



UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA METALÚRGICA E DE MATERIAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA E CIÊNCIA
DE MATERIAIS

FABRICIO DAMASCENO BRAGA

TÉCNICAS COMPUTACIONAIS APLICADAS À PREVISÃO DE
PROPRIEDADES DO SÍNTER METALÚRGICO

FORTALEZA

2021

FABRICIO DAMASCENO BRAGA

**TÉCNICAS COMPUTACIONAIS APLICADAS À PREVISÃO DE
PROPRIEDADES DO SÍNTER METALÚRGICO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia e Ciência de Materiais da Universidade Federal do Ceará, como requisito parcial para à obtenção do título de Mestre em Engenharia e Ciência de Materiais.

Orientador: Prof. Dr. Elineudo Pinho de Moura

FORTALEZA

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Biblioteca Universitária

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

- B793t Braga, Fabricio Damasceno.
Técnicas computacionais aplicadas à previsão de propriedades do sínter metalúrgico /
Fabricio Damasceno Braga. – 2022.
117 f. : il. color.
- Dissertação (mestrado) – Universidade Federal do Ceará, Centro de Tecnologia,
Programa de Pós-Graduação em Engenharia e Ciência de Materiais, Fortaleza, 2022.
Orientação: Prof. Dr. Elineudo Pinho de Moura.
1. Sinterização. 2. Propriedades do sínter. 3. Técnicas computacionais de previsão. I.
Título.

CDD 620.11

FABRICIO DAMASCENO BRAGA

TÉCNICAS COMPUTACIONAIS APLICADAS À PREVISÃO DE
PROPRIEDADES DO SÍNER METALÚRGICO

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia e Ciência de Materiais da Universidade Federal do Ceará, como requisito parcial para à obtenção do título de Mestre em Engenharia e Ciência de Materiais.

Aprovada em: _____ de _____ de _____.

BANCA EXAMINADORA

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

Prof. Dr. Andre Jalles Monteiro
Universidade Federal do Ceará (UFC)

Prof. Me. Ulisses Tetti
Companhia Siderúrgica do Pecém (CSP)

Agradecimentos

Agradeço aos meus pais, Fábio Braga e Laélia Braga, por me apoiarem em minha vida e em minhas escolhas.

Agradeço ao Prof. Dr. Elineudo Pinho de Moura pela excelente orientação.

Agradeço à Andressa Lima pelo fiel apoio nos momentos difíceis.

Agradeço aos meus amigos Jeová da Silva, Jamil Nobre e Pedro Paulo por, mesmo longe, sempre estarem perto.

Agradeço aos participantes da banca examinadora Prof. Dr. Andre Jalles Monteiro e Prof. Me. Ulisses Tetti pelo tempo, pelas valiosas colaborações, e pelas sugestões.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

“A verdadeira motivação vem da realização, desenvolvimento pessoal, satisfação no trabalho e reconhecimento.”
(Frederick Herzberg)

Resumo

Tanto a necessidade de produzir sinter metalúrgico com propriedades melhores e mais estáveis, independente das variações da qualidade dos insumos utilizados, tais como minério de ferro e coque, como a de garantir a competitividade na produção do aço diante de um cenário de constantes crises econômicas mundiais, são fatores motivadores para o desenvolvimento de modelos de previsão aplicados para indústria siderúrgica. O presente trabalho propõe o desenvolvimento de ferramentas computacionais para estimar índices de qualidade do sinter metalúrgico a partir de suas características químicas e das variáveis do processo de sinterização. Os indicadores investigados são: Índice de Resistência Shatter (SI), Índice de Redutibilidade (RI), Índice de Degradação Sob Redução (RDI) e Tamanho Médio de Partícula (MPS). Investigar a influência das variáveis de entrada nos índices de qualidade considerados, avaliar a qualidade dos modelos estimadores desenvolvidos e compará-los com resultados disponíveis na literatura científica, além de verificar a possibilidade de utilização de uma nova técnica de otimização hiperparamétrica são outros objetivos desse trabalho. Algoritmos distintos foram utilizados a fim de se obter o melhor modelo de predição para cada uma das respostas estudadas, são eles: Regressão Linear Múltipla (MLR), Regressão *Stepwise* (SR), Rede Neural *Percéptrons* Múltiplos com algoritmo de aprendizagem do tipo gradiente descendente com *momentum* (MLP-GDM) e Rede Neural *Percéptrons* Múltiplos com algoritmo de aprendizagem do tipo Levenberg–Marquardt (MLP-LM).

Palavras-chaves: Sinterização; Propriedades do Sinter; Técnicas computacionais de previsão.

Abstract

Both the need to produce metallurgical sinter with better and more stable properties, regardless of variations in the quality of the inputs used, such as iron ore and coke, as well as the need to ensure competitiveness in steel production in the face of a scenario of constant global economic crises, are motivating factors for the development of forecast models applied to the steel industry. This work proposes the development of computational tools to estimate metallurgical sinter quality indexes from their chemical characteristics and sintering process variables. The investigated indexes are Shatter Resistance Index (SI), Reducibility Index (RI), Degradation Under Reduction Index (RDI), and Average Particle Size (MPS). Investigating the influence of input variables on the considered quality indicators, evaluating the quality of the estimating models developed and comparing them with results available in the scientific literature, as well as verifying the possibility of using a new hyperparametric optimization technique are other objectives of this work. Different algorithms were used to obtain the best prediction model for each of the studied responses, including multiple linear regression (MLR), *stepwise* regression (SR), multiple perceptron neural network with gradient descent with *momentum* learning algorithm (MLP-GDM), and multiple perceptron neural network with Levenberg–Marquardt learning algorithm (MLP-LM).

Key-words: Sintering; Sinter Properties; Predictive Computational Techniques.

Lista de ilustrações

Figura 1 – Ilustração do processo Dwigth & Lloyd	21
Figura 2 – Esquema simplificado de uma partícula-quasi (micropelota)	22
Figura 3 – Representação esquemática da máquina de sinterização e localização das zonas principais do processo	23
Figura 4 – Corte transversal e perfil térmico de um leito de sinterização	24
Figura 5 – Representação do processo de formação da microestrutura do sínter - Etapa de formação das primeiras ferritas de cálcio	25
Figura 6 – Representação do processo de formação da microestrutura do sínter - Etapa de formação do líquido primário	26
Figura 7 – Representação do processo de formação da microestrutura do sínter - Etapa de formação das ferritas de cálcio	27
Figura 8 – Representação do processo de formação da microestrutura do sínter - Etapa de formação da hematina granular	28
Figura 9 – Representação do processo de formação da microestrutura do sínter - Etapa de formação da hematita secundária	29
Figura 10 – Efeito da redutibilidade da carga no consumo de coque de alto-forno .	31
Figura 11 – Efeito da qualidade da carga ferrosa na permeabilidade do alto-forno .	32
Figura 12 – Efeito da multicolinearidade	36
Figura 13 – Técnica de redimensionamento dos valores para novos mínimos e máximos	41
Figura 14 – Representação gráfica de um neurônio artificial de três entradas e uma saída.	42
Figura 15 – Representação gráfica de uma rede de duas camadas com múltiplos neurônios e múltiplas entradas.	43
Figura 16 – Representação da Função de Ativação Linear e sua derivada.	46
Figura 17 – Representação da Função de Ativação Logística/Sigmóide e sua derivada com $c = 1$	47
Figura 18 – Representação da Função de Ativação Tangente Hiperbólica e sua derivada com $a = 1$ e $b = 1$	48
Figura 19 – Ilustração esquemática do ponto mínimo na curva de validação como critério de parada antecipada da etapa de treinamento de uma rede neural.	55
Figura 20 – Gráfico de dispersão entre os valores previstos e o valores reais de um conjunto de dados de teste	56
Figura 21 – Gráficos de dispersão, histogramas e coeficientes de correlação entre variáveis de entrada	75
Figura 22 – Representação gráfica do fator de inflação da variância.	76

Figura 23 – Influência da velocidade da esteira na frente de queima.	77
Figura 24 – Gráficos de dispersão dos melhores resultados entre os valores previstos e os valores reais dos dados de teste para cada indicador de qualidade do sínter estudado.	84
Figura 25 – Gráfico em série dos valores de teste previstos pela rede MLP do trabalho de Kinnunen e Laitinen (2005) e os valores de teste reais do indicador de qualidade RDI.	87
Figura 26 – Gráfico em série dos valores de teste previstos pela rede MLP-GDM do trabalho de Laitinen e Saxén (2007) e os valores de teste reais do indicador de qualidade TI.	88
Figura 27 – Gráfico em série dos valores de teste previstos por algoritmos de regressão polinomial do trabalho de Donskoi et al. (2007) e os valores de teste reais do indicador de qualidade TI de dois subconjuntos de testes distintos.	88
Figura 28 – Gráfico em série e de dispersão dos valores de teste previstos pela rede MLP-GDM do trabalho de Umadevi et al. (2016) e os valores de teste reais do indicador de qualidade TI.	89
Figura 29 – Gráfico em série dos valores de teste previstos pela rede MLP-GDM do trabalho de Shao et al. (2020) e os valores de teste reais do indicador de qualidade TI.	90
Figura 30 – Gráfico em série e de dispersão dos valores de teste previstos pela rede MLP-GDM do trabalho de Gao et al. (2021) e os valores de teste reais do indicador de qualidade TI.	91
Figura 31 – Análise de sensibilidade do melhor modelo de previsão do Índice de Resistência Shatter (SI).	95
Figura 32 – Análise de sensibilidade do melhor modelo de previsão do Índice de Redutibilidade (RI).	96
Figura 33 – Análise de sensibilidade do melhor modelo de previsão do Índice de Degradação Sob Redução (RDI).	97
Figura 34 – Análise de sensibilidade do melhor modelo de previsão do Tamanho Médio de Partícula (MPS).	98
Figura 35 – Representação esquemática do teste Shatter	114
Figura 36 – Representação esquemática do teste de redutibilidade	115
Figura 37 – Representação esquemática do teste de degradação sob redução	116

Lista de tabelas

Tabela 1 – Distribuição da carga metálica de um alto-forno em diversas regiões do mundo em 2019.	16
Tabela 2 – Variáveis de entrada e suas definições	65
Tabela 3 – Variáveis de saída e suas definições	66
Tabela 4 – Estatísticas descritivas das variáveis de entrada.	67
Tabela 5 – Estatísticas descritivas das variáveis de saída.	68
Tabela 6 – Hiperparâmetros considerados para as configurações da rede MLP-GDM.	70
Tabela 7 – Hiperparâmetros considerados para as configurações da rede MLP-LM.	71
Tabela 8 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão linear múltipla proposta. . .	74
Tabela 9 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão linear múltipla (análise FIV) proposta.	78
Tabela 10 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão <i>stepwise</i> proposta.	79
Tabela 11 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo dentre todas as técnicas de regressão linear utilizadas.	79
Tabela 12 – Resumo dos melhores resultados de previsão das propriedades buscadas classificados de acordo com o R^2 do melhor modelo da configuração de rede MLP - GDM proposta.	81
Tabela 13 – Resumo dos melhores resultados de previsão das propriedades buscadas classificados de acordo com o R^2 do melhor modelo da configuração de rede MLP - LM proposta.. . . .	82
Tabela 14 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo dentre todas as redes neurais utilizadas. .	83
Tabela 15 – Resumo da diferença absoluta de performance no indicador R^2 entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.	85
Tabela 16 – Resumo da diferença absoluta de performance no indicador EPAM entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.	85
Tabela 17 – Resumo da diferença absoluta de performance no indicador EQM entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.	86

Tabela 18 – Análise SVD para determinação do número ótimo de neurônios para cada camada oculta conforme melhor modelo MLP-GDM.	94
Tabela 19 – Análise SVD para determinação do número ótimo de neurônios para cada camada oculta conforme melhor modelo MLP-LM.	94

Lista de abreviaturas e siglas

ADALINE	Elementos adaptativos lineares, tradução do inglês <i>Adaptive Linear Elements</i>
BTP	Ponto de Queima, tradução do inglês <i>Burning-Through-Point</i>
EPAM	Erro Percentual Absoluto Médio
EQM	Erro Quadrático Médio
GDM	Gradiente Descendente com <i>Momentum</i>
LM	Levenberg–Marquardt
logsig	Função de ativação log-sigmoidal
MADALINE	Elementos adaptativos lineares múltiplos, tradução do inglês <i>Multiple Adaptive Linear Elements</i>
MLP	Percétrons Múltiplas Camadas, tradução do inglês <i>Multilayer Perceptrons</i>
MPS	Tamanho Médio de Partícula, tradução do inglês <i>Average Particle Size</i>
R^2	Coefficiente de Determinação
RDI	Índice de Degradação Sob Redução, tradução do inglês <i>Degradation Under Reduction Index</i>
RI	Índice de Redutibilidade, tradução do inglês <i>Reducibility Index</i>
RLM	Regressão Linear Múltipla
RNA	Redes Neurais Artificiais
SI	Índice de Resistência Shatter, tradução do inglês <i>Shatter Resistance Index</i>
SVD	Decomposição em valores singulares, tradução do inglês <i>Singular Value Decomposition</i>
tanh	Função de ativação tangente hiperbólica

Sumário

	Introdução	16
1	OBJETIVOS	18
1.1	OBJETIVO GERAL	18
1.2	OBJETIVOS ESPECÍFICOS	18
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	SINTERIZAÇÃO DO MINÉRIO DE FERRO	19
2.1.1	Visão geral	19
2.1.2	Matérias-primas	19
2.1.3	Processo	20
2.1.3.1	Dosagem	21
2.1.3.2	Homogeneização	21
2.1.3.3	Granulação (micropelotização)	22
2.1.3.4	Aglomerção a quente	23
2.1.3.4.1	Zona ou mistura Úmida	24
2.1.3.4.2	Zona de secagem	24
2.1.3.4.3	Zona de reação	24
2.1.3.4.4	Zona de resfriamento	30
2.1.3.5	Britagem e resfriamento	30
2.1.4	Qualidade do sinter	30
2.1.4.1	Índice de Resistência Shatter (SI)	30
2.1.4.2	Índice redutibilidade (RI)	31
2.1.4.3	Índice de Degradação Sob Redução (RDI)	32
2.1.4.4	Tamanho Médio de Partícula (MPS)	33
2.2	REGRESSÃO LINEAR MÚLTIPLA	33
2.2.1	Definição	34
2.2.2	Conceitos principais	34
2.2.2.1	Modelo	34
2.2.2.2	Método de solução	34
2.2.3	Multicolinearidade	35
2.2.4	Fator de inflação de variância (FIV)	37
2.2.5	Regressão <i>stepwise</i>	37
2.3	REDES NEURAIS ARTIFICIAIS	37
2.3.1	Breve histórico	37
2.3.2	Aplicações	39

2.3.3	Definição	39
2.3.4	Conceitos principais	40
2.3.4.1	Pré-processamento de dados	40
2.3.4.2	Neurônios	41
2.3.4.3	Funções de ativação (de transferência)	45
2.3.4.4	Índice de aprendizado e sua otimização	48
2.3.4.4.1	Gradiente descendente com <i>momentum</i>	51
2.3.4.4.2	Levenberg–Marquardt	51
2.3.4.5	Treinamento, validação e teste	53
2.3.4.6	Critérios de parada	54
2.3.4.7	Análise pós-treinamento	55
2.3.4.7.1	Coefficiente de determinação (R^2)	57
2.3.4.7.2	Erro percentual absoluto médio (EPAM)	57
2.3.4.7.3	Erro quadrático médio (EQM)	58
2.4	SVD NA BUSCA DO NÚMERO ÓTIMO DE NEURÔNIOS	58
2.4.1	Contexto	58
2.4.2	Conceito e aplicação	59
3	METODOLOGIA	64
3.1	DADOS	64
3.1.1	Aquisição	64
3.1.2	Pré-processamento	64
3.1.3	Estatística descritiva	66
3.2	EXPERIMENTOS COMPUTACIONAIS	68
3.2.1	Regressão linear múltipla	69
3.2.2	MLP - GDM	69
3.2.3	MLP - LM	71
3.2.4	Indicadores de comparação entre experimentos	72
3.2.5	Análise SVD para otimização do número ótimo de neurônios	72
3.2.6	Análise de sensibilidade do melhor modelo	72
4	RESULTADOS E DISCUSSÕES	73
4.1	REGRESSÃO LINEAR MÚLTIPLA	73
4.2	REDES NEURAIS ARTIFICIAIS	80
4.3	COMPARAÇÃO DE RESULTADOS	84
4.3.1	Comparação entre MLP e RLM	84
4.3.2	Comparação com outros autores	86
4.4	OTIMIZAÇÃO DO NÚMERO DE NEURÔNIOS OCULTOS	93
4.5	ANÁLISE DE SENSIBILIDADE	95

5	CONCLUSÃO	103
	REFERÊNCIAS	104
	ANEXOS	113
	ANEXO A – TESTE SHATTER	114
	ANEXO B – TESTE DE REDUTIBILIDADE	115
	ANEXO C – TESTE DE DEGRADAÇÃO SOB REDUÇÃO	116

Introdução

Até a década de 50, a maior parte do material ferroso carregado no alto-forno possuía granulometria entre 10 e 120 mm. O material grosseiro sempre foi indesejável, seja pelo alto esforço mecânico que os equipamentos devem suportar no seu manejo, seja pela ineficiência com que as reações podem ocorrer no seu núcleo. Já os finos, na época, geravam a diminuição da permeabilidade do leito, causando sérios problemas na circulação dos gases e correto funcionamento do forno. Atualmente, devido aos avanços tecnológicos, mais de 70% do conteúdo ferroso de entrada de um alto-forno pode ser composto por sinter formado por finos abaixo de 10 mm (FERNANDEZ-GONZALEZ *et al.*, 2017c). Nesse contexto, mecanismos de aglomeração como a sinterização e a pelotização desempenham papel fundamental.

Além do aspecto granulométrico, é importante ressaltar que o alto-forno também precisa ser carregado com materiais de propriedades mecânicas e metalúrgicas adequadas e homogêneas. Sabe-se que para seu correto funcionamento, um processo estável é a chave para a eficiência econômica e energética. Daí, diz-se que uma planta de sinterização é de extrema importância nas mais modernas siderúrgicas (DAWSON, 1993).

A Tabela 1 mostra que a sinterização é um dos processos de aglomeração que mais se destaca a nível mundial. Em parte, isso se deve ao fato de ser um processo flexível, permitindo alterações no balanço de minério de ferro e qualidade química e física do sinter através da manipulação de entradas como: combustíveis, fundentes e aditivos.

Tabela 1 – Distribuição da carga metálica de um alto-forno em diversas regiões do mundo em 2019.

País/Região	% Sinter	% Pelota	% Granulado
China	78	12	10
Coreia do Sul	70	6	24
Japão	66	12	22
Europa	64	26	10

Fonte: adaptado de Ferrexpo (2020).

Segundo Fernandez-Gonzalez *et al.* (2017a), essencialmente, a sinterização resume-se a um processo térmico de aglomeração com fusão parcial de uma mistura de minérios de ferro, fundentes, aglomerantes, aditivos e combustíveis. Seu objetivo principal é de fabricar

um produto com as características térmicas, mecânicas, físicas e químicas adequadas à alimentação do alto-forno.

A eficiência da sinterização e a qualidade final do sínter são diretamente afetados pela composição química da matéria-prima de entrada do processo, representada basicamente por diversos tipos de *sinters feeds*. Estes, por sua vez, apresentam uma gama de propriedades que afetam importantes parâmetros como: consumo de combustível, resistência mecânica do produto, comportamento metalúrgico e produtividade. Paralelo a isso, variáveis de processo como: velocidade da esteira de alimentação, densidade da carga e temperatura da frente de queima, dentre diversas outras, também possuem seu papel nas propriedades finais do produto da sinterização.

Apesar de haver inúmeros trabalhos que propõem um entendimento geral dos complexos fenômenos envolvidos na sinterização como [Nath, Silva e Chakraborti \(1997\)](#), [Mitterlehner et al. \(2004\)](#), [Yang et al. \(2004\)](#), não é tarefa fácil modelar, em paralelo, reações químicas, passagem de gases pelo leito, trocas de calor entre gases, sólidos e líquidos, fusão, solidificação e processos de aglomeração. Tudo isso, levando em conta um processo real, complexo e instável.

Algumas pesquisas já utilizaram métodos computacionais para prever indicadores de qualidade do sínter como: [Umadevi et al. \(2016\)](#), [Shao et al. \(2020\)](#), [Donskoi et al. \(2007\)](#), [Gao et al. \(2021\)](#), que modelaram um sistema de previsão do Índice de Tamboramento (TI), uma propriedade bastante estudada sobre a resistência do sínter; [Kinnunen e Laitinen \(2005\)](#), que buscaram prever a produtividade e o Índice de Degradação Sob Redução (RDI) simultaneamente, assim como [Laitinen e Saxén \(2007\)](#); [Kumar et al. \(2017\)](#), que usou uma rede neural de camada única e algoritmo de aprendizagem comum para prever o Tamanho Médio de Partícula (MPS), o TI e o RDI, chegando, em dados de treinamento, a valores altos do coeficiente de determinação.

Contudo, apenas uma pequena parcela do conteúdo disponível se utiliza de algoritmos mais complexos e específicos, ou de uma maior quantidade de variáveis de entrada e abrangência de propriedades finais buscadas. Além disso, poucos autores se preocupam com uma análise mais profunda sobre os resultados, buscando discutir as relações e os resultados encontrados.

Dessa forma, o presente trabalho busca desenvolver um sistema complexo de previsão de quatro propriedades físicas e metalúrgicas do sínter: Índice de Resistência Shatter (SI), Tamanho Médio de Partícula (MPS), Índice de Degradação Sob Redução (RDI) e Índice de Redutibilidade (RI), através do processamento de dados com diversos parâmetros de entrada. Os algoritmos testados incluem: Regressão Linear Múltipla (MLR), Regressão *stepwise* (SR) Redes Neurais Artificiais MLP (Multilayer Perceptrons) com algoritmos de retropropagação do tipo Gradiente Descendente com *Momentum* (MLP-GDM) e Levenberg-Marquardt (MLP-LM).

1 OBJETIVOS

1.1 OBJETIVO GERAL

- Desenvolver um modelo de previsão que seja capaz de eficientemente relacionar parâmetros e variáveis do processo de sinterização com quatro indicadores de qualidade desejados: Índice de Resistência Shatter (SI), Índice de Redutibilidade (RI), Índice de Degradação Sob Redução (RDI) e Tamanho Médio de Partícula (MPS).

1.2 OBJETIVOS ESPECÍFICOS

- Avaliar e comparar o desempenho de cada um dos modelos de previsão criados, atentando aos valores do Coeficiente de Determinação (R^2), do Erro Percentual Absoluto Médio (EPAM) e do Erro Quadrático Médio (EQM) para os dados de teste;
- Realizar uma análise de sensibilidade do melhor modelo criado, de modo a verificar a existência de relações entre as variações dos parâmetros de entrada com as variações dos parâmetros de saída a fim de encontrar tendências e padrões que possam ser discutidos a partir de conceitos já definidos na literatura base;
- Utilizar uma metodologia baseada na Decomposição em Valores Singulares (SVD) da matriz da saída das camadas ocultas de uma rede neural previamente treinada para estimar o número ideal de neurônios ocultos.
- Analisar e comparar os resultados obtidos com trabalhos anteriores para as variáveis de saída que forem passíveis de comparação, visto que nem todas são encontradas em outros trabalhos de previsão.

2 FUNDAMENTAÇÃO TEÓRICA

Neste capítulo, aspectos relevantes do tema de dissertação são abordados de forma a conectar todos os conceitos necessários ao completo entendimento do trabalho e sua metodologia.

2.1 SINTERIZAÇÃO DO MINÉRIO DE FERRO

2.1.1 Visão geral

A sinterização do minério de ferro pode ser entendida como um fenômeno de aglomeração a quente de partículas minerais através de uma fusão incompleta, causada pela queima de um combustível sólido, tendo como resultado uma massa porosa. É um processo complexo que envolve diversos fenômenos físicos e químicos. O produto resultante é o sinter, que deve apresentar propriedades físicas, químicas e metalúrgicas compatíveis com as requeridas ao correto funcionamento do alto-forno (BALL, 1973; NAJAR; SANTOS, 1981).

De acordo com Honorato (2005), as principais características exigidas para um sinter de boa qualidade são: composição química estável; elevado teor de ferro; elevada resistência mecânica; granulometria estável e com baixa porcentagem de finos; baixa degradação sob redução e alta redutibilidade.

2.1.2 Matérias-primas

As matérias-primas básicas utilizadas na composição da mistura de sinterização são: finos de minério de ferro (*sinter feed*), combustíveis sólidos, fundentes, aditivos e subprodutos de outros processos.

A matriz de *sinter feeds* geralmente é composta por um conjunto de minérios cujas propriedades são bastante diferentes. Grandes variações químicas, físicas, metalúrgicas e mineralógicas ocorrem dentre os diversos minérios ofertados no mercado siderúrgico. Isso influencia fortemente o desempenho do processo de sinterização e algumas variáveis como: produtividade, consumo de combustível, comportamento metalúrgico e resistência mecânica do produto. Devido a essa diversidade de oferta de minérios no mercado e às suas diferentes características físicas e químicas, as sinterizações industriais adotam como prática a composição de matrizes de sinterização com variados tipos de finos. Esta

prática é bastante favorável ao processo, pois muitas vezes diferentes *sinter feeds* apresentam composições complementares e benéficas ao processo (NAJAR; SANTOS, 1981; FERNANDEZ-GONZALEZ et al., 2017c; CORES et al., 2013).

Os fundentes atuam no processo de sinterização de forma a reduzir a temperatura de fusão dos componentes, facilitando a aglomeração. Além disso, são grandes iniciadores da formação de ferritas de cálcio ($CaO.Fe_2O_3$), estruturas essenciais à resistência mecânica do sínter. Quando chegam ao alto-forno, possuem a função de facilitar a formação de escórias com baixa temperatura de fusão, onde as impurezas presentes nos minérios e demais insumos devem ser eliminadas através da formação de uma escória líquida, com baixa viscosidade e baixa temperatura de fusão (LINHARES, 2019). Segundo Ball (1973), a introdução desses compostos na mistura de sinterização, ao invés de adicioná-los diretamente no alto-forno, trouxe vantagens consideráveis para o desempenho desse reator. Calcário, dolomita, dunita, bauxita e cal são os fundentes comumente utilizados.

Se os fundentes são responsáveis pelo aporte químico do processo, os combustíveis sólidos atuam no aporte térmico, de modo a garantir que todas as reações necessárias à produção do sínter possam ocorrer. Dentre os principais combustíveis disponíveis, destacam-se o antracito e a moinha de coque (ALMEIDA, 2019; FERNANDEZ-GONZALEZ et al., 2017c).

Elementos de liga também podem ser adicionados na forma de óxidos ainda na fase de preparação devido a dois fatores: ou por serem facilmente reduzidos no processo de redução do minério de ferro; ou por ser interessante ao processo de redução. Os mais comumente utilizados são os minérios de manganês (fonte de Mn) e de ilmenita (fonte de TiO_2) (FERNANDEZ-GONZALEZ et al., 2017c).

Subprodutos (*by-products*) de outros processos da indústria siderúrgica também são utilizados na sinterização. A reciclagem desses subprodutos já vem sendo incentivada a vários anos como uma consequência de novas políticas ambientais, necessidade de redução do consumo energético e tentativa de reutilizar materiais descartados com alto teor de ferro. A maioria dos subprodutos sólidos podem ser reempregados no processo de sinterização enquanto alguns, como a escória de alto forno, são destinados à indústria cimenteira (FERNANDEZ-GONZALEZ et al., 2017c).

2.1.3 Processo

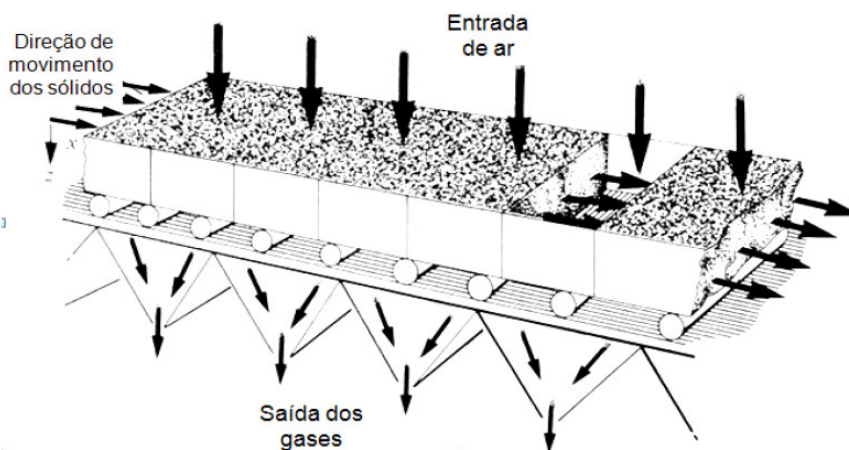
Existem dois tipos principais de sinterização: sinterização por batelada, conhecida como *Greenwalt*, que é mais antiga, necessita de baixo investimento, gera um sínter de pior qualidade e de baixa produtividade, e geralmente é utilizada em pequenas empresas; e a sinterização contínua (estudada no presente trabalho), conhecida como *Dwight & Lloyd*, que apresenta maior produtividade, necessita de maiores investimentos e é, em geral,

aplicada em produções em larga escala (PEREIRA, 2013; FERNANDEZ-GONZALEZ et al., 2017a; CORES et al., 2013).

A Figura 1 ilustra o princípio básico da sinterização contínua: ignição da superfície do leito, combustão do coque, fusão e aglomeração a quente de partículas. A mistura a ser sinterizada forma um leito, que se desloca horizontalmente, enquanto o processo de queima é feito pela reação entre o ar que atravessa a mistura e o combustível sólido, de cima para baixo (CASTRO; FIGUEIRA; TAVARES, 1985).

A operação pode ser dividida em seis etapas: dosagem, homogeneização, granulação, aglomeração a quente, britagem e resfriamento. Cada uma será abordada com mais detalhes a seguir.

Figura 1 – Ilustração do processo Dwight & Lloyd



Fonte: CASTRO, FIGUEIRA e TAVARES (1985)

2.1.3.1 Dosagem

O processo de sinterização se inicia com a dosagem de materiais, etapa de carregamento em proporções definidas das matérias primas em uma correia transportadora, formando a mistura de sinterização (*blend*). O *blend* é enviado a um processo de mistura que pode ser dividido em duas etapas: homogeneização e granulação (micropelotização).

2.1.3.2 Homogeneização

A primeira etapa, a de homogeneização, ocorre em um misturador primário (tambor rotativo) munido de ressaltos salientes, onde o *blend* é rotacionado em velocidade e inclinação pré-estabelecidas, de modo a formar um composto mais homogêneo e levemente aglomerado.

2.1.3.3 Granulação (micropelotização)

Na segunda etapa, a de granulação, o material é encaminhado a um misturador secundário com ressaltos menos salientes, onde também ocorre rotação, porém há a adição controlada de pequenas quantidades de água.

Isso resulta no processo de micropelotização, também denominado de aglomeração a frio, que dá origem aos grânulos verdes ou partículas-quasi (*quasi – particles*), usualmente conhecidos como micropelotas. Segundo os estudos de [Litster e Waters \(1988\)](#), o principal mecanismo do processo de granulação é dado pela aderência de partículas finas (aderentes) em partículas maiores (núcleos), sendo as forças de ligação principais as forças capilares, devido à presença de água na camada de finos.

O grânulo verde (micropelota) mostrado na Figura 2, pode ser definido como um núcleo de minério de ferro (que permanece sólido mesmo após a completa sinterização) circundado por finos grãos de minério e ganga de sílica. Partículas com tamanho superior a 0.7 mm desempenham o papel de núcleos, enquanto partículas com tamanho inferior a 0.2 mm atuam como finos aderentes. A parcela de partículas com dimensão entre 0.2 e 0.7 mm deve ser mínima já que afeta a permeabilidade da mistura de duas maneiras diferentes, dependendo do papel que podem desempenhar: sendo núcleos dão origem a uma partícula-quasi (micropelota) de menor tamanho, diminuindo a permeabilidade do leito; sendo finos aderentes são fracamente ligadas ao núcleo e facilmente separáveis, o que também compromete a permeabilidade do leito e resistência do sínter ([FERNANDEZ-GONZALEZ et al., 2017c](#); [CORES et al., 2013](#); [DAWSON, 1993](#)).

Figura 2 – Esquema simplificado de uma partícula-quasi (micropelota)



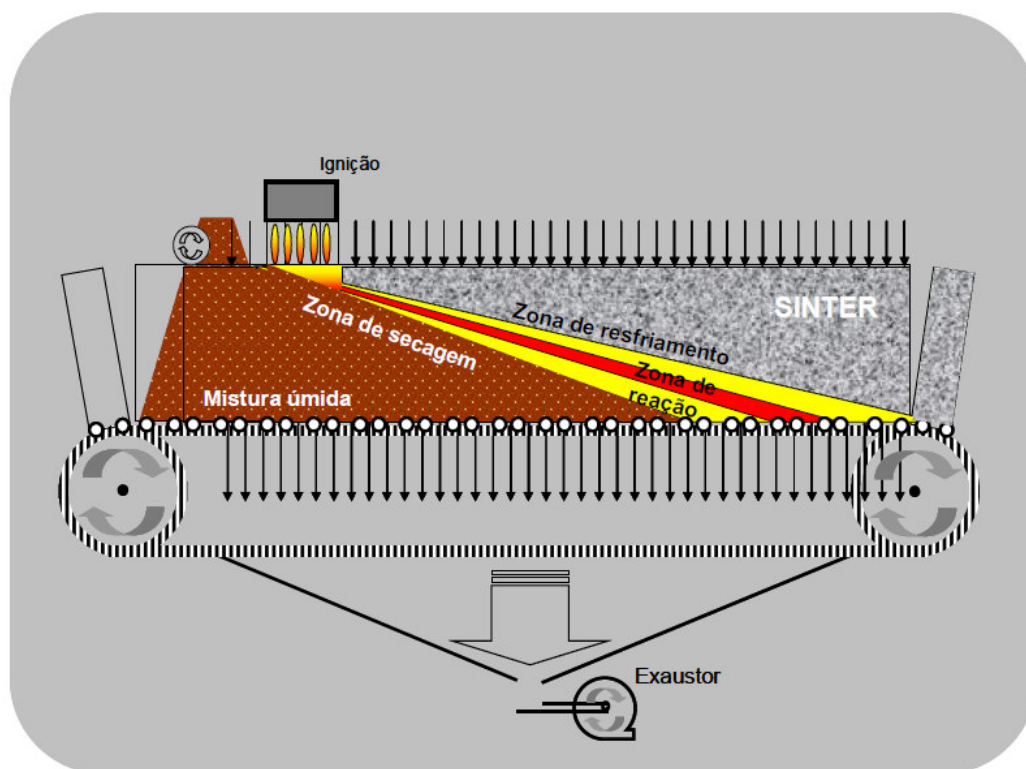
Fonte: adaptado de [Fernandez-Gonzalez et al. \(2017c\)](#)

2.1.3.4 Aglomeração a quente

Nessa etapa, o material finalmente chega à máquina principal de sinterização, ilustrada anteriormente pela Figura 1. De acordo com [CASTRO, FIGUEIRA e TAVARES \(1985\)](#), o mecanismo do processo é determinado por dois fenômenos distintos: um fenômeno físico de transferência de calor, da camada superior para a inferior, proporcionada pela sucção forçada de ar formando uma frente térmica; e um fenômeno químico de combustão do carbono contido na mistura a sinterizar que, produzindo calor, gera uma frente de reações químicas.

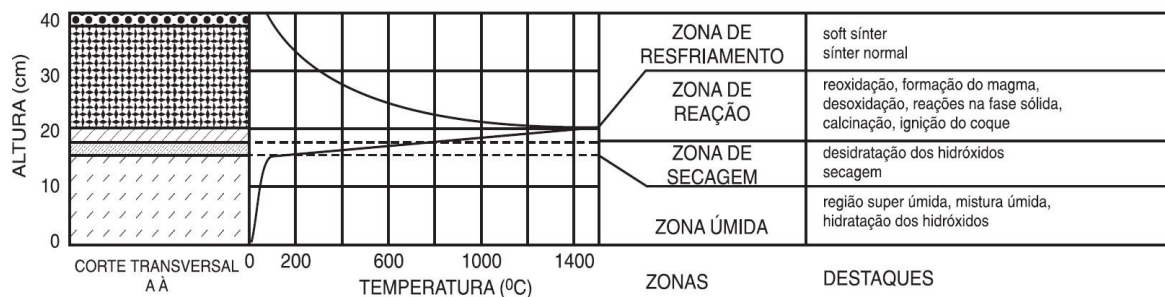
Estes fenômenos são interdependentes e definem as zonas de sinterização e o desenvolvimento do processo. O equilíbrio entre os dois determina a qualidade da queima, o desempenho, o consumo de combustível e a qualidade do produto. Existem quatro zonas principais nesse processo: úmida, de secagem, de reação e de resfriamento, representados nas Figuras 3 e 4.

Figura 3 – Representação esquemática da máquina de sinterização e localização das zonas principais do processo



Fonte: [Starling \(2012\)](#)

Figura 4 – Corte transversal e perfil térmico de um leito de sinterização



Fonte: adaptado de Honorato (2005), NAJAR e SANTOS (1981)

2.1.3.4.1 Zona ou mistura Úmida

É uma região representada pelo leito de mistura úmida. Nela, a mistura encontra-se conforme saiu da etapa de granulação. A temperatura máxima atingida é de 100° C e as principais reações que ocorrem são a hidratação da cal e a condensação dos vapores de água (NAJAR; SANTOS, 1981).

2.1.3.4.2 Zona de secagem

Nesta zona acontece a vaporização da umidade da mistura e a desidratação dos hidróxidos, com a temperatura variando de 100°C a 500°C. As principais reações que ocorrem nesta etapa são a desidratação do hidróxido de cálcio e a vaporização da umidade (NAJAR; SANTOS, 1981).

2.1.3.4.3 Zona de reação

Esta zona representa toda a região térmica de aquecimento acima de 500°C. Compreende a ignição do carbono até o início do resfriamento do sinter, ocorrendo simultaneamente reações químicas e reações aglomerantes. Dentre elas, destacam-se: combustão do carbono, decomposição dos carbonatos, reações na fase sólida, redução e reoxidação dos óxidos de ferro, reações de formação da escória e reações de aglomeração (NAJAR; SANTOS, 1981; STARLING, 2012). As reações de combustão do combustível carbonoso levam a formação de CO e CO_2 e são responsáveis pela elevação da temperatura de 500°C a 1400°C.

Já as reações de decomposição de carbonatos resultam na formação da cal, já que grande parte do carbonato usado na sinterização é calcário. Essas reações ocorrem a partir de 800°C, absorvendo e baixando a temperatura dos gases.

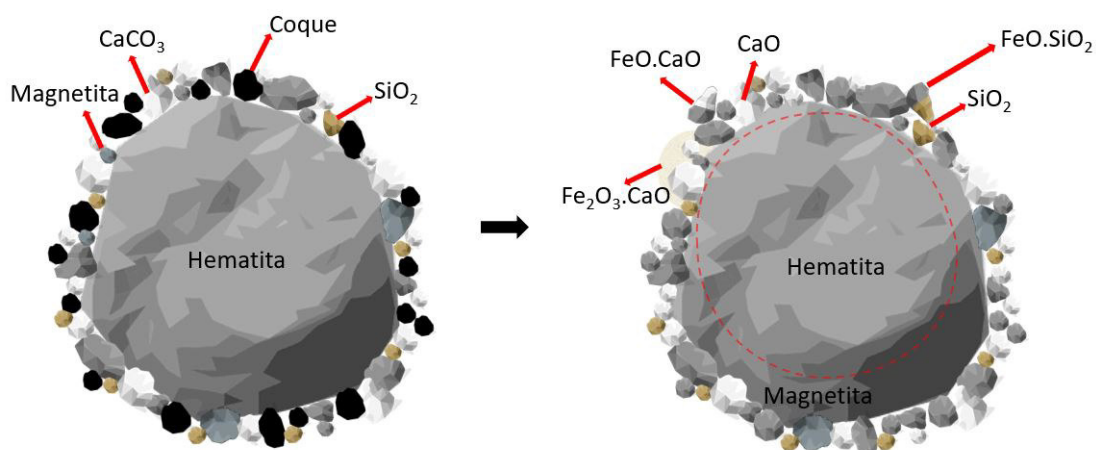
Durante o aquecimento da mistura na máquina de sinter, o contato entre as partículas de óxido de ferro, CaO , SiO_2 e outros, favorece o desenvolvimento de algumas reações no estado sólido, propiciando a formação de ferritas de cálcio e silicatos diversos.

As reações de redução e reoxidação do minério de ferro ocorrem em praticamente todo o processo, enquanto reações de formação de escória ocorrem na frente de combustão, gerando silicatos.

As reações de aglomeração a quente ocorrem em temperaturas maiores que a de fusão dos materiais e compõem uma mistura de reações entre sólido-líquido e líquido-líquido. Merecem atenção especial pois são muito importantes na definição da microestrutura final do sinter, que por sua vez é fundamental em diversas propriedades como resistência e redutibilidade (DAWSON, 1993; LU; ISHIYAMA, 2015).

Esse processo de aglomeração a quente se inicia com particulados de hematita (Fe_2O_3) da camada aderente do *sinter feed* reagindo com as partículas de cal (CaO), seja a cal adicionada ou a resultante da calcinação de calcários, em temperaturas na ordem de $1100^\circ C$, formando as primeiras ferritas de cálcio ($Fe_2O_3.CaO$), conforme Figura 5 (DAWSON, 1993).

Figura 5 – Representação do processo de formação da microestrutura do sinter - Etapa de formação das primeiras ferritas de cálcio

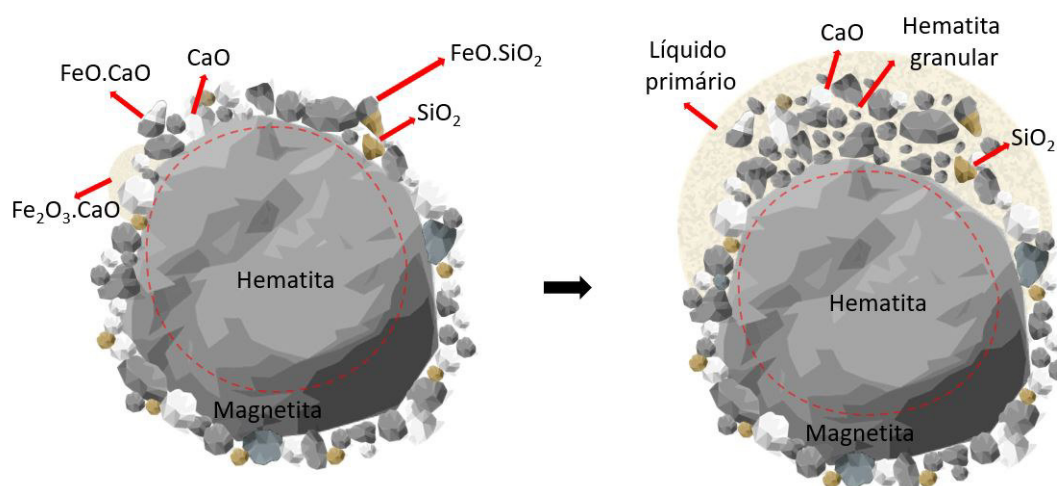


Fonte: elaborado pelo próprio autor.

A partir de $1200^\circ C$ inicia-se a formação de um líquido chamado de primário que é constituído basicamente de Fe_2O_3 e CaO . Em seguida, esse líquido passa a assimilar óxidos de ferro juntamente de outros óxidos presentes na camada aderente do aglomerado, isto é, SiO_2 , Al_2O_3 e MgO . Nesse momento, comumente ocorre a desagregação superficial das partículas policristalinas do núcleo hematítico em função da penetração do líquido

primário nos contornos de grão, formando a hematita granular, isto é, desagregados de hematita, de acordo com a Figura 6 (DAWSON, 1993; FERNANDEZ-GONZALEZ et al., 2017a; CORES et al., 2013).

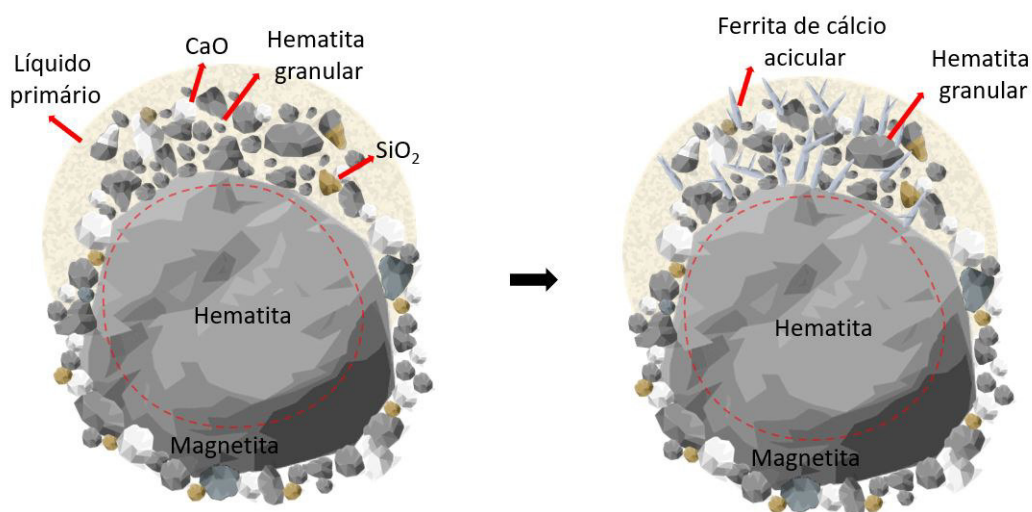
Figura 6 – Representação do processo de formação da microestrutura do sinter - Etapa de formação do líquido primário



Fonte: elaborado pelo próprio autor.

Após o líquido ter assimilado parcialmente o CaO e a Al_2O_3 , ele passa a reagir com o óxido de ferro produzindo uma estrutura de ferrita de cálcio acicular sólida (em forma de agulhas), rica em Al_2O_3 e SiO_2 , conforme Figura 7 (DAWSON, 1993; PADULA, 2009).

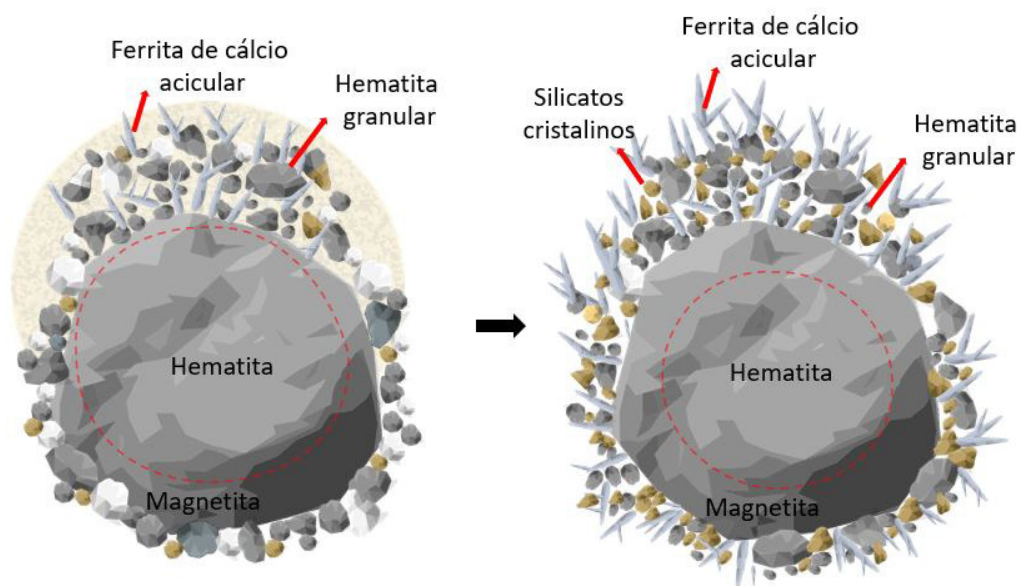
Figura 7 – Representação do processo de formação da microestrutura do sinter - Etapa de formação das ferritas de cálcio



Fonte: elaborado pelo próprio autor.

Se a temperatura máxima do perfil de queima não exceder o valor de 1300°C, ou se o tempo de retenção for curto, a microestrutura resultante será formada por um aglomerado com elevada proporção de ferrita de cálcio acicular em meio a silicatos cristalinos e partículas de hematita granular, estando toda a matriz solidificada ao redor da partícula de hematita nucleante, assim como no esquema da Figura 8. Esse produto é então chamado de sinter heterogêneo e apresenta satisfatórias propriedades metalúrgicas para uso nos altos-fornos (DAWSON, 1993; PADULA, 2009).

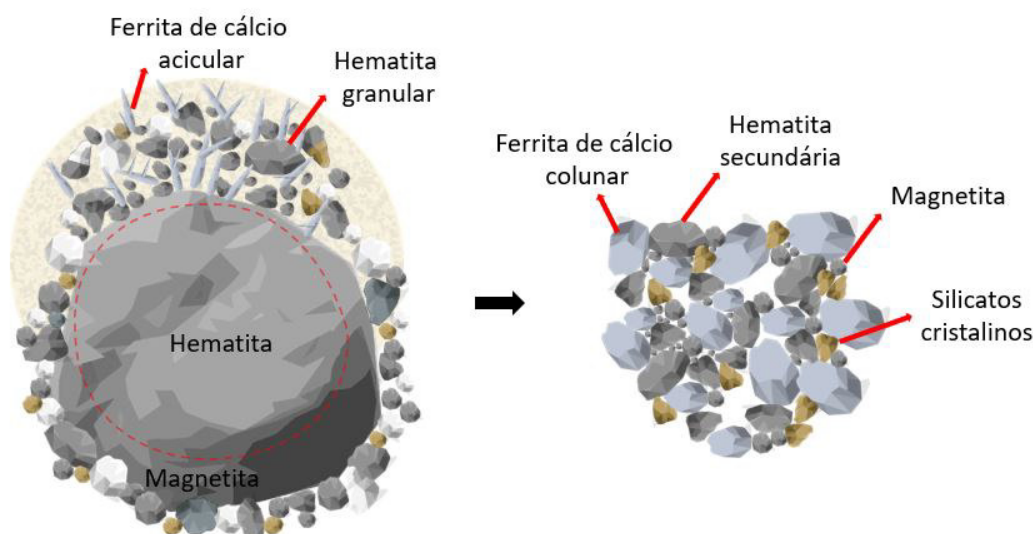
Figura 8 – Representação do processo de formação da microestrutura do sinter - Etapa de formação da hematina granular



Fonte: elaborado pelo próprio autor.

Contudo, se a temperatura máxima do perfil térmico exceder 1300°C , a ferrita de cálcio e os nucleantes de hematita se fundem e dissolvem, formando hematita secundária ou magnetita e uma escória líquida. Ao resfriar, forma-se um tipo de hematita secundária romboédrica esqueletiforme, que torna o sinter suscetível à desintegração durante redução. As ferritas de cálcio formadas são do tipo colunares, conforme Figura 9. Esse aglomerado é chamado de sinter homogêneo e não apresenta as propriedades metalúrgicas adequadas ao uso nos altos-fornos (DAWSON, 1993; PADULA, 2009).

Figura 9 – Representação do processo de formação da microestrutura do sinter - Etapa de formação da hematita secundária



Fonte: elaborado pelo próprio autor.

Muitas das fases ricas em ferritas de cálcio contêm alguma quantidade de sílica (SiO_2) e alumina (Al_2O_3) e são geralmente mencionadas como sílico-ferritas de cálcio e alumínio (SFCA). Existem duas principais: SFCA e SFCA-I. A SFCA é a de baixo teor de óxidos de ferro, exibindo forma colunar/prismática (formada em temperaturas maiores que $1300^\circ C$), sendo uma fase geralmente indesejada. Já a que mais se busca é a SFCA-I, de alto teor de óxidos de ferro e que possui morfologia acicular (formada em temperaturas menores que $1300^\circ C$) (WEBSTER et al., 2012; FERNANDEZ-GONZALEZ et al., 2017a).

A fase SFCA-I é a mais desejável para fins de ligação no aglomerado de sinter devido sua alta redutibilidade, alta resistência mecânica e baixa degradação sob redução. Todas essas características são fatores significantes na determinação da produtividade e eficiência do alto-forno. Contudo, é importante ressaltar que as temperaturas e tempos exatos de formação dessas fases dependem de muitos fatores complexos, desde os constituintes das matérias-primas até os parâmetros de processamento durante a aglomeração (FERNANDEZ-GONZALEZ et al., 2017a).

Também existem outras fases de ferrita no sinter metalúrgico: ferrita monocálcica ($CaO.Fe_2O_3$, abreviado para CF), ferrita dicálcica ($2CaO.Fe_2O_3$, abreviado para C_2F) e ferrita hemicálcica ($CaO.2Fe_2O_3$, abreviado para CF_2), etc (WEBSTER et al., 2012).

2.1.3.4.4 Zona de resfriamento

Começa abaixo de 900°C, quando se encerram as reações. Nesta fase do processo tem lugar o resfriamento do sínter, onde aparecem os primeiros cristais que, conforme o seu desenvolvimento, irão conferir determinadas propriedades ao sínter (FERNANDEZ-GONZALEZ et al., 2017a).

2.1.3.5 Britagem e resfriamento

A última etapa corresponde a um processo de adequação dimensional para correto carregamento e manuseio no forno de redução. Utilizam-se britadores para reduzir o tamanho e sopradores para auxiliar na queda de temperatura e facilitar o transporte por correias (ALMEIDA, 2019).

2.1.4 Qualidade do sínter

2.1.4.1 Índice de Resistência Shatter (SI)

A degradação do sínter durante seu transporte pode reduzir significativamente a quantidade de material ferroso disponível para carregamento, enquanto sua deterioração durante as operações iniciais do alto-forno produz finos que prejudicam a permeabilidade do leito (LU; PAN; ZHU, 2015; FERNANDEZ-GONZALEZ et al., 2017b).

Dois mecanismos são responsáveis pela degradação mecânica do sínter, um relacionado ao impacto e outro relacionado à abrasão. Se uma partícula individual for incapaz de acomodar a energia devido ao impacto durante o carregamento, bem como à abrasão entre as partículas e entre as partículas e a parede do forno, essa partícula individual provavelmente se quebrará para gerar finos. Esses finos preenchem os vazios no material de carga compactada e reduzem a permeabilidade do eixo do forno, reduzindo sua produtividade (LU; PAN; ZHU, 2015).

A resistência do material ferroso à degradação mecânica é determinada principalmente pela mineralogia e textura do material envolvido (LU; PAN; ZHU, 2015). Dukino et al. (1995) mediu a tenacidade de granulados usando uma combinação de testes de fratura e concluiu que as fases mais duras e com contornos de grãos mais resistentes produziam menos finos, sendo mais resistentes à degradação.

Existem diversos testes e índices que medem a resistência física do sínter e verificam se ele é adequado às operações no alto-forno, dentre eles, os mais comuns: índice Shatter (SI), índice de tamboramento (TI) e o índice abrasivo (AI).

A indústria que forneceu os dados do presente trabalho utiliza o índice Shatter (SI) para verificar as propriedades de resistência física do produto da sinterização. Esse índice está baseado no teste de uma norma japonesa, a [JIS-M8711:2011](#), similar à norma [ISO-3271:2015](#), utilizada no índice de tamboramento (TI). A principal diferença entre eles é que o teste Shatter considera a geração de finos provenientes de sucessivas quedas de uma amostra, enquanto o teste de tamboramento verifica a geração de finos provenientes de um tambor rotativo.

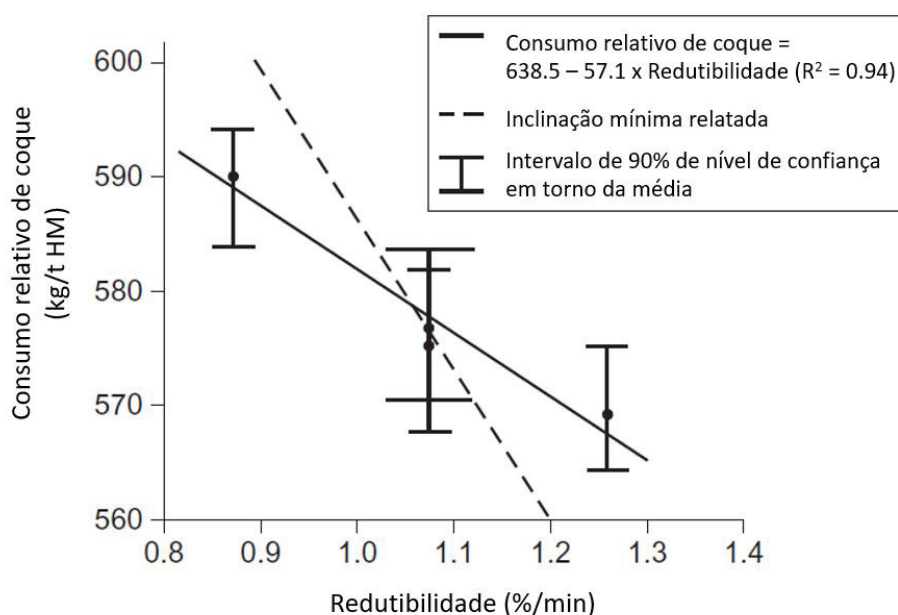
O Anexo A mostra mais detalhes da metodologia desse teste.

2.1.4.2 Índice redutibilidade (RI)

Dois tipos distintos de reações de redução ocorrem simultaneamente no alto-forno. Na parte superior próximo ao carregamento de matéria prima, os óxidos de ferro são reduzidos a wustita (FeO) pelos gases redutores, incluindo CO e H_2 , na chamada reação indireta de redução, enquanto a redução adicional para ferro metálico (Fe) ocorre principalmente em contato com combustível sólido ou coque em regiões mais internas, na chamada reação direta de redução (LU; ISHIYAMA, 2015).

A redutibilidade é importante para o alto-forno particularmente na parte inferior da zona granular. Ela pode afetar a utilização do potencial de redução de CO e H_2 no poço do forno e, portanto, o consumo de coque, conforme evidenciado na Figura 10 (LU; PAN; ZHU, 2015).

Figura 10 – Efeito da redutibilidade da carga no consumo de coque de alto-forno



Fonte: adaptado de Blattner, Ranade e Ricketts (1984), Lu, Pan e Zhu (2015)

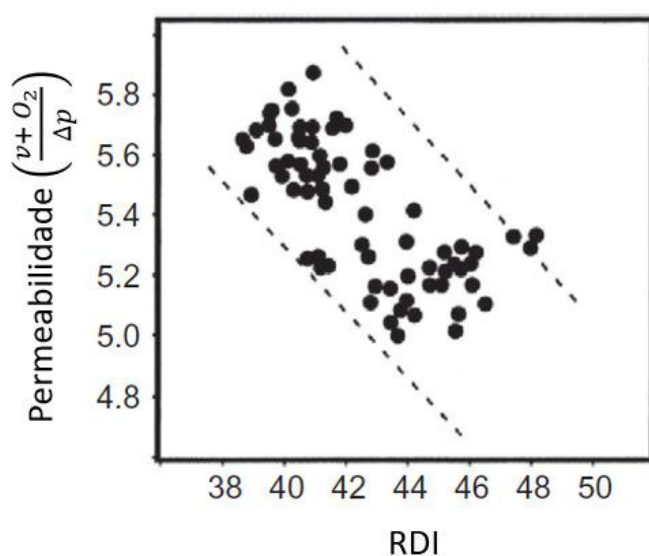
Um minério altamente redutível também resultará em menos produção de FeO à medida que desce para as zonas de mais altas temperaturas e, portanto, melhorará as propriedades de amolecimento do material da carga. Materiais ferrosos com alta redutibilidade são, portanto, desejáveis para uma operação eficiente do alto-forno. Durante a redução indireta, os óxidos de ferro são reduzidos por meio de um mecanismo de reação sólido/gás. Em temperaturas intermediárias, a taxa de redução é controlada por difusão de gases mistos e reação química na superfície dos poros do sólido. Dessa forma, a porosidade da fase sólida afeta a cinética da reação, uma vez que altera a área de superfície disponível para as reações químicas (LU; PAN; ZHU, 2015; FERNANDEZ-GONZALEZ et al., 2017b).

Os testes mais comuns para verificação de RI estão baseados nas normas ISO-7215:2015, ISO-4695:2015 e ISO-7992:2015. Mais detalhes dos métodos podem ser encontrados no Anexo B.

2.1.4.3 Índice de Degradação Sob Redução (RDI)

O RDI é projetado para avaliar o grau de geração de finos de material ferroso quando reduzido em condições semelhantes às prevalentes na zona de redução de baixa temperatura de um alto-forno. Conforme mostrado na Figura 11, a permeabilidade do alto-forno melhora consideravelmente conforme o valor de RDI do material ferroso diminui (LU; PAN; ZHU, 2015).

Figura 11 – Efeito da qualidade da carga ferrosa na permeabilidade do alto-forno



Fonte: adaptado de Nishio et al. (1982), Lu, Pan e Zhu (2015)

A desintegração de material ferroso é geralmente associada à iniciação e propagação de trincas devido à mudança de volume que acompanha a redução de hematita em magnetita e à incapacidade dos grãos de hematita ou da própria matriz de sinterização em acomodar as deformações e tensões internas resultantes de mudanças no volume e na estrutura cristalina. As trincas são formadas inicialmente nos contornos de grãos e se propagam pela matriz, levando à desintegração das partículas (LU; PAN; ZHU, 2015; FERNANDEZ-GONZALEZ et al., 2017b).

Os testes para medição desse índice baseiam-se nas normas ISO-4696-1:2015 e ISO-4696-2:2015. Maiores detalhes dessa técnica estão expostos no Anexo C.

2.1.4.4 Tamanho Médio de Partícula (MPS)

A produtividade da máquina de sinterização está relacionada diretamente com uma menor geração de finos do processo, já que uma alta parcela de finos indica uma baixa parcela de material corretamente aglomerado. Assim, uma redução na geração de finos significa um aumento na resistência do produto da sinterização, consequência de uma correta utilização da velocidade de esteira e porcentagem de ferro (FERNANDEZ-GONZALEZ et al., 2017b).

Segundo NAJAR e SANTOS (1981), o Tamanho Médio de Partícula (MPS) é um importante indicador de uma correta aglomeração do sínter. Assim como o Índice de Resistência Shatter (SI) e o Índice de Degradação Sob Redução (RDI), também está relacionado com a resistência mecânica à geração de finos.

Segundo Kumar et al. (2017), o índice MPS é calculado pela Equação 2.1, onde W1, W2, W3, W4, W5, W6, W7 e W8 correspondem às frações de partículas com tamanhos, respectivamente, de +50mm, entre 50mm e 40mm, entre 40mm e 30mm, entre 30mm e 20mm, entre 20mm e 16mm, entre 16mm e 10mm, entre 10mm e 5mm, e -5mm.

$$MPS = \frac{(55W_1 + 45W_2 + 35W_3 + 25W_4 + 18W_5 + 13W_6 + 7.5W_7 + 2.5W_8)}{100} \quad (2.1)$$

2.2 REGRESSÃO LINEAR MÚLTIPLA

A regressão linear múltipla (RLM) é usada no presente trabalho como forma de previsão inicial e de referência básica de índices de previsão. Entende-se que esse tipo de modelo é simples e de fácil interpretação, mas que na maior parte das aplicações em problemas reais e complexos, principalmente no âmbito siderúrgico/metalúrgico, possui algumas limitações.

2.2.1 Definição

A análise de regressão é uma técnica estatística para investigar e modelar a relação entre variáveis. As aplicações da regressão são numerosas e ocorrem em quase todos os campos, incluindo engenharia, ciências físicas e químicas, economia, administração, ciências biológicas e da vida e ciências sociais. Na verdade, a análise de regressão pode ser a técnica estatística mais amplamente usada (MONTGOMERY; PECK; VINING, 2012).

A RLM é um tipo de aproximação na qual os coeficientes de regressão são os termos buscados para que uma função linear seja ajustada no espaço multidimensional das variáveis em estudo, dessa forma, possuindo um significado prático/matemático: ajuste de uma reta através de pontos em um espaço n -dimensional (PITOMBO; COSTA, 2015; DEVORE, 2019).

2.2.2 Conceitos principais

2.2.2.1 Modelo

O objetivo principal desse método é estabelecer uma equação linear que possa ser utilizada para prever valores de y_i - variável dependente, a partir de variáveis independentes x_i , sendo i o coeficiente para o i -ésimo termo de um conjunto de n dados. Segundo Montgomery, Peck e Vining (2012), o modelo de regressão linear múltipla com k variáveis de regressão pode ser representado pela Equação 2.2 para $i = 1, 2, \dots, n$:

$$y_i = \beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \dots + \beta_k x_{i,k} + \epsilon_i \quad (2.2)$$

Onde β_0 representa a constante de regressão, $x_{i,j}$ representa o valor da i -ésima observação da j -ésima variável independente, de um total de n dados com k variáveis independentes, β_j representa o valor do coeficiente de regressão associado à j -ésima variável independente, y_i representa o valor da variável dependente analisada na i -ésima observação, e ϵ_i um valor aleatório (ou resíduo) do erro associado a equação no i -ésimo exemplo.

2.2.2.2 Método de solução

A estimação dos coeficientes β_j é a etapa principal da RLM, já que com esses coeficientes é possível estimar qualquer novo valor de saída a partir de qualquer novo exemplo de entrada. A solução mais comum tem como base o método dos mínimos quadrados, onde busca-se uma forma de minimizar a soma dos quadrados das diferenças entre os valores estimados e os valores reais – essa diferença também pode ser chamada de resíduo

(MONTGOMERY; PECK; VINING, 2012; DEVORE, 2019; BRAGA, 2019). A Equação 2.3 mostra o resultado do método no formato matricial.

$$\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} \quad (2.3)$$

Onde as matrizes $\boldsymbol{\beta}$, \mathbf{X} e \mathbf{Y} são definidas, respectivamente, pelas Equações 2.4, 2.5 e 2.6.

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} \quad (2.4)$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n,1} & x_{n,2} & \cdots & x_{n,k} \end{bmatrix} \quad (2.5)$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad (2.6)$$

2.2.3 Multicolinearidade

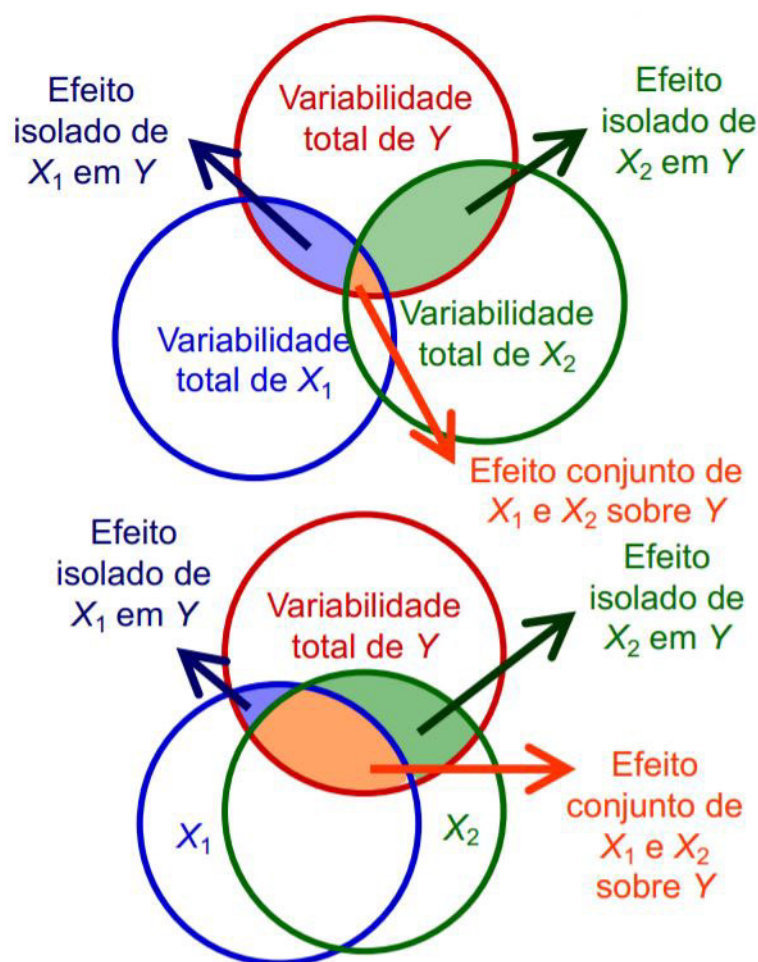
Segundo Salvian (2016), multicolinearidade é uma característica de um conjunto de dados que não apenas apresenta correlação entre as variáveis preditoras e a variável resposta, mas também apresenta forte correlação entre duas ou mais variáveis independentes, de tal forma que há redundância de informação (MINITAB, 2019).

A Figura 12 mostra graficamente a influência da multicolinearidade nos dados. Na imagem superior, observa-se os efeitos isolados predominantes de cada uma de duas possíveis variáveis preditoras sobre a variável dependente buscada. Já na imagem inferior observa-se uma mistura maior da influência das variáveis em conjunto sobre a variável dependente.

A existência de uma colinearidade exata entre duas ou mais variáveis independentes torna difícil a obtenção dos coeficientes dos parâmetros pelo método dos mínimos quadrados. O problema é que, no geral, a multicolinearidade torna as estimativas dos coeficientes de regressão insignificantes, já que cada coeficiente pressupõe, por definição,

a variação em y dada uma variação unitária em x , mantendo-se constantes as demais informações. Uma vez que, se duas variáveis independentes são fortemente correlacionadas, será muito difícil haver variação em uma sem que haja em outra, o que prejudica a RLM (MAIA, 2019; SALVIAN, 2016).

Figura 12 – Efeito da multicolinearidade



Fonte: adaptado de Maia (2019)

Para verificação da multicolinearidade deve-se, em primeiro lugar, analisar a matriz das correlações, que espelha as dependências lineares de primeira ordem entre as variáveis independentes, com atenção especial para resultados superiores a 0,70. Como também é possível ocorrer multicolinearidade, mesmo quando a matriz de correlação apresenta coeficientes de valor baixo, recomenda-se, também, verificar a correlação de cada variável com subconjuntos de outras variáveis independentes, por meio do cálculo do fator de inflação de variância. Realizar outros tipos de regressão como a regressão *stepwise* igualmente pode contornar o problema. (BIAGGI; MEDVID; ASSIS, 2017).

2.2.4 Fator de inflação de variância (FIV)

O fator de inflação da variância (FIV) avalia o quanto a variância de um coeficiente de regressão estimado aumenta se as suas preditoras estiverem correlacionadas. Se o FIV for igual a 1 não há multicolinearidade entre os fatores, mas se o FIV for maior que 1, as preditoras podem estar moderadamente correlacionadas. Um FIV entre 5 e 10 indica alta correlação, o que pode ser problemático. E se o FIV for acima de 10, pode-se assumir que os coeficientes de regressão estão mal estimados devido à multicolinearidade (BIAGGI; MEDVID; ASSIS, 2017; MINITAB, 2019).

Segundo Gujarati e Porter (2011), o FIV poder ser calculado como:

$$FIV_j = -\frac{1}{1 - R_j^2} \quad (2.7)$$

Onde FIV_j representa o fator de inflação de variância e R_j^2 o coeficiente de determinação parcial da variável x_j em relação às demais variáveis independentes.

2.2.5 Regressão *stepwise*

A regressão *stepwise* é um método sistemático para adicionar ou remover termos de um modelo linear com base em um indicador de validação do modelo, por exemplo, o coeficiente de determinação. O método começa com um modelo inicial, geralmente apenas com uma constante e sem nenhuma variável, e então compara a qualidade da regressão de diversos outros modelos, adicionando ou retirando variáveis que melhoram a qualidade da aproximação (ABBAD; TORRES, 2002).

Enquanto o FIV calcula, através de regressões auxiliares, a influência de variáveis independentes sobre as outras, a regressão *stepwise* atenta-se apenas em verificar de forma sistemática quais das variáveis independentes fornecem os melhores resultados de previsão.

2.3 REDES NEURAIS ARTIFICIAIS

2.3.1 Breve histórico

A ideia precursora das redes neurais artificiais tem início no modelo do funcionamento dos neurônios do cérebro humano, conceito esse chamado de "conexionismo". Essa abordagem foi pioneiramente elaborada em 1943, pelo neurofisiologista Warren McCulloch e pelo matemático Walter Pitts no artigo "*A logical calculus of the ideas immanent*

in nervous activity", quando, com simples circuitos elétricos, demonstraram que estruturas semelhantes a neurônios (ou unidades, como eram chamadas) que agem e interagem baseados em princípios neurofisiologicamente plausíveis, poderiam ser conectadas e, assim, receber a capacidade de realizar cálculos lógicos complexos (MCCULLOCH; PITTS, 1943). Eles começaram observando que a atividade dos neurônios tem um caráter "all-or-none" - ou seja, ou os neurônios estão mandando impulsos eletroquímicos em suas projeções longas (axônios) em direção a junções com outros neurônios (sinapses), ou estão inativos. Também observaram que, para se tornarem ativos, a quantidade de influência excitatória de outros neurônios deve atingir um certo limite e que alguns neurônios devem inibir outros. Esses princípios podem ser descritos por formalismos matemáticos, que permitiram o cálculo dos comportamentos de desdobramento de redes que obedecem a tais princípios (KRÖSE; SMAGT, 1996; BECHTEL; ABRAHAMSEN, 2002).

O próximo grande passo na pesquisa conexionista veio com a proposta do neurofisiologista Donald Hebb, em 1949, com o livro *The Organization of Behavior: A Neuropsychological Theory*, dizendo que a conexão entre dois neurônios biológicos é fortalecida, isto é, o neurônio pré-sináptico passará a ter uma influência excitatória ainda mais forte, quando ambos os neurônios estiverem simultaneamente ativos. A partir daí, ficou-se conhecida a expressão "neurônios que atuam juntos, conectam-se" (HEBB, 1949).

No final de 1959, quando os computadores já possuíam maior capacidade de processamento, Bernard Widrow e Marcian Hoff desenvolveram os modelos chamados de ADALINE e MADALINE (WIDROW; HOFF, 1960). ADALINE foi desenvolvido para reconhecer padrões binários de forma que, se estivesse lendo bits de *streaming* de uma linha telefônica, pudesse prever o próximo bit. MADALINE foi a primeira rede neural aplicada a um problema do mundo real, usando um filtro adaptativo que eliminava ecos nas linhas telefônicas e que até hoje possui uso comercial (WIDROW, 1960; WIDROW, 1962).

Contudo, nos anos posteriores de pesquisa ocorreu o chamado *Inverno da IA* (Inteligência Artificial), tradução literal do inglês *AI Winter* (HENDLER, 2008). Com o sucesso inicial dos trabalhos na área, o potencial das redes neurais foi superestimado, especialmente quando levado em conta os aspectos práticos tecnológicos da época. Expectativas não foram atendidas e questões filosóficas maiores foram surgindo. Os escritores ponderaram o efeito que as chamadas "máquinas pensantes" teriam sobre os humanos, ideias que existem ainda hoje. Além disso, o livro *Perceptrons: An introduction to computational geometry* de Minsky e Papert, em 1969, mostrou sérias limitações nos modelos neurais até então sugeridos, o que levou aos investimentos serem redirecionados e aos pesquisadores deixarem o campo de pesquisa (MINSKY; PAPERT, 1969).

Somente quando Jon Hopfield, em 1982, publicou o artigo *Neural networks and physical systems with emergent collective computational abilities*, que depois veio a ser

conhecido como Rede de Hopfield, e com anúncios do Japão na *Cooperative/Competitive Neural Networks Conference* de começar os esforços na quinta geração de redes neurais, é que os trabalhos na área foram retomados (HOPFIELD, 1982). Logo em seguida, por receio de ficar atrás nas pesquisas, o Instituto Americano de Física, em 1985, estabeleceu uma reunião anual *Neural Networks in Computing* seguida pela primeira Conferência Internacional sobre Redes Neurais pelo *Institute of Electrical and Electronic Engineers* (IEEE) em 1987 (ANDERSON, 1995).

Em 1986, com as redes neurais de múltiplas camadas em evidência, o grande problema era como estender os conceitos de Widrow e Hoff para múltiplas camadas. Três grupos independentes de pesquisadores tiveram ideias semelhantes que agora são chamadas de redes de retropropagação, traduzida de forma literal do inglês *backpropagation networks*, porque distribuem os erros de interação ao longo de toda a rede (WERBOS, 1988b; WERBOS, 1988a; WERBOS, 1990; PINEDA, 1988; RUMELHART; HINTON; WILLIAMS, 1986). A partir de então, nos anos 90, diversos trabalhos foram publicados e as redes neurais retomaram o destaque no meio científico.

2.3.2 Aplicações

Após as primeiras aplicações nos modelos ADALINE e MADALINE citadas na seção anterior, houveram outras inúmeras utilizações em diversos setores ao longo de vários anos (HAGAN et al., 2014). Dentre eles: aeroespacial, médico, defesa/segurança, automobilístico, processos industriais etc.

2.3.3 Definição

Segundo Gurney (1997), uma rede neural é um conjunto interconectado de elementos, ou nós, de processamento simples, cuja funcionalidade é vagamente baseada no neurônio biológico. A capacidade de processamento da rede é armazenada nas intensidades de conexão, ou pesos, entre as unidades, obtidos por um processo de adaptação ou aprendizado de um conjunto de padrões de treinamento.

Já para Haykin et al. (2009), é um processador distribuído maciçamente paralelo composto de unidades de processamento simples que tem uma propensão natural para armazenar conhecimento experiencial e torná-lo disponível para uso. Assemelha-se ao cérebro em dois aspectos: o conhecimento é adquirido pela rede por meio de um processo de aprendizagem; as intensidades de conexão entre os neurônios, conhecidas como pesos sinápticos, são usadas para armazenar o conhecimento adquirido.

2.3.4 Conceitos principais

2.3.4.1 Pré-processamento de dados

A etapa de pré-processamento dos dados é uma etapa crucial na utilização de algoritmos de *machine learning*, pois garantem que os dados disponibilizados aos algoritmos são completos e consistentes, de forma a obter o melhor desempenho e os melhores resultados (SILVA, 2021). Algumas das etapas chaves desse processo envolvem a limpeza, integração, transformação e redução dos dados (SIVAKUMAR; GUNASUNDARI, 2017).

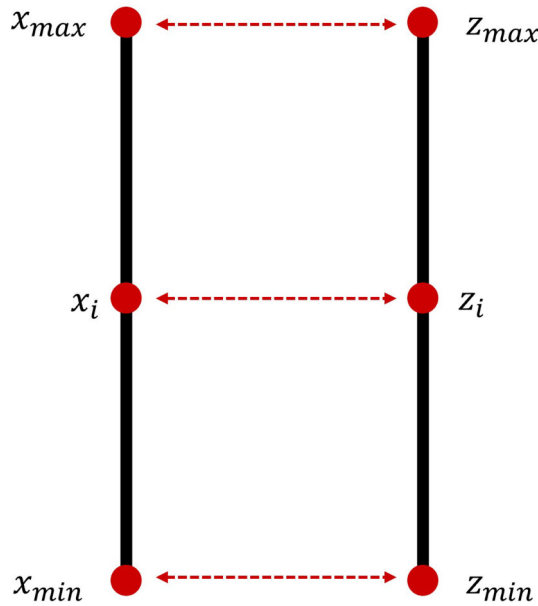
Com relação à limpeza, é comum que o banco de dados contenha informações em falta, que podem surgir durante a criação do mesmo ou através de regras de validação. De qualquer forma, é necessário efetuar o tratamento destes valores em falta. Existem várias alternativas para corrigir este problema, a mais comum é a eliminação das linhas de dados com valores em falta (SILVA, 2021; SIVAKUMAR; GUNASUNDARI, 2017).

Já sobre a integralização, em um problema real e complexo, várias fontes compõem o banco de dados final. Dessa forma, a integralização de toda essa informação faz-se necessária, de forma a combinar múltiplas fontes em uma só, sem que haja problemas de inconsistência e redundância (SILVA, 2021; SIVAKUMAR; GUNASUNDARI, 2017).

Na etapa de transformação, é comum que ocorra a chamada padronização ou normalização. Os dados são realocados da escala original para uma nova escala, geralmente próxima das escalas das funções de ativação, facilitando as operações do modelo de previsão. Um bom exemplo é a técnica de redimensionamento dos valores para novos mínimos e máximos (PATRO; SAHU, 2015).

Nessa técnica, para cada variável aleatória x_i do conjunto de dados, o respectivo valor padronizado z_i é dado pela Equação 2.8. A Figura 13 mostra o conceito dessa metodologia. Através de uma simples relação entre os valores, transforma-se x_i em z_i baseado apenas nos valores extremos de cada escala, x_{max} , x_{min} , z_{max} e z_{min} .

Figura 13 – Técnica de redimensionamento dos valores para novos mínimos e máximos



Fonte: elaborado pelo próprio autor.

$$\frac{(z_{max} - z_i)}{(x_{max} - x_i)} = \frac{(z_{max} - z_{min})}{(x_{max} - x_{min})}$$

$$z_i = \frac{-(z_{max} - z_{min})}{(x_{max} - x_{min})}(x_{max} - x_i) + z_{max} \quad (2.8)$$

Sobre a redução dos dados, em alguns casos existem dados em excesso que dificultam o processamento de tanta informação. Assim, algumas técnicas reduzem o número de instâncias fornecidas ao algoritmo. Porém, o mais comum é ter disponível menos instâncias com demasiados números de variáveis. Para evitar esse tipo de problema, análises de correlação, por exemplo, podem ser úteis a fim de se trabalhar apenas com as variáveis mais relevantes ao caso (SILVA, 2021; SIVAKUMAR; GUNASUNDARI, 2017).

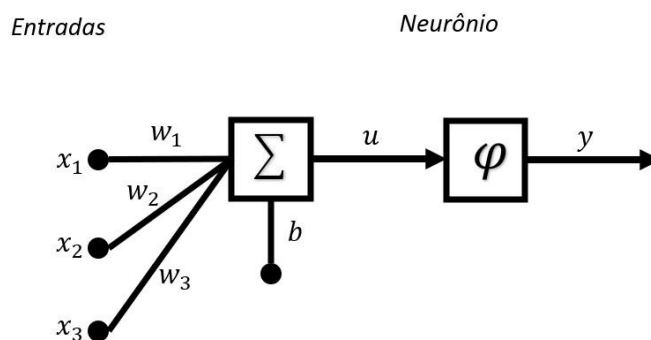
Contudo, nem sempre as análises de redução do número de variáveis trazem os resultados esperados. Dessa forma, essas técnicas devem ser utilizadas com cautela, afinal, as redes neurais artificiais são algoritmos complexos que, na maioria das vezes, conseguem conectar informações que outros algoritmos não conectam (HAYKIN et al., 2009; SILVA, 2021; SIVAKUMAR; GUNASUNDARI, 2017).

2.3.4.2 Neurônios

Uma maneira pragmática de abordagem para melhor entendimento de uma rede neural é através da análise de seus neurônios, pesos sinápticos e do fluxo de informações

(HAGAN et al., 2014; HAYKIN et al., 2009). A Figura 14 é uma representação de um neurônio artificial simples com três entradas e uma saída.

Figura 14 – Representação gráfica de um neurônio artificial de três entradas e uma saída.



Fonte: elaborado pelo próprio autor.

Onde x_i representa as entradas do neurônio e w_i os pesos sinápticos associados à tais entradas, b é o bias aplicado ao neurônio, u o sinal completo de entrada do neurônio, φ é a função de ativação aplicada em u e y a saída do neurônio.

A resposta y desse neurônio é calculada conforme Equação 2.9, onde w_i e b são parâmetros ajustáveis e, tipicamente, após escolha de uma função adequada de ativação, são modificados segundo uma regra de aprendizado definida (HAGAN et al., 2014).

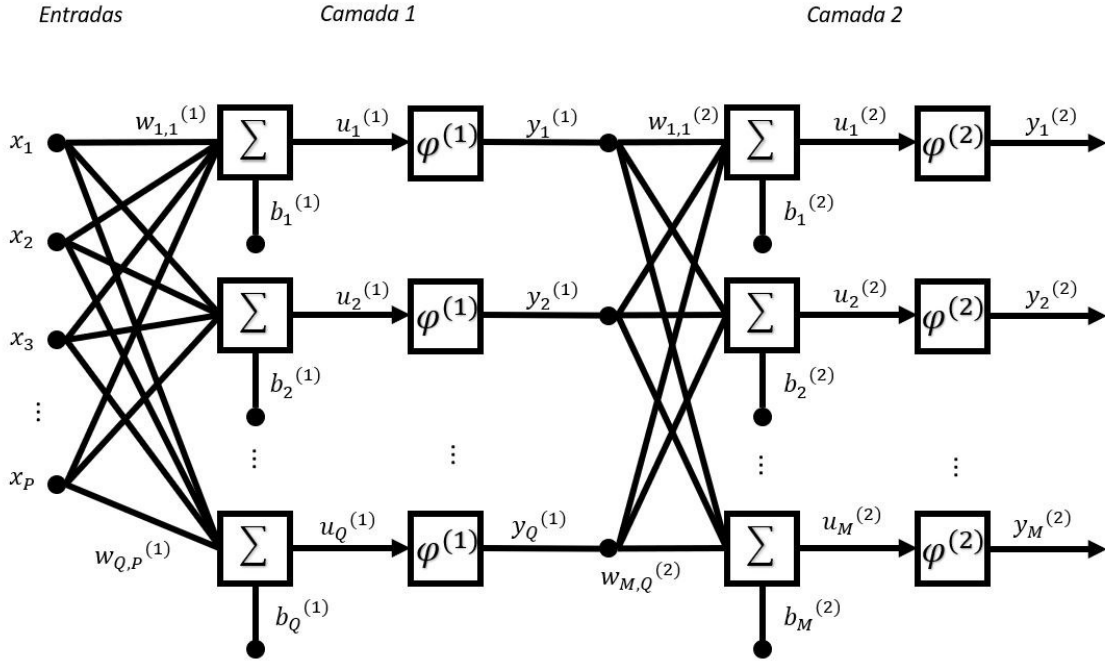
$$y = \varphi(u) = \varphi(w_1x_1 + w_2x_2 + w_3x_3 + b) \quad (2.9)$$

Outra notação usual para a resposta y de um neurônio como o mostrado na Figura 14 considera um vetor de entrada $\mathbf{x} \in \mathbb{R}^{3 \times 1}$ formado pelas entradas x_1 , x_2 e x_3 e um vetor $\mathbf{W} \in \mathbb{R}^{1 \times 3}$ formado pelos pesos sinápticos w_1 , w_2 e w_3 , de tal forma que:

$$y = \varphi(u) = \varphi(\mathbf{W}\mathbf{x} + b) \quad (2.10)$$

É possível estender facilmente o raciocínio utilizado e encontrar as relações entre múltiplos neurônios e entradas. Para uma rede com duas camadas de múltiplos neurônios e entradas, mais especificamente P entradas, Q neurônios na primeira camada e M neurônios na segunda camada, tem-se a representação conforme Figura 15 (HAGAN et al., 2014).

Figura 15 – Representação gráfica de uma rede de duas camadas com múltiplos neurônios e múltiplas entradas.



Fonte: elaborado pelo próprio autor.

A entrada da rede mostrada na Figura 15 pode ser representada por um vetor $\mathbf{X} \in \mathbb{R}^{P \times 1}$, conforme Equação 2.14, enquanto os pesos sinápticos e bias da 1ª camada oculta podem ser descritos, respectivamente, pelas matrizes $\mathbf{W}^{(1)} \in \mathbb{R}^{Q \times P}$ e $\mathbf{b}^{(1)} \in \mathbb{R}^{Q \times 1}$, de acordo com as Equações 2.12 e 2.13.

$$\mathbf{X} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_P \end{bmatrix} \quad (2.11)$$

$$\mathbf{W}^{(1)} = \begin{bmatrix} w_{1,1}^{(1)} & w_{1,2}^{(1)} & \cdots & w_{1,P}^{(1)} \\ w_{2,1}^{(1)} & w_{2,2}^{(1)} & \cdots & w_{2,P}^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ w_{Q,1}^{(1)} & w_{Q,2}^{(1)} & \cdots & w_{Q,P}^{(1)} \end{bmatrix} \quad (2.12)$$

$$\mathbf{b}^{(1)} = \begin{bmatrix} b_1^{(1)} \\ b_2^{(1)} \\ \vdots \\ b_Q^{(1)} \end{bmatrix} \quad (2.13)$$

A saída da 1ª camada oculta pode ser representada por um vetor $\mathbf{Y}^{(1)} \in \mathbb{R}^{Q \times 1}$, segundo Equação 2.14, onde o termo $y_i^{(1)}$ descreve a resposta do i -ésimo neurônio de um total de Q neurônios.

$$\mathbf{Y}^{(1)} = \begin{bmatrix} y_1^{(1)} \\ y_2^{(1)} \\ \vdots \\ y_Q^{(1)} \end{bmatrix} \quad (2.14)$$

Os pesos sinápticos e bias da 2ª camada oculta podem ser detalhados, respectivamente, pelas matrizes $\mathbf{W}^{(2)} \in \mathbb{R}^{M \times Q}$ e $\mathbf{b}^{(2)} \in \mathbb{R}^{M \times 1}$, de acordo com as Equações 2.15 e 2.16.

$$\mathbf{W}^{(2)} = \begin{bmatrix} w_{1,1}^{(2)} & w_{1,2}^{(2)} & \cdots & w_{1,Q}^{(2)} \\ w_{2,1}^{(2)} & w_{2,2}^{(2)} & \cdots & w_{2,Q}^{(2)} \\ \vdots & \vdots & & \vdots \\ w_{M,1}^{(2)} & w_{M,2}^{(2)} & \cdots & w_{M,Q}^{(2)} \end{bmatrix} \quad (2.15)$$

$$\mathbf{b}^{(2)} = \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ \vdots \\ b_M^{(2)} \end{bmatrix} \quad (2.16)$$

Já a saída da 2ª camada oculta pode ser descrita por um vetor $\mathbf{Y}^{(2)} \in \mathbb{R}^{M \times 1}$, conforme Equação 2.17, onde o termo $y_j^{(2)}$ representa a resposta do j -ésimo neurônio de um total de M neurônios.

$$\mathbf{Y}^{(2)} = \begin{bmatrix} y_1^{(2)} \\ y_2^{(2)} \\ \vdots \\ y_M^{(2)} \end{bmatrix} \quad (2.17)$$

De forma resumida, o vetor $\mathbf{Y}^{(2)} \in \mathbb{R}^{M \times 1}$ é então descrito de acordo com a Equação 2.18.

$$\mathbf{Y}^{(2)} = \varphi^{(2)}(\mathbf{W}^{(2)}\mathbf{Y}^{(1)} + \mathbf{b}^{(2)}) = \varphi^{(2)}(\mathbf{W}^{(2)}\varphi^{(1)}(\mathbf{W}^{(1)}\mathbf{X} + \mathbf{b}^{(1)}) + \mathbf{b}^{(2)}) \quad (2.18)$$

2.3.4.3 Funções de ativação (de transferência)

Intuitivamente, as funções de ativação introduzem um componente não linear nas redes neurais, que faz com que tenham a capacidade de melhor assimilar as relações entre as variáveis dependentes e independentes (MÜLLER; REINHARDT; STRICKLAND, 1995).

Considerando-se um modelo de duas camadas, conforme Figura 15, cuja resposta $\mathbf{Y}^{(2)}$, que pode ser traduzida como \mathbf{Y} , é objeto de interesse de previsão, tem-se, omitindo-se os bias $\mathbf{b}^{(1)}$ e $\mathbf{b}^{(2)}$ e sem perda de generalidade:

$$\mathbf{Y} = \varphi^{(2)}(\mathbf{W}^{(2)}\varphi^{(1)}(\mathbf{W}^{(1)}\mathbf{p})) \quad (2.19)$$

Nota-se que, retirando-se as funções de ativação $\varphi^{(1)}$ e $\varphi^{(2)}$ e considerando-se \mathbf{V} a multiplicação $\mathbf{W}^{(1)}\mathbf{W}^{(2)}$, o modelo de resposta dessa rede sem função de ativação torna-se um simples modelo de regressão linear:

$$\mathbf{Y} = \mathbf{W}^{(2)}\mathbf{W}^{(1)}\mathbf{X} = \mathbf{V}\mathbf{X} \quad (2.20)$$

Assim, essas funções mostram-se essenciais para dar capacidade representativa às redes neurais artificiais, introduzindo um componente de não linearidade, além de limitar a amplitude de resposta do neurônio para valores finitos e/ou conhecidos.

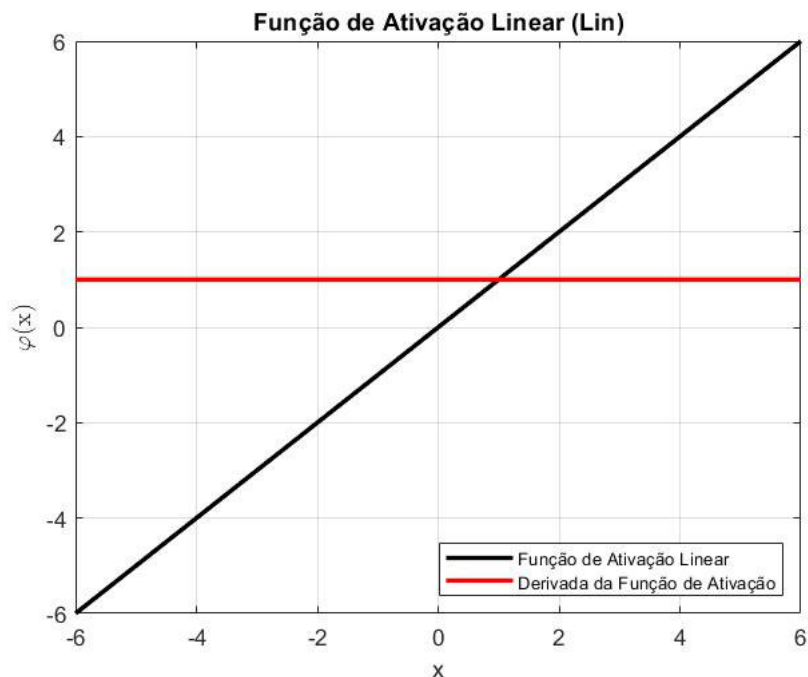
Existem diversos tipos, dentre as mais conhecidas: linear, logística e tangente hiperbólica. É importante ressaltar que para diversos tipos de regras de aprendizado, tópico que será apresentado nas seções posteriores, faz-se necessário o conhecimento da derivada da função de ativação, dessa forma, nos gráficos apresentados a seguir, mostram-se tanto a função de ativação quanto sua derivada.

- **Função de Ativação Linear (lin):** É uma das funções mais básicas já que não altera o sinal de saída do neurônio, sendo bastante utilizada nas camadas finais das redes neurais (SHARMA; SHARMA; ATHAIYA, 2020). As Equações 2.21 e 2.22 e a Figura 16 mostram a função de ativação linear e sua derivada.

$$\varphi(x) = x \quad (2.21)$$

$$\varphi'(x) = 1 \quad (2.22)$$

Figura 16 – Representação da Função de Ativação Linear e sua derivada.



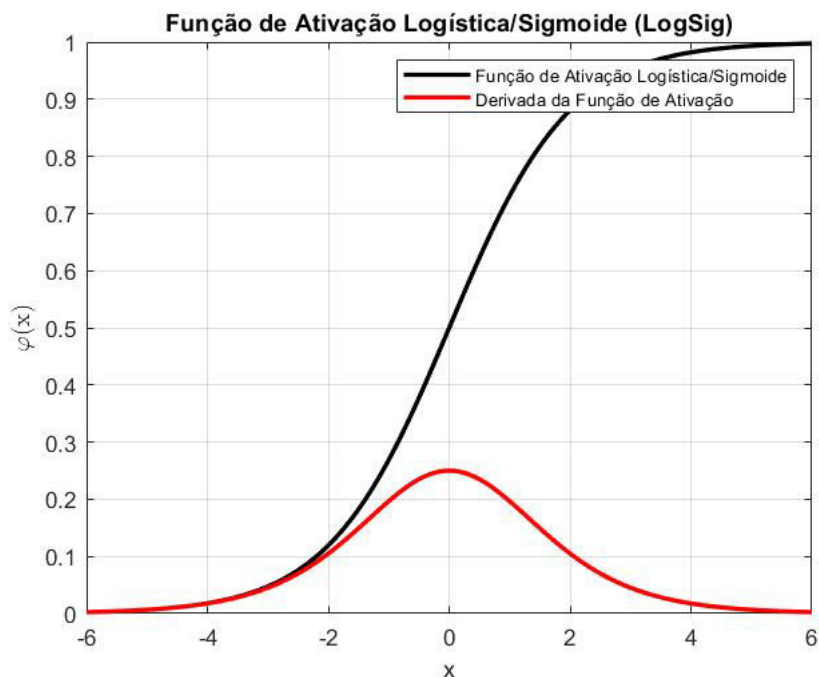
Fonte: elaborado pelo próprio autor.

- **Função de Ativação Logística/Sigmoide (logsig):** A função sigmoide já foi uma das mais utilizadas em RNAs (Redes Neurais Artificiais) por ser biologicamente plausível, considerando o comportamento binário dos neurônios na forma ativada vs. não ativada, onde 1 representa a ativação e 0 a não ativação (GERON, 2019). No entanto, sua derivada satura para valores acima de 5 e abaixo de -5, o que prejudica as regras de aprendizado da rede, dificultando o treinamento dos modelos. As Equações 2.23 e 2.24 e a Figura 17 mostram essa função e sua derivada.

$$\varphi(x) = \frac{1}{1 + e^{-x c}} \quad (2.23)$$

$$\varphi'(x) = \varphi(x) (1 - \varphi(x)) c \quad (2.24)$$

Figura 17 – Representação da Função de Ativação Logística/Sigmóide e sua derivada com $c = 1$.



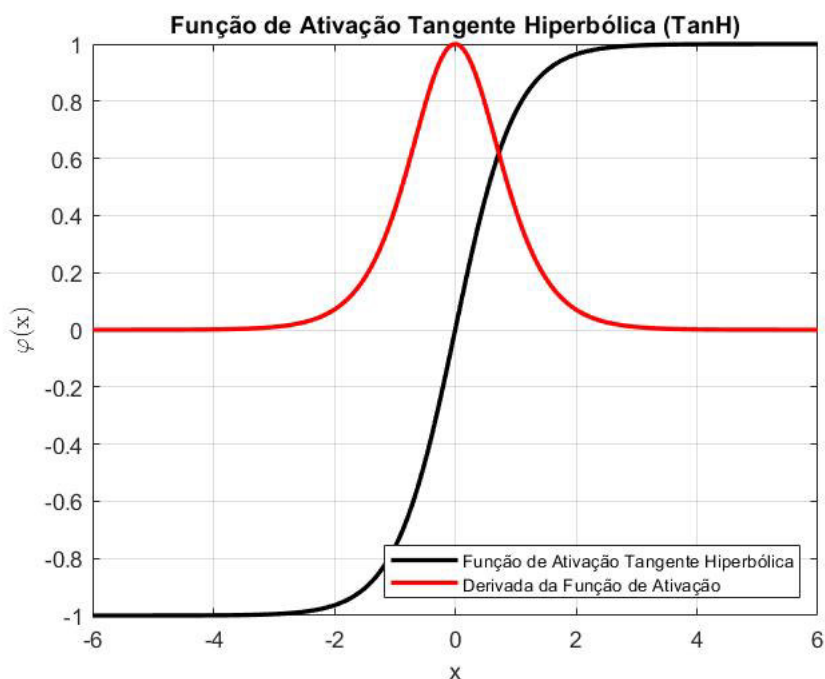
Fonte: elaborado pelo próprio autor.

- **Função de Ativação Tangente Hiperbólica (tanh):** Similar à função sigmoide, a função tangente hiperbólica também possui um formato em ‘S’, mas varia de -1 a 1, ao invés de 0 a 1 como na sigmoide. A *tanh* se aproxima mais da identidade, sendo assim, considerada por alguns autores uma alternativa mais atraente para a ativação das camadas ocultas das RNAs (NWANKPA et al., 2018). Apesar de ainda ser notável a saturação da sua derivada, ela alcança valores mais próximos de 1 quando $x = 0$, como pode ser observado nas Equações 2.25 e 2.26 e Figura 18.

$$\varphi(x) = a \tanh(x b) \quad (2.25)$$

$$\varphi'(x) = \frac{b}{a} (a - \varphi(x)) (a + \varphi(x)) \quad (2.26)$$

Figura 18 – Representação da Função de Ativação Tangente Hiperbólica e sua derivada com $a = 1$ e $b = 1$.



Fonte: elaborado pelo próprio autor.

2.3.4.4 Índice de aprendizado e sua otimização

O índice de aprendizado, também chamado de função de custo, pode ser definido como um indicador que avalia quão bem a rede neural, após treinamento, consegue prever a resposta esperada de tal forma que, quando esse índice for pequeno, significa que a rede teve um bom desempenho e quando o índice for grande, significa que a rede desempenhou de forma insuficiente. Para isso, buscam-se formas de otimização dos parâmetros do modelo (pesos sinápticos e bias) no intuito de reduzir o índice de performance (AHMADI; MOTTAGHITALAB; NARIMAN-ZADEH, 2007; HAYKIN et al., 2009; HAGAN et al., 2014).

Seja $F(x)$ o índice de desempenho da rede que se deseja minimizar, onde x é um escalar ajustável, assumindo-se que $F(x)$ seja uma função analítica para que sua derivada exista, essa função pode ser representada através de uma expansão em série de Taylor (HUMMEL; SEEBECK, 1949) sobre algum ponto x^* , conforme Equação 2.27.

$$\begin{aligned}
 F(x) = & F(x^*) + \frac{d}{dx}F(x)\Big|_{x=x^*} (x - x^*) + \frac{1}{2} \frac{d^2}{dx^2}F(x)\Big|_{x=x^*} (x - x^*)^2 + \dots \\
 & \dots + \frac{1}{n!} \frac{d^n}{dx^n}F(x)\Big|_{x=x^*} (x - x^*)^n + \dots
 \end{aligned} \tag{2.27}$$

Através de uma aproximação dessa expansão com a utilização de um número finito de termos, é possível estimar a função de custo.

Como, na aplicação real em uma rede neural, esse índice não é dependente de uma única variável, mas sim de todos os parâmetros da rede como pesos e bias, deve-se trabalhar a função no formato matricial $F(\mathbf{x})$, conforme Equação 2.28 ou, considerando $\Delta\mathbf{x} = \mathbf{x} - \mathbf{x}^*$, segundo Equação 2.29.

$$F(\mathbf{x}) = F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) + \frac{1}{2} (\mathbf{x} - \mathbf{x}^*)^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} (\mathbf{x} - \mathbf{x}^*) + \dots \quad (2.28)$$

$$F(\mathbf{x}) = F(\mathbf{x}^* + \Delta\mathbf{x}) = F(\mathbf{x}^*) + \nabla F(\mathbf{x})^T \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} + \frac{1}{2} \Delta\mathbf{x}^T \nabla^2 F(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}^*} \Delta\mathbf{x} + \dots \quad (2.29)$$

Onde \mathbf{x} pode ser entendido como um vetor de parâmetros da rede, \mathbf{x}^* um novo vetor de parâmetros da rede, $\nabla F(\mathbf{x})$ o gradiente, definido pela Equação 2.36, e $\nabla^2 F(\mathbf{x})$ a matriz Hessiana, definida pela Equação 2.31.

$$\nabla F(\mathbf{x}) = \left[\frac{\partial}{\partial x_1} F(\mathbf{x}) \quad \frac{\partial}{\partial x_2} F(\mathbf{x}) \quad \dots \quad \frac{\partial}{\partial x_n} F(\mathbf{x}) \right]^T \quad (2.30)$$

$$\nabla^2 F(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2}{\partial x_1^2} F(\mathbf{x}) & \frac{\partial^2}{\partial x_1 x_2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_1 x_n} F(\mathbf{x}) \\ \frac{\partial^2}{\partial x_2 \partial x_1} F(\mathbf{x}) & \frac{\partial^2}{\partial x_2^2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_2 x_n} F(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2}{\partial x_n \partial x_1} F(\mathbf{x}) & \frac{\partial^2}{\partial x_n \partial x_2} F(\mathbf{x}) & \dots & \frac{\partial^2}{\partial x_n^2} F(\mathbf{x}) \end{bmatrix} \quad (2.31)$$

O termo "otimização", no contexto desse estudo, pode ser entendido como a busca por um valor \mathbf{x} que minimize a função de custo $F(\mathbf{x})$. A partir de um palpite inicial \mathbf{x}_0 , essa busca ocorre em estágios de acordo com a Equação 2.32 ou 2.33, onde o vetor \mathbf{p}_k representa uma direção de busca e o escalar α_k uma taxa de aprendizado (HAGAN et al., 2014).

$$\mathbf{x}_{k+1} = \mathbf{x}_k + \alpha_k \mathbf{p}_k \quad (2.32)$$

$$\Delta\mathbf{x}_k = (\mathbf{x}_{k+1} - \mathbf{x}_k) = \alpha_k \mathbf{p}_k \quad (2.33)$$

Quando a busca é iniciada com $\mathbf{x}_k (= \mathbf{x}_0)$, procura-se atualizar esse valor inicial para \mathbf{x}_{k+1} de tal forma que a Equação 2.34 seja satisfeita:

$$F(\mathbf{x}_{k+1}) < F(\mathbf{x}_k) \quad (2.34)$$

Existem diversas maneiras de se abordar a otimização de $F(\mathbf{x})$. No presente trabalho, duas vertentes principais foram empregadas: a do Gradiente Descendente e a do Método de Newton.

No Gradiente Descendente, utiliza-se a aproximação de primeira ordem da expansão em série de Taylor mostrada na Equação 2.29 entorno do valor anterior \mathbf{x}_k , conforme Equação 2.35 (HAGAN et al., 2014; HAYKIN et al., 2009).

$$F(\mathbf{x}_{k+1}) = F(\mathbf{x}_k + \Delta\mathbf{x}_k) \cong F(\mathbf{x}_k) + \mathbf{g}_k^T \Delta\mathbf{x}_k \quad (2.35)$$

Onde \mathbf{g}_k é o gradiente calculado para \mathbf{x}_k :

$$\mathbf{g}_k \equiv \nabla F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k} \quad (2.36)$$

Para que seja possível seguir a Equação 2.34, considerando-se o escalar α_k positivo e pequeno:

$$\mathbf{g}_k^T \Delta\mathbf{x}_k = \alpha_k \mathbf{g}_k^T \mathbf{p}_k < 0 \implies \mathbf{g}_k^T \mathbf{p}_k < 0 \quad (2.37)$$

Buscando-se a direção que mais decresce o valor de $F(\mathbf{x})$, pode-se dizer que $\mathbf{p}_k = -\mathbf{g}_k$, o que transforma a Equação 2.32 em:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \alpha_k \mathbf{g}_k \quad (2.38)$$

Já no Método de Newton, é utilizada a aproximação de segunda ordem da expansão em série de Taylor mostrada na Equação 2.29 entorno do valor \mathbf{x}_k , segundo Equação 2.39 (HAYKIN et al., 2009; HAGAN et al., 2014).

$$F(\mathbf{x}_{k+1}) = F(\mathbf{x}_k + \Delta\mathbf{x}_k) \cong F(\mathbf{x}_k) + \mathbf{g}_k^T \Delta\mathbf{x}_k + \frac{1}{2} \Delta\mathbf{x}_k^T \mathbf{A}_k \Delta\mathbf{x}_k \quad (2.39)$$

Onde \mathbf{g}_k é o gradiente calculado para \mathbf{x}_k , mostrado na Equação 2.36 e \mathbf{A}_k é a matriz Hessiana:

$$\mathbf{A}_k \equiv \nabla^2 F(\mathbf{x})|_{\mathbf{x}=\mathbf{x}_k} \quad (2.40)$$

Como a função é quadrática, calculando seu gradiente com respeito à $\Delta\mathbf{x}_k$ e admitindo ser zero, tem-se que:

$$\mathbf{g}_k + \mathbf{A}_k \Delta\mathbf{x}_k = 0 \quad (2.41)$$

Resolvendo para $\Delta \mathbf{x}_k$:

$$\Delta \mathbf{x}_k = -\mathbf{A}_k^{-1} \mathbf{g}_k \quad (2.42)$$

Dessa forma, o método de Newton fica definido como:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \mathbf{A}_k^{-1} \mathbf{g}_k \quad (2.43)$$

Essas duas equações, Equação 2.38 e Equação 2.43, formam o ponto de partida para os métodos de retropropagação (*backpropagation*) utilizados no presente trabalho. Elas definem uma técnica recursiva de otimização dos parâmetros da rede, como pesos sinápticos e bias, segundo uma metodologia específica (HAGAN et al., 2014; HAYKIN et al., 2009).

Enquanto a primeira considera os termos de primeira ordem da expansão em série de Taylor da função custo e, dessa forma, necessita apenas das derivadas parciais de primeira ordem para os cálculos (gradiente \mathbf{g}_k), a segunda considera os termos de segunda ordem, sendo necessário o cálculo das derivadas parciais de segunda ordem (matriz Hessiana \mathbf{A}_k) (HAGAN et al., 2014; HAYKIN et al., 2009).

2.3.4.4.1 Gradiente descendente com *momentum*

Conforme discutido, a técnica recursiva do gradiente descendente pode ser resumida pela Equação 2.38. Contudo, uma técnica similar foi utilizada no trabalho: gradiente descendente com *momentum*.

O termo "*momentum*" refere-se a uma constante adicionada na atualização dos parâmetros que atua como um filtro, de forma a diminuir as oscilações intrínsecas do método na trajetória de busca da otimização da função de custo.

Essa técnica recursiva pode ser entendida como:

$$\mathbf{x}_{k+1} = \gamma \mathbf{x}_k - (1 - \gamma) \alpha_k \mathbf{g}_k \quad (2.44)$$

Onde γ é a constante *momentum*, α é a constante taxa de aprendizado e \mathbf{g}_k é o gradiente.

2.3.4.4.2 Levenberg–Marquardt

O Método de Newton da Equação 2.43 dá origem, através de pequenas alterações, ao método recursivo Levenberg–Marquardt.

Primeiramente, deve-se considerar que a função de custo $F(\mathbf{x})$, definida na Equação 2.39, é uma soma de funções quadráticas, conforme:

$$F(\mathbf{x}) = \sum_{i=1}^N v_i^2(\mathbf{x}) = \mathbf{v}^T(\mathbf{x})\mathbf{v}(\mathbf{x}) \quad (2.45)$$

Assim, o j -ésimo elemento do gradiente seria:

$$[\nabla F(\mathbf{x})]_j = \frac{\partial F(\mathbf{x})}{\partial x_j} = 2 \sum_{i=1}^N v_i(\mathbf{x}) \frac{\partial v_i(\mathbf{x})}{\partial x_j} \quad (2.46)$$

E, dessa forma, o gradiente é escrito como:

$$\nabla F(\mathbf{x}) = 2\mathbf{J}^T(\mathbf{x})\mathbf{v}(\mathbf{x}) \quad (2.47)$$

onde $\mathbf{J}(\mathbf{x})$ é a matriz Jacobiana definida por:

$$\mathbf{J}(\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1} v_1(\mathbf{x}) & \frac{\partial}{\partial x_2} v_1(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_n} v_1(\mathbf{x}) \\ \frac{\partial}{\partial x_1} v_2(\mathbf{x}) & \frac{\partial}{\partial x_2} v_2(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_n} v_2(\mathbf{x}) \\ \vdots & \vdots & & \vdots \\ \frac{\partial}{\partial x_1} v_N(\mathbf{x}) & \frac{\partial}{\partial x_2} v_N(\mathbf{x}) & \cdots & \frac{\partial}{\partial x_n} v_N(\mathbf{x}) \end{bmatrix} \quad (2.48)$$

Da mesma maneira, o elemento k,j da matriz Hessiana pode ser escrito como:

$$[\nabla^2 F(\mathbf{x})]_{k,j} = \frac{\partial^2 F(\mathbf{x})}{\partial x_k \partial x_j} = 2 \sum_{i=1}^N \left[\frac{\partial v_i(\mathbf{x})}{\partial x_k} \frac{\partial v_i(\mathbf{x})}{\partial x_j} + v_i(\mathbf{x}) \frac{\partial^2 v_i(\mathbf{x})}{\partial x_k \partial x_j} \right] \quad (2.49)$$

O que permite escrever a matriz Hessiana no formato:

$$\nabla^2 F(\mathbf{x}) = 2\mathbf{J}^T(\mathbf{x})\mathbf{J}(\mathbf{x}) + 2\mathbf{S}(\mathbf{x}) \quad (2.50)$$

Onde $\mathbf{S}(\mathbf{x})$ é definido por:

$$\mathbf{S}(\mathbf{x}) = \sum_{i=1}^N v_i(\mathbf{x}) \nabla^2 v_i(\mathbf{x}) \quad (2.51)$$

Seja $\mathbf{S}(\mathbf{x})$ muito pequeno, podemos aproximar a matriz Hessiana por:

$$\nabla^2 F(\mathbf{x}) \cong 2\mathbf{J}^T(\mathbf{x})\mathbf{J}(\mathbf{x}) \quad (2.52)$$

Substituindo os termos, temos finalmente, o que é chamado de método Gauss-Newton:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[2\mathbf{J}^T(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) \right]^{-1} 2\mathbf{J}^T(\mathbf{x}_k)\mathbf{v}(\mathbf{x}_k) \quad (2.53)$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{J}^T(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) \right]^{-1} \mathbf{J}^T(\mathbf{x}_k)\mathbf{v}(\mathbf{x}_k) \quad (2.54)$$

Nota-se que a vantagem desse método em relação ao Método de Newton comum é que o cálculo das derivadas de segunda ordem é desnecessário (HAGAN et al., 2014).

Um problema desse método é que a matriz $\mathbf{J}^T(\mathbf{x})\mathbf{J}(\mathbf{x})$ pode não ser invertível. Daí surge o método de Levenberg–Marquardt, que propõe uma simples modificação para melhorar a invertibilidade da matriz, através da adição de um parâmetro $\mu_k\mathbf{I}$, onde \mathbf{I} é a matriz identidade:

$$\mathbf{x}_{k+1} = \mathbf{x}_k - \left[\mathbf{J}^T(\mathbf{x}_k)\mathbf{J}(\mathbf{x}_k) + \mu_k\mathbf{I} \right]^{-1} \mathbf{J}^T(\mathbf{x}_k)\mathbf{v}(\mathbf{x}_k) \quad (2.55)$$

A constante μ_k é chamada de parâmetro de Levenberg–Marquardt (ou parâmetro LM). Geralmente inicia-se o parâmetro LM com um valor pequeno como 0.01. Caso no passo k a função de custo não seja minimizada, no passo seguinte $k+1$, μ_k é multiplicado por um fator $\theta > 1$, por exemplo, $\theta = 10$. Porém, se a função de custo é minimizada, μ_k é dividido por esse valor θ no passo seguinte (LV et al., 2017; MEDEIROS, 2008; HAGAN; MENHAJ, 1994).

Observa-se que quando μ_k torna-se grande, o método aproxima-se do método gradiente descendente com uma pequena taxa de aprendizado, enquanto μ_k próximo a zero, aproxima-se do Método de Newton (LV et al., 2017; MEDEIROS, 2008; HAGAN; MENHAJ, 1994).

2.3.4.5 Treinamento, validação e teste

A essência do algoritmo de aprendizado supervisionado, popularmente conhecido como retropropagação de erro (*error backpropagation*), é codificar um mapeamento de entrada e saída (representado por um conjunto de exemplos rotulados) nos pesos e limites sinápticos de uma rede. A expectativa é que uma rede bem treinada seja capaz de generalizar, ou seja, alcance bom desempenho durante a análise de dados não apresentados na etapa de treinamento. Nessa perspectiva, o processo de aprendizagem equivale a uma escolha de parametrização de rede para um determinado conjunto de dados. Mais especificamente, podemos ver o problema de seleção de rede como escolher, dentro de um

conjunto de estruturas de modelos candidatas (parametrizações), o “melhor” de acordo com um determinado critério (HAYKIN et al., 2009).

Normalmente, uma rede treinada com o algoritmo de retropropagação aprende em estágios, passando da realização de funções de mapeamento bastante simples para funções de mapeamento mais complexas conforme a sessão de treinamento avança, com o erro decrescendo indefinidamente conforme o número de épocas transcorre. Com uma boa generalização como objetivo, é muito difícil descobrir quando é melhor interromper o treinamento se olharmos para a curva de aprendizado do treinamento por si só. Um super-treinamento da rede (*overfitting*) pode acontecer caso a sessão de treinamento não seja interrompida no momento correto. O início do *overfitting* pode ser identificado por meio de validação cruzada (HAYKIN et al., 2009; MÜLLER; REINHARDT; STRICKLAND, 1995).

Idealmente, o modelo deve ser avaliado em amostras que não foram usadas para construir ou ajustar o modelo, de modo que forneçam um senso de eficácia do modelo. Quando uma grande quantidade de dados está disponível, um conjunto de amostras pode ser reservado para avaliar o modelo final. O conjunto de dados de “treinamento” é o termo geral para as amostras usadas para criar o modelo, enquanto o conjunto de dados de “teste” ou “validação” é usado para qualificar o desempenho (KUHN; JOHNSON, 2013b; KUHN; JOHNSON, 2013a).

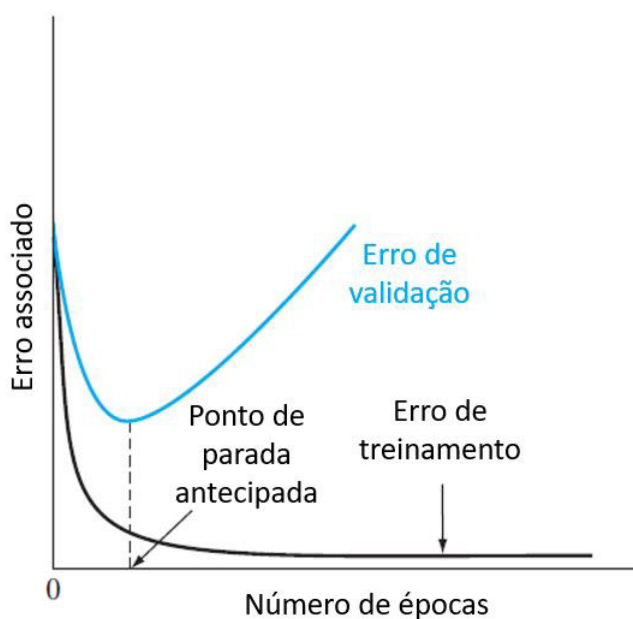
Assim, pode-se classificar três conjuntos distintos (RIPLEY, 1996):

- Conjunto de dados de treinamento: a amostra de dados usada para ajustar o modelo.
- Conjunto de dados de validação: a amostra de dados usada para fornecer uma avaliação não viesada de um ajuste de modelo no conjunto de dados de treinamento durante o ajuste dos hiperparâmetros do modelo. A avaliação torna-se mais tendencial conforme a habilidade do conjunto de dados de validação é incorporada à configuração do modelo.
- Conjunto de testes: um conjunto de exemplos usados apenas para avaliar o desempenho de uma rede treinada.

2.3.4.6 Critérios de parada

A Figura 19 mostra duas curvas esquemáticas de erro associadas aos subconjuntos de validação e de estimação em função do número de épocas de treinamento.

Figura 19 – Ilustração esquemática do ponto mínimo na curva de validação como critério de parada antecipada da etapa de treinamento de uma rede neural.



Fonte: adaptado de [Haykin et al. \(2009\)](#)

Normalmente, o modelo não se sai tão bem no subconjunto de validação quanto no subconjunto de estimativa, no qual seu modelo foi baseado. A curva de aprendizado de estimativa diminui monotonicamente por um número crescente de épocas. Em contraste, a curva de aprendizado de validação diminui monotonicamente para um mínimo e então começa a aumentar conforme o treinamento continua ([HAYKIN et al., 2009](#); [GURNEY, 1997](#)).

Quando olhamos para a curva de aprendizado de estimativa, pode parecer que poderíamos fazer melhor indo além do ponto mínimo na curva de aprendizado de validação. Na realidade, porém, o que a rede está aprendendo além desse ponto é essencialmente o ruído contido nos dados de treinamento. Essa heurística sugere que o ponto mínimo na curva de aprendizado da validação seja usado como um critério para interromper a sessão de treinamento ([HAYKIN et al., 2009](#); [MÜLLER](#); [REINHARDT](#); [STRICKLAND, 1995](#)).

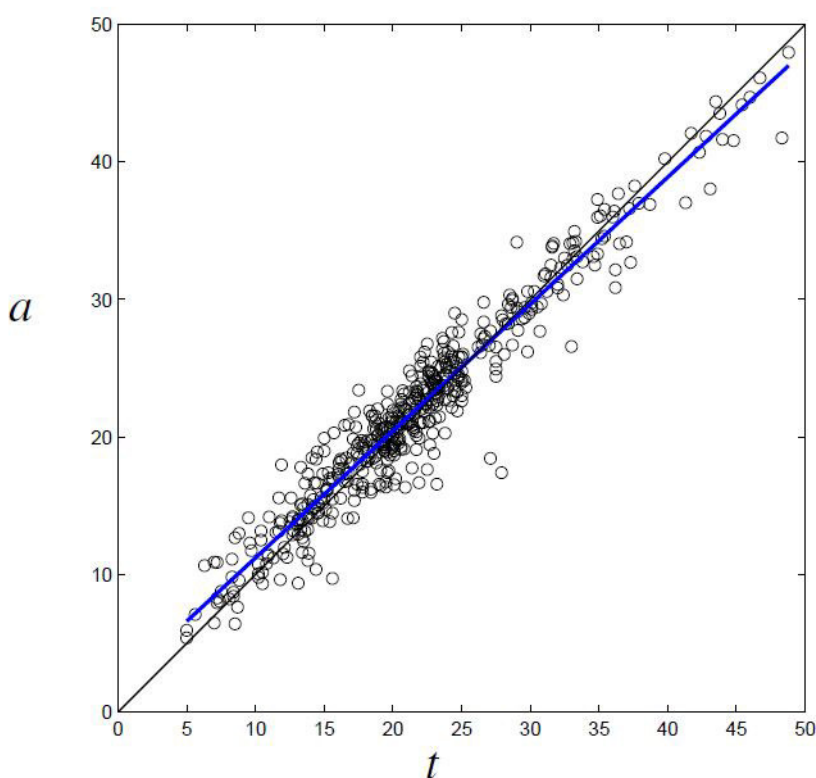
2.3.4.7 Análise pós-treinamento

Uma vez concluído o treinamento da rede neural, ou seja, após ajustar seus pesos sinápticos e bias com base em um conjunto de treinamento e utilizar o erro associado ao conjunto de validação cruzada como critério de parada antecipada, a qualidade do treinamento é verificada através da análise de um conjunto de teste composto por dados

não apresentados no treinamento da rede (HAGAN; MENHAJ, 1994; HAYKIN et al., 2009).

Existem muitas técnicas para análise pós-treinamento, porém a mais comum para fins de previsão é a análise com o gráfico de dispersão entre os valores previstos e os valores reais, conforme Figura 20 (HAGAN; MENHAJ, 1994).

Figura 20 – Gráfico de dispersão entre os valores previstos e o valores reais de um conjunto de dados de teste



Fonte: adaptado de Hagan e Menhaj (1994)

A linha azul representa uma reta de regressão entre os dados previstos, a , e os dados reais, t , a linha preta representa a correlação perfeita entre eles ($a = t$) e os círculos representam os dados. No exemplo apresentado pela Figura 20, a qualidade da previsão foi alta, apesar de não ser perfeita.

Contudo, apenas gráficos não podem medir satisfatoriamente a qualidade do modelo. Deve-se, para tanto, utilizar medidas quantitativas de verificação de performance do algoritmo. Para fins de regressão, os índices mais utilizadas são: coeficiente de determinação (R^2), erro percentual absoluto médio (EPAM) e erro quadrático médio (EQM) (MÜLLER; REINHARDT; STRICKLAND, 1995; GURNEY, 1997; HAGAN; MENHAJ, 1994).

2.3.4.7.1 Coeficiente de determinação (R^2)

Segundo [Hoffmann \(2016\)](#), o coeficiente de determinação pode ser definido conforme Equação 2.56, em que os termos SQ_{total} e SQ_{exp} são apresentados pelas Equações 2.57 e 2.58, respectivamente.

$$R^2 = \frac{SQ_{exp}}{SQ_{total}} \quad (2.56)$$

$$SQ_{total} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.57)$$

$$SQ_{exp} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (2.58)$$

R^2 representa o coeficiente de determinação, SQ_{total} representa a soma quadrática total, SQ_{exp} a soma quadrática explicada, n o número de observações, y_i o valor observado/real, \bar{y} a média das observações e \hat{y}_i o valor estimado/previsto.

Dessa forma, SQ_{total} indica a variação dos dados observados em relação a sua média, já o SQ_{exp} indica a variação dos dados calculados pela reta de regressão em relação a média dos dados reais. Juntando essas duas informações na Equação 2.56, tem-se que o coeficiente de regressão R^2 indica um percentual entre a variação que pode ser explicada pelo ajuste e a variação total dos dados, ou o quanto o modelo de previsão explica sobre a variação dos dados. Quanto mais próximo de 1 melhor é a qualidade do modelo de previsão, na medida em que os dados previstos coincidem perfeitamente com os dados reais ([MONTGOMERY; PECK; VINING, 2012](#); [HOFFMANN, 2016](#); [BRAGA, 2019](#); [LOBO, 2007](#)).

2.3.4.7.2 Erro percentual absoluto médio (EPAM)

As medições de erros estatísticos desempenham um papel crítico de precisão nas previsões, pelo rastreamento, monitoramento de exceções, e verificação da efetividade dos resultados ([LOBO, 2007](#); [SWAMIDASS, 2000](#)).

O EPAM é a média dos erros percentuais absolutos das previsões, sendo uma medida de fácil compreensão porque fornece o erro em termos de porcentagem. Além disso, como são usados erros de porcentagem absoluta, evita-se o problema de erros positivos e negativos anularem-se mutuamente. Conseqüentemente, o EPAM, definido na Equação

2.59, é uma medida comumente utilizada em previsões. Quanto menor for o EPAM, melhor será a previsão (SWAMIDASS, 2000).

$$EPAM = \frac{\sum_{i=1}^n \left(\frac{y_i - \hat{y}_i}{y_i} \right)}{n} \times 100 \quad (2.59)$$

Na Equação 2.59, y_i é o valor real/observado, \hat{y}_i é o valor calculado pelo modelo de previsão e n o número de observações disponíveis.

2.3.4.7.3 Erro quadrático médio (EQM)

Segundo Sammut e Webb (2010), o erro médio quadrático é um índice de avaliação de modelo frequentemente usada em modelos de regressão. O EQM de um modelo em relação a um conjunto de teste é a média dos erros de predição quadráticos em todas as instâncias do conjunto de teste, conforme Equação 2.60, onde y_i é o valor real/observado, \hat{y}_i é o valor calculado pelo modelo de previsão e n o número de observações disponíveis.

$$EQM = \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n} \quad (2.60)$$

Da mesma maneira que no EPAM, quanto menor o valor de EQM, melhor é a qualidade preditiva do modelo avaliado.

2.4 SVD NA BUSCA DO NÚMERO ÓTIMO DE NEURÔNIOS

2.4.1 Contexto

Diversos pesquisadores já elaboraram técnicas de estimação do número ótimo de neurônios ocultos em redes MLP a fim de reduzir uma possível busca exaustiva hiperparamétrica. Trabalhos como Gómez, Franco e Jerez (2009), Delogu, Fanni e Montisci (2008), Trenn (2008), Medeiros e Barreto (2007) abordaram algumas análises teóricas, mas não aplicaram em dados reais e complexos. Além de outros estudos, mais recentemente Teoh, Tan e Xiang (2006) utilizou a Decomposição em Valores Singulares (*Singular Value Decomposition*, SVD) na estimação de neurônios ocultos através da matriz de valores singulares de funções de ativação da camada oculta. Uma técnica muito promissora é a de Santos, Barreto e Medeiros (2010), que utiliza algumas matrizes das respostas de uma rede neural de classificação de camada oculta única em conjunto com a análise SVD e com a Análise de Componentes Principais (*Principal Component Analysis*, PCA).

O presente trabalho adapta os conceitos e métodos alcançados por Santos, Barreto e Medeiros (2010) para a estimação do número ótimo de neurônios ocultos em redes

MLP de regressão de duas camadas ocultas. Também opta por concentrar os esforços nas análises SVD das matrizes propostas pela técnica, visto que em testes iniciais, resultados para o PCA foram inconclusivos.

2.4.2 Conceito e aplicação

A SVD é uma técnica de análise multivariada de dados amplamente utilizada em redes neurais artificiais e reconhecimento de padrões em processos como compreensão de dados, seleção e extração de atributos, etc. Essa ferramenta pode ser empregada principalmente no pré-processamento de dados, podendo ser considerada como um método de transformação de variáveis correlacionadas em variáveis descorrelacionadas. Também é bastante comum o emprego desse recurso, através de uma análise de significância dos valores singulares, para a redução de dimensionalidade do conjunto de dados inicial. (SANTOS, 2010; SANTOS J. D. A. ; BARRETO; MEDEIROS, 2010; SANTOS; BARRETO; MEDEIROS, 2010).

Golub G. H.; Van Loan (1996), STRANG (1980), Lay (2011) resumem o teorema da decomposição de valores singulares da seguinte forma: seja uma matriz retangular $\mathbf{A} \in \mathbb{R}^{m \times n}$, existe um produto entre três outras matrizes que é resultado de sua fatoração, conforme Equação 2.61, em que $\mathbf{U} \in \mathbb{R}^{m \times m}$ e $\mathbf{V} \in \mathbb{R}^{n \times n}$ são matrizes ortogonais.

$$\mathbf{A} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T \quad (2.61)$$

$\mathbf{\Sigma} \in \mathbb{R}^{m \times n}$ é a matriz diagonal que contém os valores singulares (raiz quadrada dos autovalores) de \mathbf{U} e \mathbf{V} em ordem decrescente, conforme Equação 2.62, onde $r \leq \min(m, n)$ é o número de valores singulares não-nulos.

$$\mathbf{\Sigma} = \text{diag}[\sigma_1, \sigma_2, \dots, \sigma_r] \quad (2.62)$$

A variância representada pelos autovalores mais significativos é comumente chamada de variância explicada (S_{exp}^2). A partir desses autovalores já ordenados na Equação 2.62, Santos, Barreto e Medeiros (2010) utiliza um limiar $0 < \gamma' < 1$ que define um S_{exp}^2 aceitável ao problema, seguindo a Equação 2.63.

$$q_F = \arg \min_{q=1, \dots, q_0} \left\{ \frac{\sum_{i=1}^q \sigma_i^2}{\sum_{j=1}^{q_0} \sigma_j^2} \geq \gamma' \right\} \quad (2.63)$$

A Equação 2.63 estabelece então que, para um dado valor de γ' , o número de neurônios ocultos deve ser definido como o menor valor q que satisfaça a desigualdade $\sum_{i=1}^q \sigma_i^2 \setminus \sum_{j=1}^{q_0} \sigma_j^2 \geq \gamma'$.

As três matrizes nas quais essa técnica é aplicada a fim de se encontrar os valores q_F da Equação 2.63 foram definidas por Santos, Barreto e Medeiros (2010) para uma camada oculta: $\mathbf{Y}^{(h)}$, $\mathbf{E}^{(h)}$ e $\mathbf{\Delta}^{(h)}$. O procedimento adotado para a aquisição desses dados é simples: realizar o aprendizado da rede normalmente, adaptando todos os pesos sinápticos e encontrando seus valores ótimos de acordo com a técnica de solução escolhida; após isso, apresentar novamente os vetores de treinamento à rede já otimizada, contudo sem modificação dos pesos sinápticos, ou seja, apenas executa-se o chamado *feedforward* (sentido direto da rede), sem a execução do *backforward* (sentido reverso da rede); dessa forma, o único objetivo da última etapa de reapresentação dos dados de treinamento à rede é a construção das matrizes $\mathbf{Y}^{(h)}$, $\mathbf{E}^{(h)}$ e $\mathbf{\Delta}^{(h)}$. Com isso, aplica-se a análise SVD e estima-se o número ótimo de neurônios ocultos. Esse procedimento descrito acima é executado 100 vezes.

Como mencionado, no presente trabalho, a análise SVD foi adaptada a uma rede MLP de regressão com duas camadas ocultas, logo, existem seis matrizes em estudo: $\mathbf{Y}^{(h)}$, $\mathbf{E}^{(h)}$ e $\mathbf{\Delta}^{(h)}$ relacionadas à primeira camada oculta e $\mathbf{Y}^{(h)}$, $\mathbf{E}^{(h)}$ e $\mathbf{\Delta}^{(h)}$ relacionadas à segunda camada oculta.

Seja uma rede neural MLP de duas camadas, conforme já apresentado e discutido na Figura 15, com N vetores de treinamento, sendo cada vetor de dimensão P , número de neurônios ocultos da primeira camada oculta igual a Q , número de neurônios ocultos da segunda camada oculta igual a M e O neurônios de saída, pode-se mostrar nas seguintes equações como as matrizes buscadas são definidas.

Em relação à primeira camada oculta:

- $\mathbf{Y}^{(h)}[1^\circ \text{camada}]$ é uma matriz que armazena os *outputs* dos Q neurônios ocultos da primeira camada oculta, computados para os N dados de treinamento.
- Já $\mathbf{E}^{(h)}[1^\circ \text{camada}]$ é uma matriz que guarda os erros de retropropagação associados a esses Q neurônios ocultos, computando todos os N dados de treinamento.
- Enquanto $\mathbf{\Delta}^{(h)}[1^\circ \text{camada}]$ armazena os gradientes locais desses Q neurônios da primeira camada oculta também para todos os N dados de treinamento.

Conforme comentado anteriormente nas discussões sobre a Figura 15, sejam as matrizes $\mathbf{W}^{(1)} \in \mathbb{R}^{Q \times P}$, $\mathbf{b}^{(1)} \in \mathbb{R}^{Q \times 1}$ e $\mathbf{X} \in \mathbb{R}^{P \times 1}$, define-se uma matriz $\tilde{\mathbf{W}}^{(1)} \in \mathbb{R}^{Q \times (P+1)}$, conforme Equação 2.64, como sendo uma junção das matrizes $\mathbf{W}^{(1)}$ e $\mathbf{b}^{(1)}$, e uma matriz $\tilde{\mathbf{X}}^* \in \mathbb{R}^{(P+1) \times N}$, conforme Equação 2.65, como sendo uma representação adaptada de

todos os vetores de entrada \mathbf{X} de todos os N dados de treinamento.

$$\tilde{\mathbf{W}}^{(1)} = [\mathbf{b}^{(1)} \mid \mathbf{W}^{(1)}] = \begin{bmatrix} b_1^{(1)} & w_{1,1}^{(1)} & w_{1,2}^{(1)} & \cdots & w_{1,P}^{(1)} \\ b_2^{(1)} & w_{2,1}^{(1)} & w_{2,2}^{(1)} & \cdots & w_{2,P}^{(1)} \\ \vdots & \vdots & \vdots & & \vdots \\ b_Q^{(1)} & w_{Q,1}^{(1)} & w_{Q,2}^{(1)} & \cdots & w_{Q,P}^{(1)} \end{bmatrix} \quad (2.64)$$

$$\tilde{\mathbf{X}}^* = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1[1] & x_1[2] & \cdots & x_1[N] \\ x_2[1] & x_2[2] & \cdots & x_2[N] \\ \vdots & \vdots & & \vdots \\ x_P[1] & x_P[2] & \cdots & x_P[N] \end{bmatrix} \quad (2.65)$$

A matriz $\mathbf{Y}^{(h)}[1^\circ \text{camada}] \in \mathbb{R}^{Q \times N}$ é então definida como a aplicação da função de ativação da primeira camada oculta nessas matrizes definidas acima:

$$\mathbf{Y}^{(h)}[1^\circ \text{camada}] = \varphi_{(1)} [\tilde{\mathbf{W}}^{(1)} \tilde{\mathbf{X}}^*] = \begin{bmatrix} y_1^{(1)}[1] & y_1^{(1)}[2] & \cdots & y_1^{(1)}[N] \\ y_2^{(1)}[1] & y_2^{(1)}[2] & \cdots & y_2^{(1)}[N] \\ \vdots & \vdots & & \vdots \\ y_Q^{(1)}[1] & y_Q^{(1)}[2] & \cdots & y_Q^{(1)}[N] \end{bmatrix} \quad (2.66)$$

Através das matrizes $\mathbf{W}^{(2)} \in \mathbb{R}^{M \times Q}$ (matriz de pesos sinápticos associada aos M neurônios da segunda camada oculta), $\mathbf{b}^{(2)} \in \mathbb{R}^{M \times 1}$ (vetor de bias associado à segunda camada oculta), define-se a matriz $\tilde{\mathbf{W}}^{(2)} \in \mathbb{R}^{M \times (Q+1)}$ conforme Equação 2.67.

$$\tilde{\mathbf{W}}^{(2)} = [\mathbf{b}^{(2)} \mid \mathbf{W}^{(2)}] = \begin{bmatrix} b_1^{(2)} & w_{1,1}^{(2)} & w_{1,2}^{(2)} & \cdots & w_{1,Q}^{(2)} \\ b_2^{(2)} & w_{2,1}^{(2)} & w_{2,2}^{(2)} & \cdots & w_{2,Q}^{(2)} \\ \vdots & \vdots & \vdots & & \vdots \\ b_M^{(2)} & w_{M,1}^{(2)} & w_{M,2}^{(2)} & \cdots & w_{M,Q}^{(2)} \end{bmatrix} \quad (2.67)$$

Seja uma matriz $\Delta^{(2)} \in \mathbb{R}^{M \times N}$ que armazena os gradientes locais dos M neurônios ocultos da segunda camada para todos os N dados de treinamento, conforme Equação 2.70, onde $\delta_i^{(2)}[j]$ é o gradiente local do i -ésimo neurônio da segunda camada oculta do j -ésimo vetor de dado de treinamento. Define-se a matriz $\mathbf{E}^{(h)}[1^\circ \text{camada}] \in \mathbb{R}^{(Q+1) \times N}$ segundo Equação 2.69.

$$\Delta^{(2)} = \begin{bmatrix} \delta_1^{(2)}[1] & \delta_1^{(2)}[2] & \cdots & \delta_1^{(2)}[N] \\ \delta_2^{(2)}[1] & \delta_2^{(2)}[2] & \cdots & \delta_2^{(2)}[N] \\ \vdots & \vdots & & \vdots \\ \delta_M^{(2)}[1] & \delta_M^{(2)}[2] & \cdots & \delta_M^{(2)}[N] \end{bmatrix} \quad (2.68)$$

$$\mathbf{E}^h[1^\circ \text{camada}] = \left[\tilde{\mathbf{W}}^{(2)T} \mathbf{\Delta}^{(2)} \right] = \begin{bmatrix} e_0^{(1)}[1] & e_0^{(1)}[2] & \cdots & e_0^{(1)}[N] \\ e_1^{(1)}[1] & e_1^{(1)}[2] & \cdots & e_1^{(1)}[N] \\ \vdots & \vdots & & \vdots \\ e_Q^{(1)}[1] & e_Q^{(1)}[2] & \cdots & e_Q^{(1)}[N] \end{bmatrix} \quad (2.69)$$

E, por último, define-se a matriz $\mathbf{\Delta}^h[1^\circ \text{camada}] \in \mathbb{R}^{Q \times N}$ que armazena os gradientes locais dos Q neurônios ocultos da primeira camada para todos os N dados de treinamento conforme Equação 2.70, sendo $\delta_i^{(1)}[j]$ o gradiente local do i -ésimo neurônio da primeira camada oculta do j -ésimo vetor de dado de treinamento.

$$\mathbf{\Delta}^h[1^\circ \text{camada}] = \begin{bmatrix} \delta_1^{(1)}[1] & \delta_1^{(1)}[2] & \cdots & \delta_1^{(1)}[N] \\ \delta_2^{(1)}[1] & \delta_2^{(1)}[2] & \cdots & \delta_2^{(1)}[N] \\ \vdots & \vdots & & \vdots \\ \delta_Q^{(1)}[1] & \delta_Q^{(1)}[2] & \cdots & \delta_Q^{(1)}[N] \end{bmatrix} \quad (2.70)$$

Já para as matrizes relacionadas à segunda camada oculta, é possível realizar uma análise dimensional e estimar diversas informações, por exemplo: se $\mathbf{Y}^{(h)}[1^\circ \text{camada}]$ tem dimensão $\mathbb{R}^{Q \times N}$, pressupõe-se que $\mathbf{Y}^{(h)}[2^\circ \text{camada}]$ tenha dimensão $\mathbb{R}^{M \times N}$, como de fato ocorre. Além disso, se $\mathbf{Y}^{(h)}[1^\circ \text{camada}]$ é formada pela aplicação da função de ativação da primeira camada oculta $\varphi_{(1)}$ entre as matrizes $\tilde{\mathbf{W}}^{(1)}$ (matriz que armazena pesos sinápticos e bias associados aos Q neurônios da primeira camada oculta) e $\tilde{\mathbf{X}}^*$ (matriz adaptada que armazena os sinais de entrada da primeira camada oculta da rede), pressupõe-se que $\mathbf{Y}^{(h)}[2^\circ \text{camada}]$ é calculada pela aplicação da função de ativação da segunda camada oculta $\varphi_{(2)}$ entre matrizes $\tilde{\mathbf{W}}^{(2)}$ (matriz que armazena pesos sinápticos e bias associados aos M neurônios da segunda camada oculta) e $\mathbf{Y}^{(h)*}[1^\circ \text{camada}]$ (matriz adaptada que armazena os sinais de entrada da segunda camada oculta da rede).

Sendo assim, seja $\mathbf{Y}^{(h)*}[1^\circ \text{camada}] \in \mathbb{R}^{(Q+1) \times N}$ uma adaptação da matriz $\mathbf{Y}^{(h)}[1^\circ \text{camada}]$, conforme Equação 2.71, pode-se definir $\mathbf{Y}^{(h)}[2^\circ \text{camada}] \in \mathbb{R}^{M \times N}$ seguindo a Equação 2.72.

$$\mathbf{Y}^{(h)*}[1^\circ \text{camada}] = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ y_1^{(1)}[1] & y_1^{(1)}[2] & \cdots & y_1^{(1)}[N] \\ y_2^{(1)}[1] & y_2^{(1)}[2] & \cdots & y_2^{(1)}[N] \\ \vdots & \vdots & & \vdots \\ y_Q^{(1)}[1] & y_Q^{(1)}[2] & \cdots & y_Q^{(1)}[N] \end{bmatrix} \quad (2.71)$$

$$\mathbf{Y}^{(h)}[2^\circ \text{camada}] = \varphi_{(2)} \left[\tilde{\mathbf{W}}^{(2)} \mathbf{Y}^{(h)*}[1^\circ \text{camada}] \right] = \begin{bmatrix} y_1^{(2)}[1] & y_1^{(2)}[2] & \cdots & y_1^{(2)}[N] \\ y_2^{(2)}[1] & y_2^{(2)}[2] & \cdots & y_2^{(2)}[N] \\ \vdots & \vdots & & \vdots \\ y_M^{(2)}[1] & y_M^{(2)}[2] & \cdots & y_M^{(2)}[N] \end{bmatrix} \quad (2.72)$$

Seja a matriz $\tilde{\mathbf{W}}^{(3)} \in \mathbb{R}^{O \times (M+1)}$ formada pela junção das matrizes $\mathbf{W}^{(3)} \in \mathbb{R}^{O \times M}$ (matriz de pesos sinápticos associada aos O neurônios de saída) e $\mathbf{b}^{(3)} \in \mathbb{R}^{O \times 1}$ (vetor de bias associado à camada de saída), conforme Equação 2.73. Além disso, seja uma matriz $\Delta^{(3)} \in \mathbb{R}^{O \times N}$ que armazena os gradientes locais dos O neurônios de saída para todos os N dados de treinamento, conforme Equação 2.74, onde $\delta_i^{(3)}[j]$ é o gradiente local do i -ésimo neurônio da camada de saída do j -ésimo vetor de dado de treinamento. Define-se então a matriz $\mathbf{E}^{(h)}[2^\circ \text{camada}] \in \mathbb{R}^{(M+1) \times N}$ segundo Equação 2.75.

$$\tilde{\mathbf{W}}^{(3)} = \left[\mathbf{b}^{(3)} \mid \mathbf{W}^{(3)} \right] = \begin{bmatrix} b_1^{(3)} & w_{1,1}^{(3)} & w_{1,2}^{(3)} & \cdots & w_{1,M}^{(3)} \\ b_2^{(3)} & w_{2,1}^{(3)} & w_{2,2}^{(3)} & \cdots & w_{2,M}^{(3)} \\ \vdots & \vdots & \vdots & & \vdots \\ b_O^{(3)} & w_{O,1}^{(3)} & w_{O,2}^{(3)} & \cdots & w_{O,M}^{(3)} \end{bmatrix} \quad (2.73)$$

$$\Delta^{(3)} = \begin{bmatrix} \delta_1^{(3)}[1] & \delta_1^{(3)}[2] & \cdots & \delta_1^{(3)}[N] \\ \delta_2^{(3)}[1] & \delta_2^{(3)}[2] & \cdots & \delta_2^{(3)}[N] \\ \vdots & \vdots & & \vdots \\ \delta_O^{(3)}[1] & \delta_O^{(3)}[2] & \cdots & \delta_O^{(3)}[N] \end{bmatrix} \quad (2.74)$$

$$\mathbf{E}^{(h)}[2^\circ \text{camada}] = \left[\tilde{\mathbf{W}}^{(3)T} \Delta^{(3)} \right] = \begin{bmatrix} e_0^{(2)}[1] & e_0^{(2)}[2] & \cdots & e_0^{(2)}[N] \\ e_1^{(2)}[1] & e_1^{(2)}[2] & \cdots & e_1^{(2)}[N] \\ \vdots & \vdots & & \vdots \\ e_M^{(2)}[1] & e_M^{(2)}[2] & \cdots & e_M^{(2)}[N] \end{bmatrix} \quad (2.75)$$

A matriz $\Delta^h[2^\circ \text{camada}] \in \mathbb{R}^{M \times N}$ que armazena os gradientes locais dos M neurônios ocultos da segunda camada para todos os N dados de treinamento já foi definida pela Equação 2.70. Dessa forma:

$$\Delta^h[2^\circ \text{camada}] = \Delta^{(2)} = \begin{bmatrix} \delta_1^{(2)}[1] & \delta_1^{(2)}[2] & \cdots & \delta_1^{(2)}[N] \\ \delta_2^{(2)}[1] & \delta_2^{(2)}[2] & \cdots & \delta_2^{(2)}[N] \\ \vdots & \vdots & & \vdots \\ \delta_M^{(2)}[1] & \delta_M^{(2)}[2] & \cdots & \delta_M^{(2)}[N] \end{bmatrix} \quad (2.76)$$

3 Metodologia

Este capítulo visa apresentar a metodologia utilizada no trabalho, informando características dos dados e seu processamento, forma de abordagem do experimentos computacionais, suas validações, e métodos de comparação.

3.1 DADOS

3.1.1 Aquisição

A aquisição dos dados ocorreu na planta de sinterização de uma indústria siderúrgica brasileira entre os anos de 2018 e 2019, compreendendo 19 meses de processo. O conjunto de dados compreende informações sobre (i) parâmetros da máquina de sinterização, (ii) variáveis químicas do sínter e (iii) indicadores de qualidade física e química do sínter produzido. Dessa forma, bancos de dados de produção e qualidade foram compilados com base no número amostral e horários de aquisição das amostras.

Parâmetros de processo da máquina de sinterização são obtidos a cada hora, totalizando 24 instâncias ao longo de um dia. Variáveis químicas do sínter são averiguadas a cada quatro horas, totalizando 6 aferições ao longo de um dia de processo. Dos quatro índices de qualidade final do sínter obtidos pela empresa, o SI e o MPS são obtidos a cada quatro horas, totalizando 6 observações em um dia de produção, enquanto o RI e o RDI são obtidos uma vez ao dia.

Os dados foram disponibilizados pela indústria de forma que cada instância representava um dia de produção, isto é, as variáveis com mais de uma aferição por dia foram sumarizadas através de uma média aritmética.

3.1.2 Pré-processamento

Considerando que as observações foram reunidas e apresentadas em uma única tabela de dados, a etapa de pré-processamento resume-se em detectar dados inconsistentes e falta de registro de informações, e realizar a limpeza e redução de variáveis.

Ao todo, 527 instâncias foram disponibilizadas, das quais 101 amostras apresentaram problemas de incoerência de valores (valores não condizentes com a realidade) ou informações faltantes. Dessa forma, 426 instâncias foram utilizadas no trabalho.

Após uma seleção preliminar de variáveis realizada pela equipe de especialistas e

engenheiros da indústria siderúrgica onde os dados foram coletados, apenas as variáveis consideradas mais relevantes e de simples aquisição foram analisadas no presente trabalho. Aliado a isso, com base nas conclusões de trabalhos realizados com o mesmo objetivo de estimar indicadores de qualidade do sinter utilizando *machine learning*, [Kinnunen e Laitinen \(2005\)](#), [Donskoi et al. \(2007\)](#), [Laitinen e Saxén \(2007\)](#), [Umadevi et al. \(2016\)](#), [Kumar et al. \(2017\)](#), [Shao et al. \(2020\)](#), [Gao et al. \(2021\)](#), o número de variáveis de entrada principais e representativas ao problema foi reduzido a catorze. Além disso, como já mencionado, a referida indústria faz a aquisição de 4 índices principais de qualidade do sinter, que são os objetos de estudo do presente trabalho. As Tabelas 2 e 3 listam as variáveis de entrada e de saída disponíveis.

Tabela 2 – Variáveis de entrada e suas definições

Variáveis de entrada	Definição
FeO [%]	Óxido de ferro (wustita) presente no sinter
$T - Fe$ [%]	Ferro total presente no sinter
SiO_2 [%]	Óxido de silício presente no sinter
CaO [%]	Óxido de cálcio presente no sinter
Al_2O_3 [%]	Óxido de alumínio presente no sinter
MgO [%]	Óxido de magnésio presente no sinter
Basicidade Binária (CaO/SiO_2)	Razão entre os óxidos de cálcio e silício no sinter
Velocidade da Esteira [m/min]	Velocidade da máquina de sinterização
V.F.Q [m/min]	Velocidade da frente de queima
Camada de Ventilação [JPU]	Medida de permeabilidade do leito de sinterização
BTP [°C]	Temperatura na qual finaliza a queima do sinter
Altura BTP [m]	Posição da máquina onde a sinterização se concluiu
Temperatura [°C]	Temperatura do forno de ignição
Pressão do Resfriador [Mbar]	Pressão associada a taxa com a qual o sinter é resfriado

Fonte: elaborado pelo próprio autor.

Tabela 3 – Variáveis de saída e suas definições

Variáveis de saída	Definição
Índice de Resistência Shatter (SI) [%>10mm]	Resistência mecânica do sinter no teste Shatter
Índice de Redutibilidade (RI) [%]	Capacidade do sinter em ser reduzido no alto-forno
Índice de degradação sob redução (RDI) [%<2.83mm]	Geração de finos durante a redução no alto-forno
Tamanho médio de partícula (MPS) [mm]	Tamanho médio da partícula de sinter enviado ao alto-forno

Fonte: elaborado pelo próprio autor.

3.1.3 Estatística descritiva

As Tabelas 4 e 5 resumizam as estatísticas descritivas principais de todas as variáveis em estudo, informando média aritmética, desvio padrão, mínimo e máximo dos valores obtidos.

Tabela 4 – Estatísticas descritivas das variáveis de entrada.

Variáveis de entrada	Média	Desv. pad.	Mín.	Máx.	Assimetria ⁽¹⁾
% <i>FeO</i>	7,60	0,31	6,54	8,42	0,13
% <i>T – Fe</i>	57,97	0,61	56,59	59,31	0,54
% <i>SiO₂</i>	5,50	0,16	5,20	5,95	-0,19
% <i>CaO</i>	9,32	0,19	8,82	9,89	-0,37
% <i>Al₂O₃</i>	1,25	0,15	0,91	1,50	-0,53
% <i>MgO</i>	1,03	0,13	0,70	1,30	-0,31
Basicidade Binária (<i>CaO/SiO₂</i>)	1,70	0,05	1,59	1,83	0,20
Velocidade de esteira [m/min]	2,71	0,67	0,00	3,53	-0,66
V.F.Q [m/min]	17,86	2,15	9,08	21,09	-0,49
Camada de Ventilação [JPU]	28,51	2,94	18,74	37,78	-0,24
BTP [°C]	296,07	92,73	27,80	680,35	-0,87
Altura BTP [m]	88,09	8,26	5,87	97,47	-0,11
Temperatura [°C]	1122,26	163,62	74,32	1189,85	-0,34
Pressão do resfriador [Mbar]	11,78	20,24	0,08	68,16	0,42

⁽¹⁾ Coeficiente de assimetria de Pearson, conforme descrito em [Petrucci e Oliveira \(2017\)](#).

Fonte: elaborado pelo próprio autor.

Tabela 5 – Estatísticas descritivas das variáveis de saída.

Variáveis de saída	Média	Desv. pad.	Mín.	Máx.	Assimetria ⁽¹⁾
Índice de Resistência Shatter (SI) [%>10mm]	92,26	0,80	89,25	94,40	0,68
Índice de Redutibilidade (RI) [%]	59,63	2,07	55,00	64,00	-0,13
Índice de degradação sob redução (RDI) [%<2.83mm]	32,85	2,72	27,00	39,00	-0,09
Tamanho médio de partícula (MPS) [mm]	24,17	2,34	17,18	31,27	-0,08

⁽¹⁾ Coeficiente de assimetria de Pearson, conforme descrito em [Petrucci e Oliveira \(2017\)](#).

Fonte: elaborado pelo próprio autor.

3.2 EXPERIMENTOS COMPUTACIONAIS

Todos os algoritmos utilizados no presente trabalho foram escritos na linguagem MATLAB pelo próprio autor e processados com um computador de uso pessoal. As informações relevantes do hardware utilizado são: processador Intel Core i5-8400 2.80GHz, memória RAM 8.00GB, SSD 256GB.

De forma geral, as estatísticas de desempenho apresentadas ao longo deste trabalho fazem referência aos cálculos provenientes de dados do conjunto de teste, ou seja, as etapas de treinamento e/ou validação foram utilizadas única e exclusivamente para adequação de coeficientes e pesos sinápticos (ajuste do modelo de previsão). Esses conjuntos, para fins de cálculo das índices de comparação entre modelos, foram divididos entre 70% para treinamento e 30% para teste.

Para cada experimento de cada configuração dos modelos testados foram executadas 100 rodadas distintas com divisão aleatória dos conjuntos de treinamento e teste para busca de uma melhor confiabilidade estatística dos resultados. Contudo, é prática usual nos artigos de previsão comparar os melhores modelos encontrados, de modo a apresentar somente a melhor rodada que gerou, nos dados de teste, os melhores resultados. Entende-se que a rodada com melhor performance de teste gerou os coeficientes ou pesos sinápticos mais bem otimizados do modelo de previsão e que poderiam ser aplicados no processo real investigado. Assim, no presente trabalho optou-se por mostrar tanto os indicadores médios das 100 rodadas de execução quanto o resultado mais bem otimizado. Porém, como esse último pode ser comparado com outros trabalhos de previsão, um maior destaque foi dado a ele.

Optou-se por criar modelos de previsão individuais para fins de melhor comparação com trabalhos de outros autores, já que facilitam tanto a obtenção de modelos similares quanto permitem focar no indicador específico. Dessa forma, para cada algoritmo e método apresentado, quatro algoritmos preditores distintos foram criados e otimizados.

3.2.1 Regressão linear múltipla

Como mencionado no Capítulo anterior, o presente trabalho busca utilizar as regressões lineares como forma de parâmetro inicial de comparação. Três modelos principais foram utilizados:

- Regressão linear múltipla completa - todas as 14 variáveis disponíveis fazem parte do modelo de previsão.
- Regressão linear múltipla após análise do FIV - variáveis com maior FIV são excluídas do modelo.
- Regressão linear múltipla *stepwise* - através de metodologia *stepwise*, variáveis consideradas irrelevantes são excluídas do modelo.

Para entrada de dados no modelo de regressão, realizou-se redimensionamento conforme a Equação 2.8, com z_{max} igual a 1 e z_{min} igual a -1.

3.2.2 MLP - GDM

Outro modelo de previsão consiste em uma rede neural *percéptrons* múltiplas camadas (*multilayer perceptrons*) com algoritmo de aprendizagem supervisionada do tipo retropropagação de erro (*error backpropagation*) com gradiente descendente com *momentum*.

As configurações de rede estudadas consideram os seguintes hiperparâmetros: número de neurônios da primeira camada, número de neurônios da segunda camada, função de ativação da primeira camada, função de ativação da segunda camada, e constante *momentum*, conforme tabela abaixo:

Tabela 6 – Hiperparâmetros considerados para as configurações da rede MLP-GDM.

Hiperparâmetros	Valores
Número de neurônios da 1ª camada oculta	{5, 6, ..., 13, 14} ⁽¹⁾
Número de neurônios da 2ª camada oculta	{5, 6, ..., 13, 14} ⁽¹⁾
Função de ativação das camadas ocultas	{tanh, logsig}
Função de ativação das camadas de saída	{tanh, logsig}
Parâmetros a, b e c	a = {1, 1.7159} ⁽²⁾
	b = {1, 2/3} ⁽²⁾
	c = {1, 2.2} ⁽²⁾
Constante <i>Momentum</i> (γ)	{0.90, 0.95} ⁽³⁾

⁽¹⁾ O limite inferior do número de neurônios foi definido como 5 devido nos testes iniciais valores menores apresentarem resultados insatisfatórios de previsão, já o limite superior foi definido de acordo com o número máximo de variáveis de entrada, 14.

⁽²⁾ a , b e c são constantes mostradas nas Equações 2.23, 2.24, 2.25 e 2.26, conforme Haykin et al. (2009).

⁽³⁾ γ é mostrada na Equação 2.44, conforme Haykin et al. (2009).

Fonte: elaborado pelo próprio autor.

Tanto a taxa de aprendizado ($\alpha = 0.1$) quanto a divisão de treinamento|teste (80%|20%) foram mantidas constantes.

Conforme Haykin et al. (2009), os dados de treinamento são subdivididos em dados para estimação e dados para validação. O coeficiente de partição dos dados de treinamento entre esses dois conjuntos é chamado de r e é definido conforme Equação 3.1, onde W é o número de parâmetros livres da rede (conexões sinápticas).

$$r = 1 - \frac{1}{\sqrt{2W}} \quad (3.1)$$

W , por sua vez, é definido na Equação 3.2 para uma rede MLP com 2 camadas ocultas, onde p é a dimensão do vetor de entrada (número de variáveis de entrada), q_1 é o número de neurônios da 1ª camada oculta, q_2 é o número de neurônios da 2ª camada oculta e m é a dimensão do vetor de saída (número de variáveis de saída).

$$W = (p + 1)q_1 + (q_1 + 1)q_2 + (q_2 + 1)m \quad (3.2)$$

A consequência direta dessas equações é que, apesar da divisão treinamento|teste permanecer constante, o tamanho dos conjuntos para estimação e validação que compõem os dados de treinamento são variáveis.

Para fins de entrada na rede, os dados são redimensionados conforme a Equação 2.8, com z_{max} igual a 1 e z_{min} variando entre -1 e 0, a depender da zona de atuação da função de ativação utilizada, conforme discutido no Item 2.3.4.3.

3.2.3 MLP - LM

Além do gradiente descendente com *momentum*, outro algoritmo de retropropagação também foi utilizado: o algoritmo Levenberg–Marquardt.

As configurações estudadas são similares as da MLP-GDM, contudo a constante *momentum* não faz parte desse modelo. Dessa forma, a Tabela 7 apresenta os hiperparâmetros estudados.

Tabela 7 – Hiperparâmetros considerados para as configurações da rede MLP-LM.

Hiperparâmetros	Valores
Número de neurônios da 1ª camada oculta	{5, 6, ..., 13, 14} ⁽¹⁾
Número de neurônios da 2ª camada oculta	{5, 6, ..., 13, 14} ⁽¹⁾
Função de ativação das camadas ocultas	{tanh, logsig}
Função de ativação das camadas de saída	{tanh, logsig}
	a = {1, 1.7159} ⁽²⁾
Parâmetros a, b e c	b = {1, 2/3} ⁽²⁾
	c = {1, 2.2} ⁽²⁾

⁽¹⁾ O limite inferior do número de neurônios foi definido como 5 devido nos testes iniciais valores menores apresentarem resultados insatisfatórios de previsão, já o limite superior foi definido de acordo com o número máximo de variáveis de entrada, 14.

⁽²⁾ a, b e c são constantes mostradas nas Equações 2.23, 2.24, 2.25 e 2.26, conforme Haykin et al. (2009).

Fonte: elaborado pelo próprio autor.

O valor do parâmetro de Levenberg–Marquardt (μ), mostrado na Equação 2.55, foi inicializado em 0.001, enquanto o fator θ foi adotado como 10. A divisão de treinamento|teste (80%|20%) foi mantida constante, bem como a técnica de divisão dos dados de treinamento em um conjunto de estimação e um conjunto de validação.

A entrada da rede também funciona como na MLP-GDM, ou seja, os dados são redimensionados conforme Equação 2.8, com z_{max} igual a 1 e z_{min} variando entre -1 e 0, a depender da zona de atuação da função de ativação utilizada, conforme discutido no Item 2.3.4.3.

3.2.4 Indicadores de comparação entre experimentos

Para fins de comparação entre os próprios modelos de previsão do trabalho e entre trabalhos de outros autores, como já mencionado, foram utilizados os indicadores: coeficiente de determinação (R^2), erro percentual absoluto médio (EPAM) e erro quadrático médio (EQM).

Todos os índices foram calculados com os dados na escala original, ou seja, mesmo que a resposta básica de um algoritmo seja normalizada, o presente trabalho faz o redimensionamento dos dados para melhor entendimento das respostas. Alguns projetos optam por verificar os resultados ainda normalizados, por exemplo, entre 0 e 1, trazendo alguns problemas para fins de comparação no EPAM e no EQM, pois essas medidas são sensíveis à escala dos dados. Contudo, como no cálculo do coeficiente de determinação (R^2) não há diferença entre dados normalizados ou não normalizados, optou-se por adotá-lo como principal forma de avaliação entre os modelos.

3.2.5 Análise SVD para otimização do número ótimo de neurônios

O procedimento descrito no item 2.4.2 foi executado 100 vezes, de modo a permitir maior confiabilidade estatística nos resultados do número ótimo de neurônios ocultos para cada uma das matrizes analisadas. O limiar γ' descrito na Equação 2.63 foi escolhido como 0.98.

3.2.6 Análise de sensibilidade do melhor modelo

A análise de sensibilidade do melhor modelo propõe encontrar as relações de causa e efeito existentes entre cada variável de entrada com a saída. Tratando-se de um problema com múltiplas entradas, a forma mais comum de executar o procedimento é adotar valores médios para todas as variáveis. A partir daí, executa-se um estudo específico da influência na modificação de uma variável distinta por vez na variação da propriedade final. Ao término do procedimento, espera-se obter as relações causa e efeito entre cada entrada com a saída procurada. Essa análise possibilita a discussão principal do trabalho e auxilia no entendimento geral dos resultados alcançados, assim como nos estudos de [Kinnunen e Laitinen \(2005\)](#), [Laitinen e Saxén \(2007\)](#), [Umadevi et al. \(2016\)](#), [Gao et al. \(2021\)](#).

4 Resultados e Discussões

Este Capítulo visa apresentar e discutir os principais resultados alcançados no trabalho, realizando comparações tanto entre os modelos de previsão obtidos quanto resultados de trabalhos de outros autores.

4.1 REGRESSÃO LINEAR MÚLTIPLA

Dos três algoritmos propostos no item 3.2.1, o primeiro consiste na simples regressão linear múltipla com todas as variáveis 14 disponíveis.

A Tabela 8 mostra os principais resultados obtidos para cada propriedade final estudada. Estão expostos os resultados tanto da média dos 100 modelos criados, quanto do melhor modelo alcançado (classificado pelo coeficiente de determinação).

Os maiores acertos foram observados na previsão do Índice de Resistência Shatter (SI) onde, no melhor modelo, foi obtido R^2 de 0.4806, EPAM de 0.5175% e EQM de 0.3793. As saídas Índice de Degradação Sob Redução (RDI) e Tamanho Médio de Partícula (MPS) foram as outras duas propriedades com melhores acertos, tendo alcançado nos melhores modelos, respectivamente, R^2 de 0.3105, EPAM de 5.8456% e EQM de 4.5127 e R^2 de 0.2836, EPAM de 6.6896% e EQM de 4.0746. O resultado para o Índice de Redutibilidade (RI) foi o mais baixo, com R^2 de 0.1436, 2.6032% de EPAM e 3.6952 de EQM para o melhor dos 100 modelos configurados.

Tabela 8 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão linear múltipla proposta.

Regressão linear múltipla	Melhor modelo			Média dos modelos		
	Saídas	R^2	EPAM	EQM	R^2	EPAM
Índice de Resistência Shatter (SI) [%>10mm]	0.4806	0.5175%	0.3793	0.3288	0.5455%	0.4226
Índice de Redutibilidade (RI) [%]	0.1436	2.6032%	3.6952	0.0232	2.7742%	4.1366
Índice de degradação sob redução (RDI) [%<2.83mm]	0.3105	5.8456%	4.9127	0.1461	6.1451%	6.2735
Tamanho médio de partícula (MPS) [mm]	0.2836	6.6896%	4.0746	0.1300	6.9186%	4.7754

Fonte: elaborado pelo próprio autor.

A partir desses resultados, com o auxílio de uma análise da matriz de correlação e do FIV, uma nova regressão linear múltipla foi executada, porém apenas com as variáveis mais relevantes de acordo com esses índices.

A análise de matriz de correlação por si só não indica necessariamente a presença de problemas nos dados por multicolinearidade, contudo, com o auxílio do FIV, é possível estimar essa influência.

A Figura 21 mostra o que é chamado de correlograma das variáveis de entrada, contendo a matriz de correlação, os histogramas individuais e o gráfico de dispersão entre as entradas.

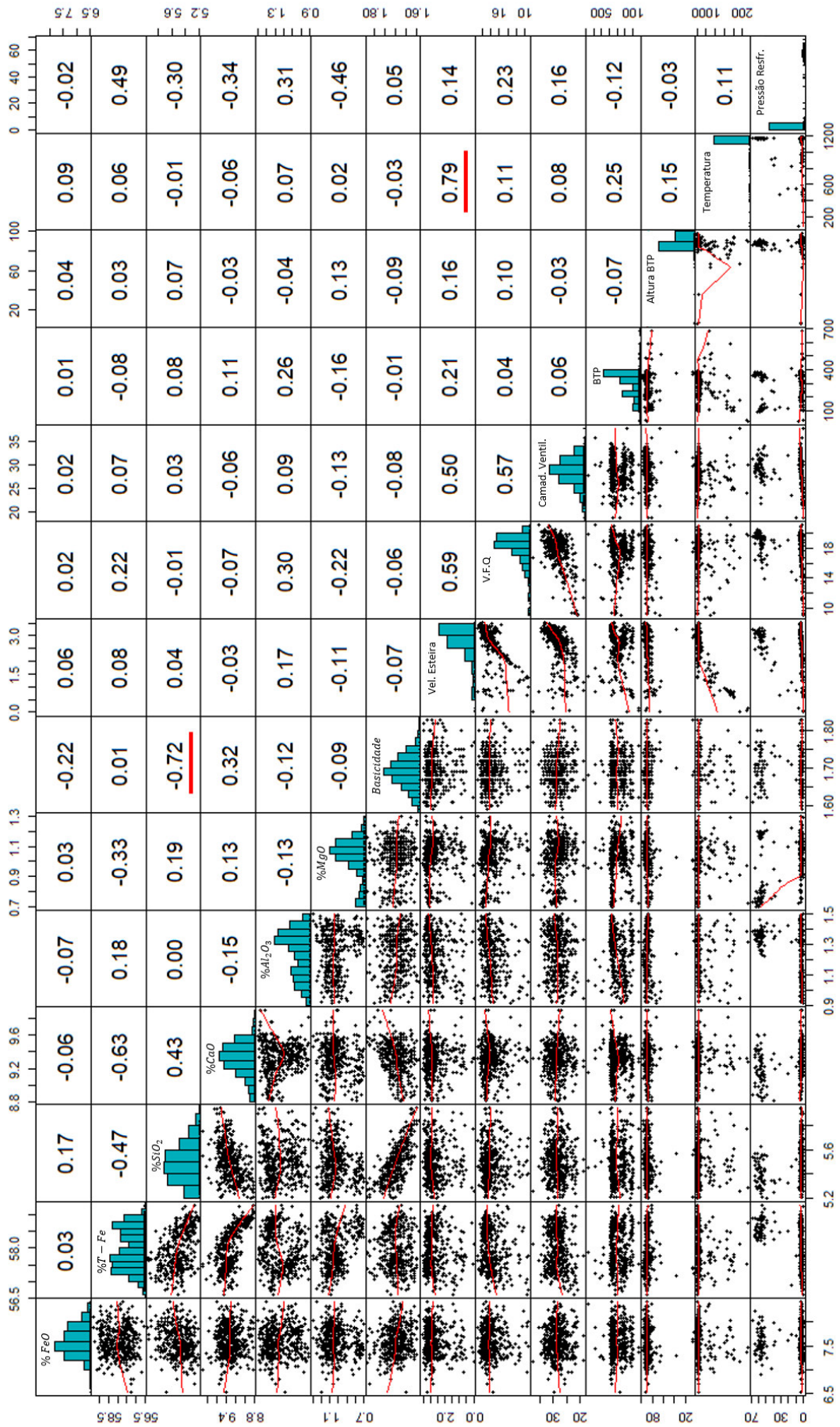
Algumas relações fortes, destacadas em vermelho, sugerem multicolinearidade entre os dados. O coeficiente de correlação entre a variável basicidade e a variável % SiO_2 é de -0.72, enquanto entre a temperatura do forno de ignição e a velocidade da esteira é de 0.79.

O FIV foi calculado conforme Equação 2.7 e está exposto na Figura 22. Observa-se um grande fator para % SiO_2 , % CaO , basicidade, velocidade da esteira e temperatura do forno de ignição. Como o FIV é um escalar, a interpretação é de que essas variáveis relacionam-se fortemente entre si e, por isso, a fim de melhorar a obtenção dos coeficientes de regressão na RLM, deve-se eliminar algumas delas.

Aliando as informações fornecidas pelo FIV e pela matriz de correlação com os conhecimentos teóricos da produção do sinter, pode-se dizer que:

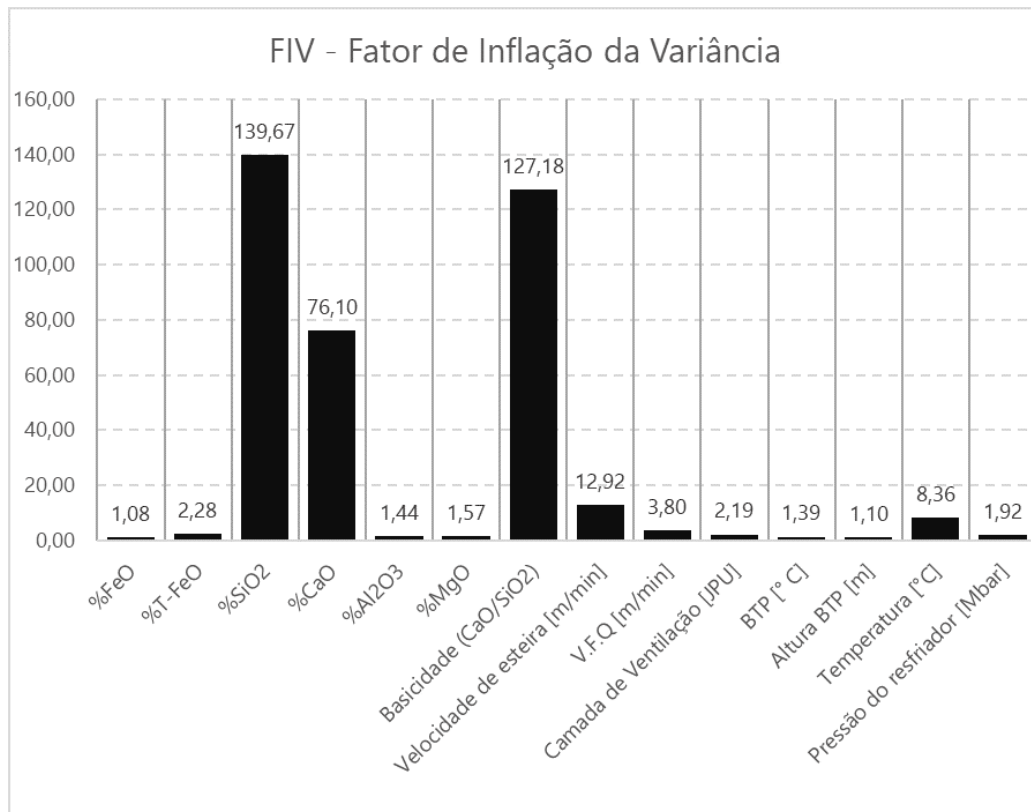
- A basicidade binária está diretamente relacionada com as variáveis % SiO_2 e % CaO ,

Figura 21 – Gráficos de dispersão, histogramas e coeficientes de correlação entre variáveis de entrada



Fonte: elaborado pelo próprio autor

Figura 22 – Representação gráfica do fator de inflação da variância.



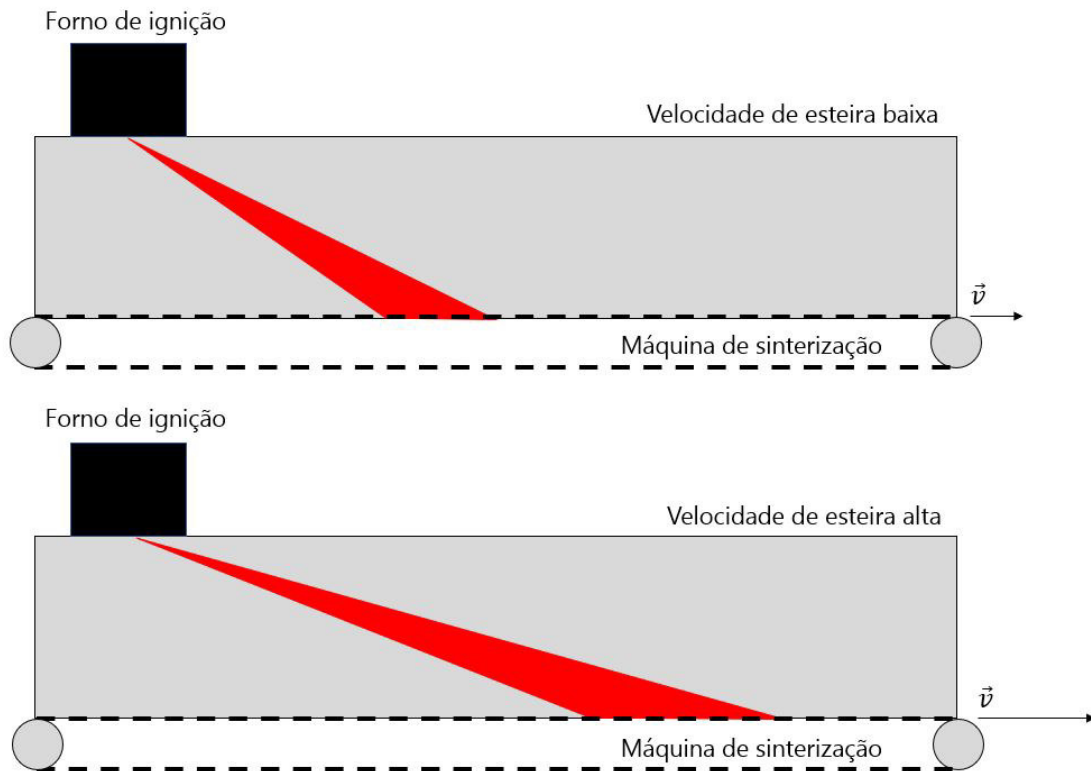
Fonte: elaborado pelo próprio autor

afinal, o seu cálculo leva em consideração o percentual desses compostos no sínter. Dessa forma, o procedimento adotado na próxima etapa é a retirada da variável basicidade binária, já que a informação fornecida por ela já está contida em %SiO₂ e %CaO.

- As variáveis temperatura e velocidade de esteira estão intimamente ligadas, pois, quanto maior a velocidade da esteira, maior deverá ser a temperatura do forno de ignição para fazer com que mais sínter seja produzido em menor tempo, ou seja, acelerar o ritmo do processo requer além de maior velocidade de esteira, uma maior energia proveniente do forno de ignição. Além disso, aumentar a velocidade da esteira induz a formação de uma frente de queima maior, e, para que a totalidade dessa nova camada tenha energia suficiente para realizar todas as reações necessárias à produção do sínter, é preciso fornecer mais energia. A Figura 23 mostra a influência da velocidade de esteira na extensão da frente de queima. O procedimento adotado foi retirar a variável velocidade de esteira, de maior FIV.

Uma nova regressão linear múltipla foi executada retirando-se duas variáveis do modelo: "basicidade binária" e "velocidade de esteira". A Tabela 9 mostra os principais resultados obtidos para cada propriedade final estudada. Assim como na Tabela 8, estão

Figura 23 – Influência da velocidade da esteira na frente de queima.



Fonte: elaborado pelo próprio autor

expostos os resultados tanto da média dos 100 modelos criados, quanto do melhor modelo alcançado (classificado pelo coeficiente de determinação). Após a eliminação das dessas variáveis observou-se uma melhor estimativa do Índice de Redutibilidade (RI) em relação ao resultado da regressão obtido com todas as 14 variáveis. O coeficiente de determinação R^2 passou de 0.1436 para 0.1810. As demais propriedades não tiveram diferenças significativas no coeficiente de determinação para os conjuntos de testes.

Tabela 9 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão linear múltipla (análise FIV) proposta.

Regressão linear múltipla (Análise FIV)	Melhor modelo			Média dos modelos		
	Saídas	R^2	EPAM	EQM	R^2	EPAM
Índice de Resistência Shatter (SI) [%>10mm]	0.4683	0.5093%	0.3893	0.3505	0.5438%	0.4061
Índice de Redutibilidade (RI) [%]	0.1810	2.5921%	3.5593	0.0418	2.7772%	4.0884
Índice de degradação sob redução (RDI) [%<2.83mm]	0.2918	6.1344%	5.9639	0.1557	6.1021%	6.2119
Tamanho médio de partícula (MPS) [mm]	0.2940	6.3095%	3.5357	0.1294	6.9186%	4.7600

Fonte: elaborado pelo próprio autor.

A última técnica utilizada para melhorar os resultados da regressão, a regressão *stepwise*, difere teoricamente da análise do FIV. Como já mencionado no item 2.2.5, enquanto o FIV calcula através de regressões auxiliares a influência de variáveis dependentes umas sobre as outras, a regressão *stepwise* atenta-se apenas em verificar de forma sistemática quais das variáveis dependentes fornecem os melhores resultados de previsão.

A regressão *stepwise* utilizada no presente trabalho continua sendo uma regressão linear múltipla na medida em que apenas adota um método sistemático para eliminar as variáveis preditoras menos importantes. O critério de classificação utilizado para adição ou remoção de variáveis foi o coeficiente de determinação.

O início da técnica é a configuração de uma RLM com uma única constante (sem levar em consideração variáveis) e, através de sucessivos passos, variáveis são adicionadas ou removidas do modelo final. O critério de adição adotado foi de aumentar o R^2 em mais que 0.5% e o de retirar uma variável foi de aumentar o R^2 em menos de 0.05%. Dessa forma, variáveis que contribuem significativamente no resultado final de previsão são adicionadas ao modelo ao passo que parâmetros que pouco influenciam no R^2 são deixados de lado. Vale ressaltar que, como o critério dos passos é o coeficiente de determinação, quanto mais variáveis o modelo contiver, analisando as relações das Equações 2.56, 2.57 e 2.58, maior seria o R^2 e, dessa forma, a regra de remoção de uma variável do modelo final foi formulada levando isso em consideração.

A Tabela 10 mostra os resultado da regressão *stepwise*. Para as propriedades SI e MPS foram encontrados os melhores coeficientes de determinação, respectivamente, 0.4987 e 0.3108.

Tabela 10 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo da regressão *stepwise* proposta.

Regressão <i>stepwise</i>	Melhor modelo			Média dos modelos		
	Saídas	R^2	EPAM	EQM	R^2	EPAM
Índice de Resistência Shatter (SI) [%>10mm]	0.4987	0.4829%	0.3013	0.3396	0.5462%	0.4234
Índice de Redutibilidade (RI) [%]	0.1490	2.6939%	3.8290	0.0236	2.7947%	4.1781
Índice de degradação sob redução (RDI) [%<2.83mm]	0.3023	5.8329%	5.2854	0.1492	6.1207%	6.2654
Tamanho médio de partícula (MPS) [mm]	0.3108	6.3160%	3.2900	0.1134	6.9598%	4.7868

Fonte: elaborado pelo próprio autor.

Tabela 11 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo dentre todas as técnicas de regressão linear utilizadas.

Saídas	Tipo de regressão	Melhor modelo		
		R^2	EPAM	EQM
Índice de Resistência Shatter (SI) [%>10mm]	Regressão <i>stepwise</i>	0.4987	0.4829%	0.3013
Índice de Redutibilidade (RI) [%]	Regressão Linear Múltipla [Análise FIV]	0.1810	2.5921%	3,5593
Índice de degradação sob redução (RDI) [%<2.83mm]	Regressão Linear Múltipla	0.3105	5.8456%	4.9127
Tamanho médio de partícula (MPS) [mm]	Regressão <i>stepwise</i>	0.3108	6.3160%	3.2900

A Tabela 11 mostra os melhores resultados alcançados em todas as técnicas de regressão. Em relação ao SI e MPS, a regressão *stepwise* foi a técnica que melhor se adaptou aos dados e possibilitou melhores coeficientes de determinação, enquanto no RI a análise FIV mostrou-se a mais eficiente na predição. A regressão linear múltipla com todas as 14 variáveis foi a que melhor se adaptou à propriedade RDI. Observa-se nos resultados uma tendência de melhor previsão no índice de qualidade SI, seguido de MPS e RDI. O índice RI mostra-se, nesses algoritmos lineares de previsão, de difícil previsibilidade com os dados disponíveis.

4.2 REDES NEURAIS ARTIFICIAIS

As Tabelas 12 e 13 mostram os 5 melhores resultados alcançados por cada rede MLP classificados de acordo com o coeficiente de determinação dos dados de teste. Assim como na RLM, são mostradas as médias dos indicadores R^2 , EPAM e EQM calculadas sobre 100 rodadas de teste e o melhor modelo obtido dentre elas.

Para facilitar as discussões de configurações de rede, o autor propõe a utilização da nomenclatura GDM - [Número de neurônios | Função de ativação das camadas ocultas | Função de ativação da camada de saída | Parâmetros a, b e c | Constante *Momentum* (γ)] para identificação do experimento nas redes MLP-GDM e de LM - [Número de neurônios | Função de ativação das camadas ocultas | Função de ativação da camada de saída | Parâmetros a, b e c] para a identificação do experimento nas redes MLP-LM.

Na MLP-GDM, os melhores resultados para o Índice de Resistência Shatter (SI) foram obtidos com uma rede GDM - [8 - 5 - 1 | tanh | tanh | 1, 1 e 0 | 0.90], alcançando R^2 de 0.7002 no melhor modelo. Para o Índice de Redutibilidade (RI), a configuração que alcançou os melhores resultados foi GDM - [8 - 9 - 1 | tanh | tanh | 1.7159, 2/3 e 0 | 0.90], obtendo no melhor modelo das 100 verificações R^2 de 0.2684. Já para o Índice de Degradação Sob Redução (RDI), a configuração GDM - [10 - 6 - 1 | tanh | tanh | 1, 1 e 0 | 0.90] foi a que melhor performou, com o melhor modelo obtendo R^2 de 0.4478. No último indicador, o Tamanho Médio de Partícula (MPS), a configuração GDM - [14 - 12 - 1 | logística | tanh | 1, 1 e 2.2 | 0.90] foi a melhor, com o modelo superior com R^2 de 0.5654.

Enquanto nas redes MLP-LM, os melhores resultados para o Índice de Resistência Shatter (SI) foram atingidos com uma rede GDM - [10 - 14 - 1 | logística | tanh | 1, 1 e 1], alcançando R^2 de 0.7410 no melhor modelo. Para o Índice de Redutibilidade (RI), a configuração que alcançou os melhores resultados foi GDM - [6 - 13 - 1 | logística | tanh | 1, 1 e 1], obtendo no melhor modelo das 100 verificações R^2 de 0.2660. Enquanto para o Índice de Degradação Sob Redução (RDI), a configuração GDM - [11 - 8 - 1 | tanh | tanh | 1.7159, 2/3 e 0] foi a que melhor performou, com o melhor modelo obtendo R^2 de 0.4388. No indicador Tamanho Médio de Partícula (MPS), a configuração GDM - [6 - 12 - 1 | tanh | tanh | 1.7159, 2/3 e 0] foi a superior, com o melhor modelo com R^2 de 0.6030.

Tabela 12 – Resumo dos melhores resultados de previsão de previsão das propriedades buscadas classificadas de acordo com o R^2 do melhor modelo da configuração de rede MLP - GDM proposta.

Saídas	Configuração da rede				Melhor modelo da configuração				Média dos modelos da configuração			
	Número de neurônios	Função das camadas ocultas	Função de saída	Parâmetros a, b e c	γ	R^2	EPAM	EQM	R^2	EPAM	EQM	
Índice de resistência shatter (SI) [%>10mm]	8 - 5 - 1	tanh	tanh	1, 1 e 0	0.90	0.7002	0.4146%	0.2112	0.4069	0.5041%	0.3663	
	9 - 10 - 1	tanh	tanh	1, 1 e 0	0.90	0.6942	0.3328%	0.1440	0.3886	0.5123%	0.3849	
	10 - 5 - 1	logística	tanh	1, 1 e 1	0.95	0.6902	0.3658%	0.1891	0.3012	0.5538%	0.4392	
	6 - 5 - 1	tanh	tanh	1, 1 e 0	0.95	0.6853	0.4017%	0.2248	0.3809	0.5154%	0.3861	
7 - 7 - 1	logística	tanh	tanh	1, 1 e 2.2	0.95	0.6803	0.3925%	0.2047	0.3938	0.5129%	0.3878	
Índice de redutibilidade (RI) [%]	8 - 9 - 1	tanh	tanh	1.7159, 2/3 e 0	0.90	0.2684	2.2995%	2.7159	0.0438	2.7900%	4.1087	
	12 - 8 - 1	logística	tanh	1, 1 e 1	0.90	0.2647	2.3066%	2.7296	0.0057	2.8331%	4.2107	
	11 - 5 - 1	tanh	logística	1.7159, 2/3 e 2.2	0.95	0.2540	2.3831%	3.0310	0.0177	2.7884%	4.0898	
	14 - 13 - 1	tanh	logística	1.7159, 2/3 e 2.2	0.90	0.2520	2.3092%	2.7765	0.0141	2.8284%	4.1791	
7 - 5 - 1	tanh	tanh	tanh	1.7159, 2/3 e 0	0.90	0.2517	2.6018%	3.4400	0.0356	2.8029%	4.1619	
Índice de degradação sob redução (RDI) [%<2.83mm]	10 - 6 - 1	tanh	tanh	1, 1 e 0	0.90	0.4478	5.7926%	4.8498	0.1060	6.3457%	6.5597	
	10 - 5 - 1	logística	tanh	1.7159, 2/3 e 2.2	0.95	0.4273	5.3687%	4.4796	0.0662	6.5082%	6.7839	
	13 - 14 - 1	logística	tanh	1, 1 e 2.2	0.95	0.4249	5.4536%	4.6506	0.0340	6.4939%	6.9788	
	11 - 10 - 1	tanh	logística	1, 1 e 1	0.95	0.4161	5.3919%	4.7216	0.1106	6.3601%	6.5193	
10 - 7 - 1	tanh	tanh	tanh	1, 1 e 0	0.95	0.4104	5.4865%	4.3999	0.0433	6.4530%	6.9817	
Tamanho médio de partícula (MPS) [mm]	14 - 12 - 1	logística	tanh	1, 1 e 2.2	0.90	0.5654	5.3116%	2.4198	0.1697	6.8067%	4.4373	
	10 - 14 - 1	logística	tanh	1, 1 e 2.2	0.90	0.5508	5.2225%	2.6153	0.1772	6.7941%	4.4316	
	13 - 10 - 1	tanh	tanh	1, 1 e 0	0.95	0.5469	5.0380%	2.4394	0.1916	6.6929%	4.3930	
	11 - 10 - 1	logística	tanh	1, 1 e 2.2	0.90	0.5431	5.3331%	2.5438	0.1679	6.8339%	4.4657	
13 - 8 - 1	tanh	tanh	tanh	1, 1 e 0	0.95	0.5414	5.0332%	2.5249	0.1718	6.8582%	4.5511	

Fonte: elaborado pelo próprio autor.

Tabela 13 – Resumo dos melhores resultados de previsão de resistência das propriedades buscadas classificadas de acordo com o R^2 do melhor modelo da configuração de rede MLP - LM proposta..

MLP - LM	Configuração da rede			Melhor modelo da configuração			Média dos modelos da configuração			
	Número de neurônios	Função das camadas ocultas	Função de saída	Parâmetros a, b e c	R^2	EPAM	EQM	R^2	EPAM	EQM
Saídas	10 - 14 - 1	logística	tanh	1, 1 e 1	0.7410	0.3651%	0.1792	0.3735	0.5164%	0.3882
Índice de resistência shatter (SI) [%>10mm]	9 - 14 - 1	tanh	tanh	1, 1 e 0	0.7297	0.4118%	0.2349	0.3563	0.5239%	0.3981
	6 - 12 - 1	tanh	tanh	1,7159, 2/3 e 0	0.6903	0.4117%	0.2066	0.3480	0.5228%	0.3992
	14 - 10 - 1	tanh	tanh	1, 1 e 0	0.6896	0.4565%	0.2740	0.3231	0.5492%	0.4373
	5 - 9 - 1	tanh	tanh	1, 1 e 0	0.6886	0.3983%	0.2103	0.3529	0.5251%	0.4033
	6 - 13 - 1	logística	tanh	1, 1 e 1	0.2660	2.5186%	3.2203	0.0398	2.9217%	4.5604
Índice de redutibilidade (RI) [%]	5 - 12 - 1	tanh	tanh	1,7159, 2/3 e 0	0.2314	2.3442%	3.2292	0.0257	2.9458%	4.6791
	9 - 5 - 1	tanh	tanh	1, 1 e 0	0.2232	2.4114%	3.3247	0.0167	2.9256%	4.5264
	8 - 14 - 1	tanh	tanh	1, 1 e 0	0.2067	2.6299%	3.6541	0.0139	2.9713%	4.6711
	14 - 12 - 1	logística	tanh	1, 1 e 2.2	0.2065	3.036%	4.4158	0.0146	2.9007%	4.5431
	11 - 8 - 1	tanh	tanh	1,7159, 2/3 e 0	0.4388	4.9452%	4.1073	0.0811	6.5684%	7.1170
Índice de degradação sob redução (RDI) [%<2.83mm]	10 - 11 - 1	tanh	tanh	1, 1 e 0	0.3903	5.2687%	4.6628	0.0427	6.6795%	7.2770
	14 - 13 - 1	tanh	tanh	1,7159, 2/3 e 0	0.3852	6.2349%	5.5172	0.0594	6.7569%	7.3992
	7 - 13 - 1	logística	tanh	1, 1 e 1	0.3719	5.3494%	4.7584	0.0452	6.6190%	7.0892
	9 - 5 - 1	tanh	tanh	1,7159, 2/3 e 0	0.3693	5.2180%	4.5835	0.0348	6.4645%	6.7987
	6 - 12 - 1	tanh	tanh	1,7159, 2/3 e 0	0.6030	4.9901%	2.1682	0.1923	6.6225%	4.273
Tamanho médio de partícula (MPS) [mm]	10 - 8 - 1	tanh	tanh	1,7159, 2/3 e 0	0.5922	6.0135%	3.1740	0.1782	6.7780%	4.3325
	13 - 6 - 1	tanh	tanh	1,7159, 2/3 e 0	0.5914	4.9788%	2.3542	0.1718	6.6837%	4.3292
	13 - 7 - 1	tanh	tanh	1,7159, 2/3 e 0	0.5878	5.2684%	2.1845	0.1714	6.8195%	4.4288
	8 - 14 - 1	logística	tanh	1, 1 e 1	0.5764	5.1625%	2.7292	0.1903	6.6901%	4.3414

Fonte: elaborado pelo próprio autor.

Tabela 14 – Resumo dos melhores resultados de previsão classificados de acordo com o R^2 do melhor modelo dentre todas as redes neurais utilizadas.

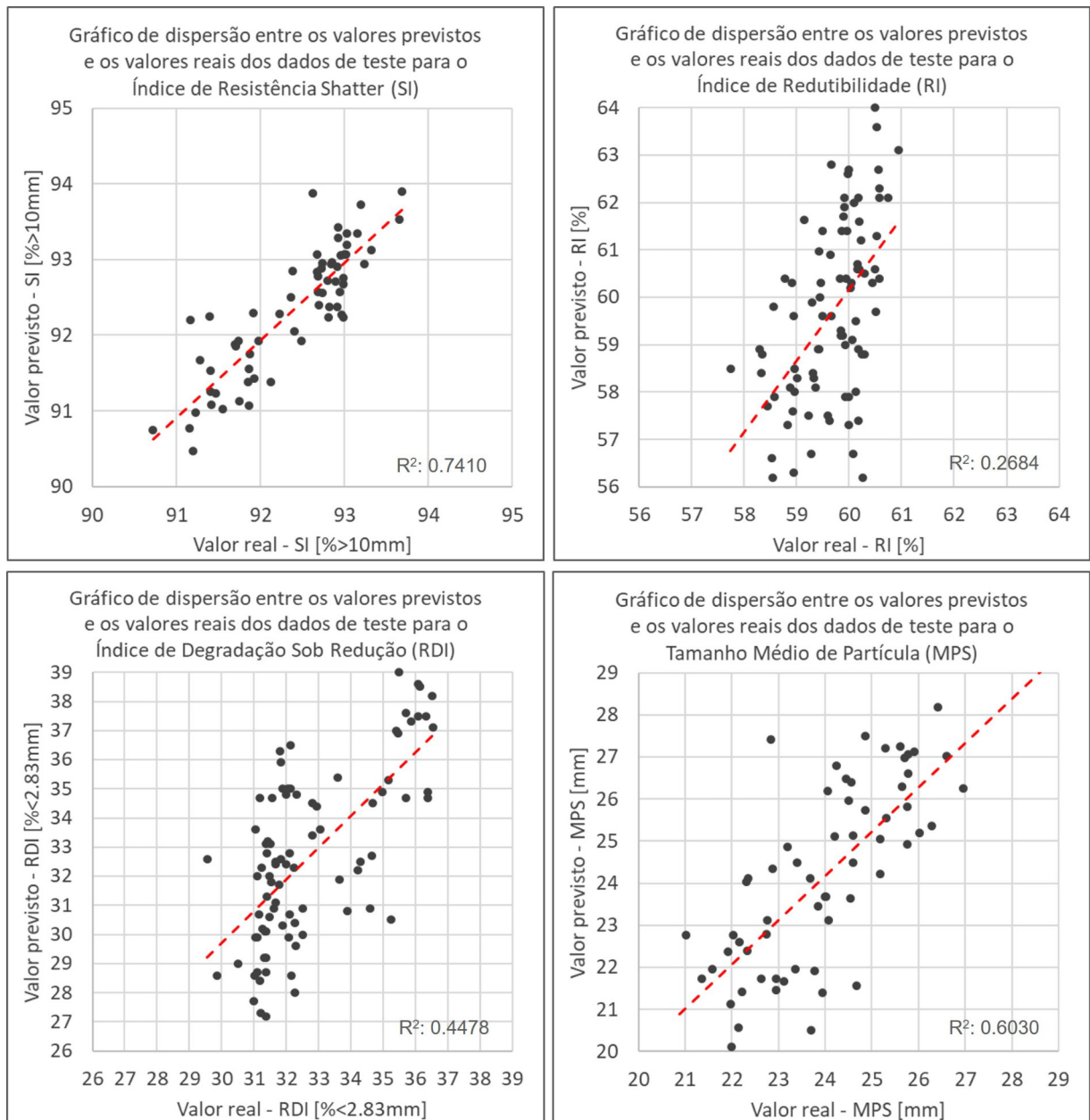
Saídas	Tipo de rede neural	Melhor modelo		
		R^2	EPAM	EQM
Índice de Resistência Shatter (SI) [%>10mm]	MLP - LM	0.7410	0.3651%	0.1792
Índice de Redutibilidade (RI) [%]	MLP - GDM	0.2684	2.2995%	2.7159
Índice de degradação sob redução (RDI) [%<2.83mm]	MLP - GDM	0.4478	5.7926%	4.8498
Tamanho médio de partícula (MPS) [mm]	MLP - LM	0.6030	4.9901%	2.1682

Fonte: elaborado pelo próprio autor.

A Tabela 14 mostra os melhores resultados alcançados dentre todas as redes testadas. Assim como nas análises de regressão lineares, o Índice de Resistência Shatter (SI) foi o indicador de qualidade mais corretamente predito, seguido do Tamanho Médio de Partícula (MPS) e pelo Índice de Degradação Sob Redução (RDI). O Índice de Redutibilidade (RI) também foi o índice que alcançou o pior desempenho.

A Figura 24 mostra os gráficos de dispersão dos dados observados (valores reais) e valores calculados pela rede com os melhores desempenhos verificados. Assim como pôde ser observado nas tabelas anteriores, nota-se melhor previsibilidade e menores erros associados à previsão do Índice de Resistência Shatter (SI).

Figura 24 – Gráficos de dispersão dos melhores resultados entre os valores previstos e os valores reais dos dados de teste para cada indicador de qualidade do sínter estudado.



Fonte: elaborado pelo próprio autor

4.3 COMPARAÇÃO DE RESULTADOS

4.3.1 Comparação entre MLP e RLM

As Tabelas 15, 16 e 17 mostram a comparação entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.

Tabela 15 – Resumo da diferença absoluta de performance no indicador R^2 entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.

Saídas	R^2		
	MLP	RLM	(MLP - RLM)
Índice de Resistência Shatter (SI) [%>10mm]	0.7410	0.4987	+0.2423
Índice de Redutibilidade (RI) [%]	0.2684	0.1810	+0.0874
Índice de degradação sob redução (RDI) [%<2.83mm]	0.4478	0.3105	+0.1373
Tamanho médio de partícula (MPS) [mm]	0.6030	0.3108	+0.2922

Fonte: elaborado pelo próprio autor.

Tabela 16 – Resumo da diferença absoluta de performance no indicador EPAM entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.

Saídas	EPAM		
	MLP	RLM	(MLP - RLM)
Índice de Resistência Shatter (SI) [%>10mm]	0.3651%	0.4829%	-0.1178%
Índice de Redutibilidade (RI) [%]	2.2995%	2.5921%	-0.2926%
Índice de degradação sob redução (RDI) [%<2.83mm]	5.7926%	5.8456%	-0.0530%
Tamanho médio de partícula (MPS) [mm]	4.9901%	6.3160%	-1.3259%

Fonte: elaborado pelo próprio autor.

Tabela 17 – Resumo da diferença absoluta de performance no indicador EQM entre os melhores resultados das regressões lineares e das redes neurais para os dados de teste.

Saídas	EQM		
	MLP	RLM	(MLP - RLM)
Índice de Resistência Shatter (SI) [%>10mm]	0.1792	0.3013	-0.1221
Índice de Redutibilidade (RI) [%]	2.7159	3.5593	-0.8434
Índice de degradação sob redução (RDI) [%<2.83mm]	4.8498	4.9127	-0.0629
Tamanho médio de partícula (MPS) [mm]	2.1682	3.2900	-1.1218

Fonte: elaborado pelo próprio autor.

Todos os indicadores de desempenho demonstraram que as redes neurais alcançaram melhor desempenho na previsão dos 4 índices de qualidade do síter. A inferioridade dos resultados obtidos pela regressão linear pode ser justificada pelas complexas relações entre os dados de entradas e saídas do processo de sinterização que não podem ser descritas/reproduzidas por relações lineares. Dessa forma, é necessário um método mais abrangente como as redes neurais que se utilizam de funções de ativação, estrutura e técnica de solução mais complexas e não lineares.

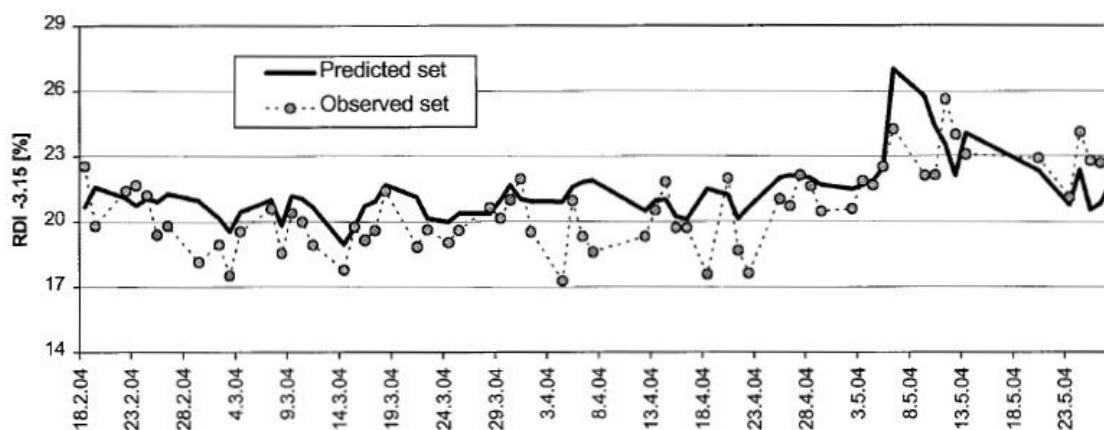
4.3.2 Comparação com outros autores

Os resultados encontrados no presente trabalho podem ser comparados a diversos outros estudos sobre previsão de índices de qualidade do síter. Uma importante observação desses estudos é que a maioria opta por utilizar um indicador de resistência mecânica do síter chamado Índice de Tamboreamento (TI), parâmetro esse não disponível nos dados utilizados nesta pesquisa. Contudo, diante da proximidade existente entre o cálculo de TI e o cálculo do Índice de Resistência Shatter (SI), ambos medidores da resistência mecânica do produto, o presente trabalho considera que os resultados preditivos alcançados por esses índices podem ser equiparáveis. Dentre as principais pesquisas nessa área, destacam-se:

- [Kinnunen e Laitinen \(2005\)](#) : Sendo um dos primeiros estudos sobre propriedades e produtividade do síter metalúrgico na área de *machine learning*, essa pesquisa abrangeu 695 instâncias referentes ao período de 2001 a 2004 de uma planta siderúrgica, incluindo até 18 variáveis de entrada na busca do Índice de Degradação Sob

Redução (RDI) e da taxa de produtividade. Apesar dos autores não detalharem a construção do algoritmo utilizado, é possível verificar a criação de redes MLP de camada única com 3 a 7 neurônios ocultos. Os resultados são fornecidos apenas com o Erro Quadrático Médio (EQM) dos dados de teste (dados normalizados) igual a 0.25 para o RDI. Como dados normalizados alteram o referencial e inviabilizam comparação com o presente trabalho, que alcançou EQM de 4.8498 em dados não normalizados, a confrontação das respostas apenas pode ser realizada com os gráficos em série disponibilizados pelos autores. A Figura 25 mostra um comparativo entre os dados previstos pela rede (linha sólida) e os dados reais (linha pontilhada), sendo possível identificar que a previsão do algoritmo, apesar de ruidosa, segue as tendências de aumento e de diminuição dos dados reais.

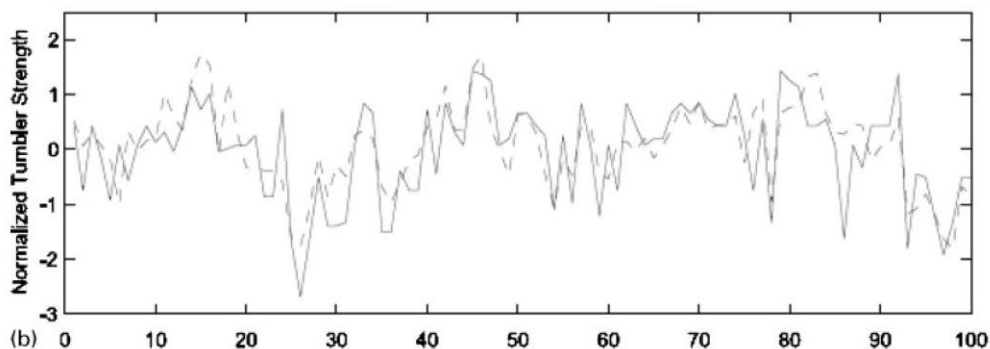
Figura 25 – Gráfico em série dos valores de teste previstos pela rede MLP do trabalho de [Kinnunen e Laitinen \(2005\)](#) e os valores de teste reais do indicador de qualidade RDI.



Fonte: ([KINNUNEN; LAITINEN, 2005](#))

- [Laitinen e Saxén \(2007\)](#) : Nessa pesquisa, os autores sumarizaram 1754 dias de processo de sinterização em um total de 1603 instâncias na busca de cinco parâmetros: produtividade, consumo de combustível, fração de finos de retorno, *low temperature breakdown* (LTB) e Índice de Tamboreamento (TI) através da modelagem de uma rede MLP - GDM de camada oculta única. Os autores optam por normalizar os dados através da chamada normalização *Z - Score*, também chamada de média-zero ou normalização unidade-variante, e informam que, para o TI, o algoritmo alcançou R^2 de 0.65. A Figura 26 mostra um comparativo, dos dados de teste, dos resultados previstos pela MLP (linha pontilhada) e dos dados reais (linha sólida).

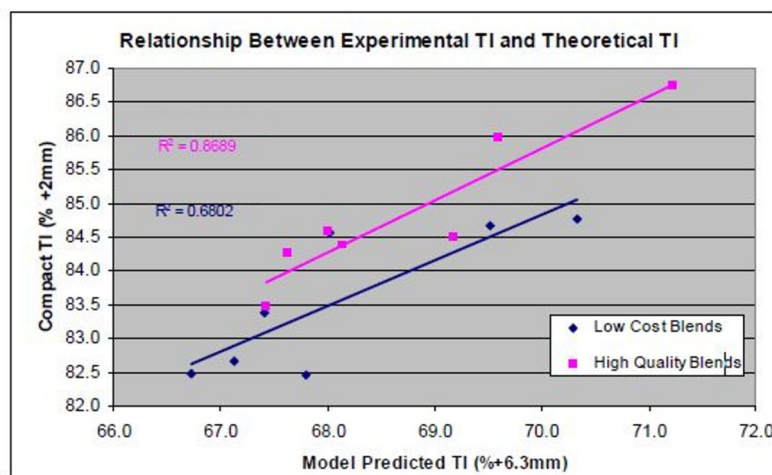
Figura 26 – Gráfico em série dos valores de teste previstos pela rede MLP-GDM do trabalho de Laitinen e Saxén (2007) e os valores de teste reais do indicador de qualidade TI.



Fonte: (LAITINEN; SAXÉN, 2007)

- Donskoi et al. (2007) : Os autores realizaram um cuidadoso filtro nos dados de uma planta de sinterização obtendo 95 instâncias com os melhores e mais estáveis parâmetros. Ao todo, 12 variáveis foram rearranjadas para o treinamento de algoritmos de regressões polinomiais na busca de indicadores como produtividade, Índice de Tamboreamento (TI), Índice de Degradação Sob Redução (RDI) e consumo de combustível. Os resultados de interesse para comparação são o R^2 para o TI, de 0.88, e o R^2 para o RDI, de 0.86. A Figura 27 mostra um comparativo entre os dados de teste previstos e os dados de teste reais do TI de dois conjuntos de dados distintos. Pode-se observar que a quantidade de dados nessa pesquisa é limitada e, por isso, pode não refletir um excelente ajuste do modelo preditivo.

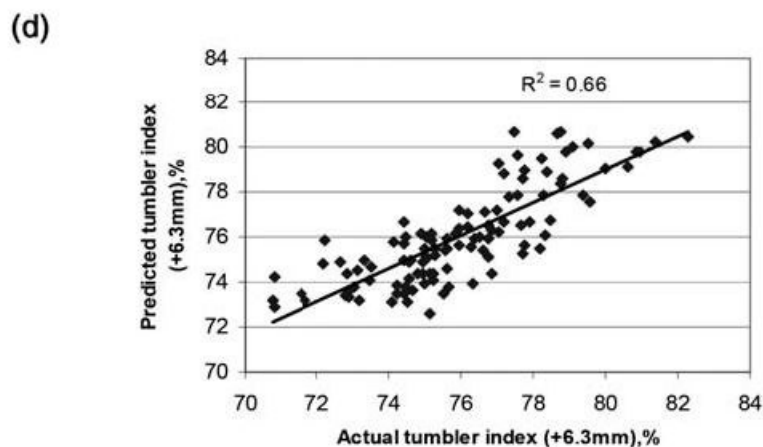
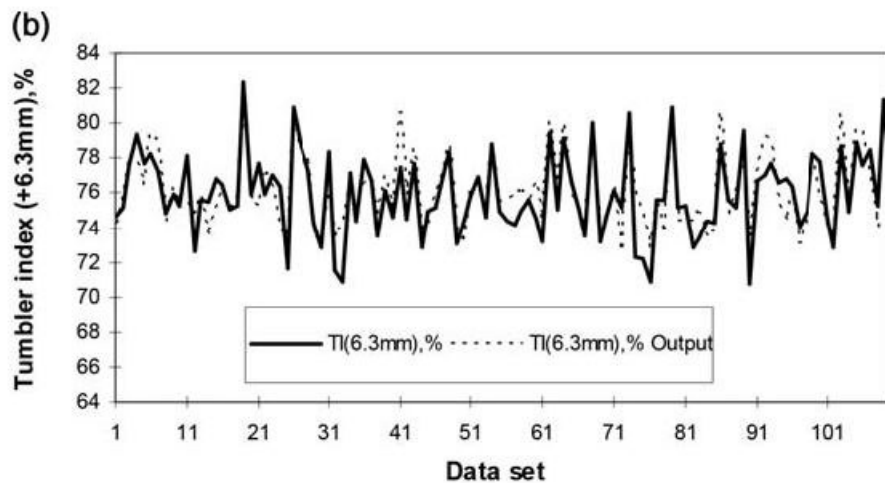
Figura 27 – Gráfico em série dos valores de teste previstos por algoritmos de regressão polinomial do trabalho de Donskoi et al. (2007) e os valores de teste reais do indicador de qualidade TI de dois subconjuntos de testes distintos.



Fonte: Adaptado de (DONSKOI et al., 2007)

- [Umadevi et al. \(2016\)](#) : Nesse estudo, os autores focaram em desenvolver uma rede MLP - GDM de camada oculta única para prever o Índice de Tamboreamento (TI) utilizando 441 instâncias de uma planta de sinterização com 9 variáveis de entrada. A Figura 28 mostra duas imagens: um gráfico em série entre os valores previstos pela rede (linha pontilhada) e os dados reais (linha sólida) e um gráfico de dispersão também entre os valores previstos e os valores reais. O melhor modelo de rede obteve R^2 de 0.66 para o TI.

Figura 28 – Gráfico em série e de dispersão dos valores de teste previstos pela rede MLP-GDM do trabalho de [Umadevi et al. \(2016\)](#) e os valores de teste reais do indicador de qualidade TI.



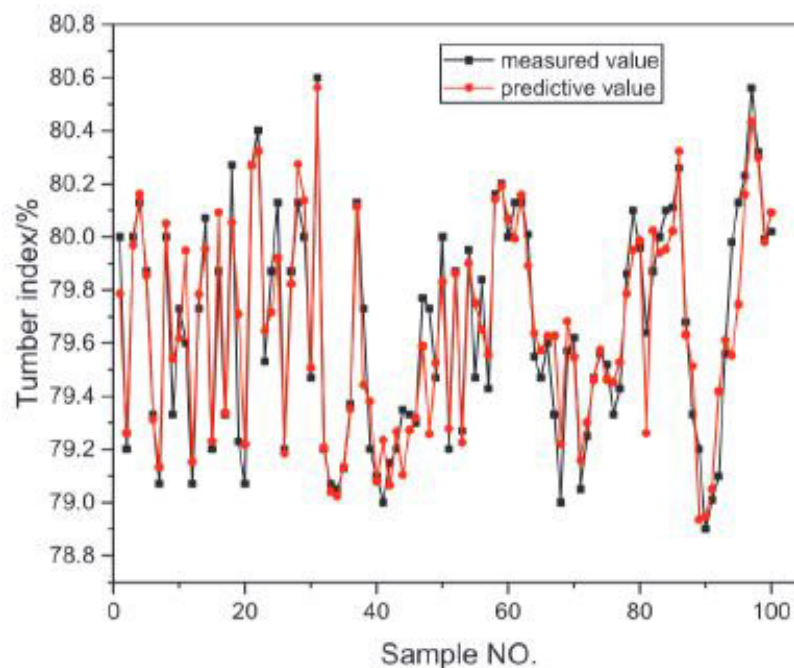
Fonte: ([UMADEVI et al., 2016](#))

- [Kumar et al. \(2017\)](#) : Os autores conseguiram sumarizar mais de 3000 instâncias de 3 anos da planta de sinterização com 11 variáveis de entrada importantes do processo para configurar uma rede MLP de camada única que conseguisse prever o MPS, o TI e o RDI. Os melhores resultados de teste dos melhores modelos e configurações de redes testados alcançaram R^2 de 0.758 para o MPS, de 0.90 para o TI e de 0.71

para o RDI. Esses resultados são bem consistentes visto a grande quantidade de dados disponíveis.

- [Shao et al. \(2020\)](#) : Nessa pesquisa realizada em uma planta siderúrgica chinesa, os autores reúnem 1000 instâncias de 15 variáveis consideradas importantes no processo na busca de diversos indicadores de qualidade do sinter, como o Índice de Tamboreamento (TI). Através de uma rede MLP de camada oculta única com um algoritmo modificado de retropropagação para mais rápida convergência de resultados, para os dados de teste, foi obtido um R^2 de 0.87. A Figura 29 mostra um comparativo entre os valores previstos pela MLP (linha vermelha) e os valores reais (linha preta) do indicador de qualidade do sinter TI.

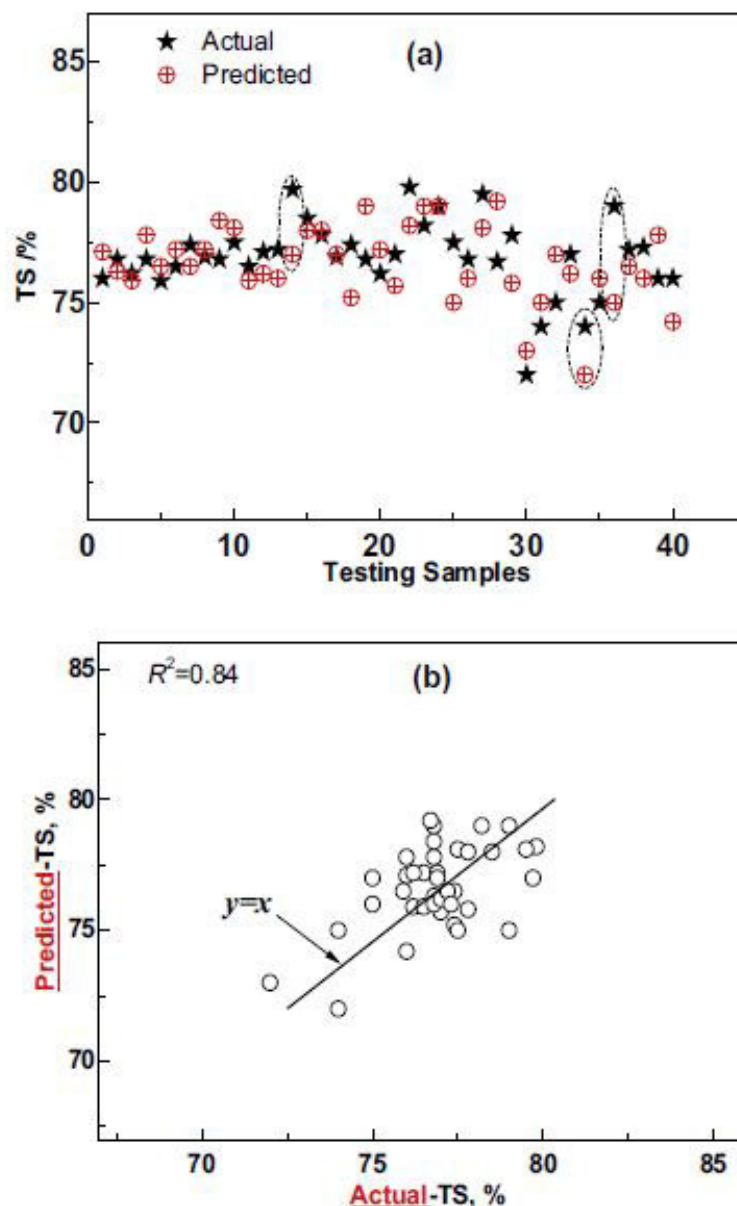
Figura 29 – Gráfico em série dos valores de teste previstos pela rede MLP-GDM do trabalho de [Shao et al. \(2020\)](#) e os valores de teste reais do indicador de qualidade TI.



Fonte: ([SHAO et al., 2020](#))

- [Gao et al. \(2021\)](#) : Os autores realizaram uma complexa metodologia de redução de variáveis via Análise de Componente Principais (PCA) e otimização hiperparamétrica de uma MLP-GDM com Algoritmos Genéticos (GA) através de um conjunto cuidadosamente filtrado de 280 instâncias de uma planta de sinterização com 19 variáveis de entrada principais na previsão do Índice de Tamboreamento (TI). O modelo superior da melhor configuração alcançou R^2 de 0.84, conforme Figura 30, onde, no gráfico em série, os valores previstos pelo algoritmo estão em vermelho e os valores reais do indicador estão em preto.

Figura 30 – Gráfico em série e de dispersão dos valores de teste previstos pela rede MLP-GDM do trabalho de Gao et al. (2021) e os valores de teste reais do indicador de qualidade TI.



Fonte: (GAO et al., 2021)

Como pode ser observado a partir dos resultados de outros autores, o Índice de Tamboreamento (TI) é um indicador bastante estudado para avaliar a resistência mecânica do sínter. O coeficiente de determinação, R^2 , alcançado nos trabalhos citados variou entre 0.65 e 0.90, dependendo da quantidade e qualidade de dados, quantidade e importância de variáveis de entrada e tipo de algoritmo de previsão. Como mencionado anteriormente, o banco de dados utilizado no presente trabalho não contém informações sobre o Índice de Tamboreamento (TI) e a resistência mecânica do sínter foi avaliada por meio do Índice de Resistência Shatter (SI). Embora esse banco de dados seja composto por apenas 426

observações, a rede MLP-LM alcançou um R^2 igual a 0.7410 para o Índice de Resistência Shatter (SI). Esse resultado pode ser considerado satisfatório, principalmente considerando o reduzido número de observações disponíveis para o desenvolvimento do modelo. O coeficiente de determinação alcançado nesta pesquisa para o Índice de Degradação Sob Redução (RDI) foi de 0.4478, contra R^2 de 0.86 e 0.71 apresentados por dois dos trabalhos encontrados durante a revisão da literatura (DONSKOI et al., 2007; KUMAR et al., 2017). Apenas um trabalho encontrado durante a revisão da literatura, Kumar et al. (2017), realizou um estudo de previsibilidade do Tamanho Médio de Partícula (MPS) e obteve R^2 de 0.758 contra R^2 igual a 0.6030 alcançado no presente estudo. A presente pesquisa também alcançou R^2 igual a 0.2684 para o Índice de Redutibilidade (RI), mas nenhum trabalho encontrado durante a revisão da literatura investigou esse indicador para permitir uma comparação.

As respostas dos algoritmos configurados para o Índice de Resistência Shatter (SI) e o Tamanho Médio de Partícula (MPS) são coerentes com outros estudos encontrados e mostram-se promissoras técnicas de avaliação preditiva da qualidade do sínter. Já as previsões tanto do Índice de Degradação Sob Redução (RDI) quanto do Índice de Redutibilidade (RI) mostram-se inferiores, seja por outros estudos já terem encontrado coeficientes de determinação mais elevados, seja pelo próprio R^2 encontrado poder ser considerado baixo e de difícil utilização prática.

Algumas hipóteses em relação ao baixo R^2 encontrado em alguns indicadores de qualidade podem ser discutidas:

- Conforme mencionado no Item 3.1.1, o registro de aquisição das variáveis, seja de entrada ou de saída, é realizado com intervalos distintos. Para alguns parâmetros é atualizado a cada hora, para outros a cada 4 horas, chegando até mesmo a cada 24 horas. Dessa forma, as 426 instâncias agrupadas representam médias aritméticas de valores para sumarizar um dia de processo. Uma média registrada a cada 24 horas pode não ser representativa das variações que ocorrem nos parâmetros do processo ou propriedades físico-químicas das materiais primas ao longo de um dia de produção.
- Outra hipótese está relacionada com a falta de variáveis significativas para a correta previsão dos indicadores. Como pode ser observado em Shao et al. (2020) e Gao et al. (2021), parâmetros relacionados à etapa de granulação (anterior à máquina de sinterização) como umidade do produto granulado, tempo de granulação e tamanho dos grânulos, além de parâmetros relacionados ao combustível da sinterização, como taxa de consumo de combustível e a fração de carbono no combustível, são variáveis relevantes e representativas para modelar o processo de sinterização que não foram fornecidas para o presente trabalho.

- Apesar de [Gao et al. \(2021\)](#) conseguir alcançar bons resultados com apenas 280 instâncias, as melhores respostas encontradas são de [Shao et al. \(2020\)](#) e [Kumar et al. \(2017\)](#), utilizando, respectivamente, 1000 e 3000 dados. Em comparação com as 426 instâncias, mais dados podem representar melhor o processo e facilitar a execução dos algoritmos, além de auxiliar na validação estatísticas das respostas.

4.4 OTIMIZAÇÃO DO NÚMERO DE NEURÔNIOS OCULTOS

O custo computacional associado a busca hiperparamétrica das redes neurais utilizadas é elevado. Mesmo com processadores atuais, a realização de uma busca exaustiva da melhor configuração da rede pode se tornar inviável caso muitos hiperparâmetros sejam considerados. Como já mencionado, no presente estudo, foram criados 2400 experimentos de redes MLP-GDM e 1200 experimentos de redes MLP-LM para cada um dos 4 indicadores de qualidade do sinter buscados.

Em relação a MLP-GDM, a duração média calculada de um experimento com 100 rodadas de verificação treinamento|teste é de 1 minuto e 30 segundos, totalizando 10 dias de execução dos procedimentos. Já para a MLP-LM, a duração média calculada de um experimento é aproximadamente 36 segundos, totalizando 2 dias de execução dos algoritmos. Ao todo, foram necessários aproximadamente 288 horas de execução para completar todos os experimentos desejados.

A técnica proposta no presente trabalho utilizando a análise SVD em diferentes matrizes das redes neurais para determinar o número ótimo de neurônios ocultos tem o potencial de reduzir para segundos a busca de hiperparâmetros ou, pelo menos, ajudar em uma melhor definição dos domínios de busca.

As Tabelas 18 e 19 mostram os resultados para as redes MLP-GDM e MLP-LM, respectivamente. As matrizes $\Delta^{(h)}$ e $E^{(h)}$ apresentam-se com problemas visto que em vários testes, a aplicação de SVD indicou um único valor singular não-nulo. Segundo [Santos, Barreto e Medeiros \(2010\)](#), [Santos \(2010\)](#), isso significa que a variância total dessas matrizes é descrita somente pelo único valor singular presente, o que leva ao número de condicionamento das matrizes a se aproximar ao infinito, sendo uma indicação de que são matrizes mal condicionadas. Dessa forma, apenas os resultados obtidos com a matriz $Y^{(h)}$ serão discutidos.

Os melhores resultados alcançados pelos modelos de MLP-GDM para os indicadores de qualidade SI, RI, RDI e MPS, respectivamente, possuem as estruturas: [8 - 5 - 1]; [8 - 9 - 1]; [10 - 6 - 1] e [14 - 12 - 1]. Comparando as Tabelas 12 e 18, observa-se que para o SI e RI, com estrutura proposta iguais de [9 - 4 - 1], a diferença entre os números de neurônios ideais e propostos é de aproximadamente 1.

Para os modelos de MLP-LM, os melhores resultados alcançados nos indicadores SI, RI, RDI e MPS possuem, respectivamente, as estruturas: [10 - 14 - 1]; [6 - 13 - 1]; [11 - 8 - 1] e [6 - 12 - 1]. Comparando as Tabelas 13 e 19, nota-se que para o RDI, a estrutura proposta foi 11 - 8 - 1, o que coincide exatamente com o melhor resultado. Para o restante dos indicadores SI, RI e MPS, com estruturas propostas [10 - 10 - 1]; [10 - 9 - 1] e [10 - 9 - 1], respectivamente, observa-se uma diferença entre os números de neurônios ideais e propostos entre 0 e 4.

Tabela 18 – Análise SVD para determinação do número ótimo de neurônios para cada camada oculta conforme melhor modelo MLP-GDM.

Técnica	Camada da rede neural	Saídas			
		SI [%>10mm]	RI [%]	RDI [%<2.83mm]	MPS [mm]
$Y^{(h)}$	1ª camada oculta	9	10	9	8
	2ª camada oculta	4	5	4	1
SVD $\Delta^{(h)}$	1ª camada oculta	4	4	5	8
	2ª camada oculta	1	1	1	2
$E^{(h)}$	1ª camada oculta	1	1	1	1
	2ª camada oculta	1	1	1	1

Fonte: elaborado pelo próprio autor.

Tabela 19 – Análise SVD para determinação do número ótimo de neurônios para cada camada oculta conforme melhor modelo MLP-LM.

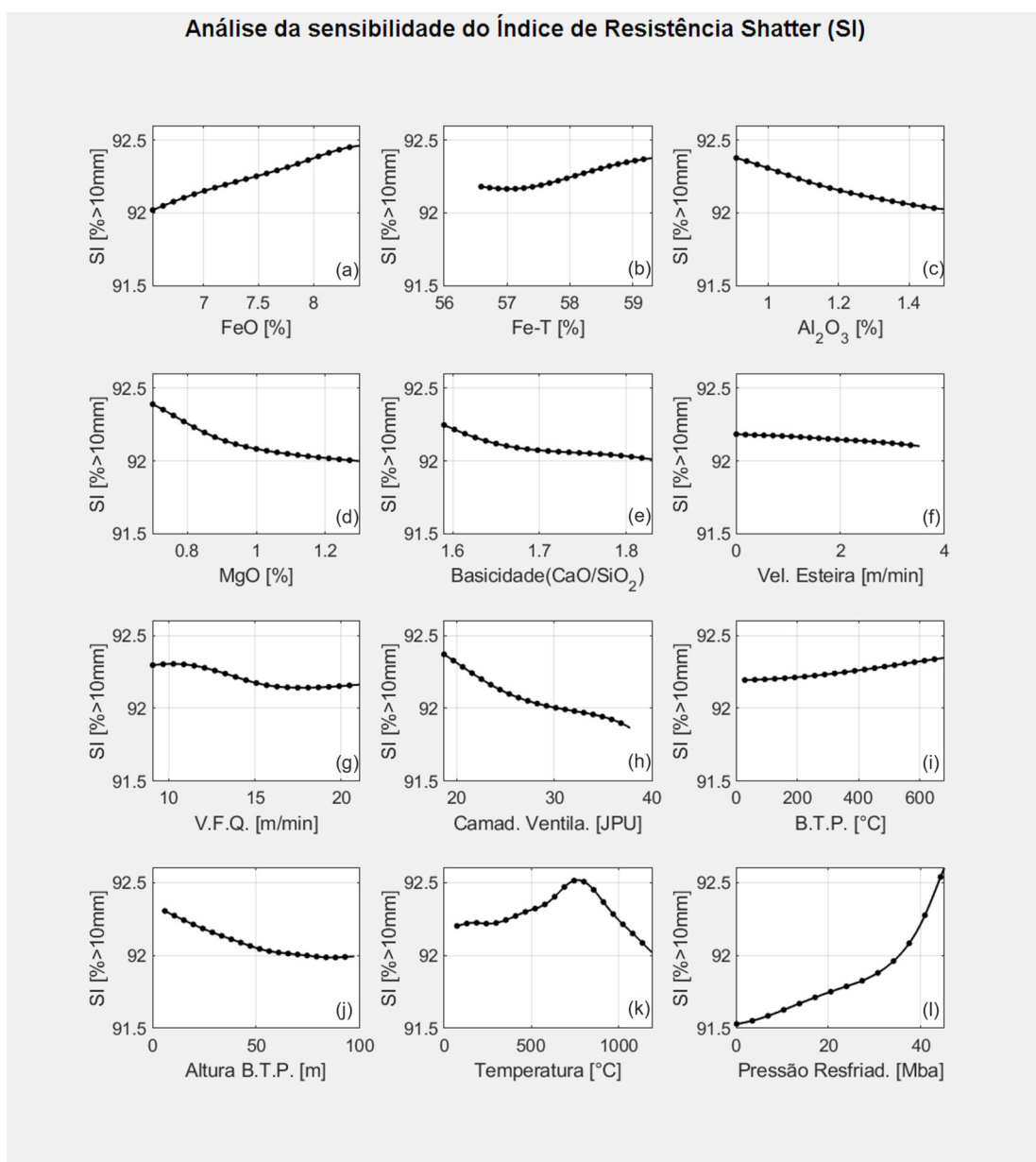
Técnica	Camada da rede neural	Saídas			
		SI [%>10mm]	RI [%]	RDI [%<2.83mm]	MPS [mm]
$Y^{(h)}$	1ª camada oculta	10	10	11	10
	2ª camada oculta	10	9	8	9
SVD $\Delta^{(h)}$	1ª camada oculta	10	9	11	11
	2ª camada oculta	6	6	3	7
$E^{(h)}$	1ª camada oculta	5	5	3	4
	2ª camada oculta	1	1	1	1

Fonte: elaborado pelo próprio autor.

4.5 ANÁLISE DE SENSIBILIDADE

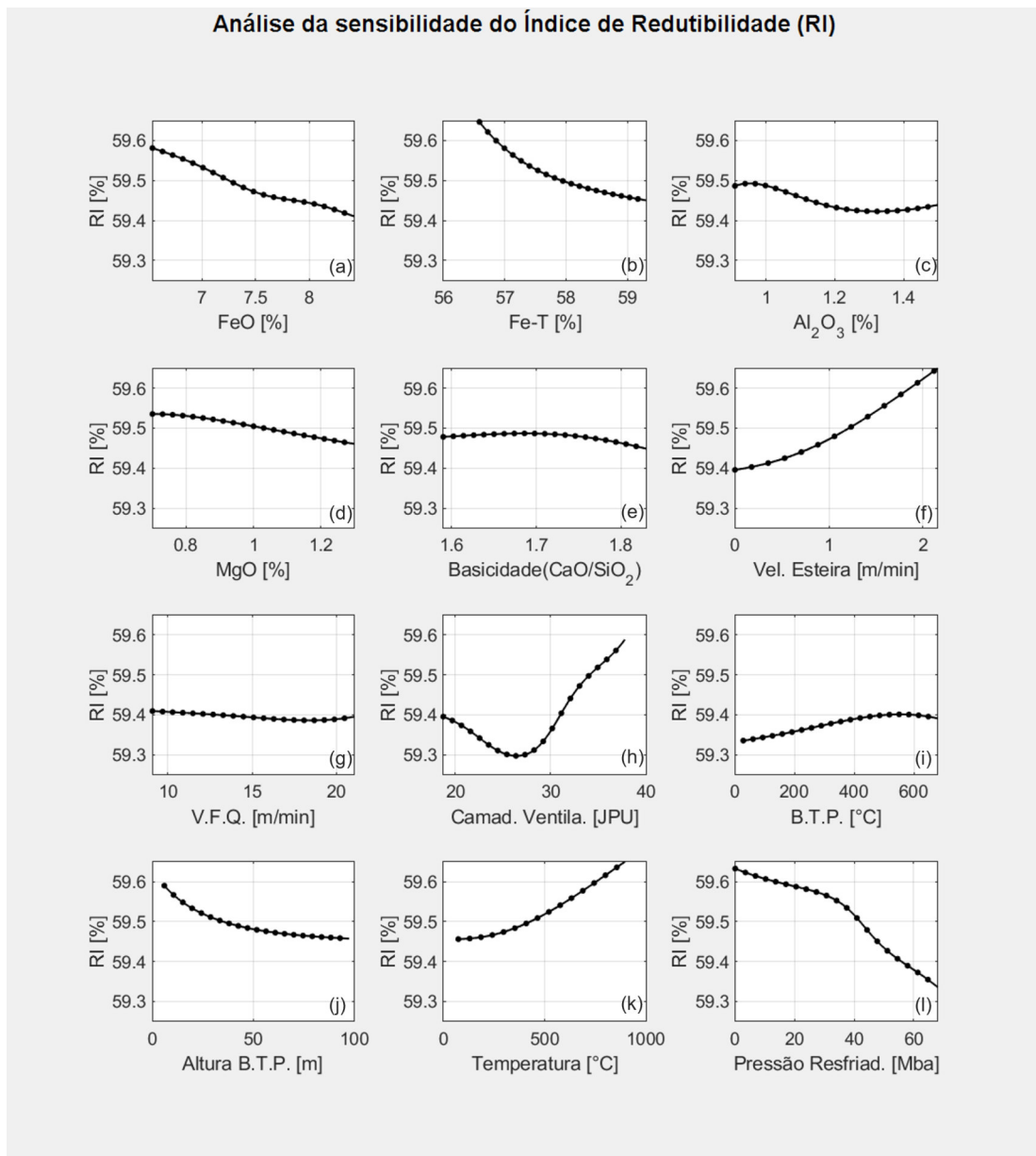
As Figuras 31, 32, 33 e 34 mostram a análise de sensibilidade dos melhores modelos de redes neurais, respectivamente, para o SI, RI, RDI e MPS. Esse estudo, contudo, deve ser analisado com cautela. As relações apontadas nos gráficos, seja diretamente proporcional, seja inversamente proporcional ou mesmo sem relação visível, são resultados aparentes e preliminares.

Figura 31 – Análise de sensibilidade do melhor modelo de previsão do Índice de Resistência Shatter (SI).



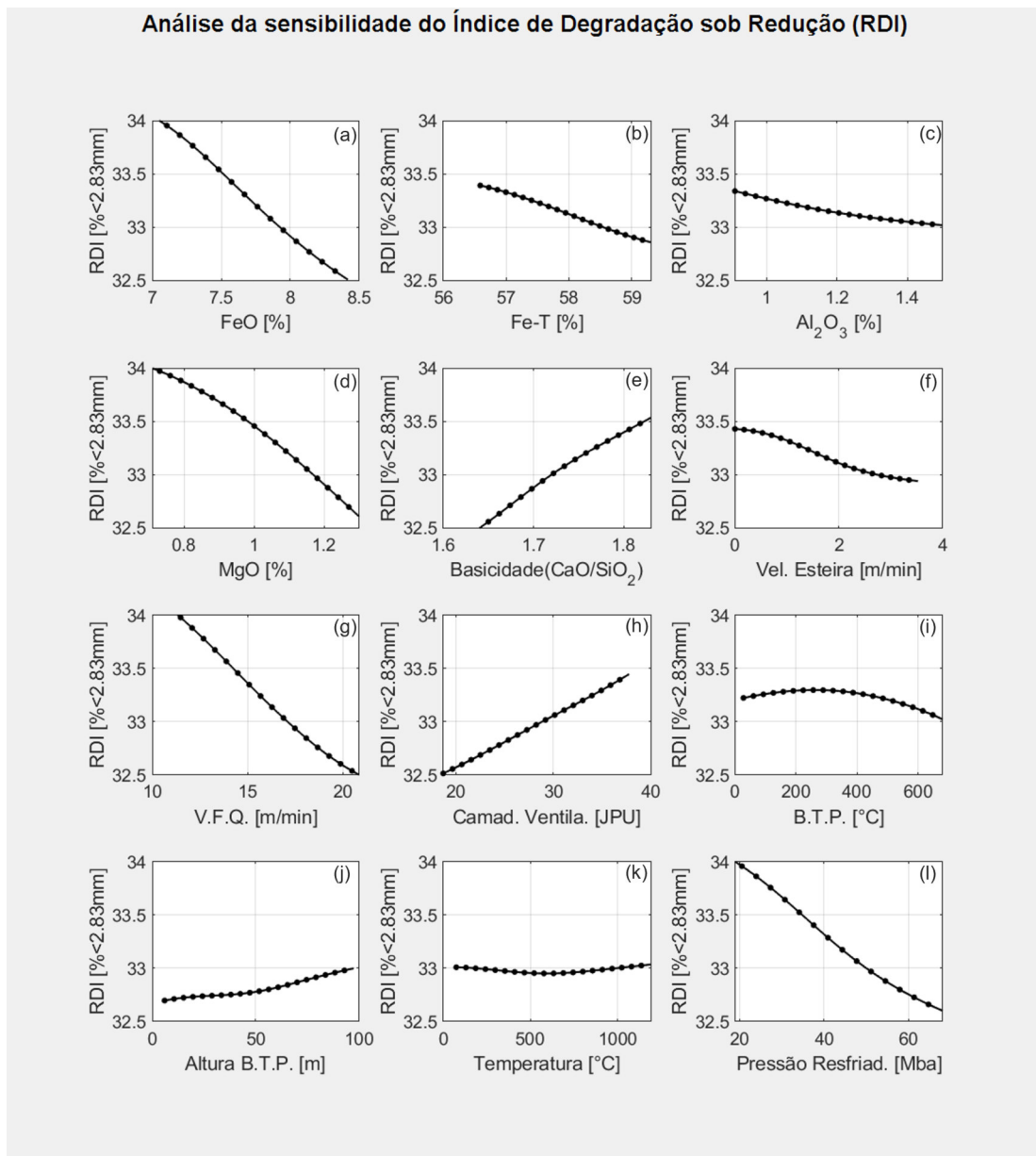
Fonte: elaborado pelo próprio autor.

Figura 32 – Análise de sensibilidade do melhor modelo de previsão do Índice de Redutibilidade (RI).



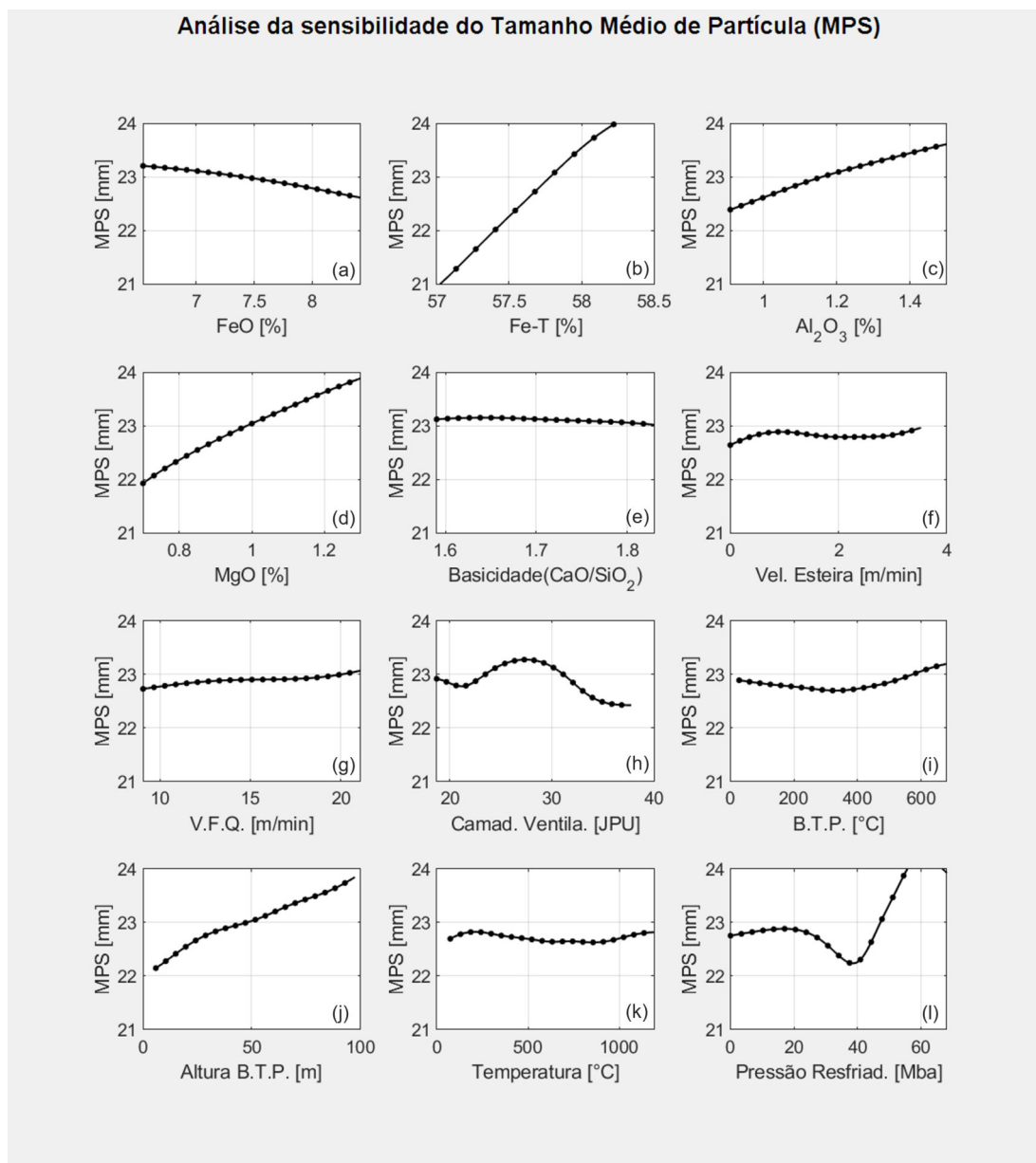
Fonte: elaborado pelo próprio autor.

Figura 33 – Análise de sensibilidade do melhor modelo de previsão do Índice de Degradação Sob Redução (RDI).



Fonte: elaborado pelo próprio autor.

Figura 34 – Análise de sensibilidade do melhor modelo de previsão do Tamanho Médio de Partícula (MPS).



Fonte: elaborado pelo próprio autor.

Como alguns modelos de previsão de indicadores de qualidade do sinter não apresentam elevado R^2 , ou mesmo como o algoritmo pode modelar erroneamente algumas das relações devido a dados não otimizados e não representativos, sabe-se que nem todas as relações entre as variáveis observadas graficamente nas análises condizem com informações técnicas reais. Os textos a seguir sumarizam as principais informações encontradas na literatura base que dão suporte a algumas das relações encontradas nas análises de sensibilidade. Apesar de nem todos os vínculos terem sido corretamente preditos, no geral, boas discussões e observações podem ser feitas, como:

- FeO [%]: A presença de FeO no sinter é um importante indicador das condições do processo de sinterização. Geralmente é associada a dois principais fatores: excesso de coque/finos de coque no *blend* inicial e maior presença da fase magnetita ($FeO.Fe_2O_3$) (FERNANDEZ-GONZALEZ et al., 2017b; UMADEVI et al., 2012; MOCHÓN et al., 2014). Segundo Umadevi et al. (2012), FeO induz a formação da fase SFCA-I (acicular, mais resistente) em detrimento da fase SFCA (colunar, menos resistente), propiciando a minimização de trincas e de poros de maior dimensão. Isso faz com que a resistência mecânica do sinter aumente com o aumento de FeO , como analisado na Figura 31 (a). Entretanto, sabe-se que o aumento de FeO no sinter está associado a uma diminuição do RI pois esse óxido reage com o SiO_2 formando a fase faialita, um silicato de pequena redutibilidade. Isso também é verificado na análise na Figura 32 (a). Além disso, como a presença de maiores quantidades de FeO também está relacionada com a presença de magnetita, cria-se uma matriz de sinter com menos hematita, o que, segundo Fernandez-Gonzalez et al. (2017b), leva a uma redução do RDI. Essa associação novamente pode ser verificada na análise de sensibilidade mostrada na Figura 33 (a). Em relação ao MPS, não foi encontrado informação na literatura que especifique como o FeO influencia essa propriedade, porém, sendo o FeO um indicador de maior resistência mecânica, estima-se que as relações encontradas para o SI sejam parecidas as encontradas para o MPS, indicador do tamanho médio das partículas finais, na medida em que um produto mais resistente tenha presumidamente partículas mais grosseiras devido menor desgaste e abrasão durante o processo produtivo.
- TFe [%]: Um alto teor Fe no sinter indica a utilização de matéria-prima de melhor qualidade no mix de minérios, ou seja, menor quantidade de impurezas como SiO_2 , Al_2O_3 , S , P , MgO e CaO . Essas impurezas dão origem a diversas fases já mencionadas ao longo do trabalho e que, geralmente, induzem a diminuição da resistência e redutibilidade do produto. Dukino et al. (1995) indica a seguinte ordem decrescente de resistência a frio nas fases do sinter: hematita primária > hematita secundária > magnetita > fases SFCA > silicatos, enquanto Dawson (1993) discute que a ordem decrescente das fases com maior redutibilidade seria: hematita primária > hematita secundária > fases SFCA > magnetita > silicatos. Como pode ser observado, as fases geradas pela ganga mineral produzem um sinter de pior qualidade (apesar de ser necessário uma quantidade mínima de fases SFCA para permitir correta composição das camadas aderentes no sinter). Assim, maiores quantidade de Fe no sinter tendem a indicar melhores propriedades finais: elevado SI, como mostrado na análise de sensibilidade da Figura 31 (b), elevado MPS, também indicado nas relações da Figura 34 (b), elevado RI e baixo RDI, observado na análise contida na Figura 33 (b) (LU; PAN; ZHU, 2015).

- MgO [%] : Segundo [Fernandez-Gonzalez et al. \(2017b\)](#), o MgO no sinter permite a obtenção de uma escória adequado no alto forno (melhora fluidez e dessulfuração), sendo adicionado como dolmita ou dunita. [Gao et al. \(2021\)](#), [Gao et al. \(2013\)](#) dizem que maiores concentrações de MgO tendem a reduzir a resistência mecânica do sinter pois o Fe_2O_3 e o MgO formam solução, com os íons Fe^{2+} e Mg^{2+} substituindo uns aos outros. Na sinterização, esses compostos formam $\{(Fe_xMg_{1-x})O.Fe_2O_3\}$, uma fase chamada minerais MF, cujo ponto de fusão é alto e certamente prejudica a atuação das fases que devem fundir para compactar e dar resistência ao sinter (fases aderentes), assim como encontrado na análise de sensibilidade mostrada na Figura 31 (d). Apesar disso, ainda existe certa divergência na atuação do MgO , com autores como [Zhang et al. \(2013\)](#) defendendo que esse composto aumenta a fluidez das fases ligantes melhorando a resistência do material. [Lu, Holmes e Manuel \(2007\)](#) estabelece que o MgO reduz o RDI pois esse composto estabiliza a fase magnetita, reduzindo a quantidade de hematita, o que já foi mencionado anteriormente como característica redutora de RDI. Esse relação também foi observada na análise de sensibilidade contida na Figura 33 (d). [Fernandez-Gonzalez et al. \(2017a\)](#) alerta que o MgO pode atuar de diversas maneiras na redutibilidade do sinter a depender dos teores de sílica, mas que de forma geral foi encontrado que a redutibilidade diminui com a adição de MgO , relação também encontrada na análise da Figura 32 (d).
- Al_2O_3 [%]: [Gao et al. \(2021\)](#), através de estudos com redes neurais, encontrou uma atuação parecida da alumina (Al_2O_3) com o MgO : quanto maior seu teor, menor a resistência mecânica do sinter. Essa relação foi observada na análise de sensibilidade da Figura 31 (c). Apesar de [Lu, Holmes e Manuel \(2007\)](#) estabelecer que o RDI aumenta com o aumento de Al_2O_3 , autores como [HSIEH \(2005\)](#), [Kinnunen e Laitinen \(2005\)](#) dizem que essa relação é fraca e pode, a depender de outros componentes estruturais, ser nula. [Yu et al. \(2015\)](#) observou que altas concentrações de alumina causavam uma gradual diminuição na redutibilidade do sinter devido a deterioração do líquido primário, o que também foi observado na análise da Figura 32 (c).
- *Basicidade*(CaO/SiO_2): [Umadevi, Sah e Mahapatra \(2014\)](#) em um estudo prático com finos de minério de ferro com alta e baixa concentração de alumina concluiu que um aumento na basicidade do sinter induzia uma aumento da resistência mecânica e diminuição do RDI. Contudo, [Fernandez-Gonzalez et al. \(2017b\)](#) mostra que a basicidade atua de diversas formas em diferentes intervalos. Em teores mais baixos, a presença de múltiplas fases, incluindo as fases SFCA e SFCA-I, com múltiplos coeficientes de expansão e pontos de fusão, induz à tensões internas e ao aparecimento de trincas, diminuindo a resistência mecânica, como observado na análise de sensibilidade da Figura 31 (e). A partir de valores mais altos de basicidade, segundo

- Tufan et al. (2013), há um estímulo para a formação de uma estrutura ótima composta de um núcleo hematítico rodeado de fases ferríticas aciculares, obtendo um sínter de boa resistência e redutibilidade.
- *Velocidade da esteira* [m/min]: Umadevi et al. (2016), Gao et al. (2021) mostram que a resistência mecânica do sínter diminui com o aumento da velocidade da esteira de sinterização. A velocidade da esteira comanda o tempo de permanência do leito sob o forno de ignição e sua zona de aumento de temperatura. Com o aumento da velocidade ocorre o aumento da taxa vertical de sinterização, que proporciona a formação do material fundido e sua ligação ao resto núcleo hematítico. Dessa forma, prejudicando a correta formação das camadas aderentes ligantes, o aumento da velocidade da esteira piora a resistência do sínter, assim como encontrado na análise de sensibilidade da Figura 31 (f).
 - *V.F.Q (Velocidade da Frente de Queima)* [m/min]: segundo Fernandez-Gonzalez et al. (2017a), a velocidade da frente de queima tem uma grande influência na qualidade da sinterização, produtividade e tempo de sinterização. Essa é a região onde as partículas de coque estão em combustão, e possui duas fronteiras, uma onde o coque começa a queimar e outra onde ele termina. As propriedades dessa zona são críticas porque determinam o calor transferido para o leito, sendo responsável pela resistência do leito de sinterização formado. Zhou et al. (2015) afirma que até certo valor, a V.F.Q. não prejudica a resistência final do sínter, mas que a partir daí, existe uma tendência de atenuação dessa indicador, assim como a relação estabelecida pela análise de sensibilidade encontrada na Figura 31 (g).
 - *Camada de Ventilação* [JPU]: a camada de ventilação é uma medida da permeabilidade do leito de sinterização que, segundo Zhou et al. (2020), é um forte parâmetro de todo o processo. Um aumento na permeabilidade do leito é desejável para uma maior troca de calor do topo ao fundo do leito e para obter melhor produtividade e qualidade do sínter.
 - *B.T.P (Burn Through Point)* [°C]: segundo Umadevi et al. (2016), o B.T.P. indica a temperatura de conclusão do processo sinterização porque é a medida de quanto a frente de queima progrediu através do leito de sinterização, sendo detectado por sensores de temperatura sob o leito de sinterização. O B.T.P é determinado em grande parte pela temperatura do forno de ignição, altura do leito, V.F.Q. e velocidade da máquina de sinterização. O coque é um fornecedor de energia essencial para assegurar a propagação da frente de calor através do leito após a zona de ignição. Para obtenção de alta resistência mecânica do sínter, é necessário manter um alto B.T.P. através do ajustes dos parâmetros mencionado acima. A relação diretamente proporcional entre essas duas variáveis foi encontrada na análise da Figura 31 (i).

- *Temperatura do forno de ignição* [°C]: conforme [Gao et al. \(2021\)](#), um aumento da temperatura do forno de ignição induz a um aumento da resistência mecânica do sinter. [Umadevi et al. \(2016\)](#) afirma que esse parâmetro é um dos grandes influenciadores do B.T.P. e, quanto maior a temperatura de ignição, maior o B.T.P. e melhores índices de qualidade terá o sinter produzido.
- *Pressão do Resfriador* [Mbar]: a pressão do resfriador está associada a taxa na qual o ar é succionado e o sinter é resfriado. [Umadevi et al. \(2016\)](#) destaca que esse parâmetro controla o tempo e temperaturas ótimas de sinterização, sendo função de outros fatores como permeabilidade do leito e velocidade da máquina de sinterização. O uso de sucção muito alta pode resultar em menor eficiência térmica e piores indicadores qualidade do sinter, como baixa reutilizabilidade, relação encontrada na análise de sensibilidade da Figura 32 (1).

5 Conclusão

O estudo conseguiu criar um sistema de previsão de quatro índices de qualidade do sinter metalúrgico: Índice de Resistência Shatter (SI), Índice de Redutibilidade (RI), Índice de Degradação Sob Redução (RDI) e Tamanho Médio de Partícula (MPS), composto de redes neurais MLP-GDM e MLP-LM capaz de relacionar diversos parâmetros e variáveis de um complexo processo de sinterização.

Os indicadores de desempenho das respostas obtidas com as redes neurais para o SI ($R^2:0.7410$) e o MPS ($R^2:0.6030$) são equiparáveis e, em alguns casos, superiores a resultados de trabalhos anteriores. Enquanto para o RI ($R^2:0.2684$) e RDI ($R^2:0.4478$), as respostas foram insatisfatórias para fins de utilização prática. Contudo, é importante destacar que foram encontrados apenas dois trabalhos que abordaram a previsibilidade do RDI e nenhum sobre o RI, ou seja, o estudo desses indicadores é novo e mostra-se um desafio. Algumas hipóteses podem ser levantadas quanto ao baixo coeficiente de determinação desses dois indicadores de qualidade: a aquisição de dados em intervalos distintos, com possíveis ajustes no processo que podem não ser levados em consideração; a falta de algumas variáveis significativas relacionadas à etapa de granulação e ao combustível de sinterização; a baixa quantidade de dados disponíveis em comparação com os melhores resultados da literatura base.

Análises de sensibilidade dos melhores modelos de redes neurais foram realizadas, sendo possível verificar as tendências e padrões entre as variações dos dados de entrada e de saída. Diversas relações foram identificadas e discutidas de acordo com a literatura base.

Em relação à técnica de análise SVD, dois pontos principais podem ser mencionados:

- A adaptação desse tipo de análise em uma rede MLP de classificação de camada oculta única, proposta por Santos, Barreto e Medeiros (2010), para uma rede MLP de regressão com duas camadas ocultas, mostrou-se viável;
- A técnica mostrou o potencial de reduzir para segundos a busca de hiperparâmetros ou, pelo menos, ajudar em uma melhor definição dos domínios dessa busca. Para ambas as redes testadas, em alguns indicadores de qualidade, a otimização conseguiu aproximar o número ótimo de neurônios proposto com o número de neurônios estimados pelas redes de forma que mostra-se útil e deve ser melhor analisada em trabalhos futuros.

Referências

- ABBAD, G.; TORRES, C. V. Regressão múltipla stepwise e hierárquica em psicologia organizacional: aplicações, problemas e soluções. *Estudos de Psicologia (Natal)*, SciELO Brasil, v. 7, p. 19–29, 2002. Disponível em: <<https://doi.org/10.1590/S1413-294X2002000300004>>.
- AHMADI, H.; MOTTAGHITALAB, M.; NARIMAN-ZADEH, N. Group method of data handling-type neural network prediction of broiler performance based on dietary metabolizable energy, methionine, and lysine. *Journal of Applied Poultry Research*, Elsevier, v. 16, n. 4, p. 494–501, 2007. Disponível em: <<https://doi.org/10.3382/japr.2006-00074>>.
- ALMEIDA, E. S. *Previsão de propriedades físicas e metalúrgicas do sinter através de redes neurais artificiais e regressão linear múltipla*. 2019. Monografia (Bacharel em Engenharia Metalúrgica), Universidade Federal do Ceará, Fortaleza, CE. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/33363>>.
- ANDERSON, J. A. *An introduction to neural networks*. 3rd. ed. Massachusetts: MIT press, 1995. Disponível em: <<https://mitpress.mit.edu/books/introduction-neural-networks>>.
- BALL, D. F. *Agglomeration of iron ores*. [S.l.]: American Elsevier Publishing Company, 1973.
- BECHTEL, W.; ABRAHAMSEN, A. *Connectionism and the mind: Parallel processing, dynamics, and evolution in networks*. 2nd. ed. Cambridge, MA: Blackwell Publishing, 2002. Disponível em: <<https://psycnet.apa.org/record/2002-01429-000>>.
- BIAGGI, M.; MEDVID, M.; ASSIS, C. Fator de inflação da variância e regressões auxiliares para diagnóstico do problema de multicolinearidade nos modelos de regressão. In: *Anais XIX Congresso Brasileiro de Engenharia de Avaliações e Perícias*. [s.n.], 2017. p. 1–26. Disponível em: <<https://ibape-nacional.com.br/biblioteca/wp-content/uploads/2017/08/025.pdf>>.
- BLATTNER, J.; RANADE, M.; RICKETTS, J. Evaluation of the effects of pellet properties on inland's 'a' blast furnace performance. In: . [s.n.], 1984. p. 267–271. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-0021727428&partnerID=40&md5=dd8b5a7fd6d8ba163e449ec4387715f5>>.
- BRAGA, F. D. *Uso de redes neurais de funções de base radial e regressão linear múltipla para a previsão de propriedades mecânicas de vergalhões de aço*. 2019. Monografia (Bacharel em Engenharia Metalúrgica), Universidade Federal do Ceará, Fortaleza, CE. Disponível em: <<http://www.repositorio.ufc.br/handle/riufc/49234>>.
- CASTRO, L. F. A.; FIGUEIRA, R. M.; TAVARES, R. P. *Princípios básicos e processos de fabricação do gusa ao aço líquido*. Belo Horizonte, MG, 1985. 25-31 p.
- CORES, A. et al. Iron ore sintering. part 1. theory and practice of the sintering process. *Dyna*, v. 80, n. 180, p. 152–171, 2013. Disponível em: <<https://www.researchgate.net/>

publication/265477115_Iron_ore_sintering_Part_1_Theory_and_practice_of_the_sintering_process>.

DAWSON, P. Recent development in iron ore sintering. *Ironmaking and Steelmaking*, v. 20, n. 2, p. 135–159, 1993.

DELOGU, R.; FANNI, A.; MONTISCI, A. Geometrical synthesis of mlp neural networks. *Neurocomputing*, v. 71, p. 919–930, 01 2008. Disponível em: <<http://dx.doi.org/10.1016/j.neucom.2007.02.006>>.

DEVORE, J. L. *Probabilidade e estatística para engenharia e ciências*. 9th. ed. Cengage Learning Edições Ltda., 2019. ISBN 978-85-221-2803-7. Disponível em: <<https://www.cengage.com.br/ls/9788522128037/>>.

DONSKOI, E. et al. Mathematical modeling and optimization of iron ore sinter properties. *Israel Journal of Chemistry*, v. 47, p. 373–379, 12 2007. Disponível em: <<http://dx.doi.org/10.1560/IJC.47.3-4.373>>.

DUKINO, R. et al. Fracture behaviour of three australian iron ores. *Transactions of the Institution of Mining and Metallurgy, Section C: Mineral Processing and Extractive Metallurgy*, v. 104, p. C11–C19, 1995. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-0029215870&partnerID=40&md5=80d06aba8d1ec195a8b5e6dc672a2516>>.

FERNANDEZ-GONZALEZ, D. et al. Iron ore sintering: Process. *Mineral Processing and Extractive Metallurgy Review*, Taylor & Francis, v. 38, n. 4, p. 215–227, 2017. Disponível em: <<https://doi.org/10.1080/08827508.2017.1288115>>.

FERNANDEZ-GONZALEZ, D. et al. Iron ore sintering: Quality indices. *Mineral Processing and Extractive Metallurgy Review*, Taylor & Francis, v. 38, n. 4, p. 254–264, 2017. Disponível em: <<https://doi.org/10.1080/08827508.2017.1323744>>.

FERNANDEZ-GONZALEZ, D. et al. Iron ore sintering: Raw materials and granulation. *Mineral Processing and Extractive Metallurgy Review*, Taylor & Francis, v. 38, n. 1, p. 36–46, 2017. Disponível em: <<https://doi.org/10.1080/08827508.2016.1244059>>.

FERREXPO. *2019 Full Year Results*. Baar, Switzerland, 2020. Disponível em: <https://www.ferrexpo.com/media/ag3c2xzf/2019-full-year-results_0.pdf>.

GAO, Q. et al. A forecast model of the sinter tumble strength in iron ore fines sintering process. *Powder Technology*, v. 390, p. 256–267, 2021. ISSN 0032-5910. Disponível em: <<https://doi.org/10.1016/j.powtec.2021.05.063>>.

GAO, Q.-j. et al. Effects of mgo containing additive on low-temperature metallurgical properties of oxidized pellet. *Journal of Iron and Steel Research International*, Springer, v. 20, n. 7, p. 25–28, 2013. Disponível em: <[https://doi.org/10.1016/S1006-706X\(13\)60121-1](https://doi.org/10.1016/S1006-706X(13)60121-1)>.

GERON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. 2nd. ed. O'Reilly Media, 2019. ISBN 9781492032649. Disponível em: <<https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/>>.

GOLUB G. H.; VAN LOAN, C. F. *Matrix Computations*. 3rd. ed. Baltimore, Maryland: The John-Hopkins University Press, 1996.

GUJARATI, D. N.; PORTER, D. C. *Econometria básica*. 5th. ed. Amgh Editora, 2011. Disponível em: <<https://lemos.pro.br/wp-content/uploads/2019/04/Gujarati-Econometria-B\C3\%A1sica-5a-1.pdf>>.

GURNEY, K. *An introduction to neural networks*. 1st. ed. London, UK: CRC press, 1997. Disponível em: <<https://doi.org/10.1201/9781315273570>>.

GÓMEZ, I.; FRANCO, L.; JEREZ, J. Neural network architecture selection: Can function complexity help? *Neural Processing Letters*, v. 30, p. 71–87, 10 2009. Disponível em: <<http://dx.doi.org/10.1007/s11063-009-9108-2>>.

HAGAN, M. T. et al. *Neural network design*. 2nd edition. ed. Martin Hagan, 2014. Disponível em: <<https://hagan.okstate.edu/nnd.html>>.

HAGAN, M. T.; MENHAJ, M. B. Training feedforward networks with the marquardt algorithm. *IEEE transactions on Neural Networks*, IEEE, v. 5, n. 6, p. 989–993, 1994. Disponível em: <<https://doi.org/10.1109/72.329697>>.

HAYKIN, S. S. et al. *Neural networks and learning machines*. 3rd. ed. New York, NY: Prentice Hall, 2009. ISBN 9780131293762. Disponível em: <<https://www.pearson.com/us/higher-education/program/Haykin-Neural-Networks-and-Learning-Machines-3rd-Edition/PGM320370.html?tab=order>>.

HEBB, D. O. *The organization of behavior: a neuropsychological theory*. 1st. ed. New York, NY: J. Wiley; Chapman & Hall, 1949. Disponível em: <<https://psycnet.apa.org/record/1950-02200-000>>.

HENDLER, J. Avoiding another ai winter. *IEEE Intelligent Systems*, IEEE Educational Activities Department, v. 23, n. 2, p. 2–4, 2008. Disponível em: <<https://doi.org/10.1109/MIS.2008.20>>.

HOFFMANN, R. *Análise de regressão: uma Introdução à Econometria*. São Paulo, SP: Universidade de São Paulo, 2016. ISBN 978-85-921057-0-9. Disponível em: <<https://doi.org/10.11606/9788592105709>>.

HONORATO, E.; RANGEL, N. *Curso: utilização do minério de ferro na siderurgia*. Ipatinga, MG, 1993. 6-27 p.

HONORATO, E. P. *Adequação granulométrica das matérias-primas e do sistema desegregação contínua (ISF), para melhorias na produtividade e qualidade do sinter para os altos-fornos*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2005. Disponível em: <<http://hdl.handle.net/1843/BUOS-8DRJBV>>.

HOPFIELD, J. J. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the national academy of sciences*, National Acad Sciences, v. 79, n. 8, p. 2554–2558, 1982. Disponível em: <<https://doi.org/10.1073/pnas.79.8.2554>>.

- HSIEH, L.-H. Effect of raw material composition on the sintering properties. *ISIJ International*, v. 45, n. 4, p. 551–559, 2005. Disponível em: <<https://doi.org/10.2355/isijinternational.45.551>>.
- HUMMEL, P. M.; SEEBECK, C. L. A generalization of Taylor's expansion. *The American Mathematical Monthly*, Taylor & Francis, v. 56, n. 4, p. 243–247, 1949. Disponível em: <<https://doi.org/10.1080/00029890.1949.11999368>>.
- ISO-3271:2015. *Iron ores for blast furnace and direct reduction feedstocks — Determination of the tumble and abrasion indices*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62135.html>>.
- ISO-4695:2015. *Iron ores for blast furnace feedstocks — Determination of the reducibility by the rate of reduction index*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62141.html>>.
- ISO-4696-1:2015. *Iron ores for blast furnace feedstocks — Determination of low-temperature reduction-disintegration indices by static method — Part 1: Reduction with CO, CO₂, H₂ and N₂*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62136.html>>.
- ISO-4696-2:2015. *Iron ores for blast furnace feedstocks — Determination of low-temperature reduction-disintegration indices by static method — Part 2: Reduction with CO and N₂*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62137.html>>.
- ISO-7215:2015. *Iron ores for blast furnace feedstocks — Determination of the reducibility by the final degree of reduction index*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62142.html>>.
- ISO-7992:2015. *Iron ores for blast furnace feedstocks — Determination of reduction under load*. Geneva, CH, 2015. International Organization for Standardization (ISO). Disponível em: <<https://www.iso.org/standard/62143.html>>.
- JIS-M8711:2011. *Iron ore sinter — Determination of shatter strength*. Tokyo, Japan, 2011. Japanese Standards Association (JSA). Disponível em: <https://global.ihs.com/doc_detail.cfm?document_name=JIS\%20M\%208711&item_s_key=00504147>.
- KINNUNEN, K.; LAITINEN, P. Investigation of sinter plant production rate and rdi by neural networks. *Metallurgical Research & Technology*, EDP Sciences, v. 102, n. 5, p. 361–371, 2005. Disponível em: <<https://doi.org/10.1051/metal:2005130>>.
- KRÖSE, B.; SMAGT, P. V. D. *An introduction to neural networks*. 8th. ed. Amsterdam, Holand: The University of Amsterdam, 1996. Disponível em: <<https://www.infor.uva.es/~teodoro/neuro-intro.pdf>>.
- KUHN, M.; JOHNSON, K. Over-fitting and model tuning. In: *Applied Predictive Modeling*. New York, NY: Springer New York, 2013. p. 61–92. Disponível em: <https://doi.org/10.1007/978-1-4614-6849-3_4>.
- KUHN, M.; JOHNSON, K. A short tour of the predictive modeling process. In: *Applied Predictive Modeling*. New York, NY: Springer New York, 2013. p. 19–26. ISBN 978-1-4614-6849-3. Disponível em: <https://doi.org/10.1007/978-1-4614-6849-3_2>.

- KUMAR, V. et al. Prediction of iron ore sinter properties using statistical technique. *Transactions of the Indian Institute of Metals*, The Indian Institute of Metals, v. 70, p. 1661–1670, 2017. Disponível em: <<https://doi.org/10.1007/s12666-016-0964-y>>.
- LAITINEN, P. J.; SAXÉN, H. A neural network based model of sinter quality and sinter plant performance indices. *Ironmaking & Steelmaking*, Taylor & Francis, v. 34, n. 2, p. 109–114, 2007. Disponível em: <<https://doi.org/10.1179/174328107X155312>>.
- LAY, D. C. *Linear Algebra and Its Applications*. 4th. ed. Baltimore, Maryland: Pearson Education, 2011.
- LINHARES, F. M. *Utilização de maiores percentuais de pellet feed na etapa de aglomeração a frio da sinterização de minério de ferro a partir da adição de aglomerantes*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2019. Disponível em: <<http://hdl.handle.net/1843/33589>>.
- LITSTER, J.; WATERS, A. Influence of the material properties of iron ore sinter feed on granulation effectiveness. *Powder Technology*, Elsevier, v. 55, n. 2, p. 141–151, 1988. Disponível em: <[https://doi.org/10.1016/0032-5910\(88\)80097-4](https://doi.org/10.1016/0032-5910(88)80097-4)>.
- LOBO, B. A. Métodos estatísticos de previsão – Análise de Erros. Faculdade de Engenharia FEUP, Universidade do Porto, Portugal. Unpublished. 2007. Disponível em: <https://sigarra.up.pt/feup/pt/conteudos_service.conteudos_cont?pct_id=34385&pv_cod=1112qaTawyc8>.
- LU, L.; HOLMES, R.; MANUEL, J. Effects of alumina on sintering performance of hematite iron ores. *Isij International - ISIJ INT*, v. 47, p. 349–358, 04 2007.
- LU, L.; ISHIYAMA, O. 14 - iron ore sintering. In: LU, L. (Ed.). *Iron Ore*. Woodhead Publishing, 2015. p. 395–433. ISBN 978-1-78242-156-6. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9781782421566000149>>.
- LU, L.; PAN, J.; ZHU, D. 16 - quality requirements of iron ore for iron production. In: LU, L. (Ed.). *Iron Ore*. Woodhead Publishing, 2015. p. 475–504. ISBN 978-1-78242-156-6. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B9781782421566000162>>.
- LV, C. et al. Levenberg–marquardt backpropagation training of multilayer neural networks for state estimation of a safety-critical cyber-physical system. *IEEE Transactions on Industrial Informatics*, IEEE, v. 14, n. 8, p. 3436–3446, 2017. Disponível em: <<https://doi.org/10.1109/TII.2017.2777460>>.
- MAIA, A. G. *Econometria: conceitos e aplicações*. Saint Paul Editora, 2019. Disponível em: <<https://saintpaul.com.br/produto/econometria-conceitos-e-aplicacoes-1o-edicao-2017/>>.
- MCCULLOCH, W. S.; PITTS, W. A logical calculus of the ideas immanent in nervous activity. *The bulletin of mathematical biophysics*, Springer, v. 5, n. 4, p. 115–133, 1943. Disponível em: <<https://doi.org/10.1007/BF02478259>>.
- MEDEIROS, C.; BARRETO, G. Pruning the multilayer perceptron through the correlation of backpropagated errors. In: . [s.n.], 2007. p. 64 – 69. ISBN 978-0-7695-2976-9. Disponível em: <<http://dx.doi.org/10.1109/ISDA.2007.156>>.

MEDEIROS, L. A. d. S. *Acelerando o método de Levenberg-Marquardt para a minimização da soma de quadrados de funções com restrições em caixa*. Tese (Doutorado) — Universidade Estadual de Campinas, Campinas, São Paulo, 2008. Disponível em: <https://www.ime.unicamp.br/~chico/tese_luiz_medeiros.pdf>.

MINITAB, E. *Basta! Lidando com a multicolinearidade na análise de regressão*. 2019. Acesso: 20/10/2021. Disponível em: <<https://blog.minitab.com/pt/basta-lidando-com-a-multicolinearidade-na-analise-de-regressao>>.

MINSKY, M.; PAPER, S. *Perceptrons: An introduction to computational geometry*. MIT Press, Cambridge, MA, 1969. Disponível em: <<https://mitpress.mit.edu/books/perceptrons>>.

MITTERLEHNER, J. et al. Modeling and simulation of heat front propagation in the iron ore sintering process. *ISIJ International*, v. 44, n. 1, p. 11–20, 2004. Disponível em: <<https://doi.org/10.2355/isijinternational.44.11>>.

MOCHÓN, J. et al. Iron ore sintering part 2. quality indices and productivity. *Dyna*, v. 81, n. 183, p. 168–177, 2014. Disponível em: <<https://doi.org/10.15446/dyna.v81n183.41568>>.

MONTGOMERY, D. C.; PECK, E. A.; Vining, G. G. *Introduction to linear regression analysis*. 5th. ed. John Wiley & Sons, 2012. ISBN 978-0-470-54281-1. Disponível em: <<https://www.wiley.com/en-us/Introduction+to+Linear+Regression+Analysis+%2C+5th+Edition-p-9780470542811>>.

MÜLLER, B.; REINHARDT, J.; STRICKLAND, M. T. *Neural networks: an introduction*. 2nd. ed. New York, NY: Springer Science & Business Media, 1995. Disponível em: <<https://doi.org/10.1007/978-3-642-57760-4>>.

NAJAR, F.; SANTOS, J. *Curso de sinterização*. 1981. USIMINAS, Ipatinga, MG.

NATH, N. K.; SILVA, A. J. da; CHAKRABORTI, N. Dynamic process modelling of iron ore sintering. *Steel Research*, v. 68, p. 285–292, 1997. Disponível em: <<https://doi.org/10.1002/srin.199701791>>.

NISHIO, H. et al. Testing methods of sinter. In: . [s.n.], 1982. p. 90–97. Disponível em: <<https://www.scopus.com/inward/record.uri?eid=2-s2.0-0020266954&partnerID=40&md5=f835b427c1ba8d0aff9fa7dd52209216>>.

NWANKPA, C. et al. Activation functions: Comparison of trends in practice and research for deep learning. *arXiv.org*, 2018. Disponível em: <<https://arxiv.org/abs/1811.03378>>.

PADULA, V. P. *Caracterização tecnológica de um Sinter Feed hematítico contendo diferentes proporções e tipos de sílica nas frações aderentes, intermediárias e nucleantes*. Dissertação (Mestrado) — Universidade Federal de Ouro Preto, Ouro Preto, MG, 2009. Disponível em: <<http://www.repositorio.ufop.br/handle/123456789/2398>>.

PATRO, S. G.; SAHU, K. K. Normalization: A preprocessing stage. *IARJSET*, v. 2, n. 3, p. 20–22, 2015. Disponível em: <<https://doi.org/10.17148/IARJSET.2015.2305>>.

- PEREIRA, M. J. *Calcinação e sinterização de finos de minério de manganês sílico-carbonatado de Morro da Mina*. Dissertação (Mestrado) — Universidade Federal de Ouro Preto, Ouro Preto, MG, 2013. Disponível em: <<http://www.repositorio.ufop.br/jspui/handle/123456789/3319>>.
- PETRUCCI, E.; OLIVEIRA, L. A. de. Coeficientes de assimetria e curtose nos dados de vazão média mensal da bacia do rio preto - ba. *Os Desafios da Geografia Física na Fronteira do Conhecimento*, v. 1, p. 158–170, 2017. Disponível em: <<https://doi.org/10.20396/sbgfa.v1i2017.1785>>.
- PINEDA, F. Generalization of back propagation to recurrent and higher order neural networks. In: ANDERSON, D. (Ed.). *Neural Information Processing Systems*. American Institute of Physics, 1988. Disponível em: <<https://proceedings.neurips.cc/paper/1987/file/735b90b4568125ed6c3f678819b6e058-Paper.pdf>>.
- PITOMBO, C. S.; COSTA, A. S. G. d. Aplicação conjunta de modelos não paramétricos e paramétricos para previsão de escolha modal. *Journal of Transport Literature*, SciELO Brasil, v. 9, p. 30–34, 2015. Disponível em: <<https://doi.org/10.1590/2238-1031.jtl.v9n1a6>>.
- RIPLEY, B. D. *Pattern Recognition and Neural Networks*. Cambridge, MA: Cambridge University Press, 1996. ISBN 9780511812651. Disponível em: <<https://doi.org/10.1017/CBO9780511812651>>.
- RUMELHART, D. E.; HINTON, G. E.; WILLIAMS, R. J. Learning representations by back-propagating errors. *nature*, Nature Publishing Group, v. 323, n. 6088, p. 533–536, 1986. Disponível em: <<https://doi.org/10.1038/323533a0>>.
- SALVIAN, M. LCE 5801 regressão e covariância – Multicolinearidade. Universidade de São Paulo, Piracicaba, São Paulo, Unpublished. 2016. Disponível em: <https://edisciplinas.usp.br/pluginfile.php/2340848/mod_resource/content/0/Mayara_Multicolinearidade.pdf>.
- Mean squared error. In: SAMMUT, C.; WEBB, G. I. (Ed.). *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2010. p. 653–653. ISBN 978-0-387-30164-8. Disponível em: <https://doi.org/10.1007/978-0-387-30164-8_528>.
- SANTOS, J. D. A. *Sobre a Estimação do Número de Neurônios Ocultos da Rede MLP: Uma Nova Técnica Baseada em PCA e SVD*. Dissertação (Mestrado) — Universidade Federal do Ceará, Fortaleza, Ceará, 2010. Disponível em: <http://www.dominiopublico.gov.br/pesquisa/DetalheObraForm.do?select_action=&co_obra=199810>.
- SANTOS, J. D. A.; BARRETO, G. A.; MEDEIROS, C. M. S. Estimating the number of hidden neurons of the mlp using singular value decomposition and principal components analysis: A novel approach. In: *2010 Eleventh Brazilian Symposium on Neural Networks*. [s.n.], 2010. p. 19–24. Disponível em: <<https://doi.org/10.1109/SBRN.2010.12>>.
- SANTOS J. D. A. ; BARRETO, G. A.; MEDEIROS, C. M. S. Uma nova técnica de estimação de neurônios ocultos da rede mlp usando decomposição em valores singulares e análise de componentes principais. In: *XVIII Congresso Brasileiro de Automática (CBA'2010) – Mato Grosso do Sul: Bonito, 1994*. Mato Grosso do Sul: [s.n.], 2010. v. 1, p. 1741–1748.

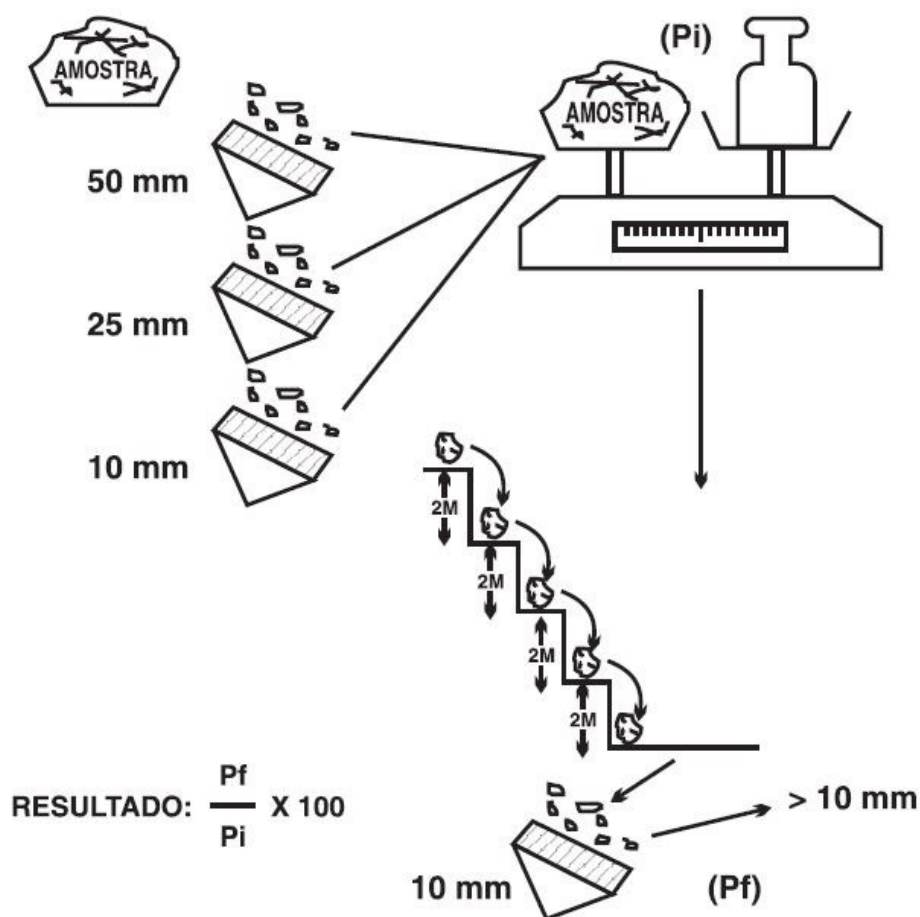
- SHAO, H. et al. Application of artificial neural networks for prediction of sinter quality based on process parameters control. *Transactions of the Institute of Measurement and Control*, v. 42, n. 3, p. 422–429, 2020. Disponível em: <<https://doi.org/10.1177/0142331219883501>>.
- SHARMA, S.; SHARMA, S.; ATHAIYA, A. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, v. 04, p. 310–316, 2020. Disponível em: <<http://dx.doi.org/10.33564/IJEAST.2020.v04i12.054>>.
- SILVA, D. F. B. F. d. *Pré-processamento de Dados e Comparação entre Algoritmos de Machine Learning para a Análise Preditiva de Falhas em Linhas de Produção para o Controle*. Dissertação (Mestrado) — Instituto Superior de Engenharia do Porto, Porto, Distrito do Porto, 2021. Disponível em: <<http://hdl.handle.net/10400.22/18266>>.
- SIVAKUMAR, A.; GUNASUNDARI, R. A survey on data preprocessing techniques for bioinformatics and web usage mining. *International Journal of Pure and Applied Mathematics*, v. 117, n. 20, p. 785–794, 2017. Disponível em: <<https://acadpubl.eu/jsi/2017-117-20-22/articles/20/68.pdf>>.
- STARLING, A. R. d. A. R. *Avaliação de metodologias de simulação física do processo desinterização*. Dissertação (Mestrado) — Universidade Federal de Minas Gerais, Belo Horizonte, MG, 2012. Disponível em: <<http://hdl.handle.net/1843/BUOS-8Z8M6H>>.
- STRANG, G. *Linear Algebra and its Applications*. 2nd. ed. Academic Press, 1980. Disponível em: <<https://www.sciencedirect.com/book/9780126736601/linear-algebra-and-its-applications>>.
- Mape (mean absolute percentage error). In: SWAMIDASS, P. M. (Ed.). *Encyclopedia of Production and Manufacturing Management*. Boston, MA: Springer US, 2000. p. 462–462. ISBN 978-1-4020-0612-8. Disponível em: <https://doi.org/10.1007/1-4020-0612-8_580>.
- TEOH, E. J.; TAN, K. C.; XIANG, C. Estimating the number of hidden neurons in a feedforward network using the singular value decomposition. *IEEE Transactions on Neural Networks*, v. 17, n. 6, p. 1623–1629, 2006. Disponível em: <<https://doi.org/10.1109/TNN.2006.880582>>.
- TRENN, S. Multilayer perceptrons: Approximation order and necessary number of hidden units. *IEEE Transactions on Neural Networks*, v. 19, n. 5, p. 836–844, 2008. Disponível em: <<https://doi.org/10.1109/TNN.2007.912306>>.
- TUFAN, B. et al. The influence of raw material composition on the quality of sinter. *Life Science Journal*, v. 10, p. 584–594, 01 2013.
- UMADEVI, T. et al. Optimisation of feo in iron ore sinter at jsw steel limited. *Ironmaking & Steelmaking*, Taylor Francis, v. 39, n. 3, p. 180–189, 2012. Disponível em: <<https://doi.org/10.1179/1743281211Y.0000000080>>.
- UMADEVI, T. et al. Studies on parameters affecting sinter strength and prediction through artificial neural network model. *Mineral Processing and Extractive Metallurgy*, Taylor & Francis, v. 125, n. 1, p. 32–38, 2016. Disponível em: <<https://doi.org/10.1179/1743285515Y.0000000020>>.

- UMADEVI, T.; SAH, R.; MAHAPATRA, P. C. Influence of sinter basicity (cao/sio₂) on low and high alumina iron ore sinter quality. *Mineral Processing and Extractive Metallurgy*, Taylor Francis, v. 123, n. 2, p. 75–85, 2014. Disponível em: <<https://doi.org/10.1179/1743285514Y.0000000052>>.
- WEBSTER, N. A. S. et al. Silico-ferrite of calcium and aluminum (sfca) iron ore sinter bonding phases: new insights into their formation during heating and cooling. *Metallurgical and Materials Transactions B*, Springer, v. 43, n. 6, p. 1344–1357, 2012. Disponível em: <<https://doi.org/10.1007/s11663-012-9740-5>>.
- WERBOS, P. J. Backpropagation: Past and future. In: IEEE. *Proceedings of the Second International Conference on Neural Network*. 1988. v. 1, p. 343–353. Disponível em: <<https://doi.org/10.1109/ICNN.1988.23866>>.
- WERBOS, P. J. Generalization of backpropagation with application to a recurrent gas market model. *Neural networks*, Elsevier, v. 1, n. 4, p. 339–356, 1988. Disponível em: <[https://doi.org/10.1016/0893-6080\(88\)90007-X](https://doi.org/10.1016/0893-6080(88)90007-X)>.
- WERBOS, P. J. Backpropagation through time: what it does and how to do it. *Proceedings of the IEEE*, IEEE, v. 78, n. 10, p. 1550–1560, 1990. Disponível em: <<https://doi.org/10.1109/5.58337>>.
- WIDROW, B. *An adaptive 'ADALINE' neuron using chemical 'memistors'*. Stanford, CA, 1960. Disponível em: <<https://www-isl.stanford.edu/~widrow/papers/t1960anadaptive.pdf>>.
- WIDROW, B. Generalization and information storage in network of 'adaline' neurons. *Self-organizing systems-1962*, Spartan Books, p. 435–462, 1962. Disponível em: <<https://isl.stanford.edu/~widrow/papers/c1961generalizationand.pdf>>.
- WIDROW, B.; HOFF, M. E. *Adaptive switching circuits*. Stanford, CA, 1960. Disponível em: <<https://apps.dtic.mil/sti/pdfs/AD0241531.pdf>>.
- YANG, W. et al. Modeling of combustion and heat transfer in an iron ore sintering bed with considerations of multiple solid phases. *ISIJ International*, v. 44, n. 3, p. 492–499, 2004. Disponível em: <<https://doi.org/10.2355/isijinternational.44.492>>.
- YU, W.-t. et al. The effects of high al₂o₃ on the metallurgical properties of sinter. In: _____. [s.n.], 2015. p. 419–425. ISBN 978-3-319-48601-7. Disponível em: <http://dx.doi.org/10.1007/978-3-319-48191-3_51>.
- ZHANG, G. et al. Optimization of dolomite usage in iron ore sintering process. *ISIJ International*, v. 53, n. 9, p. 1515–1522, 2013. Disponível em: <<https://doi.org/10.2355/isijinternational.53.1515>>.
- ZHOU, H. et al. Improvement in the permeability of sintering beds by drying treatment after granulating sinter raw materials containing concentrates. *Advanced Powder Technology*, v. 31, n. 8, p. 3297–3306, 2020. ISSN 0921-8831. Disponível em: <<https://doi.org/10.1016/j.appt.2020.06.017>>.
- ZHOU, H. et al. Effect of flame-front speed on the pisolite-ore sintering process. *Applied Thermal Engineering*, v. 75, p. 307–314, 2015. ISSN 1359-4311. Disponível em: <<https://doi.org/10.1016/j.applthermaleng.2014.09.050>>.

Anexos

ANEXO A – Teste Shatter

Figura 35 – Representação esquemática do teste Shatter

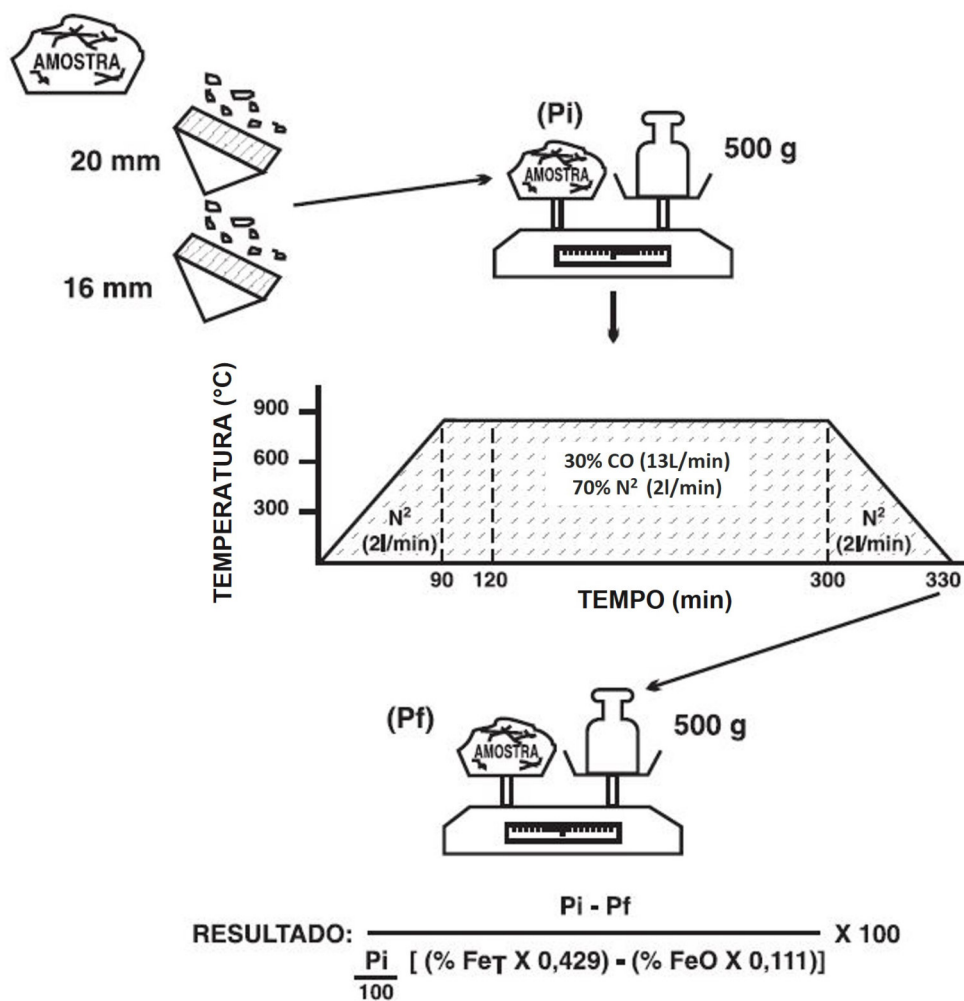


Fonte: [Honorato e Rangel \(1993\)](#)

A Figura 35 mostra esquematicamente as etapas do teste Shatter. A partir do conhecimento da distribuição granulométrica do material nas malhas de 10mm, 25mm e 50mm, forma-se uma amostra que sofre quatro quedas sucessivas de uma altura de 2m. O resultado passa por uma peneira de 10mm e o peso final da amostra após o teste é comparado com o peso inicial para fins de cálculo do índice.

ANEXO B – Teste de Redutibilidade

Figura 36 – Representação esquemática do teste de redutibilidade

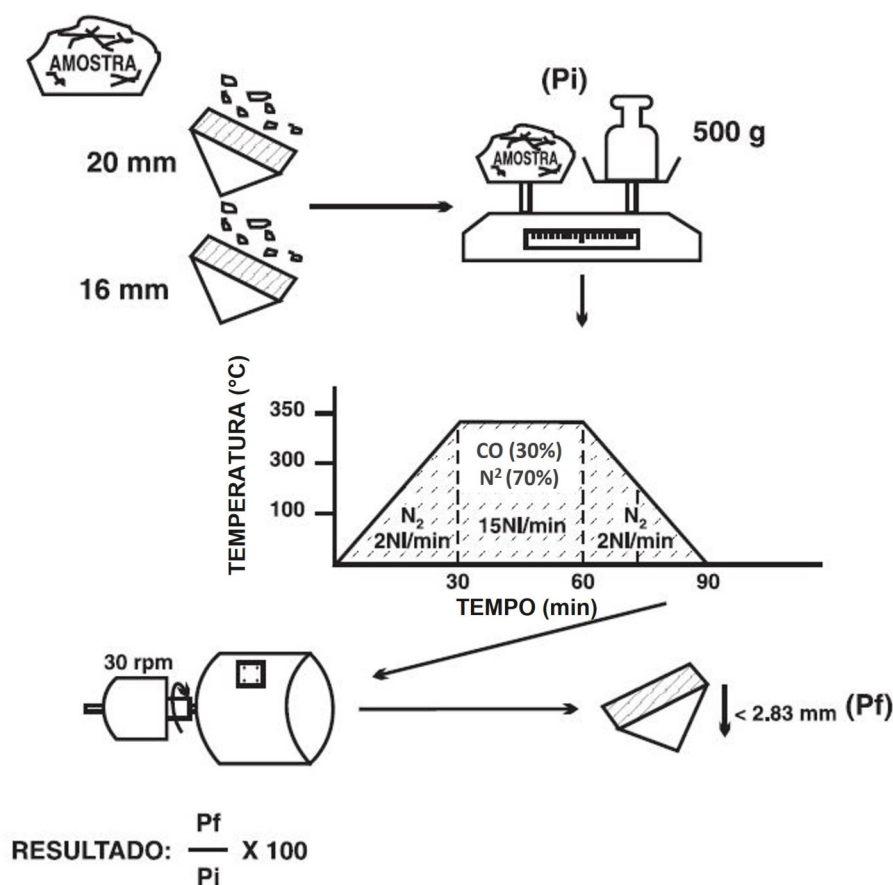


Fonte: [Honorato e Rangel \(1993\)](#)

A Figura 36 mostra as principais etapas de um teste de redutibilidade. A amostra original é classificada na faixa granulométrica entre 16mm a 20mm, pesada e então colocada em um forno submetida à temperaturas variadas, sofrendo redução pela ação de gases como N_2 e CO , conforme mostra a figura. Ao final, a amostra é pesada e o índice é calculado.

ANEXO C – Teste de Degradação Sob Redução

Figura 37 – Representação esquemática do teste de degradação sob redução



Fonte: Honorato e Rangel (1993)

A Figura 37 exemplifica as principais etapas de um teste de degradação sob redução. A amostra original é classificada na faixa granulométrica entre 16mm a 20mm, pesada e então colocada em um forno submetida à temperaturas variadas, sofrendo redução pela ação de gases como N_2 e CO , conforme mostra a figura. Essa etapa diferencia-se do teste de redutibilidade pelas menores temperatura e tempo de exposição da amostra ao ambiente redutor. Acrescenta-se ao teste um tamboramento final seguido de peneiramento em malha de 2.83mm. Ao final, a amostra é pesada e o índice calculado.