveira, Tacin e Gattás (2016), com S_0 = 111,5 g/L, X_0 = 25 g/L, P_0 = 0 g/L e μ_{max} = 0,5 h^{-1}

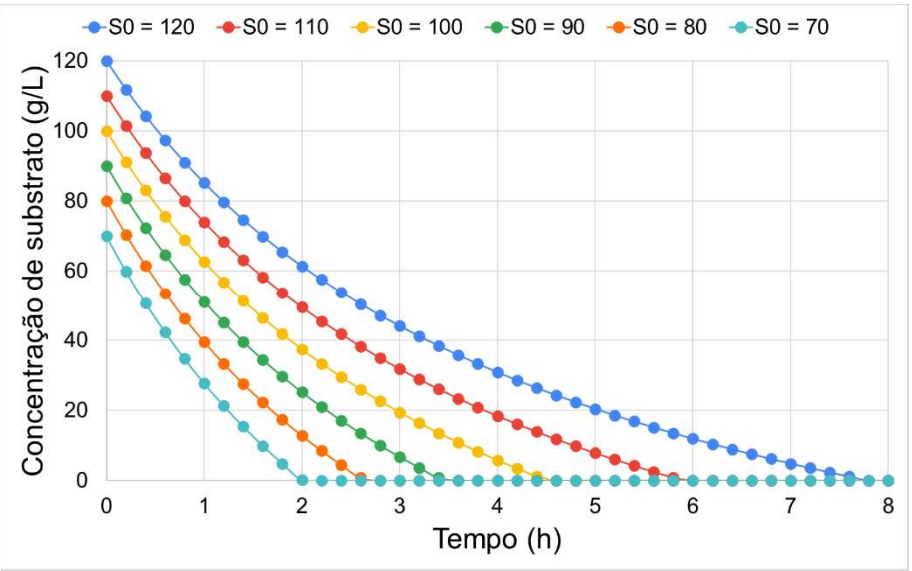


Fonte: Oliveira, Oliveira, Tacin e Gattás (2016, p. 3).

Observa-se que o código foi capaz de reproduzir os dados obtidos por Oliveira, Oliveira, Tacin e Gattás (2016). Sendo assim, foi possível extrapolar o modelo para diferentes concentrações iniciais de substrato e de biomassa, assim como a presença ou ausência de contaminação. As Figuras 19, 20 e 21 traz a comparação entre os perfis de concentração de substrato, produto e biomassa,

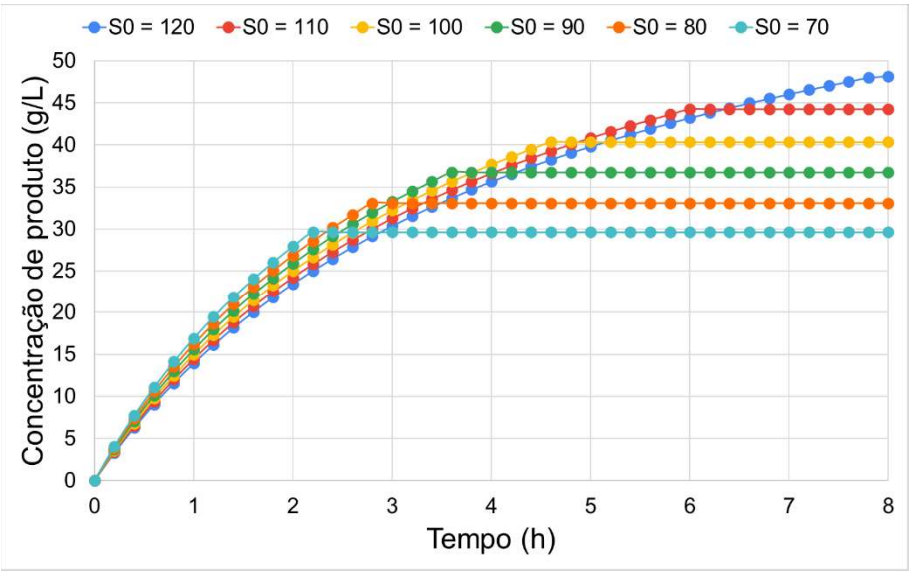
respectivamente, para diferentes valores de concentração inicial de substrato, com concentração inicial de biomassa fixa em 25 g/L.

Figura 19 – Concentração de substrato em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato



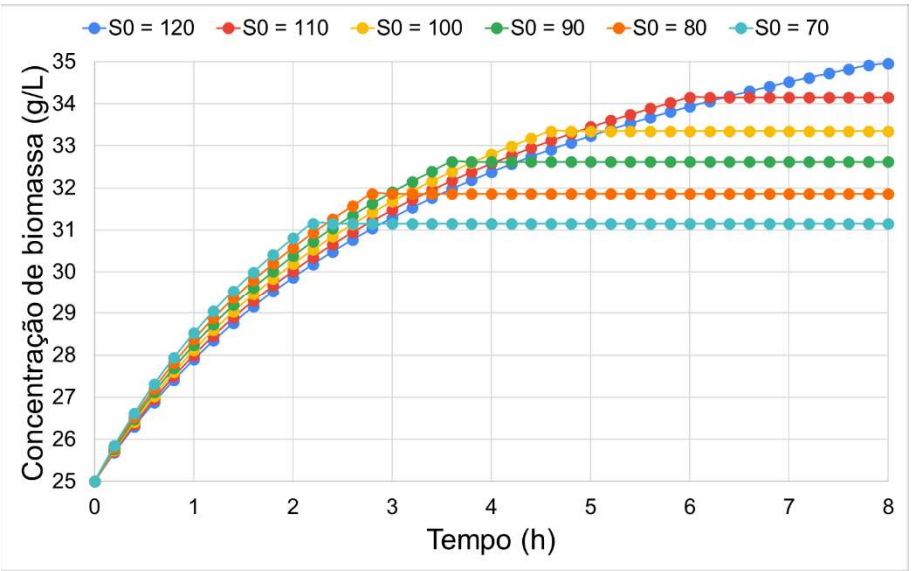
Fonte: A autora (2023).

Figura 20 – Concentração de produto em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato



Fonte: A autora (2023).

Figura 21 – Concentração de biomassa em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de substrato

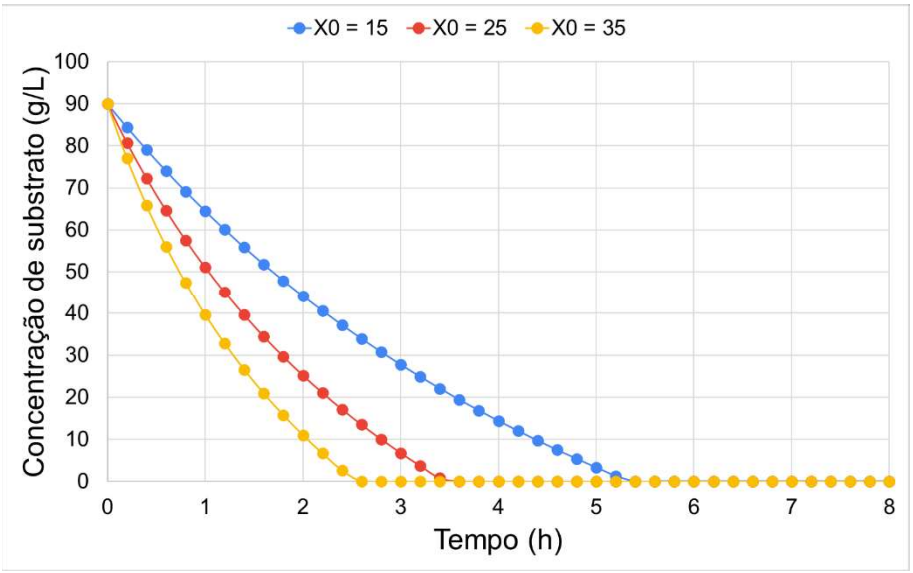


Fonte: A autora (2023).

Pode-se observar por meio das Figuras acima que o aumento da concentração inicial de substrato causa aumentos no tempo necessário para se consumir todo o substrato, na produção de etanol e no crescimento celular. Esses resultados encontram-se de acordo com o observado experimentalmente neste trabalho.

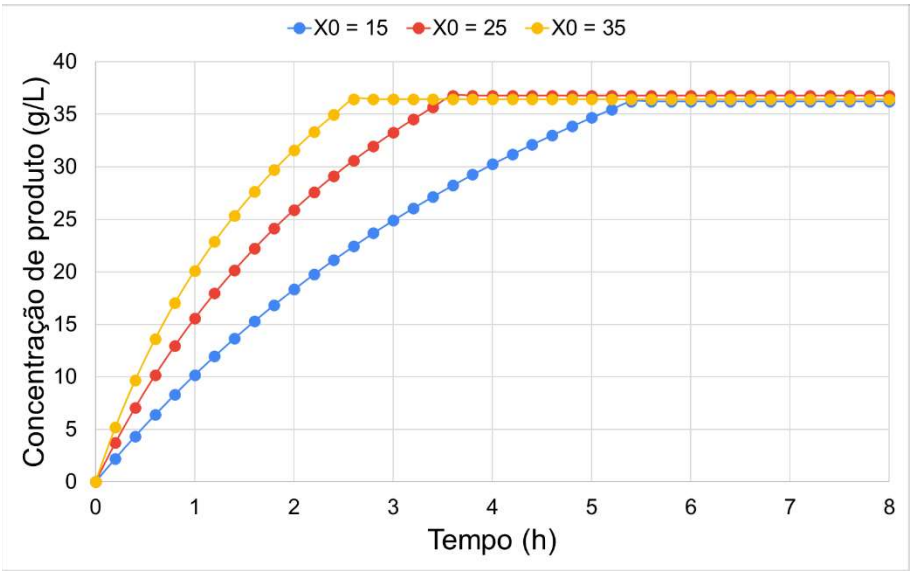
Também foi analisada a influência da concentração inicial de biomassa nos perfis de concentração de substrato, produto e biomassa em função do tempo, mantendo-se a concentração inicial de substrato fixa em 90 g/L (Figuras 22, 23 e 24).

Figura 22 – Concentração de substrato em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa



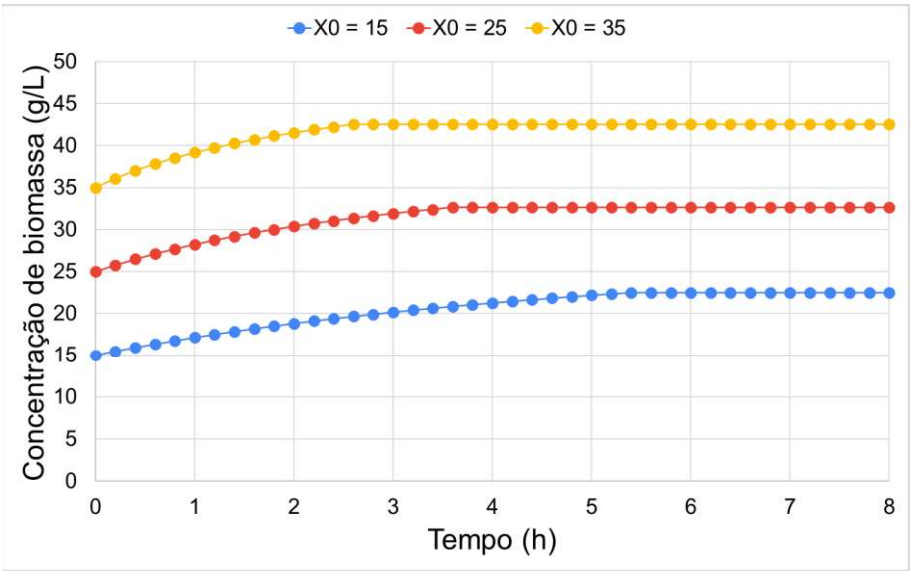
Fonte: A autora (2023).

Figura 23 – Concentração de produto em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa



Fonte: A autora (2023).

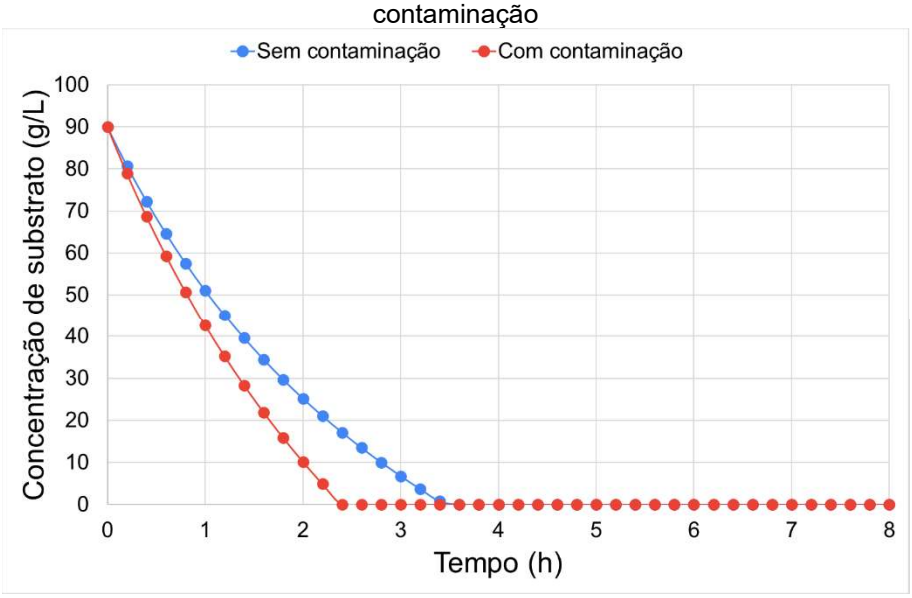
Figura 24 – Concentração de biomassa em função do tempo obtida através da simulação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para diferentes valores de concentração inicial de biomassa



Fonte: A autora (2023).

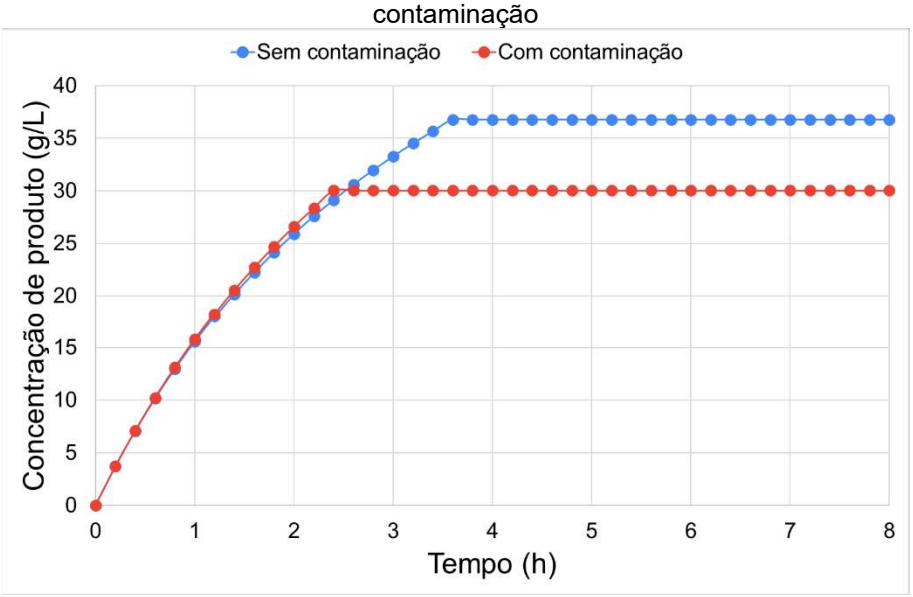
Observa-se que a redução na concentração de biomassa resulta em uma diminuição na velocidade de consumo de substrato e de formação de produto, o que era esperado visto que há menos células para metabolizar os açúcares. Ainda, verificou-se a influência da contaminação nos perfis de substrato, produto e biomassa em função do tempo, com X_0 = 25 g/L e S_0 =90 g/L (Figuras 25, 26 e 27).

Figura 25 – Concentração de substrato em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação



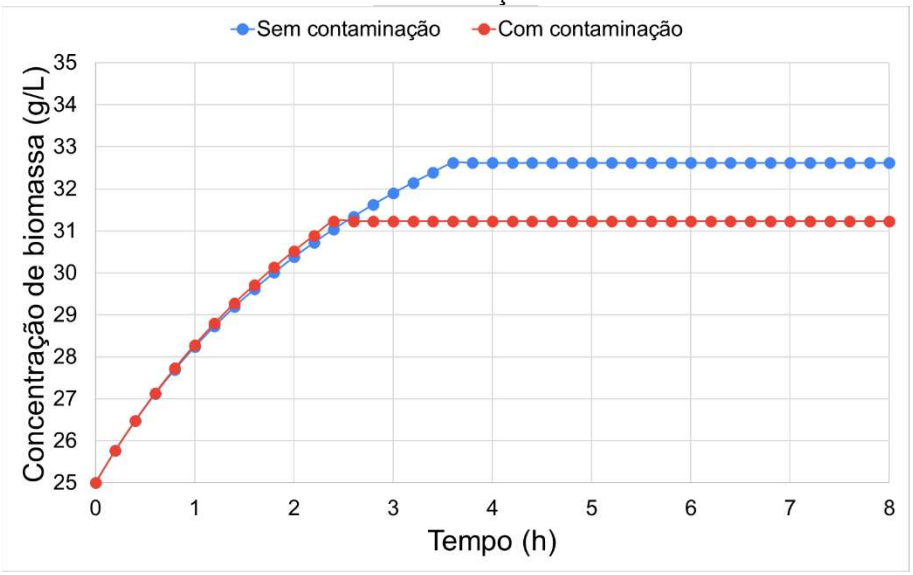
Fonte: A autora (2023).

Figura 26 – Concentração de produto em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação



Fonte: A autora (2023).

Figura 27 – Concentração de biomassa em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) com e sem contaminação



Fonte: A autora (2023).

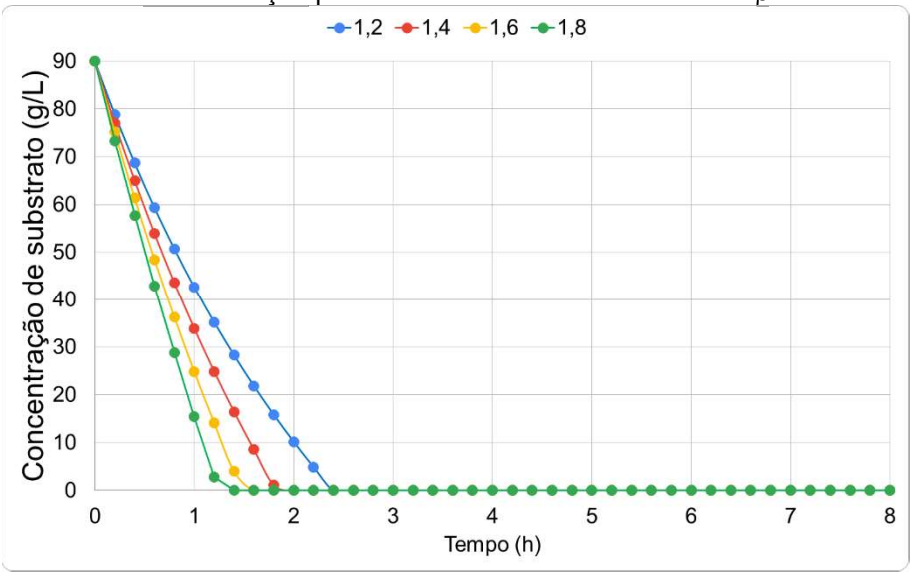
Assim como o observado experimentalmente neste trabalho, a contaminação causa uma redução na produção de etanol para uma mesma concentração inicial de substrato, visto que parte do substrato é metabolizada pela espécie contaminante de modo a formar outros produtos.

A redução da concentração de leveduras em casos de contaminação também é esperada, uma vez que as substâncias produzidas pelas bactérias contaminantes podem inibir o crescimento da *S. cerevisiae*, como reportado por Seo *et al.* (2020).

Para o modelo em questão, a concentração de biomassa diz respeito somente à levedura. Experimentalmente, não foi possível reproduzir esse resultado, uma vez que a metodologia da variação do peso seco não permite a diferenciação entre os microrganismos.

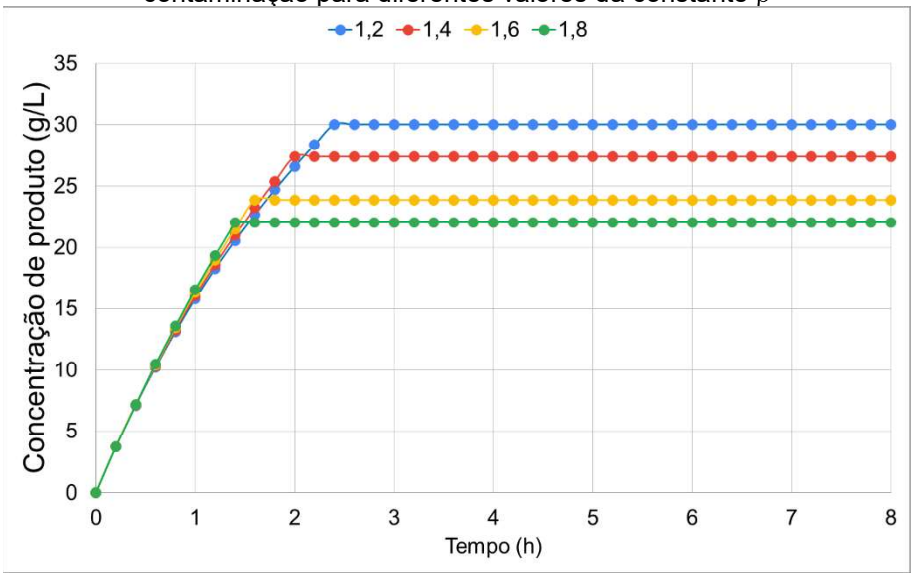
Por fim, foi verificado o impacto do valor da constante β nos perfis de substrato, produto e biomassa ao longo do tempo (Figuras 28, 29 e 30).

Figura 28 – Concentração de substrato em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β



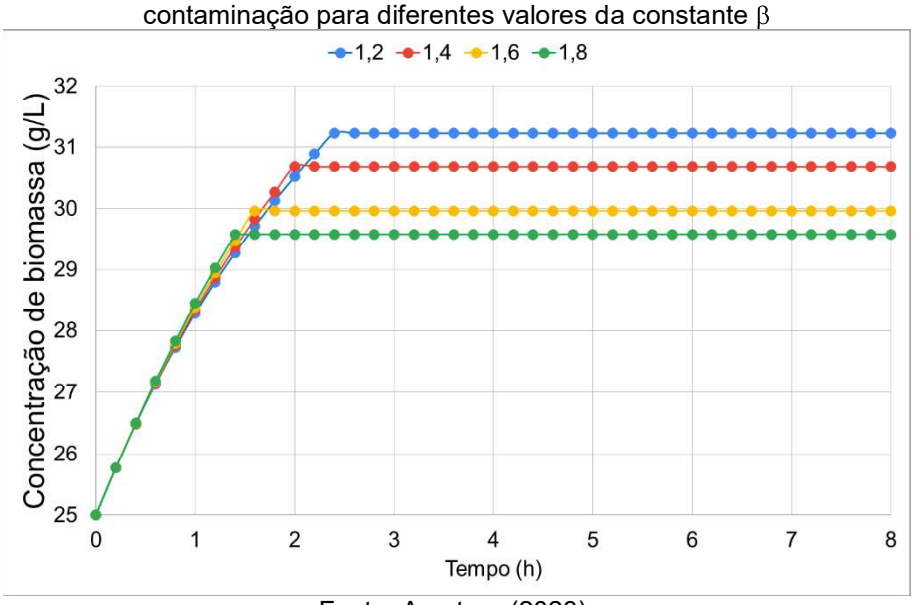
Fonte: A autora (2023).

Figura 29 – Concentração de produto em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β



Fonte: A autora (2023).

Figura 30 – Concentração de biomassa em função do tempo obtida através da simulação da modificação do modelo proposto por Oliveira, Oliveira, Tacin e Gattás (2016) para representar a contaminação para diferentes valores da constante β



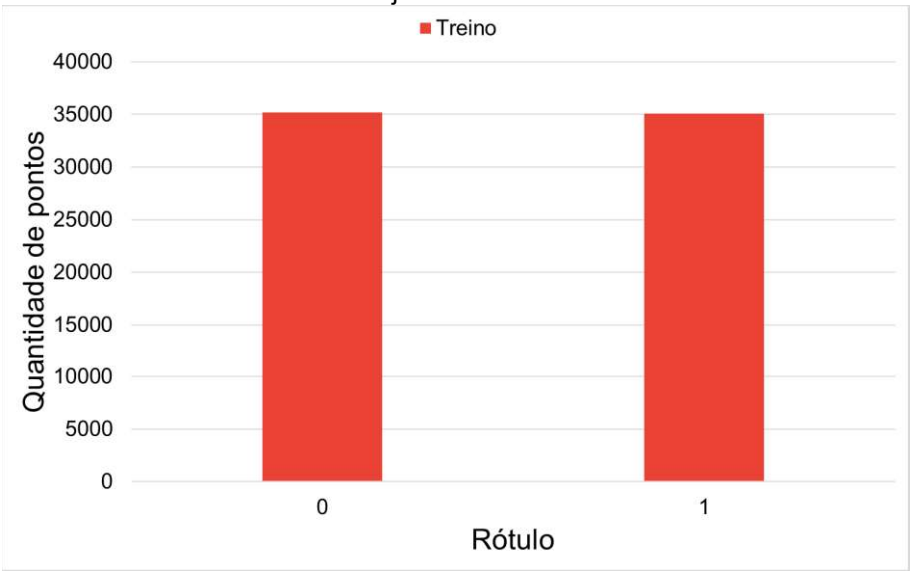
Fonte: A autora (2023).

É possível observar a mesma tendência independentemente do valor da constante, uma menor produção de etanol e uma inibição do crescimento da levedura. Sendo assim, poderia ter sido escolhido qualquer valor acima de 1 para β. Isso porque o intuito da simulação é gerar dados para validar o uso de aprendizado de máquina na identificação dessas alterações.

4.4 ETAPA 4: IDENTIFICAÇÃO DE CONTAMINAÇÃO EM PROCESSOS FERMENTATIVOS POR INTELIGÊNCIA ARTIFICIAL

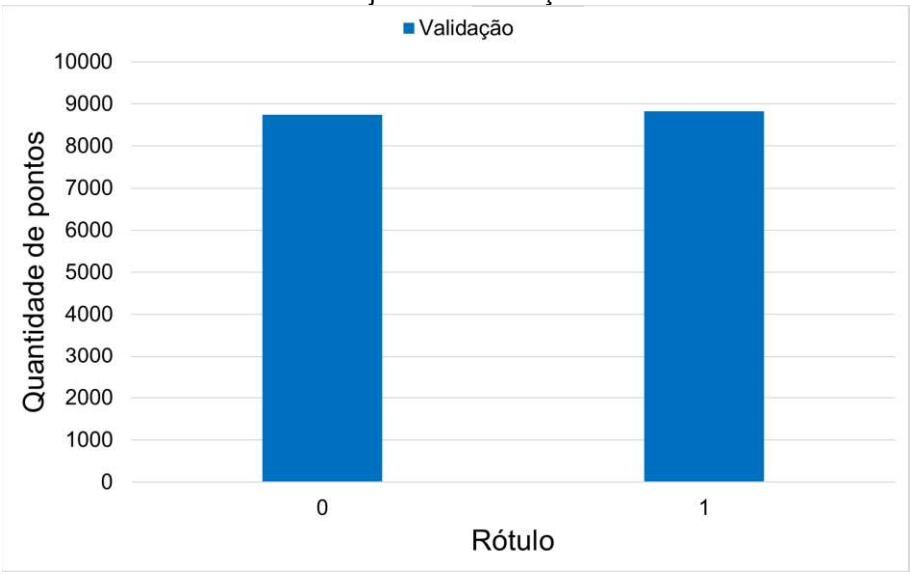
As Figuras 31 e 32 trazem as proporções de dados provenientes de simulações sem contaminação e com contaminação para os conjuntos de treino e validação, respectivamente. O rótulo 0 refere-se a simulações sem contaminação e o rótulo 1 indica existência de contaminação.

Figura 31 – Quantidade de pontos provenientes de simulações com (1) e sem (0) contaminação no conjunto de treino



Fonte: A autora (2023).

Figura 32 – Quantidade de pontos provenientes de simulações com (1) e sem (0) contaminação no conjunto de validação



Fonte: A autora (2023).

Observa-se que ambos os conjuntos apresentam proporções entre as duas classificações muito semelhantes, o que era de se esperar visto que a escolha dos pontos foi feita aleatoriamente. Então, utilizou-se o conjunto de treino para o treinamento de um modelo de redes neurais MLP com o valor *default* da biblioteca para todos os hiperparâmetros (Apêndice G). As acurácias obtidas para os conjuntos de treino e validação encontram-se na Tabela 7.

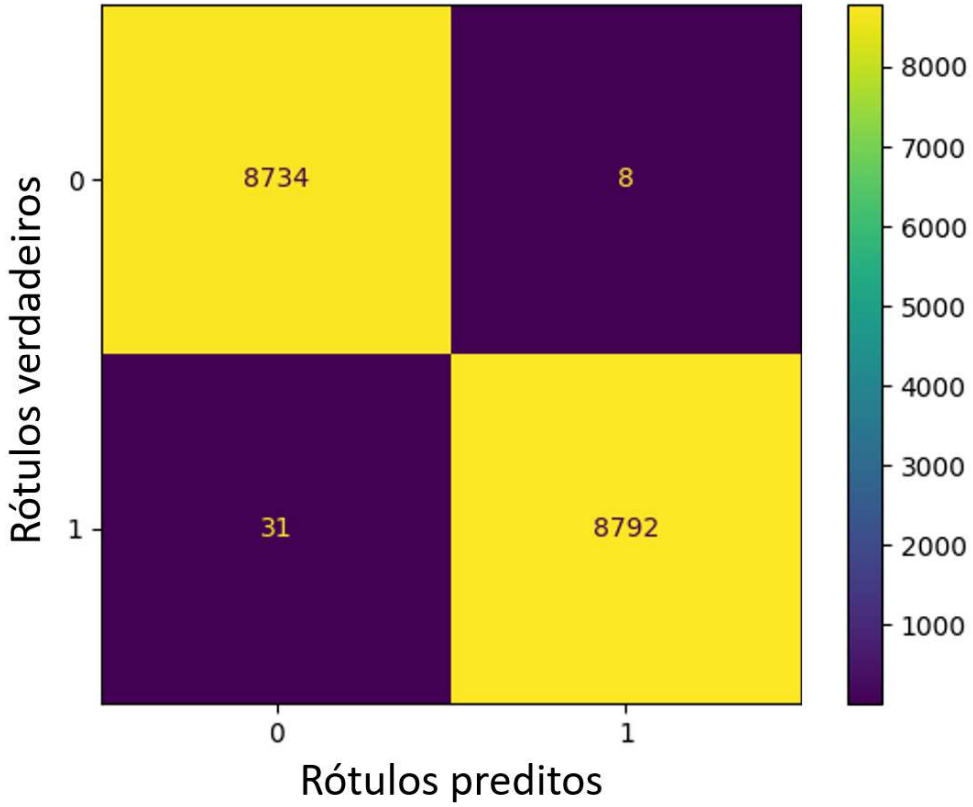
Tabela 7 – Acurácias obtidas para os conjuntos de teste e de validação com o modelo de redes neurais MLP

Conjuntos	Acurácia
Treino	0,9980
Validação	0,9978

Fonte: A autora (2023).

Para ambos os conjuntos, a acurácia do modelo foi muito próxima de 1, indicando que o modelo possui uma excelente capacidade preditiva. Também foi avaliada a matriz de confusão para o conjunto de validação (Figura 33).

Figura 33 – Matriz de confusão para o conjunto de validação com modelo de redes neurais MLP



Fonte: A autora (2023).

A partir da Figura 33, é possível verificar que as predições incorretas do modelo (em roxo) corresponderam majoritariamente à classificação como sem contaminação em situação com contaminação. A fim de melhorar o resultado obtido, foi utilizada a validação cruzada para ajuste de um novo modelo de redes neurais do tipo MLP. A acurácia de cada iteração encontra-se na Tabela 8, assim como a média das cinco iterações e o desvio padrão.

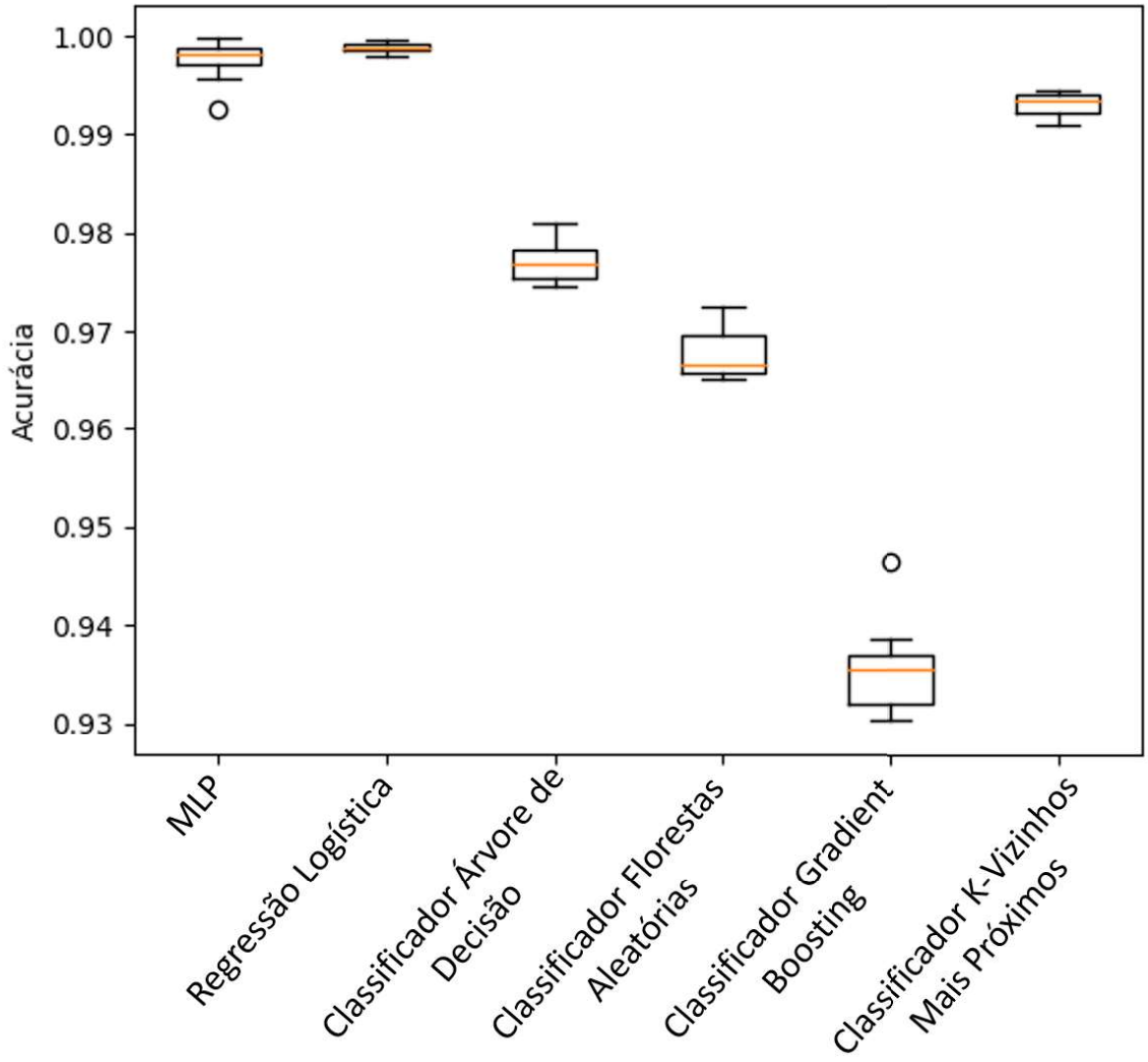
Tabela 8 – Acurácias obtidas para cada uma das iterações da metodologia de validação cruzada com o modelo de redes neurais MLP, a média e o desvio padrão desses valores

Iteração 1	Iteração 2	Iteração 3	Iteração 4	Iteração 5	Média	Desvio padrão
0,9982	0,9989	0,9978	0,9971	0,9984	0,9981	0,0006

Fonte: A autora (2023).

Comparando com o valor obtido sem validação cruzada (0,9978), observa-se que só não foi observado melhora em uma das iterações. No geral, obteve-se uma melhor acurácia aplicando esse método. Também, utilizou-se outros cinco modelos de classificação com o intuito de comparar a performance de cada um deles. A Figura 34 traz a comparação entre a acurácia obtida em cada caso.

Figura 34 – Comparação entre a acurácia obtida ao utilizar diferentes modelos de classificação



Fonte: A autora (2023).

É possível perceber que o modelo que apresentou a melhor performance foi o de regressão logística, seguido pelas redes neurais MLP. Por fim, foi realizada uma sintonização dos hiperparâmetros do modelo MLP, sendo eles o número de camadas ocultas, a taxa de aprendizagem e o tamanho dos mini lotes para otimizadores estocásticos. A combinação que apresentou melhor desempenho encontra-se na Tabela 9.

Tabela 9 – Combinação dos hiperparâmetros do modelo MLP que apresentou melhor desempenho e a acurácia da validação cruzada

Número de camadas ocultas	**Taxa de aprendizagem**	**Tamanho dos mini lotes**	**Acurácia**
47	0,00101	187	0,99895

Fonte: A autora (ano).

Observa-se uma pequena melhora na acurácia, apesar de que o modelo já apresentava um ótimo desempenho. Dessa forma, foi verificado que é possível aplicar modelos de aprendizado de máquina na detecção de problemas em processos industriais. No entanto, para que o modelo proposto possa ser aplicado diretamente a processos reais, é necessário um banco de dados maior e, principalmente, mais diversificado para treiná-lo. Esse banco de dados deve ser representativo da realidade, de forma a garantir uma maior aplicabilidade do modelo em diferentes circunstâncias.

5 CONCLUSÃO

O presente trabalho teve como objetivo principal estudar o processo fermentativo de produção do etanol em diferentes condições e aplicar modelos de aprendizado de máquina na identificação de problemas nesse processo industrial. Dessa forma, foram realizados cinco experimentos utilizando a levedura *Saccharomyces cerevisiae* com diferentes fontes de carbono e sem ou com contaminação, simulada através da bactéria *Bacillus clausii*. No entanto, verificou-se que a metodologia de medição da biomassa por peso seco não foi adequada. Ainda, foi observado que a fermentação continuou nas amostras refrigeradas, ao longo das duas semanas em que as análises cromatográficas foram realizadas. Sendo assim, sugere-se um ajuste da metodologia com congelamento das amostras, garantindo a parada do processo fermentativo. Também foi observado que a presença de contaminação bacteriana levou a uma menor produção de etanol, o que pode ser explicado pela competição entre a levedura e a bactéria pelo substrato, o qual é metabolizado de maneiras diferentes por cada uma delas.

Outros experimentos foram realizados com a mesma espécie de levedura e diferentes concentrações de glicose no meio, utilizando novas metodologias para determinação da concentração de biomassa, de glicose e de etanol. A partir dos resultados, foram determinadas as velocidades específicas máximas para cada experimento, as quais foram iguais a 0,4054, 0,2515 e 0,2946 h^{-1} para os experimentos com concentração inicial de glicose iguais a 50, 100 e 150 g/L, respectivamente. Ainda, foram calculados os fatores de conversão de substrato em produto e células, as produtividades em produto e células e o rendimento. Os dois experimentos com maiores concentrações iniciais de glicose apresentaram excelentes resultados, com rendimentos acima de 90%.

Também foi desenvolvido um código na linguagem de programação Fortran 90 para simular o processo fermentativo a partir do modelo cinético proposto por Oliveira, Oliveira, Tacin e Gattás (2016). Verificou-se que o programa foi capaz de reproduzir os dados obtidos por ele, permitindo assim a extrapolação para diferentes valores de concentração inicial de biomassa e de substrato. Ainda, foi implementada uma modificação no modelo para simular a contaminação no processo, a partir da adição de um termo que leve a um maior consumo de substrato sem correspondente conversão em produto.

Com os dados simulados, utilizou-se um modelo de aprendizado de máquina por redes neurais do tipo MLP para o seguinte problema de classificação, a presença ou não de contaminação no sistema, obtendo uma acurácia igual a 0,9978. Então, aplicou-se a metodologia da validação cruzada para melhorar esse resultado, uma vez que o banco de dados utilizado não foi muito grande. Verificou-se uma melhora, com acurácia igual a 0,9981. Também se avaliou a aplicação de diferentes modelos com validação cruzada, sendo a regressão logística o que apresentou melhor desempenho, seguida do MLP. Por fim, foi realizada uma sintonização de três hiperparâmetros do modelo MLP, levando a uma acurácia final igual a 0,99895. Sendo assim, observa-se o grande potencial dessa ferramenta na detecção de problemas em processos industriais, ressaltando a importância de um grande número de dados diversificados para treinamento.

REFERÊNCIAS

AMERICANO DA COSTA, M. V.; PASAMONTES, M.; NORMEY-RICO, J. E.; GUZMAN, J. L.; BERENGUEL, M. Advanced control strategy combined with solar cooling for improving ethanol production in fermentation units. **Industrial & Engineering Chemistry Research**, v. 53, n. 28, p. 11384-11392, 2014.

AMIRIAN, M. M.; IRWIN, A. J.; FINKEL, Z. V. Extending the Monod model of microbal growth with memory. **Frontiers in Marine Science**, v. 9, p. 963734, 2022.

BARCHI, A. C.; ITO, S.; ESCARAMBONI, B.; DE OLIVA NETO, P.; HERCULANO, R. D.; MIRANDA, M. C. R.; PASSALIA, F. J.; ROCHA, J. C.; NÚÑEZ, E. G. F. Artificial intelligence approach based on near-infrared spectral data for monitoring of solid-state fermentation. **Process Biochemistry**, v. 51, n.10, p. 1338-1347, 2016.

BASSO, L. C.; BASSO, T. O.; ROCHA, S. N. Ethanol production in Brazil: the industrial process and its impact on yeast fermentation. **Biofuel Production-Recent Developments and Prospects**, v. 1530, p. 85-100, 2011.

BHATAWADEKAR, M.; GAIKAR, S.; HADKAR, P.; HASHMI, R.; KHAN, F.; RANE, R. Single Cell Protein Production with Lactobacillus Species and Bacillus Clausii on Media Containing Sugarcane Bagasse Extract & Fish Waste Hydrolysate. **International Journal of Emerging Technologies and Innovative Research**, v. 7, n. 11, p. 66-71, 2020.

BORGES, P. C. S. **Otimização dinâmica da fermentação alcoólica no processo em batelada alimentada.** 2008. 162f. Dissertação (Mestrado em Engenharias) Universidade Federal de Uberlândia, Uberlândia, MG, 2008. Disponível em: https://repositorio.ufu.br/handle/123456789/15105

BRANCO, P.; DINIZ, M.; ALBERGARIA, H. Evaluation of the biocontrol potential of a commercial yeast starter against fuel-ethanol fermentation contaminants. **Fermentation**, v. 8, n. 5, p. 233, 2022.

CAÑAS, H.; MULA, J.; DÍAZ-MADROÑERO, M.; CAMPUZANO-BOLARÍN, F. Implementing industry 4.0 principles. **Computers & Industrial Engineering**, v. 158, p. 107379-107395, 2021.

CARVALHO-NETTO, O. V.; CARAZZOLLE, M. F.; MOFATTO, L. S.; TEIXEIRA, P. J.; NORONHA, M. F., CALDERÓN; L. A.; MIECZKOWSKI, P. A.; ARGUESO, J. L.; PEREIRA, G. A. Saccharomyces cerevisiae transcriptional reprograming due to bacterial contamination during industrial scale bioethanol production. **Microbial Cell Factories**, v. 14, p. 1-13, 2015.

CIOFFI, R.; TRAVAGLIONI, M.; PISCITELLI, G.; PETRILLO, A.; DE FELICE, F. Artificial intelligence and machine learning applications in smart production: Progress, trends, and directions. **Sustainability**, v. 12, n. 2, p. 492-517, 2020.

DAS, S.; DEY, A.; PAL, A.; ROY, N. Applications of artificial intelligence in machine learning: review and prospect. **International Journal of Computer Applications**, v. 115, n. 9, p. 31-41, 2015.

DEVOS, R. J. B.; COLLA, L. M. Simultaneous saccharification and fermentation to obtain bioethanol: a bibliometric and systematic study. **Bioresource Technology Reports**, v. 17, p. 100924, 2022.

DU, Y-H.; WANG, M. Y.; YANG, L. H.; TONG, L. L.; GUO, D. S.; JI, X. J. Optimization and scale-up of fermentation processes driven by models. **Bioengineering**, v. 9, n. 9, p. 473, 2022.

HAENLEIN, M.; KAPLAN, A. A brief history of artificial intelligence: On the past, present, and future of artificial intelligence. **California management review**, v. 61, n. 4, p. 5-14, 2019.

HEO, S.; LEE, J. H. Fault detection and classification using artificial neural networks. **IFAC-PapersOnLine**, v. 51, n. 18, p. 470-475, 2018.

KHAN, M. S. I.; ISLAM, N.; UDDIN, J.; ISLAM, S.; NASIR, M. K. Water quality prediction and classification based on principal component regression and gradient boosting classifier approach. **Journal of King Saud University-Computer and Information Sciences**, v. 34, n. 8, p. 4773-4781, 2022.

KHANZODE, K. C. A.; SARODE, R. D. Advantages and disadvantages of artificial intelligence and machine learning: A literature review. **International Journal of Library & Information Science**, v. 9, n. 1, p. 30-36, 2020.

LEE, Y.; CHO, H. J.; CHOI, J. I.; WOO, H. M. Hybrid Embden–Meyerhof–Parnas pathway for reducing CO2 loss and increasing the Acetyl-CoA levels during microbial fermentation. **ACS Sustainable Chemistry & Engineering**, v. 9, n. 36, p. 12394-12405, 2021.

LINEWEAVER, H.; BURK, D. The determination of enzyme dissociation constants. **Journal of the American Chemical Society**, v. 56, n. 3, p. 658-666, 1934.

LIU, D.; XU, L.; XIONG, W.; ZHANG, H. T.; LIN, C. C.; JIANG, L.; XU, B. Fermentation process modelling with Levenberg-Marquardt algorithm and Runge-Kutta method on etanol production by Saccharomyces cerevisiae. **Mathematical Problems in Engineering**, v. 2014, p. 1-10, 2014.

LO, N. G.; FLAUS, J. M.; ADROT, O. Review of machine learning approaches in fault diagnosis Applied to IoT systems. **2019 International Conference on Control, Automation and Diagnosis (ICCAD)**. IEEE, pp. 1-6, 2019. Disponível em: https://doi.org/10.1109/ICCAD46983.2019.9037949

MADIGAN, M. T.; MARTINKO, J. M.; BENDER, K. S.; BUCKLEY, D. H.; STAHL, D. A. **Microbiologia de Brock**. 14a edição, Porto Alegre: Artmed, 2016.

MOHAMED, A. E. Comparative study of four supervised machine learning techniques for classification. **International Journal of Applied Science and Technology**, v. 7, n. 2, p. 5-18, 2017.

MONOD, J. The growth of bacterial cultures. **Annual Review of Microbiology**, v. 3, n. 1, p. 371-394, 1949.

NASTESKI, V. An overview of the suervised machine learing methods. **Horizons. b**, v. 4, p. 51-62, 2017.

OLIVEIRA, S. C.; OLIVEIRA, R. C.; TACIN, M. V.; GATTÁS, E. A. L. Kinetic modeling and optimization of a batch etanol fermentation process. **J. Bioprocess. Biotech**, v. 6, n. 266, p. 2-8, 2016.

OTTERSTEDT, K.; LARSSON, C.; BILL, R. M.; STÅHLBERG, A.; BOLES, E.; HOHMANN, S.; GUSTAFSSON, L. Switching the mode of metabolism in the yeast Saccharomyces cerevisiae. **EMBO reports**, v. 5, n. 5, p. 532-537, 2004.

PARAPOULI, M.; VASILEIADIS, A.; AFENDRA, AS.; HATZILOUKAS E. Saccharomyces cerevisiae and its industrial applications. **AIMS Microbiol**., v. 6, n. 1, p. 1-31, 2020.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, É. Scikit-learn: Machine learning in Python. **The Journal of Machine Learning Research**, v. 12, p. 2825-2830, 2011.

PHILLIPS, J.; CRIPPS, E.; LAU, J. W.; HODKIEWICZ, M. R. Classifying machinery condition using oil samples and binary logistic regression. **Mechanical Systems and Signal Processing**, v. 60, p. 316-325, 2015.

RAY, S. A quick review of machine learning algorithms. **2019 International conference on machine learning, big data, cloud and parallel computing (COMITCon).** IEEE, p. 35-39, 2019. Disponível em: https://doi.org/10.1109/COMITCon.2019.8862451

RIVERA, E. C.; YAMAKAWA, C. K.; GARCIA, M. H.; GERALDO, V. C.; ROSSELL C. E. V.; MACIEL FILHO, R.; BONOMI, A. A procedure for estimation of fermentation kinetic parameters in fed-batch bioetanol production process with cell recycle. **Chemical Eng**, v. 32, p. 1369-1374, 2013.

ROELL, G. W.; SATHISH, A.; WAN, N.; CHENG, Q.; WEN, Z.; TANG, Y. J.; BAO, F. S. A comparative evaluation of machine learning algorithms for predicting syngas fermentation outcomes. **Biochemical Engineering Journal**, v. 186, p. 108578-108587, 2022.

SAFARI, M.; PARVINNIA, E.; HADDAD, A. K. Industrial intrusion detection based on the behavior of rotating machine. **International Journal of Critical Infrastructure Protection**, v. 34, p. 100424, 2021.

SARKER, I. H. Machine Learning: Algorithms, real-world applications and research directions. **SN computer Science**, v. 2, n. 3, p. 160-180, 2021.

SEO, S.-O.; PARK, S.-K.; JUNG, S.-C.; RYU, C.-M.; KIM, J.-S. Anti-Contamination Strategies for Yeast Fermentations. **Microorganisms**, v. 8, n. 2, p. 274-286, 2020. https://doi.org/10.3390/microorganisms8020274

SHARMA, R.; GARG, P.; KUMAR, P.; BHATIA, S.K.; KULSHRESTHA, S. Microbial Fermentation and Its Role in Quality Improvement of Fermented Foods. **Fermentation**, v. 6, n. 4, p. 106-126, 2020. https://doi.org/10.3390/fermentation6040106

SILVA, A. P. M. D.; SICA, P.; PIRES, L. D. A. N.; SPIRONELLO, L.; MOTA, L. A.; PEIXOTO, G. T.; CALEGARI, R. B.; BASSO, T. O.; TONSO, A.; GOMES, M. P.; SOMESSARI, S. L.; DUARTE, H. G.; SOMESSARI, E. S. R.; CARVALHO, R. S.; BAPTISTA, A. S. Integration of Corn and Cane for Ethanol Production: Effects of Lactobacilli Contamination on Fermentative Parameters and Use of Ionizing Radiation Treatment for Disinfection. **Fermentation**, v. 9, n. 2, p. 89, 2023.

SONEGO, J. L. S.; LEMOS, D. A.; PINTO, C. E. M.; CRUZ, A. J. G.; BADINO, A. C. Extractive fed-batch etanol fermentation with CO2 stripping in a bubble column bioreactor: experimente and modeling. **Energy & Fuels**, v. 30, n. 1, p. 748-757, 2016.

STARZAK, M.; KRYZSTEK, L; NOWICKI, L.; MICHALSKI, H. Macroapproach kinetics of etanol fermentation by Saccharomyces cerevisiae: experimental studies and mathematical modelling. **The Chemical Engineering Journal and the Biochemical Engineering Journal**, v. 54, n. 3, p. 221-240, 1994.

TORRES, J. F.; HADJOUT, D.; SEBAA, A.; MARTÍNEZ-ÁLVAREZ, F.; TRONCOSO, A. Deep learning for time series forecasting: a survey. **Big Data**, v. 9, n. 1, p. 3-21, 2021.

TORTORA, G. J.; FUNKE, B. R.; CASE, C. L. **Microbiologia**. 12a edição, Porto Alegre: Artmed, 2017.

TSE, T. J.; WIENS, D. J.; REANEY, M. J. T. Production of Bioethanol—A Review of Factors Affecting Ethanol Yield. **Fermentation**, v. 7, n. 4, p. 268-286, 2021. https://doi.org/10.3390/fermentation7040268

VABALAS, A.; GOWEN, E.; POLIAKOFF, E.; CASSON, A. J. Machine learning algorithm validation with a limited sample size. **PloS one**, v. 14, n. 11, p. e0224365, 2019.

VARIZE, C. S.; BUCKER, A.; LOPES, L. D.; CHRISTOFOLETI-FURLAN, R. M.; RAPOSO, M. S.; BASSO, L. C.; STAMBUK, B. U. Increasing ethanol tolerance and ethanol production in an industrial fuel etanol Saccharomyces cerevisiae strain. **Fermentation**, v. 8, n. 10, p. 470, 2022. https://doi.org/10.3390/fermentation8100470

VIDAL, M. F. **Agroindústria - Etanol**. Caderno Setorial ETENE, Fortaleza: Banco do Nordeste do Brasil, n. 237, p. 1-16, 2022. Disponível em: https://www.bnb.gov.br/s482-dspace/handle/123456789/1409. Acesso em: 21 jun. 2023.

VOIDAROU, C.; ANTONIADOU, M.; ROZOS, G.; TZORA, A.; SKOUFOS, I.; VARZAKAS, T.; LAGIOU, A.; BEZIRTZOGLOU, E Fermentative foods: Microbiology, biochemistry, potential human health benefits and public health issues. **Foods**, v. 10, n. 1, p. 69, 2020.

XU, P. Analytical solution for a hybrid Logistic-Monod cell growth model in batch and continuous stirred tank reactor culture. **Biotechnology and Bioengineering**, v. 117, n. 3, p. 873-878, 2020.

XU, L. D.; XU, E. L.; LI, L. Industry 4.0: state of the art and future trends. **International journal of production research**, v. 56, n. 8, p. 2941-2962, 2018.

ZANARDI, M. S.; CARVALHO, R. V.; COSTA JUNIOR, E. F. Descrição matemática e simulação de modelos cinéticos na produção de etanol. **e-xacta**, v. 9, n. 2, p. 21-31, 2016.

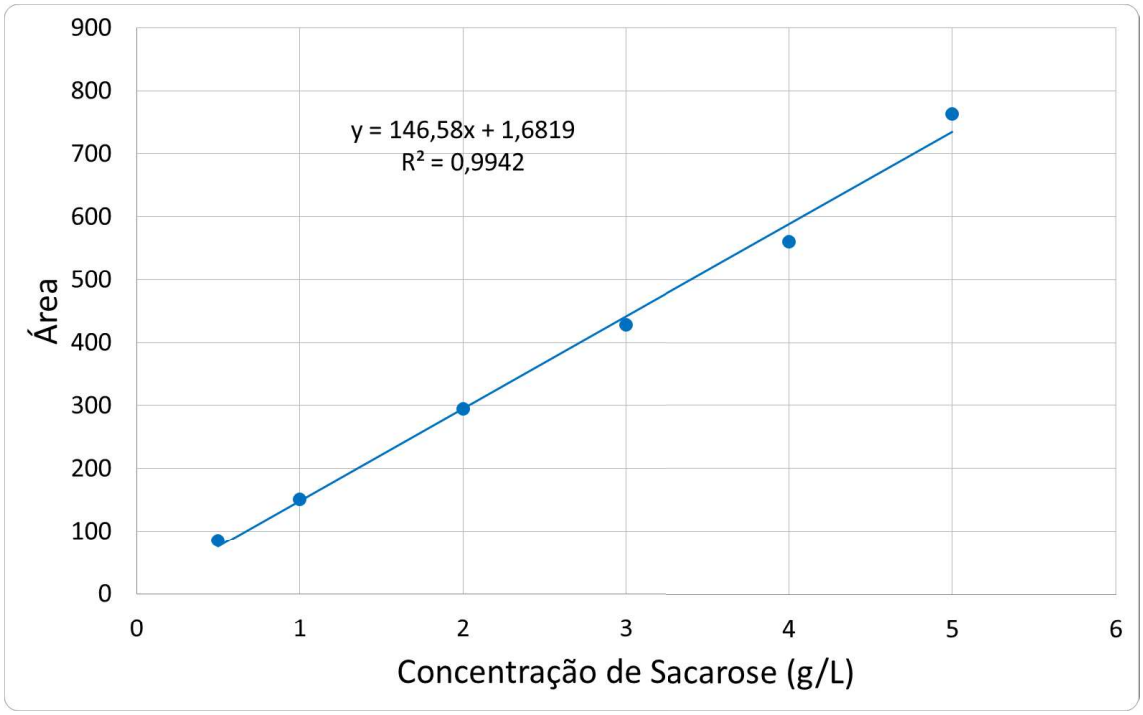
ZENTOU, H.; ZAINAL ABIDIN, Z.; YUNUS, R.; AWANG BIAK, D. R.; ABDULLAH ISSA, M.; YAHAYA PUDZA, M. A new model of alcoholic fermentation under a byproduct inhibitory effect. **ACS Omega**, v. 6, n. 6, p. 4137-4146, 2021.

ZENTOU, H.; ZAINAL ABIDIN, Z.; YUNUS, R.; AWANG BIAK, D. R.; ZOUANTI, M.; HASSANI, A. Modelling of molasses fermentation for bioethanol production: A comparative investigation of Monod and Andrews models accuracy assessment. **Biomolecules**, v. 9, n. 8, p. 308, 2019.

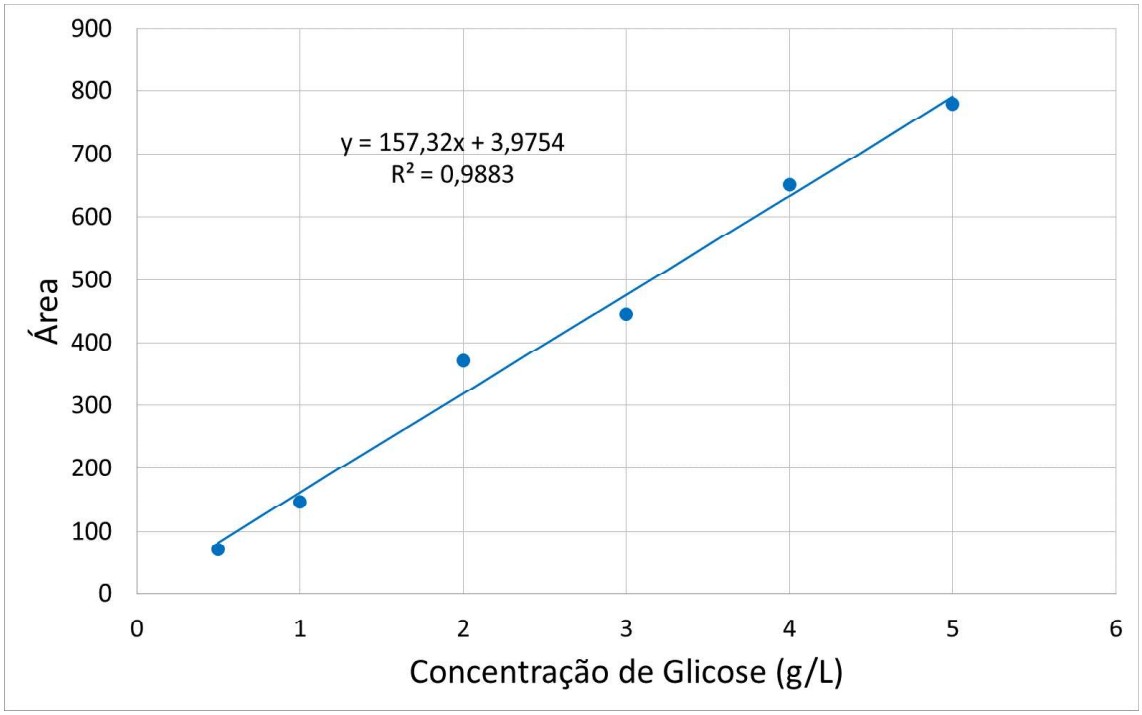
ZHANG, Y.; SU, M.; WANG, Z.; NIELSEN, J.; LIU, Z. Rewiring regulation on respiro-fermentative metabolism relieved crabtree effects in Saccharomyces cerevisiae. **Synthetic and Systems Biotechnology**, v. 7, n. 4, p. 1034-1043, 2022.

ZHU, X.; REHMAN, K.U.; WANG, B.; SHAHZAD, M. Modern Soft-Sensing Modeling Methods for Fermentation Processes. **Sensor**, v. 20, n. 6, p. 1771-1795, 2020. https://doi.org/10.3390/s20061771

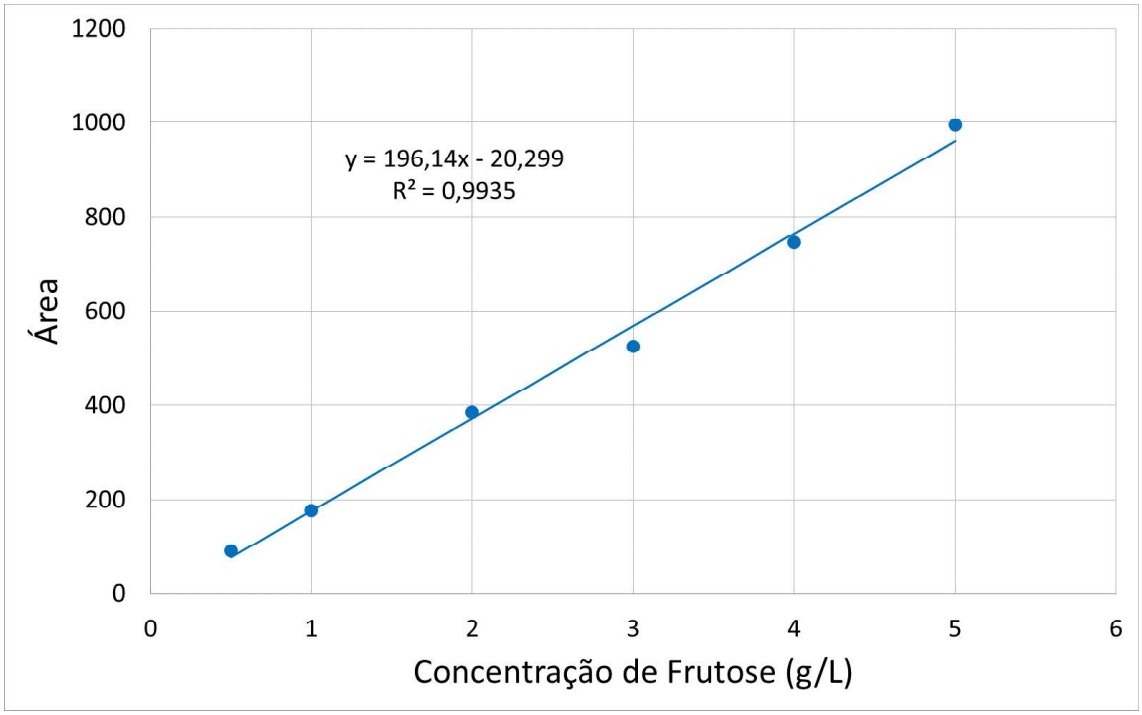
APÊNDICE A – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE SACAROSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA



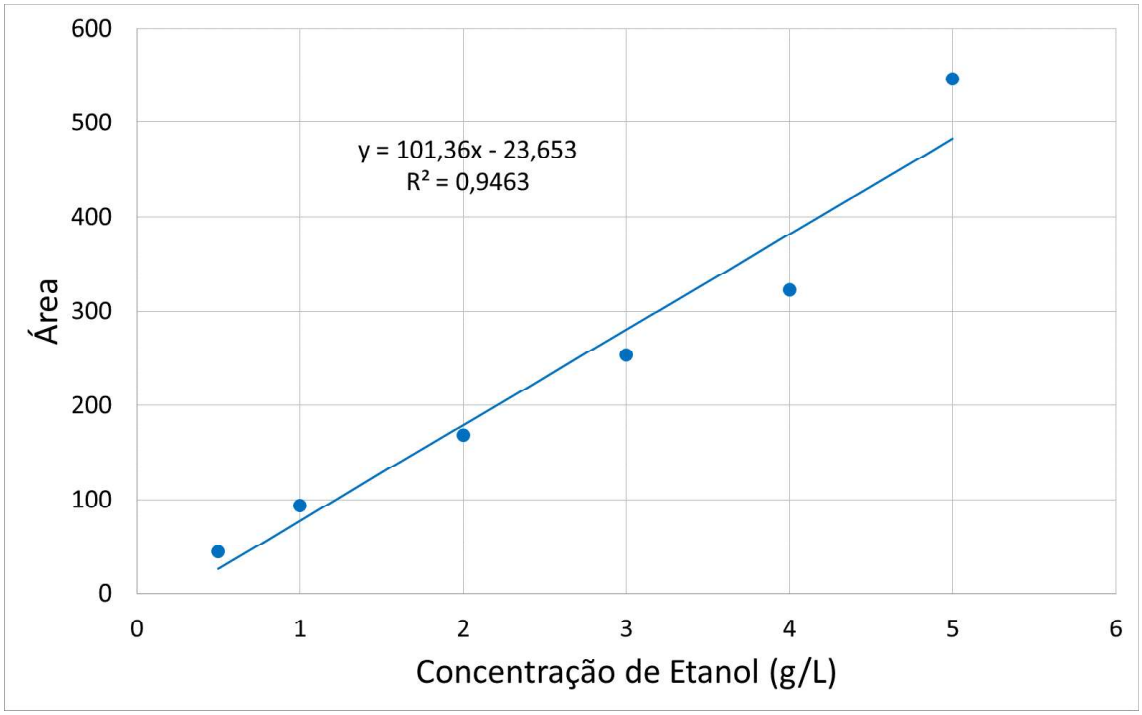
APÊNDICE B – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE GLICOSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA



APÊNDICE C – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE FRUTOSE EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA



APÊNDICE D – CURVA DE CALIBRAÇÃO DE SOLUÇÃO DE ETANOL EM CROMATÓGRAFO LÍQUIDO DE ALTA EFICIÊNCIA



APÊNDICE E – CÓDIGO EM FORTRAN 90 PARA SIMULAÇÃO DO PROCESSO FERMENTATIVO EM DIFERENTES CONDIÇÕES

```
program  reator

  integer :: i, j

  open(1, file='simul.xvg', status='new', iostat=ierr)

  do i=70, 120 ! S0

     do j=15, 35 ! X0

        call normal(i/1.0, 0/1.0, j/1.0)

     end do

  end do

  do i=70, 120 ! S0
     do j=15, 35 ! X0
        call contaminado(i/1.0, 0/1.0, j/1.0)

     end do

  end do

  close(1)

end program reator

subroutine normal(S0, P0, X0)

  real :: S0, X0, P0, S, P, X
```

```
real :: Yps, alfa, mimax, ks, ki, pmax, n
real :: gP, mi, pi, sigma
real :: t, dt
integer :: label

label = 0
Yps = 0.4
alfa = 4.87
mimax = 0.25
ks = 0.0061 ! g/L
ki = 139.7  ! g/L
pmax = 94.2 ! g/L
n = 4.12
dt = 0.2
t = 0.0
X = X0
S = S0
P = P0

do while(t < 8)

  t = t + dt
  gP = (1 - P/pmax)**n
  mi = mimax*S*gP/(ks + S + S*S/ki)
  pi = alfa*mi
  sigma = pi/Yps

  S = S - sigma*X*dt
  if (S<0) S = 0
  P = P + pi*X*dt
  X = X*exp(mi*dt)

  write(1,"(i1,7f8.3)") label, X0, S0, P0, t, X, S, P
```

```
  end do

end subroutine normal

subroutine contaminado(S0, P0, X0)

  real :: S0, X0, P0, S, P, X
  real :: Yps, alfa, mimax, ks, ki, pmax, n
  real :: gP, mi, pi, sigma, beta
  real :: t, dt
  integer :: label

  label = 1
  Yps = 0.4
  alfa = 4.87
  mimax = 0.25
  ks = 0.0061 ! g/L
  ki = 139.7  ! g/L
  pmax = 94.2 ! g/L
  n = 4.12
  dt = 0.2
  beta = 1.2
  t = 0.0
  X = X0
  S = S0
  P = P0

  do while(t < 8)

    t = t + dt
    gP = (1 - P/pmax)**n
    mi = mimax*S*gP/(ks + S + S*S/ki)
    pi = alfa*mi
    sigma = pi/Yps
```

```
    S = S - beta*sigma*X*dt
    if (S<0) S = 0
    P = P + pi*X*dt
    X = X*exp(mi*dt)

    write(1,"(i1,7f8.3)") label, X0, S0, P0, t, X, S, P

  end do

end subroutine contaminado
```

APÊNDICE F – CÓDIGO EM PYTHON PARA TREINO E VALIDAÇÃO DE MODELO DE APRENDIZADO DE MÁQUINA

```
import pandas as pd
import matplotlib.pyplot as plt

import sklearn.model_selection
import sklearn.metrics
import sklearn.ensemble

from sklearn.model_selection import train_test_split
from sklearn.neighbors import KNeighborsClassifier
from sklearn.neural_network import MLPClassifier

import sklearn.linear_model
import sklearn.tree
import sklearn.neural_network
import sklearn.neighbors

import optuna

new = open("/kaggle/input/dados-tcc-2/dados.xvg")

teste = new.read()
opa = []
opa = teste.split('], [')
aux = opa[0][2:]
opa[0] = aux

dados  = dict()
l_label, l_X0, l_S0, l_P0, l_t, l_X, l_S, l_P = [], [], [], [], [], [], [] ,[]

size = len(opa)/42
for i in range(int(size)):
```

```
    for j in range(42):

        aux = opa[(i*42)+j][1:62]
        aux2 = (aux.split('", '"))
        l_label.append(float(aux2[0]))
        l_X0.append(float(aux2[1]))
        l_S0.append(float(aux2[2]))
        l_P0.append(float(aux2[3]))
        l_t.append(float(aux2[4]))
        l_X.append(float(aux2[5]))
        l_S.append(float(aux2[6]))
        l_P.append(float(aux2[7]))

#=========================================
dados["label"] = l_label
dados["X0"] = l_X0
dados["S0"] = l_S0
dados["P0"] = l_P0
dados["t"] = l_t
dados["X"] = l_X
dados["S"] = l_S
dados["P"] = l_P

df = pd.DataFrame(data=dados)

X, y = df.iloc[:,1:].values, df.iloc[:,0].values
X_treino, X_valid, y_treino, y_valid = sklearn.model_selection.train_test_split(X, y,
                                                        test_size=0.2,
                                                        random_state=0)
count_0 = 0
count_1 = 0

for i in y_valid:
```

```
    if (i>0):
        count_1 = count_1 + 1
    else:
        count_0 = count_0 + 1

count_1, count_0

count_1 = 0
count_0 = 0

for i in y_treino:

    if (i>0):
        count_1 = count_1 + 1
    else:
        count_0 = count_0 + 1

count_1, count_0

m = sklearn.neural_network.MLPClassifier(random_state=0)

%time m.fit(X_treino, y_treino)

y_pred = m.predict(X_valid)

acc_tr = sklearn.metrics.accuracy_score(y_treino, m.predict(X_treino))
acc_val = sklearn.metrics.accuracy_score(y_valid, y_pred)

print(f'Acurácia do treino: {acc_tr}')
print(f'Acurácia da validação: {acc_val}')

confusao = sklearn.metrics.confusion_matrix(y_valid, y_pred)
confusao
```

```
disp = sklearn.metrics.ConfusionMatrixDisplay(confusion_matrix=confusao)
disp.plot()

plt.savefig("confusion-matrix.png")

resultado = sklearn.model_selection.cross_val_score(m, X_treino, y_treino, scoring='accuracy')
resultado
resultado.mean(), resultado.std()

# especificando modelos

modelos = [

    sklearn.neural_network.MLPClassifier(random_state=0),
    sklearn.linear_model.LogisticRegression(max_iter = 200),
    sklearn.tree.DecisionTreeClassifier(),
    sklearn.ensemble.RandomForestClassifier(),
    sklearn.ensemble.GradientBoostingClassifier(),
    sklearn.neighbors.KNeighborsClassifier(),

]

#lista para guardar resultados
results = [0]*len(modelos)

print('Modelo: média, desvio-padrão\n----------')

for i in range(len(modelos)):

    # efetuando a validação cruzada!
    results[i] = sklearn.model_selection.cross_val_score(modelos[i],
                                            X_treino, y_treino,
                                            cv=10,
```

```
                                        scoring='accuracy',
                                        n_jobs=-1)

    # imprimindo resultados
    print(f'{modelos[i].__class__.__name__}: {results[i].mean():.3}, {results[i].std():.3}')

# plotando resultados
fig, ax = plt.subplots()
ax.boxplot(results)

# formatando gráfico
ax.set_xticklabels([modelos[i].__class__.__name__ for i in range(len(modelos))],
                   rotation = 45, ha="right")
ax.set_ylabel("Acurácia")
plt.savefig("comp.png")

# função objetivo para otimização de hiperparâmetros
def objetivo(trial):

    # colocaremos dois modelos pra brigar: MLP e RF
    classifier_name = trial.suggest_categorical("classifier", ["MLP"])

    if classifier_name == 'MLP':

        # hiperparâmetros de busca para o MLP
        mlp_hidden_size = trial.suggest_int("mlp_hidden_size", 2,400)
        mlp_lr = trial.suggest_float("mlp_lr", 1e-3, 2e-1, log = True)
        mlp_batch_size = trial.suggest_int("mlp_batch_size", 8, 200)

        # modelo MLP
        m = sklearn.neural_network.MLPClassifier(hidden_layer_sizes =
mlp_hidden_size,
                                                 learning_rate_init = mlp_lr,
                                                 batch_size = mlp_batch_size,
```

```
                                    random_state=0)

    # retornando acurácia
    acuracias = sklearn.model_selection.cross_val_score(m, X_treino, y_treino,
n_jobs = -1, cv = 5)
    acuracia = acuracias.mean()
    return acuracia

study = optuna.create_study(direction="maximize")
study.optimize(objetivo, n_trials = 100)
study.best_params
```

APÊNDICE G – VALORES DEFAULT DOS HIPERPARÂMETROS DO MODELO DE REDES NEURAIS MLP DA BIBLIOTECA SCIKIT-LEARN

Hiperparâmetro	Valor default
Hidden layer sizes	100
Activation	‘relu’
Solver	‘adam’
Alpha	0,0001
Batch size	min(200, n_samples)
Learning rate	‘constant’
Learning rate init	0,001
Power t	0,5
Max iter	200
Shuffle	True
Random state	None
Tol	1 e-4
Verbose	False
Warm start	False
Momentum	0,9
Nesterovs momentum	True
Early stopping	False
Validation fraction	0,1
Beta 1	0,9
Beta 2	0,999
Epsilon	1 e-8
N iter no change	10
Max fun	15000