

# Predição da Temperatura do Ferro-Gusa em um Alto-Forno utilizando Redes Neurais LSTM

Rodrigo Seidel  
Programa de Pós-graduação em  
Computação Aplicada  
Instituto Federal do Espírito Santo  
Serra, ES, Brasil  
rodseidel@gmail.com

Karin Satie Komati  
Programa de Pós-graduação em  
Computação Aplicada  
Instituto Federal do Espírito Santo  
Serra, ES, Brasil  
kkomati@ifes.edu.br

Thiago Oliveira Santos  
Universidade Federal do Espírito  
Santo  
Vitória, ES, Brasil  
todsantos@inf.ufes.br

Francisco de Assis Boldt  
Instituto Federal do Espírito Santo  
Serra, ES, Brasil  
franciscoa@ifes.edu.br

Filipe Mutz  
Instituto Federal do Espírito Santo  
Serra, ES, Brasil  
filipe.mutz@ifes.edu.br

Leandro Colombi Resendo  
Instituto Federal do Espírito Santo  
Serra, ES, Brasil  
leandro@ifes.edu.br

## RESUMO

Due to the importance of the steel industry in the national economy and the inherent complexity of operating a blast furnace, it is necessary to study ways to optimize its operation and the consumption of resources, therefore, this work aims to investigate the use of the LSTM (Long Short Term Memory) neural network to perform the prediction of the next temperature of the hot metal being produced. In this way, it is possible to support the work of the blast furnace operators, in order to optimize the consumption of resources to keep the blast furnace operating. With the results obtained from the experiments using the blast furnace operating data as a time series, it is concluded that the use of LSTM is satisfactory and that improvement of these experiments will meet the needs of the steel industry. The best result for LSTM, using 2 layers and 2048 neurons, achieved a Root Mean Square Error of 11.82°C.

## KEYWORDS

Temperatura, Rede Neural, LSTM, Alto-Forno, Ferro-Gusa

## 1 INTRODUÇÃO

A siderurgia tem grande representatividade na economia do Brasil. De acordo com o Instituto Aço Brasil [1], a indústria brasileira do aço foi responsável por 1,7% da produção mundial em 2020, posicionando o país na 9ª colocação no ranking mundial de produtores do item. Na América Latina, o Brasil manteve-se na 1ª posição, com 56,1% da produção da região. No mesmo ano, a produção de ferro-gusa foi de 24.517 toneladas e a produção de aço bruto e de laminados foram de 30.971 e 21.664 toneladas, respectivamente [2].

Os alto-fornos são estruturas fundamentais na siderurgia e compõem sistemas extremamente complexos. Como ilustrado na Figura 1 (extraída de [3]), eles são construídos com placas de aço soldadas, plataformas, tubulações e possuem sua superfície interna revestida por tijolos refratários, com o objetivo de resistir às reações em altas temperaturas, proteger o corpo do alto-forno e melhorar a estabilidade da temperatura interna [3]. Operando com temperaturas de aproximadamente 1500°C, ele funciona como um reator onde inúmeras reações químicas e físicas ocorrem simultaneamente [4].

O objetivo básico do alto-forno é retirar o oxigênio dos óxidos de ferro, produzindo o ferro-gusa, uma das principais matérias primas para a produção do aço. Ele é extraído do alto-forno em intervalos

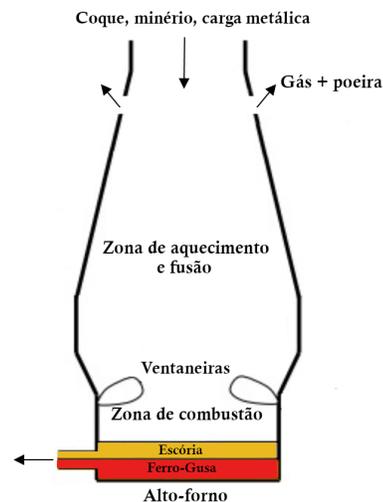


Figura 1: Estrutura típica de um alto-forno [3].

irregulares. Outro subproduto do processo é a escória que, por ter a densidade mais baixa, permanece na camada superior da mistura, e é retirada após o gusa. O alto-forno é carregado com carga metálica (sucata, sinter, pelotas, minério granulado, etc) e coque (material sólido obtido a partir do carvão mineral), inseridos no topo do alto-forno. O coque é usado para fornecer calor, aumentando a temperatura interna o suficiente para a redução do minério de ferro. O aquecimento é enriquecido com oxigênio e outros combustíveis auxiliares que são soprados no interior do alto-forno.

Além dos materiais inseridos, um dos principais fatores que influenciam o processo é a temperatura. Ela controla a taxa das reações químicas no processo de combustão e tem um efeito fundamental no produto final [5]. Se a temperatura começar a cair, a capacidade de derretimento e o processo de redução diminuirão e o equilíbrio térmico do forno será reduzido. Quando há um aumento na temperatura, a zona de derretimento torna-se instável, comprometendo a qualidade do produto, a integridade do alto-forno e aumentando a possibilidade de acidentes. Desta forma, o controle da temperatura tem um papel fundamental e influencia diversos indicadores

importantes como o custo com o consumo de energia e material, o controle do processo, que proporciona o aumento do tempo de vida do alto-forno e equipamentos envolvidos, a qualidade do produto final e a segurança. Em diversas siderúrgicas o controle de temperatura é realizado por operadores que realizam ações de ajuste com objetivo de manter a temperatura em um valor ideal de 1500°C. Estas ações são escolhidas com base em uma série de parâmetros, e cada uma delas podem gerar efeitos de curto prazo (minutos), médio prazo (1 ou 2 horas) ou longo prazo (até 8 horas) [6].

As informações coletadas para apoiar as ações de controle da temperatura são originadas de diversas fontes e contabilizam centenas de variáveis de processo que são periodicamente monitoradas e armazenadas. São coletadas temperaturas medidas por termopares espalhados por toda a carcaça do forno, dados dos materiais e medições da composição química dos gases no topo do alto-forno, resultados de cálculos clássicos realizados pela equipe de operação, variáveis de atuação da equipe de operação, pressões medidas por sensores, temperatura dos gases em diversas posições do alto-forno, carga térmica das ventaneiras e a temperatura do ferro-gusa que está sendo produzido [7]. Uma característica importante no alto-forno é a existência de “atraso” de até 8 horas entre as ações dos operadores e a mudança da temperatura do ferro-gusa produzido. Existe, portanto, uma correlação temporal, não óbvia, entre as variáveis de controle e a temperatura do metal quente, resultando em mudanças ao longo do tempo.

É com base nestas informações que operadores do alto-forno interagem com o processo. Contudo, ações humanas em um ambiente tão complexo podem levar à resultados subótimos por diversos fatores como excesso de tarefas e cansaço. Assim, é possível que através da automatização deste processo com o uso de métodos baseados em otimização seja possível alcançar aumento de produtividade, redução de custos, redução de acidentes, padronização de atuação entre equipes e aumento de tempo de vida útil dos equipamentos. Para construção de um sistema automático de otimização é necessário simular o efeito das ações dos operadores, de forma que, seja possível identificar qual o impacto da alteração de uma ou mais determinadas variáveis na temperatura.

Portanto, este trabalho propõe e avalia modelos para predição da temperatura do ferro-gusa produzido pelo alto-forno a partir de diversos dados coletados enquanto o equipamento estava em uso. Dentre estes dados estão o resultado de ações realizadas por operadores com objetivo de controlar a temperatura. Como as ações realizadas em um instante influenciam o processo por um tempo, a predição de temperatura se configura como um problema de regressão em séries temporais. Dados históricos sobre o processo foram coletados, pré-processados e utilizados para treinar redes neurais do tipo *Long Short-Term Memory* (LSTM) [8, 9] na tarefa de predição de temperatura do ferro-gusa. Resultados experimentais mostram que o modelo é capaz de realizar predições com raiz do erro médio quadrático (*Root Mean Square Error* - RMSE) de 11,82°C. Foram realizadas diversas avaliações comparativas e estudos de ablação (*ablation studies*) para averiguar a qualidade do modelo proposto e os resultados foram positivos.

O restante do artigo está organizado da seguinte forma: na Seção 2 é feita uma apresentação de trabalhos correlatos recentes. A Seção 3 apresenta a descrição do conjunto de dados e os tratamentos realizados. A Seção 4 apresenta os modelos utilizados para predição

da temperatura. Na Seção 5, são apresentados os experimentos quantitativos e qualitativos realizados. Ao final, na Seção 6, temos as conclusões e proposta para trabalhos futuros.

## 2 TRABALHOS RELACIONADOS

Devido à importância do alto-forno para a indústria siderúrgica, vários estudos foram realizados sobre suas características. Em [10], foi apresentado um método para a predição do consumo de combustível, teor de sílica (visando estabilidade térmica do processo) e teor de enxofre (visando garantir qualidade do produto a ser entregue a aciaria). Para desenvolver o sistema, foram comparadas 3 redes neurais artificiais com características diferentes e um *ensemble* formado pelas 3. *Ensembles* são conjuntos de modelos com características distintas que trabalham de forma combinada para obter resultados melhores [11]. Utilizando como métrica o RMSE, foi demonstrado que o comitê obteve melhores resultados na predição do consumo de combustível, teor de sílica e teor de enxofre.

O trabalho [12] propôs um método de previsão da temperatura do alto-forno baseado em *ensembles* de modelos do tipo *extreme learning machines* [13]. As características de entrada foram selecionadas usando um algoritmo de otimização evolucionário multi-objetivo. O modelo preditivo foi treinado buscando maximizar a acurácia e a robustez. A efetividade do método de predição proposto é verificado com experimentos baseados em dados de produção. Nos experimentos, os autores afirmam que obtiveram bons resultados utilizando 19 variáveis de entrada.

Em [14], é apresentado um guia operacional do controle de temperatura de um alto-forno. Baseado em um modelo transiente, o controle com modelo preditivo não-linear (*Non-linear Model Predictive Control* - NMPC) e a estimativa de horizonte móvel (*moving horizon estimation* - MHE) foram adotados para fornecer aos operadores ações de controle adequadas, minimizando a influência indesejável de perturbações. Os resultados da validação *online* demonstraram que o sistema de orientação operacional desenvolvido reduziu com sucesso a variância da temperatura do metal em 1,9°C. Destacamos que este foi o único trabalho encontrado que se propõe a criar recomendações para os operadores de alto-forno para o controle da temperatura.

No trabalho [15], também foi indicada a importância da predição da temperatura do metal quente em um alto-forno com objetivo de manter a operação estável no processo de fabricação de ferro. No trabalho foram avaliados diversos modelos preditivos, entre eles LSTM, regressão por vetores de suporte (*Support Vector Regression* - SVR) [11], florestas aleatórias (*Random Forest* - RF) [11], CNN, entre outros modelos. Foram utilizados como métricas de avaliação o RMSE e o  $R^2$ . Os resultados demonstraram que redes neurais rasas são preferíveis em detrimento de redes profundas para a previsão instantânea de temperatura devido ao seu custo reduzido e que regressão por processos Gaussianos e SVR são preferíveis para previsões de temperatura 1h ou 2h à frente.

Em [16], os autores comparam o modelo auto-regressivo integrado de médias móveis (*Autoregressive Integrated Moving Average* - ARIMA) com as redes neurais LSTM e *Bidirectional LSTM* (BiLSTM) [17] no contexto de predição de valor de ações utilizando uma série temporal univariada de dados financeiros. Os resultados experimentais foram positivos em relação ao uso da LSTM e da BiLSTM (esta

obteve melhores resultados) demonstrando que é possível utilizar as redes neurais para aprendizado de séries temporais.

Diferente dos trabalhos que encontramos na literatura, neste trabalho é proposta a utilização de redes neurais baseadas em LSTM para predição da próxima temperatura do ferro-gusa produzido pelo alto-forno usando como entrada dados coletados do alto-forno dispostos como uma série temporal multivariada. A performance da rede neural é comparada com diversos *baselines* e estudos de ablação são realizados.

### 3 BASE DE DADOS

Esta seção descreve o procedimento para coleta e pré-processamento da base de dados utilizada para treinar e avaliar os modelos. Durante a produção do ferro-gusa, diversos dados são coletados sobre o estado do alto-forno e das substâncias que entram e saem do equipamento. Todos os dados, exceto aqueles relacionados à introdução (carga) e coleta (drenagem) de materiais são adquiridos a cada 30 minutos. Os dados de carga são coletados sempre que materiais são introduzidos no topo do alto-forno e os dados de drenagem do ferro-gusa são coletados no momento em que ele é vazado do alto-forno. Estas atividades podem ser realizadas em intervalos que variam de poucos minutos até uma hora. A temperatura do ferro-gusa, a variável que desejamos prever neste trabalho, é medida durante o vazamento do material.

Para uniformizar o período das medições e considerando a frequência da maioria dos sensores, foi convencionado que pacotes de dados seriam produzidos a cada 30 minutos. Um pacote de dados é um vetor de números reais contendo medições das variáveis de interesse. Nos casos em que mais de uma medição foi realizada para uma variável no período, é retornada a média das medições. Durante toda a coleta da base de dados, não houve falta de dados nos intervalos de 30min. Não foram coletadas amostras nos momentos em que o alto-forno estava fora de uso. Este fato pode ser identificado através de uma das variáveis, o volume de sopro. Pelo valor desta variável é possível avaliar se o alto-forno está sendo utilizado ou não.

A base de dados foi coletada no período de janeiro de 2019 a maio de 2021 e possui 40.013 amostras. Inicialmente, cada amostra possuía um total de 430 atributos provenientes de sensores (termopares na carcaça do alto-forno, sensores no topo do equipamento para medição de pressão, temperatura e composição química dos gases), ações de controle de operadores e resultados de cálculos matemáticos comumente utilizados para avaliação do estado do processo. Com objetivo de reduzir o número de atributos e identificar aqueles mais relacionados com os objetivos deste trabalho, foram realizadas entrevistas com especialistas e engenheiros de processo. Como resultado, foram selecionados 92 dos 439 atributos para compor a base de dados final. Os atributos são descritos na Tabela 1. A coluna *Nome* da tabela indica a categoria da fonte de dados, a coluna *Qtd* indica o número de valores naquela categoria, e a coluna *Descrição* traz uma explicação dos dados coletados. A variável objetivo, a temperatura do ferro-gusa, faz parte destes atributos.

Os dados foram organizados em quatro conjuntos sem interseção, a saber treino, validação, teste de desenvolvimento, e teste final. A divisão foi feita com base nos períodos de aquisição. O conjunto de treino compreende o período de 01/01/2019 até 08/10/2020 e

**Tabela 1: Atributos selecionados para a base de dados.**

Nome	Qtd	Descrição
Temperaturas	28	Temperaturas de sensores na região inferior do alto-forno.
Cargas Térmicas	38	Cargas térmicas medidas pelos gases injetados no alto-forno por ventaneiras existentes ao redor do equipamento.
Indicadores Calculados	10	Indicadores produzidos por modelos metalúrgicos, sendo eles: índice de permeabilidade, taxa de carvão O <sub>2</sub> , vazão gás rampa, relação CO/CO <sub>2</sub> , velocidade de produção de ferro-gusa, coque <i>rate</i> , PCI <i>rate</i> calculado, redução direta, perda térmica do <i>stave</i> e umidade na carga (kg/tonelada gusa).
Carga de Materiais	4	Dados dos materiais carregados no topo do alto-forno: taxa de coque da carga, taxa de escória da carga, relação M/C (minério/coque) e umidade do coque.
Operação	8	Dados de operação do alto-forno: volume de sopro, umidade do ar de sopro, temperatura do ar de sopro, taxa de coque, taxa de O <sub>2</sub> , pressão do ar de sopro, pressão do topo e composição de H <sub>2</sub> no topo.
Drenagem	4	Dados coletados na drenagem do ferro-gusa: identificador do furo de drenagem, se é a primeira drenagem do furo, casa de corrida (lado do alto-forno) do furo e temperatura do material (alvo de predição deste trabalho).

contém com 29.455 amostras. O conjunto de validação compreende o período de 08/10/2020 até 20/11/2020 e contém com 2000 amostras. Ele foi utilizado para calibração de hiperparâmetros e para seleção da rede neural com melhor generalização para ser retornada como resultado do processo de treinamento. É comum que no processo de construção de um método baseado em aprendizado de máquina, sejam realizadas várias iterações de treinamento e avaliação dos modelos com objetivo de buscar estratégias que levem à uma maior performance preditiva. Se o conjunto de teste for utilizado nestas avaliações, é possível que a configuração final do modelo seja superajustada para este conjunto. Para prevenir este efeito, utilizamos dois conjuntos de teste, um durante o desenvolvimento do modelo e outro apenas na avaliação final. O conjunto de teste de desenvolvimento compreende o período de 20/11/2020 até 31/12/2020 e contém 2000 amostras. O conjunto de teste final compreende o período de janeiro a maio de 2021 e totaliza 6.558 amostras.

Os valores dos sensores e das temperaturas por vezes alcançavam valores na ordem dos milhares. Antes de usar os dados como entrada para a rede neural, eles foram normalizados. Considere a situação em que os dados de um sensor são, em geral, próximos de zero e raramente existem picos com valores grandes. Mesmo com a normalização tradicional (de tornar a média zero e o desvio padrão unitário [18]), os valores do sensor nos picos podem continuar

tendo magnitude significativamente maiores que 1. Estes valores podem introduzir instabilidades no treinamento e nas predições feitas pelas redes neurais. Assim, foi realizada a normalização a partir da subtração da média, seguida pela divisão pelo máximo valor absoluto de forma a garantir que a magnitude máxima seria 1. Os parâmetros utilizados na normalização foram obtidos a partir do conjunto de treinamento e utilizados para normalizar todos os conjuntos.

#### 4 MODELOS

As redes neurais do tipo *Long Short-Term Memory* (LSTM) são redes recorrentes muito utilizadas para processamento de séries temporais [8, 9, 19]. Devido à sua capacidade de armazenar informações em sua memória interna e ao uso de funções de ativação não-lineares em cada camada, o modelo LSTM pode capturar tendências não-lineares nos dados e lembrar informações anteriores por um longo tempo.

A Figura 2 ilustra as células das redes LSTM. Elas usam três tipos de portas de dados, as portas de entrada (*input gate*), esquecimento (*forget gate*) e saída (*output gate*). Estas estruturas controlam que informações entram e saem da memória e os valores retornados pela rede, respectivamente [9].

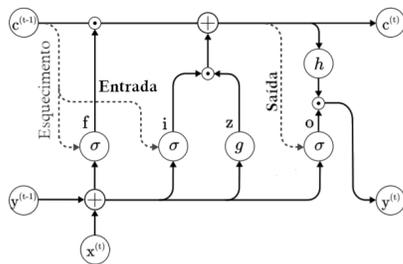


Figura 2: Célula LSTM

Neste trabalho, as redes neurais recebem como entrada os últimos 32 pacotes de dados, coletados a cada meia-hora, e produzem como saída uma estimativa da temperatura de ferro-gusa que será obtida na próxima medição. É importante enfatizar que as temperaturas anteriores fazem parte dos pacotes de dados. A escolha por utilizar 32 pacotes foi baseado no fato de que existe “atraso” entre a ação sobre as variáveis de controle e a temperatura do ferro-gusa. Ações realizadas em até 8 horas antes da medição podem influenciar a temperatura do produto [10]. Há uma correlação de tempo complexa entre as variáveis de controle e a temperatura do metal quente, resultando em mudanças ao longo do tempo. A janela de 32 pacotes corresponde a 16 horas de operação. Foram utilizadas 16h ao invés de 8h para tornar o sistema robusto a possíveis atrasos nas medições. Portanto, para prever a temperatura do instante  $t$ , a rede neural recebe como entrada as medições no intervalo de  $t - 16h$  até  $t$ , inclusive as temperaturas, mas, obviamente, excluindo a temperatura a ser predita.

Além da LSTM tradicional, também foi avaliada a performance da *Bidirectional LSTM* (BiLSTM) [17]. Este modelo processa a série temporal na direção crescente de tempo e na direção decrescente e integra os dados em seu estado interno. Esta estratégia permite que

a rede utilize dados do início e do fim da série em suas predições sem a necessidade de memorizar estas informações por muitos passos.

Foram avaliadas diversas configurações de hiperparâmetros das redes neurais incluindo o uso de 1 ou 2 camadas e números de neurônios por camadas que variavam de 2 até 2048. O estado da rede ao final do processamento da série temporal é utilizado como entrada para uma camada de saída com 1 neurônio e função de ativação linear.

Também foram realizados experimentos adicionais com a LSTM com duas camadas de 2048 neurônios utilizando técnicas de aumento de dados (*data augmentation*) que geram padrões sintéticos visando aumentar a capacidade de generalização de modelos treinados e reduzir *overfitting* [20]. Os resultados destes experimentos indicaram que o uso destas técnicas adicionais não levaram a obtenção de melhores resultados.

Para o treinamento das redes, foi utilizado o otimizador Adam [21] e *batches* de tamanho 32. As redes foram treinadas por 50 épocas com taxa de aprendizado com decaimento linear iniciando em  $1e - 3$  e com valor final de  $1e - 6$  e foi verificada convergência nos gráficos de evolução do erro. Durante o treinamento, a cada época, foi avaliado o erro no conjunto de validação e ao final do treinamento foi selecionado o modelo com melhor performance neste conjunto de validação. Este modelo foi utilizado então para avaliação usando o conjunto de teste de desenvolvimento. Apenas no fim do desenvolvimento, o conjunto de teste final foi utilizado para avaliar os modelos.

A performance das redes neurais LSTM e BiLSTM foram comparados com alguns *baselines* para averiguar a qualidade de suas predições. O primeiro *baseline* é um método que prediz a temperatura do instante  $t$  como sendo igual à temperatura do instante  $t - 1$ . Métodos de aprendizado de máquina para séries temporais comumente recorrem à esta estratégia visto que ela, em geral, leva à bons valores nas métricas de avaliação. Contudo, esta função não é desejável na prática, uma vez que ela desconsidera as tendências reais dos dados e funções mais sofisticadas que podem ser calculadas a partir das entradas.

O segundo *baseline* é a predição da temperatura como sendo a média móvel da temperatura do ferro-gusa. A média móvel consiste em calcular a média aritmética das observações mais recentes da série temporal [22]. De acordo com [23], a média móvel pode ser um método eficiente de estimativa quando a variável é estacionária, i.e., quando seus valores variam em torno de um valor médio. Foram avaliadas médias móveis considerando 4, 8 e 16 valores de temperatura mais recentes. Com janelas menores que 4, a média móvel passa a se comportar como o método que repete as temperaturas anteriores. O melhor resultado foi obtido usando a janela de 4 valores, mas também são reportados os valores com a janela de tamanho 16 visto que é o tamanho de janela usado pelos demais modelos.

Por fim, foi realizado um estudo de ablação para verificar se as variáveis medidas estavam de fato contribuindo para predições feitas pelas redes neurais. Para isto, uma versão da LSTM foi treinada usando como entrada apenas as 32 temperaturas anteriores (sem os dados das demais variáveis). Neste experimento, foi utilizada uma rede com 2 camadas LSTM e foram avaliadas configurações com 128 e 1024 neurônios. O melhor resultado foi alcançado com

128 neurônios e, portanto, os resultados são reportados para esta configuração.

## 5 RESULTADOS E DISCUSSÕES

Foram realizados experimentos quantitativos e qualitativos para avaliar a performance preditiva dos modelos. Os experimentos quantitativos consistiram em medir o erro das predições usando como métricas o RMSE, o coeficiente de determinação  $R^2$  (*R squared*) que indica o quão melhor o modelo explica os dados em comparação com o a média, e o erro absoluto relativo mediano (*Median Relative Absolute Error* - MdRAE), que compara o desempenho de dois modelos [24, 25]:

$$MdRAE = Median_{i=1,n} \left( \frac{|y_i - \hat{y}_i|}{|y_i - \hat{b}_i|} \right) \quad (1)$$

sendo  $y_i$  o valor real da observação,  $\hat{y}_i$  o valor calculado pelo modelo e  $\hat{b}_i$  o valor calculado por um dado modelo de referência.

Também foram medidas as porcentagens de amostras cujos erros ficaram abaixo de 7°C e abaixo de 20°C. Os experimentos qualitativos consistiram na inspeção visual das temperaturas reais e preditas para verificar se os modelos não recorreram a estratégias indesejáveis para minimizar o erro, e.g., repetir o último valor de temperatura.

Com objetivo de verificar se o treinamento das redes neurais foram bem sucedidos e o nível de generalização para dados diferentes daqueles usados no treinamento, a Tabela 2 traz o RMSE de diversas configurações das redes para os conjuntos de treino, validação e teste final. Os resultados para o conjunto de teste de desenvolvimento não foram reportados por serem considerados redundantes dado o conjunto de teste final. A coluna *NC* representa o número de camadas das redes e a coluna *NN* representa o número de neurônios. A tabela permite avaliar se as redes foram capazes de aprender o conjunto de treino, se houve generalização para dados novos considerando a performance nos conjuntos de validação e teste final, e se os resultados para o conjunto de validação foram superotimistas dado que foram escolhidos os modelos com melhor performance neste conjunto.

Tabela 2: Resultado do treinamento das LSTMs.

Tipo	NC	NN	RSME (Treino)	RSME (Val.)	RSME (Teste)
LSTM	1	4	11,54°C	14,07°C	14,49°C
LSTM	1	16	10,91°C	13,12°C	15,03°C
LSTM	1	64	9,86°C	15,16°C	17,34°C
LSTM	1	256	8,91°C	13,01°C	12,97°C
LSTM	2	4	12,68°C	13,86°C	18,58°C
LSTM	2	8	12,10°C	13,87°C	15,81°C
LSTM	2	32	10,78°C	13,86°C	17,77°C
LSTM	2	64	9,90°C	13,73°C	14,89°C
LSTM	2	256	8,88°C	12,78°C	14,01°C
LSTM	2	1024	7,73°C	12,11°C	12,39°C
<b>LSTM</b>	<b>2</b>	<b>2048</b>	<b>9,08°C</b>	<b>11,99°C</b>	<b>11,82°C</b>
BiLSTM	2	1024	8,11°C	12,03°C	12,49°C

Tabela 3: Performance dos modelos no conjunto de teste.

Método	RSME	$R^2$	% ≤ 7°C	% ≤ 20°C
LSTM (ablação)	9,61°C	0,83	74,51	94,20
Repetição	10,20°C	0,82	76,66	93,51
Média Móvel J4	12,75 °C	0,71	56,11	89,30
Média Móvel J16	16,42°C	0,53	37,08	80,03
LSTM	11,82°C	0,75	53,73	91,49
BiLSTM	12,49°C	0,72	49,51	90,55

Tabela 4: Comparativo de desempenho entre modelos no conjunto de teste utilizando a métrica MdRAE

Tipo	NC	NN	MdRAE Repeti- ção	MdRAE Média Móvel J4	MdRAE Média Móvel J16	MdRAE LSTM (Ab- lação)
LSTM	1	4	0,42	0,44	0,49	1,25
LSTM	1	16	0,44	0,46	0,53	1,35
LSTM	1	64	0,53	0,56	0,65	1,65
LSTM	1	256	0,35	0,37	0,42	1,11
LSTM	2	4	0,61	0,62	0,72	1,92
LSTM	2	8	0,48	0,52	0,59	1,54
LSTM	2	32	0,57	0,59	0,69	1,84
LSTM	2	64	0,46	0,48	0,54	1,39
LSTM	2	256	0,41	0,43	0,49	1,31
LSTM	2	1024	0,34	0,36	0,41	1,10
<b>LSTM</b>	<b>2</b>	<b>2048</b>	<b>0,32</b>	<b>0,34</b>	<b>0,39</b>	<b>1,01</b>
BiLSTM	2	1024	0,35	0,37	0,42	1,10

Como pode ser observado analisando a coluna do RMSE no conjunto de treino, todas as configurações foram capazes de aprender o conjunto de treino com uma boa performance. Como esperado, o erro neste conjunto diminuiu à medida que o número de neurônios aumentava. Contudo, utilizar 2048 neurônios não levou a redução do erro em relação ao uso de 1024 neurônios na LSTM com 2 camadas. Esta configuração junto com a BiLSTM foram aquelas que alcançaram menor RMSE no conjunto de treino.

O aumento no número de neurônios também levou a ganhos de performance nos conjuntos de validação e teste como pode ser observado analisando as últimas colunas da tabela. Este fato sugere que não houve superajuste (*overfitting*) aos dados de treinamento. Na verdade, a LSTM com 2 camadas e 2048 neurônios foi a rede com menor erro nestes conjuntos, mesmo sendo aquela com maior número de parâmetros. Não foram avaliadas redes maiores por limitações dos equipamentos disponíveis.

A Tabela 3 compara a performance dos modelos neurais com os *baselines* propostos no conjunto de teste final. A linha *LSTM (ablação)* apresenta os resultados para a LSTM que recebe como entrada somente as temperaturas, sem utilizar nenhuma outra variável do processo. A linha *Repetição* traz os resultados para o método que repete a última temperatura medida. As linhas *Média Móvel J4* e *Média Móvel J16* trazem os resultados obtidos com janelas de tamanho 4 e 16, respectivamente. Por fim, as linhas *LSTM* e *BiLSTM* trazem os melhores valores alcançados usando estes modelos.



Figura 3: Repetição da Última Temperatura

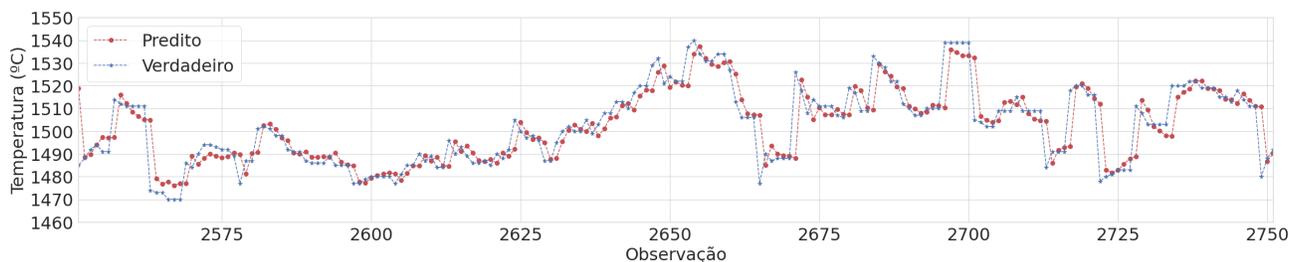


Figura 4: Ablação da LSTM usando apenas a temperatura como entrada.

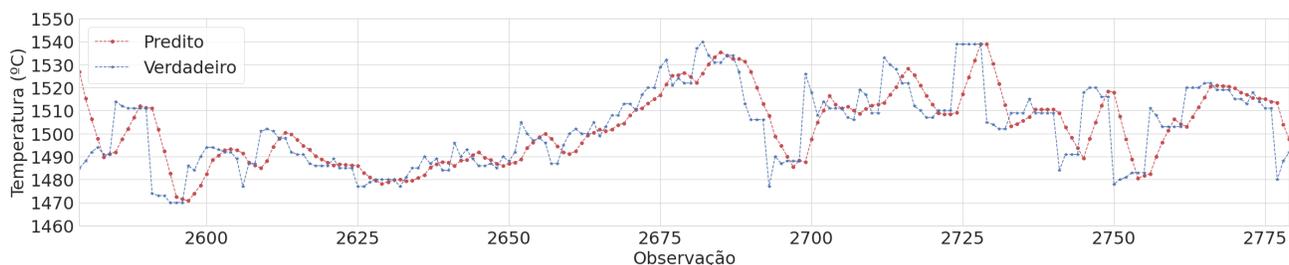


Figura 5: Média móvel com janela de tamanho 4.



Figura 6: Média móvel com janela de tamanho 16.

A Tabela 4 compara o desempenho dos modelos neurais com os *baselines* e o modelo *LSTM* (*ablação*) utilizando a métrica MdRAE, que compara o desempenho de dois modelos. Observa-se que o desempenho da *LSTM* em comparação aos *baselines* é melhor (MdRAE < 1). Entretanto, na comparação com o modelo *LSTM* (*ablação*) observou-se melhor desempenho do modelo *LSTM* *ablação* (MdRAE > 1), exceto no caso da *LSTM* com 2 camadas e 2048 neurônios, no

qual observou-se um desempenho quase equivalente (MdRAE = 1) entre os 2 modelos.

As Figuras 3 a 10 comparam a evolução real da temperatura e os valores preditos pelos modelos em 200 amostras do conjunto de teste final. O mesmo período é exibido em todas as figuras para permitir a comparação das predições. Nas figuras, a legenda indica os modelos que geraram as curvas, o eixo x é o número da amostra e o eixo y é a temperatura.

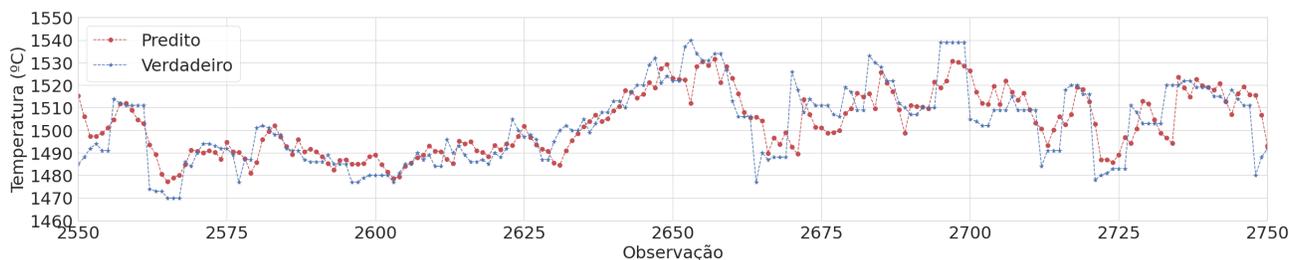


Figura 7: Rede neural com 2 camadas LSTM com 1024 neurônios.

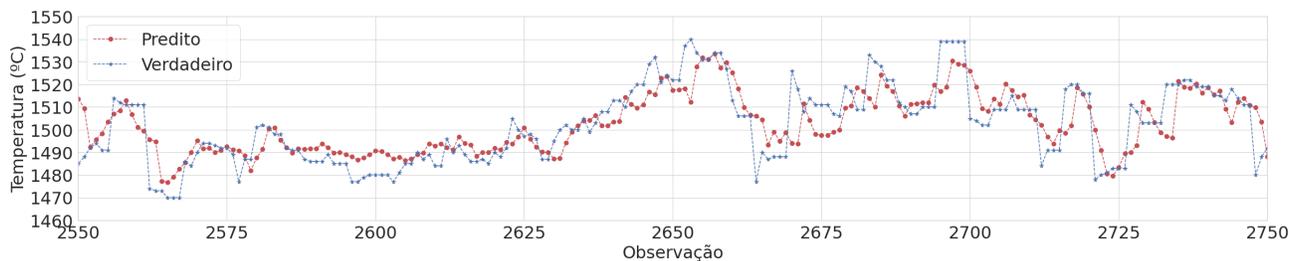


Figura 8: Rede neural com 2 camadas LSTM com 2048 neurônios.

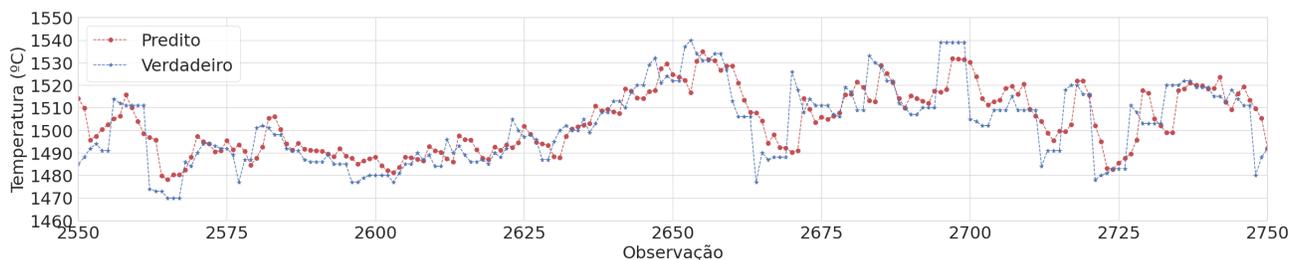


Figura 9: Rede neural com 2 camadas BiLSTM com 1024 neurônios.

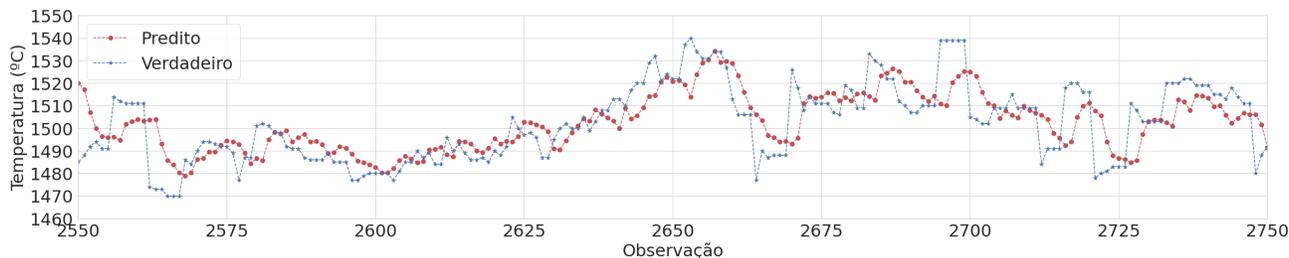


Figura 10: Rede neural com 1 camada LSTM com 4 neurônios.

Os métodos que alcançaram menor RMSE no teste final foram, respectivamente, a LSTM treinada utilizando apenas as temperaturas como entrada com RMSE de 9,61°C e  $R^2$  de 0,83, e o método que repete a última temperatura com RMSE de 10,20°C e  $R^2$  de 0,82. Apesar dos bons valores nas métricas, pode ser notado comparando a Figura 3 e a Figura 4 que esta versão da LSTM convergiu para o

comportamento de repetir a temperatura anterior. Este fato se manifesta no gráfico como um desalinhamento nas curvas causado pelo atraso nas previsões. Em geral, esta não é uma estratégia desejável.

O melhor resultado com a média móvel foi alcançado realizando a média de 4 valores anteriores com RMSE de 12,75°C e  $R^2$  de 0,71. A Figura 5 e a Figura 6 comparam as previsões obtidas usando as janelas de tamanho 4 e 16, respectivamente. Observa-se uma grande suavização de picos e vales em relação às temperaturas reais.

Além disso, é visível a existência de um atraso nas previsões que se torna mais evidente com o aumento da janela. A média móvel não considera tendências para prever valores futuros nem utiliza informações das variáveis de controle. Assim como os métodos anteriores, ela estima o futuro com base em valores passados o que não é desejável.

Dentre as redes neurais que receberam as entradas completas, a que apresentou o menor erro foi a que utilizou 2 camadas LSTM com 2048 neurônios. Ela alcançou um RMSE de  $11,82^{\circ}\text{C}$  e  $R^2$  de 0,75. As configurações com 2 camadas LSTM e BiLSTM com 1024 neurônios também foram satisfatórias com RMSE de  $12,39^{\circ}\text{C}$  e  $12,49^{\circ}\text{C}$ , respectivamente. As Figuras 7, 8 e 9 comparam as previsões feitas por estes 3 modelos. As curvas para as redes com 2 camadas LSTM com 1024 e 2048 neurônios são similares e não aparentam possuir atraso. O fato destas redes apresentarem valores pequenos de erro e, aparentemente, não terem aprendido funções triviais e indesejáveis, indicam que elas são os melhores modelos para uso em um contexto real. A curva da rede com camadas BiLSTM aparenta atraso em alguns trechos do início e meio do gráfico.

A rede com uma camada LSTM e quatro neurônios apresentou RMSE maior que as anteriores, mas ainda assim razoáveis. Em princípio, elas poderiam ser consideradas como alternativas de baixo custo em contextos com recursos limitados. A Figura 10 traz a curva que descreve as previsões aprendidas por este modelo. Embora a curva apresente atrasos eventuais e uma suavização esperada devido ao número limitado de parâmetros, a rede é capaz de prever os valores em vários trechos.

## 6 CONCLUSÃO E TRABALHOS FUTUROS

Esse trabalho avaliou a utilização de redes neurais baseadas em LSTM no problema de previsão da temperatura de ferro-gusa produzido por um alto-forno. Uma base de dados foi construída, pré-processada e utilizada para treinar e avaliar os modelos. Inicialmente, a base possuía mais de 430 valores de variáveis do processo e após entrevistas com especialistas e engenheiros de processo, foram selecionadas 92 características.

Experimentos compararam a performance das redes neurais com *baselines* que prediziam a temperatura como sendo igual à anterior e usando a média móvel das temperaturas anteriores. Além disso, foi realizado um estudo de ablação em que as redes não receberam dados das variáveis do processo, mas apenas as temperaturas anteriores. A utilização da rede neural LSTM se confirmou como melhor alternativa frente aos *baselines* utilizados. A LSTM que recebe somente a temperatura como entrada apresentou menor RMSE, porém ela adotou a estratégia de repetir as temperaturas anteriores. Entende-se que não é um comportamento esperado para a tarefa objetivo. As previsões realizadas pela LSTM que recebe todos os dados teve um comportamento mais próximo do esperado.

Em trabalhos futuros serão avaliados outros modelos para previsão da temperatura e a utilização de técnicas de *bootstrapping* ou *boosting*. Além disso, serão coletados dados de outros alto-fornos para verificar se os modelos não aprenderam características específicas e únicas do alto-forno utilizado no projeto.

## AGRADECIMENTOS

Os autores agradecem ao apoio da FAPES e CAPES (processo 2021-2S6CD, nº FAPES 132/2021) por meio do PDGP (Programa de Desenvolvimento da Pós-Graduação, Parcerias Estratégicas nos Estados). A profª Karin Komati agradece ao CNPq pela Bolsa de Produtividade DT-2 (308432/2020-7) e à FAPES pela concessão de Auxílio Taxa de Pesquisa (nº FAPES 293/2021).

## REFERÊNCIAS

- [1] Instituto Aço Brasil. Mercado brasileiro do aço - análise setorial e regional, 2021. URL [https://acobrasil.org.br/site/wp-content/uploads/2021/08/MBA\\_Ediç-ao-2021.pdf](https://acobrasil.org.br/site/wp-content/uploads/2021/08/MBA_Ediç-ao-2021.pdf).
- [2] Portal Siderurgia Brasil. Anuário brasileiro da siderurgia 2021, 2020. URL [https://siderurgiabrasil.com.br/wp-content/uploads/2020/05/gc2021\\_site.pdf](https://siderurgiabrasil.com.br/wp-content/uploads/2020/05/gc2021_site.pdf).
- [3] Anderson Badia da Silva. Utilização de carepas como componente da carga de um forno elétrico a arco. page 22, 2012.
- [4] J Haapakangas. Coke properties in simulated blast furnace conditions. *INVESTIGATION ON HOT STRENGTH, CHEMICAL REACTIVITY AND REACTION MECHANISM*, 2016.
- [5] ASL Z Majedi and A Salem. Investigation of the flame temperature for some gaseous fuels using artificial neural network. 2010.
- [6] Priyanga Chandrasekar, Kai Qian, Hossain Shahriar, and Prabir Bhattacharya. Improving the prediction accuracy of decision tree mining with data preprocessing. In *2017 IEEE 41st Annual Computer Software and Applications Conference (COMPSAC)*, volume 2, pages 481–484, 2017. doi: 10.1109/COMPSAC.2017.146.
- [7] José Díaz, F. Javier Fernández, and Inés Suárez-Ramón. Hot metal temperature prediction at basic-lined oxygen furnace (bof) converter using ir thermometry and forecasting techniques. *Energies*, 12:3235, 08 2019. doi: 10.3390/en12173235.
- [8] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
- [9] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [10] Leonard de Araujo Carvalho. Redes neurais artificiais para modelagem de altos-fornos. 2019.
- [11] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.
- [12] Tenghui Hu, Xianpeng Wang, Yao Wang, Zhiming Dong, and Xinyu Zhuang. Prediction of blast furnace temperature based on evolutionary optimization. In *Evolutionary Multi-Criterion Optimization: 11th International Conference, EMO 2021, Shenzhen, China, March 28–31, 2021, Proceedings 11*, pages 759–768. Springer, 2021.
- [13] Guang-Bin Huang, Dian Hui Wang, and Yuan Lan. Extreme learning machines: a survey. *International journal of machine learning and cybernetics*, 2(2):107–122, 2011.
- [14] Yoshinari Hashimoto, Yuki Okamoto, Tatsuya Kaise, Yoshitaka Sawa, and Manabu Kano. Practical operation guidance on thermal control of blast furnace. *ISIJ International*, 59(9):1573–1581, 2019. doi: 10.2355/isijinternational.ISIJINT-2019-119.
- [15] Xinmin Zhang, Manabu Kano, and Shinroku Matsuzaki. A comparative study of deep and shallow predictive techniques for hot metal temperature prediction in blast furnace ironmaking. *Computers & Chemical Engineering*, 130:106575, 2019.
- [16] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. The performance of lstm and bilstm in forecasting time series. In *2019 IEEE International Conference on Big Data (Big Data)*, pages 3285–3292. IEEE, 2019.
- [17] Alex Graves and Jürgen Schmidhuber. Framewise phoneme classification with bidirectional lstm and other neural network architectures. *Neural networks*, 18(5-6):602–610, 2005.
- [18] Scikit Learn. `sklearn.preprocessing.standardScaler`, 2021. URL <https://scikit-learn.org/stable/modules/generated/sklearn.preprocessing.StandardScaler.html#sklearn.preprocessing.StandardScaler>.
- [19] Greg Van Houdt, Carlos Mosquera, and Gonzalo Nápoles. A review on the long short-term memory model. *Artif. Intell. Rev.*, 53(8):5929–5955, 2020.
- [20] Brian Kenji Iwana and Seiichi Uchida. An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, 16(7):e0254841, 2021.
- [21] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [22] Pedro A Morettin and Clélia MC Toloi. *Análise de séries temporais: modelos lineares univariados*. Editora Blucher, 2018.
- [23] Ramon Gomes da Silva, Alef Berg de Oliveira, Igor Cruz da Silva, and Thulio de Oliveira. Application of a demand forecasting model in a rental company of billiard tables. 2016.
- [24] Rob J Hyndman and Anne B Koehler. Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688, 2006.
- [25] sktime. `MedianRelativeAbsoluteError`, 2022. URL [https://www.sktime.org/en/stable/api\\_reference/auto\\_generated/sktime.performance\\_metrics.forecasting.MedianRelativeAbsoluteError.html](https://www.sktime.org/en/stable/api_reference/auto_generated/sktime.performance_metrics.forecasting.MedianRelativeAbsoluteError.html).