



UNIVERSIDADE FEDERAL DO CEARÁ

CENTRO DE TECNOLOGIA

DEPARTAMENTO DE ENGENHARIA METALÚRGICA E DE MATERIAIS

JOÃO VICTOR BARROSO XAVIER

**AVALIAÇÃO DO USO DE UMA REDE NEURAL ARTIFICIAL E REGRESSÃO
LINEAR MÚLTIPLA PARA PREVISÃO DO TEOR DE SILÍCIO NO FERRO-GUSA
UTILIZANDO VARIÁVEIS OPERACIONAIS DO ALTO-FORNO.**

FORTALEZA

2023

JOÃO VICTOR BARROSO XAVIER

**AVALIAÇÃO DO USO DE UMA REDE NEURAL ARTIFICIAL E REGRESSÃO
LINEAR MÚLTIPLA PARA PREVISÃO DO TEOR DE SILÍCIO NO FERRO-GUSA
UTILIZANDO VARIÁVEIS OPERACIONAIS DO ALTO-FORNO**

Trabalho de conclusão de curso apresentado ao curso de engenharia metalúrgica do Centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia Metalúrgica.

Orientadora: Prof^ª. Dra. Janaína Gonçalves
Maria da Silva Machado

FORTALEZA

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas

Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

- X19a Xavier, João Victor Barroso.
Avaliação do uso de uma rede neural artificial e regressão linear múltipla para previsão do teor de silício no ferro-gusa utilizando variáveis operacionais do alto-forno / João Victor Barroso Xavier. – 2023.
59 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Metalúrgica, Fortaleza, 2023.
Orientação: Profa. Dra. Janaina Gonçalves Maria da Silva Machado.
1. Teor de silício. 2. Redes neurais artificiais. 3. Regressão Linear múltipla. 4. Ferro-gusa. I. Título.
CDD 669
-

JOÃO VICTOR BARROSO XAVIER

**AVALIAÇÃO DO USO DE UMA REDE NEURAL ARTIFICIAL E REGRESSÃO
LINEAR MÚLTIPLA PARA PREVISÃO DO TEOR DE SILÍCIO NO FERRO-GUSA
UTILIZANDO VARIÁVEIS OPERACIONAIS DO ALTO-FORNO**

Trabalho de conclusão de curso apresentado ao curso de engenharia metalúrgica do centro de Tecnologia da Universidade Federal do Ceará, como requisito parcial à obtenção do grau de bacharel em Engenharia Metalúrgica.

Aprovada em: ___/___/___

BANCA EXAMINADORA

Prof^a. Dra. Janaína Gonçalves Maria da Silva Machado (orientadora)
Universidade Federal do Ceará (UFC)

Prof. Dr. Elineudo Pinho de Moura
Universidade Federal do Ceará (UFC)

Eng. André Wulff Hirano
ArcelorMittal Pecém

A Deus.

Aos meus pais, José Nazareno e
Geovana Barroso.

AGRADECIMENTOS

Aos meus pais, José Nazareno e Geovana Barroso, por todo o apoio durante a minha trajetória. Agradeço do fundo do meu coração.

À minha família, em especial a Rosangela Cláudia e Rogério Barroso, pela grande presença em minha vida. Agradeço por todo apoio.

Aos meus colegas do Laboratório de Caracterização de Materiais, Breno Saraiva, Carlos Samuel, Guilherme Schmalz, Pablo Leão, Pedro Pinheiro, Rodrigo Loureiro e Suyanny Freire, pelos momentos de grande companheirismo. A jornada foi mais leve com vocês.

Ao meu amigo, João Pedro Teles, pela grandíssima amizade. Sempre vou lembrar dos momentos engraçados.

Aos colaboradores do Laboratório de Caracterização de Materiais, em especial a Flávio Herculano e Jorge Luiz, pelas grandes ensinamentos e constante apoio. Ganhei muito convivendo com vocês e sou grato por tudo.

Ao Prof. Hamilton Ferreira Gomes de Abreu, por ter me dado a oportunidade de atuar no Laboratório de Caracterização de Materiais. Obrigado pelas orientações durante a minha graduação.

Ao Prof. Elineudo Pinho de Moura, pelo apoio prestado durante a realização deste trabalho. A sua ajuda foi imprescindível.

À Prof^a. Janaína Gonçalves Maria da Silva Machado, por ter aceitado ser a minha orientadora durante este trabalho.

À Universidade Federal do Ceará, por trazer transformação ao povo cearense de forma pública, gratuita e de qualidade.

À minha avó, Maria Neuza, por tudo que realizou. Esteja em paz. (*In memoriam*).

“A morte é mais leve que a pluma. O dever, mais pesado que a montanha.” (Robert Jordan)

RESUMO

A crescente competição no setor siderúrgico levou os produtores de aço a buscarem diversas formas de melhorar a qualidade dos seus produtos. A entrega de um ferro-gusa com composição química dentro do esperado para a aciaria é de fundamental importância para esse objetivo. Nesse contexto, um dos principais elementos químicos que servem como indicador da qualidade do ferro-gusa é o silício. Dessa forma, muitos estudos foram desenvolvidos com o passar das décadas com foco na previsibilidade e controle do teor de silício no ferro-gusa. Recentemente, a aplicação de modelos computacionais para essa tarefa é crescente, sobretudo, o uso de redes neurais artificiais. No presente trabalho, avaliou-se a utilização de uma rede neural artificial do tipo *multilayer perceptron* (MLP) na construção de um modelo que utilizou variáveis operacionais como dados de entrada e o comparou com outro construído a partir de uma regressão linear múltipla. Para a obtenção dos resultados da regressão linear múltipla e da rede neural utilizaram-se os *softwares* Excel e Matlab, respectivamente. Para o primeiro, todos os dados foram utilizados como entrada e para o segundo, os dados foram divididos em conjuntos de treinamento e teste, sendo selecionados aleatoriamente 100 vezes para cada configuração de hiperparâmetros experimentada. A rede neural artificial obteve um coeficiente de correlação (R^2) médio de 0.2615, o qual é mais elevado do que o obtido pela regressão linear múltipla de 0.0971, demonstrando que ela possui uma melhor capacidade de previsão do teor de silício no ferro-gusa frente as variáveis apresentadas.

Palavras-chave: Teor de silício, Redes neurais artificiais, Regressão linear múltipla, Ferro-gusa

ABSTRACT

The increasing competition in the steel sector has led steel producers to look for different ways to improve the quality of their products. The delivery of hot metal with a chemical composition as expected for the steel shop is of fundamental importance for this objective. In this context, one of the main chemical elements used as an indicator of the quality of hot metal is silicon. Therefore, many studies have been developed over the decades focusing on the predictability and control of silicon content in hot metal. Recently, the application of computational models for this task is increasing, especially the use of artificial neural networks. In the present work, the use of a Multilayer Perceptron neural network (MLP) was evaluated in the construction of a model that used operational variables as input data and compared it with another built from a multiple linear regression. To obtain the results of the multiple linear regression and the neural network, Excel and Matlab software were used, respectively. For the first, all data was used as input and for the second, the data was divided into training and test sets, being randomly selected 100 times for each hyperparameter configuration experimented. The artificial neural network obtained an average correlation coefficient (R^2) of 0.2615, which is higher than that obtained by multiple linear regression of 0.0971, demonstrating that it has a better ability to predict the silicon content in hot metal in front of the variables presented.

Keywords: Silicon content, Artificial neural networks, Multiple linear regression, Hot metal.

LISTA DE FIGURAS

Figura 01 - Diferentes regiões de um alto-forno.....	15
Figura 02 - Visão geral dos equipamentos auxiliares do alto-forno.....	16
Figura 03 - Diagrama de Ellingham.....	18
Figura 04 - Representação esquemática de um neurônio artificial.....	29
Figura 05 - Arquitetura feedforward de camada simples.....	32
Figura 06 - Arquitetura feedforward de camadas múltiplas.....	32
Figura 07 - Rede de arquitetura recorrente.....	33
Figura 08 - Arquitetura em estrutura reticulada.....	34
Figura 09 – Representação de um adaline.....	36
Figura 10 - Processo de convergência do adaline.....	37
Figura 11 - Representação do algoritmo de backpropagation.....	38
Figura 12 - Processo de convergência com aplicação de termo de momentum.....	42
Figura 13 - Representação gráfica da técnica de pesquisa em grade.....	43
Figura 14 - Dados do ritmo de produção do alto-forno antes da retirada de <i>outliers</i>	47
Figura 15 - Dados do ritmo de produção do alto-forno depois da retirada de <i>outliers</i>	47
Figura 16 - Matriz de correlação entre as variáveis de entrada do modelo linear.....	51
Figura 17 - Previsão da saída desejada para o modelo linear.....	52
Figura 18 - Erro cometido pelo modelo linear.....	52
Figura 19 - Previsão da saída desejada para o experimento 392 da rede neural.....	53
Figura 20 - Erro cometido pela rede neural no experimento 392.....	54

LISTA DE TABELAS

Tabela 01 - Resumo dos dados.....	44
Tabela 02 - Principais métricas de desempenho da regressão linear múltipla.....	48
Tabela 03 - Resultados obtidos a partir da regressão linear múltipla.....	49
Tabela 04 - Configurações dos experimentos com os 3 maiores R^2 médios obtidos.....	53
Tabela 05 - Configurações dos experimentos com os 3 maiores R^2 máximos obtidos.....	53

SUMÁRIO

1	INTRODUÇÃO.....	13
2	REVISÃO BIBLIOGRÁFICA.....	14
2.1	O alto-forno.....	14
2.1.1	O processo de redução de minérios no interior do alto-forno.....	16
2.1.2	Mecanismos de incorporação do silício no ferro-gusa.....	19
2.2	Estudos anteriores sobre a previsão do teor de silício no ferro-gusa.....	20
2.3	Regressão linear múltipla e parâmetros para avaliação de modelos lineares.....	21
2.3.1	Regressão linear múltipla.....	21
2.3.2	Estatísticas para a regressão linear múltipla.....	24
<i>2.3.2.1</i>	<i>Graus de liberdade.....</i>	<i>24</i>
<i>2.3.2.2</i>	<i>Funções de soma dos quadrados.....</i>	<i>25</i>
<i>2.3.2.3</i>	<i>Erro padrão.....</i>	<i>25</i>
<i>2.3.2.4</i>	<i>Teste de significância global do modelo de regressão linear múltipla.....</i>	<i>26</i>
<i>2.3.2.5</i>	<i>Teste de significância para os coeficientes da regressão linear múltipla.....</i>	<i>27</i>
<i>2.3.2.6</i>	<i>Coefficiente de correlação.....</i>	<i>27</i>
2.4	Redes neurais artificiais.....	28
2.4.1	O neurônio artificial.....	28
2.4.2	Arquiteturas de redes neurais.....	31
<i>2.4.2.1</i>	<i>Arquitetura feedforward</i>	<i>31</i>
<i>2.4.2.2</i>	<i>Outros tipos de arquiteturas.....</i>	<i>33</i>
2.4.3	Formas de aprendizagem.....	34
2.4.4	Algoritmo de treinamento.....	35

2.4.4.1	<i>Perceptron e Adaline</i>	35
2.4.4.2	<i>Rede Multilayer perceptron (MLP)</i>	37
2.4.2.3	<i>Aprimoramento do algoritmo de backpropagation</i>	41
2.4.5	Hiperparâmetros das redes neurais	42
3	METODOLOGIA	44
3.1	Aquisição e tratamento de dados	44
3.2	Regressão linear múltipla	45
3.3	Rede neural artificial	46
4	RESULTADOS E DISCUSSÃO	47
5	CONCLUSÃO	56
	REFERÊNCIAS	58

1 INTRODUÇÃO

O aço é um produto siderúrgico presente em diversos setores da sociedade, sendo imprescindível para todas as economias do mundo e o desenvolvimento tecnológico das nações. Nesse contexto, as usinas siderúrgicas, onde o aço é produzido, promovem esforços constantes para melhorar a qualidade de seus produtos. O alto-forno é uma das etapas mais importantes dentro do processo siderúrgico, pois nele é produzida a principal matéria-prima para a fabricação do aço, o ferro-gusa.

A qualidade do ferro-gusa é fundamental para o sucesso do processo siderúrgico. A composição química é extremamente importante para as etapas seguintes do processo industrial. Na aciaria, aonde o ferro-gusa será efetivamente transformado em aço, o teor de silício presente no primeiro trata-se de uma fonte de calor importante para a manutenção de altas temperaturas, segundo Hwang *et al.* (2010).

Em seus estudos, Pérez *et al.* (2010). afirma que um excesso de silício pode provocar a necessidade de realizar um processo conhecido como “dupla escória”, no qual é necessário injetar oxigênio no metal para produzir uma primeira escória, inapropriada para a continuidade do ciclo siderúrgico, pois ela é muito rica em sílica e altamente viscosa. Em seguida, esta é vazada do convertedor L-D e dá-se continuidade à produção convencional do aço. Caso a primeira escória não seja vazada, pode haver o aprisionamento de gases e consequente borbulhamento da escória, o que pode levar a acidentes e emissão de gases tóxicos. Com isso, entende-se a importância do controle do teor de silício no ferro-gusa desde o alto-forno.

A quantidade de silício presente no ferro-gusa é oriunda da matéria-prima carregada no alto-forno, logo, há diversos campos de pesquisa que atuam com foco nela, entretanto, uma operação eficiente deste reator metalúrgico também é imprescindível para a entrega de um produto de qualidade para a aciaria. A partir dessa premissa, diversos modelos teóricos e computacionais foram desenvolvidos utilizando diferentes variáveis com o intuito de controlar o teor de silício no ferro-gusa. Atualmente, o uso de redes neurais artificiais é uma das técnicas computacionais mais utilizadas para essa tarefa. Nesse contexto, o presente trabalho buscou avaliar o uso de uma rede neural artificial de arquitetura *multilayer perceptron* (MLP) para prever o teor de silício no ferro-gusa utilizando variáveis operacionais, comparando-a com um modelo linear gerado por regressão linear múltipla.

2 REVISÃO BIBLIOGRÁFICA

2.1 O alto-forno

O alto-forno é um dos principais reatores do processo siderúrgico. A sua função é transformar as matérias primas carregadas em seu interior em ferro-gusa por meio da redução de minérios. O ferro-gusa é uma liga composta majoritariamente por ferro e uma quantidade significativa de carbono (por volta de 4 a 4.5%), contendo outros elementos residuais como silício, manganês, fósforo e enxofre (Vieira, 2012). Para estimar o teor de silício no ferro-gusa, primeiramente é necessário entender quais fatores influenciam nesse processo de incorporação e para isso, é preciso conhecer as diferentes divisões do alto-forno e o funcionamento do processo de redução.

Segundo Peacey (1979), o alto-forno é dividido fisicamente em goela, cuba, ventre, rampa e cadinho. A goela é a região superior do forno por onde a carga é despejada no seu interior. A cuba se localiza imediatamente após a goela e possui o formato de tronco de cone para que a carga caia livremente pelo forno enquanto a temperatura dela sobe durante a queda. O ventre, ou rampa paralela, é caracterizado por paredes paralelas entre si e corresponde à região de maior diâmetro do alto-forno, bem como a região onde a carga começa a amolecer e posteriormente, fundir. Na rampa, as paredes do forno assumem a geometria de um tronco de cone invertido, o que ajuda na sustentação da carga e no gotejamento de gusa e escória para a última região do forno. O cadinho, é o local onde o coque não queimado na zona das ventaneiras é depositado, além disso, é nesta região em que o ferro-gusa e a escória líquidos são vazados para fora do alto-forno.

Conforme Mourão *et al.* (2007), o alto-forno pode ser dividido internamente em 5 zonas distintas, baseadas em estudos realizados no Japão na década de 70:

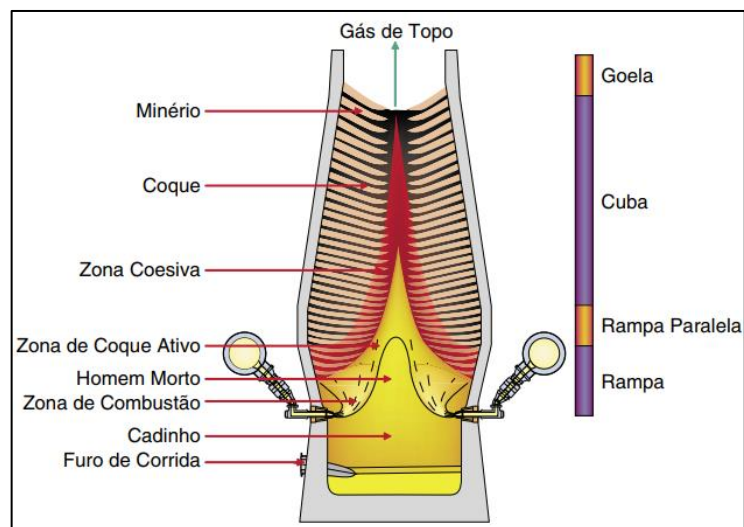
- a) Zona granular - possui camadas alternadas de coque e minério, tais quais como foram carregados. Nesta região, a matéria está presente nas fases sólida e gasosa.
- b) Zona coesiva - é constituída de minério amolecido e fundido e camadas de coque sólido, por onde o fluxo gasoso passa preferencialmente.
- c) Zona de gotejamento – encontra-se logo abaixo da zona coesiva, sendo composta por coque no estado sólido pelo qual o ferro-gusa líquido atravessa os interstícios,

resultando na incorporação de diversos elementos de liga ao ferro-gusa. Além disso, a zona de gotejamento é subdividida em região de coque ativo e o “homem morto”.

- d) Zona de combustão – é localizada em frente às ventaneiras, onde a temperatura no interior do forno atinge os níveis mais elevados devido à queima do coque para a geração do gás reductor.
- e) Cadinho – trata-se da região na qual o coque, em forma granulada, se deposita com o ferro-gusa e a escória líquidos, ocupando o espaço intersticial e separando-se por diferença de densidade.

A Figura 01 ilustra as diferentes regiões físicas e zonas internas de um alto-forno moderno:

Figura 01 - Diferentes regiões de um alto-forno.

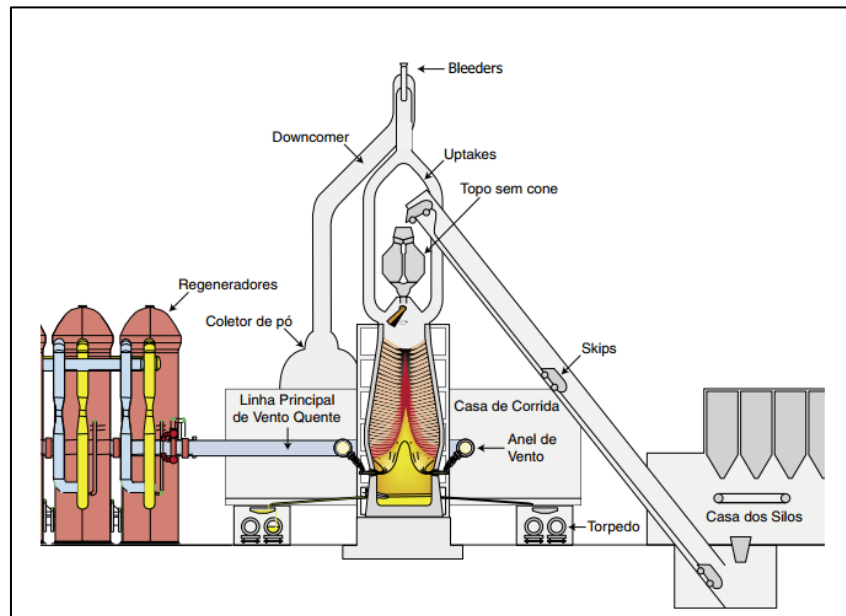


Fonte: retirado de Geerdes, Vliet e Toxopeus (2004)

Além do corpo principal do alto-forno, este necessita de outros equipamentos para operar de forma adequada. Os regeneradores são necessários para pré-aquecer o ar injetado no interior do alto-forno pelas ventaneiras. As matérias primas ficam armazenadas na casa de silos, onde é possível programar a quantidade adequada de cada componente da carga e direcionar ela para o alto-forno por meio de um sistema de esteiras transportadoras ou carros guiados (*skips*). No topo do alto-forno, há um sistema de descarregamento para garantir a manutenção da pressão interna e garantir uma distribuição homogênea da carga durante a queda desta. Ainda no topo do alto-forno, também há um sistema de coleta dos gases ascendentes que são direcionados via tubulação para um coletor de pó e para um sistema de lavagem de gases.

À medida que o cadinho acumula ferro-gusa e escória, ele é vazado na sala de corrida para um canal que separa a escória do ferro-gusa devido a diferença de densidade de ambos e direciona ele para o carro torpedo, a fim de transportar o ferro-gusa para a aciaria. A escória é direcionada para a granulação, onde é resfriada com água e armazenada para que possa ser aproveitada pela indústria de construção civil. A Figura 02 demonstra de forma geral os equipamentos que atuam em conjunto com o alto-forno.

Figura 02 -Visão geral dos equipamentos auxiliares do alto-forno.



Fonte: retirado de Geerdes, Vliet e Toxopeus (2004)

2.1.1 O processo de redução do minério de ferro no interior do alto-forno

Segundo Mourão *et al.* (2007), o alto-forno é carregado através da goela com uma mistura de matérias primas como a carga metálica, a qual pode ser constituída de sinter, pelotas, minério granulado e o combustível sólido (coque ou carvão vegetal), além de fundentes e injeções auxiliares à exemplo de gás natural, carvão pulverizado, etc.

Ghosh e Chatterjee (2007) afirmam que a formação do gás redutor é fundamental para o funcionamento do alto-forno. Para isso, ar pré-aquecido nos regeneradores entre 1000 e 1100°C é injetado através das ventaneiras, sendo transportado por meio de tubulações. A reação exotérmica do carvão do coque e do carvão com o oxigênio gera gás CO₂ e produz calor.

Baseando-se em McLean (1999). A reação exotérmica do coque e do carvão na zona de combustão ocorre em duas etapas. Primeiro, ocorre a reação exotérmica:



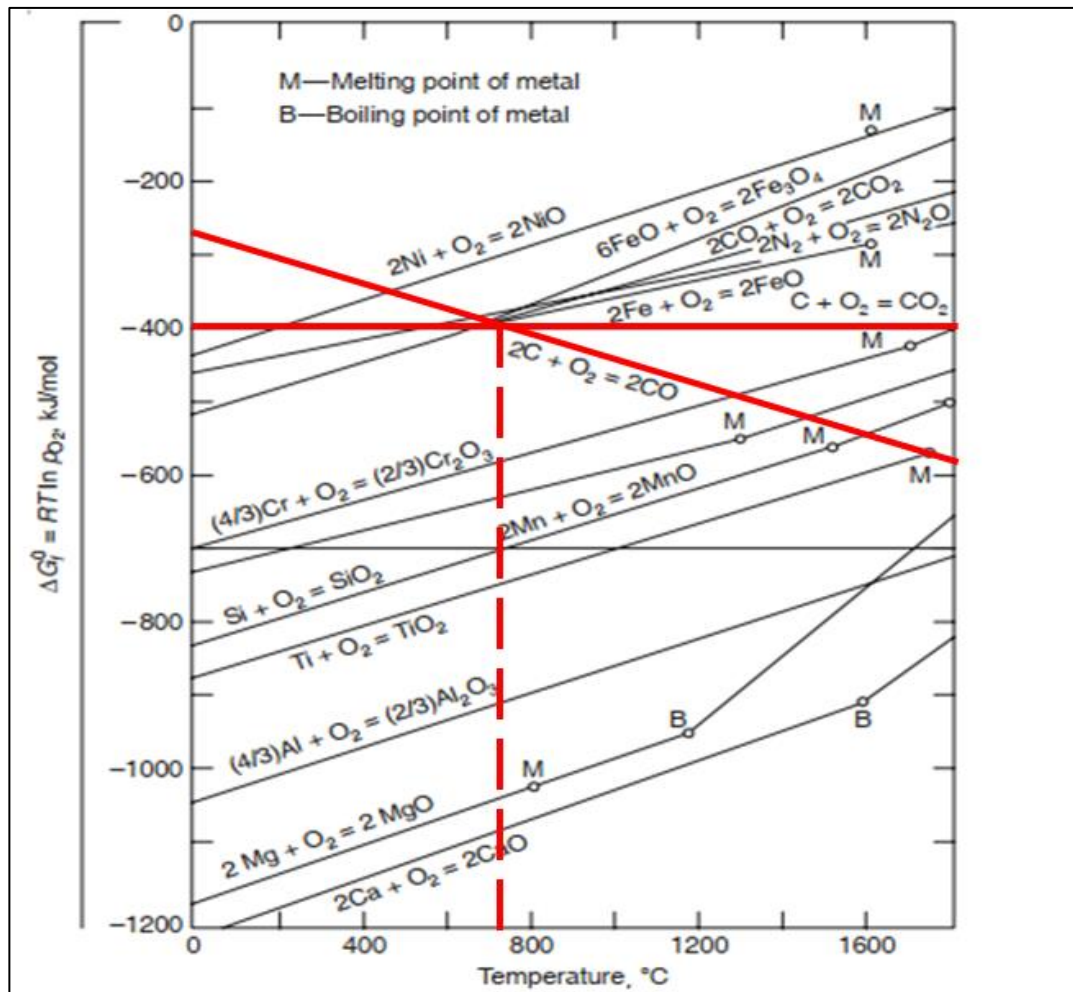
Esse processo ocorre em uma região muito próxima das ventaneiras, na qual há um excesso de oxigênio devido ao sopro e as temperaturas chegam até 2000°C. Em seguida, o dióxido de carbono reage com o coque por meio da reação:



Essa reação é conhecida como a reação de Boudouard e possui caráter endotérmico. Ela ocorre em temperaturas acima de 1000°C e é responsável por aumentar o volume de gás redutor no interior do alto-forno.

O controle da temperatura é fundamental para o processo por conta do seu efeito sobre a reação de Boudouard. Ao se observar a Figura 03, é possível ver que as linhas de formação dos óxidos de carbono se cruzam aproximadamente na temperatura de 700°C. No diagrama de Ellingham, a reação com maior estabilidade é aquela que possui uma energia livre de Gibbs mais negativa, dessa forma, abaixo de 700°C a reação de formação de CO₂ é mais estável e acima de 700°C, o monóxido de carbono passa a ser mais estável. A eficiência do processo do alto-forno é continuamente monitorada medindo-se a composição química do gás do topo. Ela é expressa como a utilização de gás: o percentual de gás CO que foi transformado em CO₂ (Geerdes; Vliet; Toxopeus, 2004).

Figura 03 - Diagrama de Ellingham

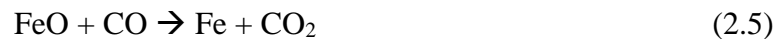
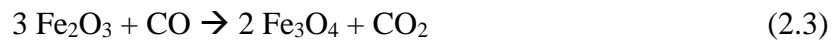


Fonte: Adaptado de Ghosh e Chatterjee (2007)

Depois de gerado, o CO ascende pelo alto-forno, dando continuidade ao processo de produção do ferro-gusa. Biswas (1981) relata que o processo de redução acontece em 3 zonas.

Na primeira zona, que recebe o nome de zona de fusão e abrange a região do alto-forno localizada de 3 a 5 metros acima do nível das ventaneiras, a matéria fundida atinge uma temperatura um pouco acima de 1400°C, enquanto os gases resfriam para uma faixa entre 800 e 1000°C. Nessa zona há uma região de coque ativo na periferia do alto-forno devido à queima contínua de coque na zona de combustão, a qual permite a descida constante da carga e no centro há uma coluna central compacta de coque inativo chamada “homem-morto”. Essa coluna pode flutuar no ferro líquido que há no cadinho ou tocar na base do alto-forno. A ascensão dos gases ocorre devido aos espaços intersticiais no coque, que atuam como um distribuidor de gás no alto-forno.

Na segunda zona, chamada de zona de reserva térmica, as temperaturas dos sólidos e gases são quase idênticas e na faixa de 800 a 1000°C, é nesta zona onde ocorre a maior parte das reações de redução indireta abaixo, sendo interessante que ela se estenda de forma que seja fornecido tempo suficiente para a redução da wustita (comumente descrita como FeO) ocorrer.



Na terceira zona, chamada de zona de preparação ou pré-aquecimento que compreende a região superior do forno, onde a temperatura do gás está na faixa de 100 a 250°C e a temperatura dos sólidos cai de 800°C até a temperatura ambiente, a carga perde a sua umidade e sofre as primeiras reações de redução causadas pelos gases ascendentes que vem da zona de zona de reserva térmica.

2.1.2 Mecanismos de incorporação do silício no ferro-gusa

Segundo Mourão *et al.* (2008), as principais fontes de silício são a cinza do coque e a ganga do minério de ferro, os quais contém SiO₂. A incorporação do silício no ferro-gusa pode acontecer por duas rotas. A primeira ocorre indiretamente por meio da redução da sílica presente na escória pelo monóxido de carbono na zona de combustão conforme a reação abaixo:



Uma vez gerado, o gás SiO reage com o carbono do ferro-gusa, produzindo silício através da reação:



Na segunda rota, o silício é incorporado diretamente da escória para o ferro-gusa por meio da reação:



Segundo alguns autores, essa reação é muito lenta quando comparada com a rota indireta. Geerdes, Vliet e Toxopeus (2004) salientam ainda que, uma vez que o ferro é gotejado e atravessa a camada de escória, há uma possibilidade de re-oxidação do silício, o qual retorna para a escória por meio da reação:



Com isso, o teor de silício no ferro-gusa pode ser influenciado por uma série de fatores que variam desde a composição da matéria-prima até a modificação em parâmetros operacionais do alto-forno, desde que estes interfiram na dinâmica dos mecanismos de incorporação de silício no ferro-gusa.

É bastante reportado na literatura que o silício é um bom indicador do estado térmico do alto-forno. Além disso, Araújo (1998 *apud* Aguiar, 2016) salienta que teores muito elevados de silício no ferro-gusa (acima de 0,6%) são prejudiciais para o convertedor na aciaria, pois provoca uma maior adição de fundentes, aumentando a incidência de projeções, riscos operacionais e reduzindo o rendimento metálico no processo. No caso inverso, quando há teores de silício muito baixos (menores do que 0,15%), há uma redução do balanço térmico no convertedor, reduzindo a adição de sucata. Sendo assim, o controle do teor de silício no ferro-gusa em faixas de valores mais estreitas acarreta vantagens no processo siderúrgico.

2.2 Estudos anteriores sobre a previsão do teor de silício no ferro-gusa.

Uma vez entendidas as estruturas básicas do alto-forno, o processo de redução dos minérios no seu interior e as principais rotas de incorporação do teor de silício no ferro-gusa, é necessário buscar na literatura quais os principais parâmetros utilizados em estudos anteriores, uma vez que esta é uma tarefa difícil e como será visto, diferentes autores utilizaram parâmetros diferentes à medida que o entendimento sobre os fenômenos internos do alto-forno foi evoluindo, bem como as técnicas de monitoramento deles.

Bulsari e Sáxen (1991) buscaram demonstrar a viabilidade de uma rede neural com uma arquitetura *feedforward multilayer perceptron* (MLP) para tentar prever o teor de silício a partir dos dados disponíveis em sua época, como o teor de silício de amostragens anteriores e alguns parâmetros operacionais. Esse estudo concluiu que a utilização de redes neurais artificiais produz resultados muito melhores do que os métodos estatísticos convencionais.

Medeiros *et al.* (2006) realizaram um estudo onde descreveram a utilização de um modelo baseado em redes neurais e no conhecimento na prática da fabricação de ferro-gusa. Aplicou-se uma rede neural artificial com arquitetura *feedforward* com 23 neurônios na camada de entrada, 23 neurônios na camada escondida e 5 neurônios na cada de saída para prever (1) o teor de silício no ferro-gusa, (2) o coeficiente de partição de enxofre entre a escória e o ferro-gusa, (3) o parâmetro de resistência ao fluxo de ar, (4) a perda térmica para o balanço de massa e (5) o parâmetro de incompatibilidade no equilíbrio para o balanço de massa do oxigênio utilizando 23 variáveis como dados de entrada relacionadas com *slag rate*, a basicidade B4 da escória, a temperatura do sopro, *small coke rate*, *injected coal rate*, a taxa de injeção de hematita, a composição química da cinza do carvão pulverizado injetado e a quantidade de oxigênio presente nele, propriedades mecânicas e a reatividade do coque. O trabalho salientou a dificuldade de se prever o teor de silício no ferro-gusa devido à natureza complexa do problema.

Saxén e Pettersson (2007) utilizaram redes neurais artificiais para a previsão do teor de silício no ferro-gusa, mas também lançaram mão de um algoritmo de poda para encontrar as variáveis de entrada mais relevantes para estudo e os seus atrasos no tempo, conseguindo detectar relações relevantes entre a perda de calor das ventaneiras, o volume total de sopro e a permeabilidade do gás com relação ao teor de silício do ferro-gusa.

Diniz (2018) sintetizou diversos estudos anteriores para buscar entender quais seriam as melhores variáveis para a construção de um modelo que pudesse prever o teor de silício no ferro-gusa. Em seguida, também foi realizado um estudo de estacionariedade e linearidade da série temporal de silício para uma melhor aplicação de um algoritmo de poda e de uma rede neural de arquitetura recorrente com atrasos no tempo.

Fontes, Vasconcelos e Brito (2020) propuseram um algoritmo baseado em técnicas de clusterização e redes neurais artificiais do tipo NARX, encontrando uma relação entre a taxa de injeção de carvão PCI, *Coke Rate*, volume, temperatura e umidade de sopro, taxa de enriquecimento de oxigênio, temperatura teórica de chama, percentual de redução direta, perda de calor e permeabilidade e o teor de silício no alto-forno.

2.3 Regressão linear múltipla e parâmetros para avaliação de modelos lineares

No presente trabalho, buscou-se comparar o desempenho entre um modelo linear gerado a partir de uma regressão linear múltipla e um modelo não linear utilizando uma rede neural artificial para demonstrar a eficiência deste último método. Para isso, é necessário entender o funcionamento básico de ambos. A presente seção dedica-se a descrever a fundamentação teórica da regressão linear múltipla e os parâmetros estatísticos que avaliam a sua qualidade, os quais podem ser obtidos via *software*.

2.3.1 Regressão linear múltipla

O desenvolvimento a seguir tem como base o texto de Montgomery e Runger (2009). Para ajustar um modelo com múltiplas variáveis à uma reta, utiliza-se a regressão linear múltipla onde variável de saída ou dependente está ligada às variáveis de entrada ou independentes por meio da expressão:

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon \quad (2.10)$$

Onde Y representa a variável de saída estimada por um modelo com k variáveis, ε representa o erro ou resíduo entre os valores estimados e reais, e que os parâmetros β_j , $j=0,1,2, \dots, k$, são chamados de coeficientes da regressão, nos quais também há o interesse de se estimar. Para essa tarefa, utiliza-se método dos mínimos quadrados (*least squares*).

Considerando $n > k$ observações ou conjecturas, tem-se que a equação 2.10 pode ser reescrita de forma:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_k x_{ik} + \varepsilon_i, \text{ para } i = 1, 2, \dots, n. \quad (2.11)$$

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i, \text{ para } i = 1, 2, \dots, n. \quad (2.12)$$

Com isso, a função soma dos quadrados dos erros (*Sum Squared Errors*) é dada por:

$$SSE = \sum_{i=1}^n \varepsilon^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j \cdot x_{ij} \right)^2 \quad (2.13)$$

$$\frac{\partial SSE}{\partial \beta} = 0 \quad (2.20)$$

Resolvendo-se a equação 2.20 de forma similar ao procedimento escalar, obtém -se as equações normais de mínimos quadrados para a abordagem matricial que possuem a forma:

$$X'X\beta = X'Y \quad (2.21)$$

Para encontrar os estimadores dos coeficientes da regressão linear, basta isolar o β , obtendo -se:

$$\beta = (X'X)^{-1}X'Y \quad (2.22)$$

A equação 2.22 pode ser desenvolvida da seguinte forma:

$$\begin{bmatrix} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{bmatrix} \quad (2.23)$$

Ao desenvolver as operações matriciais na equação 2.23, percebe-se as similaridades com as equações 2.16 e 2.17 do procedimento escalar. A matriz $(X'X)$ possui características especiais, pois os elementos de sua diagonal representam a soma dos quadrados dos elementos de X coluna a coluna, os elementos fora dela são a soma do produto vetorial nas colunas de X. Com isso, o modelo ajustado da regressão possui a forma escalar:

$$Y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij}, \text{ para } i = 1, 2, \dots, n. \quad (2.24)$$

A qual, convertendo-se na sua forma matricial, resulta em:

$$Y = \beta X \quad (2.25)$$

2.3.2 Estatísticas para a regressão linear múltipla

Atualmente, é possível realizar a regressão linear múltipla por meio de *softwares*. Apesar de ser uma forma rápida de se obter os coeficientes de regressão mencionados na sessão anterior, é importante possuir uma boa fundamentação teórica em conceitos estatísticos para conseguir avaliar o quão bem o modelo gerado conseguiu prever o comportamento dos

dados. As seções a seguir buscam revisar brevemente alguns conceitos que podem ser úteis para essa tarefa.

2.3.2.1 Graus de liberdade

Com base em Darlington e Hayes (2016), um modelo contendo k regressores possui $k + 1$ graus de liberdade, um para cada regressor além do coeficiente da regressão β_0 , no entanto, raramente é interessante testar hipóteses sobre este último. Com isso, diz-se que um modelo com k regressores possui k graus de liberdade de regressão, do inglês *regression degrees of freedom*, o que será abreviado neste trabalho para $GL_{\text{regressão}}$.

Quando tratamos de $n = k + 1$ observações, em quase todos os casos, o erro ou resíduo será zero (Darlington; Hayes, 2016). Assim, quando trabalhamos com um modelo com $n > k + 1$ observações, diz-se que haverá $n - (k+1)$ ou $n - k - 1$ graus de liberdade residual, do inglês *residual degrees of freedom*, o que será abreviado neste trabalho para $GL_{\text{resíduos}}$. Ele também pode ser visto como a quantidade de graus de liberdade que é possível adicionar ao modelo antes de exaurir a sua capacidade de informar o quão bem o modelo de uma amostra se ajusta ao modelo da população (Darlington; Hayes, 2016).

A quantidade total de graus de liberdade do modelo, do inglês *total degrees of freedom*, será a soma entre a quantidade de graus de liberdade da regressão e dos resíduos, ou seja, $n - 1$. No presente trabalho, esse termo será referido abreviadamente como GL_{total} .

2.3.2.2 Funções de soma dos quadrados

Na subseção 2.3.1, trabalhou-se com a função soma dos quadrados dos erros (SSE) para encontrar as equações normais dos mínimos quadrados. A SSE se relaciona com a soma total dos quadrados (SST) por meio da seguinte relação:

$$SST = SSR + SSE \quad (2.26)$$

$$\sum_{i=1}^n (y_i - Y)^2 = \sum_{i=1}^n (y_i^* - Y)^2 + \sum_{i=1}^n (y_i - y_i^*)^2 \quad (2.27)$$

Onde y_i e y_i^* representam os valores desejados e estimados, respectivamente, durante a i -ésima observação e Y representa a média de y_i ou y_i^* . A sigla SSR refere-se à soma dos quadrados da regressão. Spiegel (1993) afirma que o segundo termo à direita da igualdade é denominado como a variação não explicada, enquanto o primeiro pode ser referenciado como a variação explicada.

2.3.2.3 Erro padrão

Conforme Montgomery e Runger (2009), os erros padrões dos parâmetros da regressão podem ser ditos como uma medida do quão precisa foi a estimação do modelo para cada parâmetro. É possível obter o erro padrão de cada coeficiente a partir da matriz de variância e covariância, a qual pode ser calculada por:

$$\sigma^2(X'X)^{-1} \quad (2.28)$$

Em que σ^2 trata-se da variância dada pela razão entre SSE e os graus de liberdade dos resíduos, $n-k-1$. A equação 2.28 resulta em uma matriz na qual os elementos da sua diagonal principal são as variâncias dos parâmetros de regressão. A partir delas, para calcular o erro padrão, basta proceder conforme a equação:

$$EP_{\beta_i} = \sqrt{\sigma^2 C_{ii}} \quad (2.29)$$

Onde C_{ii} representa o elemento da diagonal principal da matriz $(X'X)^{-1}$ correspondente ao i -ésimo coeficiente da regressão. Os outros elementos fora da diagonal principal da matriz gerada pela equação 2.28 são ditas as covariâncias entre as variáveis do modelo

2.3.2.4 Teste de significância global do modelo de regressão linear múltipla.

A principal forma de verificar o nível de significância do modelo linear é através do teste F. Este teste possui uma hipótese nula $H_0: \beta_0 = \beta_1 = \dots = \beta_k = 0$ e uma hipótese alternativa $H_a: \beta_0 \neq \beta_k \neq 0$ onde k representa o número de preditores. Para a realização do teste, é necessário utilizar a seguinte relação:

$$F_{calculado} = \frac{QM_{regressão}}{QM_{residual}} = \frac{\frac{SSR}{GL_{regressão}}}{\frac{SSE}{GL_{resíduos}}} \quad (2.30)$$

Com base em Montgomery e Runger (2009) e Darlington e Hayes (2016), $F_{calculado}$ é comparado com os valores de presentes na tabela de valores da distribuição F de Fisher-Snedecor, a qual está amplamente presente em livros de estatística. Quando $F_{calculado}$ é menor do que $F_{tabelado}$, aceita-se a hipótese nula, o que quer dizer que o modelo proposto não possui significância. Do contrário, é possível rejeitar a hipótese nula e afirmar que pelo menos um dos coeficientes da regressão é diferente de zero e β_0 .

Outra forma de rejeitar a hipótese nula pode ser encontrada ao calcular-se o P-valor por meio das tabelas de distribuição F, utilizando o F_{calc} . O P-valor refere-se uma medida da força dos resultados frente a uma decisão de rejeitar-se ou não a hipótese nula (Arsham, 1988). Com isso, quando o P-valor é menor que o nível de significância (α), é possível rejeitar a hipótese nula.

No caso da regressão linear múltipla, o teste F não é suficiente para avaliar a qualidade geral do modelo, pois há k regressores, logo, é preciso realizar um teste de hipótese específico para cada um deles.

2.3.2.5 Teste de significância para os coeficientes da regressão linear múltipla

Para verificar se os regressores são estatisticamente significantes para o modelo, é necessário realizar um teste t para cada um deles. Esse teste considera que a hipótese nula $H_0: \beta_i = 0$ e a hipótese alternativa $H_a: \beta_i \neq 0$ para $i = 1, 2, \dots, k$. Para a realização do teste, é necessária a seguinte expressão:

$$t_{calculado} = \frac{\beta_i - \beta_i^*}{EP_{\beta_i}} \quad (2.31)$$

Onde, β_i^* e EP_{β_i} representam o valor que se supõe que o coeficiente β_i possua e o erro padrão do estimador β_i , respectivamente. De forma análoga ao teste F, é preciso comparar o $t_{calculado}$ com o $t_{tabelado}$ obtido a partir de tabelas de valores para a distribuição t de Student. O P-valor também pode ser adquirido por meio delas com um procedimento semelhante ao utilizado para o teste F.

2.3.2.6 Coeficiente de correlação

O coeficiente de correlação r pode ser obtido através da matriz de variância e covariância descrita pela equação 2.28 e ele nos diz o quão forte é a correlação entre duas variáveis, podendo variar entre -1 e +1, onde valores próximos desses extremos indicam correlações muito fortes. Segundo Darlington e Hayes (2016), o coeficiente de correlação r pode ser definido como a razão entre a covariância entre duas variáveis e as suas respectivas variâncias.

Outra métrica importante para a regressão linear múltipla é a medida de o quão bem o modelo explica os dados, para isso, é possível utilizar o coeficiente de correlação de Pearson ao quadrado (R^2). Segundo Orlov (1996), este pode ser calculado por meio da seguinte expressão:

$$R^2 = \frac{SSR}{SST} \quad (2.32)$$

Infelizmente, R^2 possui uma desvantagem, pois quanto mais regressores são adicionados ao modelo, maior será o seu valor (Martins; Domingues, 2014). Uma forma encontrada para corrigir esse comportamento é utilizando a equação abaixo:

$$R^2_{ajustado} = 1 - (1 - R^2) \frac{GL_{total}}{GL_{residuos}} \quad (2.33)$$

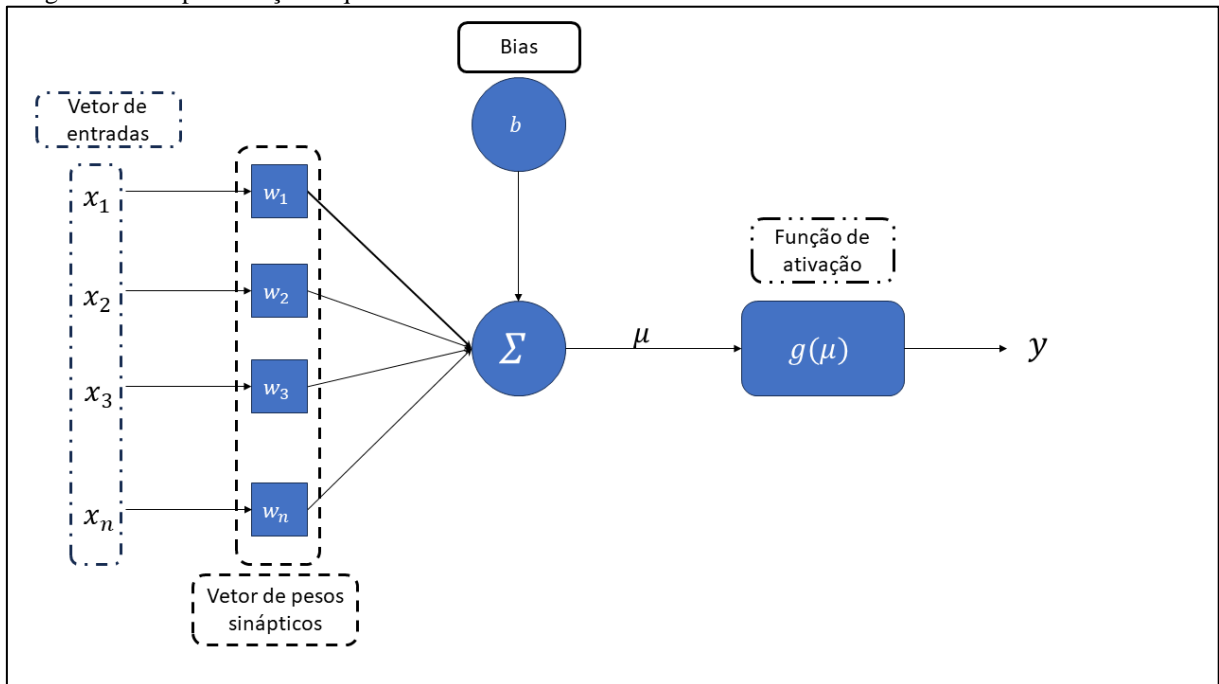
2.4 Redes neurais artificiais

Redes neurais artificiais são modelos computacionais inspirados no sistema nervoso dos seres vivos que possuem a capacidade de aquisição e manutenção do conhecimento, podendo ser definidas como um conjunto de unidades de processamento que são interligadas por muitas interconexões (sinapses) (Silva; Spatti; Flauzino, 2016). Esta seção dedica-se a dar uma breve explicação sobre como as redes neurais funcionam e os tipos existentes.

2.4.1 O neurônio artificial

Na literatura, a mais clássica abordagem no ensino sobre algoritmos de aprendizagem começa no entendimento da unidade básica de processamento das redes neurais. A Figura 04 ilustra um neurônio artificial.

Figura 04 – Representação esquemática de um neurônio artificial.



Fonte: O próprio autor.

Com base em Silva, Spatti e Flauzino (2016), o neurônio artificial é composto pelos seguintes elementos:

- Vetor de entradas: trata-se da informação adquirida pelo meio externo. Geralmente, eles são normalizados para incrementar a eficiência computacional dos algoritmos de aprendizagem.
- Vetor de pesos sinápticos: são os parâmetros que permitirão quantificar a relevância de cada um dos elementos do vetor de entrada quanto à funcionalidade do respectivo neurônio.
- Combinador linear: é responsável por realizar a soma ponderada dos sinais de entrada por seus respectivos pesos sinápticos.
- Bias: também conhecido como viés, este elemento especifica qual é o patamar apropriado para que o resultado do combinador linear possa gerar um valor de

disparo em direção à saída do neurônio. Haykin (2001) afirma que o bias possui a função de aumentar ou diminuir o valor da saída do combinador linear.

- e) Função de ativação: é responsável por limitar a saída do combinador linear dentro de uma amplitude de valores razoável a serem assumidos pelo próprio conjunto imagem da função escolhida.
- f) Sinal de saída: trata-se do valor final produzido pelo neurônio, dado o vetor de entrada. Ele pode ser usado como elemento do vetor de entrada de outros neurônios interligados com estes

Sintetizando as ideias de Haykin (2001) e Silva, Spatti e Flauzino (2016), o neurônio artificial pode ser definido matematicamente em duas equações:

$$\mu_k = \sum_{j=1}^m w_{kj} x_j \quad (2.34)$$

$$y_k = g(\mu_k + b_k) \quad (2.35)$$

Na equação 2.34, μ_k representa a saída do combinador linear, w_{kj} representa o peso sináptico do k -ésimo neurônio que irá ponderar o k -ésimo elemento do vetor de entrada. Na equação 2.35, y_k e b_k representam a saída e o bias do k -ésimo neurônio respectivamente. Alguns autores consideram o bias como um elemento do vetor de entradas, denominado x_0 , o qual será pelo peso sináptico w_{k0} de valor unitário.

As funções de ativação podem ser divididas em parcialmente e totalmente diferenciáveis, segundo Silva, Spatti e Flauzino (2016). As últimas são a classe de funções de ativação mais comuns para problemas complexos e devem ser mais detalhadamente discutidas.

A função totalmente diferenciável mais simples é a função linear, sendo responsável apenas por transmitir os seus valores de entrada para a sua imagem, logo, ela pode ser definida matematicamente por:

$$g(\alpha\mu) = \alpha\mu \quad (2.36)$$

$$g'(\alpha\mu) = \alpha \quad (2.37)$$

Onde α e μ representam a inclinação da reta e um valor qualquer no intervalo $[-\infty, +\infty]$, respectivamente. A função sigmoide é um dos tipos de função de ativação totalmente diferenciável mais comum, sendo expressa por:

$$g(\alpha, \mu) = \frac{1}{1 + e^{-\alpha\mu}} \quad (2.38)$$

$$g'(\alpha, \mu) = \alpha g(\mu)(1 - g(\mu)) \quad (2.39)$$

Onde α é o parâmetro de inclinação da função sigmoide. Esta produz valores contínuos no intervalo $[0,1]$, muitas vezes, porém, é interessante que a função de ativação produza saídas no intervalo $[-1,1]$. Para suprir essa necessidade, outra função muito comum utilizada em redes neurais artificiais é a tangente hiperbólica, expressa por:

$$g(\alpha, \beta, \mu) = \alpha \frac{(1 - e^{-\beta\mu})}{(1 + e^{-\beta\mu})} \quad (2.40)$$

$$g'(\alpha, \beta, \mu) = 1 - g(\mu)^2 \quad (2.41)$$

Em que α e β representam valores constantes. A seleção da função de ativação adequada é fundamental para facilitar o processo de convergência da rede neural.

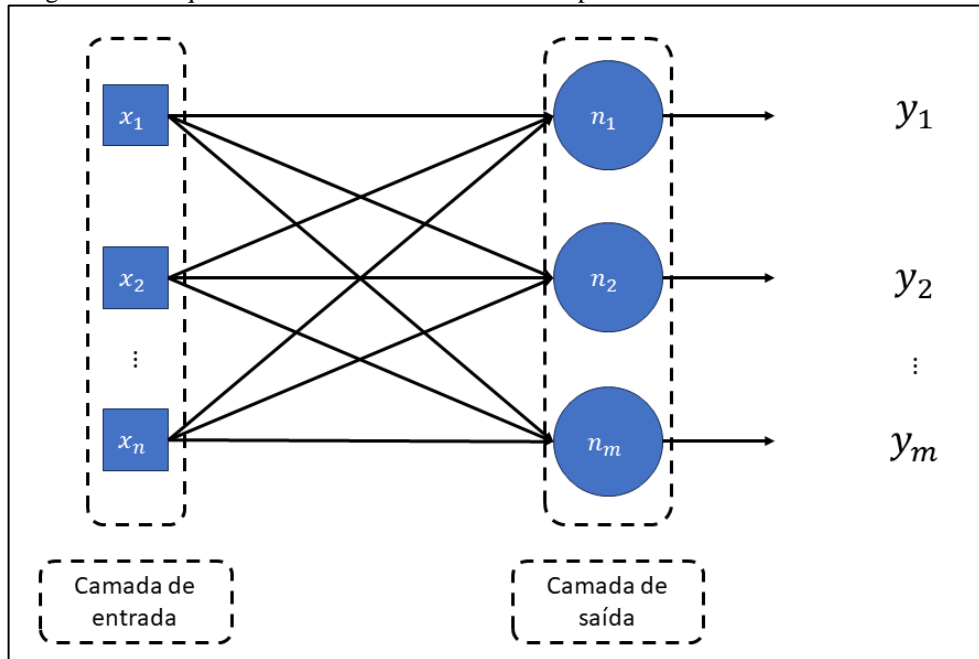
2.4.2 Arquiteturas de redes neurais

A arquitetura de uma rede neural artificial define a forma como os seus neurônios estão dispostos uns em relação aos outros, segundo Silva, Spatti e Flauzino (2016). Uma arquitetura é composta pela camada de entrada da rede, responsável por receber as informações do meio externo, n camadas intermediárias ou ocultas, as quais realizam quase todo o processamento interno da rede e a camada de saída, que irá receber as saídas das camadas ocultas e produzirá o resultado da rede. A seleção de uma determinada arquitetura é feita levando em consideração o tipo de tarefa a ser realizada e a complexidade dela.

2.4.2.1 Arquitetura *feedforward*

Esse tipo de arquitetura é caracterizado pelo fluxo unidirecional da informação. Ele pode ainda ser subdividido em redes de camada simples e camadas múltiplas. A primeira consiste apenas nas camadas de entrada e saída, sendo utilizadas para classificação de padrões e filtragem linear. A Figura 05 ilustra um exemplo deste tipo de arquitetura

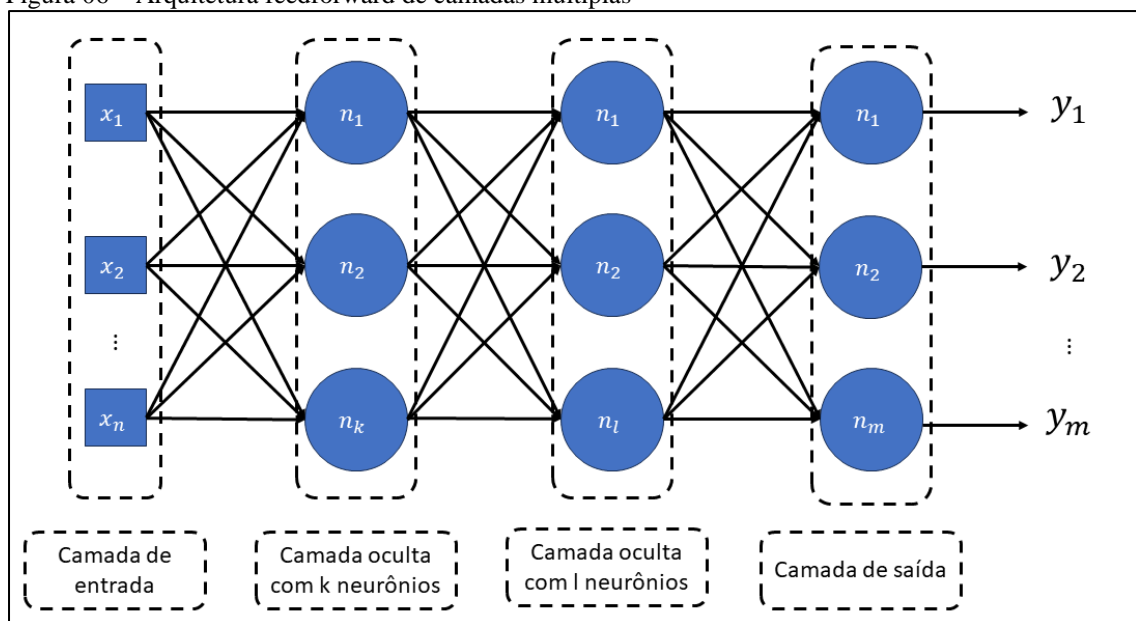
Figura 05 – Arquitetura feedforward de camada simples.



Fonte: Adaptado de Silva, Spatti e Flauzino (2016)

A arquitetura *feedforward* com múltiplas camadas permite solucionar questões mais complexas. Quando uma rede possui mais de uma camada oculta, ela é chamada de profunda e é daqui que se origina o termo *deep learning*. O número de camadas ocultas depende da complexidade do problema e da qualidade dos dados disponíveis, segundo Silva, Spatti e Flauzino (2016). A Figura 06 demonstra uma rede *feedforward* de camadas múltiplas.

Figura 06 – Arquitetura feedforward de camadas múltiplas

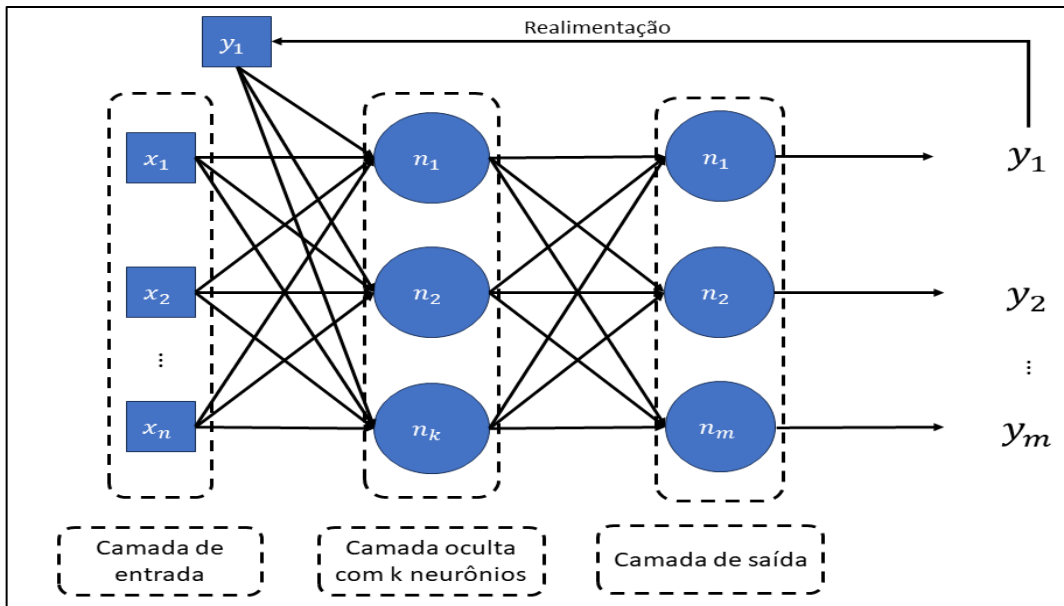


Fonte: Adaptado de Silva, Spatti e Flauzino (2016)

2.4.2.2 Outros tipos de arquiteturas

Redes neurais são objeto de estudo desde a década de 40 com os estudos de Warren McCulloch, segundo Babini e Marranghello (2007). Com isso, é notável que outras arquiteturas foram sendo criadas e testadas para solucionar problemas cada vez mais complexos. A arquitetura recorrente, é bastante utilizada para a previsão de séries temporais, devido à sua característica de realimentação, conforme Silva, Spatti e Flauzino (2016). A Figura 07 traz um exemplo desse tipo de rede:

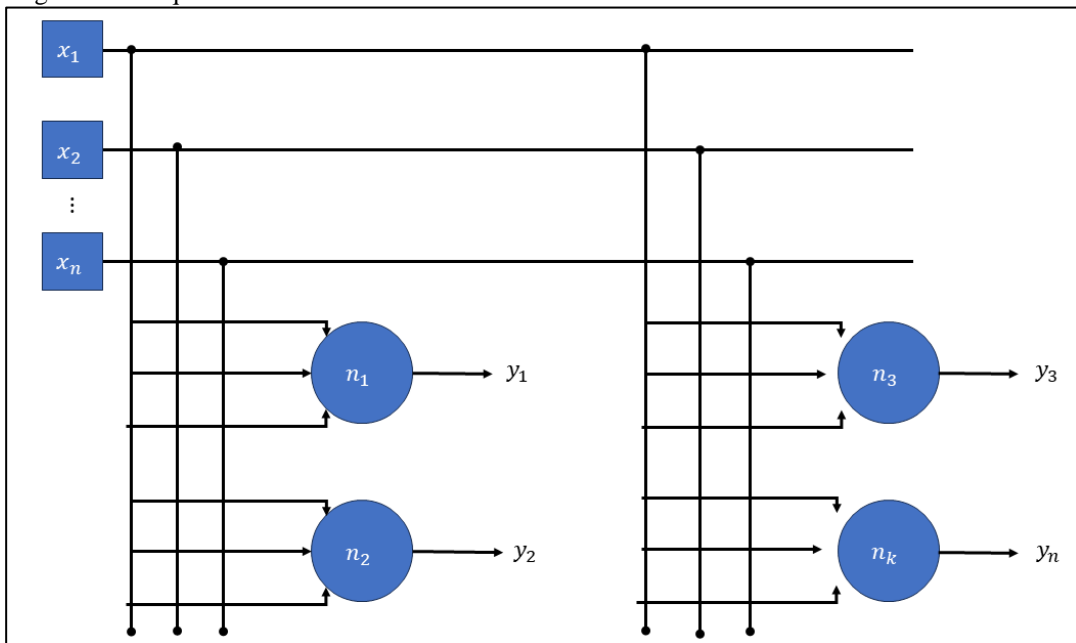
Figura 07 – Rede de arquitetura recorrente



Fonte: Adaptado de Silva, Spatti e Flauzino (2016)

Outro tipo de arquitetura de redes é a de estrutura reticulada, em que há uma consideração espacial dos neurônios, pois ela interfere no processo de aprendizagem. Suas principais aplicações estão em resolução de problemas de agrupamento (*clustering*), reconhecimento de padrões, grafos etc. A Figura 08 traz um exemplo desse tipo de arquitetura.

Figura 08 - Arquitetura em estrutura reticulada



Fonte: Adaptado de Silva, Spatti e Flauzino (2016)

2.4.3 Formas de aprendizagem

Uma rede neural artificial aprende por meio da atualização de seus pesos sinápticos e biases de forma iterativa, através de épocas de treinamento. Em uma única época de treinamento, todas as observações destinadas para a atualização de pesos e biases, também conhecidas como conjunto de treinamento, devem ser apresentadas à rede.

O treinamento é dito supervisionado quando as saídas desejadas que correspondem a cada elemento da camada de entrada são apresentadas à rede para cada época de treinamento, ou seja, a rede aprende como se houvesse um “professor” ensinando a ela. Em outros casos, não se apresenta as saídas desejadas à rede e ela deve reconhecer os padrões presentes no próprio conjunto de treinamento.

Algumas variações do treinamento supervisionado são o aprendizado com reforço, onde a rede tenta ajustar os pesos sinápticos e biases baseada em quaisquer informações do meio externo por meio de tentativa e erro, já que a única resposta disponível é se a saída gerada é satisfatória ou não. Alternativamente, também é possível treinar a rede com lotes de observações, onde ao invés de atualizar os pesos sinápticos e biases uma observação por vez, estes são atualizados em batelada, gerando lotes com n observações. Nesse caso, uma época de treinamento se dá quando todos os lotes são apresentados à rede.

Depois que uma rede neural está devidamente treinada, apresenta-se um novo conjunto de dados, conhecido como conjunto de teste, que será utilizado apenas para verificar o comportamento da rede neural frente a dados desconhecidos, ou seja, busca-se avaliar a capacidade de generalização da rede.

2.4.4 Algoritmo de treinamento.

Esta seção irá se dedicar à descrição de como as redes neurais com arquitetura *feedforward* utilizadas no presente trabalho atualizam os seus parâmetros (pesos sinápticos e biases) com o intuito de se obter as saídas desejadas. Para isso, será necessária uma breve introdução ao processo de treinamento dos *perceptrons*, uma vez que cada neurônio da rede neural atua como um. Por essa razão, as redes neurais com muitas camadas também são conhecidas como *perceptrons* de múltiplas camadas, do inglês *Multilayer Perceptron* (MLP).

2.4.4.1 Perceptron e Adaline

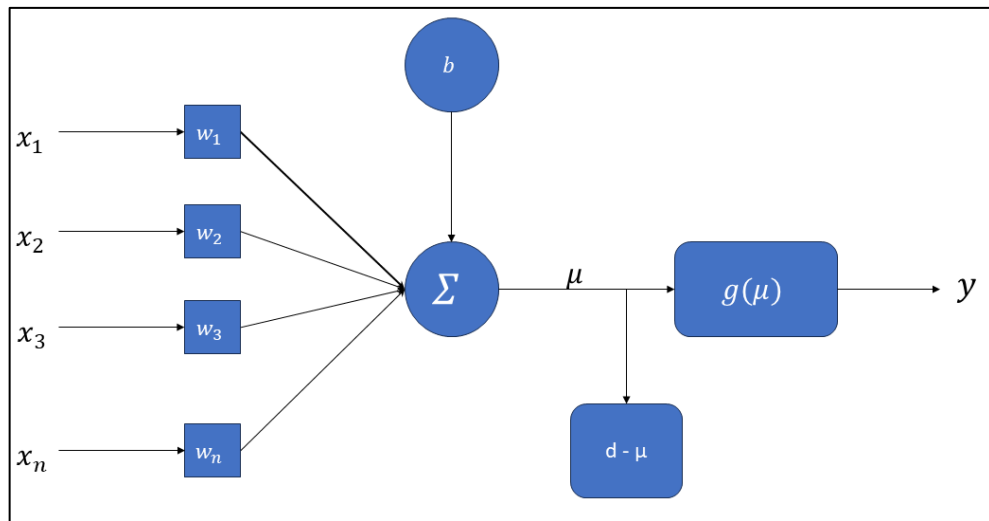
A Figura 04 apresentada na seção 2.4.1 ilustra o exemplo de um *perceptron*. Quando ele é utilizado para propósitos classificatórios com duas classes, ele realiza o ajuste de seus parâmetros por meio do processo de aprendizagem de Hebb (Silva; Spatti; Flauzino, 2016). Nesse processo, os parâmetros são ajustados quando há uma diferença entre a saída desejada e aquela produzida pela rede por meio da expressão:

$$w_i^l = w_i^{l-1} + \eta(d^k - y)x^k \quad (2.42)$$

Onde w_i representa o peso sináptico ligado ao i -ésimo elemento do vetor de entrada, η representa a taxa de aprendizagem que informa o quão rápido a rede irá convergir, d^k e x^k representam o k -ésimo par entrada-saída. Vale salientar que “ l ” se refere à época de treinamento e os primeiros elementos dos vetores de peso e entrada referem-se ao bias e ao valor unitário, respectivamente. O *perceptron* continua o seu treinamento até que não exista diferença entre a saída produzida e a desejada, segundo Silva, Spatti e Flauzino (2016).

O adaline (*adaptive linear unit*), idealizado por Widrow e Hoff (1960) representa uma evolução do *perceptron*, apesar de suas similaridades. No adaline, representado pela Figura 09, no adaline, representado pela Figura 09, o erro entre o resultado do combinador linear e a saída desejada será utilizado para ajustar os pesos sinápticos e biases por meio de um método chamado de gradiente descendente, cujo intuito é obter parâmetros que minimizem o erro produzido.

Figura 09 – Representação de um adaline.



Fonte: Adaptado de Silva, Spatti e Flauzino (2016)

Para isso, considera-se que o erro quadrático médio obtido a partir de p amostras se dá por:

$$E(w) = \frac{1}{2} \sum_{k=1}^p (d^k - \mu)^2 \quad (2.43)$$

$$E(w) = \frac{1}{2} \sum_{k=1}^p (d^k - \sum_{i=1}^n w_i x_i - b)^2 \quad (2.44)$$

Para a obtenção de parâmetros ótimos, é necessário calcular o gradiente do erro em relação aos parâmetros, ou seja, calcular a seguinte expressão:

$$\nabla E(w) = \frac{\partial E(w)}{\partial w} \quad (2.45)$$

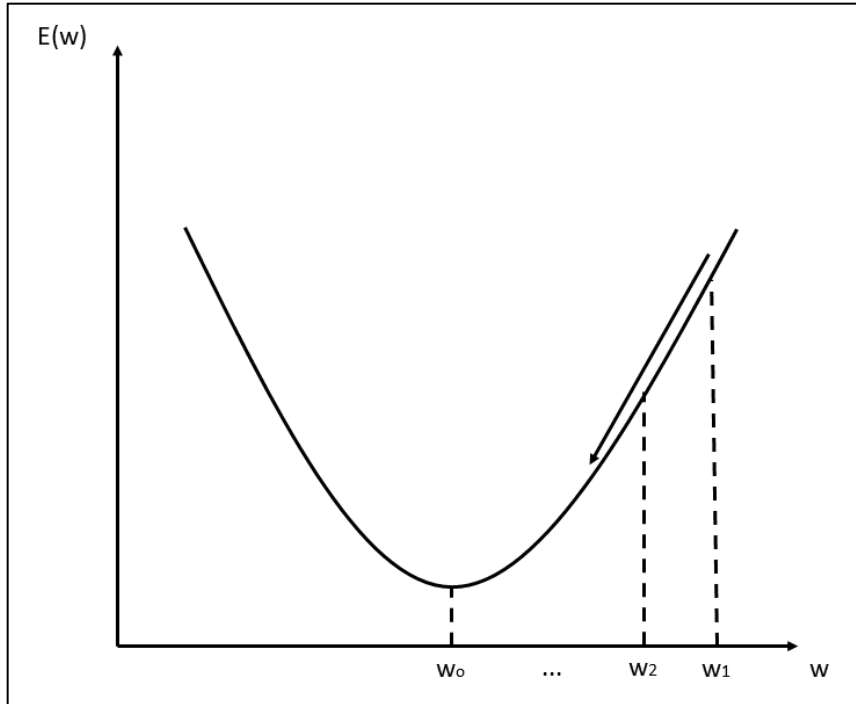
$$\nabla E(w) = - \sum_{k=1}^p (d^k - \mu) x^k \quad (2.46)$$

Com isso, deve-se ajustar os parâmetros da rede na direção oposta ao gradiente, a fim de minimizar o erro. Esse procedimento pode ser realizado da seguinte maneira:

$$w_i^l = w_i^{l-1} - \eta \nabla E(w) \quad (2.47)$$

A Figura 10 traz à tona uma interpretação geométrica do processo de convergência do adaline.

Figura 10 – Processo de convergência do adaline.



Fonte: O próprio autor

2.4.4.2 Rede Multilayer perceptron (MLP)

Após tornar claro o funcionamento do processo de otimização dos parâmetros do *perceptron* e do *adaline*, o treinamento da MLP se torna mais fácil de ser descrito. Diferentemente das redes anteriores, o MLP apresenta camadas intermediárias que se situam entre a camada de entrada e a camada de saída, a qual pode conter diversos neurônios. O algoritmo de treinamento das redes MLP se chama *backpropagation* e este será descrito com base em Silva, Spatti e Flauzino (2016).

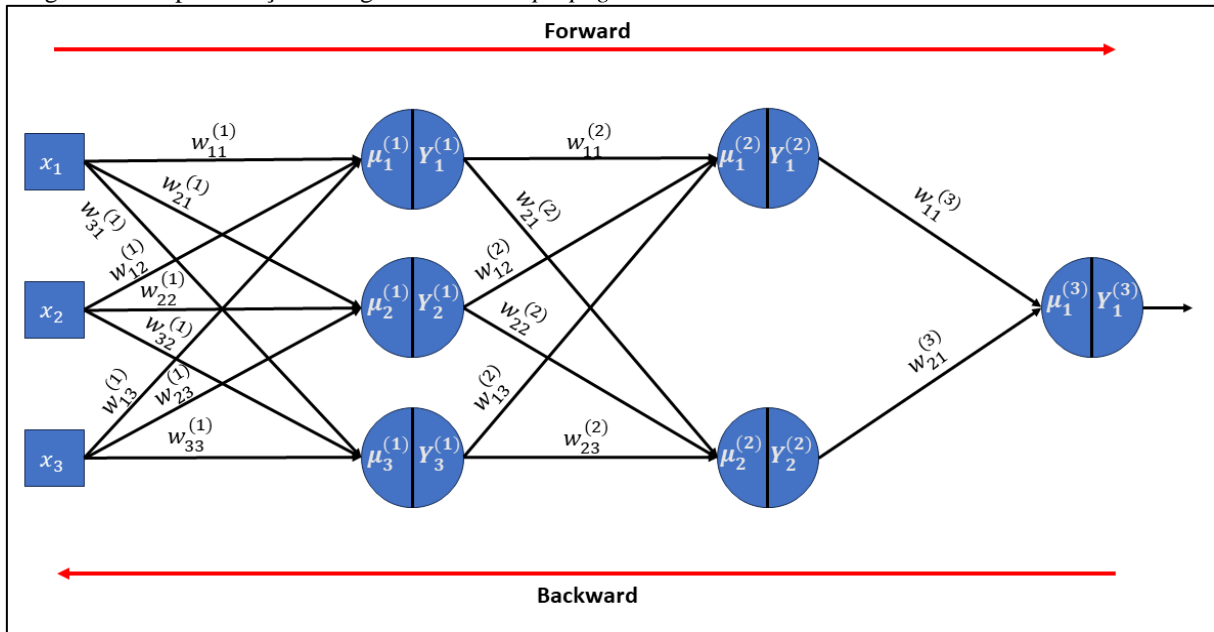
Supondo-se uma rede neural com a configuração apresentada na Figura 11, é necessário realizar as seguintes considerações

- Chama-se $w_{ji}^{(L)}$ a matriz de pesos sinápticos que liga os neurônios da camada L a camada L-1, com seus neurônios sendo representados por “j” e “i”,

respectivamente. Nesse desenvolvimento, os biases estão embutidos na matriz de pesos

- b) O vetor de entrada das camadas será representado $\mu_j^{(L)}$, onde a entrada do j -ésimo neurônio da camada L corresponde ao somatório ponderado entre os pesos do referido neurônio e as saídas da camada L-1.
- c) Y_j^L corresponde ao vetor de saídas da camada L, onde a saída do j -ésimo neurônio corresponde a $g(\mu_j^L)$.

Figura 11– Representação do algoritmo de *backpropagation*.



Fonte: O próprio autor.

Na etapa denominada de *forward*, cada neurônio processa as entradas recebidas e cria uma saída que servirá de entrada para os neurônios da próxima camada. Esse processo se repete até a camada de saída, de onde será produzida a saída final da rede, da qual é obtido o erro quadrado médio em relação à saída desejada, conforme a equação abaixo.

$$E(k) = \frac{1}{2} \sum_{j=1}^{n_3} (d_j(k) - Y_j^{(3)}(k))^2 \quad (2.48)$$

Para p amostras o erro é dado por:

$$E_M = \frac{1}{p} \sum_{k=1}^p E(k) \quad (2.49)$$

Em sequência, a etapa de *backward* envolve a retropropagação do erro obtido pela rede até a primeira camada oculta, com o intuito de ajustar os pesos sinápticos e biases de

cada camada até valores ótimos. Para encontrar os parâmetros ótimos da camada de saída é necessário calcular o gradiente da saída da rede em relação à matriz de pesos da camada de saída, em um procedimento similar ao realizado no treinamento do adaline, conforme a expressão abaixo:

$$\nabla E = \frac{\partial E}{\partial w_{ji}^3} \quad (2.50)$$

Uma vez que a saída gerada pelo neurônio da camada de saída trata-se de uma composição de funções, a equação 2.50 pode ser desenvolvida utilizando a regra da cadeia como:

$$\frac{\partial E}{\partial w_{ji}^3} = \frac{\partial E}{\partial Y_j^{(3)}} \frac{\partial Y_j^{(3)}}{\partial \mu_j^{(3)}} \frac{\partial \mu_j^{(3)}}{\partial w_{ji}^3} \quad (2.51)$$

Cada termo da equação 2.51 pode ser resolvido por meio das seguintes expressões:

$$\frac{\partial E}{\partial Y_j^{(3)}} = -\left(d_j(k) - Y_j^{(3)}(k)\right) \quad (2.52)$$

$$\frac{\partial Y_j^{(3)}}{\partial \mu_j^{(3)}} = g'(\mu_j^{(3)}) \quad (2.53)$$

$$\frac{\partial \mu_j^{(3)}}{\partial w_{ji}^3} = Y_i^{(2)} \quad (2.54)$$

Dessa forma, para atualizar os parâmetros da terceira camada, é necessário se deslocar no sentido contrário ao gradiente, conforme ilustrado na Figura 09, logo, surge a expressão:

$$w_{ji}^{(3)}(t+1) = w_{ji}^{(3)}(t) - \eta \frac{\partial E}{\partial w_{ji}^3} \quad (2.55)$$

Na equação 2.55, “t” representa a contagem de épocas de treinamento, uma vez que o processo se dá iterativamente. Dando continuidade à retropropagação do erro, é necessário ajustar os valores dos parâmetros da segunda camada escondida para que alcancem seus valores ótimos por meio do cálculo do gradiente do erro em relação à matriz de pesos da segunda camada, logo, de maneira similar ao procedimento realizado na camada de saída, tem-se:

$$\frac{\partial E}{\partial w_{ji}^2} = \frac{\partial E}{\partial Y_j^{(2)}} \frac{\partial Y_j^{(2)}}{\partial \mu_j^{(2)}} \frac{\partial \mu_j^{(2)}}{\partial w_{ji}^2} \quad (2.56)$$

O lado direito da equação pode ser desmembrado em:

$$\frac{\partial E}{\partial Y_j^{(2)}} = \sum_{k=1}^{n3} \frac{\partial E}{\partial \mu_k^{(3)}} w_{kj}^{(3)} \quad (2.57)$$

$$\frac{\partial Y_j^{(2)}}{\partial \mu_j^{(2)}} = g'(\mu_j^{(2)}) \quad (2.58)$$

$$\frac{\partial \mu_j^{(2)}}{\partial w_{ji}^{(2)}} = Y_i^{(1)} \quad (2.59)$$

O termo $\frac{\partial E}{\partial \mu_k^{(3)}}$ da equação 2.57 é o gradiente local em relação ao k -ésimo neurônio da camada de saída, logo, é possível reescrever ela como:

$$\frac{\partial E}{\partial Y_j^{(3)}} = - \sum_{k=1}^{n3} \delta_k^{(3)} w_{kj}^{(3)} \quad (2.60)$$

Silva, Spatti e Flauzino (2016) reiteram como esta etapa leva em consideração a matriz de pesos da camada de saída já ajustada anteriormente. Em seguida, repete-se o procedimento da equação 2.55:

$$w_{ji}^{(2)}(t+1) = w_{ji}^{(2)}(t) - \eta \frac{\partial E}{\partial w_{ji}^{(2)}} \quad (2.61)$$

A última etapa da fase de *backward* refere-se à atualização dos pesos da primeira camada intermediária e assim como nas demais camadas, pretende-se encontrar os valores ótimos para os parâmetros dessa camada a fim de minimizar o erro produzido pela rede. Assim, gera-se a expressão para o cálculo do gradiente do erro em relação aos parâmetros da primeira camada:

$$\frac{\partial E}{\partial w_{ji}^{(2)}} = \frac{\partial E}{\partial Y_j^{(1)}} \frac{\partial Y_j^{(1)}}{\partial \mu_j^{(1)}} \frac{\partial \mu_j^{(1)}}{\partial w_{ji}^{(1)}} \quad (2.62)$$

Semelhante às etapas anteriores, é possível desmembrar a equação 2.62 em:

$$\frac{\partial E}{\partial Y_j^{(1)}} = \sum_{k=1}^{n3} \frac{\partial E}{\partial \mu_k^{(2)}} w_{kj}^{(2)} \quad (2.63)$$

$$\frac{\partial Y_j^{(1)}}{\partial \mu_j^{(1)}} = g'(\mu_j^{(1)}) \quad (2.64)$$

$$\frac{\partial \mu_j^{(1)}}{\partial w_{ji}^{(1)}} = x_i \quad (2.65)$$

O termo $\frac{\partial E}{\partial \mu_k^{(2)}}$ representa o gradiente local da segunda camada escondida em relação ao erro da rede. Com isso, obtém-se:

$$\frac{\partial E}{\partial Y_j^{(3)}} = - \sum_{k=1}^{n_3} \delta_k^{(2)} w_{kj}^{(2)} \quad (2.66)$$

Mais uma vez, Silva, Spatti e Flauzino (2016) ressaltam que o cálculo do gradiente da primeira camada oculta leva em consideração os parâmetros ajustados da segunda camada escondida. Finalmente, é possível ajustar os parâmetros da primeira camada escondida por meio de:

$$w_{ji}^{(1)}(t+1) = w_{ji}^{(1)}(t) - \eta \frac{\partial E}{\partial w_{ji}^{(1)}} \quad (2.67)$$

É perceptível que o processo de ajuste dos parâmetros das camadas escondidas é bastante semelhante, sendo reproduzível para qualquer quantidade de camadas escondidas, entretanto, geralmente só são utilizadas até 2 camadas escondidas.

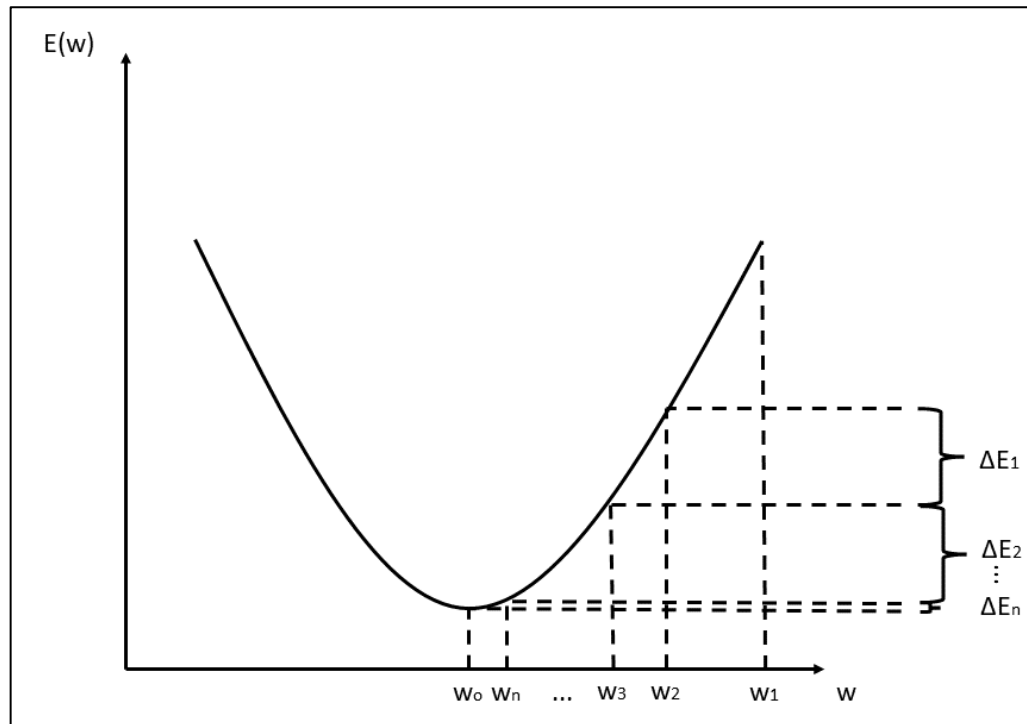
2.4.2.3 Aprimoramento do algoritmo de *backpropagation*.

Desde a sua concepção, diversos autores testam diferentes formas de acelerar o processo de convergência do algoritmo de *backpropagation*. A inserção de um termo de *momentum* na atualização dos parâmetros da rede imprime mais velocidade ao processo, conforme abaixo:

$$w_{ji}^{(L)}(t+1) = w_{ji}^{(L)}(t) - \eta \frac{\partial E}{\partial w_{ji}^{(L)}} + \alpha(w_{ji}^{(L)}(t) - w_{ji}^{(L)}(t+1)) \quad (2.68)$$

Em que α é chamado de taxa de *momentum*. Quanto maior for a diferença entre os parâmetros de épocas de treinamento sucessivas, maior será a contribuição do termo de *momentum*, logo, à medida que a rede se aproxima da convergência, essa contribuição se torna menor. Isso resulta em grandes avanços no início do treinamento. A Figura 12, ilustra esse processo.

Figura 12 – Processo de convergência com aplicação de termo de *momentum*.



Fonte: O próprio autor.

2.4.5 Hiperparâmetros das redes neurais

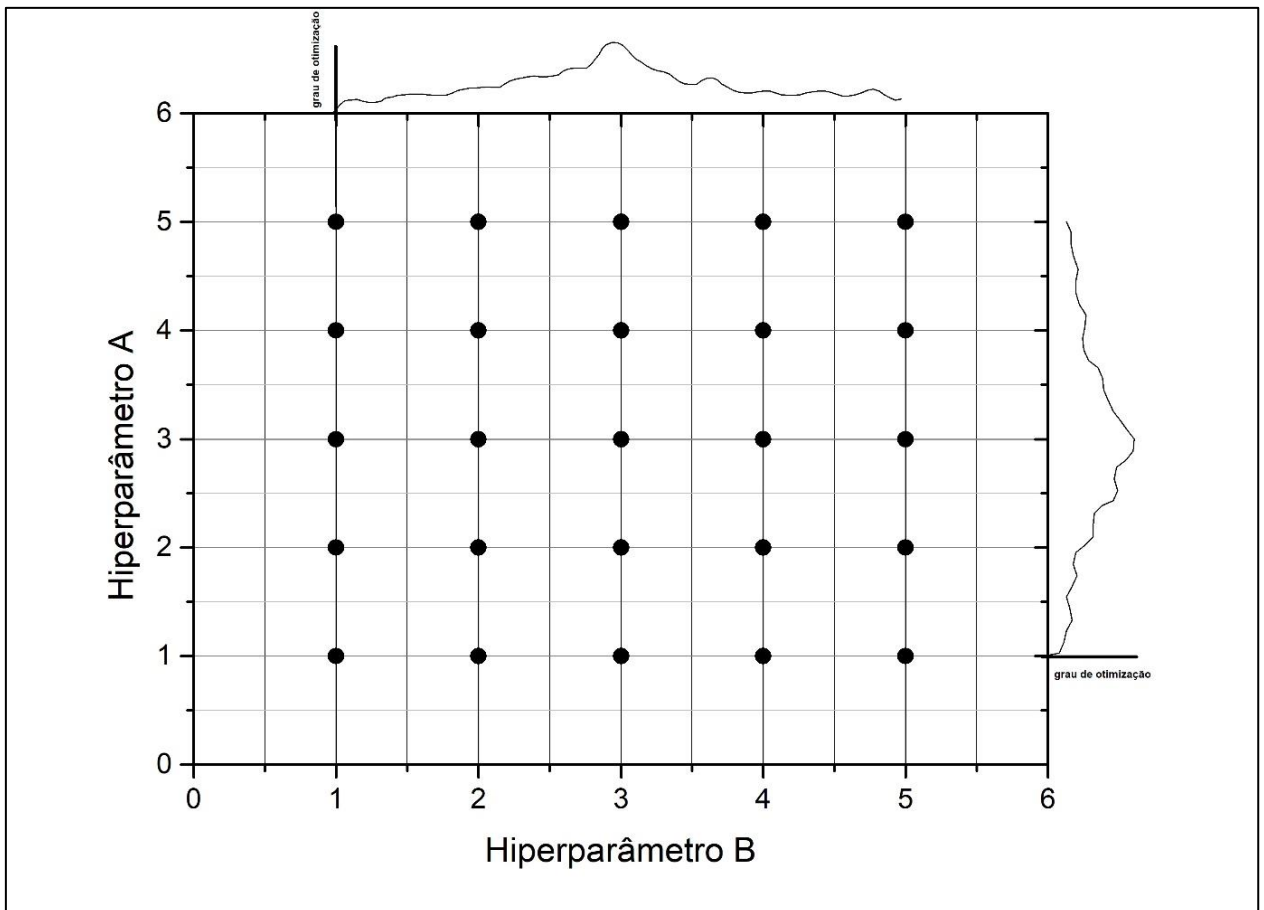
Segundo Goodfellow, Bengio e Courville (2016), os hiperparâmetros são as características da rede que controlam o comportamento do aprendizado. O bom ajuste dos hiperparâmetros de uma rede neural contribui para o sucesso dela. O efeito de cada hiperparâmetro no treinamento é listado abaixo:

- a) Taxa de aprendizagem: ela dita o quão rápido a rede neural irá convergir, segundo Silva, Spatti e Flauzino (2016). Valores muito elevados de taxa de aprendizagem podem levar a rede neural à divergência.
- b) Tamanho das camadas: diversos estudiosos, vêm tentando estabelecer heurísticas para determinar o número de neurônios, mas ainda não há uma forma de determinar a quantidade ideal deles. É correto afirmar que quanto mais complexo o problema a ser resolvido, mais neurônios serão necessários.
- c) Taxa de momentum: é responsável por acelerar o processo de convergência da rede. Normalmente, o seu valor varia de 0 a 1, segundo Villanueva (2011).

- d) Funções de ativação: são responsáveis por introduzir a não linearidade nas redes neurais. A escolha delas, bem com a de suas constantes, depende da faixa de valores em que se encontram os dados disponíveis.

A aplicação de diversas técnicas para melhorar a eficiência das redes neurais leva à criação de outros hiperparâmetros. Há diversas técnicas relacionadas à procura dos melhores hiperparâmetros. Segundo Bergstra *et al.*, a procura em grade (*grid search*) trata-se de uma pesquisa exaustiva, onde todos os valores de hiperparâmetros em uma faixa previamente estabelecida são testados, possuindo grandes desvantagens à medida que o número de hiperparâmetros cresce. A Figura 13, ilustra o conceito da técnica de procura em grade.

Figura 13 – Representação gráfica da técnica de pesquisa em grade.



Fonte: O próprio autor

3 METODOLOGIA

3.1 Aquisição e tratamento de dados

Os dados utilizados para a realização do presente trabalho foram disponibilizados por uma usina siderúrgica integrada localizada no Brasil. O conjunto de dados obtido possui 2465 observações ou pares entrada-saída registrados de forma *online* que correspondem a 13 parâmetros relacionados com a operação do alto-forno, seus produtos e carregamento, monitorados durante o período de 3 meses. Houve períodos de funcionamento pleno e anormalidade durante a operação do alto-forno, estes últimos compreendem paradas programadas e não programadas que geraram dados anômalos (*outliers*) que foram removidos antes de serem aplicados em qualquer modelo. Para garantir a completa remoção desse tipo de dado, as observações de até 3 horas antes e após os períodos de parada programada também foram removidos. Dessa forma, foram mantidas 2188 observações para análise. A Tabela 01 resume os dados mantidos.

Tabela 01 - Resumo dos dados

Parâmetro a ser estimado				
Nome dos parâmetros	Mínimo	Máximo	Média	Desvio Padrão
Teor de silício do ferro-gusa (%)	0.0100	1.3800	0.6127	0.2378
Parâmetros de entrada				
Nome dos parâmetros	Mínimo	Máximo	Média	Desvio Padrão
Ritmo de produção (ton/dia)	3921.2185	9011.2295	6439.8611	868.0800
Temperatura de chama (°C)	2035.0281	2277.4483	2133.2981	32.4222
Pressão de sopro (<i>bar</i>)	0.8938	3.6850	3.0775	0.3218
Temperatura de sopro (°C)	963.0625	1098.4847	1040.5079	23.8489
Sínter na carga (%)	29.7273	69.9948	44.4185	8.3311
Ferro no sinter (%)	49.0000	51.0000	49.8612	0.7265
Fuel rate (Kg/ton de gusa)	427.1126	615.647	521.4388	29.7841
Utilização do gás (%/%)	40.6498	51.1242	47.0233	1.7065
Volume de gás das ventaneiras (Nm ³ /h)	5508.9074	8928.3745	7577.7808	687.8576
Temperatura <i>liquidus</i> da escória (°C)	1336.1889	1401.7012	1369.5673	10.9958
Basicidade B2 da escória (%/%)	0.9100	1.3100	1.1090	0.0656
H ₂ utilizado (%)	52.2870	93.0119	72.4906	6.5767

Fonte: O próprio autor.

Para uma melhor compreensão dos dados, é válido esclarecer os alguns pontos sobre as variáveis:

- a) Apesar da unidade da variável “ritmo de produção do forno” ser de toneladas/dia, as observações foram registradas hora a hora durante todo o período citado anteriormente, bem como as demais variáveis.
- b) A variável “utilização de gás” é adimensional e se refere à quantidade de gás CO que foi transformado em gás CO₂, conforme elucidado em Geerdes, Vliet e Toxopeus (2004).
- c) A variável “Temperatura *liquidus* da escória” foi obtida a partir da sua fórmula teórica e do acompanhamento online dos fatores que a afetam.
- d) A variável “basicidade B2 da escória” é adimensional e diz respeito à razão entre os percentuais de CaO e SiO₂ presentes na escória de acordo com Geerdes, Vliet e Toxopeus (2004).
- e) A variável “H₂ utilizado” refere-se ao gás hidrogênio no topo do alto-forno.
- f) As variáveis “Sinter na carga” e “Ferro no sinter” foram considerados parâmetros operacionais devido à possibilidade de ajuste durante o período de operação.

3.2 Regressão linear múltipla

Para comparação entre modelos, foi realizada uma regressão linear múltipla entre o teor de silício do ferro-gusa e as demais variáveis, utilizando todas as observações presentes no conjunto de dados e um nível de confiança de 95%, através do *software* Microsoft Office Excel, por meio a ferramenta “Regressão” disponível no suplemento “Análise de dados”. Ela fornece dados importantes como os valores dos coeficientes de cada variável de entrada e de seus respectivos testes de significância, o coeficiente de correlação (R^2), o resultado do teste F para avaliar a significância global do modelo e os testes t para o nível de significância individual de cada variável. Para complementar a análise, a ferramenta “Correlação” que pertence ao mesmo suplemento foi utilizada para a obtenção da matriz de correlação entre as variáveis.

Dessa forma, pretende-se comprovar a eficiência do uso de redes neurais artificiais para uma tarefa de estimação do teor de silício do ferro-gusa em comparação com técnicas que fornecem modelos lineares.

3.3 Rede neural artificial

O presente trabalho utilizou uma rede neural de arquitetura *feedforward* totalmente conectada com duas camadas escondidas e um único neurônio na camada de saída implementada por meio do *software* MATLAB para a tarefa de estimação do teor de silício presente no ferro-gusa. Antes de serem apresentados à rede neural, os dados precisam ser normalizados para facilitar o aprendizado da rede, conforme Sola e Sevilla (1997). Esse procedimento é realizado de forma que os dados são transportados para um novo intervalo de valores, os quais podem ser $[-1,1]$ ou $[0,1]$. No presente trabalho, quando a função sigmoide foi utilizada como função de ativação nos neurônios da camada oculta, as entradas foram normalizadas no intervalo $[0,1]$ e quando a função tangente hiperbólica foi utilizada, o mesmo procedimento foi realizado utilizando o intervalo $[-1,1]$.

Diversas combinações de hiperparâmetros foram testadas com a arquitetura citada anteriormente, dentre os quais estão:

- a) O número de neurônios nas camadas escondidas, variando entre 6 e 15.
- b) A função de ativação das camadas escondidas, optando-se entre as funções tangente hiperbólica e sigmoide logística.
- c) A função de ativação da camada de saída, optando-se entre as funções linear, tangente hiperbólica e sigmoide logística.
- d) As constantes da função tangente hiperbólica “ α ” entre 1 e 1.17159 e “ β ” entre 1 e 0.666.
- e) A constante de ativação da função sigmoide logística variando entre 1 e 2.2.
- f) A taxa de *momentum* da rede.

Dessa forma, foram testadas 3400 combinações dos hiperparâmetros mencionados acima e para cada combinação, foram gerados 100 conjuntos aleatórios ao dividir o banco de dados original de modo que 80% das observações foram utilizadas para treinar a rede neural e 20% foram destinadas para teste, a fim de avaliar o comportamento da rede perante dados que não foram apresentados a ela antes.

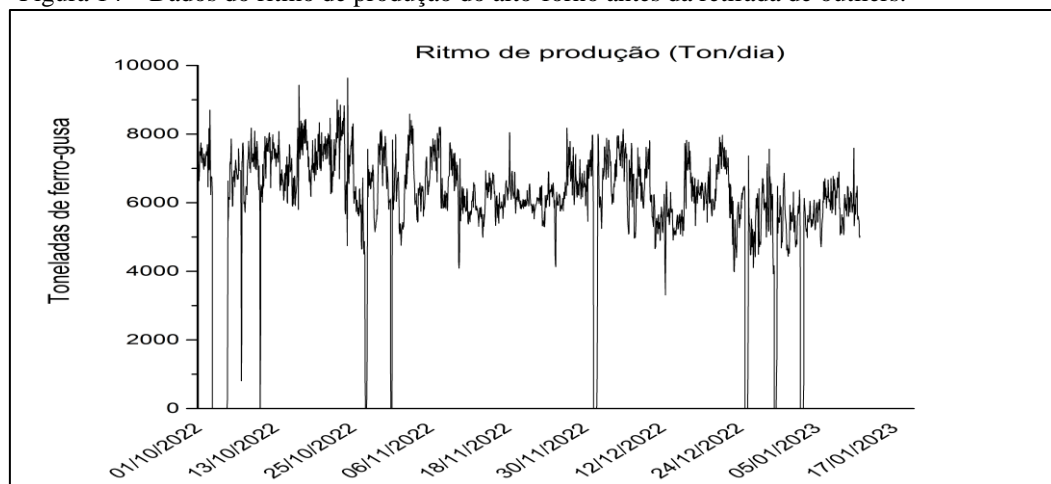
O coeficiente de correlação (R^2) foi utilizado como métrica para comparar a saída estimada pela rede e a saída verdadeira do conjunto de teste. Com isso, obtiveram-se a média

e o desvio padrão dos valores de coeficiente de correlação para cada combinação de hiperparâmetros.

4 RESULTADOS E DISCUSSÃO

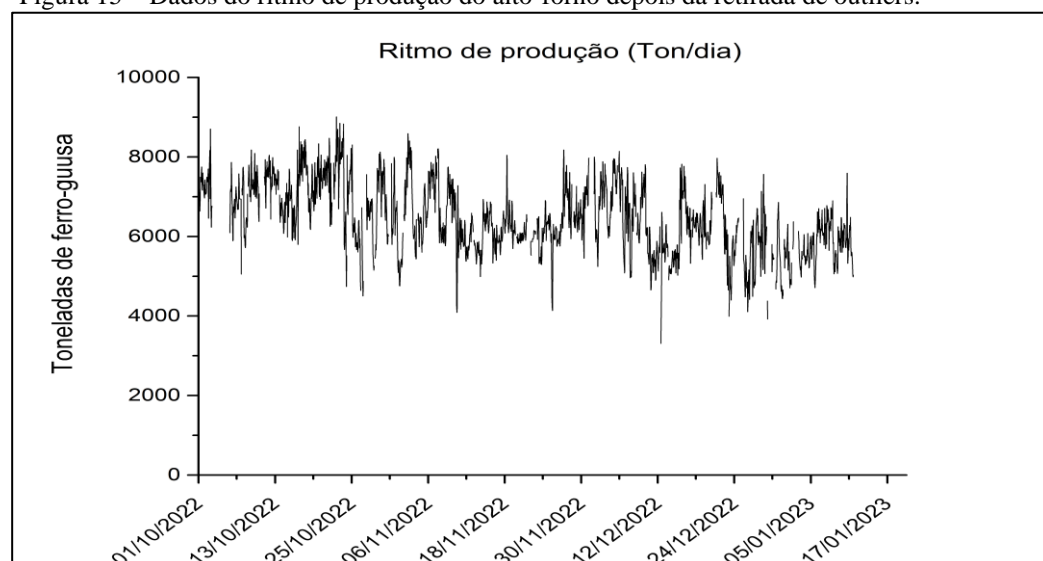
As figuras 14 e 15 ilustram o efeito da retirada das observações anômalas relacionadas às paradas do alto-forno durante o período de estudo, tendo a variável “Ritmo de produção” como exemplo.

Figura 14 – Dados do ritmo de produção do alto-forno antes da retirada de outliers.



Fonte: O próprio autor.

Figura 15 – Dados do ritmo de produção do alto-forno depois da retirada de outliers.



Fonte: O próprio autor.

A remoção de *outliers*, tanto para a regressão linear múltipla como para a rede neural, melhora a qualidade dos modelos gerados a partir dos dados restantes. Essa remoção foi realizada em todas as variáveis. A seguir, a Tabela 02 sintetiza as principais métricas utilizadas para avaliar o desempenho global da regressão linear múltipla.

Tabela 02 – Principais métricas de desempenho da regressão linear múltipla.

Estatísticas de regressão	
Coeficiente de correlação (R^2)	0.0971
Teste F de significância global do modelo	19.5016
P-valor	6.3151×10^{-41}

Fonte: O próprio autor

O R^2 obtido pela regressão linear múltipla indica que o modelo não conseguiu explicar adequadamente o comportamento dos dados, prevendo apenas aproximadamente 9,7% do fenômeno. Esse resultado está de acordo com o esperado uma vez que os fenômenos e reações que ocorrem no interior do alto-forno são complexos e não lineares.

O valor do teste F calculado e o P-valor obtidos indicam que há indícios de que é possível rejeitar a hipótese nula, pois o primeiro é maior do que o F tabelado, considerando 12 graus de liberdade no numerador e 2175 no denominador da tabela de distribuição de Fisher-Snedecor, o que representa um F tabelado entre 1,83 e 1,75. Já o P-valor é menor do que o nível de significância $\alpha = 0,05$.

Com isso, é possível afirmar que pelo menos um dos coeficientes da regressão é diferente de zero, mas apenas esta informação é insuficiente para analisar a qualidade do modelo linear, sendo necessária uma análise individual dos coeficientes da regressão. Em seguida, a Tabela 03 apresenta os coeficientes obtidos pela regressão linear múltipla e seus respectivos testes de significância individuais.

Tabela 03 – Resultados obtidos a partir da regressão linear múltipla.

Nome da variável	Coefficientes	Erro padrão	Estatística t	P-valor
Interseção	7.1169	0.9602	7.4119	1.7733×10^{-13}
Ritmo de Produção	-1.2236×10^{-5}	1.1582×10^{-5}	-1.0565	0.2909
Temperatura da chama	-8.199×10^{-4}	2.0383×10^{-4}	-4.0226	5.95201×10^{-5}
Temperatura de sopro	2.9822×10^{-4}	2.3072×10^{-4}	1.2926	0.1963
Ferro no Sinter	-0.0987	0.0116	-8.4838	3.9681×10^{-17}
Fuel Rate	-0.0015	2.6794×10^{-4}	-5.5236	3.7185×10^{-8}
Utilização do gás	0.0052	0.0039	1.2957	0.1952
Volume de gás das ventaneiras	-1.5111×10^{-4}	2.6181×10^{-5}	-5.7720	8.9542×10^{-9}
Temperatura liquidus da escoria	-1.5442×10^{-4}	4.5461×10^{-4}	-0.3397	0.7341
Sinter na carga	-5.9702×10^{-4}	7.5595×10^{-4}	-0.7898	0.4298
Basicidade B2 da escória	0.1058	0.0825	1.2815	0.2002
H ₂ utilizado	0.0079	0.0011	6.6287	4.2574×10^{-11}
Pressão de Sopro	0.3758	0.0425	8.8378	1.9698×10^{-18}

Fonte: O próprio autor.

Para um nível de confiança de 95%, as variáveis que apresentam um P-valor menor do que 0.05 são consideradas significantes para o modelo (Arsham, 1988). Além disso, é notável que estas também obtiveram um valor (em módulo) de estatística t superior a 1.911, o que também evidencia a significância delas para o nível de confiança citado anteriormente. Dessa forma, percebe -se que as variáveis “Temperatura de chama”, “Ferro no sinter”, “Fuel Rate”, “Volume de gás nas ventaneiras”, “H₂ utilizado” e “Pressão de sopro” são significantes para o modelo, entretanto, isso não significa que as demais variáveis são descartáveis, mas apenas que elas não possuem uma relação linear significativa com o teor de silício no ferro-gusa.

Analisando os coeficientes obtidos pela regressão linear múltipla, observa-se que as variáveis “ritmo de produção”, “temperatura de sopro”, “percentual de ferro no sinter”, “sinter na carga” e o “volume de gás nas ventaneiras” apresentaram um comportamento dentro do esperado, conforme abaixo:

- a) O “ritmo de produção” contribuiu negativamente para o modelo. Segundo Saxén e Pettersson (2007), baixos teores de silício podem resultar em resfriamento, o que pode gerar perda de ritmo. Além disso, quanto é provável que quanto maior o ritmo de produção, menor será o tempo disponível para a incorporação do silício no ferro fusa.
- b) A “temperatura de sopro” mais elevada provoca uma elevação do teor de silício no ferro-gusa segundo o modelo. De acordo com Vieira (2012), temperaturas de sopro

mais altas contribuem para uma maior temperatura de chama, pois há o favorecimento da formação do gás SiO .

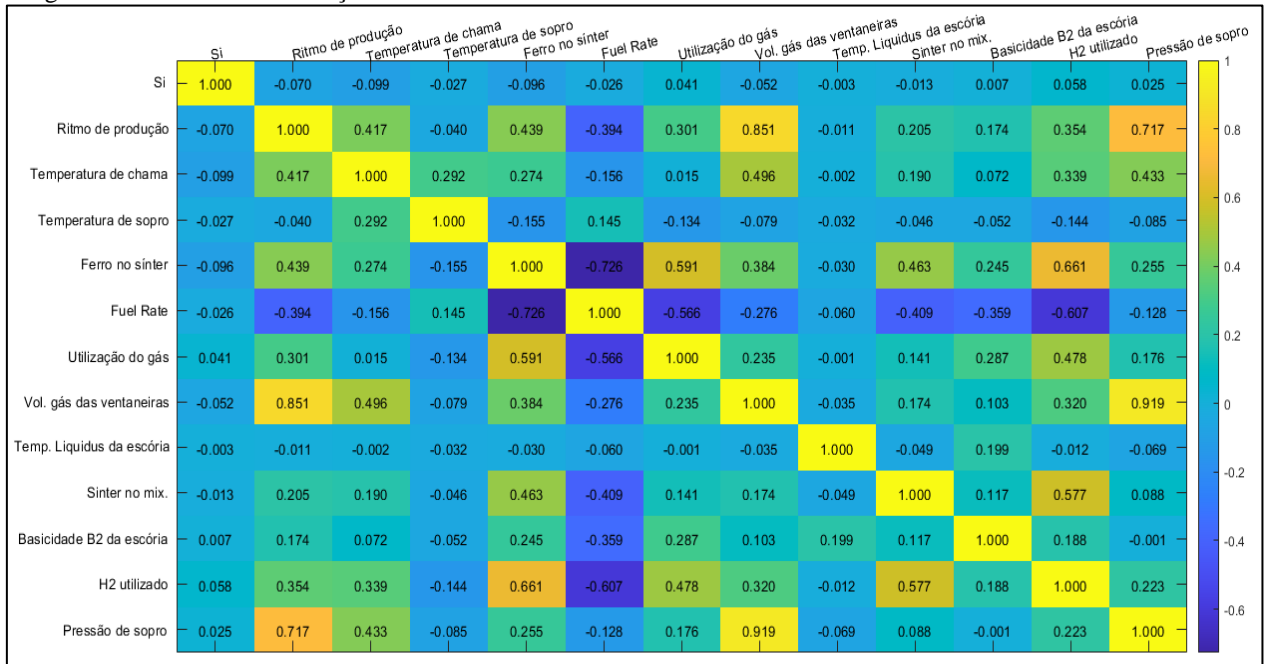
- c) Os coeficientes do “percentual de ferro no sinter” e de “sinter na carga” possuem uma contribuição negativa para o modelo, uma vez que uma maior quantidade de ferro na carga que entra no alto-forno significa um menor teor de sílica entrando no sistema, o que leva a uma menor incorporação no ferro-gusa.
- d) O “volume de gás nas ventaneiras” influencia o “ritmo de produção”, dessa maneira, quando o primeiro aumenta, o teor de silício também é reduzido, o que é confirmado pelo modelo.

As variáveis “temperatura da chama”, “pressão de sopro” e a “basicidade B2 da escória” apresentaram um comportamento fora do esperado, conforme as razões abaixo:

- a) O coeficiente da “temperatura de chama” possui contribuição negativa. Segundo David (2013), o aumento da temperatura deve favorecer a cinética da formação do gás SiO , que leva à incorporação de silício no ferro-gusa.
- b) A influência da “Pressão de sopro” sobre o modelo foi positiva, mas Diniz (2018) relata que a pressão de sopro desfavorece a formação de gás SiO , o que reduz a incorporação de silício no ferro-gusa.
- c) O coeficiente da “Basicidade B2 da escória” contribuiu positivamente para o modelo. Geerdes, Vliet e Toxopeus (2004) definem a basicidade binária como a razão entre os teores de CaO e SiO_2 na escória. Segundo David (2013), ambas as rotas direta e indireta dependem da quantidade de SiO_2 disponível na escória para serem transferidos para o ferro-gusa. Dessa forma quanto maior a basicidade binária, menor o teor de SiO_2 na escória e conseqüentemente, menor a quantidade deste disponível para ser transferida para o ferro-gusa.

É provável que o comportamento fora do esperado ocorra porque as variáveis utilizadas para a construção do modelo linear não são totalmente independentes umas das outras, conforme a matriz de correlação exposta na Figura 16, onde é possível ver correlações médias e altas entre as variáveis de entrada.

Figura 16 – Matriz de correlação entre as variáveis de entrada do modelo linear



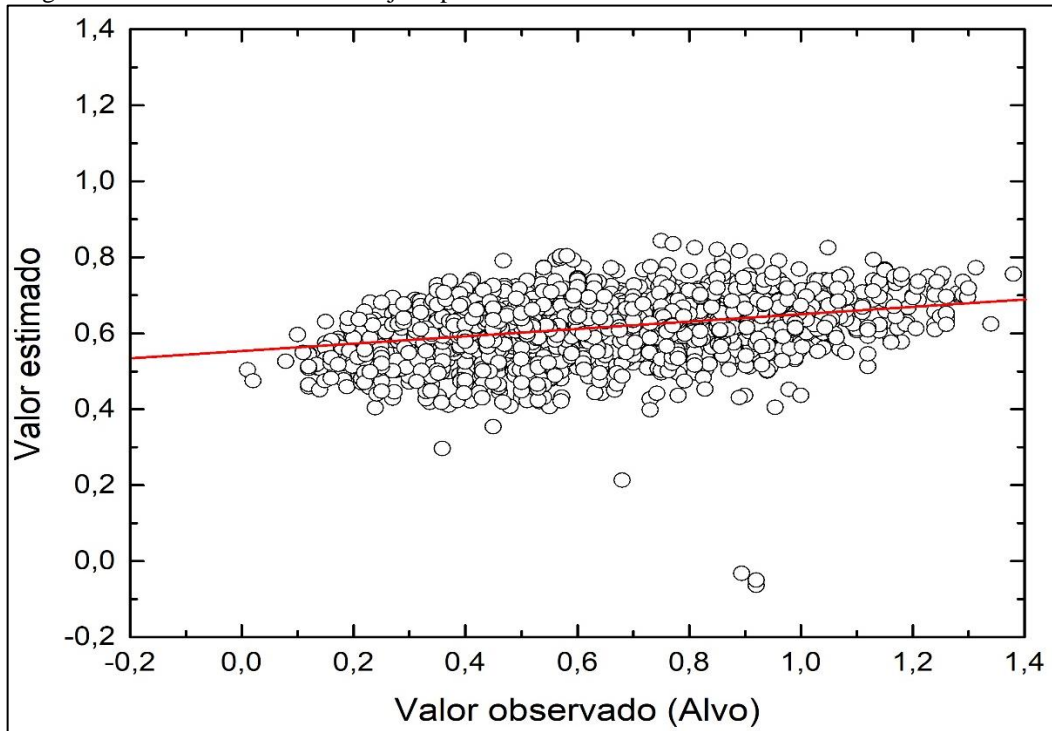
Fonte: O próprio autor.

Além disso, também é provável que isso corra por conta da influência de outras variáveis presentes durante o processo de redução, mas que não foram registradas para o presente trabalho.

As variáveis “Fuel rate”, “utilização do gás”, “H₂ utilizado” e “Temperatura *liquidus* da escória” são dependentes de fatores internos do alto-forno, bem como o estado termodinâmico deste, logo, a influência teórica destas variáveis é bastante condicional. Vale ainda salientar que há diversos estudos sobre a previsão do teor de silício no ferro-gusa, no entanto, eles frequentemente levam em consideração variáveis diferentes e que eles também são fortemente dependentes das condições em que os altos fornos se encontravam quando foram monitorados.

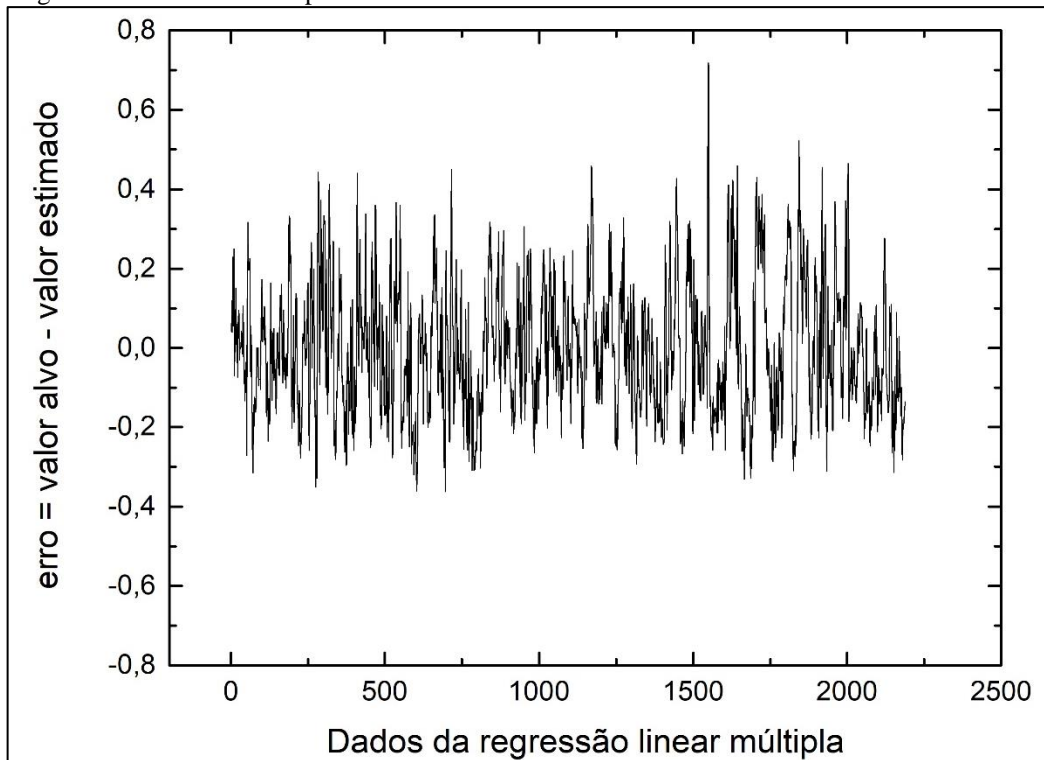
Quanto ao desempenho da previsão do modelo linear, as figuras 17 e 18 ilustram a previsão da saída desejada para o modelo linear e o erro cometido por essa previsão, respectivamente. Na Figura 17, percebe-se que a linha de tendência possui uma inclinação muito pequena e que alguns pontos estão muito afastados dela, o que é consequência do mal desempenho do modelo linear. Na Figura 18, é perceptível que a regressão linear múltipla cometeu erros bastante expressivos, chegando até uma diferença de 0.7 entre o valor real e o valor previsto pelo modelo linear.

Figura 17 – Previsão da saída desejada para o modelo linear.



Fonte: O próprio autor

Figura 18 – Erro cometido pelo modelo linear



Fonte: O próprio autor.

Quanto ao desempenho da rede neural, as Tabelas 04 e 05 indicam as 3 configurações de hiperparâmetros da rede neural que resultaram nos maiores R^2 médios e R^2 máximos,

respectivamente em relação ao conjunto de teste. O R^2 médio é de maior interesse quando comparado com o R^2 máximo, pois busca-se uma configuração que obtenha resultados consistentes frente a novos dados apresentados pela rede.

Tabela 04 – Configurações dos experimentos com os 3 maiores R^2 médios obtidos

experimento	Neurônios na primeira camada oculta	Neurônios na segunda camada oculta	função das camadas ocultas oculta	função da camada de saída	a	b	c	taxa de <i>momentum</i>	R^2 médio	R^2 máximo
392	15	13	sigmoide	linear	0	0	2.2	0.95	0.2615	0.4215
352	14	13	sigmoide	linear	0	0	2.2	0.95	0.2601	0.4089
245	12	7	tanh	linear	1	1	0	0.95	0.2599	0.4189

Fonte: O próprio autor.

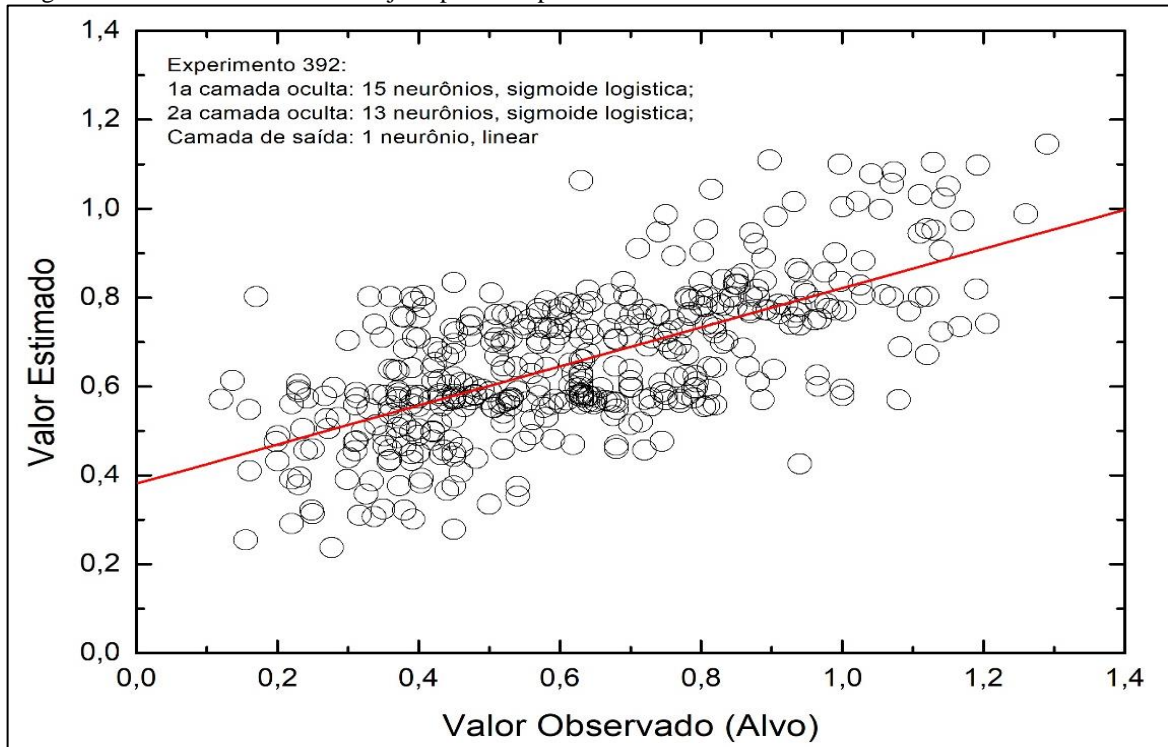
Tabela 05 – Configurações dos experimentos com os 3 maiores R^2 máximos obtidos

experimento	Neurônios na primeira camada oculta	Neurônios na segunda camada oculta	função das camadas ocultas oculta	função da camada de saída	a	b	c	taxa de <i>momentum</i>	R^2 médio	R^2 máximo
249	12	8	tanh	linear	1	1	0	0.95	0.2434	0.4919
1654	11	11	tanh	tanh	1	1	0	0.95	0.2153	0.4727
323	14	6	sigmoide	linear	0	0	1	0.95	0.2149	0.4683

Fonte: O próprio autor.

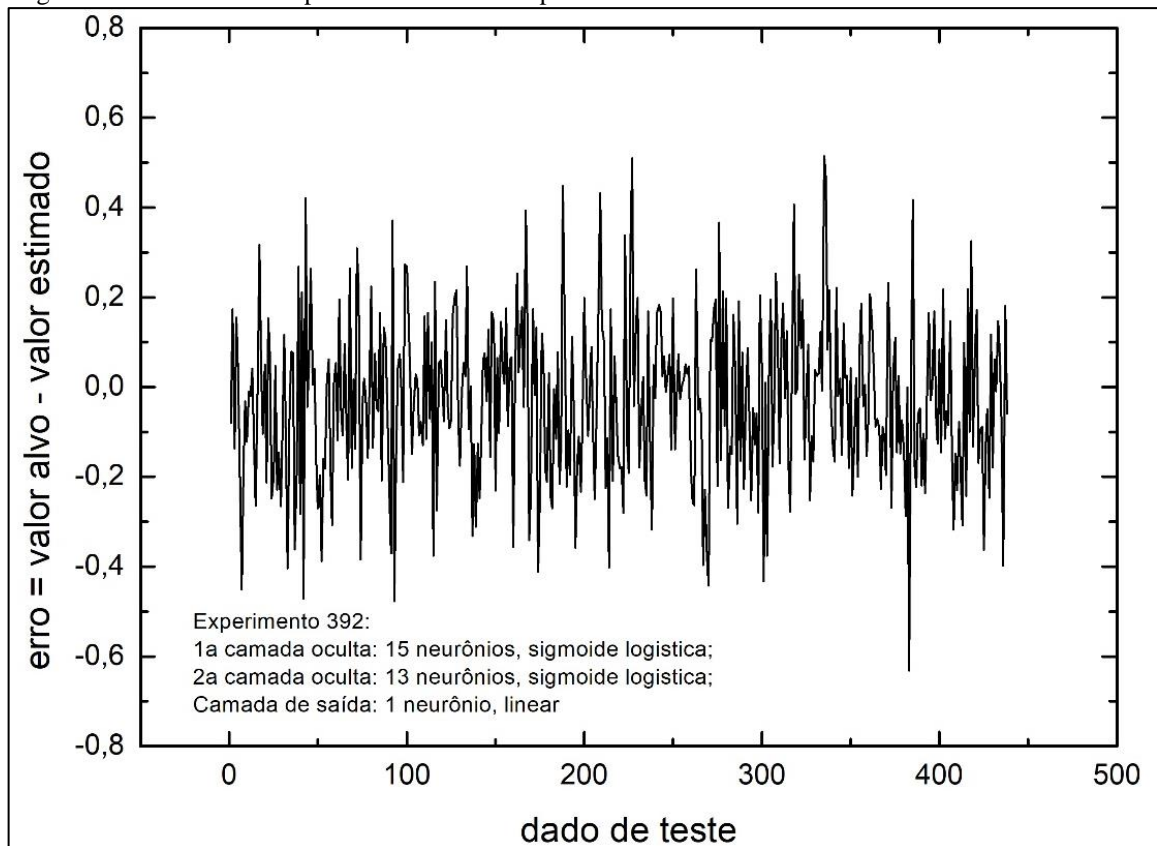
Os termos “a”, “b” e “c” representam as constantes das funções de ativação. Os maiores R^2 médios foram obtidos quando a função sigmoide e linear foram selecionadas como função de ativação das camadas escondidas e da camada de saída, respectivamente. A configuração que obteve o R^2 médio mais elevado possui 15 e 13 neurônios na primeira e na segunda camada oculta, respectivamente. As figuras 19 e 20 ilustram o desempenho de previsão das saídas desejadas do conjunto de teste para essa configuração e o erro cometido por essa previsão.

Figura 19 – Previsão da saída desejada para o experimento 392 da rede neural.



Fonte: O próprio autor.

Figura 20 – Erro cometido pela rede neural no experimento 392



Fonte: O próprio autor.

A rede neural artificial conseguiu prever cerca de 26% do comportamento dos dados, o que é evidenciado pela inclinação mais acentuada da linha de tendência presente na Figura 19 quando comparada com a Figura 17. Esse valor representa uma melhora considerável em relação ao modelo linear, no entanto, ele ainda é considerado baixo quando se considera o ambiente industrial. Além disso, percebe-se que a previsão realizada pela rede neural, evidenciada na figura 18, cometeu erros similares à regressão linear múltipla.

A previsibilidade da rede neural poderia ser melhor, no entanto, é necessário considerar alguns fatores. Dados relacionados com a matéria prima são constantemente citados em estudos que também possuem o objetivo de prever o teor de silício no ferro-gusa e a sua ausência pode explicar o desempenho da rede neural no presente trabalho. A quantidade de dados utilizados também pode ter impactado negativamente no desempenho da rede neural, pois quanto mais dados disponíveis, melhor será a capacidade da rede em aprender o comportamento dos dados e no caso da previsão de silício, geralmente estuda-se o período equivalente a um ano ou dois de operação.

Outros fatores que podem ser observados em estudos sobre o teor de silício no ferro-gusa envolvendo redes neurais são o uso de outros tipos de arquiteturas, como a recorrente, e a presença de técnicas para o pré-processamento dos dados. Com isso, ressalta-se que a utilização de redes neurais para a previsão do teor de silício no alto-forno possui bastante potencial, mas que são necessários estudos mais aprofundados no futuro.

5 CONCLUSÃO

O presente trabalho demonstrou os resultados obtidos por uma rede neural do tipo *multilayer perceptron* (MLP) para a tarefa de previsão do teor de silício no ferro-gusa utilizando parâmetros operacionais do alto-forno, comparando o seu resultado com um modelo linear.

A etapa de aquisição e pré-tratamento dos dados demonstrou-se vital para o processo. A remoção de *outliers* realizada no presente trabalho contribuiu para que houvesse melhora na qualidade dos dados, no entanto, apenas isso demonstrou-se ser insuficiente para garantir uma boa capacidade de previsão da rede neural. Nesse sentido, percebe-se que estudos mais avançados utilizam técnicas de pré-processamento dos dados antes da sua aplicação em redes neurais, a fim de melhorar o desempenho delas.

O modelo linear gerado a partir da regressão linear múltipla apresentou uma baixíssima capacidade de prever o comportamento dos dados, o que já era esperado, obtendo um coeficiente de correlação (R^2) de apenas 0.0971, ou seja, o modelo linear conseguiu prever apenas 9,7% do comportamento dos dados. Com o modelo linear, identificou-se que 5 das variáveis eram significativas para o modelo e, além disso, não foi possível identificar corretamente o comportamento de todas as variáveis utilizadas. É provável que isso ocorra por causa de correlação entre as variáveis utilizadas no presente trabalho e outras que não foram registradas, mas que também influenciam no teor de silício do ferro-gusa.

É notável que a aplicação da rede neural artificial conseguiu um desempenho superior ao modelo linear, obtendo um coeficiente de correlação (R^2) de 0.2615, ou seja, o modelo conseguiu prever aproximadamente 26% do comportamento dos dados. No entanto, a rede neural não conseguiu obter resultados aplicáveis em casos reais, pois a diferença entre os valores reais e estimados ainda foi muito elevada. Isso pode ocorrer por conta da correlação entre as variáveis do processo registradas para o presente trabalho e outras não registradas, mas também pode ser o efeito da utilização de uma arquitetura não adequada para a tarefa de estimação do teor de silício no ferro-gusa e da quantidade de dados utilizada para treinar a rede neural. Outros estudos utilizam arquiteturas recorrentes para a mesma tarefa e capturam uma quantidade de dados que corresponde a 1 ou 2 anos de operação.

A boa operação do alto-forno é de grande importância para a qualidade da entrega de ferro-gusa para a aciaria, no entanto, quando se trata da previsão do teor de silício deste, variáveis relacionadas com a matéria prima e condições ambientais são necessárias para a obtenção de modelos aplicáveis ao ambiente industrial. Portanto, a previsão do teor de silício no ferro-gusa trata-se de uma tarefa complexa, na qual o entendimento dos fatores operacionais e termodinâmicos que influenciam a incorporação do silício, a escolha cuidadosa e o pré-tratamento dos dados e a aplicação de uma técnica que consiga identificar as relações não lineares que ocorrem no interior do alto-forno são fundamentais para realizar essa tarefa com sucesso.

REFERÊNCIAS

AGUIAR, Péricles Guimarães Oliveira. **Balço de massa integrado do ferro-gusa entre altos-fornos e Aciaria da Aperam South America**. 2013, Dissertaço (Mestrado em Engenharia Metalúrgica, Materiais e de Minas) – Programa de pós-graduaço em Engenharia Metalúrgica, Materiais e de Minas. Universidade Federal de Minas Gerais. Belo Horizonte, 2016

ARSHAM, Hoseim. **Kuiper's P- value as a measuring tool and decision procedure for the goodness-of-fit test**. Journal of Applied Statistics, v. 15, p.131-135, 1988.

BABIBI, Maurizio; MARRANGHELLO, Norian. **Introduço às Redes Neurais Artificiais**. São Paulo: Cultura Acadêmica, São Paulo: Laboratório Editorial do IBILCE, 2007.

BERGSTRA, James; BARDENET, Rémi; BENGIO, Yoshua; KÉGL, Balázs. **Algorithms for Hyper-parameter Optimization**. In: Advances in Neural Information Processing Systems 24. Granada, **Proceedings** [...], Granada: Curran Associates inc, 2011.

BISWAS, Anil K. **Principles of Blast Furnace Ironmaking: Theory and practice**. Brisbane, Australia: Cootha Publishing House, 1981.

BULSARI, A.B.; SAXÉN, H. **Artificial neural networks for predicting silicon content in raw iron from blast furnaces**. In: Dehne, F., Fiala, F., Koczkodaj, W.W. (eds) Advances in Computing and Information — ICCI '91. ICCI 1991. Lecture Notes in Computer Science, vol 497. Springer, Berlin, Heidelberg.

DARLINGTON, Richard B.; HAYES, Andrew F. **Regression analysis and linear models: Concepts, applications, and implementation**. 1 ed. New York: The Guilford Press, 2016.

DAVID, Sayd Farage. **Análise da influência dos fatores operacionais dos altos fornos sobre o teor de silício e enxofre no ferro gusa por meio de redes neurais artificiais**. 2013, Dissertaço (Mestrado em engenharia metalúrgica e de materiais) – Programa de pós-graduaço em Engenharia Metalúrgica e de materiais. Instituto Federal do Espírito Santo, Vitória, 2013.

DINIZ, Ana Paula Miranda. **Modelo de previsão do conteúdo de silício no ferro-gusa usando redes neurais artificiais**. 2018, Dissertaço (Mestrado em engenharia elétrica) – Centro tecnológico, Universidade Federal do Espírito Santo, Vitória, 2018.

FONTES, Diane Otilia Lima; VASCONCELOS, Luiz Gonzaga Sales; BRITO, Romildo Pereira. **Blast furnace hot metal temperature and silicon content prediction using soft sensor based on fuzzy C-means and exogenous nonlinear autoregressive models**. Elsevier, v. 141, p. 107028, 2020.

GEERDES, M.; VLIET, C. V. D.; TOXOPEUS, H. **Modern Blast Furnace Ironmaking**. Cidade de IJmuiden: IOS Pr Inc, 2004.

GHOSH, Ahindra; CHATTERJEE, Amit. **Ironmaking And Steelmaking: Theory and practice**. New Delhi: PHI Learning Private Limited, 2007.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Massachusetts: MIT Press, 2016.

HAYKIN, Simon. **Redes Neurais: princípios e prática**. 2 ed. Porto Alegre: Bookman, 2001.

- HWANG, Win-jay; CHANG, Ling-Ku; CHOU, Wen-Hsien. **Technique of Blowing High Silicon Hot Metal by Double Slag Process in BOF**. China Steel Technical Report, v. 23, p. 1-5, 2010.
- MARTINS, Gilberto de Andrade; DOMINGUES, Osmar. **Estatística Geral e Aplicada**. 5 ed. São Paulo: Atlas S.A, 2014.
- MCLEAN, Alex. Blast furnace reactions *In*: FISCHER, Randy (Coord.); GORDS, Iron.; BROWN, Rick; KUUKSMAN, Peter; POVEROMO, Joe; RORICK, Fred; SOSTAR, Steve. **Blast furnace ironmaking: principles, design, and raw materials**. Hamilton, Canada: McMaster University, 1999, p. 3-3 – 3-31.
- MEDEIROS, F.T.P. de; CUNHA, A. Pitasse da; FILETI, A.M. Fratini. **Neural modeling as a tool to support blast furnace iron making**. *In*: International Symposium on Advanced Control of Chemical Processes, 2004, Gramado. **Anais** [...], Gramado: IFAC, 2006, p. 777 – 782.
- MONTGOMERY, Douglas C.; RUNGER, George C. **Estatística aplicada e probabilidade para engenheiros**. 4 ed. Rio de Janeiro: LTC, 2009.
- MOURÃO, Marcelo Breda (Coord.). **Introdução à Siderurgia**. São Paulo: ABM, 2007.
- ORLOV, Michael L. **Multiple linear regression analysis using Microsoft excel**. Chemistry Department, Oregon State University, 1996
- PEACEY J. G.; DAVENPORT W. G. **The Iron Blast Furnace: Theory and Practice**. 1.ed. London: Pergamon Press, 1979.
- PÉREZ, Jorgerlina; DONAYO, Roberto; GOMEZ, Adolfo; DATA, Agustina, BALANTE, Walter. **Refining high-silicon hot metal in an oxygen converter. New process to decrease slopping and fume emission**. *Revue de Métallurgie*, v. 107, p 319-328, 2010
- SAXÉN, Henrik; PETERSSON, Frank. **Nonlinear Prediction of the Hot Metal Silicon Content in the Blast Furnace**. *ISIJ International*, v. 47, p. 1732-1737, 2007.
- SILVA, IVAN Nunes da; SPATTI, Danile Hernane; FLAUZINO, Rogério Andrade. **Redes Neurais Artificiais para a engenharia e ciências aplicadas: fundamentos teóricos e aspectos práticos**. 2 ed. São Paulo: ArtLiber Editora LTDA, 2016.
- SOLA, J.; SEVILLA, J. **Importance of Input Data Normalization for the Application of Neural Networks to Complex Industrial Problems**. *IEEE transactions on nuclear science*, v. 44, p 1464 – 1468, 1997.
- SPIEGAL, Murray R. **Estatística**. 3 ed. São Paulo: Pearson Macron Books, 1993.
- VIEIRA, D. H. **Avaliação de Parâmetros Operacionais que Influenciam a Taxa de Injeção de Carvão Pulverizado em Altos-Fornos a Carvão Vegetal**. 2012, Dissertação (Mestrado em Engenharia Metalúrgica e de Minas) – Curso de Pós-Graduação em Engenharia Metalúrgica e de Minas. Universidade Federal de Minas Gerais, Belo Horizonte, 2012.
- VILLANUEVA, Wilfredo Jaime Puma. **Síntese automática de redes neurais artificiais com conexões a frente arbitrárias**. 2011, Tese (Doutorado em engenharia elétrica) – Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas, Campinas, 2011.

WIDROW, Bernard; HOFF, Marcian E. **Adaptative switching circuits.** *In:* Western Electronic Show and Convention, 1960, Los Angeles. **Proceedings** [...], New York: Institute of Radio Engineers, 1960, p. 96 – 104.