

Ana Paula Miranda Diniz

**Detecção de Eventos de Obstrução em Válvula  
no Processo de Lingotamento Contínuo com  
Uso de Aprendizado de Máquinas**

Brasil

2024



Ana Paula Miranda Diniz

## **Deteccção de Eventos de Obstrução em Válvula no Processo de Lingotamento Contínuo com Uso de Aprendizado de Máquinas**

Tese de doutorado apresentado ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para defesa da tese de doutorado em Engenharia Elétrica.

Universidade Federal do Espírito Santo – Ufes

Departamento de Engenharia Elétrica

Programa de Pós-Graduação

Orientador: Patrick Marques Ciarelli

Coorientador: Evandro Ottoni Teatini Salles

Brasil

2024



Ficha catalográfica disponibilizada pelo Sistema Integrado de  
Bibliotecas - SIBI/UFES e elaborada pelo autor

---

D585d      Diniz, Ana Paula Miranda, 1991-  
Detecção de Eventos de Obstrução em Válvula no Processo de  
Lingotamento Contínuo com Uso de Aprendizado de Máquinas  
/ Ana Paula Miranda Diniz. - 2024.  
157 f. : il.

Orientador: Patrick Marques Ciarelli.  
Coorientador: Evandro Ottoni Teatini Salles.  
Tese (Doutorado em Engenharia Elétrica) - Universidade  
Federal do Espírito Santo, Centro Tecnológico.

1. Siderurgia. 2. Aprendizado do computador. 3. Redes  
neurais (Computação). I. Ciarelli, Patrick Marques. II. Salles,  
Evandro Ottoni Teatini. III. Universidade Federal do Espírito  
Santo. Centro Tecnológico. IV. Título.

CDU: 621.3

---

Ana Paula Miranda Diniz

# Detecção de Eventos de Obstrução em Válvula no Processo de Lingotamento Contínuo com Uso de Aprendizado de Máquinas

Tese de doutorado apresentado ao Programa de Pós-Graduação em Engenharia Elétrica do Centro Tecnológico da Universidade Federal do Espírito Santo, como requisito parcial para defesa da tese de doutorado em Engenharia Elétrica.

Trabalho aprovado. Brasil, 25 de abril de 2024:



Documento assinado digitalmente  
**PATRICK MARQUES CIARELLI**  
Data: 10/07/2024 16:38:26-0300  
Verifique em <https://validar.iti.gov.br>

---

**Patrick Marques Ciarelli**  
Orientador  
Universidade Federal do Espírito Santo



Documento assinado digitalmente  
**EVANDRO OTTONI TEATINI SALLES**  
Data: 10/07/2024 19:28:37-0300  
Verifique em <https://validar.iti.gov.br>

---

**Evandro Ottoni Teatini Salles**  
Coorientador  
Universidade Federal do Espírito Santo



Documento assinado digitalmente  
**JOSE LEANDRO FELIX SALLES**  
Data: 11/07/2024 13:54:06-0300  
Verifique em <https://validar.iti.gov.br>

---

**José Leandro Félix Salles**  
Universidade Federal do Espírito Santo



Documento assinado digitalmente  
**JORGE LEONID ACHING SAMATELO**  
Data: 11/07/2024 14:00:27-0300  
Verifique em <https://validar.iti.gov.br>

---

**Jorge Leonid Aching Samatelo**  
Universidade Federal do Espírito Santo



Documento assinado digitalmente  
**DANIEL CRUZ CAVALIERI**  
Data: 11/07/2024 15:42:42-0300  
Verifique em <https://validar.iti.gov.br>

---

**Daniel Cruz Cavalieri**  
Instituto Federal do Espírito Santo



Documento assinado digitalmente  
**ALESSANDRO DO NASCIMENTO VARGAS**  
Data: 13/07/2024 13:32:00-0300  
Verifique em <https://validar.iti.gov.br>

---

**Alessandro do Nascimento Vargas**  
Universidade Tecnológica Federal do Paraná

Brasil  
2024

*Aos meus amados pais e aos meus orientadores, que tanto admiro, dedico o resultado do esforço realizado ao longo deste percurso.*



# Agradecimentos

Em primeiro lugar, agradeço à Deus por me dar forças, iluminar os meus caminhos e guiar os meus passos, me dando confiança frente aos desafios e adversidades. Sem Ele, nada disso teria sido possível.

Agradeço ao meu orientador, Professor Dr. Patrick Marques Ciarelli, por aceitar conduzir o meu trabalho, por toda atenção dedicada a mim e pelas valiosas contribuições. Ao meu coorientador, Professor Dr. Evandro Ottoni Teatini Salles, pelos ensinamentos e, principalmente, pela confiança. Ao Professor Dr. Klaus Fabian Côco, que participou ativamente no desenvolvimento desse trabalho com sugestões e compartilhamento de ideias, mantendo-me motivada durante todo o processo. Obrigada por acreditarem em mim, mesmo quando eu achei que não fosse conseguir!

Agradeço aos meus amados pais, Marco Antônio Rodrigues Diniz e Maria Elvira Miranda, por serem os meus maiores incentivadores e me ensinarem, desde cedo, o valor da educação. Este trabalho é a prova de que seus esforços ao longo de toda essa jornada valeram a pena.

Gratidão aos meus amigos e familiares por todo apoio nesta caminhada. Ao meu namorado, Giovanni Torres Corso, que me encorajou nos últimos anos, sou grata por suportar as minhas incontáveis crises de estresse e compreender a minha ausência em diversos momentos.

Pelo suporte e concessão da bolsa de estudos, agradeço à CAPES e ao PPGEE. Por fim, agradeço à empresa siderúrgica que forneceu os dados, porém, por motivo de confidencialidade de dados não pode ser citada ou mesmo ter seus dados expostos nesta tese.



*“Não busco compreender para crer, mas creio para compreender.”*  
*(Santo Anselmo)*



# Resumo

O processo de lingotamento contínuo, usado na fabricação de placas de aço, é atualmente a maneira mais econômica e eficaz de produção dentro da indústria. Embora o lingotamento contínuo seja um processo amplamente usado, alguns problemas associados ao processo ainda não foram resolvidos, sendo um deles a obstrução da válvula submersa (*Submerged Entry Nozzle* - SEN), que controla o fluxo de aço entre o distribuidor e o molde. Essa obstrução, também chamada de *clogging*, além de comprometer a qualidade do produto, resulta em menor rendimento do processo, acarretando prejuízos. Deste modo, a detecção do *clogging* é de fundamental importância, pois ações de controle podem permitir que o sistema opere por mais tempo. Neste trabalho, são apresentadas e comparadas três metodologias baseadas em aprendizagem de máquina e *Deep Learning* com o objetivo de detectar as ocorrências de *clogging* a partir de dados históricos de variáveis do processo. Em geral, os classificadores apresentaram resultados bastante promissores em dados reais com desequilíbrio de classes. Em especial, o método empregando análise espaço-temporal, utilizando quatro variáveis do processo, obteve um desempenho notavelmente superior em relação aos demais, alcançando médias de Precisão e *Recall*, respectivamente, de 95,53% e 97,33%. Visando reduzir as taxas de falsos positivos e negativos, foi implementada uma heurística de pós-processamento aplicada à saída do modelo, sendo alcançado uma Precisão e um *Recall*, respectivamente, de aproximadamente 98% e 99%. Até o conhecimento da autora, esses resultados nunca foram encontrados na literatura. Apesar de uma comparação detalhada seja inviabilizada devido às diferenças entre as bases de dados e a inacessibilidade a elas, a modelagem aqui proposta atingiu patamares de desempenho superiores quando comparados aos resultados encontrados na literatura para solucionar este tipo de problema da indústria. Os resultados elevados e inéditos obtidos neste trabalho, portanto, contribuirão tanto para a melhoria da qualidade do produto final, como para a redução de custos associados à produção do aço, uma vez que a classificação mais efetiva das ocorrências de *clogging* poderá auxiliar os operadores no planejamento de ações corretivas.

**Palavras-chaves:** Lingotamento Contínuo. Bocal da Válvula Submersa. *Clogging*. Aprendizado de Máquinas. *Deep Learning*.



# Abstract

The continuous casting process, used in the manufacture of steel plates, is currently the most economical and efficient way of production within the industry. Although continuous casting is a widely used process, some problems associated with the process have not yet been resolved, one of them being the obstruction of the Submerged Entry Nozzle (SEN), which controls the flow of steel between the tundish and the mold. This obstruction, also called clogging, not only impairs the quality of the product but also results in lower process yield, resulting in losses. Thus, clogging detection is of fundamental importance, because control actions can allow the system to operate for a longer time. In this work, methodologies based on Machine Learning and Deep Learning are presented and compared to detect the occurrences of clogging from historical data of process variables. In general, the performance of the classifiers achieved very promising results in real data with unbalanced classes. In particular, the method employing spatiotemporal analysis, using four process variables, obtained a remarkably superior performance when compared to the others, reaching averages of Precision and Recall, respectively, of 95.53% and 97.33%. To reduce the false positive and negative rates, a post-processing heuristic was implemented and applied to the model output, achieving a Precision and a Recall, respectively, of approximately 98% and 99%. To the best of our knowledge, these results have never been found in the literature. Although a detailed comparison is unfeasible due to the differences between the datasets and their inaccessibility, the modeling proposed here reached higher performance levels when compared to the results found in the literature to solve this industry's problem. The high and unprecedented results obtained in this work, therefore, will contribute both to the improvement of the quality of the final product and to the reduction of costs associated with steel production, since the more effective classification of clogging occurrences can help operators in the corrective action planning.

**Key-words:** Continuous casting. Submerged Nozzle Valve. Clogging. Machine Learning. Deep Learning.



# Lista de ilustrações

|                                                                                                                                                                                                                                                                                                                                                                                          |    |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Figura 1 – Resumo esquemático de um processo de lingotamento contínuo. . . . .                                                                                                                                                                                                                                                                                                           | 40 |
| Figura 2 – Resumo esquemático da passagem do aço do distribuidor para o molde com controle do fluxo realizado através da válvula gaveta. . . . .                                                                                                                                                                                                                                         | 41 |
| Figura 3 – Ângulo da porta do bocal da SEN. . . . .                                                                                                                                                                                                                                                                                                                                      | 48 |
| Figura 4 – Amostra de <i>clogging</i> no bocal do distribuidor acima da válvula gaveta. . . . .                                                                                                                                                                                                                                                                                          | 49 |
| Figura 5 – Exemplo do comportamento da velocidade do veio e da abertura da válvula gaveta em um processo (a) sem e (b) com a ocorrência de <i>clogging</i> . . . . .                                                                                                                                                                                                                     | 50 |
| Figura 6 – Flutuações no nível do molde (a) com <i>clogging</i> e (b) sem <i>clogging</i> . . . . .                                                                                                                                                                                                                                                                                      | 51 |
| Figura 7 – Efeitos do <i>clogging</i> inicial observados pelo padrão de fluxo de quatro bocais semelhantes, diferenciados pela geometria próxima à válvula gaveta, em que o arredondamento dos cantos da cerâmica é representada pelo desenho avermelhado, enquanto que as linhas de fluxo são representadas pelas linhas pretas. . . . .                                                | 52 |
| Figura 8 – Relação entre espessura do <i>clogging</i> e a posição da válvula deslizante. . . . .                                                                                                                                                                                                                                                                                         | 55 |
| Figura 9 – Comparativo dos resultados dos testes desenvolvidos em (IKÄHEIMONEN et al., 2002). . . . .                                                                                                                                                                                                                                                                                    | 56 |
| Figura 10 – Exemplo dos resultados obtidos no teste da RNA com dados de casos bem-sucedidos (IKÄHEIMONEN et al., 2002). . . . .                                                                                                                                                                                                                                                          | 57 |
| Figura 11 – Variação da abertura da válvula em relação à velocidade do veio e peso do aço no distribuidor. . . . .                                                                                                                                                                                                                                                                       | 58 |
| Figura 12 – Aplicação do índice de <i>clogging</i> aos dados da planta. . . . .                                                                                                                                                                                                                                                                                                          | 59 |
| Figura 13 – Relação entre a taxa de fluxo, velocidade do veio ( $Vel$ ), nível do molde e diâmetro do bocal ( $D$ ). . . . .                                                                                                                                                                                                                                                             | 60 |
| Figura 14 – Relação entre a abertura da válvula, velocidade do veio, diâmetro do bocal ( $D$ ) e nível do molde ( $Niv$ ). . . . .                                                                                                                                                                                                                                                       | 61 |
| Figura 15 – Evolução do <i>clogging</i> no bocal. (a) Vista da região de entupimento da seção vertical e uma seção transversal A-A do bico; (b) Vista ampliada das linhas de fluxo e uma seção parcialmente obstruída a 200 segundos. A seção entupida é mostrada em amarelo. A magnitude da velocidade do fluxo é mostrada em uma escala de cores ao longo das linhas de fluxo. . . . . | 62 |
| Figura 16 – (a) A massa total do entupimento calculada numericamente e a taxa de deposição de massa; (b) Três períodos de evolução do entupimento: (1) <i>Covering</i> , (2) <i>Bulging</i> e (3) <i>Branching</i> . . . . .                                                                                                                                                             | 63 |
| Figura 17 – Exemplo de uma rede MLP com 2 camadas ocultas contendo, respectivamente, 4 e 3 neurônios. . . . .                                                                                                                                                                                                                                                                            | 71 |
| Figura 18 – Fluxo de processamento do Algoritmo de <i>BackPropagation</i> . . . . .                                                                                                                                                                                                                                                                                                      | 72 |

|                                                                                                                                                                                                                                                                                                                                                      |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 19 – Operação de convolução com $\text{kernel } 3 \times 3$ e $\text{stride}$ de 1. . . . .                                                                                                                                                                                                                                                   | 73  |
| Figura 20 – Forma das funções de ativação: (a) ReLU, (b) Tangente Hiperbólica e (c) Sigmoides . . . . .                                                                                                                                                                                                                                              | 74  |
| Figura 21 – Exemplo de <i>MaxPooling</i> com filtro $2 \times 2$ e $\text{stride}$ igual a 2. . . . .                                                                                                                                                                                                                                                | 74  |
| Figura 22 – Exemplo de composição de Camada FC: <i>Flatten</i> , oculta e saída . . . . .                                                                                                                                                                                                                                                            | 75  |
| Figura 23 – Exemplo de (a) conectividade esparsa da rede CNN e (b) conectividade global da rede MLP. . . . .                                                                                                                                                                                                                                         | 77  |
| Figura 24 – Campos receptivos locais de uma rede CNN. . . . .                                                                                                                                                                                                                                                                                        | 77  |
| Figura 25 – Exemplo de CNN com três mapas de características. . . . .                                                                                                                                                                                                                                                                                | 78  |
| Figura 26 – A estrutura de uma célula LSTM. . . . .                                                                                                                                                                                                                                                                                                  | 79  |
| Figura 27 – Estrutura interna da ConvLSTM . . . . .                                                                                                                                                                                                                                                                                                  | 81  |
| Figura 28 – Estrutura genérica de uma rede ConvLSTM. . . . .                                                                                                                                                                                                                                                                                         | 82  |
| Figura 29 – Matriz de confusão. . . . .                                                                                                                                                                                                                                                                                                              | 83  |
| Figura 30 – Exemplo típico do comportamento dos sinais do processo com base nos dados brutos disponibilizados pela indústria. . . . .                                                                                                                                                                                                                | 88  |
| Figura 31 – Processo de interpolação do período de troca de distribuidor: (a) Identificação do período de troca de distribuidor; (b) Região de interpolação (aproximação); (c) Interpolação e agrupamento das variáveis de abertura da válvula (aproximação). . . . .                                                                                | 92  |
| Figura 32 – Gráficos <i>boxplot</i> das variáveis $x_1(n)$ , $x_2(n)$ e $x_3(n)$ correspondentes à abertura da válvula, nível do molde e velocidade do veio, respectivamente. . . . .                                                                                                                                                                | 93  |
| Figura 33 – Exemplo didático de ocorrência inicial do <i>clogging</i> observada a partir das variáveis do processo selecionadas após o tratamento. . . . .                                                                                                                                                                                           | 94  |
| Figura 34 – Exemplo didático do início do <i>clogging</i> sem que houvesse qualquer percepção de alteração no nível do molde. . . . .                                                                                                                                                                                                                | 95  |
| Figura 35 – Exemplo didático do início de duas ocorrências de <i>clogging</i> entre a ocorrência do <i>unclogging</i> . . . . .                                                                                                                                                                                                                      | 96  |
| Figura 36 – Exemplo didático do início do <i>clogging</i> após constantes oscilações das variáveis de abertura da válvula e nível do molde. . . . .                                                                                                                                                                                                  | 97  |
| Figura 37 – Relação entre entradas deslocadas no tempo e a saída. . . . .                                                                                                                                                                                                                                                                            | 98  |
| Figura 38 – Histograma das classes do conjunto de dados. . . . .                                                                                                                                                                                                                                                                                     | 99  |
| Figura 39 – (a) Histograma da variável $x_1(n)$ referente à abertura da válvula (existem <i>outliers</i> imperceptíveis nas faixas de 0 a 40 e 85 a 100); (b) Destaque para a distribuição da variável $x_1(n)$ desprezando os <i>outliers</i> conceituais, acompanhada por uma representação gaussiana baseada na média e no desvio padrão. . . . . | 102 |

|                                                                                                                                                                                                                                                                                                                                        |     |
|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Figura 40 – (a) Histograma da variável $x_2(n)$ referente ao nível do molde, sendo possível observar pontos de <i>outliers</i> conceituais; (b) Destaque para a distribuição da variável $x_2(n)$ desprezando os <i>outliers</i> conceituais, acompanhada por uma representação gaussiana baseada na média e no desvio padrão. . . . . | 102 |
| Figura 41 – Distribuição de Velocidade do Veio ( $x_3(n)$ ) - Análise de Histograma e Representação de Gaussianas. . . . .                                                                                                                                                                                                             | 104 |
| Figura 42 – Histograma das variáveis após a aplicação da padronização por <i>z-score</i> : (a) Abertura da válvula; (b) Nível do molde; (c) Velocidade do veio. . . . .                                                                                                                                                                | 105 |
| Figura 43 – Estrutura genérica dos classificadores LSTM usados nesta pesquisa. . . . .                                                                                                                                                                                                                                                 | 107 |
| Figura 44 – Estrutura genérica dos classificadores CNN-LSTM. . . . .                                                                                                                                                                                                                                                                   | 108 |
| Figura 45 – Estrutura genérica dos classificadores ConvLSTM. . . . .                                                                                                                                                                                                                                                                   | 109 |
| Figura 46 – Transformações do conjunto de entrada. . . . .                                                                                                                                                                                                                                                                             | 109 |
| Figura 47 – (a) Média e (b) Desvio Padrão dos oito classificadores LSTM que obtiveram os melhores desempenhos. . . . .                                                                                                                                                                                                                 | 115 |
| Figura 48 – (a) Média e (b) Desvio Padrão dos oito classificadores CNN-LSTM que obtiveram os melhores desempenhos. . . . .                                                                                                                                                                                                             | 115 |
| Figura 49 – (a) Média e (b) Desvio Padrão dos oito classificadores ConvLSTM que obtiveram os melhores desempenhos. . . . .                                                                                                                                                                                                             | 116 |
| Figura 50 – Heurística de pós-processamento aplicada ao modelo ConvLSTM 7. . . . .                                                                                                                                                                                                                                                     | 124 |
| Figura 51 – Curvas de Treinamento e Validação dos modelos LSTM 1, LSTM 2, LSTM 3 e LSTM 4. . . . .                                                                                                                                                                                                                                     | 141 |
| Figura 52 – Curvas de Treinamento e Validação dos modelos LSTM 5, LSTM 6, LSTM 7 e LSTM 8. . . . .                                                                                                                                                                                                                                     | 142 |
| Figura 53 – Curvas de Treinamento e Validação dos modelos CNN-LSTM 1, CNN-LSTM 2, CNN-LSTM 3 e CNN-LSTM 4. . . . .                                                                                                                                                                                                                     | 143 |
| Figura 54 – Curvas de Treinamento e Validação dos modelos CNN-LSTM 5, CNN-LSTM 6, CNN-LSTM 7 e CNN-LSTM 8. . . . .                                                                                                                                                                                                                     | 144 |
| Figura 55 – Curvas de Treinamento e Validação dos modelos ConvLSTM 1, ConvLSTM 2, ConvLSTM 3 e ConvLSTM 4. . . . .                                                                                                                                                                                                                     | 145 |
| Figura 56 – Curvas de Treinamento e Validação dos modelos ConvLSTM 5, ConvLSTM 6, ConvLSTM 7 e ConvLSTM 8. . . . .                                                                                                                                                                                                                     | 146 |
| Figura 57 – Curvas de Treinamento e Validação dos modelos MLP com 16, 32, 64 e 128 neurônios na camada oculta, respectivamente. . . . .                                                                                                                                                                                                | 147 |
| Figura 58 – Curvas de Treinamento e Validação dos modelos MLP com 256, 512, 1024 e 2048 neurônios na camada oculta, respectivamente. . . . .                                                                                                                                                                                           | 148 |
| Figura 59 – Curvas de Treinamento e Validação dos modelos MLP com 3072, 4096 e 5120 neurônios na camada oculta, respectivamente. . . . .                                                                                                                                                                                               | 149 |



# Lista de tabelas

|           |                                                                                                                                                                                                 |     |
|-----------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabela 1  | – Literaturas de referência: Resumo . . . . .                                                                                                                                                   | 67  |
| Tabela 2  | – Variáveis do processo pré-selecionadas. . . . .                                                                                                                                               | 88  |
| Tabela 3  | – Seleção final das variáveis de entrada após o tratamento dos dados. . .                                                                                                                       | 91  |
| Tabela 4  | – Divisão do conjunto de dados e quantidade de amostras por classe. . .                                                                                                                         | 101 |
| Tabela 5  | – Estatística descritivas das variáveis $x_1(n)$ , $x_2(n)$ e $x_3(n)$ . . . . .                                                                                                                | 103 |
| Tabela 6  | – Hiperparâmetros usados para treinar os modelos. . . . .                                                                                                                                       | 106 |
| Tabela 7  | – Seleção dos oito modelos LSTM que obtiveram os melhores desempenhos na fase de treinamento e validação. . . . .                                                                               | 111 |
| Tabela 8  | – Seleção dos oito modelos CNN-LSTM que obtiveram os melhores desempenhos na fase de treinamento e validação. . . . .                                                                           | 111 |
| Tabela 9  | – Seleção dos oito modelos ConvLSTM que obtiveram os melhores desempenhos na fase de treinamento e validação. . . . .                                                                           | 111 |
| Tabela 10 | – Comparação entre os oito classificadores LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos. . . . .                                   | 113 |
| Tabela 11 | – Comparação entre os oito classificadores CNN-LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos. . . . .                               | 114 |
| Tabela 12 | – Comparação entre os oito classificadores ConvLSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos. . . . .                               | 114 |
| Tabela 13 | – Resultado do teste de hipóteses para comparação dos valores de <i>Recall</i> dos modelos ConvLSTM 1, ConvLSTM 2 e ConvLSTM 6, comparados ao modelo ConvLSTM 7. . . . .                        | 118 |
| Tabela 14 | – Tamanho dos arquivos de cada modelo em [kB]. . . . .                                                                                                                                          | 118 |
| Tabela 15 | – Tempo médio de operação dos classificadores cujos modelos obtiveram os melhores desempenhos. . . . .                                                                                          | 119 |
| Tabela 16 | – Comparação entre os desempenhos dos classificadores MLP contendo uma única camada oculta, seguida da quantidade de parâmetros e do tempo médio para o treinamento completo do modelo. . . . . | 121 |
| Tabela 17 | – Comparação entre os desempenhos dos classificadores MLP enviesados, contendo uma única camada oculta. . . . .                                                                                 | 121 |
| Tabela 18 | – Comparação entre os critérios de desempenho antes e depois da aplicação da heurística de pós-processamento aplicada ao modelo ConvLSTM. . .                                                   | 123 |
| Tabela 19 | – Comparativo entre as métricas de desempenho do modelo ConvLSTM 7 com aquelas obtidas pelos modelos encontrados na literatura. . . . .                                                         | 125 |

|                                                                                                                                                                                                                        |     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Tabela 20 – Modelos MLP com duas camadas ocultas variando o número de neurônios de 8 a 256 cada. . . . .                                                                                                               | 151 |
| Tabela 21 – Desempenhos dos classificadores MLP com duas camadas ocultas seguido do tempo de treinamento. . . . .                                                                                                      | 151 |
| Tabela 22 – Os oito modelos LSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída. . . . .                                                                             | 153 |
| Tabela 23 – Os oito modelos CNN-LSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída. . . . .                                                                         | 153 |
| Tabela 24 – Os oito modelos ConvLSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída. . . . .                                                                         | 154 |
| Tabela 25 – Os oito classificadores LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída. .   | 154 |
| Tabela 26 – Os oito classificadores CNN-LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída. | 155 |
| Tabela 27 – Os oito classificadores ConvLSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída. | 155 |

# Lista de abreviaturas e siglas

|          |                                                   |
|----------|---------------------------------------------------|
| AUC      | <i>Area Under the Curve</i>                       |
| BPTT     | <i>Backpropagation Through Time</i>               |
| ConvLSTM | <i>Convolutional Long Short-Term Memory</i>       |
| CNN      | <i>Convolutional Neural Network</i>               |
| FC       | <i>Fully Connected</i>                            |
| FIS      | <i>Fuzzy Inference System</i>                     |
| FN       | <i>False Negative</i>                             |
| FP       | <i>False Positive</i>                             |
| IA       | Inteligência Artificial                           |
| LASCUS   | <i>LAbeled Som Classification Unbalanced Sets</i> |
| LSTM     | <i>Long Short Term Memory</i>                     |
| MAE      | <i>Mean Absolute Error</i>                        |
| MCC      | <i>Matthews Correlation Coefficient</i>           |
| MLP      | <i>MultiLayer Perceptrons</i>                     |
| ReLU     | <i>Rectified Linear Unit</i>                      |
| RNA      | Rede Neural Artificial                            |
| RNN      | <i>Recurrent Neural Network</i>                   |
| SEN      | <i>Submerged Entry Nozzle</i>                     |
| SH       | <i>Shroud Holder Gap</i>                          |
| SOM      | <i>Self-Organized Map</i>                         |
| SS       | <i>Lower Submerged Entry Nozzle</i>               |
| TN       | <i>True Negative</i>                              |
| TP       | <i>True Positive</i>                              |
| UN       | <i>Upper Tundish Well Nozzle</i>                  |
| UP       | <i>Upper Plate Gap</i>                            |



# Lista de símbolos

|                      |                                                                                                                    |
|----------------------|--------------------------------------------------------------------------------------------------------------------|
| $D$                  | Diâmetro do bocal da válvula                                                                                       |
| $Vel$                | Velocidade do aço no veio                                                                                          |
| $Niv$                | Nível do Molde                                                                                                     |
| $N_T$                | Número total de amostras                                                                                           |
| $X$                  | Conjunto de dados de entrada                                                                                       |
| $g(X)$               | Função que mapeia o espaço de entrada $X$ para o espaço de saída $Y$                                               |
| $Y$                  | Conjunto de dados da saída desejada                                                                                |
| $y(t)$               | Saída desejada para o tempo $t$                                                                                    |
| $\hat{y}(t)$         | Saída computada pela rede para o tempo $t$                                                                         |
| $e(t)$               | Erro para o tempo $t$                                                                                              |
| $x_{in}(t)$          | Padrão de Entrada para o tempo $t$                                                                                 |
| $z_{i,j,m}^l$        | $m$ -ésimo mapa de características da $l$ -ésima camada com valor da característica localizada na posição $(i, j)$ |
| $\mathbf{x}_{i,j}^l$ | Entrada centrada na posição $(i, j)$ referente à $l$ -ésima camada                                                 |
| $\mathbf{w}_m^l$     | $m$ -ésimo filtro da $l$ -ésima camada                                                                             |
| $b_m^l$              | <i>Bias</i> do $m$ -ésimo filtro da $l$ -ésima camada                                                              |
| $ReLU(x)$            | Função de ativação ReLU                                                                                            |
| $\tanh(\cdot)$       | Função de Ativação Tangente Hiperbólica                                                                            |
| $\varphi(\cdot)$     | Função de Ativação Sigmoide                                                                                        |
| $Softmax(y_c)$       | Função <i>Softmax</i> calculada sobre uma saída $y_z$                                                              |
| $N_c$                | Número de Classes                                                                                                  |
| $t$                  | Tempo                                                                                                              |
| $f(t)$               | Sinal do <i>Forget gate</i> de uma célula LSTM                                                                     |
| $i(t)$               | Sinal do <i>Input gate</i> de uma célula LSTM                                                                      |

|                   |                                                        |
|-------------------|--------------------------------------------------------|
| $o(t)$            | Sinal do <i>Output gate</i> de uma célula LSTM         |
| $x(t)$            | Sinal de Entrada                                       |
| $C(t)$            | Sinal do estado de memória atual de uma célula LSTM    |
| $h(t)$            | Sinal do estado oculto atual de uma célula LSTM        |
| $\mathbf{W}_{fx}$ | Pesos de entrada em relação ao $f(t)$                  |
| $\mathbf{W}_{fh}$ | Pesos recorrentes em relação ao $f(t)$                 |
| $\mathbf{b}_f$    | <i>Bias</i> do <i>forget gate</i>                      |
| $\mathbf{W}_{ix}$ | Pesos de entrada em relação ao $i(t)$                  |
| $\mathbf{W}_{ih}$ | Pesos recorrentes em relação ao $i(t)$                 |
| $\mathbf{b}_i$    | <i>Bias</i> do <i>input gate</i>                       |
| $\tilde{C}(t)$    | Estado Candidato                                       |
| $\mathbf{W}_{cx}$ | Pesos de entrada em relação ao $\tilde{C}(t)$          |
| $\mathbf{W}_{ch}$ | Pesos recorrentes em relação ao $\tilde{C}(t)$         |
| $\mathbf{b}_c$    | <i>Bias</i> do Estado Candidato                        |
| $\mathbf{W}_{ox}$ | Pesos de entrada em relação ao $o(t)$                  |
| $\mathbf{W}_{oh}$ | Pesos recorrentes em relação ao $o(t)$                 |
| $\mathbf{b}_o$    | <i>Bias</i> do <i>output gate</i>                      |
| $k$               | Quantidade de subconjuntos empregados no <i>k-fold</i> |
| $A_{cc}$          | Acurácia                                               |
| $P_{rc}$          | Precisão                                               |
| $R_{cl}$          | <i>Recall</i>                                          |
| $F_1$             | F1-Score                                               |
| $DI$              | Distribuidor em Operação                               |
| $DI5$             | Distribuidor de número 5                               |
| $DI6$             | Distribuidor de número 6                               |
| $V5$              | Veio de número 5                                       |

|                |                                                                                                                              |
|----------------|------------------------------------------------------------------------------------------------------------------------------|
| $TpDI$         | Temperatura do aço no distribuidor amostrada manualmente                                                                     |
| $TpDI5$        | Temperatura no $DI5$                                                                                                         |
| $TpDI6$        | Temperatura no $DI6$                                                                                                         |
| $AbDI5$        | Percentual da abertura total da válvula gaveta do $DI5$                                                                      |
| $AbDI6$        | Percentual da abertura total da válvula gaveta do $DI6$                                                                      |
| $AbDI$         | Percentual da abertura total da válvula gaveta                                                                               |
| $r$            | Índice identificador de cada variável de entrada ( $r \in \{1, 2, 3, 4\}$ )                                                  |
| $n$            | Instante de tempo discretizado                                                                                               |
| $T_s$          | Período de amostragem                                                                                                        |
| $x_r(n)$       | $r$ -ésima variável de entrada no instante $n$                                                                               |
| $\mu_r$        | Média da variável $x_r(n)$                                                                                                   |
| $\sigma_r$     | Desvio Padrão da variável $x_r(n)$                                                                                           |
| $x_{rz}(n)$    | $r$ -ésima variável de entrada normalizada pelo $z$ -score                                                                   |
| $N$            | Número total de amostras (sequências) contidas em um <i>batch</i>                                                            |
| $n_{frames}$   | Número de subsequências de uma janela temporal                                                                               |
| $n_{row}$      | Dimensão dos dados                                                                                                           |
| $n_{col}$      | Número de intervalos de tempo de uma subsequência de entrada                                                                 |
| $n_{channels}$ | Número de atributos de entrada                                                                                               |
| $ClgMed$       | Quantidade média de amostras rotuladas como <i>clogging</i> dentro da janela de avaliação da heurística de pós-processamento |
| $LimMin$       | Limiar Mínimo aplicado na heurística de pós-processamento                                                                    |
| $LimMax$       | Limiar Máximo aplicado na heurística de pós-processamento                                                                    |
| $J_1$          | Janela aplicada na heurística de pós-processamento entre as amostras do instante $n$ até $n + 44$                            |
| $J_2$          | Janela aplicada na heurística de pós-processamento entre as amostras do instante $n - 45$ até $n - 1$                        |
| $n_{j1}$       | Quantidade de amostras classificadas como <i>clogging</i> de maneira consecutiva a partir do início da janela $J1$ .         |

$n_{j2}$  Quantidade de amostras classificadas como *clogging* de maneira consecutiva a partir do final da janela  $J2$ .

$y_{prev}(n)$  Saída da classificação da rede neural no temp  $n$

# Sumário

|            |                                                                            |           |
|------------|----------------------------------------------------------------------------|-----------|
| <b>1</b>   | <b>INTRODUÇÃO</b>                                                          | <b>29</b> |
| <b>1.1</b> | <b>Caracterização do Problema</b>                                          | <b>30</b> |
| <b>1.2</b> | <b>Motivação</b>                                                           | <b>31</b> |
| <b>1.3</b> | <b>Hipótese</b>                                                            | <b>32</b> |
| <b>1.4</b> | <b>Objetivos</b>                                                           | <b>33</b> |
| <b>1.5</b> | <b>Método Proposto</b>                                                     | <b>34</b> |
| <b>1.6</b> | <b>Produção Bibliográfica</b>                                              | <b>36</b> |
| <b>1.7</b> | <b>Estrutura da Tese</b>                                                   | <b>36</b> |
| <b>2</b>   | <b>O CLOGGING DA VÁLVULA SUBMERSA NO PROCESSO DE LINGOTAMENTO CONTÍNUO</b> | <b>39</b> |
| <b>2.1</b> | <b>O Processo de Lingotamento Contínuo</b>                                 | <b>39</b> |
| 2.1.1      | A Válvula Submersa                                                         | 40        |
| <b>2.2</b> | <b>O Clogging da Válvula Submersa</b>                                      | <b>42</b> |
| 2.2.1      | As Causas do Clogging                                                      | 43        |
| 2.2.2      | As Origens de Clogging                                                     | 43        |
| 2.2.3      | As Consequências do Clogging                                               | 45        |
| 2.2.4      | Formas de Prevenir o Clogging                                              | 45        |
| 2.2.4.1    | Injeção de Argônio                                                         | 46        |
| 2.2.4.2    | Tratamento com Cálcio                                                      | 47        |
| 2.2.4.3    | Tipos de Bocais                                                            | 47        |
| 2.2.5      | Exemplo de Análise de Segmento Obstruído                                   | 48        |
| 2.2.6      | Deteção do Clogging                                                        | 50        |
| <b>2.3</b> | <b>Trabalhos Relacionados</b>                                              | <b>54</b> |
| <b>3</b>   | <b>APRENDIZAGEM PROFUNDA</b>                                               | <b>69</b> |
| <b>3.1</b> | <b>Aprendizagem Profunda em Tarefas de Classificação</b>                   | <b>70</b> |
| <b>3.2</b> | <b>Redes Neurais MultiLayer Perceptron (MLP)</b>                           | <b>70</b> |
| 3.2.1      | Convolutional Neural Network (CNN)                                         | 72        |
| 3.2.2      | Long Short Term Memory (LSTM)                                              | 78        |
| 3.2.3      | Convolutional Long Short-Term Memory (ConvLSTM)                            | 80        |
| <b>3.3</b> | <b>Validação dos Modelos</b>                                               | <b>82</b> |
| <b>3.4</b> | <b>Métricas de Desempenho do Classificador</b>                             | <b>83</b> |
| <b>4</b>   | <b>METODOLOGIA</b>                                                         | <b>87</b> |
| <b>4.1</b> | <b>Análise Descritiva das Variáveis do Processo</b>                        | <b>87</b> |

|       |                                                                                                                               |     |
|-------|-------------------------------------------------------------------------------------------------------------------------------|-----|
| 4.2   | O Tratamento do Conjunto de Dados . . . . .                                                                                   | 90  |
| 4.3   | Exemplos de Ocorrência de <i>Clogging</i> no Conjunto de Dados . . . . .                                                      | 93  |
| 4.3.1 | Caso 1 . . . . .                                                                                                              | 93  |
| 4.3.2 | Caso 2 . . . . .                                                                                                              | 94  |
| 4.3.3 | Caso 3 . . . . .                                                                                                              | 95  |
| 4.3.4 | Caso 4 . . . . .                                                                                                              | 96  |
| 4.4   | A Formação das Matrizes de Entradas . . . . .                                                                                 | 97  |
| 4.5   | A Organização dos Dados . . . . .                                                                                             | 99  |
| 4.5.1 | O Balanceamento dos Dados . . . . .                                                                                           | 99  |
| 4.5.2 | A Divisão do Conjunto de Dados . . . . .                                                                                      | 100 |
| 4.5.3 | A Normalização dos Dados . . . . .                                                                                            | 100 |
| 4.6   | Os Classificadores . . . . .                                                                                                  | 104 |
| 4.6.1 | LSTM . . . . .                                                                                                                | 106 |
| 4.6.2 | CNN-LSTM . . . . .                                                                                                            | 106 |
| 4.6.3 | ConvLSTM . . . . .                                                                                                            | 107 |
| 4.6.4 | MLP . . . . .                                                                                                                 | 110 |
| 4.7   | A Seleção dos Classificadores . . . . .                                                                                       | 110 |
| 5     | RESULTADOS . . . . .                                                                                                          | 113 |
| 5.1   | Heurística de Pós-Processamento . . . . .                                                                                     | 122 |
| 6     | CONCLUSÃO . . . . .                                                                                                           | 127 |
|       | REFERÊNCIAS . . . . .                                                                                                         | 131 |
|       | APÊNDICE A – CURVAS DE TREINAMENTO E VALIDAÇÃO<br>DAS REDES LSTM . . . . .                                                    | 141 |
|       | APÊNDICE B – CURVAS DE TREINAMENTO E VALIDAÇÃO<br>DAS REDES CNN-LSTM . . . . .                                                | 143 |
|       | APÊNDICE C – CURVAS DE TREINAMENTO E VALIDAÇÃO<br>DAS REDES CONVLSTM . . . . .                                                | 145 |
|       | APÊNDICE D – CURVAS DE TREINAMENTO E VALIDAÇÃO<br>DAS REDES MLP COM 1 CAMADA OCULTA . . . . .                                 | 147 |
|       | APÊNDICE E – MODELO MLP COM DUAS CAMADAS OCULTAS                                                                              | 151 |
|       | APÊNDICE F – RESULTADOS COMPARATIVOS DOS MODE-<br>LOS APÓS A REDUÇÃO DO NÚMERO DE NEURÔ-<br>NIOS DA CAMADA DE SAÍDA . . . . . | 153 |

# 1 Introdução

Desde o seu surgimento, a indústria passou por profundas transformações, aperfeiçoando a sua forma de produção até chegar no modelo atual de Indústria 4.0, também chamada de Quarta Revolução Industrial. A principal característica da Indústria 4.0 é o uso de produtos e processos mais autônomos e eficientes, promovendo soluções customizadas para a produção, logística e clientes, especialmente pela aplicação da Inteligência Artificial (IA) (KOCH et al., 2014; AMARAL; GASPAROTTO, 2021).

A ascensão da IA no meio industrial oferece à indústria a possibilidade do aumento de produtividade e controle de cada processo da cadeia produtiva. Essas ferramentas auxiliam, por exemplo, na identificação de falhas que seriam humanamente imperceptíveis, devido ao número considerável de informações disponíveis. Além disso, muitas dessas ferramentas contam com baixo custo de implementação e resultados rápidos (AMARAL; GASPAROTTO, 2021).

Uma das principais vantagens da IA é a sua capacidade de aprender com dados as relações entre as variáveis e ajustar-se automaticamente às condições do processo, proporcionando uma maior flexibilidade e agilidade na produção. Isso contrasta com abordagens mais tradicionais, que muitas vezes exigem intervenção manual e reprogramação frequente para lidar com novas situações e requisitos.

A IA pode potencializar a capacidade analítica processual e previsibilidade com alta precisão e assertividade. Como consequência, a IA não apenas contribui para o aumento da capacidade produtiva do processo, como também na redução dos custos produtivos, garantindo produtos de qualidade superior. Como reflexo, o consumidor final é favorecido (AMARAL; GASPAROTTO, 2021).

É importante reconhecer, no entanto, que a IA também pode apresentar desafios e limitações, como a complexidade computacional e a necessidade de dados de alta qualidade para treinamento, o que muitas vezes as indústrias carecem. Apesar disso, acredita-se que os benefícios proporcionados pela aplicação da IA na Indústria 4.0 superam esses desafios, tornando-a uma ferramenta indispensável para o futuro da produção industrial.

O mercado siderúrgico é um setor economicamente importante devido ao encadeamento de diversos setores produtivos, sendo o responsável por fornecer insumos para, por exemplo, a construção civil, o setor automobilístico e o de utensílios domésticos. Atualmente, o Brasil é o nono maior produtor de aço bruto mundial, com capacidade instalada para produção de 48,9 milhões de toneladas de aço por ano (VIANA, 2020).

Com o desenvolvimento de tecnologias siderúrgicas e a busca pela eficiência ener-

gética, sobretudo, motivados pela Indústria 4.0, o mercado tem exigido produtos de alta qualidade para aplicações cada vez mais específicas. Neste contexto, o processo de lingotamento contínuo conquistou um papel de destaque na siderurgia, uma vez que o conhecimento e o controle de suas variáveis são de grande importância para se produzir aços de qualidade.

Este processo é responsável pela solidificação do aço líquido que flui verticalmente do distribuidor para o molde, por meio da válvula submersa. O fluxo de aço nesta válvula é controlado pela válvula gaveta, localizada na saída do distribuidor. Embora esta tecnologia esteja bastante difundida no Brasil, os metalurgistas convivem com uma série de problemas devido à formação de óxidos sólidos e a solidificação do aço nas paredes e no bocal da válvula submersa, provocando o seu entupimento, fenômeno conhecido como *clogging* (IKÄHEIMONEN et al., 2002; NASCIMENTO, 2008).

## 1.1 Caracterização do Problema

Ikäheimonen et al. (2002) relatam que existem vários fatores que contribuem para a formação do *clogging*, dentre eles, fatores metalúrgicos, hidrodinâmicos, termodinâmicos, de materiais, além de distúrbios imprevisíveis e falhas operacionais.

O *clogging* é um fenômeno que dificulta o controle do fluxo de metal líquido ao longo do processo e, por sua vez, pode aumentar a frequência de interrupções da operação para a troca e/ou manutenção dos bocais e distribuidores ou, em casos mais críticos, provocar a interrupção do processo. Essas atuações extras aumentam o custo operacional e diminuem a produtividade (THOMAS; BAI, 2001). O *clogging* também afeta adversamente a qualidade do produto, alterando o padrão de fluxo e degradando a qualidade interna do produto final, principalmente, quando grandes pedaços dessas impurezas se quebram e entram no fluxo corrente do aço (RACKERS, 1995; RACKERS; THOMAS, 1995; THOMAS; BAI, 2001).

Na prática, as siderurgias não contam com ferramentas eficazes disponíveis que possam quantificar corretamente essa obstrução durante a operação de lingotamento (ROUT et al., 2013). Na literatura é possível encontrar diferentes pesquisas relacionadas à detecção das ocorrências de *clogging* no lingotamento contínuo. Apesar de haver diferentes fatores de formação do *clogging*, essas pesquisas geralmente estabelecem uma correlação entre o fenômeno de *clogging* com as variáveis de grau de abertura da válvula e velocidade de escoamento do aço no veio, este último referente ao aço em processo de solidificação. Essas análises frequentemente se baseiam em modelos descritos por equações da física (THOMAS; BAI, 2001; IKÄHEIMONEN et al., 2002; VANNUCCI; COLLA, 2011; ROUT et al., 2013; BARATI et al., 2018; KONG; CANG, 2019; KONG; CHEN; CANG, 2019; PELLEGRINI et al., 2019).

Considerando que o evento de *clogging* é um tipo de falha operacional, o pequeno número de ocorrências frente aos dados operacionais normais acaba comprometendo a habilidade dos classificadores durante a fase de aprendizagem (VANNUCCI; COLLA, 2011). Somando-se a esses fatores, devido às condições hostis, variações na produção, falhas de processo e/ou sensores, dados geralmente ruidosos, contendo valores discrepantes (*outliers*) e até mesmo ausentes contribuem para que a precisão preditiva desejada pela indústria<sup>1</sup> dificilmente possa ser garantida (IKÄHEIMONEN et al., 2002; VANNUCCI; COLLA, 2011; ROUT et al., 2013; RAJENDRA; RACHAPPA; ANIL, 2016; KONG; CANG, 2019; KONG; CHEN; CANG, 2019; PELLEGRINI et al., 2019; DINIZ et al., 2021).

## 1.2 Motivação

Este estudo foi motivado pela crescente demanda da indústria siderúrgica na detecção das ocorrências iniciais de obstruções no SEN (*Submerged Entry Nozzle*, em português: bocal de entrada submerso) no processo de lingotamento contínuo. Esta detecção permite que ações corretivas possam ser realizadas ao longo do processo de obstrução ocasionando em um maior tempo de operação do sistema, uma descarga de impurezas melhor distribuída ao longo do lingotamento e um maior intervalo de substituição da válvula submersa. Essas medidas contribuem significativamente para o melhor aproveitamento da vida útil da válvula, desempenhando um papel fundamental na economia de produção e na melhoria da qualidade do produto final.

Um aspecto motivacional está no desafio de encontrar uma solução baseada no uso de aprendizado de máquinas, uma vez que os trabalhos encontrados na literatura, em geral, propõem soluções orientadas direta ou indiretamente por equações físicas e, conseqüentemente, algumas simplificações se tornam inevitáveis. Além disto, o uso de técnicas de Inteligência Artificial permite a extração sistemática de informações inerentes ao processo a partir dos dados, aperfeiçoando os padrões operacionais e beneficiando a qualidade do produto final.

Conforme discutido por Yang et al. (2021), com o avanço acelerado da tecnologia da informação, espera-se que a disponibilidade de dados de processos industriais cresça exponencialmente. Neste sentido, buscando-se alavancar de modo eficiente o *big data* industrial, de maneira a auxiliar na tomada de decisões em tempo real, as indústrias precisam incorporar o conhecimento de campo ao processo de modelagem orientado aos dados. Por conseguinte, essas ações ajudarão a impulsionar a competitividade da indústria nacional, colaborando para a implementação de ações da Indústria 4.0, o que justifica a

---

<sup>1</sup> É importante ressaltar que não existe um valor máximo ou mínimo universalmente aceitável nesse contexto. Em vez disso, a precisão preditiva é adaptada às características específicas de cada planta e aos objetivos particulares da indústria. O principal objetivo é obter previsões com baixas taxas de falsos negativos, garantindo que problemas reais sejam identificados.

busca pelo desenvolvimento e a realização de pesquisas no país.

Outro aspecto desafiador está na natureza dos dados que, por se tratar de situações reais em uma tarefa de classificação de falhas, há um acréscimo de contingências operacionais e ruídos, além de um desequilíbrio grande da quantidade de amostras entre as classes. Como resultado, a precisão preditiva desejada pela indústria torna-se desafiadora tendo que atingir taxas de falsos negativos (taxas de falhas não detectadas) relativamente baixas.

A complexidade deste estudo, portanto, reside na estimação do momento inicial da obstrução, antes mesmo que as variáveis manifestem visualmente um comportamento anômalo do sistema. Assim, abordar o problema levando-se em consideração as não-linearidades e variações no tempo dos parâmetros do processo pressupõe a busca por soluções através de modelos complexos de aprendizado de máquinas, permitindo atingir níveis significativos de Precisão e *Recall*<sup>2</sup>.

Os resultados deste projeto de pesquisa poderão favorecer ganhos de produtividade e redução dos custos. Considerando uma usina siderúrgica com capacidade de produção anual de cerca de 7,5 milhões de toneladas de aço, é possível estimar uma economia de mais de US\$15 milhões por ano. Diante dos fatos mencionados, este trabalho é motivado não apenas pelo estímulo à competitividade da indústria nacional frente ao mercado siderúrgico mundial, como também pelo estreitamento da relação da indústria e a universidade (ROUT et al., 2013; JÚNIOR; SILVA; SILVA, 2018).

## 1.3 Hipótese

Considerando o levantamento bibliográfico previamente realizado (RACKERS, 1995; THOMAS; BAI, 2001; IKÄHEIMONEN et al., 2002; VANNUCCI; COLLA, 2011; ROUT et al., 2013; KONG; CHEN; CANG, 2019; PELLEGRINI et al., 2019), juntamente com a análise dos dados fornecidos pela indústria (para este trabalho) e conhecimento prático dos especialistas do processo, foram identificadas as variáveis que possivelmente possuem uma relação com as ocorrências de *clogging* no lingotamento contínuo.

Neste sentido, a aplicação de uma ferramenta capaz de identificar precocemente o início das obstruções pode permitir que o operador tenha tempo suficiente para tomar medidas corretivas durante o processo. Conforme abordado por Pellegrini et al. (2019) e Cemernek et al. (2021), em cenários industriais de grande complexidade com múltiplas variáveis interconectadas, os modelos de aprendizado de máquina, em especial as técnicas

---

<sup>2</sup> Precisão é a proporção de exemplos classificados como positivos que realmente são positivos. *Recall* é a taxa de classificação de verdadeiros positivos, representando a proporção de amostras positivas corretamente identificadas por um modelo ou sistema (FAWCETT, 2006; ROGERS; GIROLAMI, 2016). Mais detalhes sobre o cálculo dessas métricas e informações adicionais estão no Capítulo 3.

de *Deep Learning*, têm se destacado por sua capacidade em identificar padrões e relações complexas entre as variáveis.

Isto posto, essa tese fundamenta-se na seguinte hipótese:

*“Supõe-se que a modelagem baseada em Deep Learning possa resultar em modelos de classificadores capazes de identificar o início de ocorrências de clogging com altos níveis de precisão e baixas taxas de falsos negativos. Esta suposição será investigada considerando quatro variáveis do processo: abertura da válvula submersa, nível do molde, velocidade do veio e distribuidor em operação.”*

## 1.4 Objetivos

Pretende-se obter um modelo intrinsecamente orientado a dados considerando a não-linearidade e variabilidade das condições do processo, capaz de efetuar a identificação precoce dos eventos de *clogging* na válvula submersa. Detectar essas obstruções precocemente possibilita a intervenção imediata do operador, evitando paradas de processo não programadas, prolongamento da vida útil da válvula, assegurando a qualidade do produto final.

É importante ressaltar que este trabalho não busca determinar o local exato dentro da válvula submersa onde o fenômeno de *clogging* está ocorrendo. Da mesma forma, não se dedica a investigar as causas específicas desse fenômeno nem estabelecer valores para variáveis em pontos específicos do sistema. O objetivo central deste estudo é identificar o início do processo de obstrução, permitindo ao operador tomar medidas corretivas apropriadas.

As informações obtidas a partir da modelagem proposta poderão contribuir para o aperfeiçoamento dos padrões operacionais e na melhoria da qualidade do aço produzido pelo processo de lingotamento contínuo, colaborando para a implementação de ações da Indústria 4.0, e, por conseguinte, ajudando a aumentar a competitividade da indústria nacional.

Assim, o presente estudo tem como objetivo principal o desenvolvimento de um modelo baseado em aprendizado de máquinas para a identificação do início do *clogging* na válvula submersa do distribuidor, no processo de lingotamento contínuo. Especificamente, este trabalho propõe:

- Estudo dos fatores que exercem maior influência sobre a obstrução da válvula submersa do distribuidor no lingotamento;

- Análise da viabilidade de uso dos dados fornecidos e aplicação de técnicas para a realização do pré-processamento dos dados, garantindo que não ocorra a redução das informações importantes ao processo;
- Implementação e comparação de modelos baseados em aprendizado de máquina, sobretudo *Deep Learning*, capazes de caracterizar o início da ocorrência de uma obstrução ao longo do processo;
- Aplicação de redes *MultiLayer Perceptrons* (MLP) como linha de base para comparação dos modelos propostos;
- Implementação de uma heurística de pós-processamento visando reduzir a quantidade de falsos positivos e negativos.

## 1.5 Método Proposto

Ao discutirem o uso de redes neurais em problemas industriais complexos, por exemplo, [Pellegrini et al. \(2019\)](#) e [Cemernek et al. \(2021\)](#) sugerem as soluções em aprendizado de máquinas como abordagens promissoras. Partindo desta premissa, esta pesquisa analisa o desempenho de modelos baseados em *Deep Learning* aplicados aos dados históricos coletados ao longo de seis meses em uma indústria siderúrgica, buscando uma modelagem que melhor se adapte às constantes variações do sistema e que seja capaz de identificar de maneira precisa o início do *clogging*.

Neste estudo, será feita uma comparação entre o desempenho geral de classificadores utilizando arquiteturas envolvendo *Long Short-Term Memory* (LSTM) e Redes Neurais Convolucionais (*Convolutional Neural Networks* – CNN). Para avaliar a eficiência dos modelos propostos, modelos MLP serão utilizados como linha de base para comparação.

Além da LSTM tradicional, que leva em conta apenas as informações temporais das variáveis, será testada uma abordagem híbrida, denominada CNN-LSTM. Em especial, esta abordagem explora a capacidade das camadas convolucionais em aprender a representação interna de dados de séries temporais, bem como a capacidade das LSTM em identificar dependências de curto e longo prazo. [Agga et al. \(2021\)](#) destacam que a capacidade de extrair características temporais e espaciais torna a aplicação desses modelos híbridos interessante não apenas para problemas envolvendo séries temporais, como também em diferentes aplicações como processamento de linguagem natural, previsões de mercado e problemas de visão computacional.

Também será avaliado o desempenho de uma variante do LSTM, denominada *Convolutional Long Short-Term Memory* (ConvLSTM), em que as combinações lineares das células LSTM são substituídas por operações de convolução. Ao fazer isso, o ConvLSTM se torna capaz de processar características intrínsecas das séries temporais, usando a operação

de convolução internamente nas próprias unidades LSTM (BROWNLEE, 2018; AGGA et al., 2021).

A seleção das variáveis de entrada dos modelos será baseada em experimentos realizados na literatura (IKÄHEIMONEN et al., 2002; VANNUCCI; COLLA, 2011; ROUT et al., 2013; KONG; CHEN; CANG, 2019; KONG; CANG, 2019; PELLEGRINI et al., 2019), e de acordo com as constatações de especialistas do processo, os quais classificam como *clogging* o evento correspondente à abertura da válvula sem um aumento significativo do nível, podendo ou não haver variação da velocidade do lingotamento. Na maioria dos casos, a abertura da válvula ocorre após oscilações crescentes, que podem ou não refletir em oscilações do nível.

A natureza dos dados, no entanto, não permite o uso de regras triviais para classificar as ocorrências de *clogging*, uma vez que uma série de fatores podem afetar o processo, conforme exposto no estudo de Thomas e Bai (2001). Neste contexto, inicialmente será realizada a análise descritiva das variáveis, sendo computados intervalos de valores teóricos e reais e seus respectivos percentis. A partir de então será considerado o comportamento de cada variável com o início da ocorrência do *clogging*, bem como a sua viabilidade na aplicação ao modelo.

Cabe salientar que a utilização das variáveis relacionadas à temperatura, pressão e vazão de argônio, embora sejam importantes para a identificação do *clogging* (SZEKERES, 1992; RACKERS, 1995; RACKERS; THOMAS, 1995; THOMAS; BAI, 2001; IKÄHEIMONEN et al., 2002; SAHAI; EMI, 2008; VANNUCCI; COLLA, 2011; VANNUCCI; COLLA; CATENI, 2015; PELLEGRINI et al., 2019; YANG; OLIA; THOMAS, 2021) não estão disponibilizadas de maneira adequada no conjunto de dados disponíveis e, consequentemente, a possível aplicação se tornou inviável.

Tendo em vista as possíveis falhas de processo e/ou sensores e condições hostis inerentes ao processo, dados serão pré-processados para o tratamento de amostras ausentes e *outliers*, através de filtros de mediana móvel e utilização de valores de máximo e mínimo teóricos, respectivamente. Em relação aos ruídos, será considerada a capacidade das redes neurais em lidar com dados ruidosos, deste modo, visando preservar as informações inerentes ao processo, portanto, inicialmente, nenhum tipo de filtragem será aplicado.

Considerando que a ocorrência de *clogging* é classificada como uma falha operacional, espera-se que exista um certo desequilíbrio entre as classes de amostras (*clogging* e operação normal). Deste modo, técnicas de balanceamento de dados ou de seleção de dados deverão ser aplicadas ao conjunto de dados de treinamento, de maneira a garantir que o desequilíbrio não interfira no desempenho dos modelos testados.

Pretende-se então desenvolver uma modelagem que seja capaz de lidar com dados reais, obtidos diretamente do processo, alcançando os padrões de desempenho esperados

pela indústria, com baixas taxas de falsos negativos. Este é um resultado desejado no diagnóstico de falhas, uma vez que é preferível que os operadores recebam um alarme falso de *clogging* (falso positivo) ao invés do evento ocorrer sem qualquer aviso prévio (falso negativo).

O uso de um modelo baseado em aprendizado de máquinas e que se adapte às diferentes condições do lingotamento contínuo pode inclusive fornecer aos operadores ferramentas para a simulação, previsão e estudo de testes hipotéticos em processos, que seriam dispendiosos de serem criados na prática (YANG et al., 2021). Vale destacar que, até o presente momento, não foram encontrados trabalhos na literatura que explorem a detecção das ocorrências iniciais de *clogging* exclusivamente orientado aos dados do processo. Além disso, os estudos até o momento realizados não conseguiram alcançar resultados com baixas taxas de falsos positivos e, especialmente, falsos negativos em aplicações reais.

## 1.6 Produção Bibliográfica

Durante o doutorado e o desenvolvimento da pesquisa, foram aceitos em conferências e periódicos os seguintes artigos:

- DINIZ, A. P. M., CIARELLI, P. M., SALLES, E. O., COCO, K. F. Heat Transfer in Steel Ladles: Models and Applications. Congresso Brasileiro de Automática (CBA), v. 2, n. 1, 2020.
- DINIZ, A. P. M., CIARELLI, P. M., SALLES, E. O., COCO, K. F. Long short-term memory neural networks for clogging detection in the submerged entry nozzle. Decision Making: Applications in Management and Engineering, v. 5, n. 1, p. 154-168, 2022.
- DINIZ, A. P. M., CIARELLI, P. M., SALLES, E. O., COCO, K. F. Use of deep neural networks for clogging detection in the Submerged Entry Nozzle of the continuous casting. Expert Systems with Applications, v. 238, p. 121963, 2024.

## 1.7 Estrutura da Tese

Este trabalho possui a seguinte organização. No Capítulo 2 será apresentado o processo de lingotamento contínuo, bem como discutidas as causas e efeitos da ocorrência de *clogging* neste processo. No Capítulo 3 serão descritas as técnicas baseadas em *Deep Learning* empregadas no trabalho, tais como as redes LSTM, CNN-LSTM e ConvLSTM, além das métricas de desempenho utilizadas na tarefa de classificação. No Capítulo 4, o conjunto de dados é apresentado juntamente com o passo a passo da metodologia

---

proposta. Os resultados e as observações relevantes serão discutidos no Capítulo 5. Por fim, o Capítulo 6 traz as considerações finais e a proposta de trabalhos futuros.



## 2 O Clogging da Válvula Submersa no Processo de Lingotamento Contínuo

Para abordar o problema do *clogging* no processo de lingotamento, faz-se necessário entender como é uma planta de lingotamento contínuo, seu funcionamento e, especificamente, o funcionamento da válvula, bem como outras características importantes desse processo de produção de placas de aço.

### 2.1 O Processo de Lingotamento Contínuo

O lingotamento contínuo é uma das etapas mais importantes do processo siderúrgico de fabricação do aço. Este é o último estágio pós aciaria, cujo objetivo principal é a solidificação do aço líquido, resultando na placa ou tarugo de aço como produto final para o transporte (MOURÃO et al., 2011).

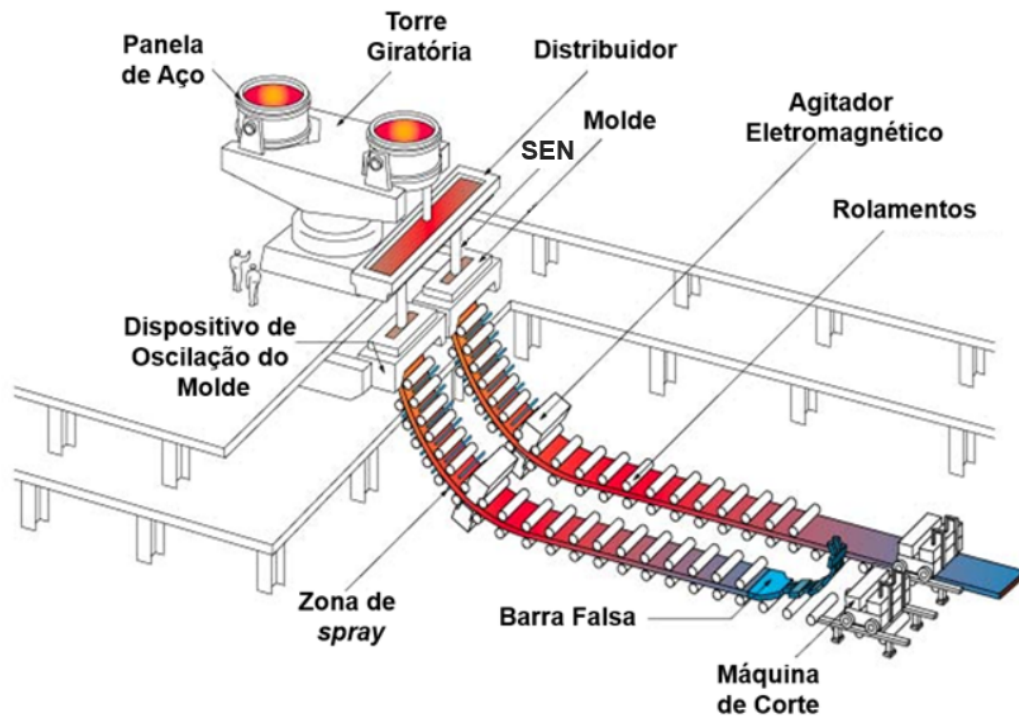
O princípio básico do lingotamento contínuo consiste no vazamento vertical do aço líquido, a partir de uma panela posicionada sobre o distribuidor, que é um mecanismo aplicado para alimentar os moldes de cobre utilizando tubos de imersão (veja esquemático na Figura 1). O aço derretido é, então, introduzido na cavidade do molde através do bocal de entrada submerso (*Submerged Entry Nozzle* – SEN), também chamado de válvula submersa.

O distribuidor aqui tem papel fundamental para a não interrupção do lingotamento, bem como o molde de envase tem grande importância na remoção adicional de inclusões de partículas sólidas, como os óxidos de metais (RACKERS, 1995; MOURÃO et al., 2011). Os distribuidores alternam no molde de maneira a mantê-lo sempre cheio de aço líquido. A troca do distribuidor é uma operação que leva em média 5 minutos.

Para que a solidificação avance, o material que sai do molde é resfriado pelo uso de jatos de água, distribuídos uniformemente ao longo da extensão da máquina (Zona de *Spray*), reduzindo progressivamente a temperatura do aço. Para auxiliar no processo de escoamento, o molde oscila a aproximadamente 1 [Hz] e um pó lubrificante é adicionado para reduzir o atrito nas paredes do molde. Segundo Mourão et al. (2011), estes dois princípios são fundamentais para evitar a adesão da camada solidificada ao molde.

O lingote formado é então conduzido por rolos de suporte até a parte final do equipamento. Nesta região, normalmente, toda a solidificação deverá ter ocorrido e o resfriamento prossegue por radiação livre. Por fim, esta placa é seccionada em comprimentos pré-definidos, conforme a aplicação final do produto (MOURÃO et al., 2011). A velocidade

Figura 1 – Resumo esquemático de um processo de lingotamento contínuo.



Fonte: Adaptada de [Ishii et al. \(2006\)](#).

média de lingotamento fica em torno de  $1,2 [m/min]$  e as placas são cortadas, em média, a cada  $10 [m]$ , pesando cada uma, aproximadamente, 23 toneladas.

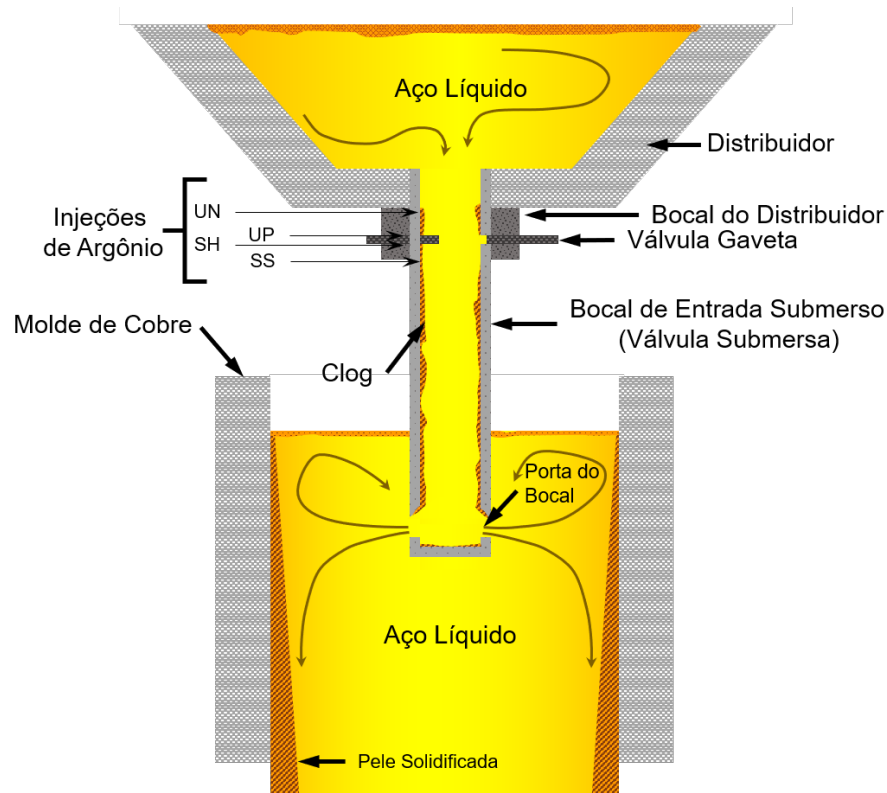
[Arrivabene, Machado e Sobrinho \(2003\)](#) relatam que o lingotamento contínuo foi o avanço tecnológico mais importante desde os anos sessenta, uma vez que ele permitiu a substituição do lingotamento convencional, que é um processo mais lento e oneroso. Essa mudança possibilitou a redução de tempo e de custos, promovendo uma significativa melhora na qualidade do aço. Apesar disso, alguns problemas apresentados nesta etapa da produção ainda não foram resolvidos como, por exemplo, defeitos nos produtos (quebras, inclusões de partículas, porosidades e segregação<sup>1</sup>) e problemas operacionais (obstruções e entupimento da válvula).

### 2.1.1 A Válvula Submersa

A válvula submersa possui um papel vital para a qualidade do produto final do lingotamento. Ela busca evitar a incorporação de oxigênio, proveniente do ar, no aço e consequente formação de óxidos durante a transferência do metal líquido entre o distribuidor e o molde. Os óxidos podem acarretar vários problemas na placa de aço reduzindo sua qualidade em termos de composição química e de conformação mecânica ([RACKERS, 1995](#); [IKÄHEIMONEN et al., 2002](#); [YANG; OLIA; THOMAS, 2021](#)).

<sup>1</sup> Distribuição não uniforme dos elementos ou componentes no aço.

Figura 2 – Resumo esquemático da passagem do aço do distribuidor para o molde com controle do fluxo realizado através da válvula gaveta.



Fonte: Elaborada pela autora.

A Figura 2 ilustra um esquema típico do fluxo de aço do distribuidor para o molde. O fluxo de aço líquido passa através do bocal do distribuidor, sendo regulado pela abertura da válvula gaveta (RACKERS, 1995).

A válvula gaveta é um dispositivo formado por três placas sobrepostas. Em cada uma dessas placas há um orifício idêntico, de modo que, quando a placa central realiza um movimento horizontal, a área de intersecção dos orifícios dessas placas se converte na área de passagem de aço (SALLES; ALMEIDA, 2016; PEREIRA et al., 2021).

O fluxo de aço segue então pelo bocal de entrada submerso (válvula submersa), chegando ao molde de cobre onde começa a ser refrigerado para início do processo de solidificação e conformação da placa (RACKERS, 1995).

As injeções de argônio podem ocorrer na parte superior do bocal conectado ao distribuidor (*Upper Tundish Well Nozzle* – UN), na folga da placa superior da válvula deslizante (*Upper Plate Gap* – UP), no suporte de apoio da placa deslizante (*Shroud Holder Gap* – SH) e na parede da parte baixa do bocal de entrada submerso (*Lower Submerged Entry Nozzle* – SS).

Um pequeno fluxo de argônio sempre deve ser mantido em torno das juntas, especialmente entre a placa deslizante da válvula (UN), a placa inferior (SH) e o suporte

do bocal de entrada submerso (SS) para garantir que o gás aspirado seja argônio e não ar (THOMAS; BAI, 2001).

Observa-se que o bocal divide o aço fundido em duas direções. O fluxo direto para baixo deve ser evitado, pois assim os compostos de alumínio na escória permaneceriam no núcleo do veio da placa. A dispersão lateral tende a jogar os óxidos para as regiões externas da placa de aço final, facilitando sua retirada durante a escarfagem<sup>2</sup>.

Ao longo do processo, parte dos óxidos e de outras impurezas do aço irão se depositar nas paredes da válvula e do bocal, desenvolvendo, gradativamente, uma obstrução (*Clog*). Tais impurezas podem ser provocadas pela reoxidação (YANG; OLIA; THOMAS, 2021), passagem de escória da panela ou distribuidor, bem como pelo processo de refino deficiente (aço muito oxidado e/ou sem tempo suficiente para flotação de inclusões). Esta obstrução altera o padrão de fluxo e afeta adversamente a qualidade interna do produto final, principalmente, quando grandes pedaços dessas impurezas se quebram e entram no fluxo de corrente do aço (THOMAS; BAI, 2001).

## 2.2 O Clogging da Válvula Submersa

O *clogging* é definido como depósito de material nas paredes da SEN (*Submerged Entry Nozzle*) ou na própria válvula gaveta restringindo a passagem do aço entre o distribuidor e o molde. Em consequência, o *clogging* acarreta uma série de problemas de qualidade (RACKERS; THOMAS, 1995; THOMAS; BAI, 2001; SAHAI; EMI, 2008).

Conforme a obstrução progride, a válvula gaveta vai se abrindo para manter a vazão de aço, entretanto, quando a válvula atinge sua posição máxima (grau de obstrução elevado), ela perde a capacidade de controle do fluxo de aço, então a produção deve ser interrompida para a substituição do conjunto distribuidor (SEN e/ou da válvula gaveta) de forma não planejada (RACKERS, 1995; THOMAS; BAI, 2001).

O tempo de vida média de operação do conjunto distribuidor é de 450 minutos devido à obstrução, isto é, a cada 8 corridas, aproximadamente, é feita a troca do conjunto distribuidor, embora a sua vida útil seja de até 900 minutos (aproximadamente 16 corridas). Em contrapartida, o SEN é trocado, em média, a cada 4 corridas, apesar de possuir vida útil de 600 minutos (em torno de 10 corridas). Este processo chega a parar o escoamento do veio em até 1 minuto (RACKERS, 1995; THOMAS; BAI, 2001; IKÄHEIMONEN et al., 2002). É interessante salientar que, embora existam outros fatores que determinem as trocas, todos esses fatores são planejados, diferentemente do *clogging*, que é um evento de troca não planejado, que tem como algumas das consequências o aumento na frequência das trocas e a redução da vida útil do SEN.

<sup>2</sup> Envolve a remoção de uma pequena camada superficial de aço para eliminar defeitos superficiais presentes na placa (TAVARES, 2019).

### 2.2.1 As Causas do Clogging

Conforme evidencia [Ikäheimonen et al. \(2002\)](#), existem vários fatores que contribuem para a formação do *clogging*, a citar: fatores metalúrgicos (limpidez do aço e teor de alumínio), fatores hidrodinâmicos (taxa de fluxo do aço e geometria do bocal), fatores termodinâmicos (aço frio e fenômenos de transferência de calor no interior do bocal), material do bocal, bem como distúrbios imprevisíveis e falhas operacionais.

[Rackers e Thomas \(1995\)](#) afirmam que para eventos de *clogging* constituídos por aço solidificado e/ou produtos de reação, o transporte e mecanismo de adesão são considerados simples, tendo em vista que o fenômeno do entupimento ocorre na parede do bocal. Já para o *clogging* contendo inclusões<sup>3</sup>, os processos de inclusão e transporte são mais complexos.

A presença de uma subcamada viscosa de aço mais próxima à parede do bocal dificulta o transporte de inclusões para a parede. Como o fluxo nessa camada é predominantemente paralelo à parede, pouca força motriz está presente para promover a deposição de inclusões. Assim, algumas teorias foram propostas afim de explicar a ocorrência do transporte suficiente de inclusões nessa subcamada para promover o entupimento do bocal.

Em geral, as teorias se baseiam no movimento irregular do fluxo de metal e características de forma e de rugosidade da superfície do bocal, todavia, o processo de fixação das inclusões na parede do bocal se deve basicamente à tensão superficial do aço influenciada, principalmente, pela temperatura, que provoca uma força atrativa entre a inclusão e a parede do bocal ([WILSON et al., 1987](#); [UEMURA et al., 1992](#); [SAHAI; EMI, 2008](#)).

### 2.2.2 As Origens de Clogging

De maneira geral, existem quatro tipos de *clogging*, cada um com diferentes origens, e, na prática, o fenômeno se manifesta pela combinação de duas ou mais origens ([RACKERS, 1995](#); [RACKERS; THOMAS, 1995](#); [THOMAS; BAI, 2001](#); [SAHAI; EMI, 2008](#); [YANG; OLIA; THOMAS, 2021](#)).

[Thomas e Bai \(2001\)](#) classificam os eventos de *clogging* de acordo com seu mecanismo de formação: o transporte de óxidos presentes no aço para a parede do bocal, aspiração de ar no bocal, reação química entre o refratário do bocal e o aço e entre o aço e o aço solidificado no bocal, no entanto, eles afirmam que as causas exatas podem ser difíceis de

<sup>3</sup> As inclusões são materiais compostos formados pela associação metal/não-metal originados pelo processo de refino, solidificação do aço ou ainda de fontes externas. De acordo com as suas origens, as inclusões podem ser classificadas como exógenas ou endógenas. As inclusões endógenas resultam principalmente da desoxidação e solidificação do aço, enquanto que as inclusões exógenas são geradas, por exemplo, pela reoxidação e por interações mecânicas do aço líquido com os refratários ([ZHANG; THOMAS, 2003](#)). De acordo com [Millman \(2004\)](#), como as inclusões endógenas são intrínsecas ao processo de fabricação de aço, elas não podem ser totalmente eliminadas. As inclusões exógenas, todavia, podem ser eliminadas por meio do controle do processo.

identificar.

Em especial, é possível distinguir as obstruções por aglomeração de produtos da desoxidação, acúmulo de aço sólido, aglomeração de óxidos complexos e acúmulo de produtos de reação (RACKERS, 1995). A seguir são descritas cada uma das causas:

- **Aglomeração de Produtos da Desoxidação:** Este tipo de obstrução pode surgir pelo acúmulo de produtos formados nas fases anteriores do processo de produção do aço, bem como pela reoxidação formada antes do aço líquido entrar no bocal do distribuidor (por exemplo, alumina, titânio e zircônio), devido à exposição do aço ao ar (YANG; OLIA; THOMAS, 2021), aprisionamento de escória, inclusões exógenas de outras fontes e reações químicas (THOMAS; BAI, 2001). O óxido de alumínio é o principal fator de *clogging* nos processos de lingotamento contínuo do aço (BARATI et al., 2018; KONG; CANG, 2019);
- **Acúmulo de Aço Sólido:** Embora as perdas térmicas dos refratários dos bocais sejam pequenas, a solidificação de aço no bocal, denominada *skulling*, ocorre quando o pré-aquecimento do bocal foi inadequado, ocorrendo normalmente no início do lingotamento ou se a vazão é muito lenta. Esses problemas são mais prováveis quando ocorre o descontrole térmico no refino secundário, isto é, o aço líquido que chega ao distribuidor apresenta uma temperatura inferior à visada (THOMAS; BAI, 2001);
- **Aglomeração de Óxidos Complexos:** Esses óxidos são materiais não metálicos não resultantes do processo de desoxidação. Szekeres (1992) e Thomas e Huang (1993) sugerem que o fluxo do molde seja atraído para o topo dos bocais, devido ao padrão de fluxo de recirculação na parte superior do molde, assim como há tendência do fluxo de revestir o bocal. De acordo com Byrne, Cramb e Fenicle (1985), uma vez depositado no bocal, o fluxo assimila as partículas de desoxidação, aumentando assim o volume de entupimento;
- **Acúmulo de Produtos de Reação:** Este acúmulo ocorre devido à formação de uma fina camada composta pelos produtos de desoxidação (por exemplo, alumina, titânia e zircônia) que se depositam na parede do bocal. Este tipo de acúmulo é resultante da reação entre o desoxidante e *i*) o ar que passa para o bocal, *ii*) o oxigênio liberado pelo aço e *iii*) o oxigênio gerado da decomposição do refratário. Esses mecanismos são consistentes com as observações relatadas de aumento das obstruções à medida que a concentração de alumínio solúvel cresce (FUKUDA; UESHIMA; MIZOGUCHI, 1992; YANG; OLIA; THOMAS, 2021).

### 2.2.3 As Consequências do *Clogging*

De acordo com [Rackers \(1995\)](#), as consequências do *clogging* incluem a redução da produtividade, aumento dos custos associados à produção e redução da qualidade do produto. A seguir, são explicadas com maiores detalhes as consequências:

- **Redução da Produtividade:** Os eventos de *clogging* aumentam a frequência de interrupções da operação para a troca e/ou manutenção dos bocais e distribuidores. Essas atuações extras elevam o custo operacional e reduzem a produtividade ([THOMAS; BAI, 2001](#));
- **Aumento dos Custos de Produção:** Alguns autores relatam que o entupimento do bocal limita a vida útil do distribuidor, podendo reduzir o tempo de vida em até a metade. Sendo assim, o entupimento do bico resulta em custos adicionais para a restauração do distribuidor, bem como para a substituição do bocal ([SIDMAR et al., 1988](#); [SCHMIDT; RUSSO; BEDERKA, 1991](#));
- **Decréscimo da Qualidade do Produto:** As obstruções alteram o padrão de fluxo dos bocais e as características do jato fazendo com que pequenas inclusões sólidas se quebrem e entrem no fluxo corrente do aço, alterando a composição do fluxo e ocasionando defeitos no produto final. Problemas de qualidade também podem surgir pelo transitório no nível do molde que ocorre quando o dispositivo de controle de fluxo precisa compensar a obstrução ([THOMAS; BAI, 2001](#)).

### 2.2.4 Formas de Prevenir o *Clogging*

Conforme discutido por [Rackers \(1995\)](#) e [Sahai e Emi \(2008\)](#), a maneira mais intuitiva de se reduzir o *clogging* é através da diminuição da concentração de produtos de desoxidação e formação de produtos de reoxidação. Neste contexto, importantes aspectos da chamada “Siderurgia Limpa”, revisados por [Byrne, Cramb e Fenicle \(1985\)](#) e [Szekeres \(1992\)](#), poderiam prevenir o *clogging*, incluindo:

- **Práticas de Refino na Panela:** Um tratamento de desgaseificação à vácuo que produza maior grau de limpidez do aço do que o borbulhamento de argônio;
- **Prevenção de Reoxidação:** Garantia de um derramamento submerso do aço para o distribuidor, com devido isolamento da superfície do distribuidor e uso de refratários não porosos, sem rachaduras e com juntas estanques;
- **Remoção de Inclusões:** Aplicação de padrões ideais de fluxo do distribuidor, filtração e uso de técnicas eletromagnéticas para removerem as inclusões do aço;

- **Prevenção de Arrastamento de Fluxo:** O derramamento submerso da panela para o distribuidor e o impedimento da transferência de escória, promovem a redução da quantidade de inclusões exógenas no material fundido.

Visando a redução da deposição das inclusões arrastadas, várias técnicas são utilizadas, dentre elas estão a injeção de argônio, o tratamento com cálcio e a alteração dos tipos de bocais.

#### 2.2.4.1 Injeção de Argônio

A [Figura 2](#) mostra os locais onde são injetados o argônio na válvula submersa. A injeção de gás argônio através da parede do bocal na taxa aproximada de 5 litros por minuto contribui na prevenção do *clogging*, uma vez que:

- Um filme de argônio é formado na parede do bocal, o que impede a inclusão de entrar em contato com a parede ([HEASLIP et al., 1987](#));
- As bolhas do argônio liberam as inclusões do bocal por turbulência, promovendo a sua flotação ([TAI; CHEN; CHOU, 1985](#); [HEASLIP et al., 1987](#); [CRAMB; JIMBO, 1989](#));
- O vácuo parcial no interior do bico diminui com a injeção do argônio, reduzindo assim a aspiração de ar através do bocal ([HEASLIP et al., 1987](#); [SASAKA et al., 1991](#); [SZEKERES, 1992](#));
- O argônio, por ser um gás nobre, não reage com o aço, e também é capaz de prevenir a reação química entre o aço e o refratário ([TAI; CHEN; CHOU, 1985](#)).

Embora o argônio altere o padrão de fluxo no molde ([ASSAR; DAUBY; LAWSON, 2000](#)), o excesso de argônio pode fazer com que as fases de líquido e gás se separem, resultando em problemas de fluxo instável no molde.

As desvantagens da injeção de argônio ainda incluem o aumento dos defeitos de qualidade, desenvolvimento de erosão na linha de escória devido ao aumento das flutuações no nível do molde ([SZEKERES, 1992](#)), aprisionamento de bolhas pela solidificação da superfície do aço ([BESSHO et al., 1991](#); [SASAKA et al., 1991](#); [SZEKERES, 1992](#)) e rachaduras nos bocais em decorrência da alta contrapressão ou do choque térmico pela diminuição da temperatura do bocal ([SASAKA et al., 1991](#)).

[McPherson e McLean \(1992\)](#) e [Rackers e Thomas \(1995\)](#) acreditam que a injeção de argônio tende a deslocar um dado problema de entupimento para um local diferente. Eles também supõem que a queda da pressão de argônio nos pontos de injeção tenha relação com a formação de alumina no bocal.

#### 2.2.4.2 Tratamento com Cálcio

O *clogging* de alumina pode ser reduzido com a adição de cálcio ao aço, prevenindo-se assim a solidificação da alumina (BOLGER, 1994), contudo, as desvantagens desse tratamento são:

- A adição insuficiente de cálcio pode aumentar o *clogging* se comparado à condição de ocorrência sem o tratamento, devido à modificação incompleta das inclusões de alumina e formação de aluminatos de cálcio sólido, com ponto de fusão elevado (WILSON et al., 1987; CHOUDHARY; GHOSH, 2008);
- A introdução de cálcio no processo de lingotamento contínuo pode aumentar tanto a probabilidade quanto a taxa de erosão do refratário, dependendo das condições específicas de operação e da composição dos materiais envolvidos (SZEKERES, 1992);
- Aumento dos custos de produção;
- O tratamento com cálcio não funcionará para aços com alto teor de enxofre, uma vez que o cálcio reage preferencialmente com o enxofre ao invés do alumínio (LARSEN; FRUEHAN, 1990), no entanto, McPherson e McLean (1992) observaram que o tratamento com cálcio ainda pode ser bem-sucedido se o enxofre for adicionado após o tratamento com cálcio.

#### 2.2.4.3 Tipos de Bocais

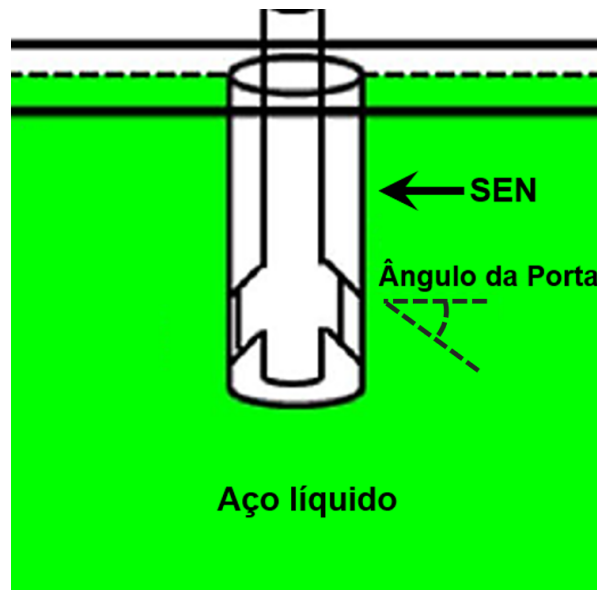
Outro mecanismo que contribui para a redução da ocorrência de *clogging* é o uso de bocais com superfícies mais lisas, com menor condutividade térmica e geometria que proporcione menor ângulo de contato do aço (SIDMAR et al., 1988; FUKUDA; UESHIMA; MIZOGUCHI, 1992; MCPHERSON; MCLEAN, 1992; SZEKERES, 1992; RACKERS; THOMAS, 1995).

O escoamento do aço líquido é importante para a qualidade no lingotamento contínuo e é capaz de influenciar no crescimento da casca de solidificação do molde. O fluxo de aço dentro dos moldes é determinado principalmente pela estrutura e profundidade submersa do SEN, sob a ação da geometria de molde e velocidade de lingotamento (THOMAS; ZHANG, 2001).

Através de experimentos e simulações numéricas, Lee e Yi (2020) propõem critérios de estabilidade de fluxo como funções do ângulo da porta do bocal do SEN (Figura 3) e o correlacionam com os padrões e velocidade de fluxo.

Para ângulos de abertura do bocal do SEN na faixa de 0° a 15°, as velocidades de fluxo na zona de circulação mostraram comportamento instável e fortes oscilações. Por

Figura 3 – Ângulo da porta do bocal da SEN.



Fonte: Adaptada de Lee e Yi (2020, p. 2).

outro lado, quando o ângulo de abertura da porta tornou-se superior a  $20^\circ$ , a velocidade do fluxo diminuiu e a magnitude da oscilação foi reduzida drasticamente, apresentando então comportamento estável (LEE; YI, 2020).

A escolha da ação corretiva depende da classe do aço e da causa exata do problema específico de *clogging* que está sendo considerado. Neste sentido, o primeiro passo é identificar a obstrução, monitorando parâmetros importantes durante o vazamento e através de exame visual, microscópico e químico do material entupido (RACKERS, 1995; THOMAS; BAI, 2001).

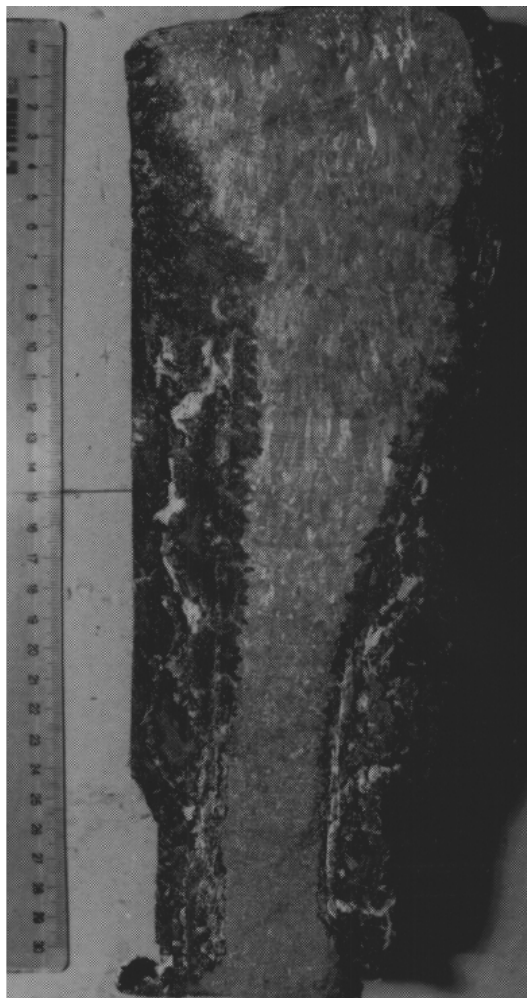
### 2.2.5 Exemplo de Análise de Segmento Obstruído

Muitas informações podem ser extraídas por meio de uma cuidadosa inspeção visual e análise do próprio entupimento. Um exemplo de *clogging* no bocal do distribuidor acima da válvula gaveta é ilustrado na Figura 4. Obstruções como esta são particularmente impactantes porque exigem a troca do distribuidor (RACKERS, 1995; THOMAS; BAI, 2001).

Conforme é observado na Figura 4, as regiões mais escuras nas laterais geralmente são compostas por alumina misturada ao aço solidificado. Uma análise da estrutura da alumina da imagem revelou que provavelmente ela tenha se originado de produtos de desoxidação que se agruparam e sintetizaram. Algumas regiões severamente entupidadas não contêm aço e parecem resultar da escória que ficou presa no distribuidor durante as trocas de painéis (THOMAS; BAI, 2001).

Ainda em análise da Figura 4, a forma do entupimento parece seguir o contorno do

Figura 4 – Amostra de *clogging* no bocal do distribuidor acima da válvula gaveta.



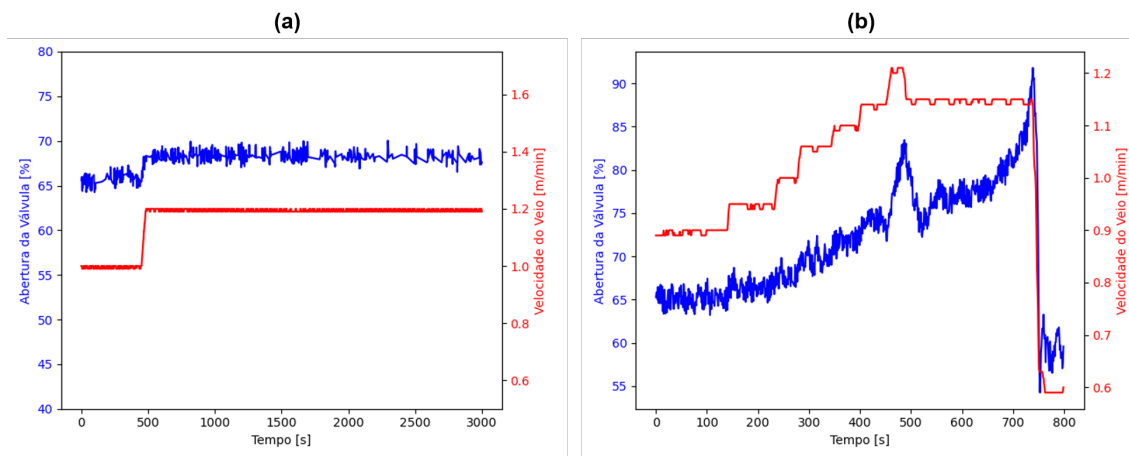
Fonte: [Thomas e Bai \(2001, p. 6\)](#).

fluxo de aço através do bocal. A forma no topo pode ter sido alterada durante o vazamento pela atuação de uma varada cega<sup>4</sup> mal sucedida, por exemplo. A forma assimétrica da porção inferior sugere que o material de obstrução tende a se acumular nas regiões mais estagnadas, logo acima da abertura da válvula gaveta. Também é possível sugerir que pequenas perdas térmicas do aço estagnado dentro dessas regiões podem permitir a solidificação, fortalecendo a ação do entupimento ([RACKERS, 1995](#); [THOMAS; BAI, 2001](#)).

Em resumo, esse tipo de análise é vital para se identificar as causas do *clogging*, sobretudo, para que ações corretivas apropriadas possam ser tomadas na planta. No entanto, também é importante prever a ocorrência do *clogging* durante o lingotamento, para que possíveis problemas de qualidade no produto possam ser minimizados ou pelo menos antecipados ([THOMAS; BAI, 2001](#)).

<sup>4</sup> Procedimento em que uma pessoa utiliza-se de uma vara de aço para bater nas inclusões presas, no bocal da válvula, resultando na sua liberação por força mecânica aplicada pela vara nas inclusões.

Figura 5 – Exemplo do comportamento da velocidade do veio e da abertura da válvula gaveta em um processo (a) sem e (b) com a ocorrência de *clogging*.



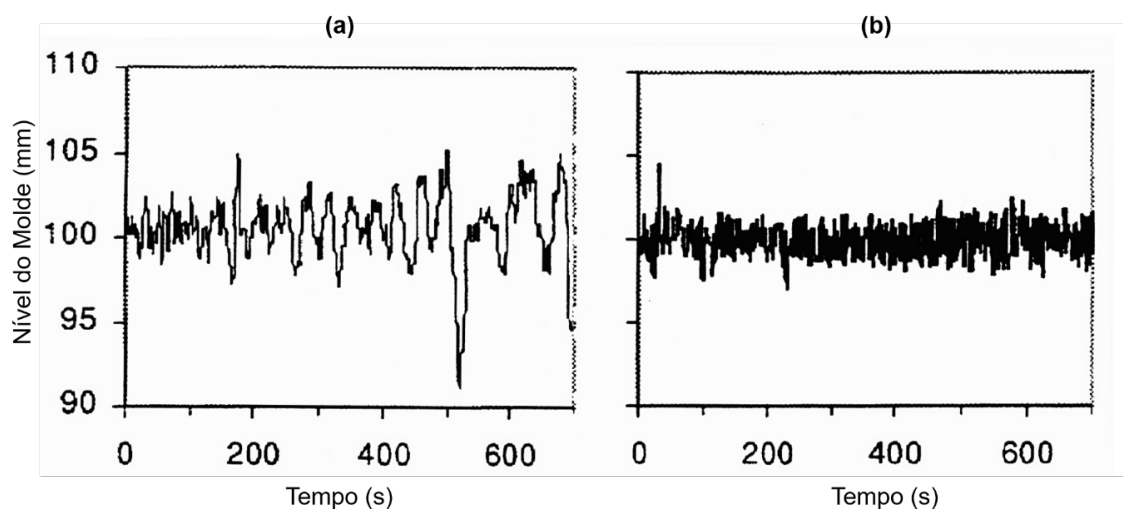
Fonte: Elaborada pela autora.

## 2.2.6 Detecção do *Clogging*

Os autores [Ikäheimonen et al. \(2002\)](#) indicam que existem algumas variáveis que permitem determinar a iminência de *clogging*. Segundo eles, as experiências mostraram que a velocidade do veio e a abertura da válvula gaveta podem dar a primeira indicação de obstrução. A [Figura 5](#) ilustra um exemplo do comportamento das variações da velocidade do veio e da abertura da válvula. A diferença entre os comportamentos das variáveis sem e com a presença do *clogging*, [Figura 5\(a\)](#) e [Figura 5\(b\)](#), respectivamente, é bastante perceptível.

Na [Figura 5\(a\)](#) o processo segue normalmente e mesmo com o aumento da velocidade, aproximadamente na amostra 500, não há alteração significativa da abertura da válvula. Já na [Figura 5\(b\)](#), o aumento gradual da abertura da válvula em conjunto com o aumento da velocidade do veio, entre os instantes 100 e 450, indicam uma ação de controle do processo frente ao acúmulo de material característico da formação do *clogging* na seção da válvula. A partir do instante 500, no entanto, a abertura da válvula volta a aumentar, mas desta vez, a velocidade do veio se mantém constante, o que indica que a obstrução da válvula progride. Próximo ao instante 700, ocorre a obstrução completa da válvula, levando a uma redução abrupta na velocidade e interrompendo o processo por volta do instante 750.

[Ikäheimonen et al. \(2002\)](#), entretanto, ressaltam que o comportamento da velocidade do veio e da abertura da válvula gaveta não é capaz de informar com exatidão quanto tempo o processo pode continuar e quando o bocal deve ser trocado, sendo necessária a utilização de outras variáveis do processo para uma análise mais apurada. Neste sentido, a literatura sugere que o *clogging* pode ser melhor detectado durante o processo de lingotamento contínuo através do monitoramento simultâneo de vários parâmetros em tempo real, tais

Figura 6 – Flutuações no nível do molde (a) com *clogging* e (b) sem *clogging*.

Fonte: Adaptada de [Thomas e Bai \(2001\)](#).

como a temperatura do aço, a contrapressão de argônio, a captação de nitrogênio e as flutuações no nível do molde, além da posição do controle do fluxo (abertura da válvula) em relação à velocidade de vazamento ([RACKERS, 1995](#); [THOMAS; BAI, 2001](#); [YANG; OLIA; THOMAS, 2021](#)).

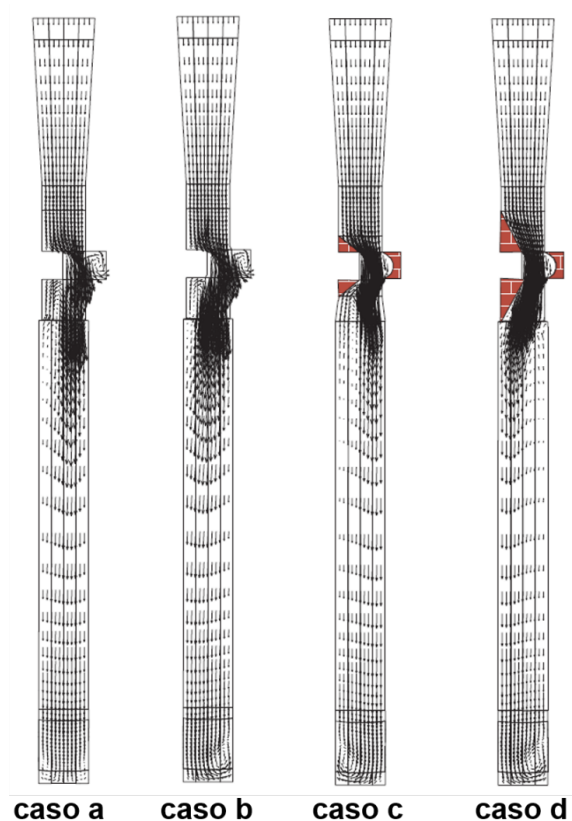
A contrapressão de argônio é um excelente indicador precoce de possíveis problemas de aspiração de ar. Quedas abruptas ou súbitas na contrapressão indicam que o argônio não está fluindo corretamente e pode haver problemas como rachaduras ou vazamento no refratário, que permitem a aspiração do ar. Um aumento na contrapressão pode indicar entupimento nos poros dos bocais, ocasionando maior resistência à injeção de argônio, o que caracteriza formação de inclusões na superfície do bocal ([THOMAS; BAI, 2001](#); [YANG; OLIA; THOMAS, 2021](#)).

[Thomas e Bai \(2001\)](#) destacam que aumentos no teor de nitrogênio entre o aço e o molde indicam problemas de reoxidação no processo. Outro indicativo do avanço na formação do *clogging* é o aumento nas flutuações no nível do aço no molde, conforme pode ser observado na [Figura 6](#).

O nível de aço no molde é uma das variáveis mais importantes em processo de lingotamento contínuo, podendo afetar na qualidade do aço produzido em relação às propriedades mecânicas e metalúrgicas e, consequentemente, afetando no seu valor comercial. Alguns autores sugerem que o nível de aço no molde deve ser mantido dentro de uma faixa específica de variação de cerca de 10 [mm]. Se ultrapassado, a qualidade pode ser prejudicada ([SALLES; ALMEIDA, 2016](#); [PEREIRA et al., 2021](#)).

Pequenas mudanças na forma da passagem do fluxo causada por *clogging* ou erosão do bocal da válvula submersa provocam alterações consideráveis na pressão de descida

Figura 7 – Efeitos do *clogging* inicial observados pelo padrão de fluxo de quatro bocais semelhantes, diferenciados pela geometria próxima à válvula gaveta, em que o arredondamento dos cantos da cerâmica é representado pelo desenho avermelhado, enquanto que as linhas de fluxo são representadas pelas linhas pretas.



Fonte: Adaptada de [Thomas e Bai \(2001\)](#).

([THOMAS; BAI, 2001](#); [YANG; OLIA; THOMAS, 2021](#); [CHANG; CHEN; REN, 2021](#)). Para exemplificar, a [Figura 7](#) exibe vetores de fluxo através de quatro bocais semelhantes, diferenciados pela geometria próxima à válvula gaveta. Em cada caso, zonas de recirculação turbulentas com alta concentração de gás são encontradas logo acima e abaixo da porta e em sua cavidade.

As zonas geram uma grande queda de pressão, exigindo uma abertura de válvula de 52%, sem obstruções (caso a). Uma leve erosão provocada pelo aço que flui pode arredondar os cantos cerâmicos (caso b), diminuindo a queda de pressão e exigindo que a válvula se feche suavemente. Como alternativa, a obstrução tende a se acumular inicialmente nas regiões de recirculação (caso c). Seguindo esta lógica, o *clogging* inicial pode diminuir a queda de pressão total no bocal, conseqüentemente, a válvula deve ser totalmente fechada para contornar isso. O avanço do *clogging* produz quedas de pressão ainda menores (caso d).

Neste sentido, devido às significativas variações na resistência do fluxo em resposta

a pequenas alterações no formato da obstrução, o controle do fluxo através da abertura da válvula gaveta nem sempre apresenta uma resposta adequada. Essas variações resultam em flutuações na taxa de fluxo, que por sua vez causam oscilações no nível do molde (THOMAS; BAI, 2001; CHANG; CHEN; REN, 2021). Em um estudo mais recente, Chang, Chen e Ren (2021) comprovaram essas constatações por meio da análise de imagens térmicas na superfície do molde de aço ao longo do processo de lingotamento contínuo. Salles e Almeida (2016) e Pereira et al. (2021) também elucidam a importância do monitoramento do nível do molde como forma de garantir a qualidade do produto final.

É interessante destacar que embora seja possível observar uma potencial ocorrência do *clogging* inicial através do monitoramento das flutuações no nível do molde, essas flutuações, por sua vez, são afetadas por diferentes distúrbios, portanto, a análise isolada desta variável não é totalmente viável para mensurar de maneira fidedigna a ocorrência inicial do *clogging* (THOMAS; BAI, 2001).

Com a elevação do acúmulo de alumina, a obstrução promove uma maior resistência ao fluxo, assim é necessário promover um aumento da abertura da válvula de maneira a manter a vazão constante do aço líquido através do bocal. Esse evento pode ser melhor explorado para inferir a severidade do entupimento em tempo real, comparando a taxa de fluxo medida com a taxa de fluxo prevista para a determinada posição da válvula gaveta na ausência de *clogging*.

Ainda, de acordo com Thomas e Bai (2001), a fração do fluxo ou “fator de *clogging*” pode ser obtido empiricamente ou com o auxílio de modelos computacionais, levando-se em conta a relação entre a velocidade de vazamento, a vazão da injeção de argônio, a posição de abertura do portão e o nível no distribuidor para a geometria especificada na ausência de *clogging*.

Em conformidade com as constatações de Thomas e Bai (2001), a taxa de fluxo de aço é especialmente sensível as alterações na abertura das válvulas entre 50% e 100%. A injeção de argônio no bocal do distribuidor superior (UN) tende a diminuir a velocidade de lingotamento, a menos que a abertura da válvula aumente para compensar. Vale ressaltar, no entanto, que a abertura da válvula é menos sensível à ocorrência do *clogging* quando o diâmetro do furo do bocal é grande e a abertura da válvula é pequena. Desta forma, o início do *clogging* é de difícil detecção, apenas a partir de alterações na posição da válvula.

Thomas e Bai (2001) demonstram que, para as condições específicas, o acúmulo de material obstrutor não apresenta efeitos significativos no fluxo de aço até que a abertura linear da válvula ultrapasse 60%. Assim, observando-se apenas a taxa de fluxo de aço no bocal, não é possível perceber a obstrução até quase 50% da seção do bocal estar cheia de material obstrutor, logo, a severidade das obstruções inferida a partir de alterações na taxa de fluxo para uma determinada posição da válvula deve ser usada com muita cautela, com base em cuidadosa modelagem, validação e interpretação conjunta de outros sinais.

## 2.3 Trabalhos Relacionados

Com objetivo de estudar e prever os eventos de *clogging* no lingotamento contínuo, diferentes pesquisas estão sendo desenvolvidas. Em sua totalidade, as pesquisas correlacionam o evento de *clogging* com a abertura da válvula e a velocidade do veio (IKÄHEIMONEN et al., 2002; ROUT et al., 2013; RAJENDRA; RACHAPPA; ANIL, 2016; BARATI et al., 2018; KONG; CANG, 2019; KONG; CHEN; CANG, 2019; WANG et al., 2021).

Rackers e Thomas (1995) analisaram as causas, efeitos e possíveis soluções para os eventos de *clogging*. Foram discutidos os resultados de uma análise unidimensional em estado estacionário da perda de calor do fluxo de aço através da parede da válvula, baseando-se nos princípios físicos do processo. O estudo mostrou que inicialmente o aço pode se solidificar e acumular, ocasionando uma obstrução relativamente pequena. Como efeito da perda de calor do aço através da parede da válvula, as partículas de aço solidificado se depositam na parede e, conseqüentemente, a obstrução aumenta.

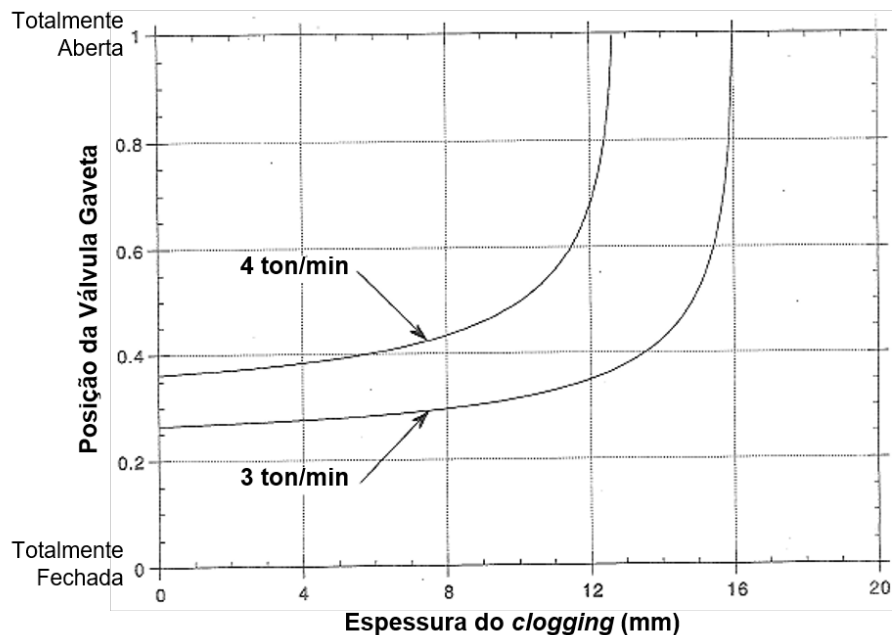
Muitas considerações foram feitas para a composição dos modelos fenomenológicos. A transferência de calor através do entupimento e da parede do bico foi assumida ocorrer apenas por condução, enquanto que a perda de calor para o ambiente foi assumida ocorrer por radiação. As dimensões, propriedades do material e condições de lingotamento foram todas constantes apresentadas em tabelas (RACKERS; THOMAS, 1995).

Ainda assim, Rackers e Thomas (1995) puderam verificar que a condutividade do refratário, a temperatura de solidificação e o grau de limpidez podem afetar diretamente na solidificação do aço, alterando a taxa de deposição do produto de desoxidação. Os experimentos mostraram ainda que este fenômeno ocorrerá em maior grau no início do lingotamento, uma vez que o bocal e os materiais do entorno funcionarão como dissipadores de calor. Deste modo, as investigações experimentais demonstraram que a transferência de calor pode desempenhar um papel importante no processo de *clogging*.

Os pesquisadores também desenvolveram um modelo simplificado com base nas leis da física para o bocal do distribuidor, válvula gaveta e SEN, com o objetivo de quantificar a relação entre a posição da válvula gaveta e o grau de obstrução. Para tanto, os bocais foram modelados como um tubo áspero cujo raio era diminuído em uma quantidade igual à espessura do entupimento (RACKERS; THOMAS, 1995).

A relação entre espessura do *clogging* e a posição da válvula deslizante necessária para manter a taxa de fundição é mostrada na Figura 8. É possível observar que foi necessário um pequeno deslocamento da porta deslizante até que o entupimento atingisse uma espessura crítica e, a partir de então, um pequeno aumento na espessura do entupimento exigiu uma grande variação na posição da porta deslizante (RACKERS; THOMAS, 1995).

A razão subjacente para esse comportamento é que, para uma taxa constante de

Figura 8 – Relação entre espessura do *clogging* e a posição da válvula deslizante.

Fonte: Adaptada de Rackers e Thomas (1995, p. 733).

vazamento, a queda da pressão através de um tubo áspero é inversamente proporcional à quinta potência do raio do tubo. Deste modo, o resultado apresentado revelou que a utilização apenas da posição da válvula gaveta não é um bom indicador da extensão do *clogging* inicial (RACKERS; THOMAS, 1995).

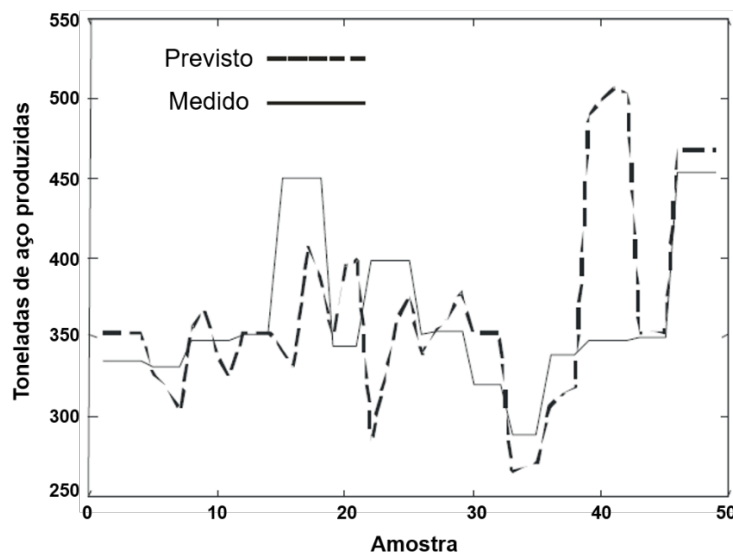
Rackers e Thomas (1995) também constataram situações severas de entupimento considerando a razão entre a alumina depositada e a vazão de alumina. Assumindo que o *clogging* seja composto por produtos de desoxidação, eles observaram que cerca de 16% da alumina que passa pelo bocal fica depositada. Isso indica que, na ausência de contramedidas de *clogging* (como a injeção de argônio, por exemplo), o transporte das inclusões para a parede do bocal é bastante significativo.

No estudo de Ikäheimonen et al. (2002), desenvolvido a partir de dados de uma planta real, os pesquisadores mostraram que a velocidade do veio e a abertura da válvula são os primeiros indicadores do desenvolvimento gradual do *clogging*. Outro indicativo do avanço na formação do *clogging* foi o aumento nas flutuações no nível do aço no molde.

Foi proposta uma análise com 67 variáveis do processo para determinar por quanto tempo o lingotamento pode continuar e quando o bocal deverá ser trocado. Todas as principais temperaturas e composições foram incluídas, a partir dos conversores até o lingotamento, juntamente com outras variáveis do processo e algumas variáveis adicionais que foram calculadas a partir das observações do processo (IKÄHEIMONEN et al., 2002).

Os dados foram coletados a partir de dois veios, contendo 5800 corridas de cada,

Figura 9 – Comparativo dos resultados dos testes desenvolvidos em (IKÄHEIMONEN et al., 2002).



Fonte: Adaptada de Ikäheimonen et al. (2002, p. 146).

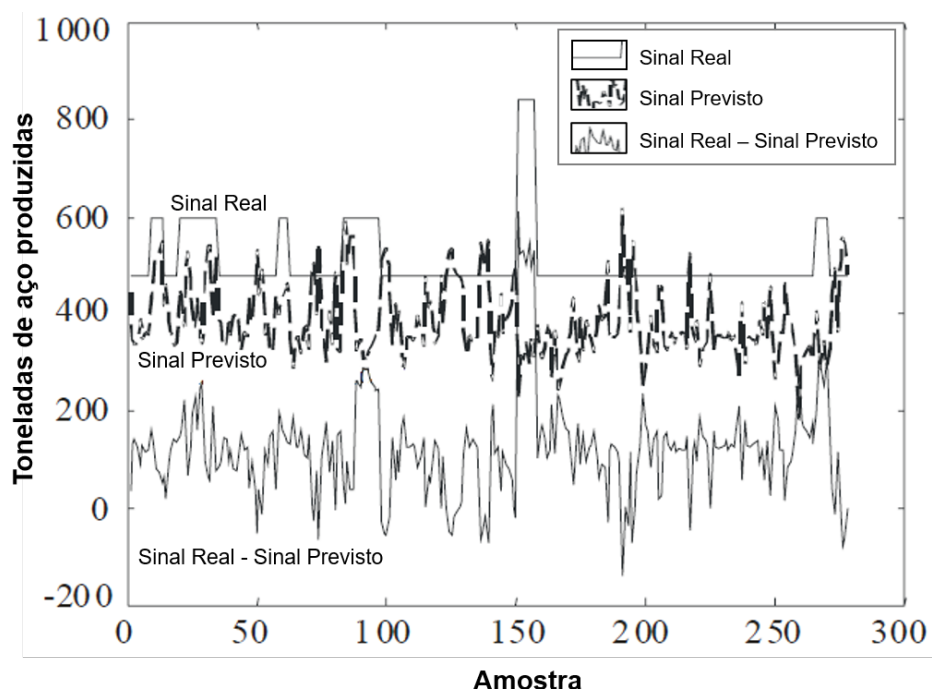
contudo, somente as corridas afetadas pelo alumínio foram utilizadas na modelagem. As corridas nitidamente afetadas pelo silício e contendo sérios distúrbio foram desprezadas sem mais considerações. A análise de correlação foi feita separadamente para duas máquinas e quatro graus de aço, resultando em oito grupos de dados (IKÄHEIMONEN et al., 2002).

As variáveis de entrada do modelo foram selecionadas de acordo com a correlação, de modo que apresentassem correlações acima de 0,2 com a quantidade de aço produzido e que tivesse correlação com as demais variáveis menor que 0,5. Essa análise resultou em 10 variáveis de entrada usadas em uma Rede Neural Artificial (RNA) do tipo *feedforward* com uma camada oculta para prever a continuidade do lingotamento antes de haver a troca do bocal. Os resultados mostraram-se promissores e os melhores modelos previram o número de toneladas fundidas dentro de uma faixa de  $\pm 60$  toneladas em torno do valor correto em mais de 80% dos casos. Um comparativo entre a variável medida e a previsão pode ser visualizado na Figura 9 (IKÄHEIMONEN et al., 2002).

Os autores observaram, todavia, diferenças significativas entre as variáveis selecionadas pelos modelos para cada um dos distribuidores e casos avaliados. Outro ponto destacado é que os modelos foram treinados usando apenas determinados casos de *clogging*, assim, ao testar os modelos com casos bem sucedidos (sem ocorrência de *clogging*), os resultados não forneceram conclusões claras ou impactos significativos nas previsões dos modelos. Conforme ilustra a Figura 10, as quantidades de aço produzido, em toneladas, previstas pelos modelos permaneceram inferiores aos valores reais na maior parte do tempo (IKÄHEIMONEN et al., 2002).

Partindo da premissa de que monitorar a abertura e o fechamento da válvula

Figura 10 – Exemplo dos resultados obtidos no teste da RNA com dados de casos bem-sucedidos (IKÄHEIMONEN et al., 2002).



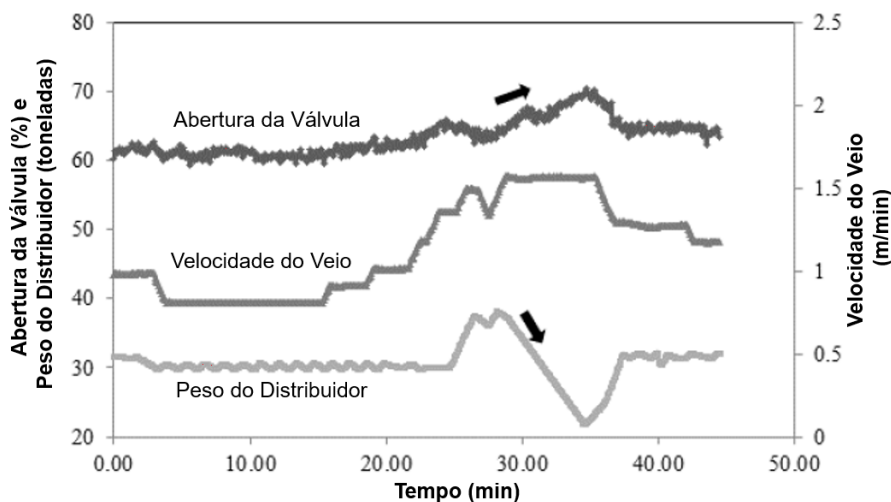
Fonte: Adaptada de Ikäheimonen et al. (2002, p. 147).

durante o lingotamento é a maneira mais comum de prever a quantidade de obstrução dentro do SEN, Rout et al. (2013) propuseram um preditor linear baseado em equações básicas da hidrodinâmica e regressão linear para estimar uma espécie de porcentagem de *clogging* em tempo real durante o lingotamento.

Ao longo da pesquisa, contudo, Rout et al. (2013) constataram que a estimativa com base apenas na posição da válvula poderia fornecer uma falsa indicação de obstrução. A Figura 11 mostra o exemplo de um caso em que houve um aumento na abertura da válvula em paralelo à redução do peso do aço no distribuidor e ao aumento controlado da velocidade do veio, mas que não houve ocorrência de *clogging*. Partindo deste princípio, além da abertura da válvula, os pesquisadores empregaram em sua modelagem outras variáveis do processo, tais como, características físicas da válvula, velocidade do veio e peso de aço no distribuidor.

A abordagem proposta por Rout et al. (2013) visava recalibrar o modelo a cada operação devido ao desgaste da válvula por erosão, alterando o valor de abertura teórica como referência de início de *clogging*. Uma comparação quantitativa foi feita entre o entupimento real e o previsto pelo modelo. Os pesquisadores alegam que uma superestimação de 5% pelo modelo deve-se à deposição de alguma obstrução no distribuidor e na entrada da válvula, onde até mesmo a verificação física era muito difícil. De todo modo, o índice

Figura 11 – Variação da abertura da válvula em relação à velocidade do veio e peso do aço no distribuidor.



Fonte: Adaptada de [Rout et al. \(2013\)](#).

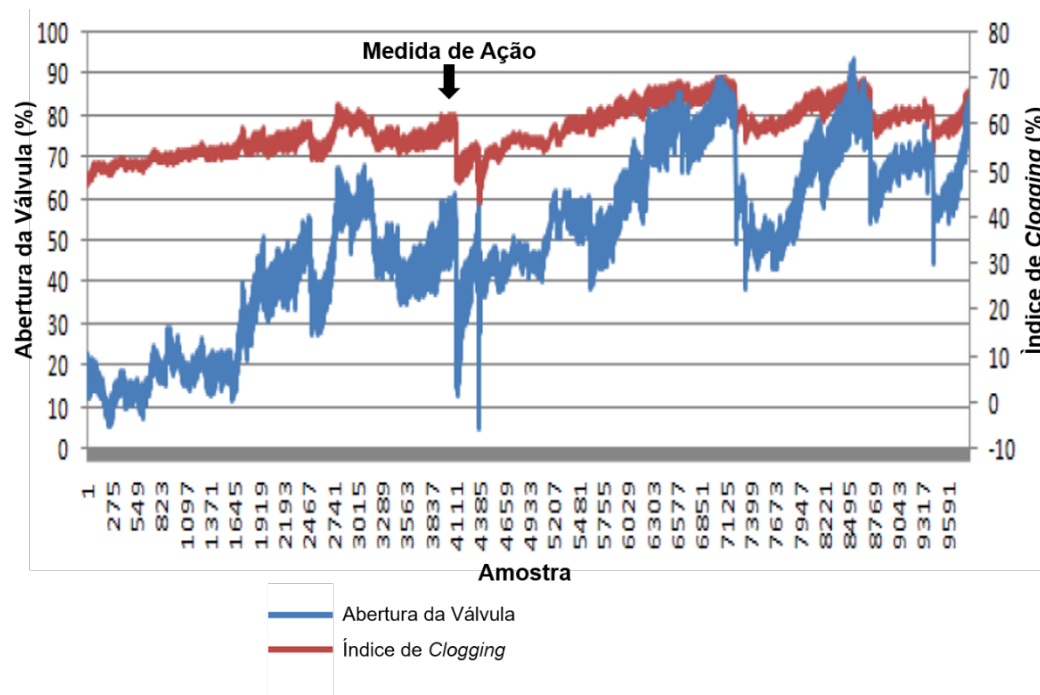
de *clogging*<sup>5</sup> obtido foi útil para fornecer aos operadores os indicativos de necessidade de trocas e manutenções no distribuidor.

[Rajendra, Rachappa e Anil \(2016\)](#) propuseram um método baseado na geometria dos elementos físicos da planta e nas características físico-químicas do aço para determinar a ocorrência dos eventos de obstrução, através de um índice de *clogging* baseado na abertura da válvula em condições normais de operação. O modelo foi testado em uma planta real, provando ser uma ferramenta útil na previsão de alguns tipos de obstruções.

Com a ajuda de um *software*, eles aplicaram dados como a abertura da válvula, a área da seção do SEN e a velocidade do veio, para o cálculo da porcentagem de *clogging* e espessura de entupimento no SEN. Os resultados obtidos pelo estudo são apresentados na [Figura 12](#), em que a linha vermelha indica o índice de *clogging* e a linha azul indica a abertura da válvula, ambos em porcentagem ([RAJENDRA; RACHAPPA; ANIL, 2016](#)).

A primeira observação feita foi a comprovação de que a abertura da válvula e o índice de *clogging* estão linearmente correlacionadas. Após vários testes, [Rajendra, Rachappa e Anil \(2016\)](#) concluíram que para um índice de *clogging* em torno de 40%, ações preventivas poderiam ser tomadas pelos operadores, afim de liberar o depósito de entupimento. Essa liberação pode ser vista com a queda no índice de *clogging* na [Figura 12](#), no entanto, os pesquisadores ressaltam que, se após as medidas de controle, o índice de *clogging* não diminuir, é possível que a obstrução já tenha tomado proporções maiores, sendo então necessária a troca do SEN e/ou do conjunto do distribuidor.

<sup>5</sup> O índice de *clogging*, ou índice de obstrução, se refere a uma métrica que avalia a ocorrência e a gravidade de obstruções ou bloqueios no sistema. Geralmente, quanto maior o valor do índice de *clogging*, mais severas são as obstruções.

Figura 12 – Aplicação do índice de *clogging* aos dados da planta.

Fonte: Adaptada de [Rajendra, Rachappa e Anil \(2016, p. 13\)](#).

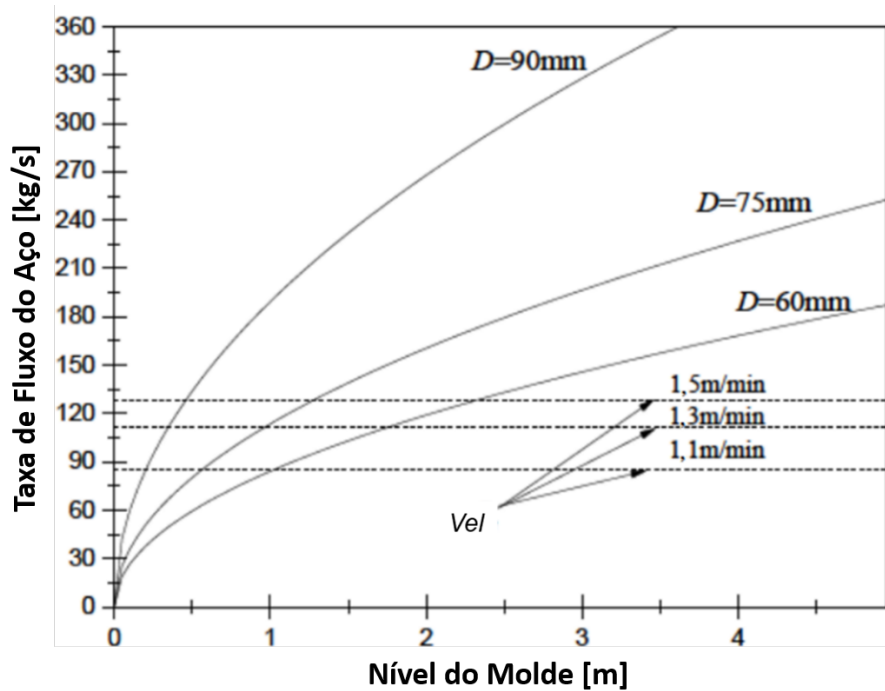
Um estudo análogo foi desenvolvido por [Kong, Chen e Cang \(2019\)](#). Eles desenvolveram um cálculo teórico das vazões que poderiam afetar a obstrução da válvula durante um processo de lingotamento contínuo do aço silício<sup>6</sup>, analisando as inclusões causadoras. Para tanto, foram analisados os tipos, formas, tamanhos e a distribuição das inclusões nas amostras de aço, que poderiam levar ao entupimento do bocal. O estudo calculou as condições normais de abertura da válvula em função da velocidade do veio e do nível do molde e mostrou que com a diminuição do nível do molde, a taxa de fluxo do aço diminui, conforme mostrado na [Figura 13](#), assim, se o nível cair de 4 [m] para 1,5 [m], haveria uma redução de cerca de 50 [kg/s] na taxa de fluxo de aço, para um diâmetro do bocal de 60 [mm].

Também é possível verificar que quanto maior for o diâmetro do bocal ( $D$ ), maior será a taxa de fluxo e a velocidade do veio. Deste modo, para um bocal com diâmetro de 90 [mm], mesmo que o nível seja inferior a 0,5 [m], a velocidade do veio poderia atingir a meta de 1,1 [m/min]. Em contrapartida, para bocais de diâmetro de 60 [mm], mesmo que o nível seja dobrado, isto é, 1 [m], a velocidade do veio não atingiria o objetivo, chegando a 0,98 [m/min] ([KONG; CHEN; CANG, 2019](#)).

Assim, com base nos cálculos desenvolvidos, a relação entre a abertura da válvula, a velocidade do veio, o diâmetro do bocal ( $D$ ) e o nível do molde ( $Niv$ ) pode ser mostrada na [Figura 14](#). Por exemplo, para um bocal de 90 [mm] de diâmetro e nível de 4 [m] (início

<sup>6</sup> Aço usado na fabricação de núcleos de transformadores.

Figura 13 – Relação entre a taxa de fluxo, velocidade do veio ( $Vel$ ), nível do molde e diâmetro do bocal ( $D$ ).



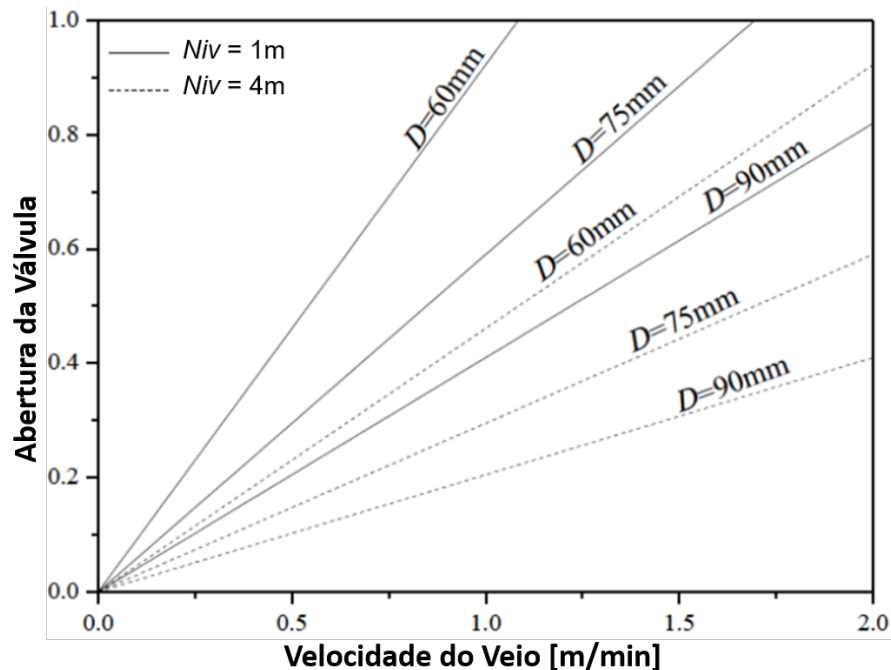
Fonte: Adaptada de Kong, Chen e Cang (2019, p. 815).

do lingotamento), uma abertura da válvula em torno de 0,22 (22%) poderá fazer com que velocidade do veio atinja 1,1 [m/min]. À medida que o nível diminui para 1 [m], a abertura da válvula deve aumentar para aproximadamente 0,45 (45%) para atingir a velocidade do veio desejada, ou seja, cerca de duas vezes a abertura inicial (KONG; CHEN; CANG, 2019).

Em análise da Figura 14, supondo que uma obstrução do bocal diminua seu diâmetro para 60 [mm] e considerando um nível do molde de 1 [m], observa-se que a válvula deve atingir a sua abertura máxima (100%) para atender à demanda de velocidade de 1,1 [m/min], no entanto, se o nível do molde diminuir ainda mais, então, mesmo que a válvula atinja uma abertura de 100%, a velocidade desejada de 1,1 [m/min] não seria atingida. Ao mesmo tempo, para manter um nível estável, a velocidade do veio deve diminuir (KONG; CHEN; CANG, 2019).

Kong, Chen e Cang (2019) validaram os efeitos de tratamentos específicos do aço na solução das obstruções. Eles também constataram que, para um bocal de diâmetro fixo, é possível encontrar o percentual de abertura da válvula apropriado para manter a velocidade do veio dentro de uma faixa desejada (Figura 14). Tomando como referência um bocal com diâmetro de 75 [mm], por exemplo, considerando a intenção de manter a velocidade do vazamento entre 1,1 [m/min] e 1,5 [m/min] em um nível de 4 [m], verifica-se que uma abertura aproximadamente na faixa de 32,46% a 44,27% garantiria a presença

Figura 14 – Relação entre a abertura da válvula, velocidade do veio, diâmetro do bocal ( $D$ ) e nível do molde ( $Niv$ ).



Fonte: Adaptada de [Kong, Chen e Cang \(2019, p. 816\)](#).

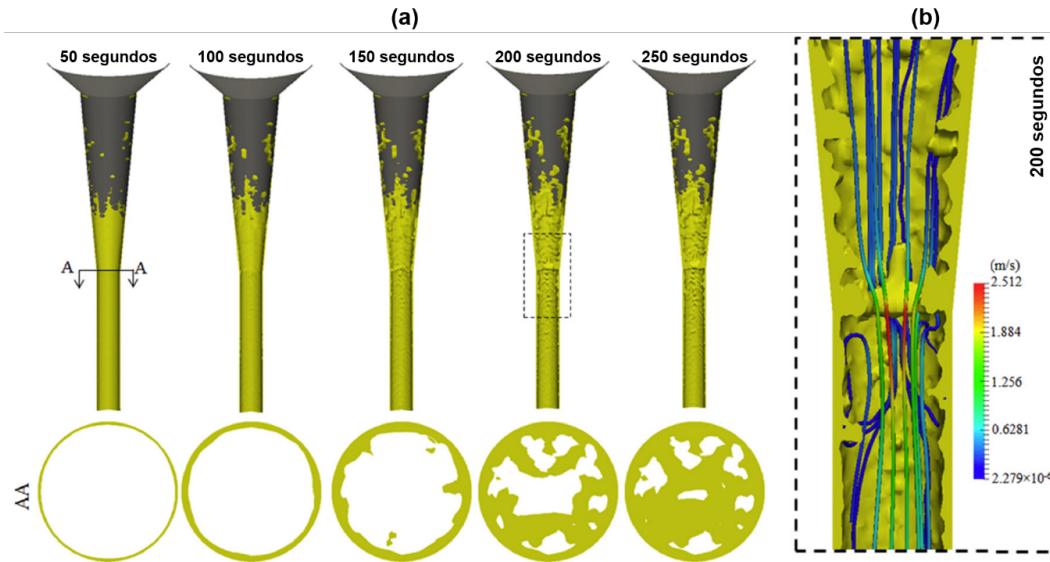
desejada de aço líquido no distribuidor. Neste contexto, eles verificaram quais condições de abertura da válvula superiores às teóricas calculadas, para uma dada velocidade e nível do molde, marca o início do fenômeno de *clogging*.

Através da reprodução de experimentos laboratoriais, [Barati et al. \(2018\)](#) apresentaram o *clogging* como um processo estocástico auto acelerado, considerando um modelo transitório bidirecional entre o crescimento da obstrução e o fluxo de aço, baseado em uma abordagem Euleriana-Lagrangiana ([BATCHELOR, 1973](#)).

A partir de então, [Barati et al. \(2018\)](#) estudaram a modificação na velocidade do fluído pela deposição sólida ao longo do tempo e estabeleceram três intervalos da formação de *clogging*:

- Antes do *clogging*: etapa em que o fluxo de metal flui suavemente ao longo do bocal;
- Início do *clogging*: momento em que uma turbulência começa a aparecer na parte superior do bocal, ficando mais forte ao longo de sua parede com a seção do bocal, sendo gradualmente estrangulada;
- Após o início do *clogging*: fase em que a característica do padrão de fluxo é marcada pelo aumento dramático da turbulência com o entupimento incremental. Neste momento, a velocidade máxima do fluxo não altera significativamente, podendo cair um pouco. Por outro lado, a energia cinética da turbulência aumenta em 5 ordens

Figura 15 – Evolução do *clogging* no bocal. (a) Vista da região de entupimento da seção vertical e uma seção transversal A-A do bico; (b) Vista ampliada das linhas de fluxo e uma seção parcialmente obstruída a 200 segundos. A seção entupida é mostrada em amarelo. A magnitude da velocidade do fluxo é mostrada em uma escala de cores ao longo das linhas de fluxo.



Fonte: Adaptada de Barati et al. (2018, p. 190).

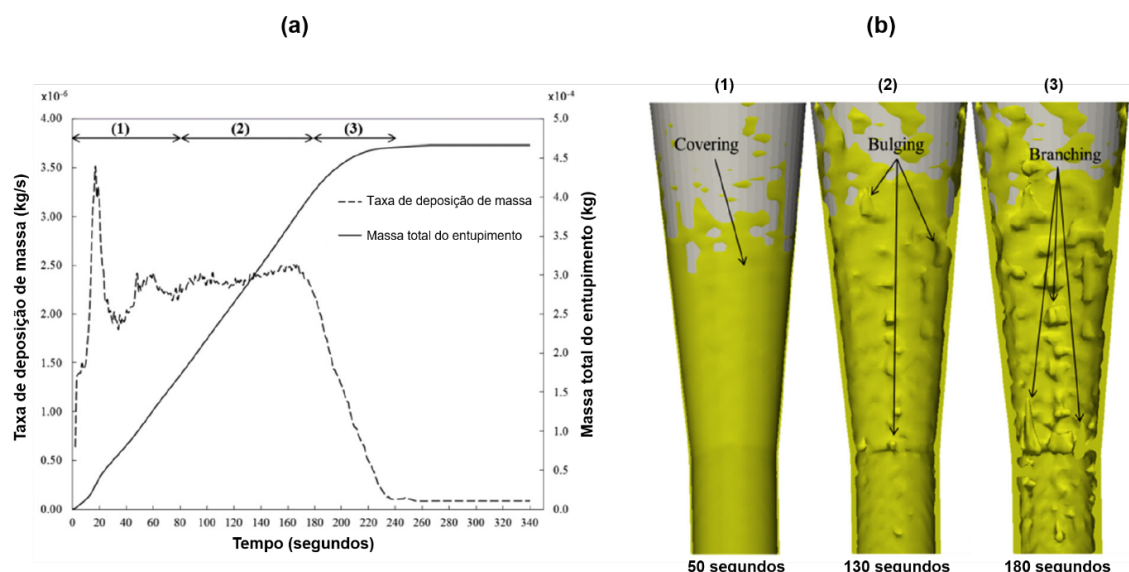
de grandeza na parte inferior do bocal devido à frente irregular do entupimento e à redução da seção transversal da passagem do fluxo.

A Figura 15 mostra a evolução do *clogging* no bocal em ambiente de laboratório. Aos 50 segundos, a seção mais estreita e uma parte da seção intermediária do bocal são cobertas por uma camada lisa de entupimento. Com a contínua deposição de partículas na parede do bocal, o *clogging* começa a crescer apresentando partículas visíveis já entre 100 a 150 segundos. Entre 200 e 250 segundos, a deposição de partículas ocorre ainda mais rápida, desenvolvendo uma estrutura ramificada (Figura 15(b)) que cresce continuamente até que a seção transversal da SEN seja completamente bloqueada (BARATI et al., 2018).

Na mesma pesquisa, Barati et al. (2018) classificaram a formação do *clogging* em 3 fases: cobertura inicial da parede do bocal pelas partículas depositadas (*Covering*), desenvolvimento de uma frente de entupimento abaulada (*Bulging*) e desenvolvimento da estrutura de entupimento ramificada (*Branching*). A Figura 16 ilustra como se desenvolve a deposição de material durante o *clogging* nas 3 fases de sua formação.

Verifica-se na Figura 16(a) que a curva da massa total de entupimento (isto é, a integral de massa das partículas depositadas na parede do bocal e na frente do entupimento) aparece relativamente suavizada, enquanto que a taxa de deposição de massa varia significativamente. Essa variação coincide com as fases de *clogging*, conforme

Figura 16 – (a) A massa total do entupimento calculada numericamente e a taxa de deposição de massa; (b) Três períodos de evolução do entupimento: (1) *Covering*, (2) *Bulging* e (3) *Branching*.



Fonte: Adaptada de Barati et al. (2018, p. 191).

mostrado na Figura 16(b).

Na fase de *Covering*, que dura em torno de 80 segundos, as partículas se depositam na parede do bocal iniciando o processo de *clogging*. Na fase de *Bulging*, que inicia em torno de 80 segundos e vai até em torno de 180 segundos, há o desenvolvimento de uma frente abaulada de depósito de partículas sólidas. Por fim, na etapa de *Branching*, que se inicia em torno dos 180 segundos e se desenvolve até em torno dos 240 segundos, ocorre o desenvolvimento da estrutura de entupimento ramificada que irá culminar com a obstrução total do bocal (BARATI et al., 2018).

Assim, conforme mostram as curvas do gráfico da Figura 16(a), constata-se que quando o *clogging* tem início, o tempo para obstrução total do bocal leva em torno de 4 minutos, com uma deposição rápida de material. Na realidade, a percepção efetiva da ocorrência de *clogging* nas variáveis do processo ocorre quando a obstrução está próxima a 50% da área efetiva do bocal, conforme apontado por Thomas e Bai (2001). Esse ponto de referência é alcançado aproximadamente 150 segundos após o início da formação do *clogging*, como indicado por (BARATI et al., 2018). Relembrando que os tempos obtidos são em experimento realizado em planta montada em ambiente de laboratório, portanto, de menores proporções que uma planta industrial do processo.

Em sequência, Barati et al. (2019) aplicam um algoritmo baseado em modelagem numérica para validar o modelo de *clogging* apresentado no trabalho anterior (BARATI et al., 2018), no que se refere à precisão e eficiência de cálculo. Devido ao grande número real de partículas no SEN e a inviabilidade em termos de capacidade computacional, um fator

artificial foi introduzido no modelo. A partir do rastreamento do movimento de uma partícula individual no aço líquido e contabilizando suas interações com o fluxo turbulento, este fator foi aplicado para então representar as partículas à medida que se depositam na parede do SEN, no entanto, a influência da introdução deste fator na precisão do cálculo não foi discutida.

Embora os mecanismos relacionados às propriedades físico-químicas, bem como à termodinâmica do sistema foram simplificados pelo modelo numérico, ele foi capaz de representar de forma eficiente o acúmulo de partículas e a formação do *clogging* obtidos através do experimento físico (BARATI et al., 2019).

Na publicação de Cemernek et al. (2021), é realizado um levantamento do emprego de técnicas de aprendizado de máquina na análise do processo de lingotamento contínuo. Dentre os temas abordados, os autores apresentam o trabalho desenvolvido por Vannucci et al. (2010), em que foram testadas várias abordagens para a identificação de padrões menos frequentes em conjuntos desbalanceados, a fim de desenvolver um classificador que atendesse aos requisitos industriais. Dentre as abordagens testadas, aquelas mais promissoras foram combinadas em uma estrutura baseada em redes neurais capaz de trabalhar com diferentes taxas de amostragem. O modelo proporcionou uma taxa de detecção de 82% para eventos frequentes e de 77% para eventos raros, com uma taxa de falsos positivos de 17%.

Posteriormente, Vannucci e Colla (2011) apresentaram um novo método denominado *Labeled Som Classification Unbalanced Sets* (LASCUS), que permitia a classificação de eventos raros, tais como o *clogging*. Neste método, o mapa auto-organizável (*Self-Organized Map* – SOM) agrupava o espaço de entrada em regiões associadas aos eventos frequentes e não frequentes, enquanto que um sistema de inferência *Fuzzy* (*Fuzzy Inference System* – FIS) classificava essas regiões com base na distribuição dos dados.

O conjunto de dados aplicado possuía cerca de 3000 amostras que descrevem as características físico-químicas do processo por intermédio de 100 variáveis. Com base nas considerações dos operadores da planta e através de uma análise de correlação linear, os autores selecionaram variáveis relacionadas às temperaturas, duração das fases de agitação do aço, bem como percentual dos principais elementos químicos do aço, tais como enxofre, alumínio e cálcio. A metodologia resultou em um *Recall* de 76,9%, Acurácia de 80,2%, com uma taxa de falsos positivos de 19,5% (VANNUCCI; COLLA, 2011).

Dando sequência às pesquisas, Vannucci, Colla e Cateni (2015) desenvolveram um método híbrido para prever a taxa de área obstruída do SEN, através de modelos de regressão. Essa modelagem era baseada em um “aluno forte” (*strong learner*), treinado utilizando todos os dados de treinamento disponíveis, e um conjunto de “alunos fracos” (*weak learners*) especializados, treinados em regiões específicas do espaço de entrada determinadas pelo SOM. Assim, quando um novo padrão era alimentado para o modelo,

o “aluno forte” e o “aluno fraco” associado operam alternativamente com base em sua confiança estimada para lidar com o padrão de entrada específico.

Essa abordagem foi projetada para explorar, para cada padrão de entrada, o aluno com melhor desempenho, superando alguns dos pontos críticos encontrados nos métodos de *ensemble* padrão. Foi verificada uma melhoria no erro médio absoluto (*Mean Absolute Error* - MAE) de 75% comparado aos resultados de uma implementação empregando apenas RNA.

No artigo apresentado por [Pellegrini et al. \(2019\)](#), os autores utilizam um *framework* que implementa um modelo preditivo de estimação *on-line* para a probabilidade de *clogging*. O modelo utilizou dados de 21.927 corridas, com 50 variáveis do processo, coletadas antes do início do lingotamento e em diversas etapas do refino secundário. Etapas preliminares, como preparação de dados, limpeza e análise das variáveis envolvidas, tiveram o apoio de especialistas do processo.

As variáveis foram agrupadas de acordo com as diferentes áreas e etapas de processo, análise química, adições de materiais, amostras de temperatura, processamento e tempos de transporte da panela. Para analisar os dados e dar uma indicação em tempo real da condição de obstrução em curso, um algoritmo de identificação foi desenvolvido com base nos sinais do processo da automação de duas máquinas de lingotamento contínuo. O índice de *clogging* estimado foi então usado como uma variável de referência em um problema de classificação de aprendizagem supervisionada baseado em *gradient boosting* ([PELLEGRINI et al., 2019](#)).

Definindo a ocorrência de *clogging* para uma corrida como a detecção de um nível de obstrução superior a um limiar predefinido, verificou-se que o *clogging* pode ser classificado como um evento raro, uma vez que ocorreu em cerca de 11,9% das corridas analisadas, com algumas diferenças entre as duas máquinas de lingotamento. Uma evidência observada na planta estudada foi a alta incidência de *clogging* em aços resulfurados, comprovando que a probabilidade de eventos de *clogging* depende, dentre outros fatores, da química do aço ([PELLEGRINI et al., 2019](#)).

Em termos de corridas identificadas como potencialmente sujeitas à ocorrência de *clogging*, o desempenho do modelo mostrou-se satisfatório, apresentando uma Área Abaixo da Curva ROC (AUC) igual a 0,8, todavia, quando testado sem restrições entre as corridas, o modelo apresentou acurácia de 75% conseguindo prever o *clogging* com uma precisão de apenas 62% ([PELLEGRINI et al., 2019](#)).

[Wang et al. \(2021\)](#) aplicaram três redes LSTM para preverem três tipos de índice de *clogging*, exclusivamente nas classes de aço tratadas com cálcio, com base nos parâmetros de operação de uma planta siderúrgica. O primeiro índice foi o mesmo aplicado por [Rajendra, Rachappa e Anil \(2016\)](#), o qual levaram em consideração a posição da válvula. Já os

cálculos dos segundo e terceiro índices foram baseados no cálculo desenvolvido por Girase, Basu e Choudhary (2007), que compararam o fluxo máximo de aço em uma posição específica da válvula, levando-se em conta a velocidade do veio, bem como a geometria e o peso do distribuidor.

Um total de 323.696 amostras tomadas a cada 3 segundos foram aplicadas aos modelos LSTM, utilizando uma janela temporal de 3 minutos e um horizonte de previsão de 36 segundos. Todas as LSTM compartilhavam dos mesmos hiperparâmetros contendo, em cada uma delas, duas camadas LSTM com 64 unidades e uma camada de saída com uma única unidade. Dentre os resultados obtidos, a rede responsável pela previsão do segundo índice de *clogging* foi a que apresentou o melhor desempenho, com  $R^2$  de 0,971. Como as redes foram treinadas com os dados de uma única classe de aço, não foi possível realizar testes sobre outros pontos de operação do sistema, tampouco quantificar a capacidade de generalização das redes para a previsão dos índices de *clogging* em outros tipos de aço (WANG et al., 2021).

Yang et al. (2021) desenvolvem um método de monitoramento e detecção do *clogging* em um processo de lingotamento contínuo baseado em gráficos causais. Ao explorar as propriedades de invariância implícitas aos dados, o método proposto remove os efeitos das perturbações do sistema, levando a um melhor desempenho de detecção. Embora o modelo apresente resultados interessantes, o método proposto não aplicou a probabilidade de ocorrência do *clogging* em modelos de classificação, impossibilitando qualquer tipo de comparação.

Na Tabela 1 são evidenciadas algumas das principais características dos trabalhos apresentados nesta seção. É importante destacar que esta tabela inclui um total de seis índices de *clogging*, identificados de (1) a (6), calculados de formas distintas e com base em variáveis específicas, de acordo com a metodologia utilizada em cada estudo. Como nem todos os estudos fornecem uma descrição detalhada dessas variáveis, optou-se por rotulá-los dessa maneira.

De maneira geral, é possível notar que não há um consenso em se definir o índice de *clogging* e poucos são os trabalhos os quais evidenciam a maneira como este índice é calculado. Dadas as informações disponíveis, observa-se que todos esses índices levam em conta fatores físicos da planta, bem como fatores químicos do aço que está sendo produzido, o que limita a aplicabilidade dos modelos propostos. A quantidade e a especificação das variáveis empregadas nos modelos também não ficam claras, o que dificulta a replicabilidade das metodologias.

Os modelos clássicos, de fato, geralmente são baseados em sistemas idealizados e carecem de precisão suficiente em tarefas complexas ou ambientes dinâmicos. Os modelos físicos, por outro lado, consideram apenas conjuntos específicos de variáveis de processo, desconsiderando as informações extraídas de fontes de dados heterogêneas. Nessas circuns-

Tabela 1 – Literaturas de referência: Resumo

| Modelo                                           | Variáveis | Metodologia                                              | Saída                                | Resultados                              |
|--------------------------------------------------|-----------|----------------------------------------------------------|--------------------------------------|-----------------------------------------|
| <a href="#">Ikäheimonen et al. (2002)</a>        | 10        | RNA                                                      | Toneladas de aço                     | 80% $\pm$ 60 toneladas                  |
| <a href="#">Vannucci e Colla (2011)</a>          | 5+        | SOM + FIS                                                | Índice de <i>Clogging</i> (1)        | Acurácia: 80,2% e <i>Recall</i> : 76,9% |
| <a href="#">Rout et al. (2013)</a>               | 4+        | Regressão Linear                                         | Índice de <i>Clogging</i> (2)        | -                                       |
| <a href="#">Vannucci, Colla e Cateni (2015)</a>  | 5+        | Híbrido (SOM + Regressão)                                | Índice de <i>Clogging</i> (1)        | Melhoria de 75% no MAE                  |
| <a href="#">Rajendra, Rachappa e Anil (2016)</a> | 4+        | Modelo matemático (Cálculo teórico)                      | Índice de <i>Clogging</i> (3)        | -                                       |
| <a href="#">Pellegrini et al. (2019)</a>         | 50        | Aprendizagem supervisionada ( <i>Gradient Boosting</i> ) | Índice de <i>Clogging</i> (4)        | Acurácia: 75% e Precisão: 62%           |
| <a href="#">Wang et al. (2021)</a>               | 6+        | LSTM                                                     | Índice de <i>Clogging</i> (3, 5 e 6) | $R^2$ : 0,971                           |

Fonte: Elaborada pela autora.

tâncias, o avanço das inovações em ferramentas computacionais e a disponibilidade de grandes quantidades de dados de processos vêm criando oportunidades promissoras de aplicações de Inteligência Artificial na indústria siderúrgica. A análise preditiva orientada por dados e o aprendizado de máquinas passou então a desempenhar um papel significativo nos avanços tecnológicos das plantas reais, uma vez que esses algoritmos são capazes de extrair sistematicamente informações inerentes ao processo a partir dos dados ([KLANKE et al., 2017](#); [PELLEGRINI et al., 2019](#); [CEMERNEK et al., 2021](#); [WANG et al., 2021](#)).



### 3 Aprendizagem Profunda

Devido à crescente complexidade dos problemas a serem resolvidos e ao grande volume de dados provenientes das mais diversas fontes, nas últimas décadas, tornou-se necessário o desenvolvimento de ferramentas computacionais mais sofisticadas, que operassem de forma mais autônoma, reduzindo cada vez mais a dependência humana. Para tanto, tornou-se primordial o aprimoramento de técnicas capazes de criar, a partir da experiência passada, modelos analíticos para resolver o problema que se deseja tratar. A este processo, capaz de aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana, dá-se o nome de Aprendizado de Máquinas (*Machine Learning*) (CARVALHO et al., 2011; BROWNLEE, 2018; CEMERNEK et al., 2021).

Por sua vez, a Aprendizagem Profunda (*Deep Learning*) é um ramo do *Machine Learning* baseado no conceito de Redes Neurais Artificiais (RNA) que, em níveis hierárquicos, são inspiradas nas redes neurais biológicas no que concerne ao processamento de dados e identificação de padrões para uso na tomada de decisões.

Considerando um conjunto de dados de entrada  $X = \{x_1, x_2, \dots, x_{N_T}\}$  e um conjunto correspondente de dados de saída  $Y = \{y_1, y_2, \dots, y_{N_T}\}$ , em que  $X$  e  $Y$  são vetores de dimensão  $N_T$ , a função que mapeia o espaço de entrada  $X$  para o espaço de saída  $Y$  é representada por  $g(\cdot)$ , conforme Equação 3.1:

$$Y = g(X) \quad (3.1)$$

A função  $g(\cdot)$ , embora desconhecida, desempenha um papel fundamental na modelagem do problema em questão, sendo essencial sua identificação e compreensão. Considerando a complexidade do problema e a ausência de conhecimento sobre sua forma exata, a escolha pelo uso de técnicas de *Machine Learning*, mais especificamente as redes neurais, justifica-se pela capacidade delas de aprenderem e modelarem relações complexas entre os dados, serem robustas a ruídos e generalizarem informações aprendidas.

Nos mais recentes estudos relativos à predição de séries temporais não-lineares, as redes neurais têm sido aplicadas com veemência, gerando resultados superiores aos tradicionais modelos físicos e/ou estatísticos (BROWNLEE, 2018; CEMERNEK et al., 2021).

Neste capítulo, serão apresentadas as principais características e modelos aplicados em *Deep Learning* para tarefas de classificação. Também serão apresentadas as formas mais relevantes de se avaliar o desempenho de um classificador.

### 3.1 Aprendizagem Profunda em Tarefas de Classificação

Conforme discutido por [Goodfellow, Bengio e Courville \(2016\)](#), o desenvolvimento do aprendizado profundo foi motivado em parte pela deficiência dos algoritmos tradicionais em generalizar certos problemas envolvendo, sobretudo, dados de alta dimensão e altamente complexos. Adicionando mais camadas e mais unidades ocultas, uma rede profunda pode representar funções de alta complexidade. O aprendizado profundo, então, fornece uma estrutura poderosa para o aprendizado supervisionado.

Inicialmente, neste capítulo, serão apresentadas as redes neurais *MultiLayer Perceptron* (MLP), que são os modelos mais tradicionais e amplamente utilizados na aprendizagem supervisionada nas tarefas de classificação e predição de séries temporais. Em sequência, serão apresentados alguns dos principais modelos empregados na aprendizagem profunda, sobretudo, em tarefas de classificação binária.

### 3.2 Redes Neurais *MultiLayer Perceptron* (MLP)

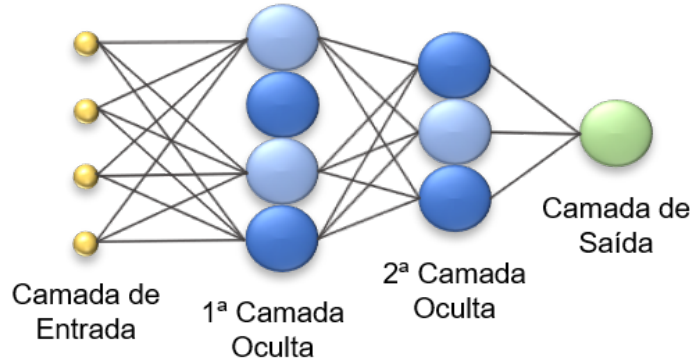
As RNA são inspiradas na arquitetura e aprendizagem do cérebro humano e funcionam como processadores constituídos de unidades de processamento simples, denominados neurônios. Através de um processo de aprendizagem, os neurônios são capazes de armazenar o conhecimento obtido através de experiências e adaptar-se perante às alterações do ambiente. Devido à sua estrutura flexível, as RNA possuem a capacidade de aproximar funções complexas, aprendendo e prevendo padrões lineares e não-lineares com precisão, mesmo nos casos em que as técnicas tradicionais falham ([HAYKIN, 2011](#)).

As redes neurais MLP são compostas por uma camada de entrada, seguida de um número arbitrário de camadas ocultas e uma camada de saída, todas elas compostas por um conjunto de nós chamados de neurônios. [Haykin \(2011\)](#) afirma, no entanto, que os neurônios das camadas ocultas se diferenciam daqueles das demais camadas por serem os responsáveis pela capacidade de aprendizagem de problemas complexos. A [Figura 17](#) ilustra a arquitetura deste tipo de rede, contendo duas camadas ocultas, cujo sinal de entrada se propaga de maneira unidirecional.

As camadas ocultas têm a função de receber os sinais da camada anterior, processá-los e enviá-los para a camada subsequente. Na prática, são permitidas quaisquer interconexões acíclicas entre os neurônios de diferentes camadas. Por fim, está a camada de saída que tem como objetivo apresentar a resposta processada pelas camadas predecessoras da rede. Este resultado pode ser desde a identificação da classe à qual pertence uma determinada entrada (classificação) ou um resultado de processamento de sinal contínuo (regressão) ([HAYKIN, 2011](#); [GOODFELLOW; BENGIO; COURVILLE, 2016](#)).

O processo de aprendizagem do MLP é realizado através do treinamento supervisio-

Figura 17 – Exemplo de uma rede MLP com 2 camadas ocultas contendo, respectivamente, 4 e 3 neurônios.



Fonte: Elaborada pela autora.

nado. De acordo com Haykin (2011), esses algoritmos fundamentam-se na minimização do sinal de erro da resposta atual calculada pela rede em relação à saída desejada. Seguindo esta análise, sejam  $y(t)$  a resposta desejada e  $\hat{y}(t)$  a resposta atual computada pela rede no instante  $t^1$ , o sinal de erro  $e(t)$  pode ser calculado através da Equação 3.2:

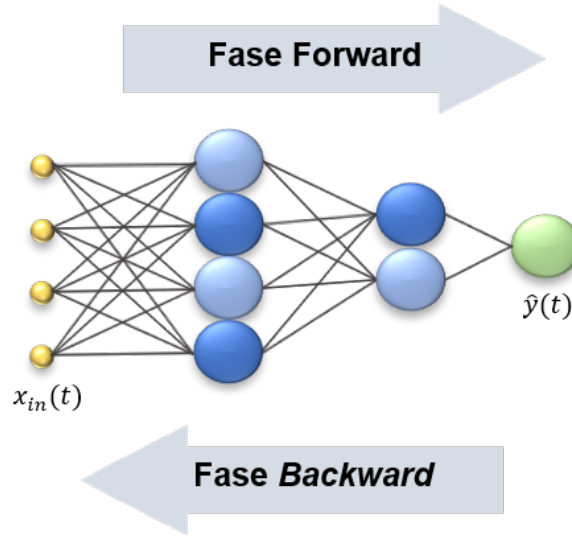
$$e(t) = y(t) - \hat{y}(t). \quad (3.2)$$

O ajuste corretivo dos pesos é realizado visando mitigar a magnitude do sinal de erro  $e(t)$ . Assim, este objetivo é então alcançado minimizando-se a função custo, que em tarefas de classificação pode ser representada pela função de entropia cruzada (*Cross-Entropy*). Essa função mede a diferença entre duas distribuições de probabilidade para uma determinada variável aleatória ou conjunto de eventos.

Um dos algoritmos mais populares devido à sua simplicidade e pioneirismo no treinamento em múltiplas camadas, foi o Algoritmo de Retropropagação de Erro ou *BackPropagation* (RUMELHART; HINTON; WILLIAMS, 1986). O *BackPropagation* ocorre em duas fases, em que cada uma delas processa o sinal percorrendo as camadas da rede em um determinado sentido. A Figura 18 apresenta um esquemático do comportamento dessas duas fases, conhecidas como *Forward* e *Backward* (CARVALHO et al., 2011).

Na fase *Forward*, o sinal de entrada é propagado das entradas até a saída da rede e, a partir da resposta  $\hat{y}(t)$  para um dado padrão de entrada  $x_{in}(t)$ , que faz parte do conjunto de dados reais, calcula-se o erro através da Equação 3.2. Em contrapartida, durante a etapa *Backward*, o sinal de erro é propagado para trás de maneira a promover o ajuste dos pesos entre as conexões da rede. Assim, os pesos sinápticos das camadas intermediárias são

<sup>1</sup> É importante ressaltar que, no contexto do treinamento de uma rede neural,  $t$  pode ser utilizado para representar o índice discreto em uma sequência de dados. Por exemplo, em séries temporais ou em modelos sequenciais, cada  $t$  pode indicar um passo na sequência, constituindo-se como um índice discreto que avança ao longo do tempo.

Figura 18 – Fluxo de processamento do Algoritmo de *BackPropagation*

Fonte: Elaborada pela autora.

atualizados, tendendo a tornar a saída da rede estatisticamente mais próxima da resposta desejada (CARVALHO et al., 2011).

### 3.2.1 Convolutional Neural Network (CNN)

A rede convolucional (*Convolutional Neural Network* – CNN) é uma variação do modelo *MultiLayer Perceptrons* (MLP), inspirada no processo biológico de processamentos de dados visuais. Nas CNN as camadas ocultas totalmente conectadas são substituídas por camadas de convolução (GOODFELLOW; BENGIO; COURVILLE, 2016).

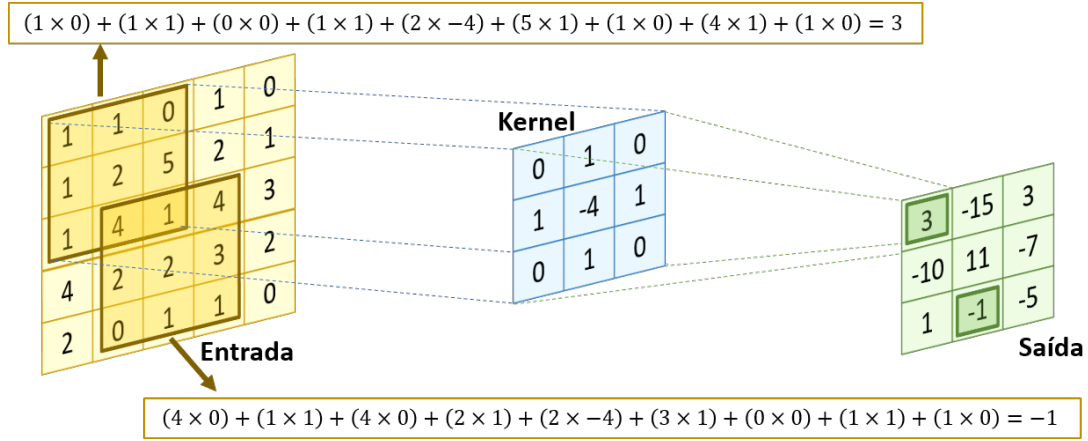
Cada camada convolucional possui um conjunto de filtros, chamados de *kernels*, que percorrem sequencialmente os dados de entrada, aplicando a operação de convolução para a extração de características locais. O tamanho desse deslizamento (passo) é chamado de *stride* e o resultado deste procedimento é denominado mapa de características (GOODFELLOW; BENGIO; COURVILLE, 2016; BROWNLEE, 2018).

A Figura 19 exemplifica a operação de convolução para o caso em que a entrada é representada por uma matriz  $5 \times 5$  e um *kernel*  $3 \times 3$ . O tamanho do *kernel* definirá o tamanho da vizinhança que cada unidade da camada irá processar.

Em termos matemáticos, o valor da característica localizada na posição  $(i, j)$  do  $m$ -ésimo mapa de características da  $l$ -ésima camada,  $z_{i,j,m}^l$ , pode ser calculado pela Equação 3.3,

$$z_{i,j,m}^l = (\mathbf{w}_m^l)^T \mathbf{x}_{i,j}^l + b_m^l, \quad (3.3)$$

em que  $\mathbf{x}_{i,j}^l$  é o elemento de entrada centrada na posição  $(i, j)$  referente à  $l$ -ésima camada,

Figura 19 – Operação de convolução com  $kernel\ 3 \times 3$  e  $stride$  de 1.

Fonte: Elaborada pela autora.

e  $\mathbf{w}_m^l$  e  $b_m^l$  são, respectivamente, o filtro e *bias* do  $m$ -ésimo filtro da  $l$ -ésima camada.

Ao longo do treinamento, os filtros das CNN são ajustados para que sejam ativados na presença de determinadas características, assim, diferentes *kernels* podem produzir diferentes mapas de características para uma mesma entrada. Neste contexto, sequências de camadas convolucionais permitem a extração de padrões mais complexos pela combinação das informações identificadas pelas camadas anteriores. A habilidade das CNN em aprender e extrair automaticamente as características intrínsecas dos dados brutos de entrada, a torna aplicável em muitos problemas envolvendo séries temporais (GOODFELLOW; BENGIO; COURVILLE, 2016; BROWNLEE, 2018).

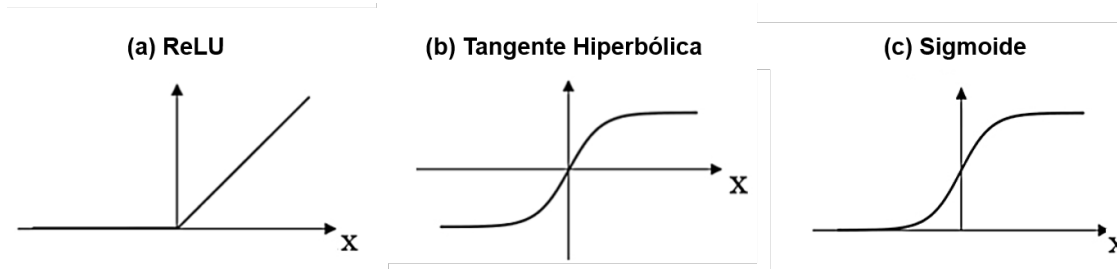
Na prática, o mapa de características é calculado em duas etapas: após ser efetuado o processo de convolução em uma matriz de entrada com um determinado filtro, é aplicada uma função não-linear aos resultados da convolução. Essas funções não-lineares são desejáveis para que a rede consiga representar não linearidades. As funções mais utilizadas são: *Rectified Linear Unit* (ReLU) (NAIR; HINTON, 2010; KRIZHEVSKY; SUTSKEVER; HINTON, 2012), Tangente Hiperbólica (LECUN et al., 1998) e Sigmoid. Essas funções são apresentadas, respectivamente, pelas Equações 3.4, 3.5 e 3.6, enquanto suas formas são ilustradas na Figura 20.

$$ReLU(x) = \max\{0, x\}, \quad (3.4)$$

$$\tanh(x) = \frac{2}{1 + \exp(-2x)} - 1, \quad (3.5)$$

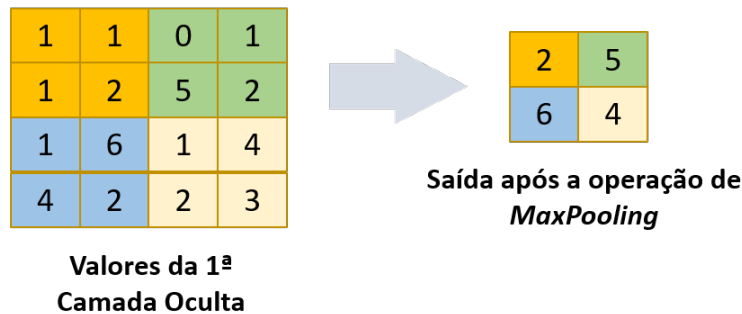
$$\varphi(x) = \frac{1}{1 + \exp(-x)}, \quad (3.6)$$

Figura 20 – Forma das funções de ativação: (a) ReLU, (b) Tangente Hiperbólica e (c) Sigmoide



Fonte: Elaborada pela autora.

Figura 21 – Exemplo de *MaxPooling* com filtro  $2 \times 2$  e *stride* igual a 2.

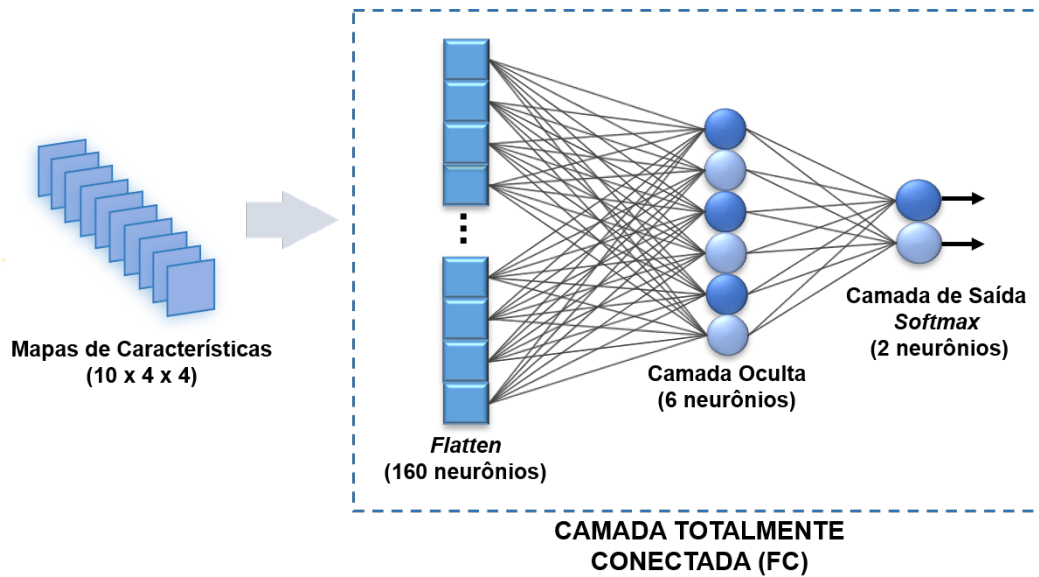


Fonte: Elaborada pela autora.

A ReLU tem apresentado melhores resultados em CNN, resultando em treinamentos mais rápidos e com desempenho parecido ou superior às outras funções de ativação, tais como a tangente hiperbólica (KRIZHEVSKY; SUTSKEVER; HINTON, 2012). Por esta razão, no presente trabalho, optou-se pelo uso da ReLU.

Em geral, a camada convolucional é seguida de uma camada de *Pooling*. O *Pooling* tem a finalidade de reduzir a dimensão da camada anterior através da seleção das informações mais significativas, fazendo com que a rede crie resistência a pequenas variações e distorções locais (GOODFELLOW; BENGIO; COURVILLE, 2016). Esta redução pode ser alcançada por meio de um *stride* não unitário, com valor usual igual a 2, por exemplo, reduzindo-se assim as dimensões espaciais pela metade. As operações típicas de *Pooling* são o *MaxPooling* e o *Average Pooling* que consistem em delimitar regiões para seleção do valor máximo e da média de cada uma delas, respectivamente (GOODFELLOW; BENGIO; COURVILLE, 2016; BUDUMA; LOCASCIO, 2017). A Figura 21 ilustra o processo de *MaxPooling*.

A maior complexidade da CNN pode acarretar em um problema conhecido como sobreajuste ou *overfitting*. Com o intuito de contornar este problema, é comum a utilização do *dropout*. O *dropout* é uma técnica de regularização em que, a cada iteração de treinamento, promove a remoção aleatória de uma porcentagem pré-estabelecida das unidades

Figura 22 – Exemplo de composição de Camada FC: *Flatten*, oculta e saída

Fonte: Elaborada pela autora.

de uma determinada camada, os adicionando novamente na iteração seguinte. Esta técnica permite que a rede aprenda atributos mais robustos, uma vez que uma unidade não poderá depender da presença específica das demais (GOODFELLOW; BENGIO; COURVILLE, 2016; BUDUMA; LOCASCIO, 2017; LABACH; SALEHINEJAD; VALAEE, 2019).

Em tarefas de classificação é necessária a aplicação de ao menos uma camada totalmente conectada (*Fully Connected* – FC). A FC é formada por uma estrutura semelhante à MLP e é a responsável por determinar a classificação dos dados a partir das informações geradas pelas CNN. Para isto, a saída da última camada convolucional tem que passar por uma camada chamada de *Flatten*, que opera uma transformação na matriz de entrada, alterando seu formato para um vetor. O *Flatten* fornece as características extraídas pelas camadas convolucionais como sequências de entrada para a camada FC para que o resultado da classificação seja obtido (GOODFELLOW; BENGIO; COURVILLE, 2016).

A Figura 22 apresenta um esquema representativo em que 10 mapas de características  $4 \times 4$  resultantes da última camada convolucional de uma rede são colocados de forma linear, pela camada *Flatten* contendo 160 unidades, para formação da entrada da camada FC. Cada unidade da camada *Flatten* está conectada a um único valor dos mapas de características e a saída da unidade é igual a entrada. A informação é transmitida à uma camada oculta composta por 6 unidades e então encaminhada à camada de saída com 2 unidades onde a resposta final do classificador é gerada.

Em tarefas de classificação, o número de unidades da camada de saída será igual ao número de classes a serem previstas. Nesta camada, é comum a utilização da função de

ativação do tipo *Softmax* que é calculada pela [Equação 3.7](#):

$$\text{Softmax}(y_c) = \frac{\exp(y_c)}{\sum_{i=1}^{N_c} \exp(y_i)}, \quad (3.7)$$

sendo  $y_c$  o vetor da  $c$ -ésima classe de saída da última camada da rede neural,  $N_c$  o número total de classes e  $c$  a classe cujo valor de saída se deseja calcular.

A função *Softmax* ajuda a mapear a resposta de uma rede neural para um vetor de saída, onde cada elemento do vetor corresponde a uma classe, e a soma dos valores do vetor é igual a 1. A entrada da rede é associada à classe correspondente ao elemento do vetor com o maior valor. A função *Softmax* é necessária ao se aplicar a função de perda de entropia cruzada<sup>2</sup> ([GOODFELLOW; BENGIO; COURVILLE, 2016](#); [CHOLLET, 2021](#)).

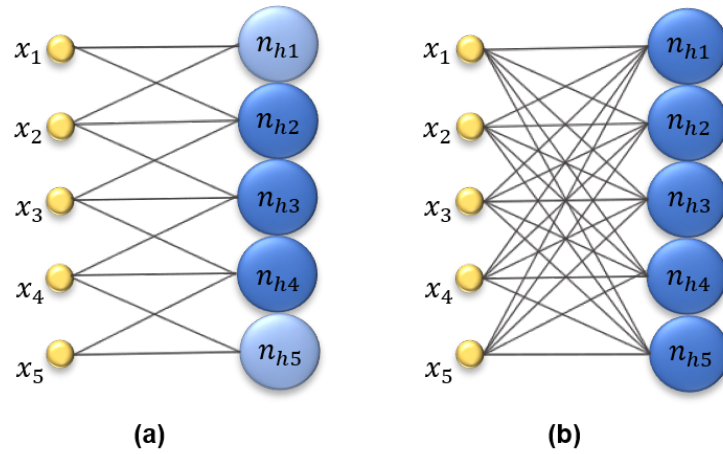
As CNN apresentam características que as diferenciam de modelos de redes neurais de arquitetura rasa pela redução dos requisitos de memória para treinamento de seus parâmetros. Estas características tornam mais viáveis as aplicações que utilizam redes CNN profundas que podem chegar a apresentar milhares de camadas de processamento, mesmo que isto corresponda a um substancial aumento da quantidade de parâmetros da rede ([GOODFELLOW; BENGIO; COURVILLE, 2016](#); [BROWNLEE, 2018](#); [LABACH; SALEHINEJAD; VALAEE, 2019](#)).

Conforme discutido por [Goodfellow, Bengio e Courville \(2016\)](#), as principais características das CNN são: conectividade esparsa, campos receptivos locais e compartilhamento de parâmetros. A seguir, cada uma das características será detalhada:

- **Conectividade Esparsa:** Uma vez que o filtro utilizado no processo de convolução apresenta dimensões menores que a matriz de entrada, ocorre uma correlação espacial. A [Figura 23\(a\)](#) apresenta um exemplo de conectividade esparsa utilizando um filtro de tamanho 3. Observa-se que apenas as unidades  $n_{h2}$ ,  $n_{h3}$  e  $n_{h4}$  da camada seguinte são afetadas pela entrada  $x_3$ . Em contrapartida, conforme é mostrado na [Figura 23\(b\)](#), em uma MLP em que há conectividade global, todas as unidades da camada seguinte são afetadas por todas as entradas.
- **Campos receptivos locais:** A região delimitada na matriz de entrada a partir da qual uma unidade obtém uma saída é chamada de campo receptivo local. Para cada unidade, uma saída é obtida considerando o pixel central da região e sua vizinhança. A composição de todas as saídas das regiões delimitadas compõem um

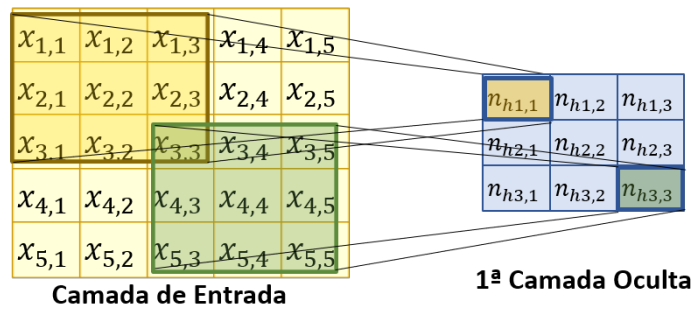
<sup>2</sup> A entropia cruzada é uma das funções de perda empregadas em problemas de otimização. Em resumo, essa função calcula a discrepância entre duas distribuições de probabilidade: a distribuição real (verdadeira) e a distribuição prevista pelo modelo. Para minimizar essa função de perda e encontrar o valor mínimo, é empregado o cálculo do gradiente, que consiste em um vetor de derivadas parciais. O gradiente representa a taxa de mudança da função de perda e direciona para o sentido em que a função reduz seu valor. Essa abordagem é fundamental no processo de otimização de modelos de aprendizado de máquina ([GORDON-RODRIGUEZ et al., 2020](#)).

Figura 23 – Exemplo de (a) conectividade esparsa da rede CNN e (b) conectividade global da rede MLP.



Fonte: Elaborada pela autora.

Figura 24 – Campos receptivos locais de uma rede CNN.

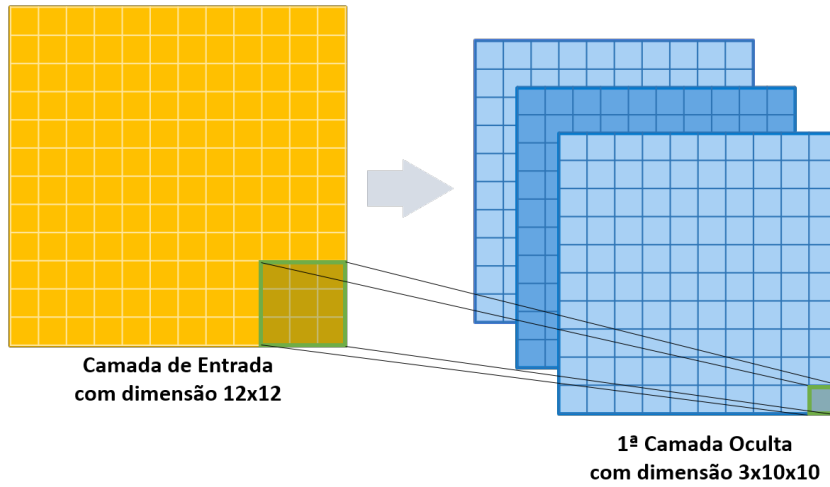


Fonte: Elaborada pela autora.

mapa de características. A Figura 24 ilustra este processo, onde existe uma matriz de entrada e uma camada com 9 unidades, cada qual com um campo receptivo  $3 \times 3$ . O campo receptivo local da unidade  $n_{h1,1}$  na matriz de entrada está destacado em amarelo, enquanto que os destacados em verde representam o campo receptivo local da unidade  $n_{h3,3}$ .

- **Compartilhamento de parâmetros:** para que as saídas das unidades da Figura 25, por exemplo, sejam as mesmas para uma mesma entrada, independente da localização dos seus campos receptivos, é necessário que eles possuam os mesmos pesos. Esse conceito é conhecido como compartilhamento de parâmetros. A Figura 25 apresenta um esquema de uma camada que obteve como saída três mapas de características, cada mapa obtido por um *kernel* diferente. Se cada *kernel* for de dimensão  $3 \times 3$ , seriam necessários  $3 \times 3 \times 3 = 27$  parâmetros (desconsiderando os *biases*). Se não houvesse o compartilhamento de pesos, além da saída não ser invariante à localização, o número de parâmetros seria igual a  $3 \times 3 \times 3 \times 10 \times 10 = 2.700$  (desconsiderando os

Figura 25 – Exemplo de CNN com três mapas de características.



Fonte: Elaborada pela autora.

*biases*). Essa comparação revela que o processo para gerar um mapa de características equivale a uma convolução entre uma matriz de entrada e um *kernel*, o qual representa um conjunto de unidades com pesos compartilhados, resultando em uma significativa redução no número de parâmetros.

### 3.2.2 Long Short Term Memory (LSTM)

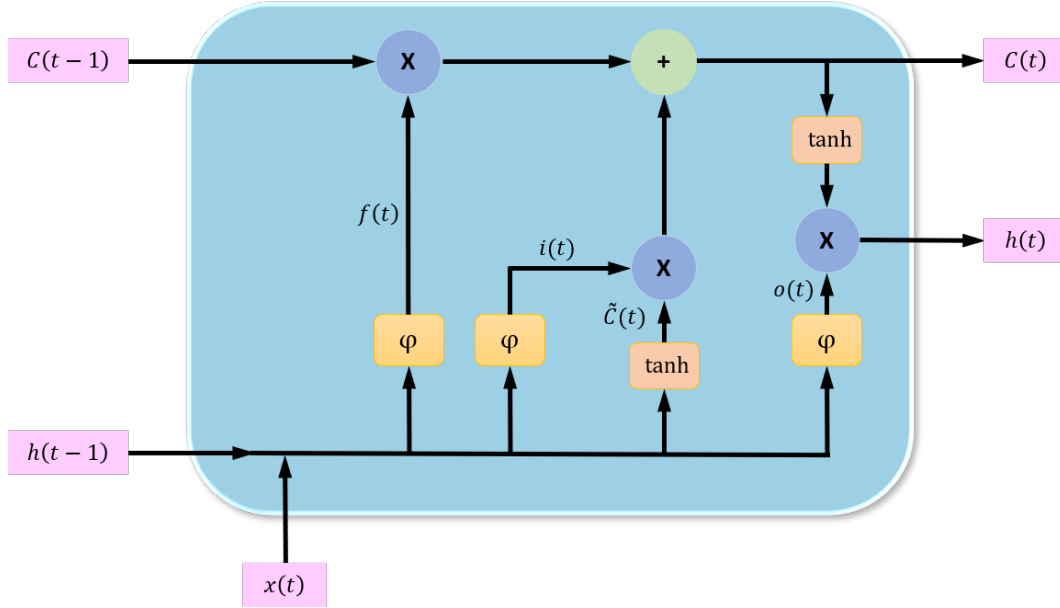
Para problemas relacionados à grandes sequências de dados é indicada a utilização da LSTM (HOCHREITER; SCHMIDHUBER, 1997). A LSTM é composta por conjunto de subredes conectadas de forma recorrente, chamadas de blocos de memória. Cada bloco contém uma ou mais células interconectadas e três unidades multiplicativas, chamadas de portas (*gates*), mais especificamente *forget gate*  $f(t)$ , *input gate*  $i(t)$ , e *output gate*  $o(t)$  (GOODFELLOW; BENGIO; COURVILLE, 2016; BUDUMA; LOCASCIO, 2017).

A arquitetura básica de uma célula LSTM está representada na Figura 26, em que  $x(t)$  corresponde ao sinal de entrada,  $C(t)$  e  $C(t-1)$  são, respectivamente, o estado da célula de memória atual e no instante anterior e  $h(t)$  e  $h(t-1)$  representam o seu estado oculto atual e instante anterior. O sinal é transmitido às três portas que controlam a informação. Para a unidade LSTM básica, suas entradas externas são seus estados de célula anterior ( $C(t-1)$  e  $h(t-1)$ ).

A primeira parte do processamento da informação dentro de uma célula é realizada no *forget gate*  $f(t)$ , cuja função é controlar quais partes do estado de longo prazo da célula devem ser esquecidas e quais devem ser mantidas. O *forget gate* é calculado conforme a Equação 3.8:

$$f(t) = \varphi(\mathbf{W}_{fx}x(t) + \mathbf{W}_{fh}h(t-1) + \mathbf{b}_f), \quad (3.8)$$

Figura 26 – A estrutura de uma célula LSTM.



Fonte: Elaborada pela autora.

em que os termos  $\mathbf{W}_{fx}$  e  $\mathbf{W}_{fh}$  correspondem, respectivamente, aos pesos de entrada e pesos recorrentes em relação ao *forget gate*,  $\mathbf{b}_f$  representa o *bias* e  $\varphi(\cdot)$  é uma função de ativação não linear que, em geral, é a função sigmoide (GOODFELLOW; BENGIO; COURVILLE, 2016).

A segunda parte do processamento consiste em decidir quais partes devem ser adicionadas ao estado de longo prazo. Esta etapa é realizada através do *input gate*  $i(t)$ , calculada através da Equação 3.9:

$$i(t) = \varphi(\mathbf{W}_{ix}x(t) + \mathbf{W}_{ih}h(t-1) + \mathbf{b}_i), \quad (3.9)$$

sendo os termos  $\mathbf{W}_{ix}$  e  $\mathbf{W}_{ih}$ , respectivamente, os pesos de entrada e recorrentes do *input gate* e  $\mathbf{b}_i$  é o *bias*. Tal como ocorre no *forget gate*, o resultado é aplicado em uma função de ativação do tipo sigmoide  $\varphi(\cdot)$  (GOODFELLOW; BENGIO; COURVILLE, 2016).

Em paralelo ao *input gate*, é determinado o estado candidato  $\tilde{C}(t)$ , representado pela Equação 3.10:

$$\tilde{C}(t) = \tanh(\mathbf{W}_{cx}x(t) + \mathbf{W}_{ch}h(t-1) + \mathbf{b}_c), \quad (3.10)$$

embora, de forma análoga,  $\mathbf{W}_{cx}$ ,  $\mathbf{W}_{ch}$ ,  $\mathbf{b}_c$  correspondam, respectivamente, aos pesos de entrada, pesos recorrentes e *bias* em relação ao estado candidato, uma função tangente hiperbólica,  $\tanh(\cdot)$ , é utilizada garantindo que os valores permaneçam na faixa entre  $-1$  e  $1$  (GOODFELLOW; BENGIO; COURVILLE, 2016).

Conforme demonstra a [Equação 3.11](#), o estado atual da célula  $C(t)$  será calculado com base no estado da célula no instante anterior  $C(t-1)$ , no *forget gate*  $f(t)$ , no *input gate*  $i(t)$  e no estado candidato  $\tilde{C}(t)$ ,

$$C(t) = f(t) \odot C(t-1) + i(t) \odot \tilde{C}(t), \quad (3.11)$$

em que  $\odot$  corresponde ao produto de Hadamard.

Por fim, *output gate*  $o(t)$  tem a função de controlar as partes que devem ser lidas e exibidas no estado atual  $h(t)$ . O *output gate* e o estado oculto da rede são calculados de acordo com as [Equações 3.12 e 3.13](#), respectivamente,

$$o(t) = \varphi(\mathbf{W}_{ox}x(t) + \mathbf{W}_{oh}h(t-1) + \mathbf{b}_o), \quad (3.12)$$

$$h(t) = o(t) \odot \tanh(C(t)), \quad (3.13)$$

sendo os termos  $\mathbf{W}_{ox}$ ,  $\mathbf{W}_{oh}$ ,  $\mathbf{b}_o$ , respectivamente, os pesos de entrada, os pesos recorrentes do *output gate* e o *bias* ([GOODFELLOW; BENGIO; COURVILLE, 2016](#)).

A LSTM evita o desaparecimento do gradiente através da comutação de suas portas, que desenvolvem uma espécie de memória temporal. Como consequência, elas são capazes de identificar quais informações temporais devem ser transmitidas ou descartadas pela rede ([BUDUMA; LOCASCIO, 2017](#)).

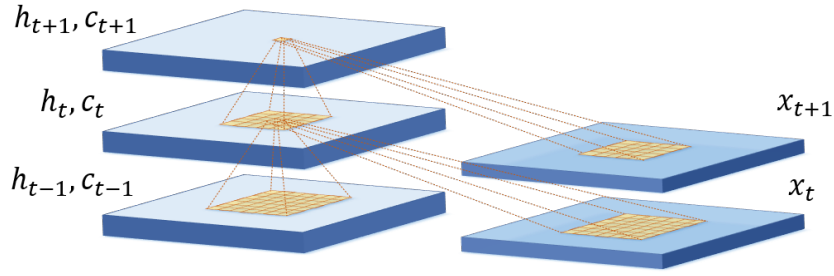
### 3.2.3 Convolutional Long Short-Term Memory (ConvLSTM)

A *Convolutional Long Short-Term Memory* (ConvLSTM) é uma variante da LSTM, projetada por [Xingjian et al. \(2015\)](#), com objetivo de capturar padrões espaço-temporais de conjuntos de dados sequenciais em grande escala. Esta variante funciona de maneira análoga a uma LSTM, diferenciando-se pelas operações de multiplicação interna de matrizes que são substituídas por convoluções. Em consequência disto, a ConvLSTM é capaz de extrair características espaciais subjacentes no tratamento de dados multivariados, o que é uma limitação dos modelos LSTM tradicionais ([AGGA et al., 2021](#)).

Uma outra diferença entre a ConvLSTM e a LSTM é o número de dimensões de entrada. As entradas e saídas de célula, estados ocultos e portas da ConvLSTM formam tensores 3D, cujas duas últimas dimensões são dimensões espaciais (linhas e colunas). O cálculo das portas da ConvLSTM é dado pelas [Equações 3.14, 3.15, 3.16 e 3.17](#).

Verifica-se que a principal diferença das [Equações 3.14, 3.15, 3.16 e 3.17](#) e as equações da LSTM tradicional ([Equações 3.8, 3.9, 3.10 e 3.12](#)) é que a operação de convolução ( $\star$ ) substitui a multiplicação das matrizes em cada porta  $f(t)$ ,  $i(t)$  e  $o(t)$ , bem

Figura 27 – Estrutura interna da ConvLSTM



Fonte: Adaptada de [Xingjian et al. \(2015, p. 4\)](#).

como no estado candidato  $\tilde{C}(t)$ . O cálculo de  $C(t)$  e  $h(t)$  permanece o mesmo das Equações 3.11 e 3.13 ([XINGJIAN et al., 2015](#)).

$$f(t) = \varphi(\mathbf{W}_{fx} \star x(t) + \mathbf{W}_{fh} \star h(t-1) + \mathbf{b}_f), \quad (3.14)$$

$$i(t) = \varphi(\mathbf{W}_{ix} \star x(t) + \mathbf{W}_{ih} \star h(t-1) + \mathbf{b}_i), \quad (3.15)$$

$$\tilde{C}(t) = \tanh(\mathbf{W}_{cx} \star x(t) + \mathbf{W}_{ch} \star h(t-1) + \mathbf{b}_c), \quad (3.16)$$

$$o(t) = \varphi(\mathbf{W}_{ox} \star x(t) + \mathbf{W}_{oh} \star h(t-1) + \mathbf{b}_o). \quad (3.17)$$

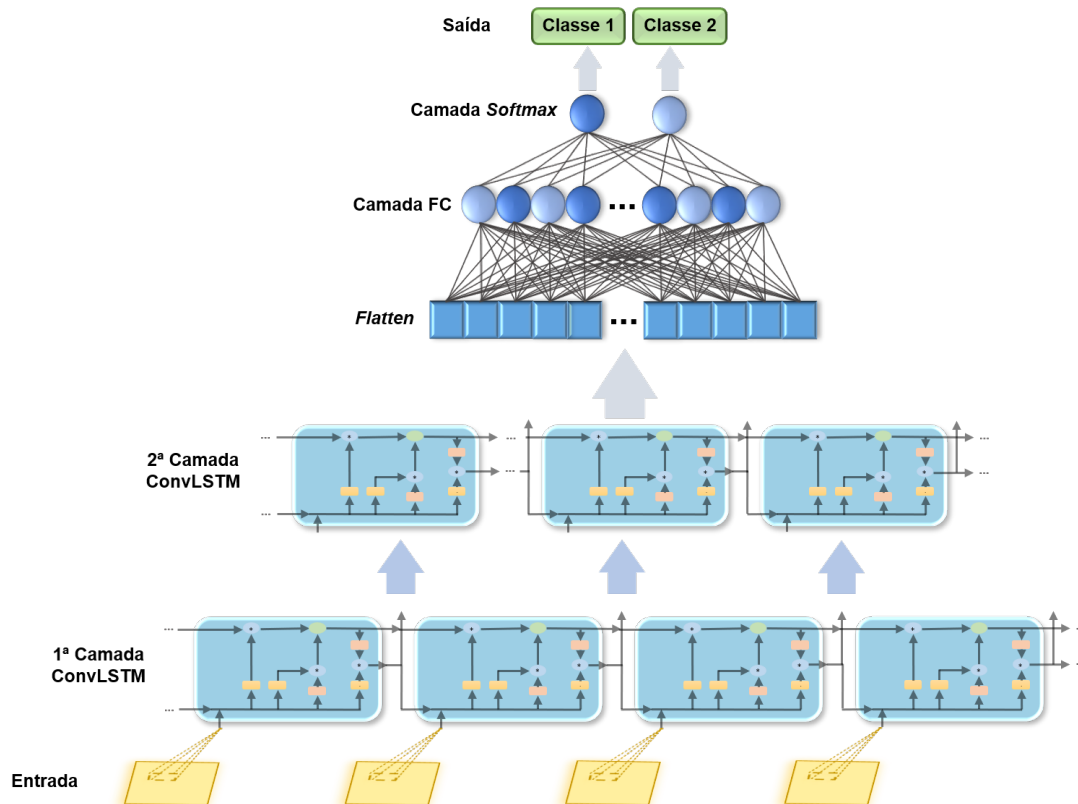
Assim como os filtros convolucionais aplicados entre a entrada e a camada oculta determinam a resolução dos mapas de características criados a partir da entrada, o tamanho do filtro convolucional entre as conexões ocultas determinam as informações agregadas que a unidade ConvLSTM recebe do estado de tempo anterior. A partir das entradas e estados passados dos vizinhos locais, a ConvLSTM determina, então, o estado futuro de uma determinada célula, conforme é apresentado na [Figura 27](#) ([XINGJIAN et al., 2015](#)).

[Xingjian et al. \(2015\)](#) constataam que a transição de estados entre instantes de tempo para uma unidade ConvLSTM pode ser interpretada como um movimento entre *frames* em uma sequência de vídeo. Considerando que a entrada desta rede será composta unicamente por variáveis temporais, os filtros de transição maiores, portanto, poderão capturar as variações mais rápidas do sinal, enquanto os filtros de transição menores serão os responsáveis pela captura de variações mais lentas.

A [Figura 28](#) apresenta uma representação genérica de uma rede neural contendo duas camadas ConvLSTM.

Uma vez que a ConvLSTM trabalha com convoluções diretamente como parte da leitura de entrada nas próprias células, ela é capaz de aplicar o compartilhamento de peso,

Figura 28 – Estrutura genérica de uma rede ConvLSTM.



Fonte: Elaborada pela autora.

da mesma maneira que as CNN, trabalhando simultaneamente com os dados de entrada como dados sequenciais (XINGJIAN et al., 2015).

### 3.3 Validação dos Modelos

A abordagem de validação dos modelos de aprendizado de máquinas geralmente envolve a divisão dos dados em três conjuntos distintos: treinamento, validação e teste. Essa divisão é crucial para avaliar a capacidade de generalização do modelo e evitar *overfitting* (HAYKIN, 2011).

O conjunto de treinamento é o responsável por atuar na etapa de aprendizagem, em que os pesos da rede são ajustados seguindo o algoritmo de aprendizagem. Para avaliar a capacidade de generalização da rede durante o processo, um conjunto de dados distinto (conjunto de validação) simula a apresentação de objetos novos ao classificador. Ao fim do treinamento, o conjunto de testes é empregado para avaliar o desempenho da rede (HAYKIN, 2011).

Uma abordagem comum, que segue essa divisão, é o método  $k$ -fold. Nele, o conjunto de treinamento é dividido em  $k$  subconjuntos mutuamente exclusivos. A cada iteração, as amostras das  $k - 1$  partições são usadas como treinamento do classificador, enquanto a

Figura 29 – Matriz de confusão.

|             |          | Classe Predita           |                          |
|-------------|----------|--------------------------|--------------------------|
|             |          | Positivo                 | Negativo                 |
| Classe Real | Positivo | Verdadeiro Positivo (TP) | Falso Negativo (FN)      |
|             | Negativo | Falso Positivo (FP)      | Verdadeiro Negativo (TN) |

Fonte: Elaborada pela autora.

partição restante é usada na validação. Esse processo é repetido por  $k$  vezes, empregando, a cada iteração, uma partição diferente para a validação e sendo calculadas as métricas de desempenho. O desempenho final do classificador na fase de treinamento é calculado pela média dos desempenhos observados sobre cada subconjunto de teste (STONE, 1974; BENGIO; GRANDVALET, 2004; HAYKIN, 2011; GOODFELLOW; BENGIO; COURVILLE, 2016).

Neste trabalho, será aplicada uma metodologia baseada na ideia do  $k$ -fold, em que a cada iteração uma porcentagem das amostras é selecionada para o treinamento (80%), outra na validação (10%) e outra na fase de teste (10%). É importante destacar que as amostras são sequenciais e mantêm a sequência temporal ao longo das iterações do processo. As métricas de desempenho são calculadas a cada iteração para avaliar a capacidade de generalização do modelo.

### 3.4 Métricas de Desempenho do Classificador

Para análise do desempenho nos problemas de classificação, é comum o emprego da matriz de confusão, ilustrada na Figura 29.

Os valores que compõem a matriz de confusão são resultantes da comparação entre a classe correta de cada uma das amostras do conjunto de teste e a classe obtida pelo classificador analisado (FAWCETT, 2006; ROGERS; GIROLAMI, 2016).

Para os classificadores implementados neste trabalho, será associada a ocorrência de *clogging* à classe positiva e a sua ausência à classe negativa. Conforme observado na Figura 29, após a classificação, os valores podem pertencer à quatro possíveis categorias (ROGERS; GIROLAMI, 2016), sendo elas:

- **Verdadeiro Positivo (*True Positive* - TP):** amostras que pertencem à classe

positiva e que foram classificadas corretamente;

- **Falso Positivo (*False Positive* - FP):** amostras que pertencem à classe negativa, mas que foram classificadas incorretamente como pertencendo à classe positiva;
- **Verdadeiro Negativo (*True Negative* - TN):** amostras que pertencem à classe negativa e que foram classificadas corretamente;
- **Falso Negativo (*False Negative* - FN):** amostras que pertencem à classe positiva, mas que foram classificadas incorretamente como pertencendo à classe negativa.

A partir da matriz de confusão é possível derivar uma série de medidas de desempenho, dentre elas:

- **Acurácia ( $A_{cc}$ ):** é uma métrica que avalia o desempenho global de um modelo. Calcula a proporção de classificações corretas, tanto de casos positivos quanto negativos. Ela é dada pela [Equação 3.18](#):

$$A_{cc} = \frac{TP + TN}{TP + FN + TN + FP} \times 100\%. \quad (3.18)$$

- **Precisão ( $P_{rc}$ ):** é a medida de exatidão do modelo, isto é, é a proporção de exemplos positivos classificados corretamente entre todos aqueles preditos como positivos. Ela é dada pela [Equação 3.19](#):

$$P_{rc} = \frac{TP}{TP + FP} \times 100\%. \quad (3.19)$$

- **Recall ( $R_{cl}$ ):** é a taxa de classificação de verdadeiros positivos em relação ao total de exemplos positivos, isto é, corresponde à taxa de amostras positivas corretamente identificadas. Ela é dada pela [Equação 3.20](#):

$$R_{cl} = \frac{TP}{TP + FN} \times 100\%. \quad (3.20)$$

- **F1-Score ( $F_1$ ):** é a combinação entre Precisão e Recall que gera um indicador da qualidade geral do modelo, conforme [Equação 3.21](#):

$$F_1 = \frac{2P_{rc}R_{cl}}{P_{rc} + R_{cl}} \times 100\%. \quad (3.21)$$

Conforme discutido por [Chicco e Jurman \(2020\)](#), ainda não há um consenso sobre a escolha de uma medida de desempenho unificada. Mesmo que a Acurácia e o F1-Score estejam entre as métricas mais populares adotadas em tarefas de classificação binária, essas

medidas podem gerar resultados excessivamente otimistas, especialmente em conjuntos de dados desequilibrados<sup>3</sup>.

Por sua vez, o *Matthews Correlation Coefficient* (MCC) é uma medida de qualidade de classificações binárias, robusta para bases desbalanceadas e mais confiável, de maneira que ela só produz uma pontuação alta se o classificador for capaz de prever corretamente a maioria das classes positivas e a maioria das classes negativas, simultaneamente. O MCC pode ser usado mesmo em casos de desbalanceamento entre classes e seu cálculo é apresentado na [Equação 3.22](#) (BOUGHORBEL; JARRAY; EL-ANBARI, 2017):

$$MCC = \frac{[(TP)(TN)] - [(FP)(FN)]}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}. \quad (3.22)$$

O MCC varia no intervalo entre  $-1$  e  $+1$ , em que os valores extremos  $+1$  e  $-1$  indicam, respectivamente, uma classificação perfeita e uma classificação totalmente incorreta, enquanto que o valor  $0$  indica uma classificação aleatória enviesada pelo desequilíbrio das classes, semelhante ao que um classificador *Dummy*<sup>4</sup>, que se baseia na classe majoritária, faria. Isto é, quanto mais próximo de zero for o MCC, menor será a correlação entre a classe prevista e a verdadeira (BOUGHORBEL; JARRAY; EL-ANBARI, 2017; CHICCO; TÖTSCH; JURMAN, 2021).

<sup>3</sup> Um conjunto de dados desequilibrado é aquele em que as classes não estão representadas de maneira equitativa, isto é, há uma predominância significativa de uma classe sobre as outras.

<sup>4</sup> O classificador *Dummy* é um modelo de aprendizado de máquinas utilizado para estabelecer uma linha de base na validação de modelos reais. Ele é gerado com regras simples e sem muita complexidade, sendo essencialmente uma referência mínima de desempenho para outros modelos. Por ser tão simples, o classificador *Dummy* geralmente apresenta um desempenho muito baixo em comparação com modelos mais sofisticados. No entanto, sua utilidade reside em fornecer uma medida mínima de desempenho esperado e em identificar possíveis problemas de desequilíbrio de classe em um conjunto de dados (MARTINO; RIZZI; MASCIOLI, 2018).



## 4 Metodologia

Nesta pesquisa, são empregados dados históricos de 6 meses de medições feitas em um processo siderúrgico de lingotamento contínuo. As variáveis foram amostradas a uma taxa de uma amostra por segundo, provenientes de dois distribuidores (*DI5* e *DI6*) operando alternadamente sobre um molde e um veio (*V5*), totalizando 4.320 horas de processo.

Devido à complexidade do processo e ao número de variáveis que podem influenciar na ocorrência de *clogging*, uma análise é necessária para seleção das variáveis de entrada dos modelos. Essa lista deve se basear em dados com o objetivo de avaliar não somente a correlação entre cada variável e sua ocorrência de *clogging*, mas também a qualidade dos dados fornecidos e a quantidade de ocorrências do evento a ser detectado.

É importante frisar que não faz parte do escopo deste trabalho definir a localização do *clogging* (isto é, se a formação ocorreu na válvula gaveta e/ou na parede da SEN, por exemplo), suas causas e/ou origens e nem estipular o fluxo de aço ou outra variável em qualquer parte do sistema. O único objetivo do trabalho é detectar o início da obstrução e sinalizar ao operador que tomará as ações necessárias, conforme Seção 2.2.4.

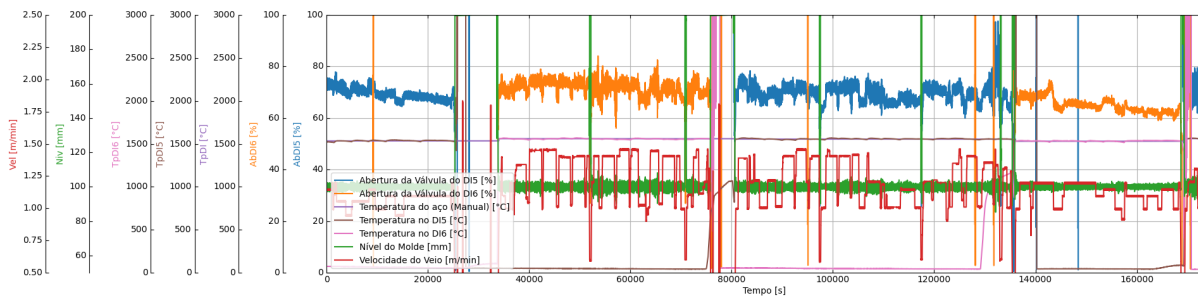
### 4.1 Análise Descritiva das Variáveis do Processo

Baseado nos experimentos realizados na literatura abordando o problema (IKÄHEIMONEN et al., 2002; VANNUCCI; COLLA, 2011; ROUT et al., 2013; KONG; CANG, 2019; KONG; CHEN; CANG, 2019; PELLEGRINI et al., 2019) e da informação de especialistas do processo, foram selecionadas as variáveis que podem estar correlacionadas aos eventos de obstrução do SEN.

O conjunto de dados foi rotulado pelos operadores do processo e, em geral, eles classificaram como *clogging* o evento correspondente às relações entre a abertura da válvula, a velocidade do veio e o nível do molde. O mecanismo fenomenológico deste tipo de evento é sempre caracterizado, inicialmente, por uma abertura da válvula sem que haja redução do nível do molde ou alteração na velocidade do veio. Na maioria das vezes, a abertura da válvula ocorre após oscilações crescentes de abertura e fechamento, podendo refletir, ou não, em oscilações perceptíveis no nível do molde.

Em conjunto com as variáveis referentes à abertura da válvula, nível do molde e velocidade do veio, a bibliografia também indica como dado complementar o uso de variáveis relacionadas à temperatura, pressão e vazão de argônio, contudo, as variáveis relacionadas ao argônio não estão disponibilizadas de maneira adequada pela indústria e,

Figura 30 – Exemplo típico do comportamento dos sinais do processo com base nos dados brutos disponibilizados pela indústria.



Fonte: Elaborada pela autora.

por conseguinte, a análise e possível seleção se tornaram inviáveis.

A Figura 30 ilustra um exemplo típico do comportamento das variáveis pré-selecionadas do processo, ao longo de dois dias, com base nos dados brutos disponibilizados pela indústria. Já a Tabela 2 apresenta o intervalo de valores teóricos e reais dessas variáveis, bem como seus respectivos percentis. A análise é feita sobre os dados brutos, na forma em que foram disponibilizados nos arquivos, não havendo qualquer tipo de tratamento prévio. Para melhor visualização, esses atributos estão organizados em grupos de acordo com a sua natureza física no processo.

As variáveis de temperatura ( $TpDI$ ,  $TpDI5$  e  $TpDI6$ ) apresentam uma alta ocorrência de *outliers* decorrentes de leituras nos sensores possivelmente ocorridas com o equipamento fora de atividade no processo. Além disto, a grandeza numérica desses

Tabela 2 – Variáveis do processo pré-selecionadas.

| TAG                                                     | Dados da Variável<br>Descrição                            | Min.<br>Teórico | Máx.<br>Teórico | Min.<br>Real | Máx.<br>Real | Percentil |          |          |
|---------------------------------------------------------|-----------------------------------------------------------|-----------------|-----------------|--------------|--------------|-----------|----------|----------|
|                                                         |                                                           |                 |                 |              |              | 25 [%]    | 50 [%]   | 75 [%]   |
| <b>Grandeza: Temperatura, Unidade de medida: [°C]</b>   |                                                           |                 |                 |              |              |           |          |          |
| $TpDI$                                                  | Temperatura do aço no distribuidor amostrada manualmente. | 1.400           | 1.600           | 0            | 143.332      | 1.540,70  | 1.551,80 | 1.559,60 |
| $TpDI5$                                                 | Temperatura no $DI5$ (medição contínua).                  | 0               | 1.800           | 0            | 65.529,80    | 45        | 1.519    | 1.556    |
| $TpDI6$                                                 | Temperatura no $DI6$ (medição contínua).                  | 0               | 1.800           | 0            | 65.513,85    | 47,56     | 1.530    | 1.557    |
| <b>Grandeza: Abertura, Unidade de medida: [%]</b>       |                                                           |                 |                 |              |              |           |          |          |
| $AbDI5$                                                 | Percentual da abertura total da válvula gaveta do $DI5$ . | 0               | 100             | -1,24        | 102,06       | 69,09     | 99,62    | 100,46   |
| $AbDI6$                                                 | Percentual da abertura total da válvula gaveta do $DI6$ . | 0               | 100             | -1,91        | 104,19       | 71,23     | 101,17   | 102,58   |
| <b>Grandeza: Nível, Unidade de medida: [mm]</b>         |                                                           |                 |                 |              |              |           |          |          |
| $Niv$                                                   | Nível de aço no molde.                                    | 0               | 200             | 0            | 200          | 99,59     | 100,34   | 103,37   |
| <b>Grandeza: Velocidade, Unidade de medida: [m/min]</b> |                                                           |                 |                 |              |              |           |          |          |
| $Vel$                                                   | Velocidade do aço no veio.                                | 0               | 2,5             | -2,52        | 2,53         | 1,067     | 1,29     | 2,53     |
| <b>Grandeza: Binária, Unidade de medida: –</b>          |                                                           |                 |                 |              |              |           |          |          |
| $DI$                                                    | Distribuidor em operação.                                 | 5               | 6               | 5            | 6            | -         | -        | -        |

Fonte: Elaborada pela autora.

valores atípicos excede a faixa de valores esperada. Por exemplo, a variável  $TpDI$ , que possui medição manual para aferição dos instrumentos, atinge patamares extremos da ordem de  $14 \times 10^4$   $^{\circ}C$ .

É observado que as variáveis  $TpDI5$  e  $TpDI6$  possuem grandezas similares no que dizem respeito à concentração e distribuição. Nessas duas variáveis, os valores acima de 4000  $^{\circ}C$  são considerados *outliers* por aplicação de técnicas matemáticas, todavia, no processo físico, as temperaturas acima de 2000  $^{\circ}C$  já são consideradas fora do escopo.

O atributo  $TpDI$  apresenta cerca de 182.614 amostras abaixo de 1400  $^{\circ}C$  e 78.928 acima de 1600  $^{\circ}C$ . Já a variável  $TpDI5$ , embora não possua valores abaixo do seu mínimo teórico, apresenta aproximadamente 13,6% dos seus dados acima de 1800  $^{\circ}C$ . O mesmo vale para a variável  $TpDI6$ , que também não possui valores abaixo de 0  $^{\circ}C$ , mas apresenta cerca de 12,7% dos seus dados acima do máximo teórico.

Desconsiderando-se os trechos com discrepâncias nas medições de temperatura durante a coleta das amostras das classes de eventos normal e *clogging*, observou-se que, no processo em estudo, a temperatura se manteve praticamente constante com um leve decaimento que não apresentou relacionamento factual com os fenômenos de operação normal e *clogging*, sendo, portanto, dispensada para a elaboração dos modelos as variáveis  $TpDI$ ,  $TpDI5$  e  $TpDI6$ .

Os dados de indicação da abertura percentual das válvulas  $AbDI5$  e  $AbDI6$  apresentam valores na faixa de 0% a 100% com *outliers* isolados. Valores que extrapolam os limites para abaixo de 0% e acima de 100% ocorrem com pequenas magnitudes, próximos a esses valores, portanto, foram ajustados para seus limites teóricos.

As medidas de nível do molde ( $Niv$ ) têm variação teórica e real na faixa de 0  $[mm]$  a 200  $[mm]$  e, conforme verificado na [Tabela 2](#), não possuem dados fora do limite estabelecido pelos especialistas do processo. Verifica-se também que seus valores estão, em grande parte, concentrados na faixa em torno de 100  $[mm]$ , com pouca variação em torno deste valor.

As medidas de velocidade do veio ( $Vel$ ) têm variação teórica na faixa de 0  $[m/min]$  a 2,5  $[m/min]$ , entretanto, há uma ocorrência elevada de medições negativas, provavelmente, devido aos intervalos de ciclos do processo ou por erros de saturação de sensores. Verifica-se também que grande parte dos seus valores se concentram na faixa entre 1,067  $[m/min]$  e 2,53  $[m/min]$ .

As variáveis  $Niv$  e  $Vel$  referem-se às medições realizadas em ambos os distribuidores, enquanto que as variáveis de abertura da válvula ( $AbDI5$  e  $AbDI6$ ) representam as medições em cada um dos distribuidores. A variável  $DI$  indica qual o distribuidor que está operando no momento e pode assumir dois valores mutuamente exclusivos, a saber, 5 ou 6. O valor 5 refere-se à operação do distribuidor 5 ( $DI5$ ), que representa 49% dos dados. Já o

valor 6, presente em 51% das amostras, indica os momentos em que o distribuidor 6 (*DI6*) está em funcionamento. Em decorrência da natureza destes dados, não foi possível realizar o cálculo dos percentis. O tratamento desses dados durante os períodos de transição é abordado em detalhes na seção seguinte.

## 4.2 O Tratamento do Conjunto de Dados

Após a seleção, a etapa de pré-processamento dos dados foi iniciada para o tratamento de *outliers* e amostras ausentes, garantindo a preservação das relações entre os atributos.

As ocorrências de *outliers* foram tratadas como erros de medição, uma vez que são casos específicos e, portanto, foram ajustadas de forma a torná-las consistentes. Para tanto, foram empregados valores previamente especificados do máximo e mínimo teóricos assumidos para cada uma das variáveis durante uma operação sem anomalias (ver [Tabela 2](#)).

Amostras com valores fora da faixa teórica foram consideradas *outliers*. Assim, se uma amostra apresenta valor abaixo do mínimo teórico, então ela deve ser ajustada para o seu valor mínimo teórico. Analogamente, se o valor assumido pelo atributo for maior do que o seu máximo teórico, a amostra deve possuir o seu máximo teórico.

Deseja-se identificar o *clogging* apenas em pontos onde não há troca entre os distribuidores, tendo em vista que esta se trata de uma situação transitória no processo. Nesses momentos, foram verificados pequenos períodos de interrupção do processo, representados pela ausência de informação que reflita no comportamento do fenômeno observado. A [Figura 31\(a\)](#) exemplifica um período de troca entre os distribuidores 6 e 5, que durou cerca de 1250 segundos (aproximadamente 20 minutos).

Em relação às amostras de abertura da válvula e velocidade do veio nos instantes de troca, é possível verificar que a velocidade cai abaixo do limite mínimo de operação e o processo é interrompido, refletindo no fechamento completo da válvula. Em consequência disto, essas amostras não possuirão relevância na tarefa de classificação. Nesses períodos também são verificados valores extremamente elevados da variável de nível, sendo então necessária a aplicação de um tratamento específico no conjunto de dados.

O tratamento dos dados consiste em eliminar as amostras referentes aos períodos de tempo em que não há operação de nenhum dos distribuidores e reagrupá-las através de uma interpolação do tipo “*spline*” (para a variável *Niv*), “*slinear*” (para as variáveis *AbDI* e *Vel*) e “*nearest*” (para a variável *DI*) compreendendo 30 pontos, dado que os sinais acham-se aproximadamente no mesmo patamar, visando garantir a continuidade e representatividade dos dados.

Tendo em vista que a variável *DI* indica o distribuidor em operação, as variáveis

*AbDI5* e *AbDI6*, que são mutuamente exclusivas, foram agrupadas formando uma única variável, a *AbDI*, cujo valor é igual a da abertura da válvula correspondente ao distribuidor em operação. A Figura 31(b) ilustra uma região de interpolação correspondente a um determinado período de troca de distribuidor, já a Figura 31(c) ilustra o resultado da interpolação.

Em relação aos dados ausentes fora dos períodos de troca de distribuidor, não foram identificadas ocorrências substanciais no período e nas variáveis analisadas. Ainda assim, os poucos períodos de ausência foram tratados por um filtro de mediana móvel<sup>1</sup>. Visando preservar as informações inerentes ao processo e levando-se em conta a capacidade das redes neurais em lidar com dados ruidosos, não foi aplicado nenhum tipo de filtragem para redução de ruídos ao conjunto de dados selecionado.

A Tabela 3 apresenta a análise descritiva das variáveis selecionadas no conjunto de entrada, após o processo de tratamento dos dados. A fim de simplificação, as TAG foram substituídas por uma nova nomenclatura, sendo representadas por  $x_r(nT_s)$ , em que  $r$  é o índice identificador de cada variável ( $r \in \{1, 2, 3, 4\}$ ) e  $T_s$  representa o período de amostragem de 1 segundo. Por simplicidade, será usada a notação  $x_r(n)$ .

Como pode ser melhor visualizado pelo gráfico *boxplot* da Figura 32(a), os valores de abertura da válvula  $x_1(n)$  estão, em grande parte, concentrados na faixa de 67% a 73%. Observa-se também que a mediana está centrada no *box*, o que demonstra a inexistência de assimetria na distribuição dos dados. Para a variável  $x_1(n)$ , a faixa interquartil compreende valores mínimos e máximos aproximados de 58% e 83%, respectivamente. É importante salientar que os valores fora da faixa na Figura 32(a), mesmo conceitualmente considerados *outliers*, eles informam a operabilidade da planta, visto que em aberturas inferiores a 60% não é possível detectar *clogging* (THOMAS; BAI, 2001).

A Figura 32(b) apresenta o gráfico *boxplot* da variável de nível do molde, representada por  $x_2(n)$ . Observa-se que a maior parte dos valores se concentram em torno de 100 [mm] e não há valores próximos de zero. Este resultado já era esperado uma vez que,

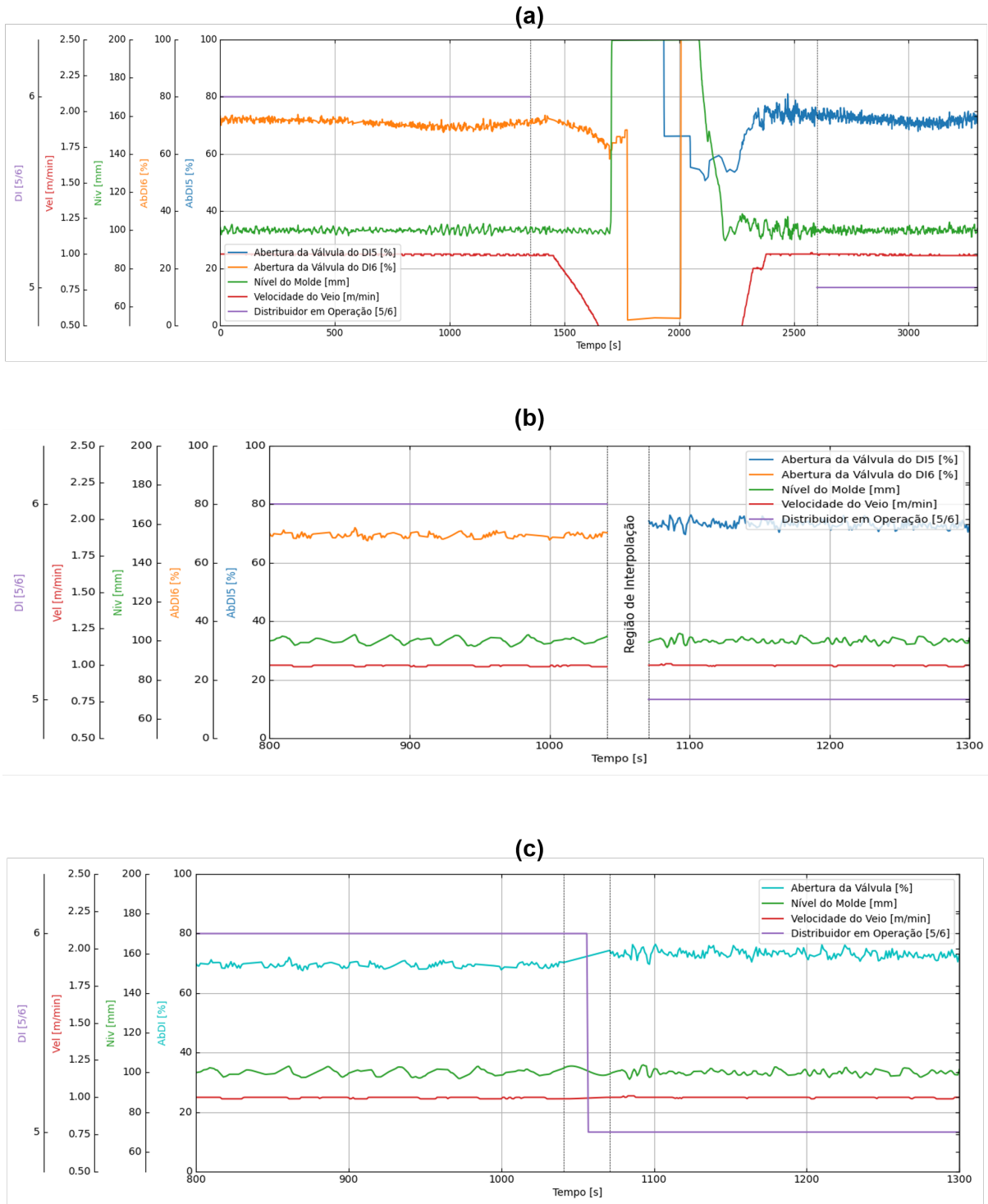
<sup>1</sup> A mediana móvel é menos sensível à presença de *outliers* do que o filtro de média móvel porque calcula o valor central, ignorando os extremos que poderiam distorcer a média. Isso garante uma resposta mais robusta e menos influenciada por dados atípicos, sendo uma escolha mais adequada em situações onde *outliers* podem afetar significativamente os resultados.

Tabela 3 – Seleção final das variáveis de entrada após o tratamento dos dados.

| TAG Anterior | Nova TAG | Descrição                | Valor Mínimo | Valor Máximo | Percentil |        |        | Média  | Desvio Padrão |
|--------------|----------|--------------------------|--------------|--------------|-----------|--------|--------|--------|---------------|
|              |          |                          |              |              | 25 [%]    | 50 [%] | 75 [%] |        |               |
| <i>AbDI</i>  | $x_1(n)$ | Abertura da Válvula      | 0            | 100          | 67,52     | 70,87  | 73,77  | 70,63  | 4,75          |
| <i>Niv</i>   | $x_2(n)$ | Nível do molde           | 34,16        | 200          | 99,38     | 100,01 | 100,65 | 100,78 | 8,72          |
| <i>Vel</i>   | $x_3(n)$ | Velocidade do veio       | 0            | 1,71         | 1,00      | 1,18   | 1,35   | 1,18   | 0,022         |
| <i>DI</i>    | $x_4(n)$ | Distribuidor em Operação | 5            | 6            | -         | -      | -      | -      | -             |

Fonte: Elaborada pela autora.

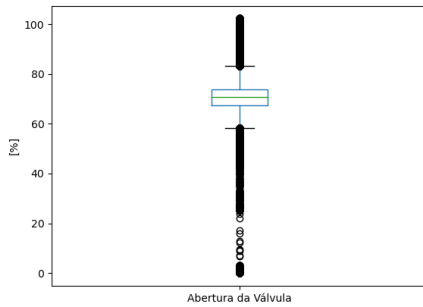
Figura 31 – Processo de interpolação do período de troca de distribuidor: (a) Identificação do período de troca de distribuidor; (b) Região de interpolação (aproximação); (c) Interpolação e agrupamento das variáveis de abertura da válvula (aproximação).



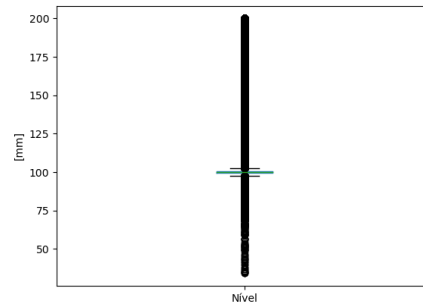
Fonte: Elaborada pela autora.

em condições de operação normal, o molde sempre estará com o nível acima de zero. Os valores fora da faixa interquartil, embora conceitualmente sejam considerados *outliers*,

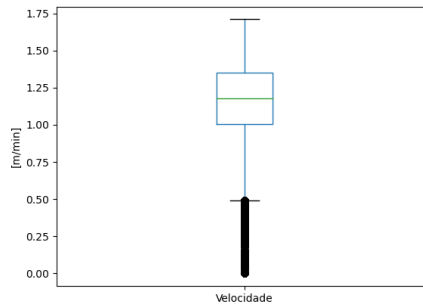
estão dentro da faixa operacional da variável.



(a)  $x_1(n)$ : abertura da válvula.



(b)  $x_2(n)$ : nível do molde.



(c)  $x_3(n)$ : velocidade do veio.

Figura 32 – Gráficos *boxplot* das variáveis  $x_1(n)$ ,  $x_2(n)$  e  $x_3(n)$  correspondentes à abertura da válvula, nível do molde e velocidade do veio, respectivamente.

Na Figura 32(c), a qual representa o *boxplot* da variável  $x_3(n)$ , correspondente à velocidade do veio, observam-se alguns pontos situados fora da faixa interquartil inferior. Esses pontos correspondem às possíveis reduções de velocidade para manutenção e/ou troca de equipamentos devido à ocorrência de *clogging*. Observa-se ainda que, após o tratamento dos dados, os seus valores reais não atingem o limiar teórico máximo e se concentram em torno da faixa de 1 [m/min] a 1,35 [m/min].

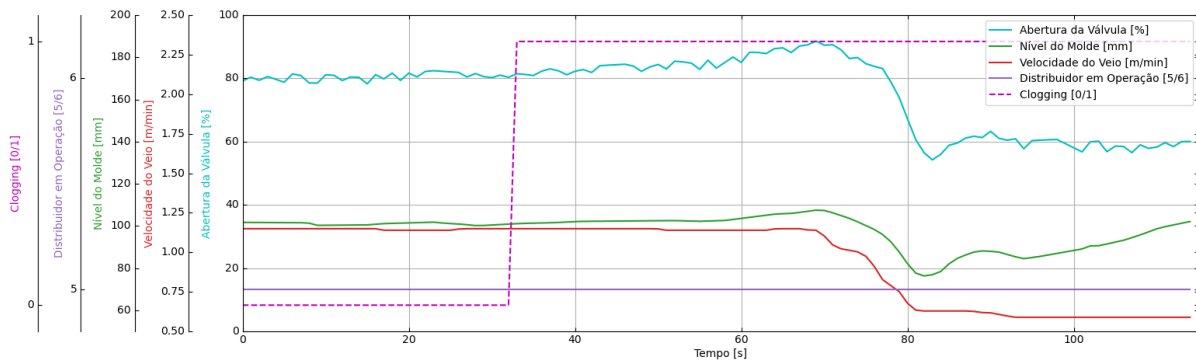
## 4.3 Exemplos de Ocorrência de Clogging no Conjunto de Dados

Nesta seção, são apresentados exemplos didáticos de ocorrências de *clogging* no conjunto de dados. Esses casos ilustrativos oferecem informações valiosas sobre como o fenômeno se desenvolve e afeta as variáveis envolvidas.

### 4.3.1 Caso 1

A Figura 33 ilustra um exemplo didático do início do *clogging* ao longo do processo correlacionando as quatro variáveis selecionadas. Partindo do instante 32, observa-se um

Figura 33 – Exemplo didático de ocorrência inicial do *clogging* observada a partir das variáveis do processo selecionadas após o tratamento.



Fonte: Elaborada pela autora.

aumento gradativo da abertura da válvula, à velocidade constante do veio, sem aumento significativo do nível do molde. Em decorrência da redução da seção da válvula causada pela obstrução de material depositante característico da formação do *clogging*, observa-se a ocorrência de um pequeno aumento do nível um pouco antes da velocidade do veio ser reduzida. O fechamento da válvula proporciona uma redução da velocidade do veio e quando a velocidade atinge o patamar inferior a  $0,6 \text{ [m/min]}$  (no instante aproximado de 80 segundos), o nível passa a subir.

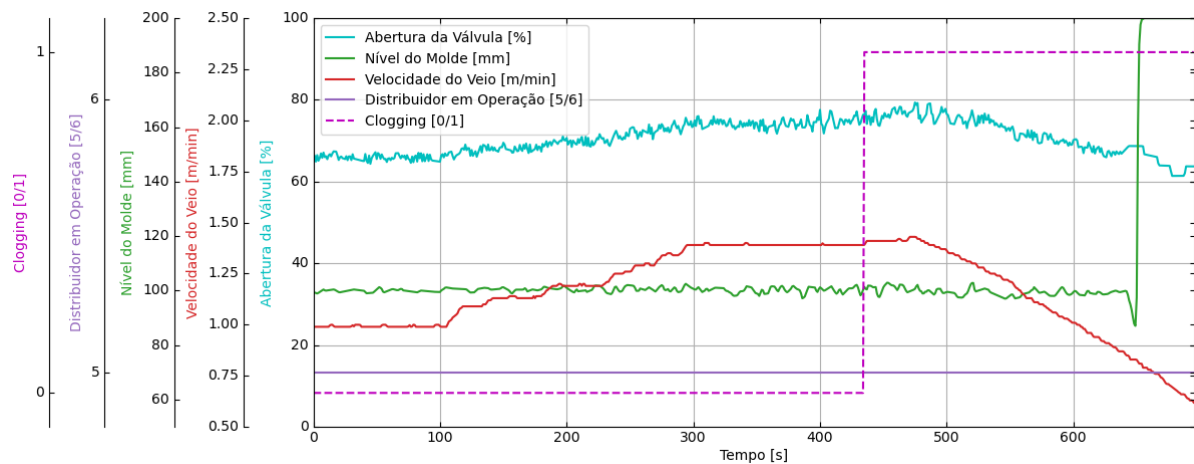
É interessante salientar que, embora as taxas de variação das variáveis apresentadas na Figura 33 possam indicar a ocorrência de *clogging*, o tipo de aço, por exemplo, pode alterar esses padrões. Além disto, as ações de controle podem estar presentes nas medições, o que pode comprometer a integridade do uso de regras triviais para a identificação do *clogging* inicial. O *clogging* também pode oscilar o nível ou apenas a abertura da válvula e, por isso, em diversos casos, não é possível a sua identificação visual.

#### 4.3.2 Caso 2

Com base na Figura 33, é esperado que na ocorrência de *clogging* ocorra inicialmente uma abertura da válvula sem que haja redução do nível do molde ou alteração na velocidade do veio. A Figura 34, contudo, exemplifica uma ocorrência inicial do *clogging* em que o mecanismo fenomenológico não pode ser percebido.

A partir do instante 100, ocorre um pequeno aumento gradual da abertura da válvula, sem uma variação média do nível do molde, possivelmente devido ao aumento da velocidade. Nota-se, a partir do instante 475, que a velocidade do veio inicia uma redução do valor de  $1,4 \text{ [m/s]}$  para  $0,65 \text{ [m/s]}$  em torno de 180 [s]. A válvula submersa, que inicialmente apresentava uma abertura de cerca de 78%, passou a iniciar o seu fechamento, sem que houvesse qualquer percepção de alteração no nível do molde. Neste caso, o *clogging* é identificado pois a taxa com que ocorre a redução da velocidade é superior à taxa do

Figura 34 – Exemplo didático do início do *clogging* sem que houvesse qualquer percepção de alteração no nível do molde.



Fonte: Elaborada pela autora.

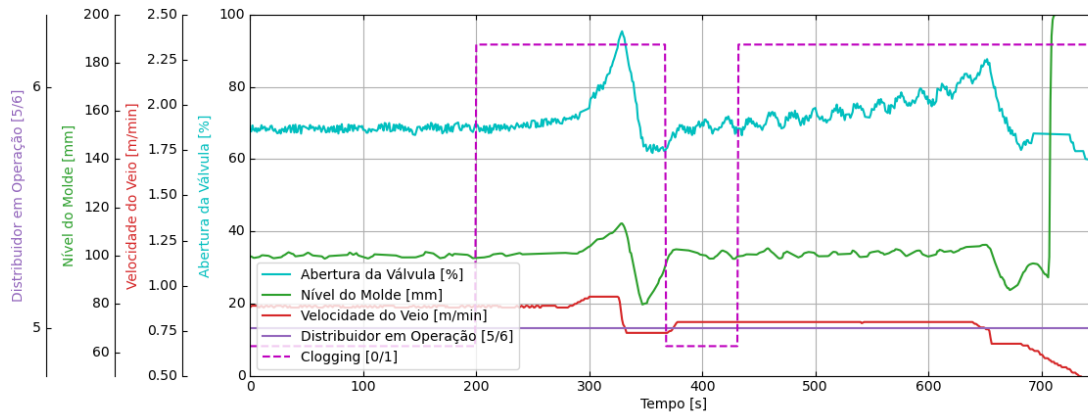
fechamento da válvula, indicando então uma deposição de material obstrutor. A partir do instante 647, o nível do molde inicia uma subida abrupta, atingindo o seu limiar de saturação de 200 [cm].

### 4.3.3 Caso 3

Já a [Figura 35](#) exemplifica uma ocorrência de *clogging* em que, partindo do instante 200, observa-se um aumento gradativo da abertura da válvula, à velocidade constante do veio, sem aumento do nível do molde. Verifica-se ainda que ocorre um pequeno aumento do nível um pouco antes da velocidade do veio ser reduzida. A redução da velocidade do veio proporciona o fechamento da válvula. Após esta ocorrência, o sistema volta a operar com velocidade reduzida em torno de 0,75 [m/min] e o nível volta ao patamar médio de 100 [mm], possivelmente em virtude de uma ação de controle. Este momento pode ser caracterizado como *unclogging*<sup>2</sup>. No instante 430, entretanto, ocorre novamente um aumento da abertura válvula, desta vez, com uma oscilação mais evidente. O nível e a velocidade permanecem constantes até que, no instante 638, a velocidade começa a decair, seguido de um pequeno aumento do nível do molde. No instante 650, a abertura da válvula atinge um novo pico que é seguido de uma redução gradual da velocidade, promovido pelo fechamento da válvula devido à obstrução, fazendo com que o nível atinja o seu limiar de saturação.

<sup>2</sup> Alguns autores definem o *unclogging* como o mecanismo de desprendimento abrupto do material obstrutor acumulado, em geral, nos casos de *clogging* causado pela deposição de produtos da reação e/ou da desoxidação (THOMAS; BAI, 2001). Conforme explicado por Rackers (1995) e Bolger (1994), a alumina acumulada, por exemplo, pode ser desprendida através da adição de cálcio ou ainda pela injeção de argônio. O *unclogging* ocasiona uma desobstrução da válvula, onde o fluxo é liberado, sendo necessária uma medida rápida de controle por parte do operador para que o processo possa ser retomado às condições anteriores de maneira eficaz (JABRI et al., 2008; GONZÁLEZ-YERO et al.,

Figura 35 – Exemplo didático do início de duas ocorrências de *clogging* entre a ocorrência do *unclogging*.



Fonte: Elaborada pela autora.

#### 4.3.4 Caso 4

Uma outra manifestação inicial do *clogging* pode ser exemplificada pela Figura 36 em que é observado um aumento na amplitude de oscilação de abertura da válvula, seguida de uma oscilação no nível do molde. Nota-se que em todo o percurso a oscilação na válvula ganha amplitude com o tempo, influenciando no nível do molde. Em aproximadamente até o instante 400 a oscilação da válvula é manifestada para uma mesma velocidade do veio e mesmo patamar de nível. A variação de amplitude do nível se tornam mais acentuada a partir do instante 400. No instante 688, ocorre uma redução da velocidade do veio e a amplitude demasiada de oscilação da válvula é compensada. A diminuição da velocidade do veio foi usada como artifício para tentar reduzir essas oscilações.

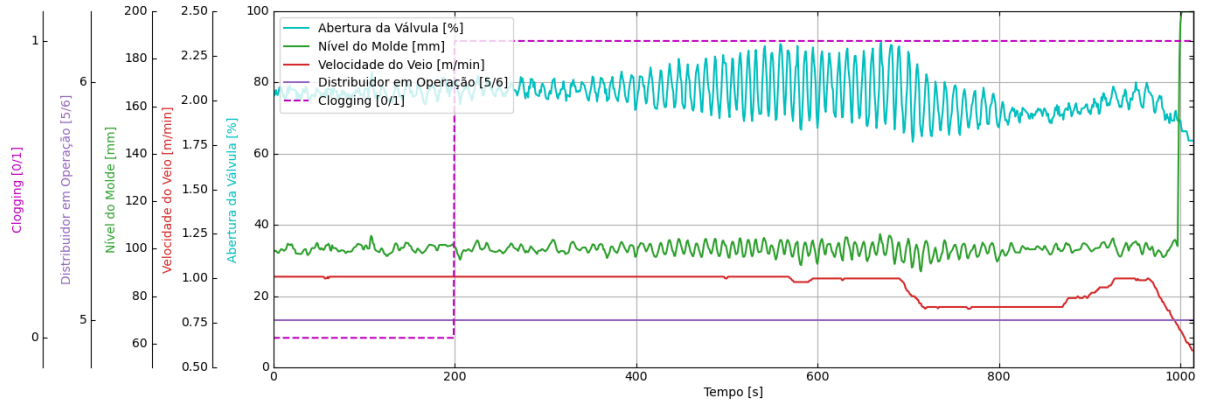
De modo geral, quando surgem as oscilações de grande amplitude na válvula e no nível do molde e, conseqüente redução da velocidade do veio, pouco tempo depois o processo é interrompido para a substituição da válvula, podendo haver uma extensão do tempo quando há o procedimento de injeção de argônio. No caso exemplificado, a partir do instante 870, a válvula novamente volta a abrir e com um aumento da velocidade do veio, que não reflete em variações abruptas do nível. A partir do instante 963, inicia-se o processo de redução da abertura da válvula e a velocidade decai, refletindo no colapso

---

2020; PEREIRA et al., 2021).

Cabe destacar que não é escopo deste trabalho a identificação do *unclogging*, uma vez que, para a ocorrência deste fenômeno é necessário que se tenha ocorrido ao menos o início de uma obstrução da SEN. Outro fator é que o *unclogging* não está associado a todos os tipo de *clogging* (RACKERS, 1995; JABRI et al., 2008; GONZÁLEZ-YERO et al., 2020; PEREIRA et al., 2021) e, portanto, sua ocorrência não é generalizada. O intuito deste estudo é identificar exclusivamente a ocorrência inicial de uma obstrução, seja ela originada pela deposição de produtos da reação e/ou aço solidificado ou ainda pela junção de outros possíveis fatores (RACKERS, 1995; RACKERS; THOMAS, 1995; THOMAS; BAI, 2001; SAHAI; EMI, 2008; YANG; OLIA; THOMAS, 2021).

Figura 36 – Exemplo didático do início do *clogging* após constantes oscilações das variáveis de abertura da válvula e nível do molde.



Fonte: Elaborada pela autora.

com a subida abrupta do nível do molde no instante 996.

A complexidade deste trabalho, portanto, está na estimação do momento inicial do fenômeno de *clogging*, antes mesmo que as variáveis sejam capazes de refletir visualmente um comportamento anômalo do sistema.

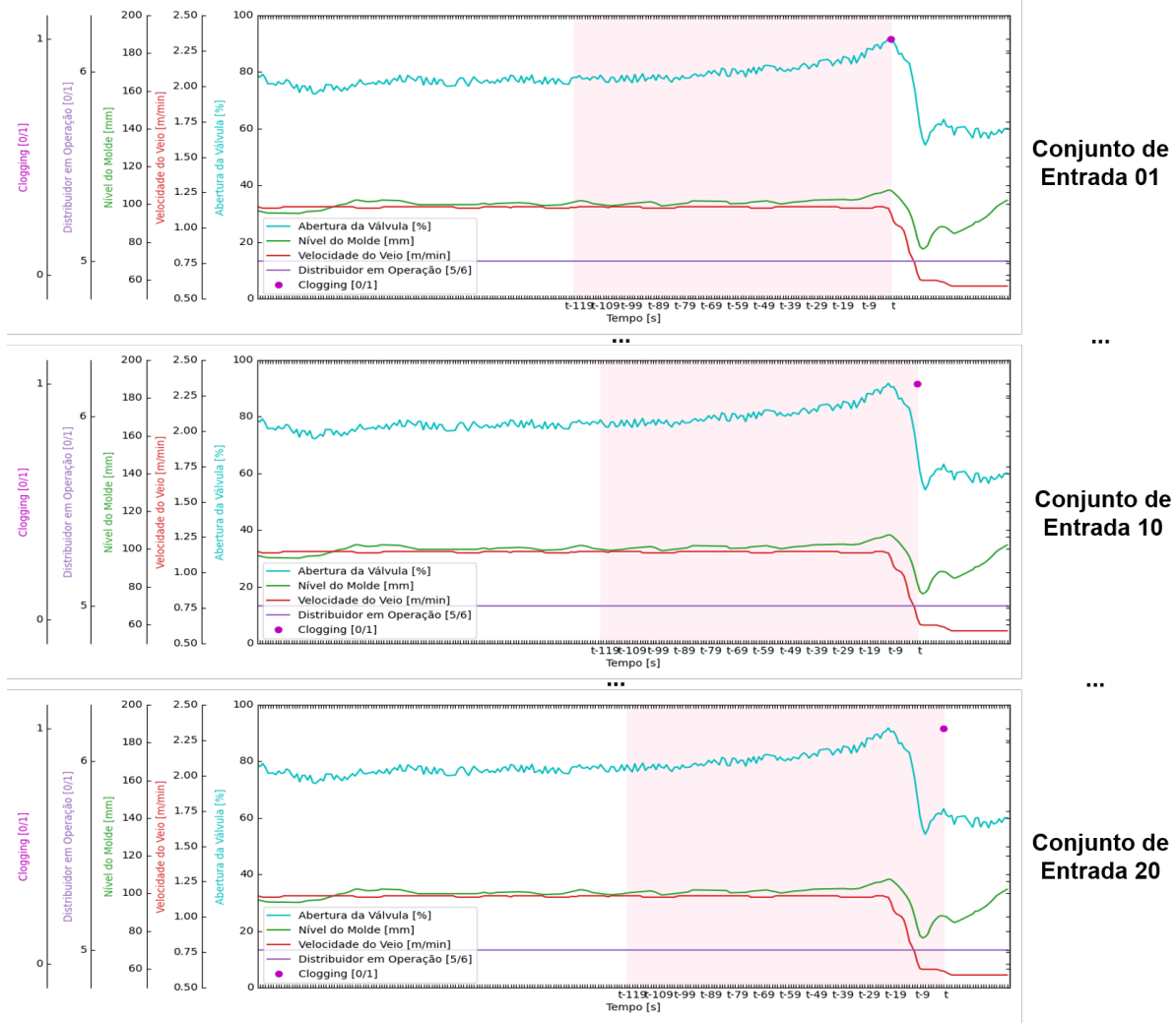
## 4.4 A Formação das Matrizes de Entradas

Em conformidade com as constatações de Barati et al. (2018), o *clogging* só é verdadeiramente percebido na seção do SEN entre 100 e 150 segundos, em proporção às dimensões do SEN. Neste contexto, levando-se em consideração que o processo de formação de *clogging* é o reflexo do comportamento passado das variáveis ao longo do tempo, é necessário organizar a entrada dos dados de forma adequada para a aplicação de modelos de detecção de eventos. O resultado é uma matriz de entradas agrupadas por meio de uma janela deslizante cobrindo 120 amostras sobrepostas com um passo de 1 amostra por *timestep*, isto é, 1 segundo por *timestep*.

Cabe destacar que, durante os experimentos, foram testados diferentes tamanhos de janela deslizante para avaliar seu impacto no desempenho do modelo. Em particular, ao testar com janelas menores, observou-se que os resultados, especialmente nos modelos LSTM, foram inferiores. Isso indica que um tamanho de janela maior pode ser necessário para capturar de maneira mais eficaz as dependências temporais dos dados, garantindo, assim, um desempenho superior do modelo.

A Figura 37 ilustra o processo da janela deslizante para a formação de 20 conjuntos de entrada. Seguindo esta análise, dada uma saída,  $y(n)$ , correspondente a uma das classes binárias da amostra  $n$ , ela estará relacionada a uma matriz de entrada,  $\mathbf{X}_{(4 \times 120)}(n)$ , dada

Figura 37 – Relação entre entradas deslocadas no tempo e a saída.



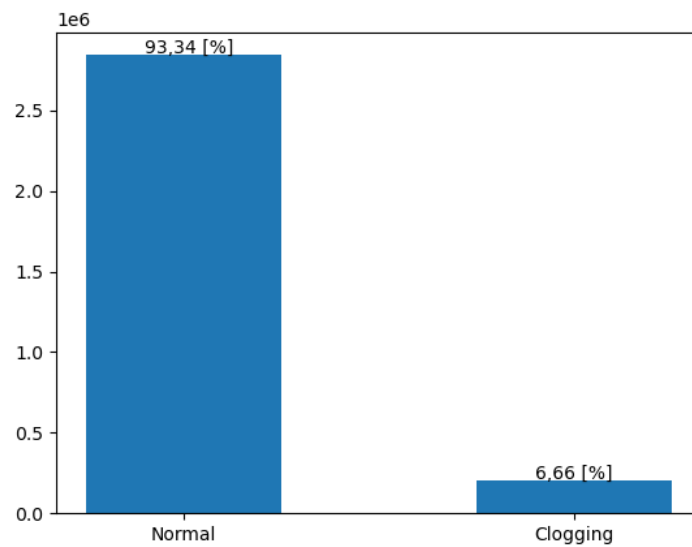
Fonte: Elaborada pela autora.

pela Equação 4.1,

$$\mathbf{X}_{(4 \times 120)}(n) = \begin{Bmatrix} x_1(n-119) & x_1(n-118) & \dots & x_1(n-2) & x_1(n-1) & x_1(n) \\ x_2(n-119) & x_2(n-118) & \dots & x_2(n-2) & x_2(n-1) & x_2(n) \\ x_3(n-119) & x_3(n-118) & \dots & x_3(n-2) & x_3(n-1) & x_3(n) \\ x_4(n-119) & x_4(n-118) & \dots & x_4(n-2) & x_4(n-1) & x_4(n) \end{Bmatrix} \quad (4.1)$$

contendo, em cada linha, uma das quatro variáveis selecionadas com amostras atrasadas em etapas de 120 segundos (de  $n - 119$  a  $n$ ). Assim, cada matriz formada será rotulada com uma das classes, isto é, com o evento que está ocorrendo no instante  $n$  (evento de *clogging* ou de operação normal, representados pelas saídas 1 e 0, respectivamente). A janela deslizante se move ao longo da série para que, a cada passo, uma nova matriz de entrada e sua respectiva classe sejam geradas. Como resultado, foram obtidas um total de 12.859.200 matrizes de entrada, isto é, 3.572 horas.

Figura 38 – Histograma das classes do conjunto de dados.



Fonte: Elaborada pela autora.

## 4.5 A Organização dos Dados

Para a aplicação em algoritmos de aprendizagem de máquinas, é importante que o conjunto de dados esteja organizado em treinamento, validação e teste. A princípio, 80% dos dados foram reservados para o treinamento, 10% para validação e 10% para os testes. Vale ressaltar que a sequência original dos dados foi mantida sem embaralhamento, de forma que os primeiros dados foram reservados para treinamento, os dados posteriores para validação e os dados mais recentes para teste. Além disso, foi necessário aplicar um balanceamento entre as classes e normalizar das variáveis de entrada. Estes procedimentos para a formação final dos conjuntos de treinamento, validação e teste são expostos a seguir.

### 4.5.1 O Balanceamento dos Dados

Conforme apresentado pelo histograma da [Figura 38](#), no conjunto de dados há um certo desequilíbrio entre as classes de *clogging* e operação normal, o que já era esperado para este tipo de problema, uma vez que o *clogging* é um evento classificado como uma anomalia.

Observa-se através da [Figura 38](#) que 6,66% das amostras referem-se à ocorrência do evento em estudo. Dada a quantidade de dados disponíveis, cerca de 3.572 horas, as matrizes de entrada puderam ser selecionadas de modo a equilibrar as classes no conjunto de treinamento. Este processo, chamado de balanceamento de classes, foi conduzido com o cuidado de assegurar que as informações relevantes para a modelagem não fossem perdidas.

Uma forma de contornar o problema do desbalanceamento entre as classes consiste

em manipular a quantidade de dados que serão efetivamente aplicados ao modelo, igualando-se o número de observações entre as classes. Um dos métodos utilizados é o *undersampling*, que consiste na redução do número de observações da classe majoritária, mantendo a quantidade de amostras da classe minoritária, de modo a equalizar as classes (JAPKOWICZ, 2000; TYAGI; REDDY et al., 2019).

Considerando que a condição de operação normal (classe majoritária) é representada por uma grande quantidade de operações do processo, cujo comportamento é repetitivo, a exclusão dessas amostras não representa significativas perdas de informação. Além de evitar a redundância, o *undersampling* tem como vantagem a redução do armazenamento dos dados (TYAGI; REDDY et al., 2019; BAGUI; LI, 2021).

É importante considerar, no entanto, que a decisão de aplicar o *undersampling* deve ser cuidadosamente avaliada para garantir que não se perca dados relevantes que poderiam contribuir para a capacidade do modelo de tomar decisões precisas. Seguindo o procedimento adotado em Diniz et al. (2024)<sup>3</sup>, o processo de balanceamento das classes do conjunto de treinamento foi conduzido de forma a manter a sequência temporal das amostras. A partir de uma inspeção visual, foram selecionadas amostras do conjunto para garantir a ordem cronológica dos dados durante o balanceamento. Dessa maneira, períodos extensos de operação normal, caracterizados por comportamento repetitivo, foram removidos. Isso reduziu a redundância e preservou a representatividade do conjunto de dados de treinamento.

#### 4.5.2 A Divisão do Conjunto de Dados

Após a fase de balanceamento no conjunto de treinamento, 162.317 amostras de cada uma das classes são mantidas no conjunto de treinamento, totalizando 324.634 amostras, equivalente à 90,17 horas. Para manter a proporção dos conjuntos de dados, são selecionadas 40.579 amostras (cerca de 11,27 horas) tanto para a fase de validação quanto para a fase de teste em cada uma das cinco rodadas de treinamento da rede.

A Tabela 4 demonstra que a divisão inicial do conjunto de dados será mantida, de modo que 80% dos dados serão utilizados para o treinamento, 10% serão utilizados na validação e os 10% restantes estarão reservados para a fase de testes.

#### 4.5.3 A Normalização dos Dados

A partir da média e do desvio padrão de cada variável (com exceção da variável binária *DI*), calculados a partir do conjunto de dados de treinamento, é realizada a padronização dos dados. Através da Equação 4.2, a técnica de padronização, também

<sup>3</sup> Mais detalhes em: <<https://github.com/apaulinhamd/Deep-Neural-Networks-For-Clogging-Detection>>.

Tabela 4 – Divisão do conjunto de dados e quantidade de amostras por classe.

| Conjunto    | Divisão [%] | Total de Amostras | Representação por Classe |              |         |            |
|-------------|-------------|-------------------|--------------------------|--------------|---------|------------|
|             |             |                   | Clogging                 | Clogging [%] | Normal  | Normal [%] |
| Treinamento | 80          | 324.634           | 162.317                  | 50           | 162.317 | 50         |
| Validação   | 10          | 40.579            | 2703                     | 6,66         | 37.876  | 93,34      |
| Teste       | 10          | 40.579            | 2703                     | 6,66         | 37.876  | 93,34      |

Fonte: Elaborada pela autora.

chamada de *z-score*, permite que as variáveis possuam média zero e desvio padrão unitário (SKIENA, 2017),

$$x_{rz}(n) = \frac{x_r(n) - \mu_r}{\sigma_r}, \quad (4.2)$$

em que  $x_r(n)$  representa a variável a ser padronizada,  $r$  é o índice identificador de cada variável ( $r \in \{1, 2, 3, 4\}$ ),  $\mu_r$  é a média da variável  $x_r(n)$  e  $\sigma_r$  é o respectivo desvio padrão, ambos calculados sobre o conjunto de treinamento.

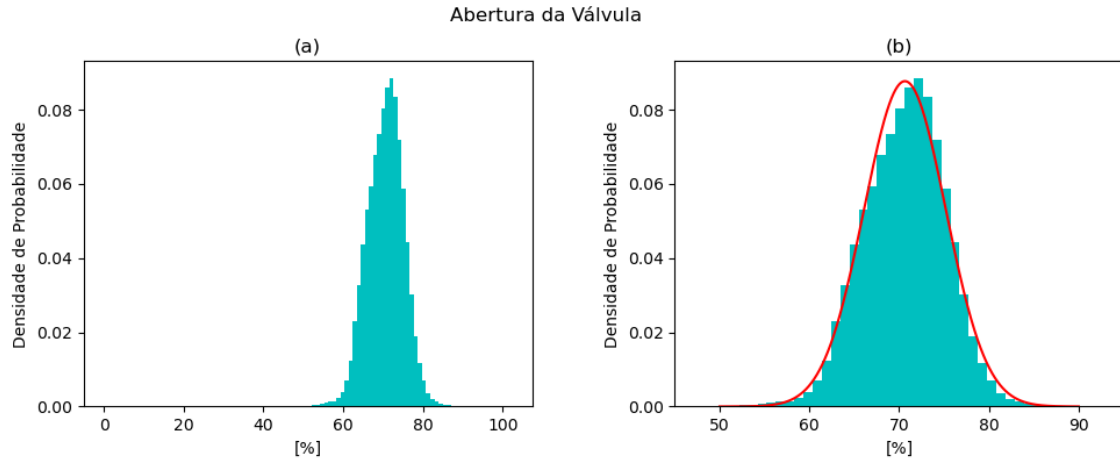
Em termos geométricos, a padronização por *z-score* translada a distribuição dos dados originais para um ponto de origem, comprimindo ou expandindo a distribuição das amostras, sem alterar a forma em que os dados originais estão distribuídos. Em análise multivariável, além disto, a padronização é preferível por permitir que as unidades de medidas se tornem comparáveis (CARVALHO et al., 2011; SKIENA, 2017; DZIERŻAK, 2019; GADEKALLU et al., 2020).

Outra característica da padronização está no fato dela ser menos afetada pelos *outliers*, dado que não há um intervalo predefinido dos valores transformados (CARVALHO et al., 2011; SKIENA, 2017). Por meio da Figura 32 é possível observar a existência de *outliers* conceituais nas séries de dados, o que apoia a escolha da padronização por *z-score*.

A aplicação da padronização por *z-score* é indicada principalmente nos casos em que os dados seguem uma distribuição gaussiana (CARVALHO et al., 2011; DZIERŻAK, 2019; GADEKALLU et al., 2020). Nesse contexto, uma análise visual dos dados<sup>4</sup>, incluindo a observação dos histogramas das variáveis de entrada  $x_1(n)$  e  $x_2(n)$ , revelou uma certa semelhança entre os dados observados e a distribuição gaussiana, embora possam apresentar algumas diferenças sutis em relação à distribuição normal padrão. As Figuras 39 e 40 ilustram esses histogramas, respectivamente, enquanto que a Tabela 5 apresenta as estatísticas descritivas das variáveis  $x_1(n)$ ,  $x_2(n)$  e  $x_3(n)$ .

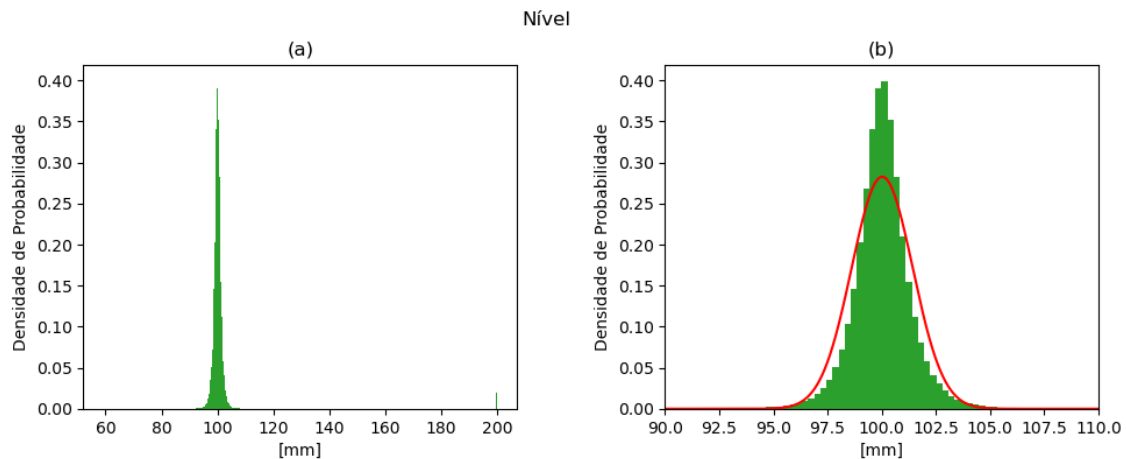
<sup>4</sup> A análise estatística realizada nesta pesquisa considerou a aplicação de testes de normalidade (GHASEMI; ZAHEDIASL, 2012; MIOT, 2017). No entanto, é de extrema importância destacar o papel significativo que o tamanho considerável da amostra desempenha na interpretação dos resultados desses testes de normalidade. Em conjuntos de dados extensos, essas análises demonstram alta sensibilidade, muitas vezes identificando pequenas discrepâncias da normalidade como estatisticamente relevantes, mesmo quando as evidências visuais sugerem o contrário. A abordagem adotada nesta

Figura 39 – (a) Histograma da variável  $x_1(n)$  referente à abertura da válvula (existem *outliers* imperceptíveis nas faixas de 0 a 40 e 85 a 100); (b) Destaque para a distribuição da variável  $x_1(n)$  desprezando os *outliers* conceituais, acompanhada por uma representação gaussiana baseada na média e no desvio padrão.



Fonte: Elaborada pela autora.

Figura 40 – (a) Histograma da variável  $x_2(n)$  referente ao nível do molde, sendo possível observar pontos de *outliers* conceituais; (b) Destaque para a distribuição da variável  $x_2(n)$  desprezando os *outliers* conceituais, acompanhada por uma representação gaussiana baseada na média e no desvio padrão.



Fonte: Elaborada pela autora.

Além disso, os resultados das estatísticas descritivas oferecem observações valiosas sobre as características de cada variável de entrada. Na variável  $x_1(n)$ , a proximidade entre a média e a mediana denota notável simetria nos dados, com um desvio padrão baixo, indicando dispersão reduzida. A assimetria levemente negativa sugere uma cauda esquerda

---

pesquisa valorizou a análise visual abrangente e o uso das estatísticas descritivas como métodos de interpretação robusta dos dados. Essa estratégia considera cuidadosamente a natureza da pesquisa e o tamanho da amostra, evitando interpretações excessivamente sensíveis dos testes de normalidade quando aplicados a conjuntos de dados extensos, embora seja importante reconhecer a possibilidade de sutis discrepâncias em relação à distribuição normal padrão.

Tabela 5 – Estatística descritivas das variáveis  $x_1(n)$ ,  $x_2(n)$  e  $x_3(n)$ .

| Variável | Média  | Mediana | Desvio Padrão | Assimetria | Curtose |
|----------|--------|---------|---------------|------------|---------|
| $x_1(n)$ | 70,63  | 70,87   | 4,75          | -0,15      | 0,31    |
| $x_2(n)$ | 100,78 | 100,01  | 8,72          | -0,14      | 2,81    |
| $x_3(n)$ | 1,18   | 1,18    | 0,022         | -0,37      | 1,59    |

Fonte: Elaborada pela autora.

ligeiramente mais longa, enquanto a curtose próxima de zero sugere semelhança com a distribuição gaussiana.

No que diz respeito à variável  $x_2(n)$ , também é observada a proximidade entre a média e a mediana, bem como um desvio padrão moderado, indicando dispersão controlada. A assimetria próxima de zero sugere uma distribuição praticamente simétrica, com caudas equilibradas. Além disso, a curtose é positiva, o que indica que as caudas da distribuição são mais concentradas do que as de uma distribuição normal, embora a diferença não seja muito significativa.

Em relação à distribuição da variável  $x_3(n)$ , conforme representada na [Figura 41](#), nota-se que a distribuição não exibe simetria perfeita e demonstra uma cauda esquerda ligeiramente mais longa. A curtose, por sua vez, indica que a distribuição é um pouco mais concentrada do que uma gaussiana.

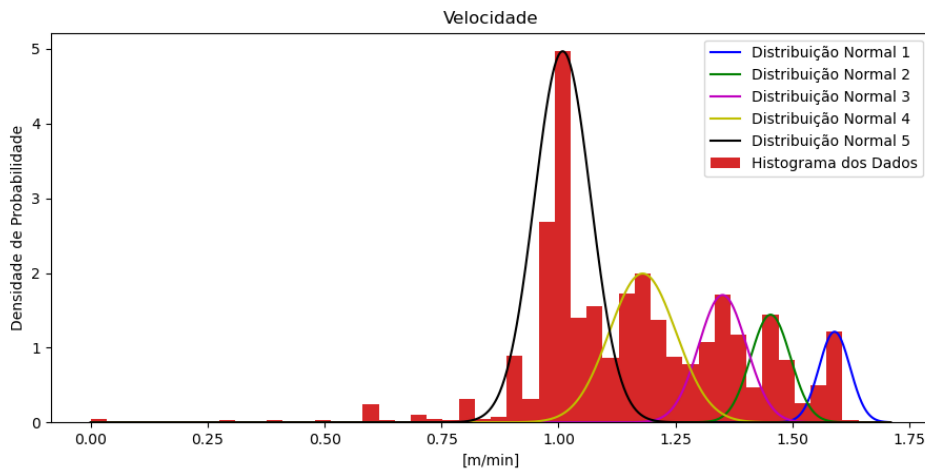
A presença de uma distribuição normal nos dados da variável  $x_3(n)$  não é aparente, notando-se a presença de picos distintos. Esses picos podem ser considerados como representações individuais de distribuições gaussianas, contribuindo para a formação de uma distribuição assimétrica negativa. A complexidade desses picos reflete a diversidade do processo em análise.

No contexto da produção de aço, a velocidade do veio desempenha um papel crítico na determinação das características físico-químicas desejáveis do produto final. Além disso, o tempo de resfriamento, que afeta diretamente a qualidade do aço, é fortemente condicionado pela velocidade do veio. Em conformidade com o que é elucidado por [Yang et al. \(2021\)](#), devido à crescente dinâmica e competição do mercado, os processos industriais geralmente apresentam frequentes transições em seus modos de operação.

Em análise da [Figura 41](#), os dados revelam que, embora uma parte significativa do processo mantenha a velocidade próxima de 1 [m/min], que também é a moda, observam-se concentrações distintas em torno de outros valores de velocidade, como 1,18 [m/min] (mediana), 1,36 [m/min], 1,45 [m/min] e 1,59 [m/min].

A [Figura 42](#) ilustra a distribuição das variáveis  $x_1(n)$  ([Figura 42 \(a\)](#)),  $x_2(n)$  ([Figura 42 \(b\)](#)) e  $x_3(n)$  ([Figura 42 \(c\)](#)) após a aplicação da padronização por *z-score* aos dados. A partir da análise visual desta figura, constata-se que de fato a forma da distribuição

Figura 41 – Distribuição de Velocidade do Veio ( $x_3(n)$ ) - Análise de Histograma e Representação de Gaussianas.



Fonte: Elaborada pela autora.

dos dados foi mantida, bem como seus *outliers* conceituais. Outro ponto que deve ser destacado é que a padronização permitiu que a amplitude dos sinais fosse ajustada para uma faixa de valores mais equitativa.

## 4.6 Os Classificadores

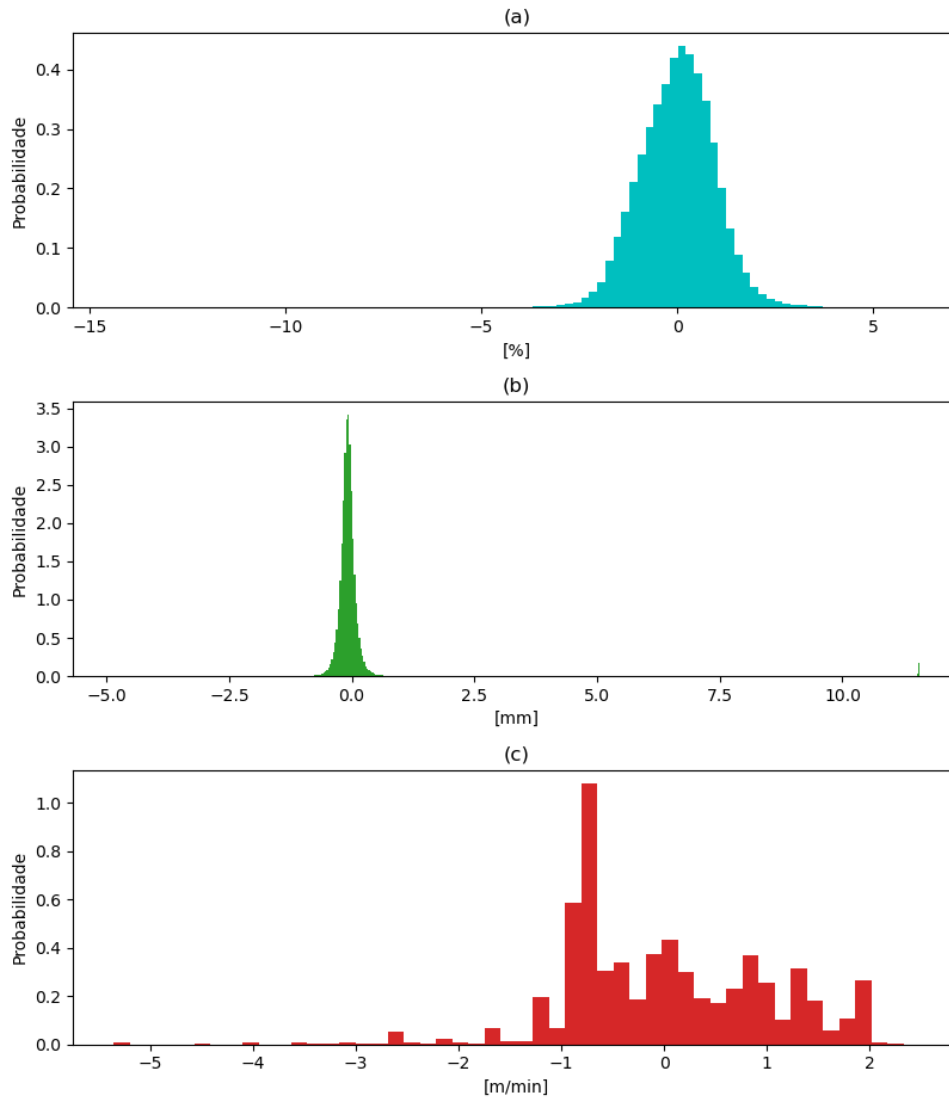
Neste trabalho, é realizada uma comparação entre o desempenho geral de classificadores utilizando as LSTM e sua variante, a ConvLSTM. Além disto, é avaliado o desempenho de uma abordagem híbrida, envolvendo CNN-1D em conjunto com LSTM, denominada CNN-LSTM. Para avaliar a eficiência dos modelos propostos, os modelos MLP foram utilizados como linha de base para comparação.

As simulações foram implementadas em um computador pessoal em ambiente Python 3.6, 64 bit, 16 GB of RAM, Intel(R) Core(TM) i7-9750H CPU @ 2.60GHz com GPU NVIDIA GeForce RTX 2060 6 GB.

Para fins comparativos, inicialmente foram realizados testes preliminares com diversas topologias construtivas de cada rede, definidas pelo método de tentativa e erro, considerando a quantidade de camadas, número de células das camadas LSTM, quantidade e tamanho dos filtros das camadas convolucionais, quantidade de unidades das camadas FC, além da utilização do *MaxPooling* e do *dropout*, a fim de se verificar como cada um dos parâmetros variantes poderiam influenciar no desempenho da rede. Esta fase de testes preliminares durou cerca de 2 meses de processamento dedicado com uso de GPU (as configurações do computador são apresentadas na Seção 4.7).

A partir da avaliação preliminar, foi definido que nas redes seriam aplicadas duas

Figura 42 – Histograma das variáveis após a aplicação da padronização por *z-score*: (a) Abertura da válvula; (b) Nível do molde; (c) Velocidade do veio.



Fonte: Elaborada pela autora.

camadas de extração de características temporais e/ou espaciais, com número de células ou filtros (dependendo do modelo) variando de 8 a 256 em incrementos de oito. Além disso, em todos os modelos, o sinal foi repassado à uma camada FC contendo 200 unidades e função de ativação ReLU, concatenada à uma camada *Softmax* com 2 unidades para gerar a saída de classificação para as séries de entrada como *clogging* ou normal. Dessa forma, é garantida que a diferença entre os desempenhos dos modelos seja resultante da distinção nas arquiteturas e não devido às melhores técnicas de otimização da rede, por exemplo.

O conjunto de dados de entrada é aplicado às redes usando lotes compostos por 1.200 amostras (*Batch Size*), entendendo amostra como uma sequência retornada pela janela deslizante. Foi escolhido um limite de 200 épocas para o treinamento, contudo, ele pode ser interrompido pela parada precoce (*Early Stopping*) com *Patience*<sup>5</sup> de 6. A

atualização dos pesos das redes (otimizador) é feita através do *Adam Algorithm*, com taxa de aprendizado de 0,0001, e a *Cross-Entropy* é aplicada como função de perda (KINGMA; BA, 2014; GOODFELLOW; BENGIO; COURVILLE, 2016). Cabe ressaltar que os hiperparâmetros foram definidos empiricamente, enquanto os não especificados foram mantidos nos valores padrão da biblioteca Keras, na versão 2.3.1.A Tabela 6 resume os hiperparâmetros empregados neste trabalho em todos os modelos.

Tabela 6 – Hiperparâmetros usados para treinar os modelos.

| Hiperparâmetro                   | Valor                |
|----------------------------------|----------------------|
| Otimizador                       | Adam                 |
| Função de Perda                  | <i>Cross-Entropy</i> |
| <i>Batch Size</i>                | 1.200                |
| Número de Épocas                 | 200                  |
| <i>Patience (Early Stopping)</i> | 6                    |
| Taxa de Aprendizado              | 0,0001               |

Fonte: Elaborada pela autora.

A seguir, serão apresentados os elementos escolhidos em cada arquitetura de rede após a realização dos testes preliminares.

#### 4.6.1 LSTM

Os modelos LSTM treinados possuem estrutura genérica similar ao ilustrado na Figura 43<sup>6</sup>. O modelo conta com duas camadas LSTM com número de células variando entre 8 e 256. Com base nos resultados dos testes preliminares, o *dropout* é aplicado tanto na segunda camada LSTM como na camada FC, respectivamente, com valores iguais a 0,5 e 0,3. Em seguida, os dados são encaminhados à camada *Softmax*.

#### 4.6.2 CNN-LSTM

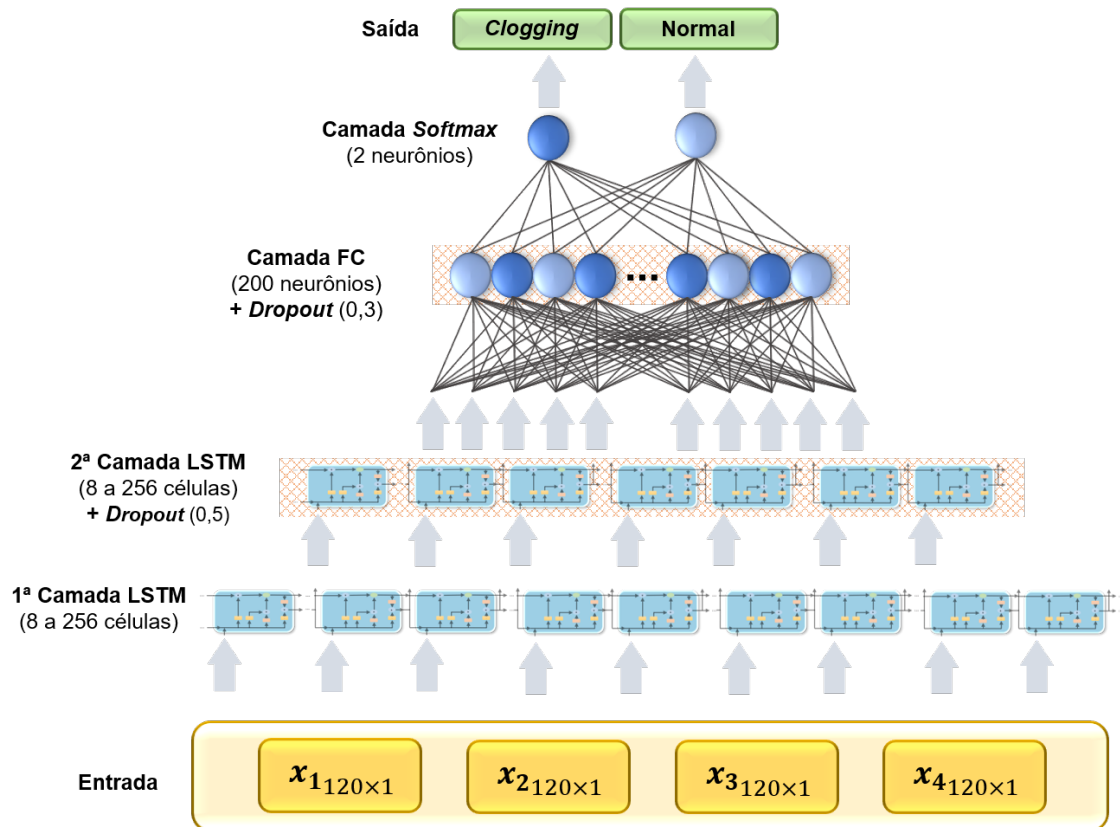
Os modelos CNN são adequados para lidar com a correlação espacial dos dados em análises de séries temporais multivariadas, embora não sejam adaptados para interpretar corretamente dependências temporais longas e complexas. Por sua vez, as redes LSTM são adaptadas para lidar com as correlações temporais de curto e longo prazo (BENGIO; COURVILLE; VINCENT, 2013; LIVIERIS; PINTELAS; PINTELAS, 2020).

Neste sentido, nas redes CNN-LSTM, as camadas convolucionais são aplicadas inicialmente para a extração de características espaciais dos dados e, sequencialmente,

<sup>5</sup> Paciência, do inglês *patience*, é o termo que se refere ao número de épocas após as quais o treinamento será interrompido se não houver melhora de resultado sobre o conjunto de validação (GOODFELLOW; BENGIO; COURVILLE, 2016; CHOLLET, 2021).

<sup>6</sup> Nas Figuras 43, 44 e 45 são representadas alegorias abstratas para representação das estruturas em camadas dos modelos e não retratam as topologias internas específicas de cada rede.

Figura 43 – Estrutura genérica dos classificadores LSTM usados nesta pesquisa.



Fonte: Elaborada pela autora.

duas camadas LSTM são aplicadas para extrair as informações temporais. A Figura 44 exemplifica a estrutura genérica das CNN-LSTM aplicadas neste trabalho<sup>6</sup>.

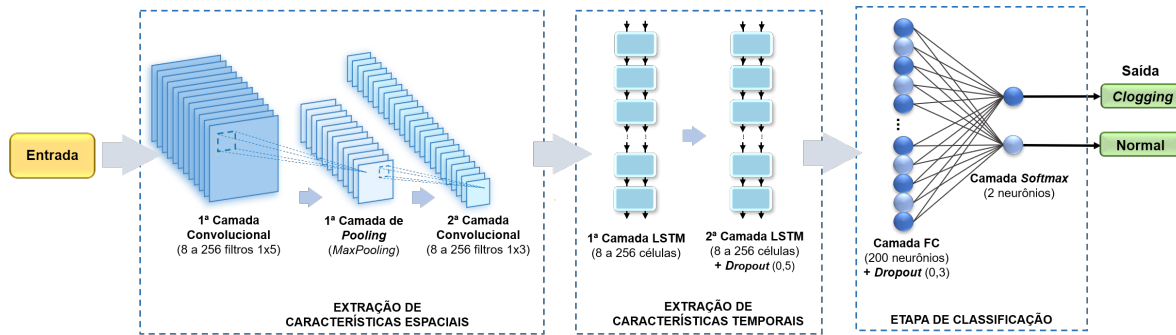
As CNN-LSTM possuem duas camadas convolucionais, contendo cada uma de 8 a 256 *kernels* de tamanhos, respectivamente, iguais a  $1 \times 5$  e  $1 \times 3$ . As camadas convolucionais são intercaladas por uma camada de *MaxPooling* com janela de 2, seguida de duas camadas LSTM, cada uma com número de células variando entre 8 e 256. Por fim, é feita a etapa de classificação através das camadas totalmente conectadas. O *dropout* também é aplicado na segunda camada LSTM e na camada FC, respectivamente, com valores iguais a 0,5 e 0,3.

Conforme observado, as CNN-1D interpretam as subsequências de entrada como blocos e, em seguida, fornece as características extraídas como sequências de entrada para o modelo LSTM. Por sua vez, a LSTM extrai suas próprias características e os encaminha à camada FC para que o resultado do classificador seja apresentado.

### 4.6.3 ConvLSTM

A rede ConvLSTM contém operação de convolução nas transições de entrada para estado e estado para estado. Ao contrário de um LSTM, que lê os dados diretamente para calcular o estado interno e as transições de estado, e ao contrário do CNN-LSTM que

Figura 44 – Estrutura genérica dos classificadores CNN-LSTM.



Fonte: Elaborada pela autora.

interpreta a saída dos modelos CNN, o ConvLSTM aplica as convoluções diretamente como parte do processamento de entrada nas próprias células (BROWNEE, 2018; SUN; ZHAO, 2020; AGGA et al., 2021).

A Figura 45 apresenta a estrutura genérica utilizada pelos modelos ConvLSTM<sup>6</sup>. Este modelo possui duas camadas ConvLSTM com número de células variando entre 8 e 256. De forma similar aos demais classificadores, a segunda camada ConvLSTM conta com um *dropout* de 0,5, ao passo que a camada FC possui um *dropout* de 0,3. O sinal é processado pelas camadas ConvLSTM e, em seguida, são encaminhados às camadas totalmente conectadas.

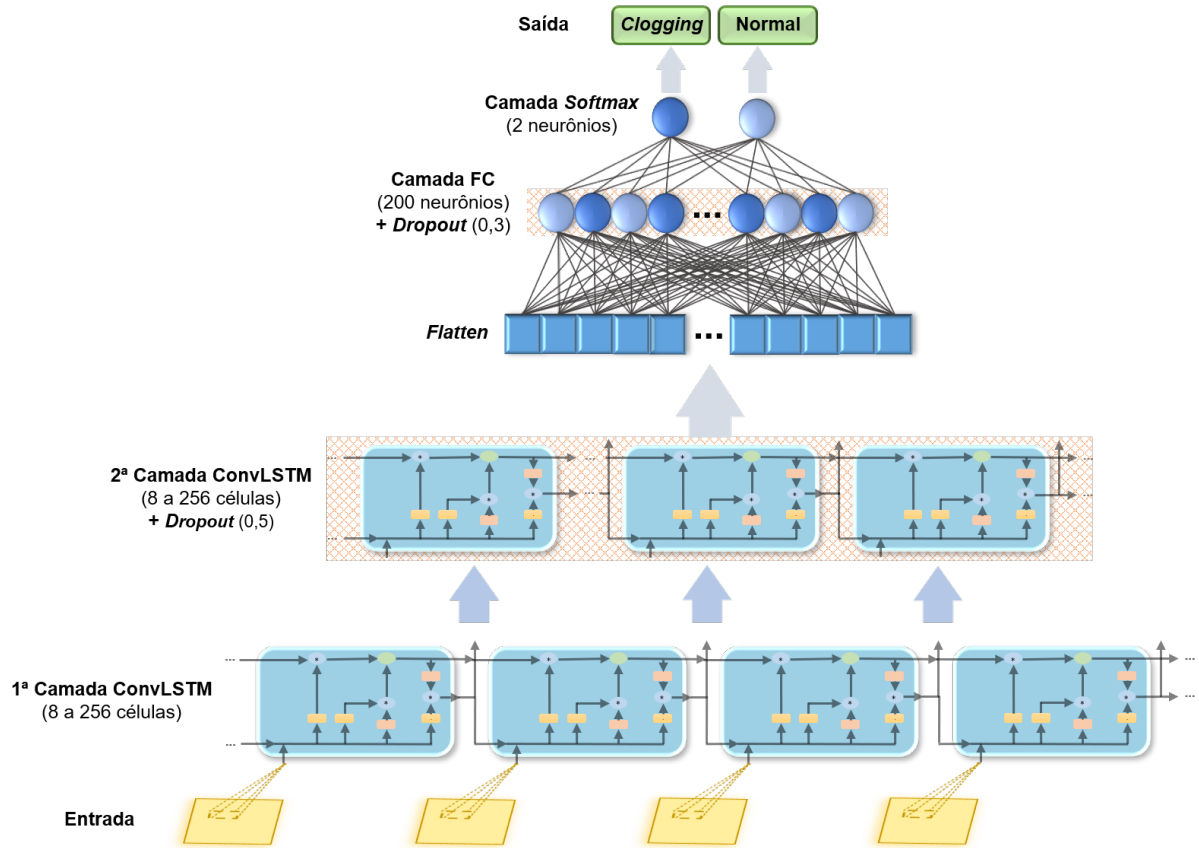
A entrada de uma LSTM é um conjunto de dados ao longo do tempo, isto é, um tensor <sup>7</sup> 3D no formato (amostras, tempo, características). Para o caso em questão, formaria um tensor na forma  $N \times 120 \times 4$ , em que  $N$  representa o número de amostras (sequências) contidas em um *batch*, 120 é o tamanho da janela temporal e 4 é a quantidade de atributos de entrada do modelo. Este também é o formato da matriz de entrada do modelo CNN-LSTM.

Por outro lado, a entrada de uma camada de convolução 2D (CNN-2D) forma um tensor 4D no formato (amostras, canais, linhas, colunas). Neste caso, considerando que o conjunto de entrada é composto por uma série temporal multivariada unidimensional, é necessário reconfigurá-lo. Para tanto, cada janela de 120 segundos de cada variável é dividida em 4 subsequências de 30 segundos, transformando a variável 1D em 2D, conforme ilustra a Figura 46, em que cada cor representa uma das quatro variáveis de entrada e  $N_T$  é o número total de amostras do conjunto de dados.

Já a entrada de uma ConvLSTM, aplicada a partir da biblioteca do Tensorflow, é

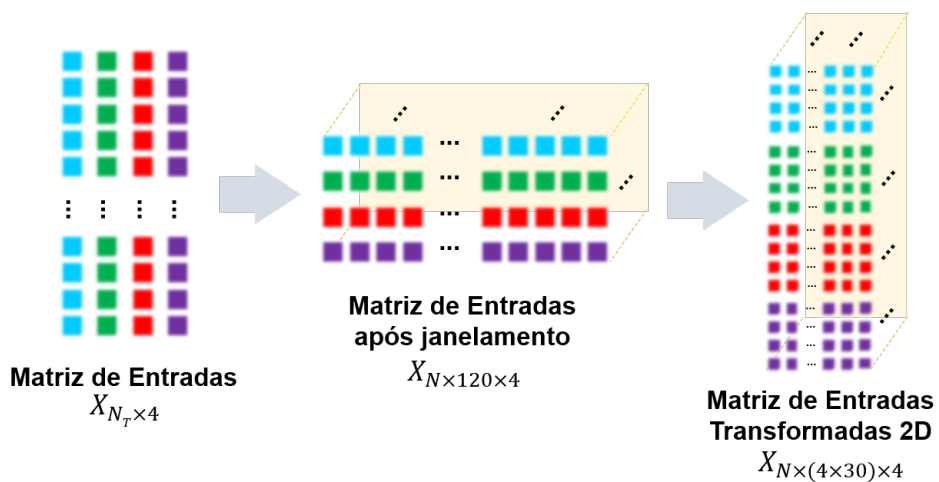
<sup>7</sup> Por definição, os tensores são objetos geométricos que simbolizam as relações lineares entre escalares, vetores e outros tensores. Deste modo, as quantidades físicas que geralmente são denominadas como escalares e vetores são, na realidade, casos particulares de tensores. Casos de estruturas matemáticas mais abstratas e complexas também podem ser tratadas como tensores. A ordem de um tensor é, portanto, representada pela dimensionalidade da matriz necessária para representá-lo (SHARIPOV, 2004).

Figura 45 – Estrutura genérica dos classificadores ConvLSTM.



Fonte: Elaborada pela autora.

Figura 46 – Transformações do conjunto de entrada.



Fonte: Elaborada pela autora.

uma combinação de imagens no tempo na forma de um tensor 5D na forma  $N \times n_{frames} \times n_{row} \times n_{col} \times n_{channels}$ , em que  $n_{frames}$  é o número de subsequências em que cada instância é dividida (isto é, a janela de 120 amostras é dividida em 4 partes de 30 segundos cada),

$n_{row}$  é a dimensão dos dados (neste caso, é 1),  $n_{col}$  é o número de intervalos de tempo em cada subsequência, neste caso 30, e  $n_{channels}$  é o número de atributos de entrada (CHOLLET, 2021). Seguindo o raciocínio, a entrada dos modelos ConvLSTM possuirá a forma  $N \times 4 \times 1 \times 30 \times 4$ .

#### 4.6.4 MLP

Os modelos MLP foram utilizados neste trabalho para a comparação com os demais modelos baseados em *Deep Learning*. Para tanto, foram testadas diferentes configurações de redes MLP contendo uma ou duas camadas ocultas com funções de ativação sigmoide, variando de 16 a 4096 o número de neurônios ocultos das respectivas camadas. Com base nos resultados dos testes preliminares, o *dropout* de 0,5 é aplicado na segunda camada MLP. Outros hiperparâmetros de rede foram mantidos como padrão da biblioteca Keras 2.3.1.

### 4.7 A Seleção dos Classificadores

Em termos estruturais, diferentes configurações de parâmetros, definidas pelo método de tentativa e erro, foram aplicadas aos três modelos de classificadores: LSTM, CNN-LSTM e ConvLSTM. Todas as redes foram testadas com duas camadas de extração de características temporais e/ou espaciais, com número de células ou filtros (dependendo do modelo) variando em incrementos de oito, usando todos os estados da LSTM.

A partir de então, foram selecionadas as oito configurações de classificadores que apresentaram os melhores desempenhos na fase de treinamento sobre o conjunto de validação. As Tabelas 7, 8 e 9 apresentam, respectivamente, as configurações de cada classificador selecionado e também a quantidade de parâmetros treinados. É interessante mencionar que, embora as configurações de rede contendo 8, 16 e 32 células nas primeiras camadas tenham sido testadas, não foram obtidos resultados relevantes. De forma análoga, outros tamanhos de *kernels* foram testados, variando entre  $1 \times 3$  e  $1 \times 11$  em cada camada, mas também não foram gerados resultados satisfatórios.

Devido à natureza estocástica do treinamento dos modelos, a fase de treinamento do classificador foi repetida 5 vezes. Em cada iteração, foi utilizado um conjunto de dados com amostras distintas selecionadas para as etapas de treino, validação e teste, garantindo que as entradas não fossem as mesmas em cada repetição, ao mesmo tempo mantendo a sequência temporal das amostras.

Tabela 7 – Seleção dos oito modelos LSTM que obtiveram os melhores desempenhos na fase de treinamento e validação.

| Modelo | Células da 1ª<br>camada LSTM | Células da 2ª<br>camada LSTM | Número de<br>Parâmetros |
|--------|------------------------------|------------------------------|-------------------------|
| LSTM 1 | 256                          | 128                          | 490.586                 |
| LSTM 2 | 128                          | 64                           | 130.906                 |
| LSTM 3 | 64                           | 32                           | 37.082                  |
| LSTM 4 | 256                          | 64                           | 362.842                 |
| LSTM 5 | 128                          | 32                           | 95.706                  |
| LSTM 6 | 256                          | 32                           | 311.258                 |
| LSTM 7 | 256                          | 16                           | 288.538                 |
| LSTM 8 | 256                          | 8                            | 277.946                 |

Fonte: Elaborada pela autora.

Tabela 8 – Seleção dos oito modelos CNN-LSTM que obtiveram os melhores desempenhos na fase de treinamento e validação.

| Modelo     | Filtros da 1ª<br>Camada CNN1D<br>(tamanho do <i>kernel</i> ) | Filtros da 2ª<br>Camada CNN1D<br>(tamanho do <i>kernel</i> ) | Células da<br>1ª Camada<br>LSTM | Células da<br>2ª Camada<br>LSTM | Número<br>de<br>Parâmetros |
|------------|--------------------------------------------------------------|--------------------------------------------------------------|---------------------------------|---------------------------------|----------------------------|
| CNN-LSTM 1 | 256 (1 × 5)                                                  | 128 (1 × 3)                                                  | 256                             | 128                             | 721.370                    |
| CNN-LSTM 2 | 128 (1 × 5)                                                  | 64 (1 × 3)                                                   | 128                             | 64                              | 188.954                    |
| CNN-LSTM 3 | 64 (1 × 5)                                                   | 32 (1 × 3)                                                   | 64                              | 32                              | 51.770                     |
| CNN-LSTM 4 | 256 (1 × 5)                                                  | 64 (1 × 3)                                                   | 256                             | 64                              | 478.874                    |
| CNN-LSTM 5 | 128 (1 × 5)                                                  | 32 (1 × 3)                                                   | 128                             | 32                              | 125.050                    |
| CNN-LSTM 6 | 256 (1 × 5)                                                  | 32 (1 × 3)                                                   | 256                             | 32                              | 369.914                    |
| CNN-LSTM 7 | 256 (1 × 5)                                                  | 16 (1 × 3)                                                   | 256                             | 16                              | 318.506                    |
| CNN-LSTM 8 | 256 (1 × 5)                                                  | 8 (1 × 3)                                                    | 256                             | 8                               | 293.570                    |

Fonte: Elaborada pela autora.

Tabela 9 – Seleção dos oito modelos ConvLSTM que obtiveram os melhores desempenhos na fase de treinamento e validação.

| Modelo     | Filtros da 1ª<br>Camada ConvLSTM<br>(tamanho do <i>kernel</i> ) | Filtros da 2ª<br>Camada ConvLSTM<br>(tamanho do <i>kernel</i> ) | Número<br>de<br>Parâmetros |
|------------|-----------------------------------------------------------------|-----------------------------------------------------------------|----------------------------|
| ConvLSTM 1 | 256 (1 × 5)                                                     | 128 (1 × 3)                                                     | 2.537.562                  |
| ConvLSTM 2 | 128 (1 × 5)                                                     | 64 (1 × 3)                                                      | 793.946                    |
| ConvLSTM 3 | 64 (1 × 5)                                                      | 32 (1 × 3)                                                      | 278.490                    |
| ConvLSTM 4 | 256 (1 × 5)                                                     | 64 (1 × 3)                                                      | 1.886.042                  |
| ConvLSTM 5 | 128 (1 × 5)                                                     | 32 (1 × 3)                                                      | 554.202                    |
| ConvLSTM 6 | 256 (1 × 5)                                                     | 32 (1 × 3)                                                      | 1.597.146                  |
| ConvLSTM 7 | 256 (1 × 5)                                                     | 16 (1 × 3)                                                      | 1.461.914                  |
| ConvLSTM 8 | 256 (1 × 5)                                                     | 8 (1 × 3)                                                       | 1.396.602                  |

Fonte: Elaborada pela autora.



## 5 Resultados

Os desempenhos dos modelos LSTM, CNN-LSTM e ConvLSTM, respectivamente, são mostrados nas Tabelas 10, 11 e 12 em termos da média e desvio padrão calculados sobre os testes realizados em cada uma das 5 iterações da fase de treinamento dos modelos. Estas tabelas destacam a porcentagem de exemplos classificados corretamente (Acurácia), a porcentagem de amostras classificadas corretamente como *clogging* dentre todas aquelas preditas como *clogging* (Precisão), a porcentagem de eventos de *clogging* detectados corretamente (*Recall*), o *F1-Score* e o MCC. Elas também ilustram o tempo médio para o treinamento completo de cada modelo. Os melhores resultados de cada arquitetura de classificação estão destacados em negrito.

As curvas de treinamento e validação obtidas ao longo dos treinamentos das redes LSTM, CNN-LSTM e ConvLSTM, respectivamente, estão apresentadas no Apêndice A, Apêndice B e Apêndice C.

Por se tratar de um problema envolvendo uma linha de produção, busca-se por metodologias mais assertivas, capazes de detectar corretamente a presença do *clogging* com a menor taxa de erros possível. Como é preferível que os operadores recebam um falso alarme de entupimento (FP) em vez de o entupimento ocorrer sem prévio aviso (FN), com base nas Equações 3.19 e 3.20, a preocupação é obter alta Precisão e, principalmente, valores mais elevados de *Recall*. De forma geral, portanto, os resultados são bastante promissores, principalmente os obtidos pelos modelos ConvLSTM.

Em análise dos classificadores LSTM (ver Tabela 10) e conforme pode ser verificado na Figura 47, a qual ilustra as médias (Figura 47 (a)) e os desvios padrão (Figura 47 (b)) dos critérios de desempenhos, percebe-se que a média do *Recall* só ultrapassou a de

Tabela 10 – Comparação entre os oito classificadores LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos.

| Modelo        | Acurácia [%]      | Precisão [%]      | <i>Recall</i> [%]  | <i>F1-Score</i> [%] | MCC                | Tempo de Treinamento [s] |
|---------------|-------------------|-------------------|--------------------|---------------------|--------------------|--------------------------|
| LSTM 1        | <b>87,56±2,54</b> | <b>86,36±3,12</b> | <b>88,09±5,65</b>  | <b>87,03±4,05</b>   | <b>0,754±0,198</b> | <b>3.224±298</b>         |
| LSTM 2        | 83,94±6,15        | 83,91±2,96        | 83,57±6,26         | 83,51±5,16          | 0,682±0,122        | 1.194±712                |
| LSTM 3        | 75,84±0,35        | 79,61±0,72        | 69,49±0,71         | 74,20±0,36          | 0,521±0,007        | 182±45                   |
| LSTM 4        | 82,40±5,11        | 83,09±5,52        | 79,71±5,15         | 81,36±5,33          | 0,631±0,161        | 3.140±178                |
| LSTM 5        | 81,52±5,61        | 83,14±1,87        | 78,81±8,55         | 80,59±6,78          | 0,636±0,111        | 493±592                  |
| <b>LSTM 6</b> | <b>89,59±6,88</b> | <b>88,98±3,95</b> | <b>89,95±11,35</b> | <b>89,28±7,86</b>   | <b>0,795±0,135</b> | <b>2.465±888</b>         |
| LSTM 7        | 81,87±8,37        | 84,77±5,44        | 76,99±13,06        | 80,47±9,57          | 0,642±0,164        | 2.119±912                |
| <b>LSTM 8</b> | <b>91,12±0,98</b> | <b>89,93±2,01</b> | <b>92,73±3,23</b>  | <b>91,24±1,07</b>   | <b>0,824±0,020</b> | <b>2.020±570</b>         |

Fonte: Elaborada pela autora.

Precisão para os modelos LSTM 1, LSTM 6 e LSTM 8.

O modelo LSTM 8, entretanto, é o que apresenta o resultado mais atrativo pois, além de apresentar um percentual superior da média dos critérios de desempenhos, os valores de desvio padrão menores garantem uma maior estabilidade dos resultados daquele modelo. Em relação ao MCC, que leva em conta todas as informações extraídas da matriz de confusão e mensura quão boa ficou a classificação, observou-se uma média 0,824 com desvio padrão em torno de 0,02 para o modelo LSTM 8.

De acordo com a Tabela 11, de uma maneira geral, os valores de desvio padrão médio dos classificadores CNN-LSTM alcançaram valores menores em comparação com aqueles verificados nos modelos LSTM. Em contrapartida, o tempo de treinamento dos modelos CNN-LSTM alcançou uma média geral de 3.000 segundos, consideravelmente superior aos do modelo LSTM (média de 1.930 segundos).

Apenas o modelo CNN-LSTM 3 não atingiu um *Recall* superior ao valor médio de

Tabela 11 – Comparação entre os oito classificadores CNN-LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos.

| Modelo            | Acurácia [%]      | Precisão [%]      | <i>Recall</i> [%] | F1-Score [%]      | MCC                | Tempo de Treinamento [s] |
|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|--------------------------|
| <b>CNN-LSTM 1</b> | <b>94,39±0,20</b> | <b>91,30±0,46</b> | <b>93,99±0,85</b> | <b>92,62±0,43</b> | <b>0,879±0,004</b> | <b>2.043±213</b>         |
| CNN-LSTM 2        | 91,90±0,49        | 89,47±0,76        | 93,18±0,69        | 91,28±0,72        | 0,840±0,010        | 2.787±421                |
| CNN-LSTM 3        | 77,82±1,67        | 79,97±0,40        | 74,25±4,62        | 77,00±0,733       | 0,559±0,003        | 4.575±1272               |
| <b>CNN-LSTM 4</b> | <b>93,87±0,23</b> | <b>91,10±0,41</b> | <b>94,31±0,58</b> | <b>93,05±0,39</b> | <b>0,879±0,004</b> | <b>3.098±182</b>         |
| CNN-LSTM 5        | 88,10±5,54        | 86,50±2,87        | 89,99±9,70        | 88,21±4,43        | 0,765±0,110        | 3.274±532                |
| <b>CNN-LSTM 6</b> | <b>93,51±0,22</b> | <b>91,82±0,43</b> | <b>94,33±0,44</b> | <b>92,73±0,43</b> | <b>0,872±0,004</b> | <b>3.015±526</b>         |
| CNN-LSTM 7        | 93,19±0,54        | 91,87±1,24        | 93,84±1,99        | 92,84±1,53        | 0,865±0,011        | 3.031±383                |
| CNN-LSTM 8        | 93,68±0,54        | 92,35±1,79        | 93,33±1,50        | 92,83±1,63        | 0,875±0,010        | 3.653±564                |

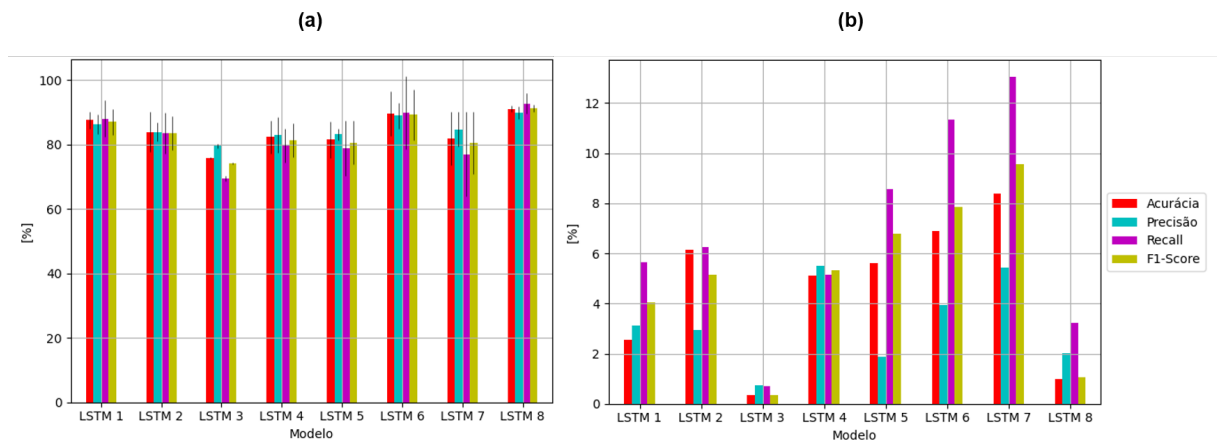
Fonte: Elaborada pela autora.

Tabela 12 – Comparação entre os oito classificadores ConvLSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo dos modelos.

| Modelo            | Acurácia [%]      | Precisão [%]      | <i>Recall</i> [%] | F1-Score [%]      | MCC                | Tempo de Treinamento [s] |
|-------------------|-------------------|-------------------|-------------------|-------------------|--------------------|--------------------------|
| <b>ConvLSTM 1</b> | <b>95,03±0,11</b> | <b>92,98±0,29</b> | <b>97,39±0,23</b> | <b>95,14±0,09</b> | <b>0,902±0,002</b> | <b>4.757±202</b>         |
| ConvLSTM 2        | 94,32±0,21        | 91,62±0,25        | 97,36±0,28        | 94,49±0,20        | 0,888±0,004        | 4.223±256                |
| ConvLSTM 3        | 90,24±0,41        | 85,94±0,50        | 96,21±0,66        | 90,79±0,39        | 0,811±0,008        | 3.708±885                |
| ConvLSTM 4        | 95,14±0,25        | 93,52±0,36        | 96,51±0,28        | 95,23±0,24        | 0,904±0,005        | 4.408±348                |
| ConvLSTM 5        | 94,06±0,29        | 91,73±0,54        | 96,85±0,32        | 94,21±0,26        | 0,883±0,006        | 3.742±278                |
| ConvLSTM 6        | 95,16±0,45        | 93,22±0,43        | 97,41±0,31        | 95,26±0,16        | 0,904±0,003        | 4.202±424                |
| <b>ConvLSTM 7</b> | <b>95,23±0,05</b> | <b>95,53±0,16</b> | <b>97,33±0,15</b> | <b>96,42±0,14</b> | <b>0,907±0,001</b> | <b>3.815±204</b>         |
| <b>ConvLSTM 8</b> | <b>95,11±0,14</b> | <b>93,31±0,40</b> | <b>97,19±0,31</b> | <b>95,21±0,14</b> | <b>0,903±0,003</b> | <b>3.754±298</b>         |

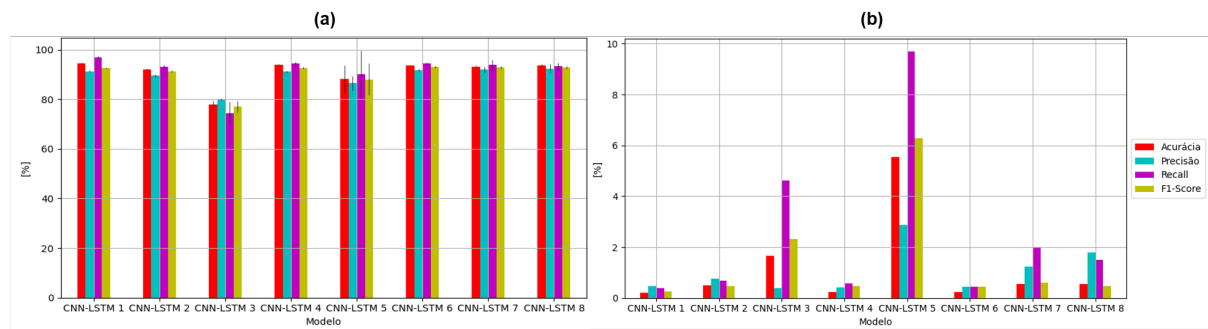
Fonte: Elaborada pela autora.

Figura 47 – (a) Média e (b) Desvio Padrão dos oito classificadores LSTM que obtiveram os melhores desempenhos.



Fonte: Elaborada pela autora.

Figura 48 – (a) Média e (b) Desvio Padrão dos oito classificadores CNN-LSTM que obtiveram os melhores desempenhos.



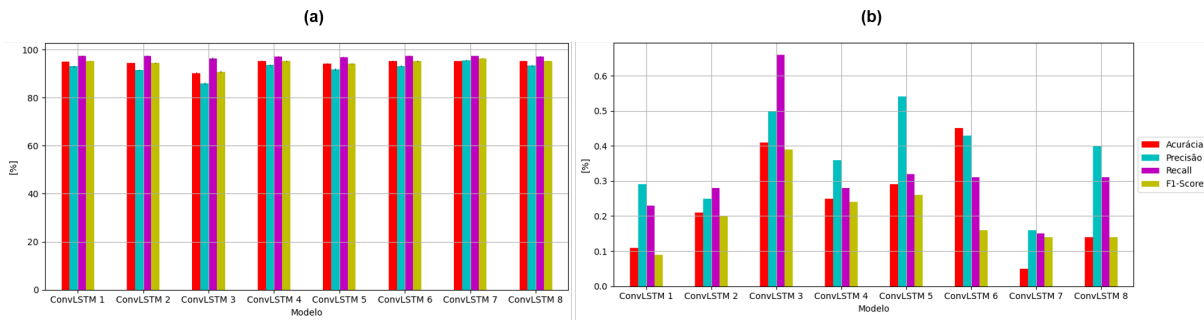
Fonte: Elaborada pela autora.

Precisão. Este também foi o modelo que exigiu um maior tempo médio de treinamento e maior variação em torno desta média, chegando a requerer em torno de 4.575 segundos para o completo treinamento, com um desvio padrão de 1.272 segundos.

Levando-se em conta os menores valores de desvio padrão, dentre os modelos que alcançaram as maiores médias apresentados pela Tabela 11, os modelos CNN-LSTM 1, CNN-LSTM 4 e CNN-LSTM 6 foram os classificadores que apresentaram os resultados mais satisfatórios. Estes resultados também podem ser observados na Figura 48, por meio das médias (Figura 48 (a)) e os desvios padrão (Figura 48 (b)) dos critérios de desempenhos. Uma característica em comum entre os três modelos é a quantidade de unidades ocultas nas primeiras camadas CNN e LSTM (256 neurônios).

Se avaliado, no entanto, o número de parâmetros desses três modelos CNN-LSTM, o modelo CNN-LSTM 6 se torna o mais atrativo, tendo em vista que este possui cerca de 369.914 parâmetros treináveis. Esta quantidade representa aproximadamente a metade dos parâmetros do modelo CNN-LSTM 1 e uma redução de quase  $\frac{1}{4}$  dos parâmetros do

Figura 49 – (a) Média e (b) Desvio Padrão dos oito classificadores ConvLSTM que obtiveram os melhores desempenhos.



Fonte: Elaborada pela autora.

modelo CNN-LSTM 4. Estas reduções, portanto, não influenciaram significativamente na variação em torno da média dos critérios de desempenho do modelo.

Por sua vez, os resultados demonstraram que o método ConvLSTM superou as demais arquiteturas na tarefa de classificação das ocorrências de *clogging*. Ao contrário do CNN-LSTM, que processa os dados em duas fases distintas, o ConvLSTM combina simultaneamente o poder das redes CNN e LSTM, realizando convoluções diretamente como parte do processamento das células LSTM e, portanto, sendo capaz de observar com mais detalhes como os dados se comportam ao longo do tempo. Ao fazer isso, eles são capazes de capturar de maneira mais eficaz as características subjacentes nos dados multidimensionais em concomitância com a identificação das dependências de curto e longo prazo (LIVIERIS; PINTELAS; PINTELAS, 2020; SUN; ZHAO, 2020; AGGA et al., 2021).

Em análise da Tabela 12, são observados valores mais elevados de *Recall* em comparação aos demais modelos. De acordo com Figura 49, de uma maneira geral, destacam-se os modelos ConvLSTM 1, ConvLSTM 7 e ConvLSTM 8, cujos valores médios de critérios de desempenho foram superiores (Ver Figura 49 (a)), seguidos de variações em torno da média menos significativas (Figura 49 (b)).

Em particular, o modelo ConvLSTM 7 obteve o melhor *Recall* médio com 97,33%, contra 94,33% do modelo CNN-LSTM 6 e 92,73% do modelo LSTM 8, por exemplo. O *F1-Score* apresentado pelo ConvLSTM 7 ficou centrado em aproximadamente 96,42% com uma variação quase desprezível em torno da média, o que indica a capacidade do classificador em distinguir as classes mais corretamente, isto é, com baixas taxas de FP e FN.

Os valores de Precisão e MCC obtidos pelo modelo ConvLSTM 7 também foram significativamente melhores, atingindo 95,53% e 0,907, respectivamente. A métrica MCC esperada para um classificador *Dummy*, que se baseia na classe majoritária, é próxima de zero. Os resultados obtidos, portanto, são superiores aos de um classificador *Dummy*.

Devido à natureza do processo, ruídos e, principalmente, ações de controle dos operadores podem estar presentes nas medições, dificultando o uso de regras simples de forma assertiva. Em virtude do desequilíbrio entre as classes, o uso de todas as amostras poderia fazer com que os modelos se comportassem de forma similar a um classificador *Dummy* que, embora pudesse resultar em altos índices de acurácia global, dificultaria a identificação de exemplos pertencentes a classes raras que, no problema em questão, representam a classe de interesse. Além disto, o uso de todos os dados implicaria em um tempo de treinamento dos modelos mais elevado.

Na prática, a análise de modelos deve ser baseada em um *trade-off* entre os resultados das métricas de desempenho, número de parâmetros envolvidos e consumo computacional. Nesse contexto, se comparados os três modelos ConvLSTM que obtiveram os melhores desempenhos (ver [Tabela 12](#)), é possível observar uma diferença de mais de um milhão de parâmetros treináveis entre os modelos ConvLSTM 1 e ConvLSTM 8, desta forma, devido ao complexidade em relação aos demais, o modelo ConvLSTM 1 torna-se menos atrativo. Em relação aos modelos ConvLSTM 7 e ConvLSTM 8, há uma diferença de 65.312 parâmetros em relação ao modelo menos complexo (ConvLSTM 8), o que acarretou em uma diferença gradual mínima entre os tempos de treinamento, no entanto, em termos de tempo de execução dos modelos, não foi registrado variações tão evidentes.

De acordo com a [Tabela 12](#), embora o modelo ConvLSTM 7 tenha apresentado um *Recall* médio inferior aos dos modelos ConvLSTM 1, ConvLSTM 2 e ConvLSTM 6, essa diferença não é significativa. Esta observação também foi constatada através do teste estatístico de Wilcoxon ([WILCOXON, 1945](#)). Para tanto, foram comparados os valores de *Recall* obtidos na fase de teste em cada uma das 5 vezes em que os modelos ConvLSTM 1, ConvLSTM 2, ConvLSTM 6 e ConvLSTM 7 foram treinados.

Considerando uma hipótese nula de que não há diferenças significativas entre os grupos de valores testados para um nível de significância de 5%, a análise dos *p-values* apresentados na [Tabela 13](#) indica que a hipótese nula não pode ser rejeitada. Os testes estatísticos, realizados comparando os valores de *Recall* dos modelos ConvLSTM 1, ConvLSTM 2 e ConvLSTM 6 com o modelo ConvLSTM 7, demonstram consistentemente *p-values* acima do nível de significância estabelecido. O modelo ConvLSTM 7 também se destaca por apresentar os menores desvios padrão em torno da média, bem como o maior valor médio de *F1-Score*. Esses resultados indicam que, estatisticamente, não há diferenças significativas entre os modelos testados com base nos valores de *Recall*, e sugerem que o ConvLSTM 7 pode ser considerado a melhor escolha dentre os modelos analisados.

A [Tabela 14](#) apresenta um comparativo entre os tamanhos dos arquivos de cada um dos modelos aqui desenvolvidos, medidos em kilobytes (kB). A partir desta tabela é possível perceber que o tamanho dos arquivos dos modelos não é tão expressivo, embora

os modelos ConvLSTM sejam os que demandam de um maior espaço para armazenamento frente aos modelos CNN-LSTM e, sobretudo, LSTM. Em especial, se comparados os modelos ConvLSTM 1, ConvLSTM 7 e ConvLSTM 8, o modelo ConvLSTM 7 ainda permanece sendo a melhor escolha, visto que o mesmo apresenta uma redução de mais de 12.000 [kB] quando comparado ao modelo mais complexo, enquanto que não há um aumento significativo de armazenamento comparado ao modelo ConvLSTM 8.

A Tabela 15 mostra o tempo médio de operação para a conclusão das fases de treinamento e teste dos classificadores baseados nos modelos LSTM, CNN-LSTM e ConvLSTM. Cada fase é composta, respectivamente, por um conjunto de 324.634 e 40.579 amostras. A tabela mostra os tempos demandados para que todas as amostras sejam processadas.

O tempo médio de treinamento dos modelos LSTM 8 é o mais curto com 2020 segundos. Já o tempo demandado para o treinamento completo do modelo ConvLSTM 7 é o mais longo dentre os classificadores, chegando a 3815 segundos, seguido da modelagem híbrida que demandou cerca de 3015 segundos para conclusão de seu treinamento. Este resultado é compreensível uma vez que os classificadores CNN-LSTM e ConvLSTM apresentam maior complexidade.

O maior tempo necessário para a conclusão do treinamento do modelo ConvLSTM

Tabela 13 – Resultado do teste de hipóteses para comparação dos valores de *Recall* dos modelos ConvLSTM 1, ConvLSTM 2 e ConvLSTM 6, comparados ao modelo ConvLSTM 7.

| Modelo     | p-value |
|------------|---------|
| ConvLSTM 1 | 0,6250  |
| ConvLSTM 2 | 0,8125  |
| ConvLSTM 6 | 0,8125  |

Fonte: Elaborada pela autora.

Tabela 14 – Tamanho dos arquivos de cada modelo em [kB].

| Modelo | Tamanho do Modelo [kB] |          |          |
|--------|------------------------|----------|----------|
|        | LSTM                   | CNN-LSTM | ConvLSTM |
| 1      | 5.790                  | 7.011    | 29.779   |
| 2      | 1.573                  | 5.347    | 9.344    |
| 3      | 474                    | 4.730    | 3.304    |
| 4      | 4.291                  | 5.667    | 22.142   |
| 5      | 1.162                  | 4.818    | 6.535    |
| 6      | 3.688                  | 4.994    | 18.757   |
| 7      | 3.422                  | 4.658    | 17.173   |
| 8      | 3.296                  | 4.490    | 16.406   |

Fonte: Elaborada pela autora.

pode ser justificado pelas operações de convolução estarem associadas às operações tanto na transição de entrada para estado quanto nas transições de estado para estado, permitindo que a representação interna dos dados multivariados seja trabalhada de forma simultânea às características temporais na classificação. Em consequência disto, este tipo de classificador chega a exigir mais de 1,88 vezes o tempo para treinamento se comparado aos modelos LSTM.

Após a conclusão do treinamento, o tempo de execução dos algoritmos é bastante reduzido. Em destaque, embora a duração média do treinamento dos ConvLSTM seja a maior, toda a fase de teste demandou 10 segundos para sua conclusão, correspondendo a cerca de 0,24 [ms] por amostra. Este resultado corresponde a menos de 1 segundo para que o operador tenha uma resposta em relação à ocorrência inicial do *clogging*, dado o comportamento das variáveis de entrada a partir daquele instante.

É importante salientar que, embora os modelos LSTM e CNN-LSTM tenham apresentado resultados significativos, os classificadores ConvLSTM desempenharam os resultados mais promissores na tarefa de classificação do *clogging*, alcançando patamares de *Recall* médio superiores aos demais e com pouca variação em torno desta média. Esta característica demonstra que são menores as chances dos modelos ConvLSTM classificarem erroneamente uma amostra referente à ocorrência de *clogging* (classe positiva) como operação normal (classe negativa).

Para fins comparativos, os modelos desenvolvidos neste trabalho foram também testados utilizando uma saída com um único neurônio ativado pela função de ativação sigmoide, ao invés de dois neurônios ativados pela função *Softmax*, conforme os testes previamente realizados. Os resultados comparativos podem ser verificados no Apêndice F. Com o intuito de preservar as características e tornar o comparativo mais justo, os hiperparâmetros permaneceram os mesmos daqueles apresentados nas Tabelas 7, 8 e 9.

Como esperado, não foram observadas diferenças significativas entre os valores médios dos critérios de desempenho dos modelos com um ou dois neurônios na camada de saída. Em relação ao desvio padrão, contudo, foi verificada uma menor variação dos modelos com dois neurônios. Os tempos de treinamento de ambas as arquiteturas permaneceram

Tabela 15 – Tempo médio de operação dos classificadores cujos modelos obtiveram os melhores desempenhos.

| Modelo   | Tempo de Treinamento [s] | Tempo de Execução [s] | Tempo de Inferência [ms] |
|----------|--------------------------|-----------------------|--------------------------|
| LSTM     | 2020                     | 13                    | 0,32                     |
| CNN-LSTM | 3015                     | 30                    | 0,73                     |
| ConvLSTM | 3815                     | 10                    | 0,24                     |

Fonte: Elaborada pela autora.

próximos, com um leve aumento após a redução do número de neurônios. É importante notar, no entanto, que os modelos com dois neurônios na camada de saída demandaram mais épocas para atingir a convergência.

Também foi verificado que a diminuição da quantidade de neurônios da camada de saída não acompanhou uma redução significativa do número de parâmetros treinados, correspondendo a uma redução média de 0,15%, 0,11% e 0,02%, respectivamente, comparado aos modelos apresentados pelas Tabelas 7, 8 e 9.

Embora a identificação do *clogging* ao longo do lingotamento contínuo seja um assunto de extrema importância, uma vez que esta reflete em uma considerável redução dos custos de produção e garante a qualidade do produto final, ao longo deste trabalho, foi observada uma certa limitação no que concerne ao fornecimento de informações precisas nos trabalhos desenvolvidos na literatura para a replicabilidade das respectivas metodologias. Outro ponto que deve ser destacado é que as questões comerciais e os protocolos de confidencialidade dos dados por parte da indústria não permitem que as bases de dados utilizadas por eles estivessem disponíveis para consulta pública e, portanto, impossibilitando a sua utilização para fins comparativos entre os modelos.

Em virtude desta limitação, como linha de base para comparação, os resultados obtidos pelos modelos baseados em aprendizagem profunda desenvolvidos neste trabalho foram comparados aos dos modelos MLP. Para tanto, foi variado o número de camadas ocultas, bem como a quantidade de neurônios em cada camada.

A Tabela 16 ilustra um comparativo entre a Precisão e o *Recall* dos modelos MLP com uma única camada oculta, nela também estão incluídas as medidas da Acurácia, *F1-Score* e MCC, além do número de parâmetros e tempo de treinamento. Nesses modelos, não foi observada a ocorrência de *overfitting* na curva de treinamento-validação (ver Apêndice D). Os modelos que obtiveram as melhores médias e, ao mesmo tempo, os menores desvios padrão estão destacados.

A Tabela 16 mostra um aumento gradual nas médias e desvios dos critérios de desempenho vinculados ao aumento do número de neurônios na camada oculta. Visando modelos que não apenas apresentaram as maiores médias dos critérios de desempenho, mas também os menores desvios padrão, os modelos MLP com 1024 e 2048 neurônios na camada oculta apresentaram os melhores desempenhos.

Em particular, o modelo MLP com 1024 neurônios apresentou uma Precisão de 77,93% com desvio padrão de 1,95% e *Recall* de 70,63% com desvio de 1,77%. Embora as redes MLP apresentem tempo de treinamento reduzido comparado aos modelos LSTM, CNN-LSTM e ConvLSTM, em nenhuma configuração o valor de *Recall* ultrapassou a Precisão. Observa-se ainda que os melhores resultados do modelo MLP apresentaram desempenhos inferiores aos dos piores modelos ConvLSTM e CNN-LSTM.

Também foram testadas redes MLP com duas camadas ocultas, variando o número de neurônios de 8 a 256 cada uma. Os critérios de desempenho destes modelos, no entanto, não apresentaram melhorias significativas comparado aos dos modelos da Tabela 16 e, por isso, não foram considerados para efeito de análise deste trabalho. Os respectivos desempenhos podem ser verificados no Apêndice E.

Para demonstrar a capacidade de generalização superior dos modelos ConvLSTM em relação a linha de base, foi realizado um treinamento enviesado nos modelos MLP. Para isso, o conjunto de testes também foi aplicado na fase de treinamento. A Tabela 17 mostra os resultados alcançados pelos modelos.

Pela Tabela 17 é verificado um aumento gradativo no desempenho associado ao aumento do número de parâmetros, bem como dos desvios padrões. Como esperado, os modelos tendenciosos superaram o desempenho dos modelos da Tabela 16, porém, as

Tabela 16 – Comparação entre os desempenhos dos classificadores MLP contendo uma única camada oculta, seguida da quantidade de parâmetros e do tempo médio para o treinamento completo do modelo.

| Núm. de Neurônios | Acurácia [%]      | Precisão [%]      | <i>Recall</i> [%] | <i>F1-Score</i> [%] | MCC                | Núm. de Parâmetros | Tempo de Treinamento [s] |
|-------------------|-------------------|-------------------|-------------------|---------------------|--------------------|--------------------|--------------------------|
| 16                | 72,07±0,38        | 74,78±1,07        | 66,63±1,28        | 70,46±0,51          | 0,444±0,006        | 7.730              | 196±55                   |
| 32                | 73,08±0,46        | 75,66±1,34        | 68,10±2,06        | 71,67±0,76          | 0,493±0,010        | 15.458             | 206±31                   |
| 64                | 73,27±0,31        | 75,64±1,33        | 68,66±1,52        | 71,98±0,33          | 0,468±0,008        | 30.914             | 207±15                   |
| 128               | 73,76±0,27        | 76,32±0,89        | 68,95±0,92        | 72,43±0,20          | 0,478±0,006        | 61.826             | 218±32                   |
| 256               | 74,58±0,46        | 76,76±2,14        | 70,74±2,53        | 73,55±0,51          | 0,494±0,011        | 123.650            | 221±21                   |
| 512               | 75,09±0,57        | 78,83±1,16        | 68,68±2,67        | 73,36±1,15          | 0,507±0,010        | 247.298            | 253±50                   |
| <b>1024</b>       | <b>75,24±0,28</b> | <b>77,93±1,95</b> | <b>70,63±1,77</b> | <b>74,02±0,90</b>   | <b>0,508±0,006</b> | 494.594            | 273±34                   |
| <b>2048</b>       | <b>75,04±0,21</b> | <b>79,03±1,41</b> | <b>68,25±1,83</b> | <b>73,21±0,51</b>   | <b>0,506±0,006</b> | 989.186            | 296±62                   |
| 3072              | 76,31±1,45        | 78,82±3,26        | 72,37±3,48        | 75,33±1,43          | 0,530±0,030        | 1.483.778          | 285±120                  |
| 4096              | 75,94±1,49        | 79,87±1,89        | 69,58±3,92        | 74,25±1,45          | 0,524±0,006        | 1.978.370          | 315±89                   |
| 5120              | 75,76±1,31        | 78,04±3,87        | 72,43±5,78        | 74,83±2,10          | 0,520±0,025        | 2.472.926          | 300±115                  |

Fonte: Elaborada pela autora.

Tabela 17 – Comparação entre os desempenhos dos classificadores MLP enviesados, contendo uma única camada oculta.

| Núm. de Neurônios | Acurácia [%]      | Precisão [%]      | <i>Recall</i> [%] | <i>F1-Score</i> [%] | MCC                |
|-------------------|-------------------|-------------------|-------------------|---------------------|--------------------|
| 16                | 75,05±0,26        | 78,06±0,88        | 69,70±0,75        | 73,63±0,07          | 0,504±0,006        |
| 32                | 75,54±0,12        | 78,49±0,42        | 70,38±0,59        | 74,21±0,20          | 0,514±0,003        |
| 64                | 76,10±0,23        | 78,99±0,88        | 71,13±1,20        | 74,56±0,44          | 0,538±0,003        |
| 128               | 76,76±0,19        | 79,66±0,67        | 71,88±1,30        | 75,56±0,44          | 0,538±0,003        |
| 256               | 78,06±0,70        | 80,59±1,28        | 74,01±3,01        | 77,10±1,23          | 0,564±0,013        |
| 512               | 78,62±1,69        | 81,20±1,59        | 74,51±3,15        | 77,68±2,03          | 0,575±0,030        |
| <b>1024</b>       | <b>79,83±0,75</b> | <b>83,15±1,76</b> | <b>74,91±1,86</b> | <b>78,78±0,81</b>   | <b>0,600±0,016</b> |
| 2048              | 79,83±2,05        | 83,22±3,86        | 75,16±2,81        | 78,89±1,95          | 0,602±0,043        |
| <b>3072</b>       | <b>81,02±2,27</b> | <b>81,88±2,60</b> | <b>79,78±3,91</b> | <b>80,76±2,54</b>   | <b>0,622±0,045</b> |
| 4096              | 78,81±1,91        | 79,89±2,26        | 77,25±5,56        | 78,39±2,66          | 0,579±0,036        |
| 5120              | 78,45±2,05        | 80,84±3,21        | 74,87±3,80        | 77,63±2,19          | 0,572±0,040        |

Fonte: Elaborada pela autora.

melhorias não foram significativas. A Precisão ainda é superior ao *Recall*, o que não é desejado para o problema em questão, além disso, os melhores resultados do modelo MLP com viés apresentam valores piores do que os do pior modelo ConvLSTM. Esses resultados demonstram a maior capacidade de generalização dos modelos baseados em *Deep Learning*, principalmente os modelos ConvLSTM, na tarefa de classificar ocorrências de *clogging*.

Por se tratar de um problema envolvendo dados reais do processo, busca-se por metodologias capazes de detectar corretamente a presença de entupimentos com a menor taxa de erro possível. Nesse contexto, o modelo ConvLSTM 7 demonstra ser a melhor escolha, por apresentar as maiores médias de desempenho e os menores desvios dessas médias, entre os modelos testados. Mesmo levando em consideração as complexidades dos classificadores, os modelos com maiores números de parâmetros ainda parecem ser os mais atrativos, visto que o tempo de processamento desses modelos não apresentou aumento significativo em relação aos demais.

## 5.1 Heurística de Pós-Processamento

Após o treinamento das redes também foi testado um tipo de heurística de pós-processamento desenvolvido a fim de reduzir as taxas de falsos positivos e falsos negativos. A partir dos dados de treinamento, foi verificado que a ocorrência do *clogging* pode variar em duração, abrangendo um período de 45 segundos a até 45 minutos, o que está diretamente relacionado à intervenção do operador, ao tipo de aço em uso e outras condições do processo. A partir dessa informação, as respostas do modelo foram avaliadas em janelas de 45 segundos, cada uma composta por 45 amostras.

Cada amostra no instante  $n$  é avaliada por uma janela deslizante que abrange as amostras de  $n$  e  $n + 44$ , denominada  $J_1$ . Em paralelo, também é analisada uma janela com as últimas 45 amostras antes de  $n$ , denominada  $J_2$ . Na janela  $J_1$ , é calculada a quantidade de amostras classificadas como *clogging* em relação ao tamanho total da janela, chamada de  $ClgMed$ . Além disso, são verificadas quantas amostras consecutivas são rotuladas como *clogging* a partir do início da janela  $J_1$  ( $n_{j1}$ ) e as amostras consecutivas classificadas como *clogging* ao final da janela  $J_2$  ( $n_{j2}$ ).

Se a taxa de ocorrência de *clogging* em  $ClgMed$  ultrapassar o limiar de 80% ( $LimMax$ ), o sistema classificará automaticamente a amostra no instante  $n$  como *clogging*. Por outro lado, se a taxa de ocorrência estiver abaixo do limiar mínimo de 30% ( $LimMin$ ), avalia-se o valor da amostra no instante  $n$  ( $y_{prev}(n)$ ). Caso esta amostra seja rotulada como *clogging* e a soma entre  $n_{j1}$  e  $n_{j2}$  seja superior a 45, a amostra no instante  $n$  é classificada como *clogging*; caso contrário, ela é rotulada como operação normal.

Nos casos em que  $ClgMed$  fica acima de  $LimMin$  e abaixo de  $LimMax$ , o sistema segue um procedimento semelhante. Mais uma vez, avalia-se o valor da amostra no instante

$n$ . Se a amostra for rotulada como *clogging*, verifica-se se a soma de  $n_{j1}$  e  $n_{j2}$  excede 45, classificando-a como *clogging* em caso positivo, e como operação normal em caso negativo. Se a amostra no instante  $n$  for rotulada como operação normal, o sistema procede verificando se todas as amostras na janela  $J_2$  foram classificadas como *clogging*, isto é,  $n_{j2} = 45$ . Se essa condição for atendida, o sistema classifica a amostra como *clogging*. No entanto, caso essa condição não seja satisfeita, a amostra é classificada como operação normal.

A Figura 50 ilustra o procedimento previamente explicado, em que, a partir da análise das janelas  $J_1$  e  $J_2$ , respectivamente, apresentadas em rosa claro e lilás, o sistema ajusta o rótulo da amostra no instante  $n$ . No exemplo, conforme apresentado na Figura 50(a),  $n$  é previamente classificada como operação normal, a janela  $J_2$  não abrange amostras rotuladas como *clogging* ( $n_{j2} = 0$ ), em contrapartida, a janela  $J_1$  conta com cerca de 10 amostras rotuladas como *clogging*, portanto,  $ClgMed \approx 0,22$ . Neste contexto, como  $ClgMed \leq 0,3$  e  $y_{prev}(n) = 0$ , a amostra permanecerá classificada como operação normal. As janelas então deslizam para as próximas amostras, até que toda a série seja analisada. Na Figura 50(b) é apresentado o resultado final após as janelas percorrerem um total de 45 amostras.

Os resultados da classificação puderam ser melhorados a partir da aplicação desta heurística de pós-processamento desenvolvida, de modo que, os níveis de *Recall* e Precisão do modelo ConvLSTM 7 atingiram 98,75% e 97,8%, respectivamente. A Tabela 18 resume um comparativo entre os critérios de desempenho antes e depois do pós-processamento.

Embora o tamanho total da janela seja de 90 segundos, o procedimento em funcionamento terá um atraso de 45 segundos entre a entrada e a resposta de saída devido ao deslocamento da janela. Apesar disso, devido à dinâmica da linha de produção e ao fato deste pós-processamento estar atuando em paralelo com a resposta instantânea, esse atraso não comprometerá a tomada de decisão do operador.

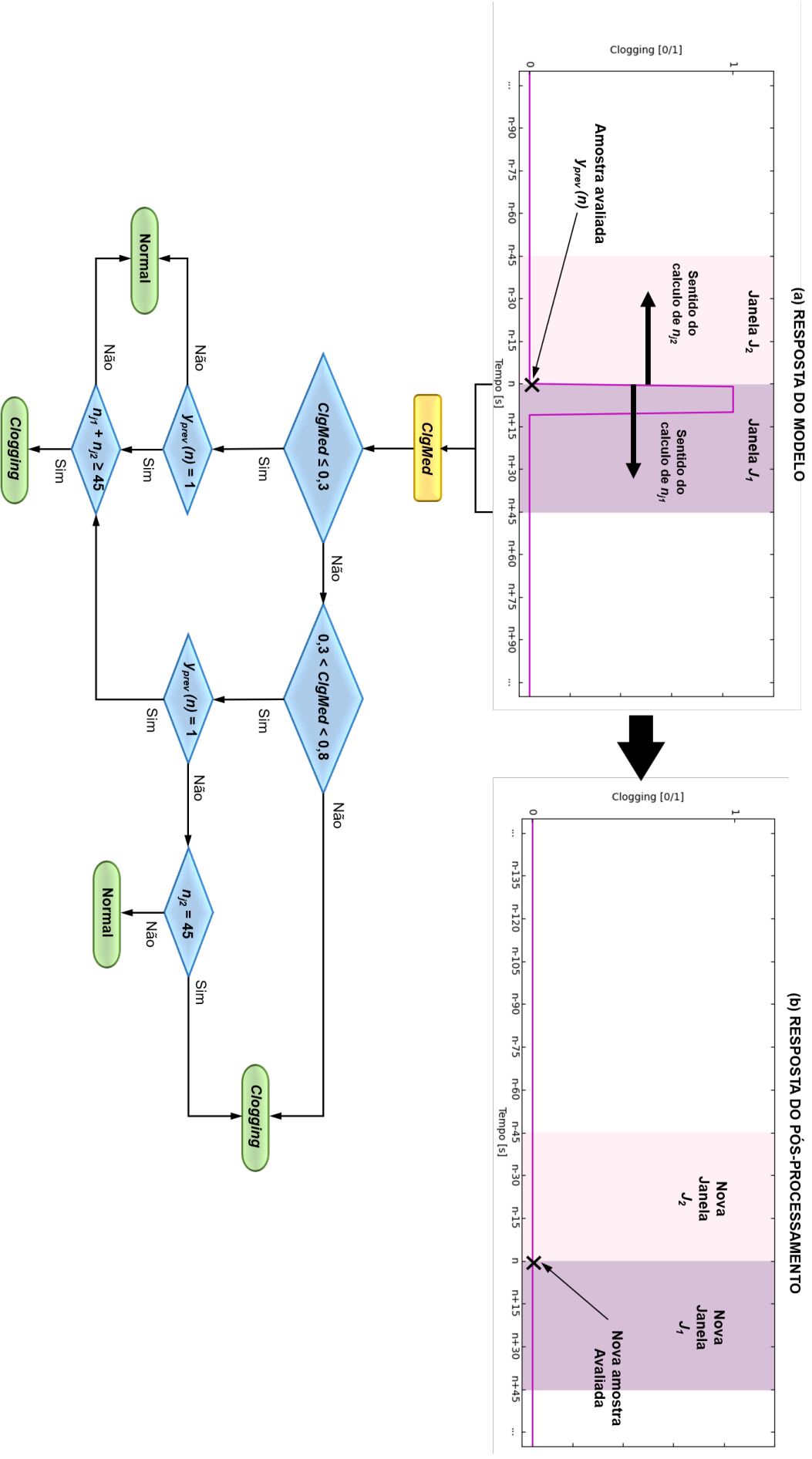
Isso se deve à natureza dinâmica da linha de produção e à operação simultânea do pós-processamento em paralelo com a resposta instantânea. Enquanto a heurística de pós-processamento está em execução em segundo plano, o sistema de classificação da ConvLSTM continua a operar, permitindo que o operador aja imediatamente, tornando

Tabela 18 – Comparação entre os critérios de desempenho antes e depois da aplicação da heurística de pós-processamento aplicada ao modelo ConvLSTM.

| Modelo<br>ConvLSTM 7 | Acurácia<br>[%] | Precisão<br>[%] | Recall<br>[%] | F1-Score<br>[%] | MCC         |
|----------------------|-----------------|-----------------|---------------|-----------------|-------------|
| Antes                | 95,23±0,05      | 95,53±0,16      | 97,33±0,15    | 96,42±0,14      | 0,907±0,001 |
| Depois               | 99,85±0,04      | 97,80±0,12      | 98,75±0,10    | 98,64±0,11      | 0,984±0,001 |

Fonte: Elaborada pela autora.

Figura 50 – Heurística de pós-processamento aplicada ao modelo ConvLSTM 7.



Fonte: Elaborada pela autora.

Tabela 19 – Comparativo entre as métricas de desempenho do modelo ConvLSTM 7 com aquelas obtidas pelos modelos encontrados na literatura.

| Modelo                                   | Acurácia [%] | Precisão [%] | Recall [%]   |
|------------------------------------------|--------------|--------------|--------------|
| ConvLSTM 7                               | <b>99,85</b> | <b>97,80</b> | <b>98,75</b> |
| <a href="#">Pellegrini et al. (2019)</a> | 75,00        | 62,00        | -            |
| <a href="#">Vannucci e Colla (2011)</a>  | 80,20        | -            | 76,90        |

Fonte: Elaborada pela autora.

o atraso praticamente imperceptível na prática. Além disso, a ocorrência de eventos de *clogging* com duração inferior a 60 segundos é relativamente rara em relação ao total de eventos.

Em resumo, embora o procedimento introduza um atraso de 45 segundos, os resultados sem pós-processamento são gerados imediatamente, enquanto os resultados com pós-processamento são gerados com esse pequeno atraso. Esse atraso não compromete a eficiência das operações nem a tomada de decisão do operador, especialmente considerando a infrequência de eventos de curta duração.

Apesar das diferenças significativas nas metodologias e conjuntos de dados, os resultados obtidos alcançaram níveis promissores de critérios de desempenho na tarefa de classificação de entupimentos em relação aos encontrados na literatura. Os resultados alcançados pelo melhor modelo deste trabalho, no caso o modelo ConvLSTM 7, excederam a Acurácia de 75% e a Precisão de 62% de [Pellegrini et al. \(2019\)](#), bem como a Acurácia de 80,2% e o *Recall* de 76,9% de [Vannucci e Colla \(2011\)](#), conforme podem ser observados na [Tabela 19](#).

Além disto, embora nesses trabalhos não sejam informados os respectivos tempos de execução de cada algoritmo para fins de comparação, o tempo de resposta obtido pelo modelo ConvLSTM 7 foi relativamente baixo, representando cerca de 8% do tempo médio estimado por [Barati et al. \(2018\)](#) para a percepção humana do *clogging* na SEN, isto é, 120 segundos. A superioridade dos classificadores ConvLSTM pode ser atribuída ao uso de um modelo especial de aprendizagem profunda que combina o poder das redes CNN e LSTM, sendo capaz de explorar padrões transversais<sup>1</sup> a partir da representação dos dados.

Com base nos dados levantados, uma corrida<sup>2</sup> leva, em média, 56,25 minutos para ser concluída. De acordo com [Rackers \(1995\)](#) e [Thomas e Bai \(2001\)](#), em geral, o SEN é trocado a cada 4 corridas, apesar de possuir vida útil de até 600 minutos (isto é, em torno de 10 corridas) e interrompendo o processo por cerca de 1 minuto. Neste contexto, a cada 4 corridas, perde-se 1 minuto de processo.

<sup>1</sup> Padrões presentes nos dados que se estendem ao longo de diferentes dimensões ou eixos.

<sup>2</sup> A corrida no contexto do processo siderúrgico se refere à produção contínua e ininterrupta do aço em lotes consecutivos.

Considerando que em uma usina siderúrgica com capacidade de produção anual de cerca de 7,5 milhões de toneladas de aço, composta por 3 máquinas de lingotamento contínuo, são realizadas 9.344 corridas/ano em cada máquina, serão perdidos anualmente mais de 7.000 minutos de processo. Isto corresponde a 33.334,72 toneladas de aço que deixaram de ser produzidas, significando uma perda de mais de US\$25 milhões/ano.

Estima-se que a aplicação de um sistema que permita identificar a ocorrência de *clogging*, conseqüentemente, contribua para o aumento da vida útil do SEN para 10 corridas, por exemplo, correspondendo a uma perda anual aproximada de 934,4 minutos por máquina, equivalendo a 13.333,88 toneladas de aço não produzidas (cerca de US\$ 10 milhões/ano). Os frutos deste trabalho poderão, portanto, favorecer ganhos de produtividade, além de uma significativa redução do custo produtivo, devido ao aumento da vida útil do SEN, podendo resultar uma economia de mais de US\$15 milhões por ano.

## 6 Conclusão

A obstrução no SEN das máquinas de lingotamento contínuo é um fenômeno de difícil identificação durante o processo e, por sua vez, afeta adversamente a qualidade do produto, alterando o padrão de fluxo do aço e degradando a qualidade interna do produto final. Além disto, o *clogging* pode acarretar em um aumento da frequência de interrupções na operação, elevando os custos operacionais devido à redução da capacidade produtiva do processo.

Na prática, as siderúrgicas não contam com ferramentas eficazes disponíveis que possam quantificar corretamente essa obstrução durante a operação de lingotamento. A minimização e o controle das perdas associadas ao fenômeno são complexos porque elas podem ocorrer por diversos mecanismos e, por vezes, de maneira muito rápida.

Visando contornar este problema, neste trabalho foram empregadas três metodologias baseadas em aprendizagem profunda com o objetivo de detectar a ocorrência de *clogging* no SEN de um processo de produção siderúrgico real. Neste contexto, o desempenho dos modelos de classificação utilizando LSTM, CNN-LSTM e ConvLSTM foram comparados na tarefa de classificação de ocorrência de entupimento, levando em consideração a Precisão, o *Recall*, F1-Score e o MCC.

Devido à complexidade do processo e o número de eventos, foi necessária uma seleção baseada em dados e na qualidade dos dados fornecidos com etapas de pré-processamento. Com base nas evidências observadas na literatura e nas análises realizadas neste trabalho, quatro variáveis foram selecionadas para compor o conjunto de dados de entrada: abertura da válvula, nível do molde, velocidade do veio e distribuidor em operação. O conjunto de dados foi pré-processado a fim de lidar com a presença de *outliers* e valores ausentes, garantindo as relações entre os atributos. Além disso, devido ao desequilíbrio inerente às classes, uma etapa cuidadosa de balanceamento foi necessária para o treinamento, para que as informações de grande importância para a modelagem não fossem desconsideradas.

Os problemas industriais requerem taxas de acerto maiores e, dentro da tarefa de classificação das ocorrências de *clogging*, os modelos ConvLSTM obtiveram desempenhos notavelmente superiores em relação às demais metodologias testadas neste trabalho. Estes modelos foram capazes de detectar uma alta porcentagem de eventos de *clogging*, com taxas de falsos negativos relativamente baixas.

O modelo ConvLSTM 7 foi escolhido levando-se em consideração não só os valores médios de *Recall* e Precisão, mas também por apresentar em geral as menores variações em torno da média dos critérios de desempenho dos demais modelos testados. Em particular, o ConvLSTM 7 obteve um valor de *Recall* médio de 97,33%, contra 94,33% do modelo

CNN-LSTM 6 e 92,73% do LSTM 8.

Visando reduzir ainda mais as taxas de falsos positivos e negativos, foi implementada uma heurística de pós-processamento aplicada em paralelo à saída do modelo ConvLSTM 7. Após a implementação, o *Recall* alcançou um nível de quase 99% enquanto a Precisão atingiu um patamar de cerca de 98% e, até o conhecimento da autora, esses resultados nunca foram encontrados na literatura para este tipo de tarefa. Devido ao tamanho da janela utilizada para o cálculo da média, a utilização desse procedimento pode resultar em um atraso de resposta de 45 segundos, porém, dada a dinâmica do processo e a utilização em paralelo com a resposta instantânea do classificador, esse atraso não afetaria a tomada de decisão do operador.

Os objetivos propostos neste trabalho foram então atingidos, tendo em vista que a partir da análise das variáveis que exercem maior influência sobre as ocorrências de *clogging* e do pré-processamento aplicado, foi possível elaborar um modelo intrinsecamente orientado aos dados, com resultados mais assertivos no que diz respeito à tarefa de identificação do *clogging* em relação aos demais classificadores testados.

Os resultados também apontam que a hipótese inicialmente declarada possa estar correta, uma vez que, a partir de uma modelagem baseada em *Deep Learning*, foi possível desenvolver um classificador para as ocorrências de *clogging* com alto grau de assertividade e baixas taxas de falsos negativos, fazendo uso de apenas quatro variáveis do processo.

Mesmo havendo diferenças significativas entre as metodologias e conjuntos de dados, os resultados alcançados neste trabalho são superiores aos encontrados na literatura para a tarefa de classificação de ocorrências de *clogging*. Em especial, o modelo ConvLSTM 7 ultrapassou a Acurácia de 75% e a Precisão de 62% de Pellegrini et al. (2019), bem como a Acurácia de 80,2% e o *Recall* de 76,9% de Vannucci e Colla (2011).

Cabe ressaltar, no entanto, que devido às questões de confidencialidade dos dados industriais, as bases de dados não costumam estar disponibilizadas. Além disto, a falta de informações mais consistentes sobre as demais metodologias empregadas na literatura tornou inviável a possibilidade de comparação direta com outras abordagens. Partindo desta premissa, os resultados da classificação dos modelos ConvLSTM foram comparados aos classificadores MLP.

Embora dois modelos de classificadores MLP tenham atingido uma média de Precisão em torno de 78% e um *Recall* próximo a 70%, é importante ressaltar que o *Recall* não ultrapassou a Precisão. Este fato pode tornar os modelos MLP menos interessantes do ponto de vista da indústria, que geralmente busca valores de *Recall* superiores aos de Precisão. Além disso, a análise revelou a predominância da superioridade dos modelos baseados em *Deep Learning* mesmo quando comparados aos classificadores MLP tendenciosos.

O fato dos valores médios de desvio padrão dos modelos ConvLSTM terem sido os menores, associado ao elevado valor médio de *Recall*, demonstra que as chances destes modelos classificarem erroneamente uma amostra referente à ocorrência de entupimento (positiva) como operação normal (negativa) são menores. Este é o resultado desejado no diagnóstico de falhas, pois é preferível que os operadores recebam um falso alarme de entupimento em vez de o evento ocorrer sem qualquer aviso prévio.

A superioridade dos classificadores ConvLSTM pode ser atribuída ao uso de um modelo especial de *Deep Learning* que combina o poder das redes CNN e LSTM, sendo capaz de explorar padrões transversais das subsequências formadas pelo tensor a partir da representação dos dados.

Embora a maior complexidade do modelo escolhido frente às demais arquiteturas testadas implique em um tempo de treinamento relativamente mais longo, o tempo médio para completar a execução dos modelos ConvLSTM na fase de teste foi drasticamente reduzido, exigindo cerca de 10 segundos para o processamento de mais de 700 minutos de processo. Levando-se em consideração todos esses fatores, é possível concluir que a modelagem ConvLSTM pode de fato atender ao interesse da indústria na tarefa de detecção do *clogging*.

Como continuação deste trabalho, pretende-se aprimorar a metodologia aqui apresentada para detectar com maior antecedência a ocorrência inicial do *clogging* no conjunto de dados avaliado e implementar o sistema na planta real. Além disso, planeja-se testar o modelo em outras plantas para avaliar sua robustez e generalização, garantindo que a metodologia seja eficaz em diferentes ambientes operacionais e com diferentes conjuntos de dados. Espera-se também que a predição de entupimento com horizontes maiores possa oferecer uma contribuição ainda mais significativa, permitindo que os operadores tenham um tempo adicional para planejar e adotar medidas preventivas de forma eficaz. É essencial considerar, no entanto, a viabilidade e os desafios dessa implementação prática.

Inicialmente, o treinamento *offline* da rede neural e sua transferência para o ambiente industrial permitiriam a conexão direta com os sensores em tempo real, facilitando a coleta e o processamento contínuo dos dados pelo modelo ConvLSTM. Durante a operação, o sistema poderia ser atualizado continuamente com base nos dados disponíveis periodicamente, o que contribuiria para aumentar sua eficácia diante das condições específicas do processo. A implementação dessa estratégia em uma planta real, entretanto, enfrentaria desafios significativos, como a integração de sistemas de coleta de dados, garantia de qualidade dos dados, adaptação da rede neural às condições da planta e segurança da informação. Seria necessário um esforço multidisciplinar para garantir o sucesso da implementação e aproveitar todo o potencial da estratégia no ambiente industrial.

Outra abordagem a ser explorada em trabalhos futuros seria o uso de outras técnicas, como a aprendizagem multimodal, *ensemble* de classificadores e *MultiScale* CNN, a fim de

compará-las com os resultados desta tese. Uma das grandes vantagens da aprendizagem multimodal, por exemplo, é a capacidade do modelo de aproveitar informações de diversas fontes (sinais, imagens, sons, etc.) disponíveis no processo ([MODI et al., 2011](#); [SONG et al., 2019](#); [XIAN; LIANG, 2022](#)). O *MultiScale* CNN, por sua vez, pode capturar diferentes níveis de detalhes nos dados ao aplicar convoluções em múltiplas escalas, proporcionando uma representação mais rica e robusta ([LIU; HSAIO; TU, 2018](#); [QIAN et al., 2020](#); [SHAO; KIM; KIM, 2020](#)).

## Referências

- AGGA, A.; ABBOU, A.; LABBADI, M.; HOUM, Y. E. Short-term self consumption pv plant power production forecasts based on hybrid cnn-lstm, convlstm models. Renewable Energy, Elsevier, v. 177, p. 101–112, 2021. Citado 5 vezes nas páginas [34](#), [35](#), [80](#), [108](#) e [116](#).
- AMARAL, H. N.; GASPAROTTO, A. M. S. Inteligência artificial: o uso da robótica indústria 4.0. Revista Interface Tecnológica, v. 18, n. 1, p. 474–486, 2021. Citado na página [29](#).
- ARRIVABENE, L. F.; MACHADO, M. L. P.; SOBRINHO, V. Siderurgia para não siderurgistas. CEFET-ES, Vitória-ES, Brasil, 2003. Citado na página [40](#).
- ASSAR, M.; DAUBY, P.; LAWSON, G. Opening the black box: Piv and mfc measurements in a continuous caster mold. In: 83 rd Steelmaking Conference. [S.l.: s.n.], 2000. p. 397–411. Citado na página [46](#).
- BAGUI, S.; LI, K. Resampling imbalanced data for network intrusion detection datasets. Journal of Big Data, SpringerOpen, v. 8, n. 1, p. 1–41, 2021. Citado na página [100](#).
- BARATI, H.; WU, M.; KHARICHA, A.; LUDWIG, A. A transient model for nozzle clogging. Powder Technology, Elsevier, v. 329, p. 181–198, 2018. Citado 8 vezes nas páginas [30](#), [44](#), [54](#), [61](#), [62](#), [63](#), [97](#) e [125](#).
- BARATI, H.; WU, M.; KHARICHA, A.; LUDWIG, A. Calculation accuracy and efficiency of a transient model for submerged entry nozzle clogging. Metallurgical and Materials Transactions B, Springer, v. 50, n. 3, p. 1428–1443, 2019. Citado 2 vezes nas páginas [63](#) e [64](#).
- BATCHELOR, G. An Introduction to Fluid Dynamics. [S.l.]: Cambridge Univ. Press, Cambridge, UK, 1973. Citado na página [61](#).
- BENGIO, Y.; COURVILLE, A.; VINCENT, P. Representation learning: A review and new perspectives. IEEE transactions on pattern analysis and machine intelligence, IEEE, v. 35, n. 8, p. 1798–1828, 2013. Citado na página [106](#).
- BENGIO, Y.; GRANDVALET, Y. No unbiased estimator of the variance of k-fold cross-validation. Journal of machine learning research, v. 5, n. Sep, p. 1089–1105, 2004. Citado na página [83](#).
- BESSHO, N.; YODA, R.; YAMASAKI, H.; FUJII, T.; NOZAKI, T.; TAKATORI, S. Numerical analysis of fluid flow in continuous casting mold by bubble dispersion model. ISIJ international, The Iron and Steel Institute of Japan, v. 31, n. 1, p. 40–45, 1991. Citado na página [46](#).
- BOLGER, D. Stopper rod and submerged nozzle design and operation in continuous casting. In: Steelmaking Conference. [S.l.: s.n.], 1994. v. 77, p. 531–537. Citado 2 vezes nas páginas [47](#) e [95](#).

- BOUGHORBEL, S.; JARRAY, F.; EL-ANBARI, M. Optimal classifier for imbalanced data using matthews correlation coefficient metric. PloS one, Public Library of Science San Francisco, CA USA, v. 12, n. 6, p. e0177678, 2017. Citado na página 85.
- BROWNLEE, J. Deep learning for time series forecasting: predict the future with MLPs, CNNs and LSTMs in Python. [S.l.]: Machine Learning Mastery, 2018. Citado 6 vezes nas páginas 35, 69, 72, 73, 76 e 108.
- BUDUMA, N.; LOCASCIO, N. Fundamentals of deep learning: Designing next-generation machine intelligence algorithms. [S.l.]: "O'Reilly Media, Inc.", 2017. Citado 4 vezes nas páginas 74, 75, 78 e 80.
- BYRNE, M.; CRAMB, A.; FENICLE, T. The sources of exogenous inclusions in continuous cast, aluminum-killed steels. In: Steelmaking Proceedings, [S.l.: s.n.], 1985. v. 68, p. 451–461. Citado 2 vezes nas páginas 44 e 45.
- CARVALHO, A.; FACELI, K.; LORENA, A.; GAMA, J. Inteligência artificial—uma abordagem de aprendizado de máquina. Rio de Janeiro: LTC, p. 45, 2011. Citado 4 vezes nas páginas 69, 71, 72 e 101.
- CEMERNEK, D.; CEMERNEK, S.; GURSCH, H.; PANDESHWAR, A.; LEITNER, T.; BERGER, M.; KLÖSCH, G.; KERN, R. Machine learning in continuous casting of steel: a state-of-the-art survey. Journal of Intelligent Manufacturing, Springer, p. 1–19, 2021. Citado 5 vezes nas páginas 32, 34, 64, 67 e 69.
- CHANG, H.-J.; CHEN, M.-H.; REN, Y.-F. Detection of mold level fluctuation in the continuous casting process with thermal image and surface disturbing evaluation. China Steel Technical Report, n. 34, p. 7–11, 2021. Citado 2 vezes nas páginas 52 e 53.
- CHICCO, D.; JURMAN, G. The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. BMC genomics, Springer, v. 21, n. 1, p. 1–13, 2020. Citado na página 84.
- CHICCO, D.; TÖTSCH, N.; JURMAN, G. The matthews correlation coefficient (mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. BioData mining, BioMed Central, v. 14, n. 1, p. 1–22, 2021. Citado na página 85.
- CHOLLET, F. Deep learning with Python. [S.l.]: Simon and Schuster, 2021. Citado 3 vezes nas páginas 76, 106 e 110.
- CHOUDHARY, S.; GHOSH, A. Thermodynamic evaluation of formation of oxide-sulfide duplex inclusions in steel. ISIJ international, The Iron and Steel Institute of Japan, v. 48, n. 11, p. 1552–1559, 2008. Citado na página 47.
- CRAMB, A.; JIMBO, I. Interfacial considerations in continuous casting. Iron Steelmaker, v. 16, n. 6, p. 43–55, 1989. Citado na página 46.
- DINIZ, A. P. M.; CIARELLI, P. M.; SALLES, E. O. T.; COCO, K. F. Use of deep neural networks for clogging detection in the submerged entry nozzle of the continuous casting. Expert Systems with Applications, Elsevier, v. 238, p. 121963, 2024. Citado na página 100.

DINIZ, A. P. M.; CÔCO, K. F.; GOMES, F. S. V.; SALLES, J. L. F. Forecasting model of silicon content in molten iron using wavelet decomposition and artificial neural networks. Metals, Multidisciplinary Digital Publishing Institute, v. 11, n. 7, p. 1001, 2021. Citado na página 31.

DZIERŻAK, R. Comparison of the influence of standardization and normalization of data on the effectiveness of spongy tissue texture classification. Informatyka, Automatyka, Pomiary w Gospodarce i Ochronie Środowiska, v. 9, 2019. Citado na página 101.

FAWCETT, T. An introduction to roc analysis. Pattern recognition letters, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado 2 vezes nas páginas 32 e 83.

FUKUDA, Y.; UESHIMA, Y.; MIZOGUCHI, S. Mechanism of alumina deposition on alumina graphite immersion nozzle in continuous caster. ISIJ international, The Iron and Steel Institute of Japan, v. 32, n. 1, p. 164–168, 1992. Citado 2 vezes nas páginas 44 e 47.

GADEKALLU, T. R.; KHARE, N.; BHATTACHARYA, S.; SINGH, S.; MADDIKUNTA, P. K. R.; RA, I.-H.; ALAZAB, M. Early detection of diabetic retinopathy using pca-firefly based deep learning model. Electronics, MDPI, v. 9, n. 2, p. 274, 2020. Citado na página 101.

GHASEMI, A.; ZAHEDIASL, S. Normality tests for statistical analysis: a guide for non-statisticians. International journal of endocrinology and metabolism, Brieflands, v. 10, n. 2, p. 486, 2012. Citado na página 101.

GIRASE, N.; BASU, S.; CHOUDHARY, S. Development of indices for quantification of nozzle clogging during continuous slab casting. Ironmaking & Steelmaking, Taylor & Francis, v. 34, n. 6, p. 506–512, 2007. Citado na página 66.

GONZÁLEZ-YERO, G.; LEYVA, R. R.; MENDOZA, M. R.; ALBERTOS, P.; CRESPO-LORENTE, A.; ALONSO, J. M. R. Neuro-fuzzy system for compensating slow disturbances in adaptive mold level control. Metals, MDPI, v. 11, n. 1, p. 56, 2020. Citado 2 vezes nas páginas 95 e 96.

GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. Deep Learning. MIT Press, 2016. (Adaptive Computation and Machine Learning series). ISBN 9780262035613. Disponível em: <<https://books.google.com.br/books?id=Np9SDQAAQBAJ>>. Citado 11 vezes nas páginas 70, 72, 73, 74, 75, 76, 78, 79, 80, 83 e 106.

GORDON-RODRIGUEZ, E.; LOAIZA-GANEM, G.; PLEISS, G.; CUNNINGHAM, J. P. Uses and abuses of the cross-entropy loss: Case studies in modern deep learning. PMLR, 2020. Citado na página 76.

HAYKIN, S. Neural Networks and Learning Machines. Pearson Education, 2011. ISBN 9780133002553. Disponível em: <<https://books.google.com.br/books?id=faouAAAAQBAJ>>. Citado 4 vezes nas páginas 70, 71, 82 e 83.

HEASLIP, L.; SOMMERVILLE, I.; MCLEAN, A.; SWARTZ, L.; WILSON, W. Model study of fluid flow and pressure distribution during sen injection–potential for reactive metal additions during continuous casting. Iron Steelmaker, v. 14, n. 8, p. 49–64, 1987. Citado na página 46.

- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. Neural computation, MIT Press, v. 9, n. 8, p. 1735–1780, 1997. Citado na página 78.
- IKÄHEIMONEN, J.; LEIVISKÄ, K.; RUUSKA, J.; MATKALA, J. Nozzle clogging prediction in continuous casting of steel. IFAC Proceedings Volumes, Elsevier, v. 35, n. 1, p. 143–147, 2002. Citado 15 vezes nas páginas 15, 30, 31, 32, 35, 40, 42, 43, 50, 54, 55, 56, 57, 67 e 87.
- ISHII, K.; ASAI, S.; MAKI, M.; FUKUSHIMA, H. An introduction to iron and steel processing. JFE 21st Century Foundation, Tokyo, p. 190–209, 2006. Citado na página 40.
- JABRI, K.; BELE, B.; MOUCHETTE, A.; DUMUR, D.; GODOY, E. Suppression of periodic disturbances in the continuous casting process. In: IEEE. 2008 IEEE International Conference on Control Applications. [S.l.], 2008. p. 91–96. Citado 2 vezes nas páginas 95 e 96.
- JAPKOWICZ, N. The class imbalance problem: Significance and strategies. In: CITESEER. Proc. of the Int'l Conf. on Artificial Intelligence. [S.l.], 2000. v. 56, p. 111–117. Citado na página 100.
- JÚNIOR, M. A. V.; SILVA, C. A.; SILVA, I. A. Hybrid model associating thermodynamic calculations and artificial neural network in order to predict molten steel temperature evolution from blowing end of a bof for secondary metallurgy. REM-International Engineering Journal, SciELO Brasil, v. 71, p. 587–592, 2018. Citado na página 32.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014. Citado na página 106.
- KLANKKE, S.; LÖPKE, M.; UEBBER, N.; ODENTHAL, H.-J.; POUCKE, J. V.; DEYNE, A. V. Y.-D. Advanced data-driven prediction models for bof endpoint detection. Association for Iron & Steel Technology 2017 Proceedings, p. 1307–1313, 2017. Citado na página 67.
- KOCH, V.; KUGE, S.; GEISSBAUER, R.; SCHRAUF, S. Industry 4.0-Opportunities and challenges of the industrial internet. Strategy& Formerly Booz & Company, PwC, 13, 1–51. 2014. Citado na página 29.
- KONG, W.; CANG, D.-Q. Research on ladle nozzle clogging during continuous casting of non-oriented electrical steel with high silicon. Archives of Metallurgy and Materials, Institute of Metallurgy and Materials Science of Polish Academy of Sciences, p. 257–264, 2019. Citado 6 vezes nas páginas 30, 31, 35, 44, 54 e 87.
- KONG, W.; CHEN, Y.-F.; CANG, D.-G. Ladle nozzle clogging during casting of silicon-steel. High Temperature Materials and Processes, De Gruyter, v. 38, n. 2019, p. 813–821, 2019. Citado 9 vezes nas páginas 30, 31, 32, 35, 54, 59, 60, 61 e 87.
- KRIZHEVSKY, A.; SUTSKEVER, I.; HINTON, G. E. Imagenet classification with deep convolutional neural networks. Advances in neural information processing systems, v. 25, p. 1097–1105, 2012. Citado 2 vezes nas páginas 73 e 74.
- LABACH, A.; SALEHINEJAD, H.; VALAEE, S. Survey of dropout methods for deep neural networks. arXiv preprint arXiv:1904.13310, 2019. Citado 2 vezes nas páginas 75 e 76.

- LARSEN, K.; FRUEHAN, R. Calcium modification of oxide inclusions. In: Steelmaking Conference Proceedings. [S.l.: s.n.], 1990. v. 73, p. 497–506. Citado na página 47.
- LECUN, Y.; BOTTOU, L.; ORR, G. B.; MÜLLER, K.-R. et al. Neural networks: Tricks of the trade. Springer Lecture Notes in Computer Sciences, v. 1524, n. 5-50, p. 6, 1998. Citado na página 73.
- LEE, W.-h.; YI, K.-w. Relationship between fluid flow stability and submerged entry nozzle port angle in a conventional slab continuous-casting mold. Metals and Materials International, Springer, p. 1–14, 2020. Citado 2 vezes nas páginas 47 e 48.
- LIU, C.-L.; HSAIO, W.-H.; TU, Y.-C. Time series classification with multivariate convolutional neural network. IEEE Transactions on industrial electronics, IEEE, v. 66, n. 6, p. 4788–4797, 2018. Citado na página 130.
- LIVIERIS, I. E.; PINTELAS, E.; PINTELAS, P. A cnn-lstm model for gold price time-series forecasting. Neural computing and applications, Springer, v. 32, n. 23, p. 17351–17360, 2020. Citado 2 vezes nas páginas 106 e 116.
- MARTINO, A.; RIZZI, A.; MASCIOLI, F. M. F. Supervised approaches for protein function prediction by topological data analysis. In: IEEE. 2018 International joint conference on neural networks (IJCNN). [S.l.], 2018. p. 1–8. Citado na página 85.
- MCPHERSON, N.; MCLEAN, A. Continuous Casting: Tundish to Mold Transfer Operations. [S.l.]: Iron and Steel Society, 1992. Citado 2 vezes nas páginas 46 e 47.
- MILLMAN, S. Clean steel-basic features and operation practices. IISI Study on Clean Steel; Wünnenberg, K., Millman, S., Eds, p. 39–60, 2004. Citado na página 43.
- MIOT, H. A. Avaliação da normalidade dos dados em estudos clínicos e experimentais. [S.l.]: SciELO Brasil, 2017. 88–91 p. Citado na página 101.
- MODI, S.; LIN, Y.; CHENG, L.; YANG, G.; LIU, L.; ZHANG, W. J. A socially inspired framework for human state inference using expert opinion integration. IEEE/ASME Transactions on Mechatronics, IEEE, v. 16, p. 874–878, 2011. Citado na página 130.
- MOURÃO, M. B.; YOKOJI, A.; MALYNOWSKYJ, A.; LEANDRO, C.; TAKANO, C.; QUITES, E.; GENTILE, E.; SILVA, G.; BOLOTA, J.; GONÇALVES, M. et al. Introdução à siderurgia. São Paulo: Associação Brasileira de Metalurgia, Materiais e Mineração, 2011. Citado na página 39.
- NAIR, V.; HINTON, G. E. Rectified linear units improve restricted boltzmann machines. In: Icml. [S.l.: s.n.], 2010. Citado na página 73.
- NASCIMENTO, R. P. Remoção de inclusões não metálicas do aço líquido através do borbulhamento de argônio pelo fundo distribuidor de lingotamento contínuo da Arcelormittal Tubarão. Dissertação (Mestrado) — Universidade Federal de Ouro Preto, Ouro Preto, 2008. Citado na página 30.
- PELLEGRINI, G.; SANDRI, M.; VILLAGROSSI, E.; CHALLAPALLI, S.; CESTARI, L.; POLO, A.; OMETTO, M. Successful use case applications of artificial intelligence in the steel industry. Iron & Steel Technology (AIST), Digital Transformations, p. 44–53, 2019. Citado 10 vezes nas páginas 30, 31, 32, 34, 35, 65, 67, 87, 125 e 128.

- PEREIRA, R. P. d. A.; ALMEIDA, G. M. d.; SALLES, J. L. F.; CUADROS, M. A. d. S.; VALADÃO, C. T.; FREITAS, R. O. d.; BASTOS-FILHO, T. A model-based predictive controller of the level of steel in the mold with disturbances using a repetitive structure. Metals, MDPI, v. 11, n. 9, p. 1458, 2021. Citado 5 vezes nas páginas [41](#), [51](#), [53](#), [95](#) e [96](#).
- QIAN, B.; XIAO, Y.; ZHENG, Z.; ZHOU, M.; ZHUANG, W.; LI, S.; MA, Q. Dynamic multi-scale convolutional neural network for time series classification. IEEE access, IEEE, v. 8, p. 109732–109746, 2020. Citado na página [130](#).
- RACKERS, K.; THOMAS, B. Clogging in continuous casting nozzles. In: CITeseer. Steelmaking Conference Proceedings. [S.l.], 1995. v. 78, p. 723–736. Citado 9 vezes nas páginas [30](#), [35](#), [42](#), [43](#), [46](#), [47](#), [54](#), [55](#) e [96](#).
- RACKERS, K. G. Mechanism and mitigation of clogging in continuous casting nozzles. Dissertação (Mestrado) — University of Illinois at Urbana-Champaign, 1995. Citado 16 vezes nas páginas [30](#), [32](#), [35](#), [39](#), [40](#), [41](#), [42](#), [43](#), [44](#), [45](#), [48](#), [49](#), [51](#), [95](#), [96](#) e [125](#).
- RAJENDRA, T.; RACHAPPA, K.; ANIL, S. Development and implementation of nozzle clogging index at continuous caster to improve length of casting. International Journal of Engineering and Technology, v. 5, n. 6, p. 10–14, 2016. Citado 6 vezes nas páginas [31](#), [54](#), [58](#), [59](#), [65](#) e [67](#).
- ROGERS, S.; GIROLAMI, M. A first course in machine learning. [S.l.]: Chapman and Hall/CRC, 2016. Citado 2 vezes nas páginas [32](#) e [83](#).
- ROUT, B. K.; SINGH, R. K.; CHOUDHARY, S.; DAS, C. L. Development and application of nozzle clogging index to improve the castability in continuous slab casting. In: International Conference on Advances in Refractories and Clean Steel Making, At Ranchi, India. [S.l.: s.n.], 2013. Citado 9 vezes nas páginas [30](#), [31](#), [32](#), [35](#), [54](#), [57](#), [58](#), [67](#) e [87](#).
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. Nature, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Citado na página [71](#).
- SAHAI, Y.; EMI, T. Tundish Technology for Clean Steel Production. [S.l.: s.n.], 2008. 1-316 p. Citado 5 vezes nas páginas [35](#), [42](#), [43](#), [45](#) e [96](#).
- SALLES, J. L. F.; ALMEIDA, G. M. de. Controle preditivo: sintonia e aplicações na siderurgia. [S.l.]: Appris Editora e Livraria Eireli-ME, 2016. Citado 3 vezes nas páginas [41](#), [51](#) e [53](#).
- SASAKA, I.; HARADA, T.; SHIKANO, H.; TANAKA, I. Improvement of porous plug and bubbling upper nozzle for continuous casting. In: Steelmaking Conf. Proc. [S.l.: s.n.], 1991. v. 74, p. 349–356. Citado na página [46](#).
- SCHMIDT, M.; RUSSO, T.; BEDERKA, D. Steel shrouding and tundish flow control to improve cleanliness and reduce plugging. Iron and Steel Society, Tundish Metallurgy., v. 2, p. 3–12, 1991. Citado na página [45](#).
- SHAO, X.; KIM, C. S.; KIM, D. G. Accurate multi-scale feature fusion cnn for time series classification in smart factory. Comput. Mater. Contin., v. 65, n. 1, p. 543–561, 2020. Citado na página [130](#).

- SHARIPOV, R. Quick introduction to tensor analysis. arXiv preprint math/0403252, 2004. Citado na página 108.
- SIDMAR, N.; HAERS, F.; BAUWENS, R.; CROOCK, R. de; MOOR, M. de. First experience in using the caster tube change device(tcd 90). In: 4 th International Conference Continuous Casting. Preprints. [S.l.: s.n.], 1988. v. 2, p. 503–514. Citado 2 vezes nas páginas 45 e 47.
- SKIENA, S. S. The data science design manual. [S.l.]: Springer, 2017. Citado na página 101.
- SONG, G. W.; TAMA, B. A.; PARK, J.; HWANG, J. Y.; BANG, J.; PARK, S. J.; LEE, S. Temperature control optimization in a steel-making continuous casting process using a multimodal deep learning approach. Steel Research International, Wiley Online Library, v. 90, p. Article 1900321, 2019. Citado na página 130.
- STONE, M. Cross-validators choice and assessment of statistical predictions. Journal of the royal statistical society: Series B (Methodological), Wiley Online Library, v. 36, n. 2, p. 111–133, 1974. Citado na página 83.
- SUN, Z.; ZHAO, M. Short-term wind power forecasting based on vmd decomposition, convlstm networks and error analysis. IEEE Access, IEEE, v. 8, p. 134422–134434, 2020. Citado 2 vezes nas páginas 108 e 116.
- SZEKERES, E. Review of strand casting factors affecting steel product cleanliness. In: 4th International Conference on Clean Steel,,(Balatonszeplak, Hungary). [S.l.: s.n.], 1992. Citado 5 vezes nas páginas 35, 44, 45, 46 e 47.
- TAI, M.-C.; CHEN, C.-H.; CHOU, C.-L. Development and benefits of four-port submerged nozzle for bloom continuous casting. Continuous Casting'85, p. 19, 1985. Citado na página 46.
- TAVARES, R. P. Influência da geometria da válvula submersa no escoamento do aço no molde de lingotamento contínuo e suas implicações na qualidade superficial de placas. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, 2019. Citado na página 42.
- THOMAS, B.; HUANG, X. Effect of argon gas on fluid flow in a continuous slab casting mold. In: 76 th Steelmaking Conference. [S.l.: s.n.], 1993. p. 273–289. Citado na página 44.
- THOMAS, B. G.; BAI, H. Tundish nozzle clogging-application of computational models. In: Steelmaking Conference Proceedings. [S.l.: s.n.], 2001. v. 84, p. 895–912. Citado 17 vezes nas páginas 30, 32, 35, 42, 43, 44, 45, 48, 49, 51, 52, 53, 63, 91, 95, 96 e 125.
- THOMAS, B. G.; ZHANG, L. Mathematical modeling of fluid flow in continuous casting. ISIJ international, The Iron and Steel Institute of Japan, v. 41, n. 10, p. 1181–1193, 2001. Citado na página 47.
- TYAGI, A. K.; REDDY, V. K. et al. Performance analysis of under-sampling and over-sampling techniques for solving class imbalance problem. 2019. Citado na página 100.

UEMURA, K.-i.; TAKAHASHI, M.; KOYAMA, S.; NITTA, M. Filtration mechanism of non-metallic inclusions in steel by ceramic loop filter. ISIJ international, The Iron and Steel Institute of Japan, v. 32, n. 1, p. 150–156, 1992. Citado na página 43.

VANNUCCI, M.; COLLA, V. Novel classification method for sensitive problems and uneven datasets based on neural networks and fuzzy logic. Applied Soft Computing, Elsevier, v. 11, n. 2, p. 2383–2390, 2011. Citado 9 vezes nas páginas 30, 31, 32, 35, 64, 67, 87, 125 e 128.

VANNUCCI, M.; COLLA, V.; CATENI, S. An hybrid ensemble method based on data clustering and weak learners reliabilities estimated through neural networks. In: SPRINGER. International Work-Conference on Artificial Neural Networks. [S.l.], 2015. p. 400–411. Citado 3 vezes nas páginas 35, 64 e 67.

VANNUCCI, M.; COLLA, V.; NASTASI, G.; MATARESE, N. Detection of rare events within industrial datasets by means of data resampling and specific algorithms. Int. J. Simul. Syst. Sci. Technol, v. 11, n. 3, p. 1–11, 2010. Citado na página 64.

VIANA, F. L. E. Indústria siderúrgica. Caderno Setorial ETENE, Banco do Nordeste do Brasil, v. 5, n. 132, p. 190–209, 2020. Citado na página 29.

WANG, R.; LI, H.; GUERRA, F.; CATHCART, C.; CHATTOPADHYAY, K. Development of quantitative indices and machine learning-based predictive models for sen clogging. In: AISTech 2021, proceedings of the iron & steel technology conference. [S.l.: s.n.], 2021. Citado 4 vezes nas páginas 54, 65, 66 e 67.

WILCOXON, F. Individual comparisons by ranking methods. Biometrics Bulletin, v. 1, p. 80–83, 1945. Citado na página 117.

WILSON, F.; HEESOM, M.; NICHOLSON, A.; HILLS, A. Effect of fluid-flow characteristics on nozzle blockage in aluminum-killed steels. Ironmaking & Steelmaking, INST MATERIALS 1 CARLTON HOUSE TERRACE, LONDON, ENGLAND SW1Y 5DB, v. 14, n. 6, p. 296–309, 1987. Citado 2 vezes nas páginas 43 e 47.

XIAN, Q.; LIANG, W. A multi-modal time series intelligent prediction model. In: SPRINGER. Proceeding of 2021 International Conference on Wireless Communications, Networking and Applications. [S.l.], 2022. p. 1150–1157. Citado na página 130.

XINGJIAN, S.; CHEN, Z.; WANG, H.; YEUNG, D.-Y.; WONG, W.-K.; WOO, W.-c. Convolutional lstm network: A machine learning approach for precipitation nowcasting. In: Advances in neural information processing systems. [S.l.: s.n.], 2015. p. 802–810. Citado 3 vezes nas páginas 80, 81 e 82.

YANG, H.; OLIA, H.; THOMAS, B. G. Modeling air aspiration in steel continuous casting slide-gate nozzles. Metals, Multidisciplinary Digital Publishing Institute, v. 11, n. 1, p. 116, 2021. Citado 8 vezes nas páginas 35, 40, 42, 43, 44, 51, 52 e 96.

YANG, S.; REBMANN, A.; TANG, M.; MORAVEC, R.; BEHRMANN, D.; BAIRD, M.; BEQUETTE, B. W. Process monitoring using causal graphical models, with application to clogging detection in steel continuous casting. Journal of Process Control, Elsevier, v. 105, p. 259–266, 2021. Citado 4 vezes nas páginas 31, 36, 66 e 103.

ZHANG, L.; THOMAS, B. G. Inclusions in continuous casting of steel. In: CITESEER. XXIV National Steelmaking Symposium, Morelia, Mich, Mexico. [S.l.], 2003. v. 26, p. 28. Citado na página [43](#).



# APÊNDICE A – Curvas de Treinamento e Validação das redes LSTM

Nas Figuras 51 e 52, é possível verificar o comportamento das curvas de treinamento e validação ao longo do processo de treinamento dos modelos LSTM contendo 2 neurônios na camada de saída, conforme as configurações apresentadas pela Tabela 7.

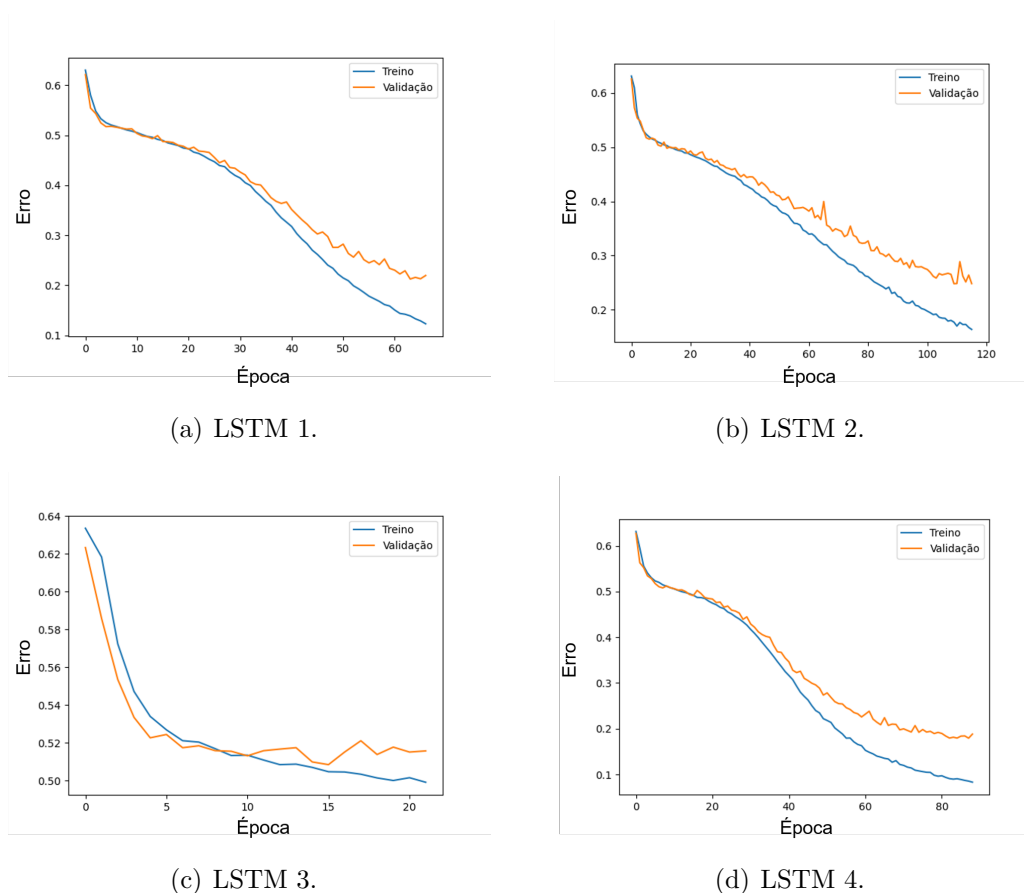
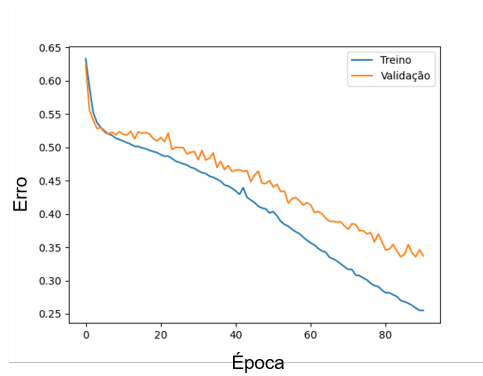
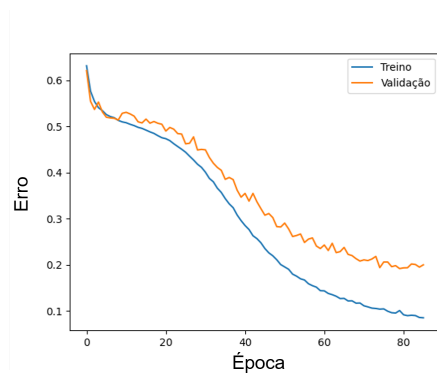


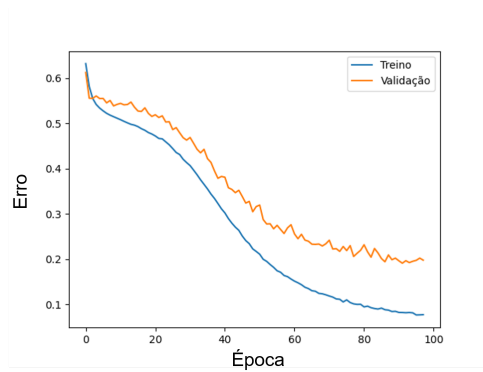
Figura 51 – Curvas de Treinamento e Validação dos modelos LSTM 1, LSTM 2, LSTM 3 e LSTM 4.



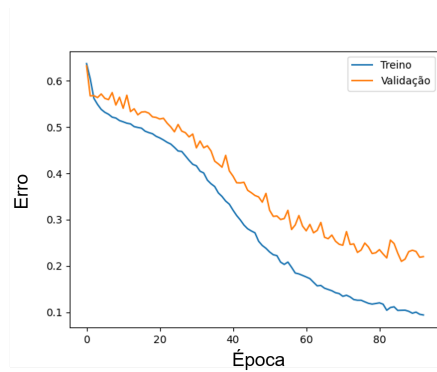
(a) LSTM 5.



(b) LSTM 6.



(c) LSTM 7.



(d) LSTM 8.

Figura 52 – Curvas de Treinamento e Validação dos modelos LSTM 5, LSTM 6, LSTM 7 e LSTM 8.

## APÊNDICE B – Curvas de Treinamento e Validação das redes CNN-LSTM

Nas Figuras 53 e 54, é possível verificar o comportamento das curvas de treinamento e validação ao longo do processo de treinamento dos modelos CNN-LSTM contendo 2 neurônios na camada de saída, conforme as configurações apresentadas pela Tabela 8.

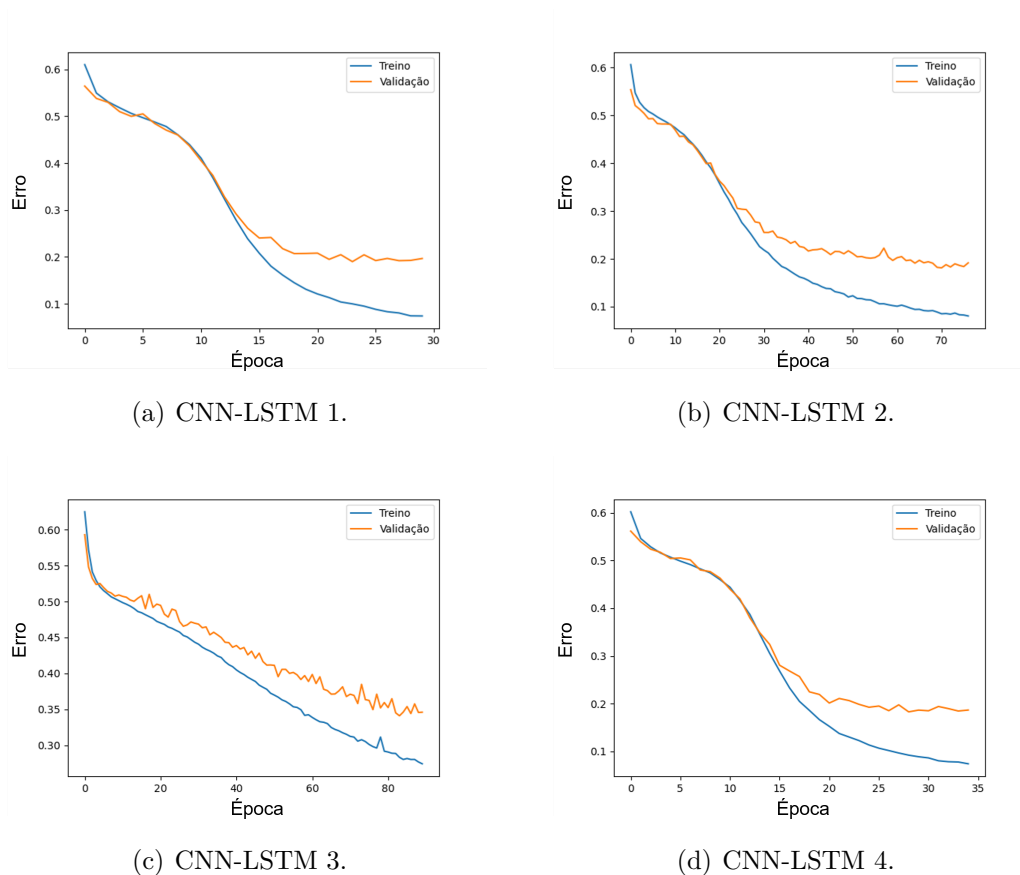
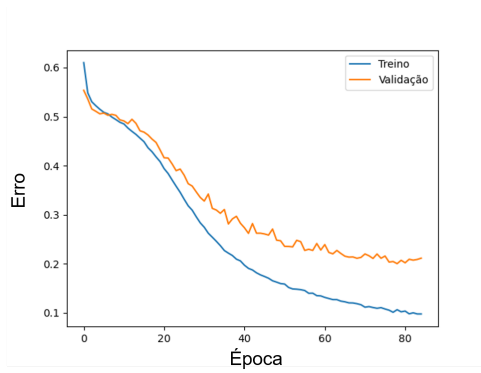
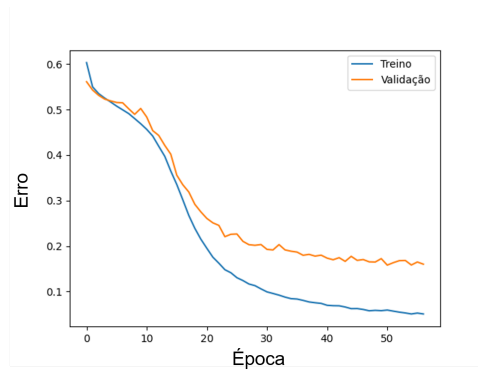


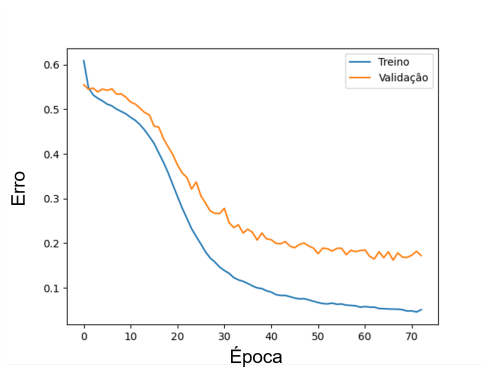
Figura 53 – Curvas de Treinamento e Validação dos modelos CNN-LSTM 1, CNN-LSTM 2, CNN-LSTM 3 e CNN-LSTM 4.



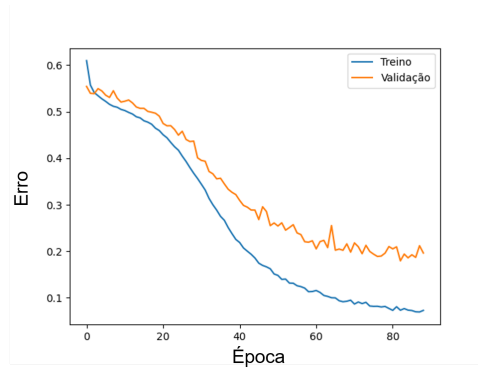
(a) CNN-LSTM 5.



(b) CNN-LSTM 6.



(c) CNN-LSTM 7.

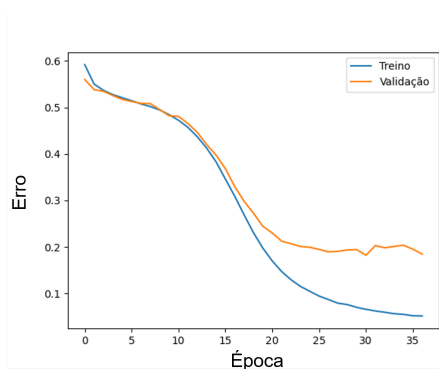


(d) CNN-LSTM 8.

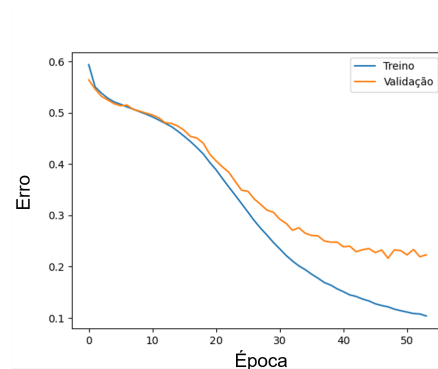
Figura 54 – Curvas de Treinamento e Validação dos modelos CNN-LSTM 5, CNN-LSTM 6, CNN-LSTM 7 e CNN-LSTM 8.

## APÊNDICE C – Curvas de Treinamento e Validação das redes ConvLSTM

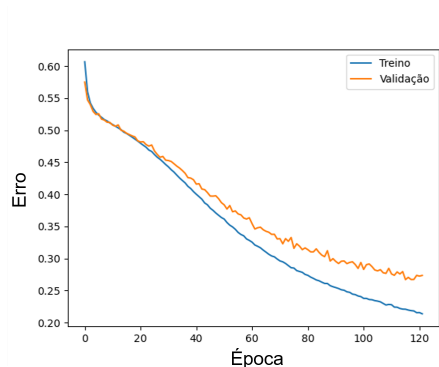
Nas Figuras 55 e 56, é possível verificar o comportamento das curvas de treinamento e validação ao longo do processo de treinamento dos modelos ConvLSTM contendo 2 neurônios na camada de saída, conforme as configurações apresentadas pela Tabela 9.



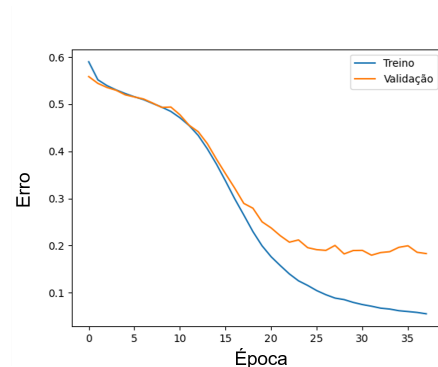
(a) ConvLSTM 1.



(b) ConvLSTM 2.



(c) ConvLSTM 3.



(d) ConvLSTM 4.

Figura 55 – Curvas de Treinamento e Validação dos modelos ConvLSTM 1, ConvLSTM 2, ConvLSTM 3 e ConvLSTM 4.

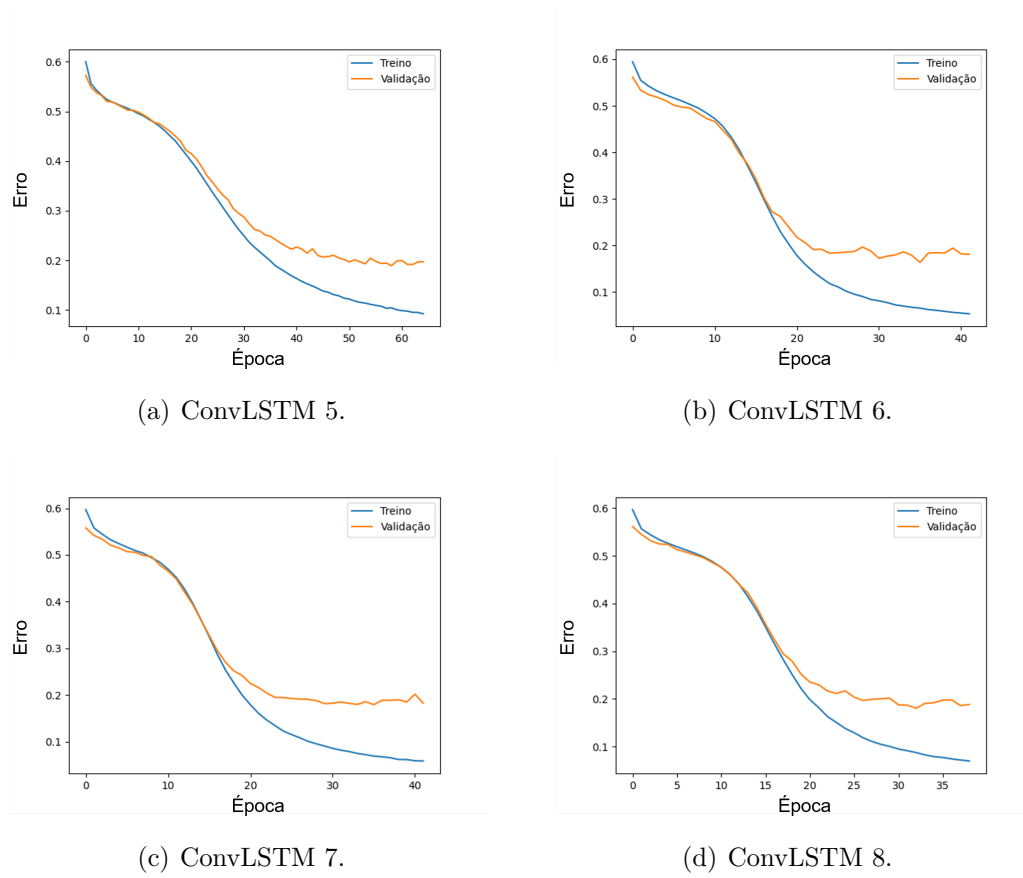
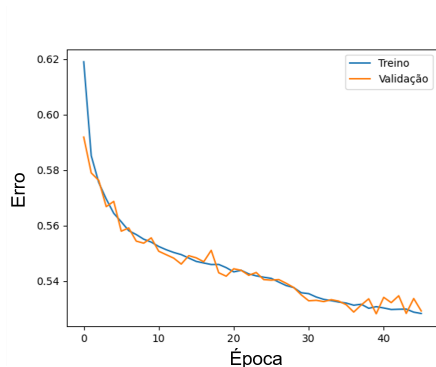


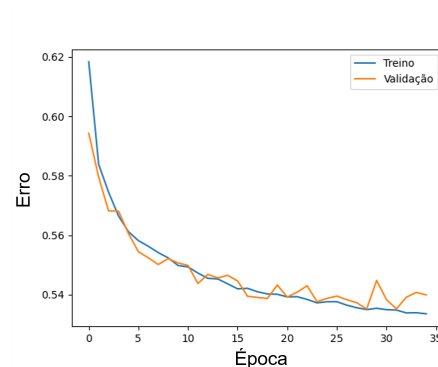
Figura 56 – Curvas de Treinamento e Validação dos modelos ConvLSTM 5, ConvLSTM 6, ConvLSTM 7 e ConvLSTM 8.

## APÊNDICE D – Curvas de Treinamento e Validação das redes MLP com 1 Camada Oculta

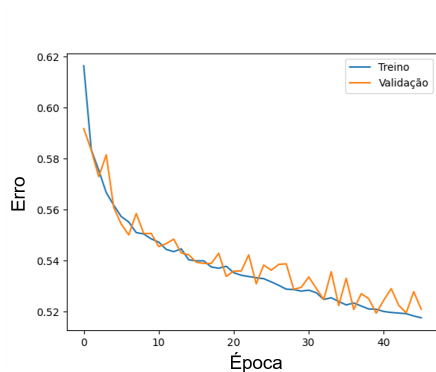
Nas Figuras 57 e 58, é possível observar o comportamento das curvas de treinamento e validação ao longo do processo de treinamento dos modelos MLP contendo 1 camada oculta e 2 neurônios na camada de saída.



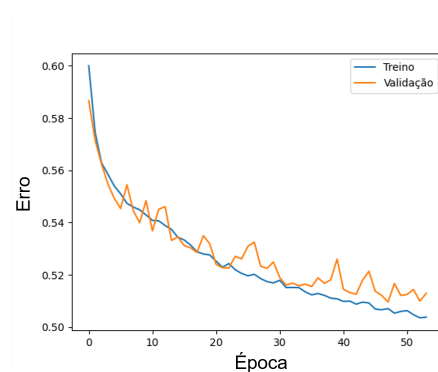
(a) MLP com 16 neurônios na camada oculta.



(b) MLP com 32 neurônios na camada oculta.

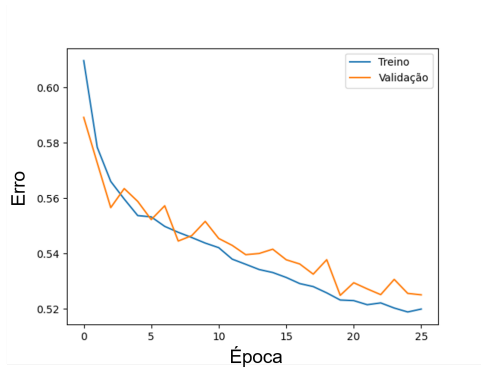


(c) MLP com 64 neurônios na camada oculta.

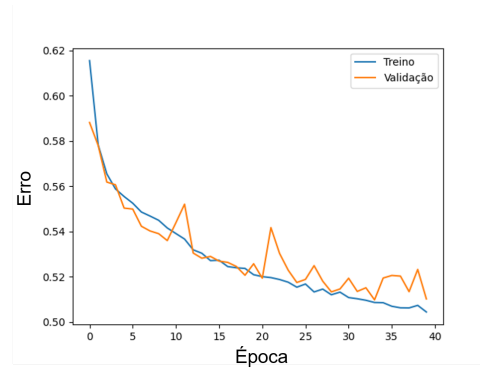


(d) MLP com 128 neurônios na camada oculta.

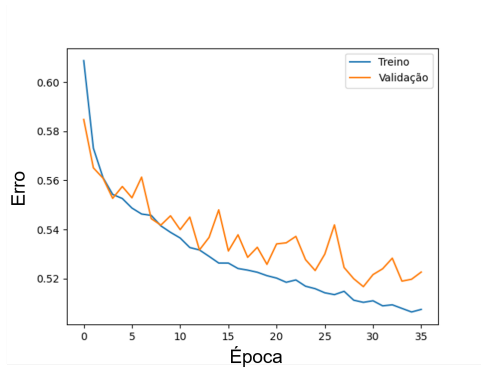
Figura 57 – Curvas de Treinamento e Validação dos modelos MLP com 16, 32, 64 e 128 neurônios na camada oculta, respectivamente.



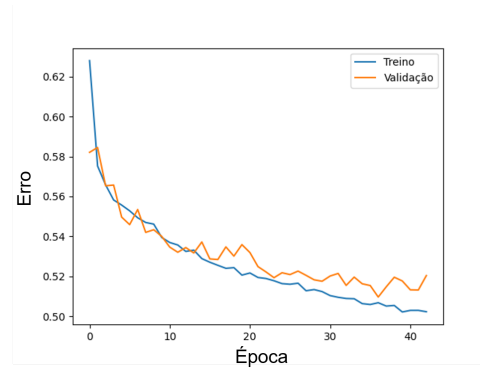
(a) MLP com 256 neurônios na camada oculta.



(b) MLP com 512 neurônios na camada oculta.

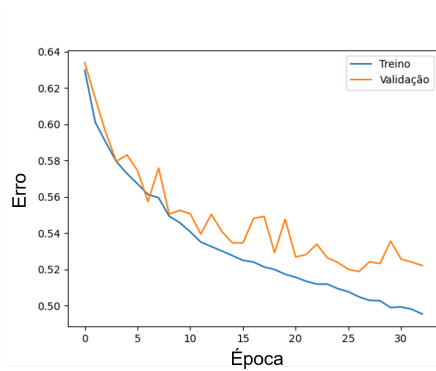


(c) MLP com 1024 neurônios na camada oculta.

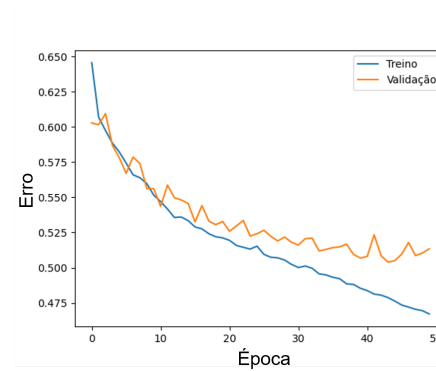


(d) MLP com 2048 neurônios na camada oculta.

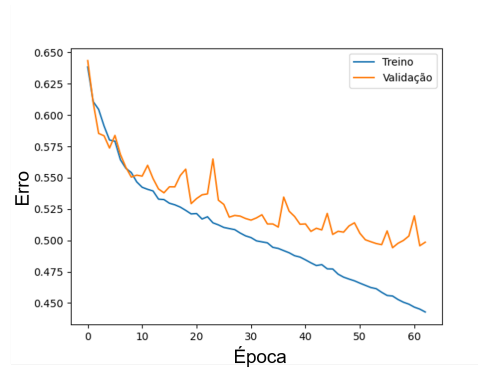
Figura 58 – Curvas de Treinamento e Validação dos modelos MLP com 256, 512, 1024 e 2048 neurônios na camada oculta, respectivamente.



(a) MLP com 3072 neurônios na camada oculta.



(b) MLP com 4096 neurônios na camada oculta.



(c) MLP com 5120 neurônios na camada oculta.

Figura 59 – Curvas de Treinamento e Validação dos modelos MLP com 3072, 4096 e 5120 neurônios na camada oculta, respectivamente.



## APÊNDICE E – Modelo MLP com duas camadas ocultas

A Tabela 20 apresenta os modelos MLP testados, os quais consistem em duas camadas ocultas configuradas conforme a estrutura (número de neurônios e quantidade de camadas ocultas) dos modelos LSTM, CNN-LSTM e ConvLSTM.

Tabela 20 – Modelos MLP com duas camadas ocultas variando o número de neurônios de 8 a 256 cada.

| Modelo | Neurônios da<br>1ª Camada Oculta | Neurônios da<br>2ª Camada Oculta | Núm. de<br>Parâmetros |
|--------|----------------------------------|----------------------------------|-----------------------|
| MLP 1  | 256                              | 128                              | 156.290               |
| MLP 2  | 128                              | 64                               | 69.954                |
| MLP 3  | 64                               | 32                               | 32.930                |
| MLP 4  | 256                              | 64                               | 139.714               |
| MLP 5  | 128                              | 32                               | 65.762                |
| MLP 6  | 256                              | 32                               | 131.426               |
| MLP 7  | 256                              | 16                               | 127.282               |
| MLP 8  | 256                              | 8                                | 125.210               |

Fonte: Elaborada pela autora.

Já na Tabela 21 são mostrados os critérios de desempenho dos modelos MLP com duas camadas ocultas e seus respectivos tempos de treinamento.

Tabela 21 – Desempenhos dos classificadores MLP com duas camadas ocultas seguido do tempo de treinamento.

| Modelo | Acurácia<br>[%] | Precisão<br>[%] | Recall<br>[%] | F1-Score<br>[%] | MCC         | Tempo de<br>Treinamento [s] |
|--------|-----------------|-----------------|---------------|-----------------|-------------|-----------------------------|
| MLP 1  | 75,85±0,65      | 77,41±0,94      | 73,04±1,55    | 76,16±1,17      | 0,518±0,009 | 232±72                      |
| MLP 2  | 75,80±0,97      | 77,78±1,02      | 72,33±1,59    | 74,95±1,25      | 0,517±0,014 | 256±47                      |
| MLP 3  | 75,11±1,21      | 77,17±1,31      | 71,32±2,37    | 74,12±1,68      | 0,504±0,011 | 239±73                      |
| MLP 4  | 75,96±0,76      | 77,07±1,89      | 72,86±2,95    | 74,90±2,30      | 0,521±0,012 | 273±78                      |
| MLP 5  | 75,42±1,27      | 77,03±2,37      | 72,46±3,31    | 74,67±2,76      | 0,509±0,011 | 268±34                      |
| MLP 6  | 75,65±2,61      | 77,37±1,43      | 72,5±2,34     | 74,85±1,78      | 0,514±0,008 | 240±50                      |
| MLP 7  | 74,49±1,68      | 76,62±1,86      | 71,68±2,45    | 74,07±2,12      | 0,512±0,009 | 252±75                      |
| MLP 8  | 74,10±2,31      | 76,11±2,35      | 71,5±2,98     | 73,73±2,62      | 0,504±0,009 | 204±87                      |

Fonte: Elaborada pela autora.



## APÊNDICE F – Resultados comparativos dos modelos após a redução do número de neurônios da camada de saída

As Tabelas 22, 23 e 24 apresentam as configurações de cada classificador contendo apenas um único neurônio na camada de saída ativado pela função *sigmóide*, bem como a quantidade de parâmetros treinados por cada modelo.

Tabela 22 – Os oito modelos LSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída.

| Modelo | Células da 1ª camada LSTM | Células da 2ª camada LSTM | Número de Parâmetros |
|--------|---------------------------|---------------------------|----------------------|
| LSTM 1 | 256                       | 128                       | 490.385              |
| LSTM 2 | 128                       | 64                        | 130.705              |
| LSTM 3 | 64                        | 32                        | 36.881               |
| LSTM 4 | 256                       | 64                        | 362.641              |
| LSTM 5 | 128                       | 32                        | 95.505               |
| LSTM 6 | 256                       | 32                        | 311.057              |
| LSTM 7 | 256                       | 16                        | 288.337              |
| LSTM 8 | 256                       | 8                         | 277.745              |

Fonte: Elaborada pela autora.

Tabela 23 – Os oito modelos CNN-LSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída.

| Modelo     | Filtros da 1ª Camada CNN1D<br>(tamanho do <i>kernel</i> ) | Filtros da 2ª Camada CNN1D<br>(tamanho do <i>kernel</i> ) | Células da 1ª Camada LSTM | Células da 2ª Camada LSTM | Número de Parâmetros |
|------------|-----------------------------------------------------------|-----------------------------------------------------------|---------------------------|---------------------------|----------------------|
| CNN-LSTM 1 | 256 (1 × 5)                                               | 128 (1 × 3)                                               | 256                       | 128                       | 721.169              |
| CNN-LSTM 2 | 128 (1 × 5)                                               | 64 (1 × 3)                                                | 128                       | 64                        | 188.753              |
| CNN-LSTM 3 | 64 (1 × 5)                                                | 32 (1 × 3)                                                | 64                        | 32                        | 51.569               |
| CNN-LSTM 4 | 256 (1 × 5)                                               | 64 (1 × 3)                                                | 256                       | 64                        | 478.673              |
| CNN-LSTM 5 | 128 (1 × 5)                                               | 32 (1 × 3)                                                | 128                       | 32                        | 124.849              |
| CNN-LSTM 6 | 256 (1 × 5)                                               | 32 (1 × 3)                                                | 256                       | 32                        | 369.713              |
| CNN-LSTM 7 | 256 (1 × 5)                                               | 16 (1 × 3)                                                | 256                       | 16                        | 318.305              |
| CNN-LSTM 8 | 256 (1 × 5)                                               | 8 (1 × 3)                                                 | 256                       | 8                         | 293.369              |

Fonte: Elaborada pela autora.

Por sua vez, as Tabelas 25, 26 e 27 ilustram os desempenhos dos modelos LSTM, CNN-LSTM e ConvLSTM, respectivamente, em termos da média e desvio padrão da Acurácia, Precisão, *Recall*, *F1-Score* e MCC, calculados sobre os testes realizados em cada uma das 5 iterações da fase de treinamento de cada modelo contendo um único neurônio

Tabela 24 – Os oito modelos ConvLSTM selecionados no Capítulo 4 para a configuração de rede com apenas um neurônio na camada de saída.

| Modelo     | Filtros da 1 <sup>a</sup><br>Camada ConvLSTM<br>(tamanho do <i>kernel</i> ) | Filtros da 2 <sup>a</sup><br>Camada ConvLSTM<br>(tamanho do <i>kernel</i> ) | Número<br>de<br>Parâmetros |
|------------|-----------------------------------------------------------------------------|-----------------------------------------------------------------------------|----------------------------|
| ConvLSTM 1 | 256 ( $1 \times 5$ )                                                        | 128 ( $1 \times 3$ )                                                        | 2.537.361                  |
| ConvLSTM 2 | 128 ( $1 \times 5$ )                                                        | 64 ( $1 \times 3$ )                                                         | 793.745                    |
| ConvLSTM 3 | 64 ( $1 \times 5$ )                                                         | 32 ( $1 \times 3$ )                                                         | 278.289                    |
| ConvLSTM 4 | 256 ( $1 \times 5$ )                                                        | 64 ( $1 \times 3$ )                                                         | 1.885.841                  |
| ConvLSTM 5 | 128 ( $1 \times 5$ )                                                        | 32 ( $1 \times 3$ )                                                         | 554.001                    |
| ConvLSTM 6 | 256 ( $1 \times 5$ )                                                        | 32 ( $1 \times 3$ )                                                         | 1.596.945                  |
| ConvLSTM 7 | 256 ( $1 \times 5$ )                                                        | 16 ( $1 \times 3$ )                                                         | 1.461.713                  |
| ConvLSTM 8 | 256 ( $1 \times 5$ )                                                        | 8 ( $1 \times 3$ )                                                          | 1.396.401                  |

Fonte: Elaborada pela autora.

na camada de saída. Essas tabelas também apresentam o tempo médio para o treinamento completo de cada modelo.

Tabela 25 – Os oito classificadores LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída.

| Modelo | Acurácia<br>[%] | Precisão<br>[%] | Recall<br>[%] | F1-Score<br>[%] | MCC         | Tempo de<br>Treinamento [s] |
|--------|-----------------|-----------------|---------------|-----------------|-------------|-----------------------------|
| LSTM 1 | 88,86±5,39      | 87,41±3,98      | 90,51±8,78    | 88,86±5,48      | 0,779±0,197 | 3019±545                    |
| LSTM 2 | 81,27±4,20      | 82,63±2,97      | 78,98±7,81    | 80,65±4,30      | 0,628±0,183 | 1847±395                    |
| LSTM 3 | 75,75±1,56      | 79,36±1,75      | 69,67±1,08    | 74,19±1,34      | 0,519±0,011 | 198±61,6                    |
| LSTM 4 | 78,94±6,48      | 81,16±4,85      | 74,96±10,06   | 77,80±6,54      | 0,582±0,179 | 3230±437                    |
| LSTM 5 | 75,90±4,29      | 79,43±3,23      | 69,99±8,12    | 74,38±4,62      | 0,520±0,106 | 1164±320                    |
| LSTM 6 | 75,57±3,40      | 78,89±4,59      | 69,62±12,96   | 73,92±6,78      | 0,513±0,109 | 2325±972                    |
| LSTM 7 | 74,67±9,51      | 76,72±5,54      | 70,93±13,84   | 73,68±7,91      | 0,495±0,211 | 2361±1029                   |
| LSTM 8 | 74,81±1,71      | 78,79±3,23      | 67,96±4,14    | 72,94±3,60      | 0,501±0,024 | 2620±824                    |

Fonte: Elaborada pela autora.

Tabela 26 – Os oito classificadores CNN-LSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída.

| <b>Modelo</b> | <b>Acurácia<br/>[%]</b> | <b>Precisão<br/>[%]</b> | <b><i>Recall</i><br/>[%]</b> | <b><i>F1-Score</i><br/>[%]</b> | <b>MCC</b>  | <b>Tempo de<br/>Treinamento [s]</b> |
|---------------|-------------------------|-------------------------|------------------------------|--------------------------------|-------------|-------------------------------------|
| CNN-LSTM 1    | 94,29±0,21              | 90,13±1,01              | 91,84±0,24                   | 90,97±0,41                     | 0,871±0,004 | 2141±310                            |
| CNN-LSTM 2    | 92,79±0,63              | 90,63±1,41              | 92,45±0,55                   | 91,53±0,79                     | 0,842±0,015 | 2810±545                            |
| CNN-LSTM 3    | 76,54±2,03              | 78,41±1,27              | 72,94±1,86                   | 75,57±1,51                     | 0,530±0,021 | 3801±870                            |
| CNN-LSTM 4    | 94,09±0,61              | 91,35±0,86              | 91,15±0,45                   | 92,24±0,59                     | 0,871±0,004 | 3952±237                            |
| CNN-LSTM 5    | 92,18±10,54             | 89,7±3,68               | 91,31±10,27                  | 90,49±5,41                     | 0,741±0,011 | 3200±333                            |
| CNN-LSTM 6    | 94,41±1,86              | 90,59±0,63              | 92,53±0,59                   | 91,54±0,61                     | 0,861±0,004 | 3070±735                            |
| CNN-LSTM 7    | 93,84±0,96              | 91,72±2,45              | 92,98±2,22                   | 92,83±2,33                     | 0,850±0,005 | 3067±335                            |
| CNN-LSTM 8    | 93,31±0,65              | 90,82±1,02              | 92,00±1,81                   | 91,40±1,32                     | 0,860±0,013 | 3864±611                            |

Fonte: Elaborada pela autora.

Tabela 27 – Os oito classificadores ConvLSTM que obtiveram os melhores desempenhos seguido do tempo médio para o treinamento completo do modelo para a configuração de rede com apenas um neurônio na camada de saída.

| <b>Modelo</b> | <b>Acurácia<br/>[%]</b> | <b>Precisão<br/>[%]</b> | <b><i>Recall</i><br/>[%]</b> | <b><i>F1-Score</i><br/>[%]</b> | <b>MCC</b>  | <b>Tempo de<br/>Treinamento [s]</b> |
|---------------|-------------------------|-------------------------|------------------------------|--------------------------------|-------------|-------------------------------------|
| ConvLSTM 1    | 95,05±0,21              | 92,92±0,61              | 96,78±0,72                   | 94,81±0,66                     | 0,879±0,004 | 5025±330                            |
| ConvLSTM 2    | 94,40±0,24              | 91,85±0,36              | 97,16±0,87                   | 94,43±0,51                     | 0,887±0,005 | 5013±321                            |
| ConvLSTM 3    | 90,66±0,58              | 84,55±0,72              | 96,15±0,86                   | 89,97±0,78                     | 0,801±0,012 | 3508±954                            |
| ConvLSTM 4    | 95,11±0,39              | 92,99±0,54              | 96,88±0,33                   | 94,89±0,41                     | 0,874±0,004 | 4602±424                            |
| ConvLSTM 5    | 94,09±0,31              | 91,41±0,49              | 96,87±0,72                   | 94,06±0,58                     | 0,864±0,006 | 3800±412                            |
| ConvLSTM 6    | 94,96±0,89              | 93,05±0,56              | 96,98±0,47                   | 94,97±0,51                     | 0,882±0,005 | 4251±390                            |
| ConvLSTM 7    | 95,01±0,17              | 93,20±0,27              | 97,02±0,76                   | 95,07±0,40                     | 0,886±0,003 | 4021±280                            |
| ConvLSTM 8    | 94,80±0,21              | 93,21±0,20              | 96,90±0,42                   | 95,01±0,27                     | 0,881±0,003 | 3989±400                            |

Fonte: Elaborada pela autora.