



**UNIVERSIDADE FEDERAL DO CEARÁ
CENTRO DE TECNOLOGIA
DEPARTAMENTO DE ENGENHARIA METALÚRGICA E DE MATERIAIS
GRADUAÇÃO EM ENGENHARIA METALÚRGICA**

FABRÍCIO DAMASCENO BRAGA

**USO DE REDES NEURAIS DE FUNÇÕES DE BASE RADIAL E REGRESSÃO
LINEAR MÚLTIPLA PARA A PREVISÃO DE PROPRIEDADES MECÂNICAS DE
VERGALHÕES DE AÇO.**

FORTALEZA

2019

FABRÍCIO DAMASCENO BRAGA

USO DE REDES NEURAIS DE FUNÇÕES DE BASE RADIAL E REGRESSÃO LINEAR
MÚLTIPLA PARA A PREVISÃO DE PROPRIEDADES MECÂNICAS DE VERGALHÕES
DE AÇO.

Monografia apresentada ao Curso de Engenharia Metalúrgica do Departamento de Engenharia Metalúrgica e de Materiais da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Engenharia Metalúrgica.

Orientador: Prof. Dr. Elineudo Pinho de Moura

FORTALEZA

2019

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)

B793u Braga, Fabrício Damasceno.

Uso de redes neurais de funções de base radial e regressão linear múltipla para a previsão de propriedades mecânicas de vergalhões de aço / Fabrício Damasceno Braga. – 2019.

64 f. : il. color.

Trabalho de Conclusão de Curso (graduação) – Universidade Federal do Ceará, Centro de Tecnologia, Curso de Engenharia Metalúrgica, Fortaleza, 2019.

Orientação: Prof. Dr. Elineudo Pinho de Moura.

1. Redes Neurais Artificiais. 2. Regressão Linear Múltipla. 3. Siderurgia. 4. Propriedades Mecânicas. I.
Título.

CDD 669

FABRÍCIO DAMASCENO BRAGA

USO DE REDES NEURAIS DE FUNÇÕES DE BASE RADIAL E REGRESSÃO LINEAR
MÚLTIPLA PARA A PREVISÃO DE PROPRIEDADES MECÂNICAS DE VERGALHÕES
DE AÇO.

Monografia apresentada ao Curso de Engenharia Metalúrgica do Departamento de Engenharia Metalúrgica e de Materiais da Universidade Federal do Ceará, como requisito parcial para obtenção do título de Bacharel em Engenharia Metalúrgica.

Aprovada em: ____/____/_____.

BANCA EXAMINADORA

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

Prof. Dr. Guilherme de Alencar Barreto
Universidade Federal do Ceará (UFC)

Ma. Raphaella Hermont Fonseca Murta
Universidade Federal do Ceará (UFC)

A Deus.

Aos meus pais, Fábio e Laélia.

AGRADECIMENTOS

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

Ao Prof. Dr. Elineudo Pinho de Moura, pela excelente orientação.

Aos participantes da banca examinadora Prof. Dr. Guilherme de Alencar Barreto e Ma. Raphaella Hermont Fonseca Murta pelo tempo, pelas valiosas colaborações e sugestões.

Aos colegas do Centro de Ensaios Não Destrutivos – CENDE/UFC, Pedro Paulo e Aldecira, por apoiarem todos os estudos no laboratório.

À minha família, em especial, Maria das Graças e Maria de Fátima pelo apoio e companheirismo.

Aos colegas da UFC, em especial, Andressa Lima, Jeová da Silva, Jamil Nobre, Dávila Oliveira, Bruno Henrique e Otílio Freire, pela amizade e companheirismo e por sempre me incentivarem nos estudos.

“O todo sem a parte não é todo, a parte sem o todo não é parte.” (Gregório de Matos)

RESUMO

Com o avanço da tecnologia torna-se cada vez mais necessária a utilização de ferramentas avançadas na busca da maximização da produção e redução dos custos nas indústrias siderúrgicas. O presente trabalho estuda dois algoritmos de previsão de dados: uma Rede Neural de Funções de Base Radial (RBFN) e uma Regressão Linear Múltipla (MLR) no intuito de prever 4 propriedades mecânicas principais resultantes da produção de vergalhões de aço: Limite de Escoamento (LE), Limite de Resistência à Tração (LR), Razão entre Limite de Resistência à Tração e Limite de Escoamento (LR/LE) e Alongamento (AL). Através de dados obtidos da aciaria e laminação, que vão desde a composição química do material até a temperaturas ao longo do processo, ritmo de produção e pressão e vazão da água no tratamento térmico final, busca-se encontrar um modelo matemático adequado que possa prever as propriedades mecânicas finais do produto resultante desse processo. O coeficiente de determinação (R^2) foi utilizado para determinar a qualidade do modelo encontrado nas diversas configurações testadas durante 100 rodadas de teste cada para os dois algoritmos, dividindo-se aleatoriamente os dados disponíveis em um conjunto de treinamento e um conjunto de teste. Para a regressão linear múltipla, encontrou-se, respectivamente, para o LE, LR, LR/LE e AL, um R^2 médio de 0,5146, 0,2388, 0,7633 e 0,5099, já para a rede neural, encontrou-se 0,8437, 0,7403, 0,8691 e 0,7015. Os resultados indicam que a RBFN conseguiu prever com mais precisão as propriedades quando comparada a MLR. As médias do R^2 de 100 rodadas da rede, considerando um problema real e complexo de uma siderúrgica, podem ser consideradas altas, representando uma ferramenta importante e de fácil implementação.

Palavras-chave: Redes Neurais Artificiais, Regressão Linear Múltipla, Siderurgia, Propriedades Mecânicas.

ABSTRACT

With the advancement of technology, it is becoming increasingly necessary to use advanced tools in the pursuit of maximizing production and cost reduction in the steel industries. The present work studies two data prediction algorithms: a Radial Base Function Neural Network (RBFN) and a Multiple Linear Regression (MLR) to predict 4 main mechanical properties resulting from the production of steel rebar: yield strength (LE), ultimate tensile strength (LR), ultimate tensile strength and yield strength (LR / LE) and elongation ratio (AL). Through data obtained from the steelmaking and rolling, ranging from the chemical composition of the material to temperatures throughout the process, production rate and water pressure and flow in the final heat treatment, we seek to find a suitable mathematical model that can predict the final mechanical properties of the product resulting from this process. The coefficient of determination (R^2) was used to determine the quality of the model found in the various configurations tested during 100 test rounds each for the two algorithms, randomly dividing the available data into one training set and one test set. For the multiple linear regression, for the LE, LR, LR / LE and AL, respectively, an average R^2 0.5146, 0.2388, 0.7633 and 0.5099 was found. For the neural network, 0.8437, 0.7403, 0.8691 and 0.7015. The results indicate that RBFN was able to more accurately predict the properties when compared to MLR. The R^2 averages of 100 network rounds, considering a real and complex problem of a steel mill, can be considered high, representing an important and easy implementation tool.

Keywords: Artificial Neural Network, Multiple Linear Regression, Siderurgy, Mechanical Properties.

LISTA DE FIGURAS

Figura 1	– Fluxo simplificado de produção do aço	18
Figura 2	– Esquema de um forno elétrico a arco	19
Figura 3	– Influência do teor de carbono nas propriedades mecânicas do aço-carbono .	21
Figura 4	– Ilustração dos principais componentes de um neurônio artificial (matemático)	29
Figura 5	– Ilustração gráfica de uma ANN	29
Figura 6	– Ilustração gráfica da função $f(x)$ definida pela equação 2.16 com variação de r sendo $c=0$ (à esquerda) e com variação de c sendo $r^2=0.9$ (à direita)	31
Figura 7	– Ilustração da função $h(x)$ constituída pela soma ponderada de funções gaussianas aleatórias	32
Figura 8	– Ilustração gráfica da função $f(x)$ definida pela Equação 2.17	33
Figura 9	– Componentes básicos de uma RBFN	34
Figura 10	– Representação da influência do bias	35
Figura 11	– Representação da grade criada pela variação de valores de hiperparâmetros no método <i>Grid Search</i>	40
Figura 12	– Representação da grade criada pela variação de valores de hiperparâmetros no método <i>Random Search</i>	41
Figura 13	– Representações gráficas do significado de R^2	43
Figura 14	– Exemplo da aplicação do método <i>Grid Search</i>	48
Figura 15	– Exemplo da aplicação do método <i>Random Search</i>	48
Figura 16	– Resultados das baterias de teste usando <i>Grid Search</i>	50
Figura 17	– Resultados das baterias de teste usando <i>Random Search</i>	52
Figura 18	– Exemplo de previsibilidade do LE de uma das rodadas de teste da RBFN ..	55
Figura 19	– Exemplo de previsibilidade do LR de uma das rodadas de teste da RBFN ...	55

Figura 20 – Exemplo de previsibilidade da razão LR/LE de uma das rodadas de teste da RBFN	56
Figura 21 – Exemplo de previsibilidade do AL de uma das rodadas de teste da RBFN ..	56
Figura 22 – Variação na previsibilidade de propriedades mecânicas conforme variação de elementos químicos no aço	58

LISTA DE TABELAS

Tabela 1 – Resumo dos dados	44
Tabela 2 – Resultados de previsão de propriedades mecânicas usando MLR	49
Tabela 3 – Regiões hiperparamétricas otimizadas	51
Tabela 4 – Melhores resultados de previsão do LE classificados de acordo com R^2 médio	53
Tabela 5 – Melhores resultados de previsão do LR classificados de acordo com R^2 médio	53
Tabela 6 – Melhores resultados de previsão do LR/LE classificados de acordo com R^2 médio	53
Tabela 7 – Melhores resultados de previsão do AL classificados de acordo com R^2 médio	54
Tabela 8 – Comparativo entre os melhores resultados encontrados por cada algoritmo de previsão	54

LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
AL	Alongamento
ANN	Redes Neurais Artificiais, tradução livre de <i>Artificial Neural Networks</i>
BOF	Forno à Oxigênio, tradução livre de <i>Basic Oxygen Furnace</i>
EAF	Forno Elétrico á Arco, tradução livre de <i>Electric Arc Furnace</i>
LE	Limite de Escoamento
LR	Limite de Resistência à Tração
LR/LE	Razão entre Limite de Resistência à Tração e Limite de Escoamento
MLP	Percéptrons de Múltiplas Camadas, tradução livre de <i>Multilayer Perceptrons</i>
MLR	Regressão Linear Múltipla, tradução livre de <i>Multilinear Regression</i>
MW	Mega Watt
NBR	Norma Brasileira Regulamentar
R^2	Coeficiente de Determinação
RBF	Função de Base Radial, tradução livre de <i>Radial Basis Function</i>
RBFN	Rede de Função de Base Radial, tradução livre de <i>Radial Basis Function Network</i>

SUMÁRIO

1	INTRODUÇÃO	14
2	REVISÃO BIBLIOGRÁFICA	16
2.1	Produção de vergalhões de aço	16
2.1.1	<i>Definição de aço</i>	16
2.1.2	<i>Indústria siderúrgica</i>	16
2.1.3	<i>Aciaria Elétrica</i>	18
2.1.4	<i>Laminação</i>	19
2.2	Propriedades mecânicas	20
2.3	Previsão de dados	22
2.3.1	<i>Problema da aproximação de funções</i>	23
2.3.2	<i>Regressão linear múltipla (MLR)</i>	24
2.3.2.1	<i>Aspectos gerais</i>	24
2.3.2.2	<i>Método dos mínimos quadrados</i>	25
2.3.3	<i>Redes neurais artificiais (ANN)</i>	28
2.3.3.1	<i>Aspectos Gerais</i>	28
2.3.3.2	<i>Redes neurais de funções de base radial (RBFN)</i>	30
2.3.3.2.1	<i>Funções de base radial (RBF)</i>	30
2.3.3.2.2	<i>Estrutura da rede</i>	33
2.3.3.2.3	<i>Método de solução</i>	36
2.3.3.3	<i>Otimização de hiperparâmetros</i>	39
2.3.3.3.1	<i>Grid Search</i>	39
2.3.3.3.2	<i>Random Search</i>	40
2.4	Estatísticas de validação	41
2.4.1	<i>Coeficiente de determinação (R^2)</i>	42
3	METODOLOGIA	43
3.1	Dados	43
3.2	Ajuste da MLR	45
3.3	Ajuste da RBFN	46
3.4	Obtenção dos hiperparâmetros	47
4	RESULTADOS E DISCUSSÕES	49
5	CONCLUSÃO	59

1 INTRODUÇÃO

Diante da crescente competitividade na indústria siderúrgica, ferramentas e soluções avançadas são criadas para buscar a maximização da produção do aço, seja através da inovação do processo com pesquisas e testes que buscam unir elementos da literatura de referência com a escala do mundo industrial, seja através do melhor entendimento do modo de produção já utilizado que demanda maior previsibilidade (RIBEIRO; VIEIRA, 2001).

Análise de dados e *machine learning* são exemplos de áreas de onde se tem encontrado proveito para essa problemática. Busca-se analisar, entender e prever, através de modelos numéricos baseados em dados de produção, características do processo que não podem ser descritas analiticamente de forma simples, visto que a natureza complexa de um processo produtivo real inviabiliza, em parte, a obtenção de modelos analíticos simplistas que consideram as relações entre entrada/insumos e saída/respostas lineares (MARCOS; ELÍAS, 2017).

O problema em foco é o de previsão de propriedades mecânicas de vergalhões de aço utilizados na construção civil a partir de parâmetros do processo siderúrgico que envolvem desde a composição química da matéria-prima até parâmetros do processo de laminiação tais como: temperatura, grau de redução de área e velocidade de produção.

As propriedades mecânicas têm papel essencial na comercialização dos produtos mencionados, visto que são os principais parâmetros restritivos, por norma NBR 7480 (ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS, 2007), que podem impedir sua venda. Outros parâmetros como composição química final e modo de operação de laminiação são livres e podem ser escolhidos a critério do fabricante.

O processo convencional da empresa que forneceu os dados de produção não se aproveita de ferramentas matemáticas mais complexas como as propostas no trabalho, tenta apenas unir o conhecimento prático, experiência e referências na literatura, para a previsão de propriedades mecânicas como: limite de escoamento, limite de resistência à tração e alongamento. Dessa forma, existem muitas perdas por produtos fora do padrão, já que não se tem uma estimativa prévia das características mecânicas do material de acordo com dados iniciais do processo.

Existem diversas formas de se obter modelos numéricos que tentam representar um processo real a nível industrial, as usadas no presente trabalho são: Regressão Linear Múltipla, tradução livre de *Multiple Linear Regression* (MLR) e Rede Neural de Funções de Base Radial, tradução livre de *Radial Basis Function Network* (RBFN). Essas ferramentas buscam encontrar

uma espécie de função de aproximação que representa uma relação entre entrada e saída de dados, sendo capazes de prever com certa precisão uma resposta a partir de um estímulo inicial de entrada.

O intuito do trabalho é comparar a qualidade da aproximação realizada por esses métodos e avalia-la por meio do coeficiente de determinação, denominado de R^2 , entre dados de produção medidos e valores estimados. Para cada método e propriedade estimada foram executadas 100 rodadas de verificação para subconjuntos de dados de treinamento e de teste sorteados aleatoriamente. As propriedades estudadas são: limite de escoamento (LE), limite de resistência à tração (LR), razão entre limite de resistência à tração e limite de escoamento (LR/LE) e alongamento (AL). Todas as verificações são realizadas através de algoritmos desenvolvidos na linguagem de programação MATLAB.

Espera-se verificar que um modelo mais complexo e otimizado como uma rede neural artificial tenha uma melhor previsibilidade de dados que um modelo simples como a regressão multilinear e, dessa forma, mostrar que a implementação de métodos mais avançados de análise de dados pode ser viável e trazer ganhos à indústria siderúrgica.

2 REVISÃO BIBLIOGRÁFICA

Este capítulo contém informações relevantes sobre o assunto abordado e serve de base para o entendimento das técnicas e métodos utilizados ao longo do trabalho. Na seção 2.1 são abordados os aspectos metalúrgicos/siderúrgicos pertinentes relacionados à produção de vergalhões de aço. Já a seção 2.2 explica a influência de alguns parâmetros nas propriedades mecânicas estudadas. As seções 2.3 e 2.4 são destinadas a explicação dos métodos computacionais utilizados, tanto aspectos teóricos quanto práticos, bem como formas de validação dos resultados.

2.1 Produção de vergalhões de aço

2.1.1 Definição de aço

Segundo Chiaverini (1977), de forma geral, o aço-carbono ou aço comum pode ser entendido como uma liga ferro-carbono contendo geralmente entre 0,008% até aproximadamente 2,11% de carbono, além de certos elementos residuais, resultantes do processo de fabricação. O limite inferior de 0,008% refere-se à máxima solubilidade do carbono no ferro à temperatura ambiente e o limite superior de 2,11% à máxima quantidade de carbono que se dissolve no ferro à 1148°C. Já os aços-liga são aços que contém outros elementos de liga ou apresenta elementos residuais acima do que são considerados normais.

2.1.2 Indústria siderúrgica

A indústria siderúrgica dedica-se à fabricação de aço na forma de produtos semiacabados, laminados, trefilados e em tubos sem costura, e é classificada como um ramo da metalurgia, no qual outras atividades estão inseridas, como: produção independente de gusa e produção de ferro-ligas (Instituto Aço Brasil, 2019).

Araújo (1997) diz que existem duas rotas básicas para a produção do aço comum nas indústrias siderúrgicas: através de fornos à oxigênio, tradução livre de *Basic Oxygen Steelmaking* (BOF) e através de fornos EAF, sendo os primeiros responsáveis por 60% da produção mundial de aço.

Dentre os principais insumos empregados na indústria siderúrgica estão, para as usinas integradas, minério de ferro e carvão, e para as usinas semi-integradas, sucata e energia

elétrica. Cada um tendo uma importância relativa dependendo da rota tecnológica de produção a ser utilizada (COSTA V.; ESCORSIM; COSTA D., 2007).

Dessa forma, pode-se classificar as usinas siderúrgicas segundo seu processo produtivo (Instituto Aço Brasil, 2019):

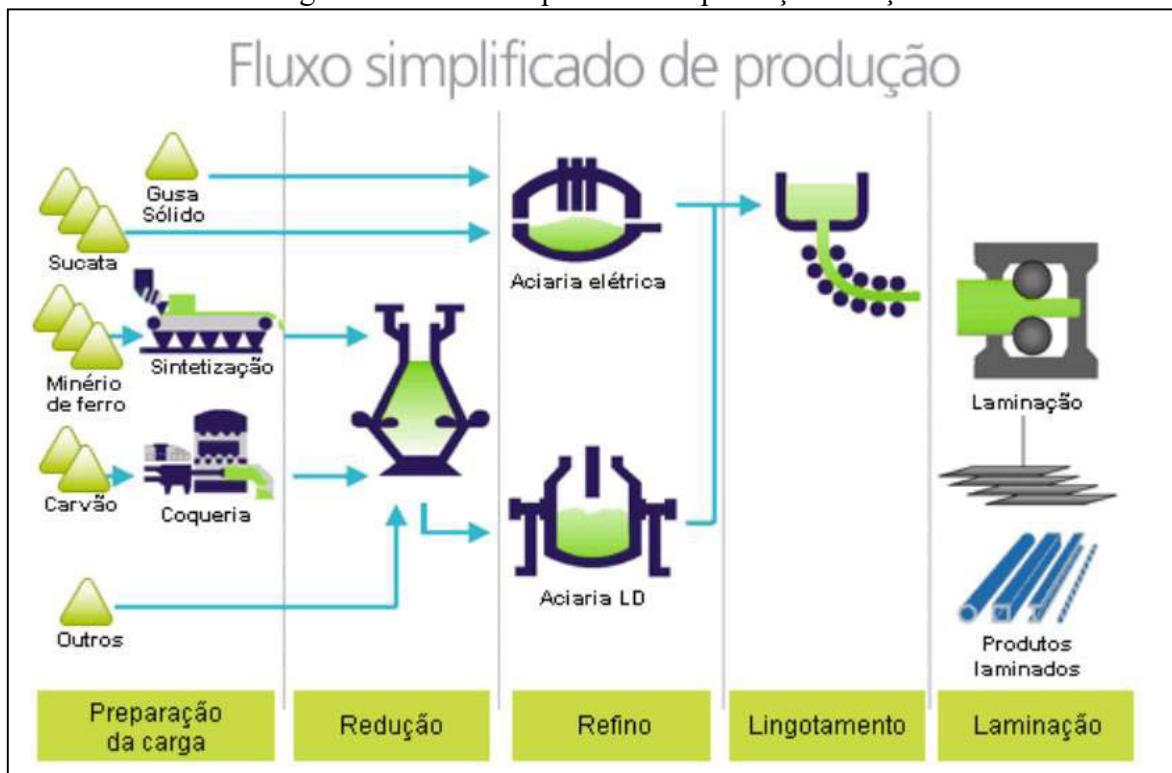
- Usinas integradas – que operam em 3 fases básicas: redução, refino e laminação; participam de todo o processo produtivo do aço; utilizam-se da rota de fornos à oxigênio.
- Usinas semi-integradas (*mini-mills*) – que operam em duas fases básicas: refino e laminação; utilizam-se da rota de fornos elétricos.

Assim, existem 3 etapas principais da produção do aço e mais uma etapa preparatória, são elas:

- Preparação da carga – Os finos de minério de ferro são aglomerados utilizando-se cal e finos de coque, resultando no sínter. O carvão, em sua maior parte mineral, é processado na coqueria e transforma-se em coque.
- Redução – Os insumos já preparados são adicionados ao alto-forno, onde também é injetado, por sopro por de baixo do forno, oxigênio a uma temperatura de 1000°C. O carvão em contato com o oxigênio produz uma reação de combustão (exotérmica), gerando calor e aquecendo a carga metálica, dando início ao processo de redução do minério de ferro em um metal líquido chamado de ferro-gusa – uma liga de ferro-carbono com alto teor de carbono (entre 3,5% e 4,5%) (SILVA; MEI, 2008).
- Refino – Aciarias a oxigênio ou elétricas transformam o gusa líquido ou sólido e a sucata de ferro em aço líquido, sendo responsáveis pela composição química final do produto. Dessa forma, podem contemplar etapas de descarburação, desoxidação, dessulfuração e desfosforilação (SILVA; MEI, 2008). A maior parte do aço líquido é solidificada no processo de lingotamento contínuo para produção de semi-acabados, lingotes e blocos.
- Laminação – Os produtos da aciaria sofrem conformações mecânicas a quente ou a frio nos chamados laminadores, que reduzem e modificam a forma do semi-acabado até a obtenção do produto final, cuja nomenclatura depende de sua forma e/ou composição química (CALLISTER JR, 2002).

A Figura 1 mostra o fluxo simplificado de produção do aço para as duas rotas básicas já citadas: a primeira, a rota da aciaria elétrica, utiliza-se de sucata e/ou gusa sólido adicionado de muita energia elétrica e uma parte energia química de combustão como forma de produção; já a segunda, a rota da aciaria à oxigênio, utiliza-se de insumos como sínter, coque, carvão, entre outros, sendo a energia química a principal fonte de energia. Nas etapas finais, tem-se o lingotamento e laminação para obtenção do produto final.

Figura 1 – Fluxo simplificado de produção do aço



Fonte: Adaptado de Instituto Aço Brasil (2019).

2.1.3 Aciaria Elétrica

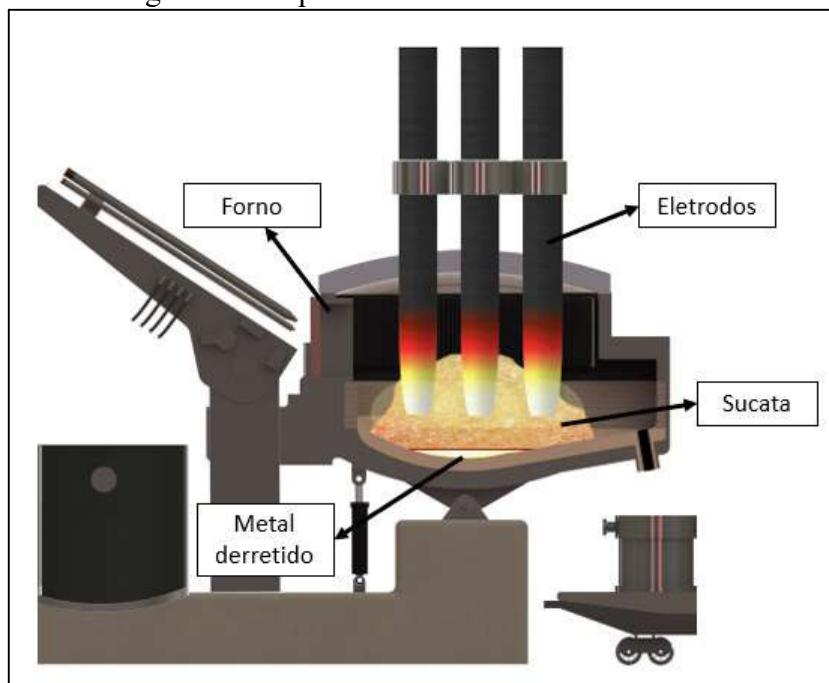
A aciaria engloba as etapas de refino e lingotamento do aço, sendo a aciaria elétrica – objeto o local de realização dessas etapas tendo início em fornos elétricos, como EAF, ou seja, faz parte do processo de uma usina semi-integrada (RIZZO, 2006).

O EAF é hoje a maneira mais comum de reciclar aço a partir de sucata. Ao derreter a sucata em um forno com a ajuda de eletrodos e uma corrente elétrica, um novo aço funcional pode ser produzido a partir de produtos antigos, o que é benéfico do ponto de vista econômico e ambiental. O calor necessário para derreter o aço é fornecido por arcos elétricos, criados entre os eletrodos e a sucata no forno. A energia elétrica dos EAF tradicionais fica na faixa de 50 a

120 MW, dependendo do tamanho do forno. Já o derretimento da sucata ocorre a uma faixa de temperatura entre 1500 e 1550°C, dependendo da composição da sucata. Após a fusão do aço e o refino do banho líquido até a temperatura e composição química desejadas, o produto é vazado a uma panela que irá realizar o refino secundário, etapas em que a composição química final do aço é obtida e, posteriormente, o lingotamento, onde ocorre a solidificação do produto na forma de tarugos ou placas (STEEL UNIVERISTY, 2019).

A Figura 2 mostra o esquema de um forno elétrico a arco e seus principais componentes.

Figura 2 – Esquema de um forno elétrico a arco



Fonte: Adaptado de Steel University (2019).

2.1.4 Laminação

Após lingotamento ocorre a etapa de laminação do aço para formação do produto final. A laminação é um processo de conformação mecânica que consiste essencialmente na passagem de um corpo sólido entre dois cilindros que giram na mesma velocidade periférica, mas em direções opostas. O corpo inicial tendo uma dimensão maior que a distância entre as superfícies laterais dos cilindros sofre uma deformação plástica na passagem, o que resulta na redução de sua seção transversal e no aumento de seu comprimento e largura. Para obter uma certa espessura a peça deve ser submetida a passagens sucessivas pelos cilindros a medida que

diminui as distâncias entre eles. A passagem da peça pelos cilindros ocorre pela ação da força de atrito que atua na superfície de contato entre as peças e os cilindros (ASM METALS, 1993).

A usina estudada no presente trabalho utiliza o processo de laminação a quente de tarugos provenientes do lingotamento contínuo do material produzido pelo EAF para a geração de vergalhões de aço. O laminador possui, ainda, um processo adicional que atua como tratamento térmico.

De forma resumida, o processo dessa siderúrgica pode ser dividido entre 4 zonas principais: forno de reaquecimento, trem de laminação inicial (zona de desbaste), trem médio e trem de acabamento. O forno de reaquecimento visa atingir temperaturas ligeiramente superiores à temperatura de recristalização do material. No trem inicial ocorre o primeiro passe de laminação, onde a conformação do metal tem início através de uma gaiola de laminação – equipamento que inclui rolamentos, cilindros e guias de rolamento. No trem médio, composto por até 4 gaiolas, ocorrem as maiores reduções de área. Finalmente, no trem de acabamento, que utiliza até 6 gaiolas, ocorrem pequenas reduções e visa dar as características finais como frisos, fitas e logotipo da empresa (ROBERTS, 1983).

A etapa final atua como um tratamento térmico, ele consiste na inserção de água pressurizada através de tubos sobre a superfície do vergalhão para que ocorra uma intensa troca de calor, seguido da atuação de sopradores de ar que retiram o excesso de água da superfície. Esse sistema de troca térmica garante que o vergalhão sofra um resfriamento homogeneamente forçado ao longo de toda a sua superfície, que produz uma camada resistente de martensita, dependendo do teor de carbono existente. Caso o teor de carbono não seja suficientemente alto, podem se formar outras camadas de microestruturas menos resistentes. Simultaneamente ao resfriamento da camada superficial, o núcleo experimenta uma taxa de resfriamento mais lenta e gera fase mais dúctil de ferrita e/ou uma microestrutura perlítica. Quando o vergalhão sai do sistema, o resfriamento forçado com água é interrompido e ele segue resfriando ao ar até a temperatura ambiente. O núcleo ainda quente transfere o calor para a superfície do vergalhão, reaquecendo-o e fazendo com que sua camada de martensita sofra revenimento, gerando as propriedades mecânicas desejadas no final do processo (MADIAS; WRIGHT; WOLKOWICZ, 2016; ARAÚJO, 2010).

2.2 Propriedades mecânicas

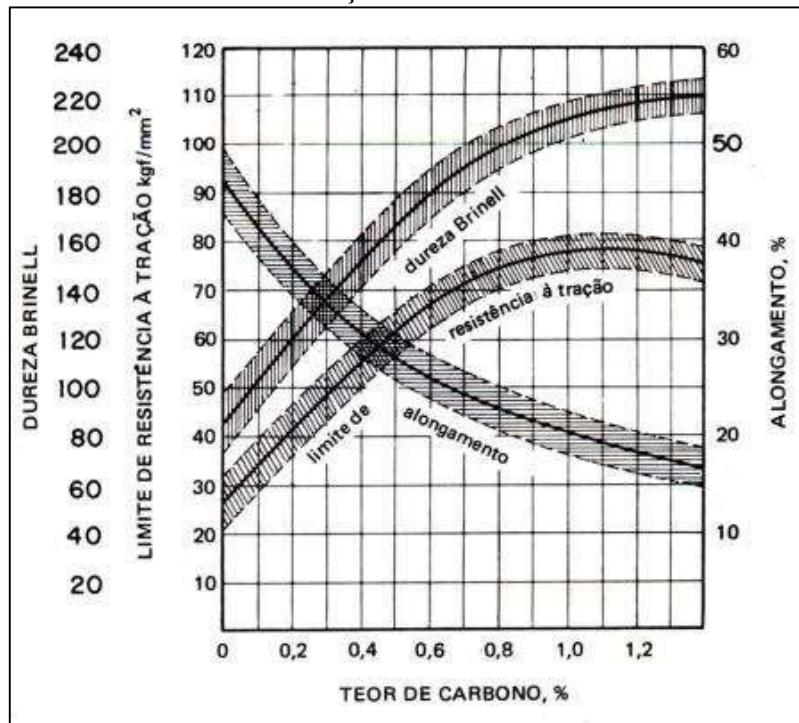
Segundo Chiaverini (1977), as propriedades mecânicas dos aços-carbono estão sujeitas basicamente a dois fatores: composição química e microestrutura. Em relação à

composição química, para os aços comerciais, o elemento mais importante é o carbono. Contudo, faz-se necessário considerar também a influência de elementos residuais. No que se refere à microestrutura, pode-se dizer que é influenciada por fatores como: velocidade de resfriamento e temperaturas de deformação.

O teor de carbono é um dos fatores com maior influência nas propriedades mecânicas dos aços comuns. Em geral, a resistência mecânica cresce (limite de resistência à tração, limite de escoamento e dureza) e a ductilidade e o alongamento diminuem à medida que o teor de carbono aumenta, conforme Figura 3. O teor desse elemento também influencia nos tratamentos térmicos posteriores, podendo facilitar ou impedir que ocorram.

O carbono, por tratar-se de aço comercial, exerce a maior das influências nas propriedades mecânicas finais. De maneira resumida, pode-se dizer que: à medida que o teor de carbono aumenta, crescem os valores representativos de resistência mecânica (limite de resistência à tração, limite de escoamento e dureza), ao passo que caem propriedades relacionadas a ductilidade como o alongamento, conforme Figura 3. Também influencia nos tratamentos térmicos posteriores, podendo facilitar ou impedir que ocorram (CHIAVERINI, 1977).

Figura 3 – Influência do teor de carbono nas propriedades mecânicas do aço-carbono



Fonte: Adaptado de Chiaverini (1977)

Outros elementos também exercem sua parcela de contribuição (CHIAVERINI, 1977):

- Enxofre: pode causar a chamada “fragilidade a quente” quando há a formação do sulfeto de ferro, o que prejudica a correta conformação mecânica do produto, também prejudica a resistência mecânica como um todo quando a porcentagem de carbono no banho metálico não é tão baixa;
- Manganês: pode formar sulfeto de manganês, prevenindo a “fragilidade a quente” e/ou óxido de manganês, contribuindo para a desoxidação. O restante do elemento incorporado ao aço se dissolve na ferrita, aumentando sua resistência mecânica;
- Silício: atua como desoxidante e o restante dissolve-se na ferrita sem afetar sua ductilidade, porém aumentando sua resistência mecânica;
- Fósforo: pode causar a chamada “fragilidade a frio”, dissolvendo-se na ferrita aumentando sua resistência mecânica e diminuindo sua ductilidade;
- Nitrogênio: quando dissolvido na ferrita em aços baixo carbono promove o endurecimento por precipitação, aumentando a resistência mecânica, contudo diminuindo drasticamente suas propriedades dúcteis;
- Outros elementos: de um modo geral, aumentam a resistência mecânica quando se dissolvem na ferrita, mesmo sem haver tratamento térmico.

2.3 Previsão de dados

Segundo Marsland (2014), o aprendizado através de dados (das máquinas) pode ser estendido, sem muita confusão e apenas de forma conceitual, para o aprendizado através de experiências (dos animais, humano). O aprendizado como um todo permite aos humanos adaptar-se e ter flexibilidade para tudo na vida, tendo base em ideias como: memória associativa, adaptação e generalização. Dessa forma, o termo *machine learning*, que na tradução literal significa “aprendizado de máquinas”, consiste em fazer com que os computadores adaptem ou modifiquem suas ações para que fiquem mais precisas, onde a precisão é medida pela forma como as ações escolhidas refletem as ações corretas.

Assim, a previsão de dados, quando relacionada a *machine learning*, representa a resposta que um algoritmo de aprendizado dá após treinamento com dados similares, em uma

tentativa de responder a um novo estímulo de entrada de uma forma parecida ao realizado durante fase de aprendizado (OBERMEYER; EMANUEL, 2016).

2.3.1 Problema da aproximação de funções

O desenvolvimento a seguir tem como base o texto de Von Zubem e Attux (2001). Estimar um parâmetro de saída por meio de uma série de parâmetros de entrada a partir de um conjunto de exemplos pode ser entendido como um problema de aproximação de uma função multidimensional. Esse processo pode ser resumido, em outras palavras, como a sintetização de uma memória associativa capaz de retornar valores aproximados do real a partir de um estímulo de entrada e, dessa forma, estender e generalizar quando novos dados são apresentados.

Em teoria, busca-se aproximar uma função $y(\mathbf{x})$ por uma função de aproximação $f(\mathbf{x}, \mathbf{w})$, onde \mathbf{w} representam parâmetros e \mathbf{x} o conjunto de dados/exemplos. Sendo assim, a função $f(\cdot)$ e os parâmetros \mathbf{w} são dois grandes objetos de estudo e identificação. Contudo, escolhida uma função $f(\cdot)$ específica, o problema é reduzido à determinação do conjunto de parâmetros \mathbf{w} que fornece os melhores resultados de aproximação em relação a $y(\mathbf{x})$ a partir dos exemplos iniciais.

Matematicamente, diz-se que, para uma função $y(\mathbf{x})$ contínua sobre o conjunto \mathbf{x} e uma função de aproximação $f(\mathbf{x}, \mathbf{w})$ que depende continuamente de \mathbf{w} e de \mathbf{x} , o problema de aproximação resume-se a determinar parâmetros \mathbf{w}^* tais que:

$$m[f(\mathbf{x}, \mathbf{w}^*), y(\mathbf{x})] < m[f(\mathbf{x}, \mathbf{w}), y(\mathbf{x})] \quad (2.1)$$

Onde $m[\cdot, \cdot]$ representa uma métrica de distância que permite validar a qualidade da aproximação.

Assim, o processo de aproximação consiste na coleta de dados de entrada/exemplos e suas respectivas saídas desejadas $\{(\mathbf{x}_1, \mathbf{m}_1), \dots, (\mathbf{x}_n, \mathbf{m}_n)\}$ e na aplicação do método de escolha de parâmetros \mathbf{w} definido.

Existem dois tipos básicos de aproximação/regressão: **paramétrica** - onde a relação entre variáveis dependentes e independentes é conhecida e, geralmente, tem uma interpretação física, como o caso de aproximação de uma reta através de pontos em um espaço bidimensional; e a **não-paramétrica** - onde há ausência do conhecimento das relações entre as variáveis, ou seja, da função que está sendo estimada, existindo um conjunto diversificado de características

e formas que ela pode assumir, sendo consequência direta disso a existência de um elevado número de parâmetros que não admitem uma interpretação física isolada, como o caso dos pesos sinápticos de uma rede neural artificial (Caçao, 2010).

2.3.2 Regressão linear múltipla (MLR)

A MLR é um tipo de aproximação/regressão paramétrica 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 no espaço (PITOMBO; COSTA, 2015).

2.3.2.1 Aspectos Gerais

Tomando-se como base o problema de aproximação de uma função, conforme mencionado no tópico anterior, tem-se a função de aproximação genérica $f(\mathbf{x}, \mathbf{w})$ onde, na MLR, define-se que $f(\cdot)$ é linear e \mathbf{w} são os coeficientes de regressão - objetos de identificação do modelo.

O objetivo principal desse método é estabelecer uma equação que possa ser utilizada para predizer 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 por:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i , \text{ para } i = 1, 2, \dots, n \quad (2.2)$$

Onde y_i representa o valor da variável dependente analisada na i -ésima observação de um total de n dados, x_i representa o valor de uma das k variáveis independentes na i -ésima observação de n dados, k representa o número de variáveis independentes, β_0 representa a constante de regressão, β_i representa o valor de um dos coeficientes de regressão, e ε_i um valor aleatório (ou resíduo) do erro associado a equação no i -ésimo exemplo de n dados.

2.3.2.2 Método dos mínimos quadrados

A estimativa dos coeficientes β_i tem como base o método dos mínimos quadrados. Através dele, 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).

A Equação 2.2 pode ser reescrita como:

$$y_i = \beta_0 + \sum_{j=1}^k \beta_j x_{ij} + \varepsilon_i \quad (2.3)$$

A partir daí, aplicando-se a função dos mínimos quadrados denominada S , tem-se:

$$S(\beta_0, \beta_1, \dots, \beta_k) = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right)^2 \quad (2.4)$$

Essa função deve ser minimizada em relação a $\beta_0, \beta_1, \dots, \beta_k$, devendo satisfazer, dessa forma, as equações:

$$\frac{\partial S}{\partial \beta_0} \Big|_{\beta_0, \beta_1, \dots, \beta_k} = -2 \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right) = 0 \quad (2.5)$$

e

$$\frac{\partial S}{\partial \beta_j} \Big|_{\beta_0, \beta_1, \dots, \beta_k} = -2 \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^k \beta_j x_{ij} \right) x_{ij} = 0 \quad (2.6)$$

Simplificando-se as duas equações acima obtém-se as equações normais dos mínimos quadrados na forma escalar:

$$n\beta_0 + \beta_1 \sum_{i=1}^n x_{i1} + \beta_2 \sum_{i=1}^n x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{ik} = \sum_{i=1}^n y_i \quad (2.7)$$

e

$$\begin{aligned} \beta_0 \sum_{i=1}^n x_{i1} + \beta_1 \sum_{i=1}^n x_{i1}^2 + \beta_2 \sum_{i=1}^n x_{i1}x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{i1}x_{ik} &= \sum_{i=1}^n x_{i1}y_i \\ \vdots &\quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \quad \vdots \end{aligned} \quad (2.8)$$

$$\beta_0 \sum_{i=1}^n x_{ik} + \beta_1 \sum_{i=1}^n x_{ik}^2 + \beta_2 \sum_{i=1}^n x_{ik}x_{i2} + \dots + \beta_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n x_{ik}y_i$$

Observa-se que existem $p = k + 1$ equações normais, uma para cada coeficiente de regressão desconhecido (incluindo a constante de regressão). A solução para o sistema formado por essas equações será os estimadores dos mínimos quadrados $\beta_0, \beta_1, \dots, \beta_k$.

Contudo, é mais conveniente trabalhar essas equações no formato matricial. A Equação 2.3 pode ser traduzida no seguinte formato matricial:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (2.9)$$

$$\text{Onde: } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}, \boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}, \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Aplicando-se na função dos mínimos quadrados:

$$S(\boldsymbol{\beta}) = \sum_{i=1}^n \varepsilon^2 = \boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})$$

$$S(\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - \boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} - \mathbf{Y}'\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta}$$

$$S(\boldsymbol{\beta}) = \mathbf{Y}'\mathbf{Y} - 2\boldsymbol{\beta}'\mathbf{X}'\mathbf{Y} + \boldsymbol{\beta}'\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \quad (2.10)$$

Essa função deve satisfazer o critério a seguir de minimização da função em relação a $\boldsymbol{\beta}$, similar ao realizado nas Equações 2.5 e 2.6:

$$\left. \frac{\partial S}{\partial \boldsymbol{\beta}} \right|_{\boldsymbol{\beta}} = -2\mathbf{X}'\mathbf{Y} + 2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{0} \quad (2.11)$$

Que simplificada, fornece a equação:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y} \quad (2.12)$$

Essa equação é chamada de equação normal dos mínimos quadrados na forma matricial, análoga a sua representação escalar nas Equações 2.7 e 2.8. Os coeficientes de regressão podem ser encontrados multiplicando-se os dois lados da equação pela inversa da matriz $\mathbf{X}'\mathbf{X}$.

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \quad (2.13)$$

Desenvolvendo a Equação 2.13 tem-se:

$$\left[\begin{array}{ccccc} n & \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i2} & \cdots & \sum_{i=1}^n x_{ik} \\ \sum_{i=1}^n x_{i1} & \sum_{i=1}^n x_{i1}^2 & \sum_{i=1}^n x_{i1}x_{i2} & \cdots & \sum_{i=1}^n x_{i1}x_{ik} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n x_{ik} & \sum_{i=1}^n x_{ik}x_{i1} & \sum_{i=1}^n x_{ik}x_{i2} & \cdots & \sum_{i=1}^n x_{ik}^2 \end{array} \right] \left[\begin{array}{c} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{array} \right] = \left[\begin{array}{c} \sum_{i=1}^n y_i \\ \sum_{i=1}^n x_{i1}y_i \\ \vdots \\ \sum_{i=1}^n x_{ik}y_i \end{array} \right] \quad (2.14)$$

Se as multiplicações entre matrizes forem realizadas, fornece as equações normais dos mínimos quadrados na forma escalar, mostradas nas Equações 2.7 e 2.8. Ou seja, os dois caminhos, escalar e matricial, fornecem a possibilidade de chegar em um conjunto de equações que pode ser resolvido através da inversa de uma matriz. Contudo, como já mencionado, é de mais fácil compreensão e visualização a solução no formato matricial, conforme Equação 2.13.

A equação de estimativa do modelo de regressão linear múltipla tem a seguinte forma:

$$y^* = \mathbf{X}'\boldsymbol{\beta} = \beta_0 + \sum_{j=1}^k \beta_j x_j \quad (2.15)$$

Onde y^* representa o valor estimado correspondente a um valor y_i do conjunto de teste $\{(x_i, y_i)\}$ para $i = 1, 2, \dots, n$.

2.3.3 Redes Neurais Artificiais (ANN)

Redes Neurais Artificiais, tradução livre de *Artificial Neural Networks* (ANN), da forma mais resumida, são algoritmos que tentam simular o processamento neurológico humano, sendo capazes de assimilar, analisar, entender e generalizar informações (RAUBER, 2005).

2.3.3.1 Aspectos Gerais

Para melhor compreensão do método de processamento desses algoritmos, faz-se necessário um breve resumo sobre o comportamento do sistema nervoso biológico e, mais especificamente, dos neurônios.

De maneira geral, o sistema nervoso pode ser organizado em diversos níveis, desde moléculas, a sinapses, neurônios, mapas e até em sistemas. Contudo, uma estrutura facilmente identificável do sistema nervoso é o neurônio.

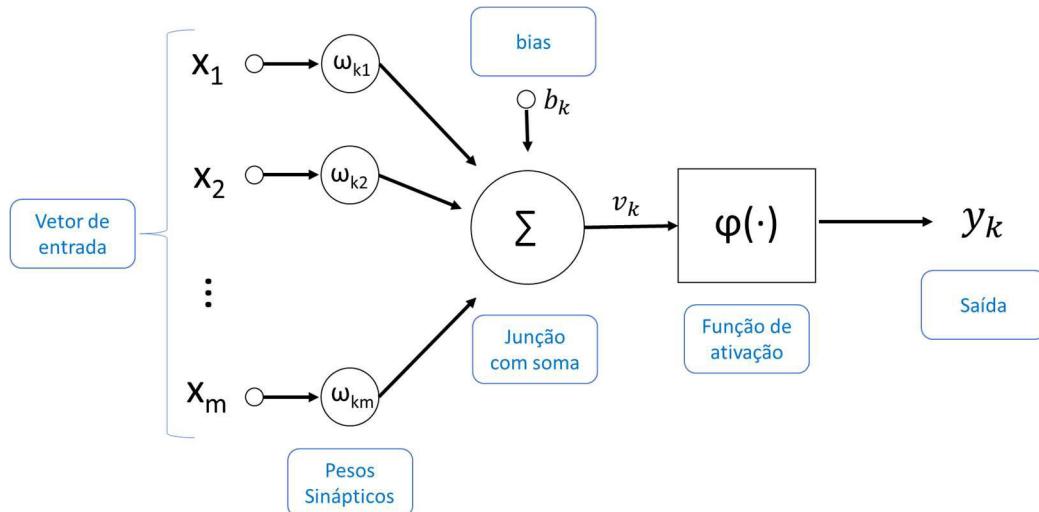
Segundo Amari et al. (2003), a expressão “o neurônio” deve ser usada com cautela, pois, biologicamente falando, existem estruturas e tipos bem distintos de neurônios. Dessa forma, a computação neurológica busca abstrair em diversos aspectos a complexidade real dos neurônios objetivando uma forma de generalizar seu comportamento.

Os neurônios são capazes de gerar, na maior parte das vezes, um sinal na forma de potencial elétrico que é utilizado para transmitir informações a outras células/neurônios. Esse sinal pode ser recebido ou mesmo enviado a diversos outros neurônios. O processo de transferência de informações entre células e a grande conexão entre os neurônios é o fundamento para a capacidade de processamento do cérebro (SQUIRE et al. 2012).

Um dos mais importantes aspectos da neurociência que é aplicado nas ANN é o conceito da efetividade variável da transmissão de sinais, permitindo que o cérebro se adapte a diferentes situações. A chamada plasticidade sináptica é a chave do aprendizado de diversos modelos, podendo ser traduzida como a modificação/atualização dos pesos sinápticos (KRIESEL, 2007).

A Figura 4 representa uma ilustração com os principais componentes de um neurônio matemático teórico.

Figura 4 – Ilustração dos principais componentes de um neurônio artificial (matemático)

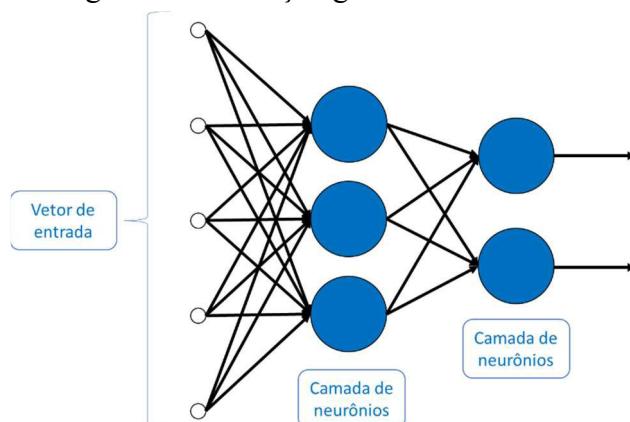


Fonte: Elaborada pelo autor.

De forma geral, um sinal de entrada chamado de vetor de entrada é transmitido ao corpo do neurônio através de pesos sinápticos que controlam o fluxo de informação de entrada. Em seguida ocorre uma soma de todos os sinais que atravessaram os pesos sinápticos e sua aplicação em uma função de ativação, que define a maneira como os dados de entrada serão processados e fornece como saída um vetor resposta do neurônio. Essa resposta é então enviada a um neurônio posterior através de outros pesos sinápticos, similar aos w_k , porém de índices w_{k+1} (HAYKIN, 2009).

Cada neurônio se comunica com um ou mais neurônios diferentes formando uma grande conexão entre essas estruturas. A Figura 5 mostra de forma gráfica uma ANN completamente conectada (à direita).

Figura 5 – Ilustração gráfica de uma ANN



Fonte: Elaborada pelo autor.

Segundo Haykin (2009), uma ANN é um processador distribuído massivamente e paralelamente, constituído por unidades de processamento simples que tem uma propensão natural para armazenar conhecimento e disponibilizá-lo para uso. Assemelha-se ao cérebro em dois aspectos: o conhecimento é adquirido pela rede a partir de seu ambiente por meio de um processo de aprendizado; e as forças de conexão interna, conhecidas como pesos sinápticos, são usadas para armazenar o conhecimento adquirido.

Segundo Kriesel (2007), o estudo das ANN é motivado por sua incrível semelhança com sistemas biológicos funcionais que, de forma geral, consiste em simples, porém grande quantidade de células que trabalham massiva e paralelamente, tendo a capacidade de aprender. O resultado do aprendizado é a capacidade da ANN em associar e generalizar informação que, em parte, leva a uma alta capacidade de tolerar erros provenientes de ruídos dos dados.

Entende-se, dessa forma, que as redes neurais artificiais são modelos de aproximação não-paramétricos, na medida em que seus pesos sinápticos não apresentam um significado físico particular em relação ao problema de aplicação. Além disso, a estimativa dos parâmetros de um modelo não-paramétrico (por exemplo, pesos de uma rede neural artificial) não é o objetivo primário desse tipo de método. O objetivo principal é estimar a “forma” da função em uma região compacta do espaço de aproximação - ou ao menos a saída para certos valores desejados de entrada (VON ZUBEN; ATTUX, 2001).

2.3.3.2 Redes Neurais de Funções de Base Radial (RBFN)

Segundo Von Zuben e Attux (2001), o que diferencia as redes de função de base radial das demais redes pode ser resumido em 3 ideias principais: as RBFN sempre apresentam um única camada intermediária, ou seja, possuem estrutura fixa; seus neurônios de saída são sempre lineares; a função de ativação utilizada pelos neurônios ocultos são apenas funções de base radial, ao invés de sigmoidais ou tangente hiperbólicas como em outros métodos.

2.3.3.2.1 Funções de Base radial (RBF)

Dessa forma, é necessário abordar inicialmente o que são Funções de Base Radial, tradução livre de *Radial Basis Functions* (RBF): são funções com características distintas - suas respostas crescem/decrecem monotonicamente em relação a um ponto central. Valores como o centro, distância até ele, formato da curva e outros, são fixos (KRIESEL, 2007).

Uma função RBF típica é a gaussiana, que no formato escalar tem a seguinte estrutura:

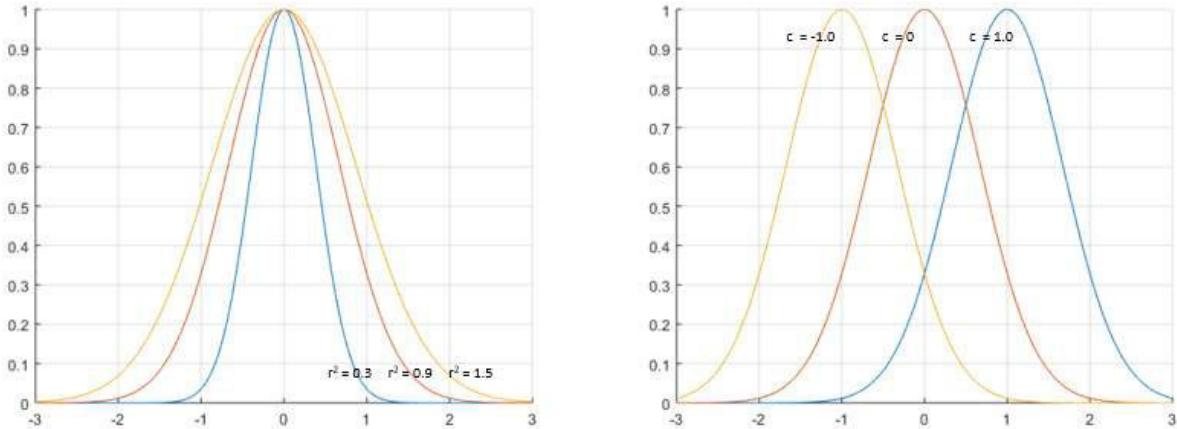
$$f(x) = \exp\left(-\frac{(x - c)^2}{r^2}\right) \quad (2.16)$$

Onde c representa o centro da função e r o raio que ela abrange. Graficamente, r define se a função é mais ou menos espalhada.

Segundo Haykin (2009), existem diversas outras funções como a multiquadrática e a multiquadrática inversa que atendem teoricamente aos requisitos de uma RBFN, contudo, optou-se no trabalho por desenvolver a função gaussiana já que grande parte das referências bibliográficas a utilizam.

Exemplificando-se a função da Equação 2.16 e considerando-se as possíveis variações em relação à c (centro) e em relação à r (raio), a Figura 6 ilustra a função $f(x)$ para diversas configurações:

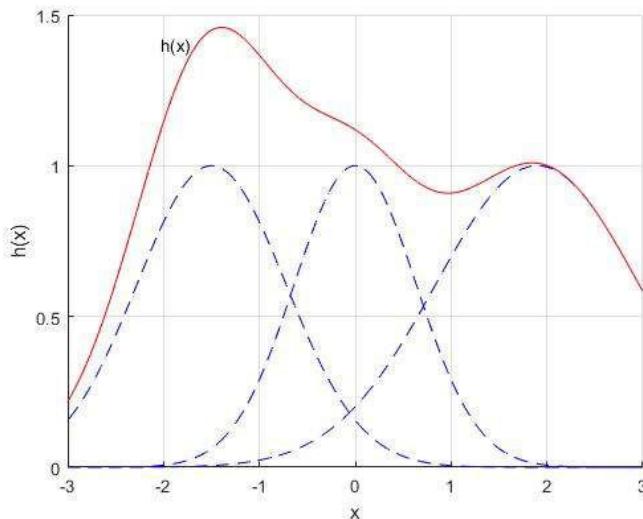
Figura 6 – Ilustração da função $f(x)$ definida pela equação 2.16 com variação de r sendo $c = 0$ (à esquerda) e com variação de c sendo $r^2 = 0.9$ (à direita)



Fonte: Elaborada pelo autor.

Dessa forma, tem-se a soma ponderada de algumas RBF como a base do entendimento e funcionamento de uma RBFN (KRIESEL, 2007). A Figura 7 ilustra o formato de uma função $h(x)$ constituída pela soma ponderada de funções gaussianas aleatórias.

Figura 7 – Ilustração da função $h(x)$ constituída pela soma ponderada de funções gaussianas aleatórias



Fonte: Elaborada pelo autor.

A função $h(x)$ pode representar uma função qualquer, dependendo apenas das configurações e quantidades de funções gaussianas que a forma.

Essa característica é de fundamental importância no problema de aproximação de uma função, conforme Seção 2.3.1, já que, analisando o problema em uma única dimensão, uma função $h(x)$ ajustada seria uma possível solução na previsão de dados de uma função $y(x)$ qualquer.

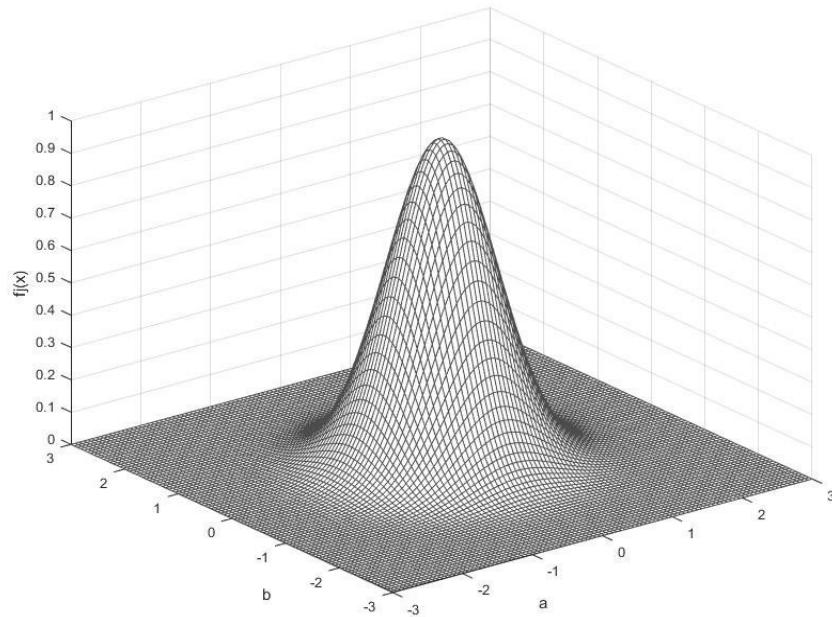
Vale ressaltar o formato multidimensional da função. Assim, a Equação 2.16 (formato escalar) pode ser traduzida para a Equação 2.17.

$$f_j(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}_j\|^2}{\sigma_j^2}\right), \text{ para } j = 1, 2, \dots, n \quad (2.17)$$

Onde $\|\cdot\|^2$ representa a norma euclidiana quadrática - uma métrica de distância entre a entrada de dados \mathbf{x} e o centro $\boldsymbol{\mu}$, σ representa o raio da função e j representa um índice qualquer de identificação da função gaussiana utilizada.

Considerando-se \mathbf{x} como um vetor de entrada de duas dimensões $-3 \leq a, b \leq 3$ e $\sigma = 0.9$, tem-se graficamente:

Figura 8 – Ilustração gráfica da função $f(x)$ definida pela Equação 2.17



Fonte: Elaborada pelo autor.

Assim, seja qual for a dimensionalidade da função trabalhada, é possível encontrar uma função $h(x)$ (soma ponderada de funções gaussianas $f(x)$ de mesma dimensionalidade) que represente essa função.

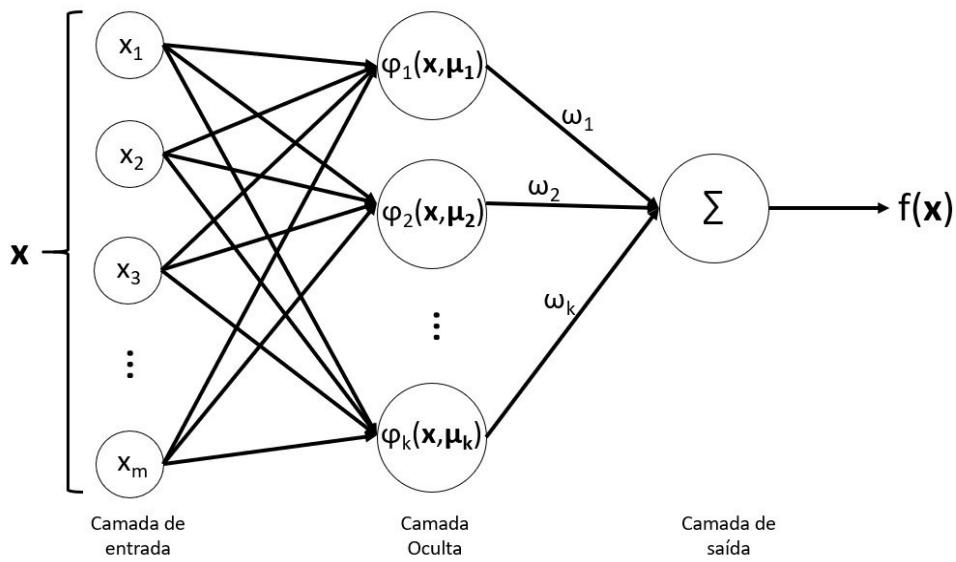
2.3.3.2.2 Estrutura da rede

Na década de 80, Broomhead e Lowe (1988) sugeriram uma nova rede neural denominada Rede Neural de Funções de Base Radial (RBFN), inovando na área de *machine learning* na medida em que as redes mais usuais eram do tipo Percéptrons de Múltiplas Camadas, tradução livre de *Multilayer Perceptrons* (MLP).

Esse trabalho foi o precursor dos estudos em RBFN, propondo ao invés de uma aproximação estocástica - através de uma técnica recursiva com o algoritmo *back-propagation* utilizado nas MLP, uma aproximação híbrida que envolve dois estágios: transformação de um espaço não-linear de dados iniciais em um espaço linearmente separável, de dimensão maior que o inicial, dadas certas condições, baseando matematicamente a justificativa para essa transformação no Teorema de Cover (COVER, 1965), seguido de um estágio final baseado no método dos mínimos quadrados (HAYKIN, 2009).

A estrutura básica da RBFN é rígida, sempre composta por 3 camadas: a primeira é a camada de entrada, onde estão os nós (unidades sensores) da rede e que ligam o espaço multidimensional das variáveis analisadas com o ambiente da rede; a segunda é a camada oculta, onde ocorre a transformação do espaço de entrada no espaço oculto de alta dimensionalidade; a terceira e última é a camada de saída, que fornece uma resposta ao sinal de entrada (HAYKIN, 2009). A Figura 9 a seguir apresenta a estrutura básica de uma RBFN:

Figura 9 – Componentes básicos de uma RBFN



Fonte: Elaborada pelo autor.

Um vetor de entrada \mathbf{x} de dimensionalidade m é inserido na primeira camada. A partir dele, cada neurônio que compõe um total de k neurônios, onde $k \leq m$, recebe esse mesmo sinal de entrada e aplica em uma função de ativação φ . Cada função de ativação tem um centro μ específico e, dessa forma, na camada oculta existem k neurônios com k funções de ativação relacionadas, cada uma com um centro μ específico. A camada de saída realiza uma soma ponderada, onde ω representa os k pesos sinápticos distintos, cada um conectando um neurônio oculto específico com a camada de saída. A camada de saída $f(\mathbf{x})$ fornece a resposta da rede dado o estímulo inicial \mathbf{x} .

A função de ativação mais utilizada na RBFN é a gaussiana, representada na Equação 2.18, similar à mostrada na Equação 2.17. Vale ressaltar que em problemas que envolvem aproximação de função o parâmetro σ , que representa o raio, pode ser mantido constante para as k funções de ativação distintas (HAYKIN, 2009; SCHWENKER; KESTLER; PALM, 2001).

$$\varphi(\mathbf{x}) = \exp\left(-\frac{\|\mathbf{x} - \boldsymbol{\mu}\|^2}{\sigma^2}\right) \quad (2.18)$$

Segundo Feng e Chou (2011), quando a rede RBFN é utilizada para fins de regressão (aproximação de função) e não para fins de classificação de dados, é comum que a saída de cada neurônio da camada oculta seja normalizada. Dessa forma, a equação de saída de um neurônio i qualquer pode ser escrita como:

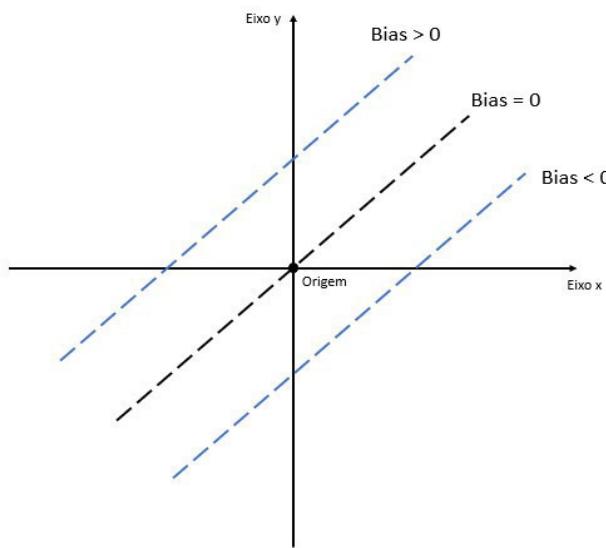
$$\Phi_i(\mathbf{x}) = \left(\frac{\varphi_i(\mathbf{x}, \boldsymbol{\mu}_i)}{\sum_{j=1}^k \varphi_j(\mathbf{x}, \boldsymbol{\mu}_j)} \right) \quad (2.19)$$

A função de saída da rede pode ser representada por:

$$f(\mathbf{x}) = \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}, \boldsymbol{\mu}) \quad (2.20)$$

É comum, ainda, a alocação do chamado “bias”, uma constante que, de forma resumida, translada as regiões de contorno de uma função a partir da origem. A Figura 10 representa a transladação feita pelo uso do bias em um problema onde a função aproximada atravessaria necessariamente a origem.

Figura 10 – Representação da influência do bias



Fonte: Elaborada pelo autor.

Assim, a Equação 2.20 pode ser reescrita como (para ω_0 representando o bias):

$$f(\mathbf{x}) = \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}, \boldsymbol{\mu}) + \omega_0 \quad (2.21)$$

2.3.3.2.3 Método de solução

No tópico anterior foi mostrado como o sinal de entrada representado pelo vetor de estímulo \mathbf{x} percorre o interior da rede RBFN e como a resposta representada por $f(\mathbf{x})$ é calculada. Agora, as equações a seguir demonstram qual o raciocínio usado para que os pesos sinápticos sejam ajustados ($\boldsymbol{\omega} \rightarrow \boldsymbol{\omega}^*$) e, dessa forma, como um novo vetor de entrada/vetor de teste \mathbf{x}^* pode ser apresentado as mesmas funções de ativação anteriores juntamente com os pesos ajustados para que um estímulo $f^*(\mathbf{x}^*)$ seja calculado com a menor diferença possível da resposta esperada y^* (KAVAKLIOGLU; KOSEOGLU; CALISKAN, 2018).

Primeiramente, considera-se o conjunto de vetores de entrada \mathbf{X} e o conjunto de respostas esperadas \mathbf{Y} com as seguintes propriedades:

$$\mathbf{X} = \begin{bmatrix} \mathbf{x}^{(1)} \\ \mathbf{x}^{(2)} \\ \vdots \\ \mathbf{x}^{(n)} \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix} \text{ e } \mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

Onde n representa o número de dados de entrada e m o número de variáveis independentes que compõem a matriz \mathbf{X} . Assim, cada vetor de entrada $\mathbf{x}^{(i)}$, composto por m variáveis de entrada, tem como resposta esperada y_i .

A partir disso, três parâmetros são fixados:

- Número de neurônios na única camada oculta existente: k
- Centros de cada neurônio: \mathbf{M}
- Raio das funções gaussianas: σ

$$\text{Onde: } \mathbf{M} = \begin{bmatrix} \boldsymbol{\mu}^{(1)} \\ \boldsymbol{\mu}^{(2)} \\ \vdots \\ \boldsymbol{\mu}^{(k)} \end{bmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1m} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{k1} & \mu_{k2} & \cdots & \mu_{km} \end{bmatrix}$$

A partir desses parâmetros da RBFN e do modo de organização da entrada \mathbf{X} e saída esperada \mathbf{Y} de dados, pode-se escrever o valor esperado y_i para cada vetor de entrada $\mathbf{x}^{(i)}$, sendo $i = \{1, 2, \dots, n\}$, como uma soma de todas as saídas dos k neurônios da camada oculta, conforme Equação 2.21, adicionado de um erro associado ϵ_i . A Equação 2.22 a seguir sumariza o que foi dito acima:

$$y_i = \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}^{(i)}, \boldsymbol{\mu}_j) + \omega_0 + \epsilon_i \quad (2.22)$$

Dessa forma, detalhando as respostas calculadas para os n dados selecionados:

$$\begin{aligned} y_1 &= \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}^{(1)}, \boldsymbol{\mu}_j) + \omega_0 + \epsilon_1 \\ y_2 &= \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}^{(2)}, \boldsymbol{\mu}_j) + \omega_0 + \epsilon_2 \\ &\vdots \\ y_n &= \sum_{j=1}^k \omega_j \Phi_j(\mathbf{x}^{(n)}, \boldsymbol{\mu}_j) + \omega_0 + \epsilon_n \end{aligned} \quad (2.23)$$

$$\begin{aligned} y_1 &= \omega_0 + \omega_1 \Phi_1^{(1)} + \omega_2 \Phi_2^{(1)} + \dots + \omega_k \Phi_k^{(1)} + \epsilon_1 \\ y_2 &= \omega_0 + \omega_1 \Phi_1^{(2)} + \omega_2 \Phi_2^{(2)} + \dots + \omega_k \Phi_k^{(2)} + \epsilon_2 \\ &\vdots \\ y_n &= \omega_0 + \omega_1 \Phi_1^{(n)} + \omega_2 \Phi_2^{(n)} + \dots + \omega_k \Phi_k^{(n)} + \epsilon_n \end{aligned} \quad (2.24)$$

Na forma matricial, pode ser visto como:

$$\mathbf{Y} = \boldsymbol{\Phi} \boldsymbol{\omega} + \boldsymbol{\epsilon} \quad (2.25)$$

Onde:

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \boldsymbol{\Phi} = \begin{bmatrix} 1 & \Phi_1^{(1)} & \Phi_2^{(1)} & \dots & \Phi_k^{(1)} \\ 1 & \Phi_1^{(2)} & \Phi_2^{(2)} & \dots & \Phi_k^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \Phi_1^{(n)} & \Phi_2^{(n)} & \dots & \Phi_k^{(n)} \end{bmatrix}, \boldsymbol{\omega} = \begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix}, \boldsymbol{\epsilon} = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix}$$

Assim como mencionado na Seção 2.3.2 sobre MLR, pode-se usar o método dos mínimos quadrados para prever ω através da pseudo inversão, análogo ao feito na Equação 2.13 para encontrar os coeficientes de regressão β :

$$\omega = (\Phi' \Phi)^{-1} \Phi' Y \quad (2.26)$$

Dessa forma, os valores calculados pela RBFN podem ser tomados como:

$$Y^{calulado} = \Phi \omega \quad (2.27)$$

Ou mais detalhadamente:

$$y_i^{calulado} = [1 \quad \Phi_1^{(i)} \quad \Phi_2^{(i)} \quad \dots \quad \Phi_k^{(i)}] \begin{bmatrix} \omega_0 \\ \omega_1 \\ \omega_2 \\ \vdots \\ \omega_k \end{bmatrix} \quad (2.28)$$

Onde:

$$\Phi_j^{(i)} = \left(\frac{\varphi_j(\mathbf{x}^{(i)}, \boldsymbol{\mu}_j)}{\sum_{z=1}^k \varphi_z(\mathbf{x}^{(i)}, \boldsymbol{\mu}_z)} \right) = \left(\frac{\exp\left(-\frac{\|\mathbf{x}^{(i)} - \boldsymbol{\mu}_j\|^2}{\sigma^2}\right)}{\sum_{z=1}^k \exp\left(-\frac{\|\mathbf{x}^{(i)} - \boldsymbol{\mu}_z\|^2}{\sigma^2}\right)} \right) \quad (2.29)$$

Sendo $\mathbf{x}^{(i)} = [x_{i1} \quad x_{i2} \quad \dots \quad x_{im}]$ um vetor de entrada qualquer com m variáveis independentes, para $i = \{1, 2, \dots, l\}$, onde l representa o número de dados de entrada que se deseja inserir para o teste da rede.

Resumidamente, a partir de dados de entrada quaisquer e fixados alguns parâmetros da rede, pode-se encontrar um Φ , calculado conforme Equação 2.29, que poderá ser utilizado para prever a saída de quaisquer outros dados de entrada ainda não fornecidos à rede (Considerando que esses dados fazem parte de um conjunto e que sejam, portanto, representativos do problema em estudo).

2.3.3.3 Otimização de hiperparâmetros

Pode-se dizer que os algoritmos de aprendizado são compostos por parâmetros e hiperparâmetros. Os primeiros são obtidos através do treinamento do algoritmo, tendo uma solução analítica (fórmula matemática) para sua correta definição, por exemplo, os pesos sinápticos de uma rede neural MLP através do algoritmo de *backpropagation*; já os últimos devem ser definidos antes mesmo do início do treinamento e estão diretamente ligados ao desempenho do algoritmo e ao seu custo computacional (ALVARENGA JÚNIOR, 2018; BERGSTRA; BENGIO, 2012).

Como o método de solução para os parâmetros já é previamente definido pelo próprio modelo de algoritmo de aprendizado, resta apenas o problema de encontrar de forma eficiente os hiperparâmetros ideais.

A forma mais simples é variar cada hiperparâmetro manualmente em busca de uma resposta satisfatória do algoritmo. Contudo, ao longo dos anos, métodos automáticos de otimização foram surgindo, entre eles: *Grid Search* (LERMAN, 1980), *Random Search* (SOLIS; WETS, 1981.), *Gradient-based Optimization* (BENGIO, 2000), *Evolutionary Computation* (FRIEDRICH; IGEL, 2005) e outros (LUO, 2016). Dois desses métodos serão apresentados à seguir.

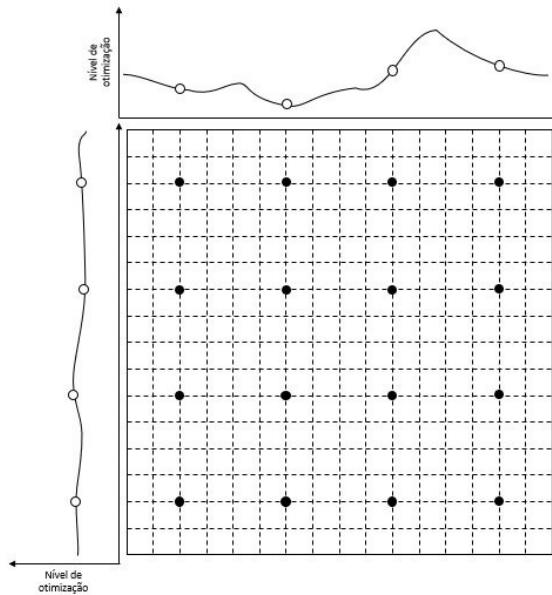
2.3.3.3.1 Grid Search

O *Grid Search* é usualmente utilizado na otimização envolvendo um número reduzidos de hiperparâmetros, pois se trata de um método bem robusto computacionalmente. Ele, de forma resumida, propõe que cada hiperparâmetro seja delimitado em um intervalo de busca. Em seguida, deve-se escolher o número de divisões a que esse intervalo será submetido, ou seja, quantos pontos dentro do intervalo serão validados pela busca. A execução desse método cria no espaço hiperparamétrico uma estrutura em grade formada pelo número de variações de cada hiperparâmetro. Um problema visível é o crescimento exponencial do custo computacional conforme crescem o número de hiperparâmetros e o número de divisões do intervalo. Apesar disso, é um método bastante utilizado e pode trazer bons resultados caso o número de pontos analisados seja suficiente para criar uma busca adequada (ALVARENGA JÚNIOR, 2018).

A Figura 11 representa a grade formada pelo método de otimização e uma possível influência de nível de otimização para cada valor de cada hiperparâmetro. Nota-se que se um

hiperparâmetro possuir picos de nível de otimização em determinado intervalo da busca, pode ser que o número de divisões da grade não permita alcançar esse valor, dessa forma, prejudicando a resposta final do algoritmo de aprendizado.

Figura 11 – Representação da grade criada pela variação de valores de hiperparâmetros no método *Grid Search*



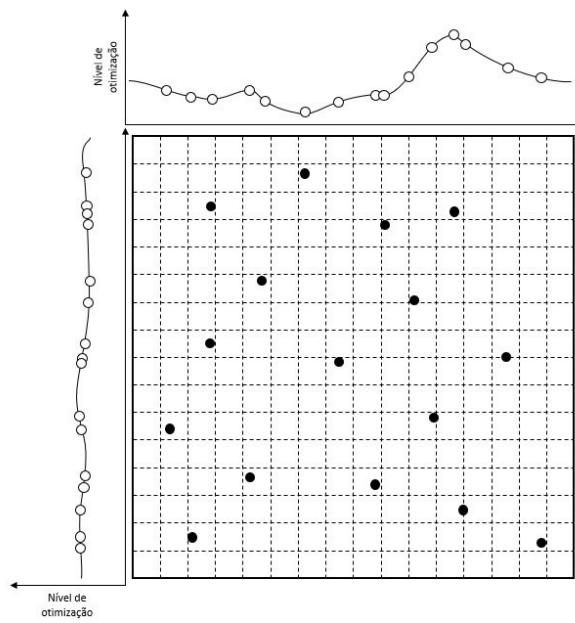
Fonte: Elaborada pelo autor.

2.3.3.3.2 Random Search

O *Random Search*, na forma mais simples, é um método que não utiliza o histórico de pontos já analisados como referencial para o próximo ponto testado, ou seja, os pontos escolhidos de um certo intervalo pré-determinado são sempre aleatórios. Dessa forma, assim como o *Grid Search*, faz-se necessário estabelecer a resolução do intervalo para definir a quantidade de possíveis pontos que poderão ser analisados. A partir disso, a escolha torna-se aleatória até que a busca termine. (BERGSTRA; BENGIO, 2012).

De forma análoga a realizada na Figura 11, a Figura 12 demonstra como a grade do método de *Radom Search* é formada e uma possível influência do nível de otimização para cada valor de cada hiperparâmetro. É perceptível, comparando-se as duas figuras, que para uma mesma quantidade de pontos, o segundo método fornece mais possibilidades de alcançar valores ideais de otimização do algoritmo e faz uma busca mais ampliada dentro dos valores possíveis. Assim, entende-se que esse método é um pouco mais eficiente e tem a capacidade de fornecer melhores resultado

Figura 12 – Representação da grade criada pela variação de valores de hiperparâmetros no método *Random Search*



Fonte: Elaborada pelo autor.

2.4 Estatísticas de validação

De forma resumida, os algoritmos utilizados no trabalho separaram os dados em um conjunto de treino - que é utilizados na solução do método, seja para encontrar coeficientes de regressão, no caso da regressão multilinear, seja para encontrar pesos sinápticos, como no caso da rede neural artificial; e um conjunto de teste – que somente é apresentado para o algoritmo após o treinamento e não servem para mudar os parâmetros já citados, mas sim para validar/testar o quanto próximo o método consegue chegar do resultado real, ou seja, qual a diferença entre uma resposta calculada pelo algoritmo e a resposta desejada, sem que as variáveis desse conjunto influenciem no aprendizado.

Existem diversas formas de quantificar a precisão da resposta e se o modelo matemático proposto é realmente válido. No presente trabalho o coeficiente de determinação (R^2) foi escolhido como estatística principal de verificação.

2.4.1 Coeficiente de determinação

Segundo Hoffman (2016), o coeficiente de determinação pode ser definido como:

$$R^2 = \frac{SQ_{exp}}{SQ_{total}} \quad (2.30)$$

Sendo:

$$SQ_{total} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.31)$$

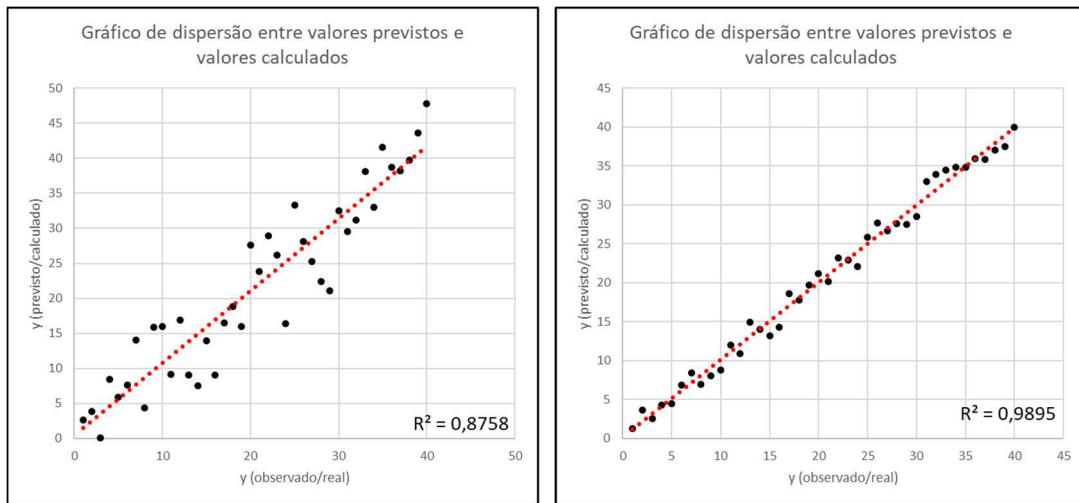
$$SQ_{exp} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (2.32)$$

onde R^2 representa o coeficiente de determinação, SQ_{total} representa a soma dos quadrados total, SQ_{exp} representa a soma dos quadrados explicada, n é o número de observações, y_i é o valor observado (real), \bar{y} é a média dessas observações e \hat{y}_i é o valor estimado (calculado/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 em relação a média dos dados reais. Juntando essas duas informações na Equação 2.30, tem-se que o coeficiente de regressão R^2 indica uma espécie de 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 (MONTGOMERY; PECK; VINING, 2012).

Em outras palavras, o R^2 informa o quão próximo estarão os pontos de um gráfico composto por dois eixos: um dos valores reais, outro dos valores calculados, de uma reta de regressão, como mostrado na Figura 13.

Figura 13 – Representações gráficas do significado de R^2



Fonte: Elaborada pelo autor.

Observa-se que quanto maior o R^2 , mais próximos os pontos estarão da reta, significando uma maior relação entre valores calculados e valores reais, o que representa um melhor ajuste do modelo matemático utilizado para previsão.

3 METODOLOGIA

Este capítulo visa abordar os métodos utilizados para realização dos testes de previsão dos algoritmos selecionados.

3.1 Dados

Os dados para realização dos experimentos foram disponibilizados por uma indústria siderúrgica localizada no Ceará, referindo-se à produção de vergalhões de aço no final de 2017. Ao todo, foram obtidas 1300 instâncias, ou seja, 1300 pares de entrada e saída de dados. Foram analisadas entradas relacionadas à composição química final do aço após passagem pela acaria elétrica, alguns parâmetros do processo de laminação – incluindo desde temperaturas em diversas etapas, redução da seção transversal do tarugo até ritmo de produção, totalizando 24 parâmetros de entrada; e buscou-se uma relação com propriedades mecânicas específicas do produto final como limite de escoamento, limite de resistência à tração e alongamento, totalizando 4 propriedades de saída. A Tabela 1 resume os dados utilizados:

Tabela 1 – Resumo dos dados

Composição Química				
Variável	Média	Desvio Padrão	Mínimo	Máximo
<i>Carbono (%)</i>	0,3216	0,0393	0,1560	0,4740
<i>Carbono Equivalente (%)</i>	0,5258	0,0539	0,2850	0,6690
<i>Cromo (%)</i>	0,1314	0,0356	0,0560	0,2710
<i>Cobre (%)</i>	0,2814	0,0403	0,1870	0,4450
<i>Manganês (%)</i>	0,9009	0,2289	0,4560	1,1490
<i>Molibdénio (%)</i>	0,0111	0,0024	0,0060	0,0290
<i>Níobio (%)</i>	0,0026	0,0012	0,0000	0,0080
<i>Níquel (%)</i>	0,0745	0,0115	0,0490	0,1430
<i>Fósforo (%)</i>	0,0064	0,0017	0,0020	0,0130
<i>Enxofre (%)</i>	0,0370	0,0056	0,0200	0,0500
<i>Silício (%)</i>	0,1566	0,0212	0,0990	0,2080
<i>Estanho (%)</i>	0,0193	0,0034	0,0110	0,0400
<i>Vanádio (%)</i>	0,0080	0,0059	0,0020	0,0190
<i>Arsênio (%)</i>	0,0040	0,0004	0,0030	0,0070
<i>Zinco (%)</i>	0,0040	0,0033	0,0000	0,0330
<i>Relação Manganês/Enxofre</i>	24,9860	7,6328	9,4990	49,3380
Processo (Laminação)				
Variável	Média	Desvio Padrão	Mínimo	Máximo
<i>Área da seção transversal (mm²)</i>	93,2485	46,4917	50,0000	199,0000
<i>Temperatura de saída do forno de reaquecimento (°C)</i>	1068,2414	17,2822	1013,3922	1125,3582
<i>Temperatura na entrada na zona de desbaste (°C)</i>	1031,9664	21,9701	742,6667	1081,4603
<i>Temperatura de entrada do acabador (°C)</i>	920,6544	21,1274	654,6667	961,8219
<i>Ritmo de produção (peças por hora)</i>	75,8365	13,0677	40,0000	100,8375
<i>Temperatura de equalização para tratamento térmico (°C)</i>	436,7833	47,5267	337,5806	584,1379
<i>Pressão da água para tratamento térmico (Bar)</i>	5,8464	4,9932	0,0000	15,5000
<i>Vazão da água para tratamento térmico (m³/h)</i>	29,2119	23,6532	0,0000	68,2222
Propriedades mecânicas				
Variável	Média	Desvio Padrão	Mínimo	Máximo
<i>Limite de escoamento (Mpa)</i>	521,7088	41,5603	421,0000	661,0000
<i>Limite de resistência à tração (Mpa)</i>	682,7937	35,0719	596,0000	913,6667
<i>Razão entre Limite de resistência e limite de escoamento</i>	1,3136	0,0723	1,1200	1,5500
<i>Alongamento (%)</i>	15,0505	1,8974	8,0000	20,0440

Fonte: Elaborada pelo autor.

Algumas observações devem ser feitas para melhor compreensão dos dados fornecidos:

- “Carbono equivalente” considera uma equação influenciada prioritariamente pelo teor de carbono, mas que leva em consideração também os elementos como silício, manganês, vanádio, cromo, níquel, e etc, conforme Kostin (2012);
- “Relação Manganês/Enxofre” considera a razão % de manganês sobre % de enxofre no aço líquido, conforme Chiaverini (1977);
- As temperaturas citadas foram medidas ao longo do processo de laminação, envolvendo as etapas já citadas na Seção 2.1.1;
- “Razão entre limite de resistência e limite de escoamento” considera a razão entre o limite de resistência à tração (MPa) e o limite de escoamento (MPa).

3.2 Ajuste da MLR

Para o algoritmo de regressão multilinear foram consideradas todas as 1300 instâncias obtidas, realizando-se 4 baterias de testes separadas, cada uma considerando uma propriedade mecânica específica, visto a impossibilidade do método na previsão de 4 saídas simultâneas.

Cada bateria de teste consiste em executar o algoritmo para 100 conjuntos de treino e teste distintos. Tais conjuntos selecionados pela separação aleatória de 80% dos dados de treino (1040) e 20% dos dados de teste (360). Na etapa de treino os coeficientes de regressão são estimados, ou seja, são utilizados os 80% dos dados e seus 24 parâmetros relacionados com 1 propriedade mecânica específica para aplicação no método de resolução proposto na Seção 2.3.2. A partir disso, 20% dos dados que não foram apresentados ao algoritmo são aplicados na Equação 2.15 para o cálculo de um saída. Essa resposta é então comparada com a resposta esperada desses dados através do coeficiente de determinação, obtendo-se através das 100 rodadas valores médios, máximos e desvios padrões dessa estatística.

Com o uso desse procedimento, pretende-se simular um ambiente real de atuação do algoritmo, onde dados “conhecidos” (dados de treino) são aplicados ao método de solução e então dados “desconhecidos” (dados de teste) são utilizados para se obter uma resposta

calculada sem que seja conhecida previamente a resposta esperada, simulando, portanto, uma previsão de resposta de propriedade mecânicas tendo como base apenas parâmetros de produção

3.3 Ajuste da RBFN

Inicialmente, os dados aplicados na RBFN são normalizados. A normalização é prática habitual dos processos envolvendo redes neurais e consiste na transformação do intervalo original dos dados em um novo intervalo fixo para todo o conjunto, facilitando o aprendizado da rede (SILVA; SCHIMIDT, 2016; MARTINELI, 1999; HAYKIN, 2009). Os intervalos mais comuns são de -1 a 1 e de 0 a 1. A Equação 3.1 detalha o método de transformação para cada variável dos dados.

$$y^{novo} = \left(\frac{(x^{orig} - x^{orig\ min})(y^{novo\ máx} - y^{novo\ mín})}{(x^{orig\ máx} - x^{orig\ mín})} \right) + y^{novo\ mín} \quad (3.1)$$

Para o presente trabalho, escolheu-se o intervalo entre -1 e 1. Exemplificando, um valor de limite de escoamento de 600 MPa é transformado, considerando o intervalo máximo original 661 MPa e mínimo 421 MPa, no valor de 0,4917.

Similar ao realizado na regressão multilinear, foram realizadas 4 baterias de testes, cada uma para uma propriedade mecânica específica, utilizando-se as 1300 instâncias e seus 24 parâmetros, separando-se 80% dos dados aleatoriamente para treino e 20% para teste. Contudo, para cada bateria de testes foram consideradas diferentes configurações de rede, ocorrendo para cada configuração 100 rodadas distintas para obtenção dos valores médios, máximos e desvio padrão dos coeficientes de determinação.

Esse procedimento de separação das configurações da rede foi realizado para possibilitar o ajuste adequado dos hiperparâmetros através dos métodos de *Grid Search* e *Random Search*.

Os hiperparâmetros considerados foram:

- Número de neurônios da camada oculta (k);
- Raio das funções gaussianas (σ).

É importante ressaltar que o número de neurônios da camada oculta define o número de centros μ existentes, um em cada neurônio, conforme mostrado na Figura 10.

O método utilizado para obtenção desses centros foi o *k-means*, explicado brevemente por Munnoli e Bapat (2013). Nele, são definidos um número k de centroides, e

então é aplicado um método de minimização de função para que a coordenada de cada centroide ao final de um número específico de interações seja tal que a distância entre ele e os pontos mais próximos a ele seja mínima. De maneira prática, basta apenas definir um $k \leq m$ e aplicar o método para que sejam obtidas as coordenadas ideais dos centroides, originando, dessa forma, o que foi chamado no presente trabalho de μ .

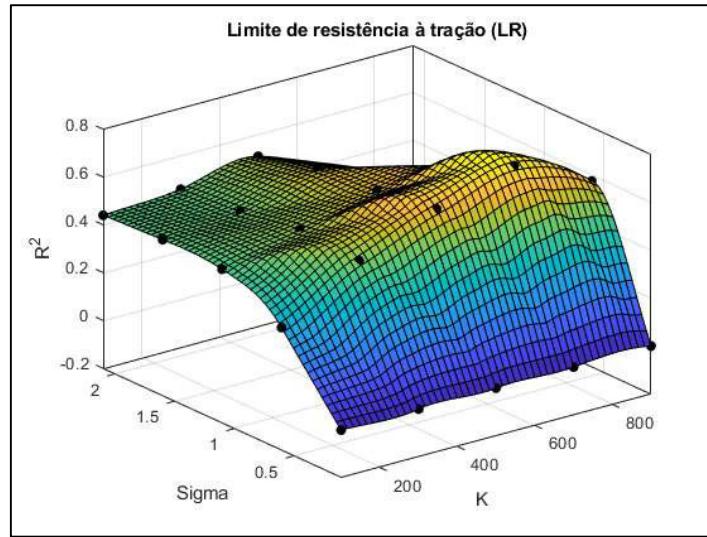
Já o valor σ foi considerado uma constante para todos os neurônios. Existem métodos analíticos de obtenção desse valor, conforme Haykin (2009). Contudo, no presente trabalho essa variável foi considerada um hiperparâmetro pois em teste iniciais foi verificado que os métodos analíticos não estavam fornecendo bons resultados. Dessa forma, σ foi obtido apenas através de sua variação nos métodos *Grid Search* e *Random Search*.

3.4 Obtenção dos hiperparâmetros

O número de neurônios da camada oculta e o raio das funções gaussianas foram obtidos utilizando uma combinação dos dois métodos de otimização de hiperparâmetros citados. Uma primeira análise foi executada com o método *Grid Search* para encontrar uma região hiperparamétrica ótima de atuação do algoritmo, em seguida, o método de *Random Search* foi utilizado para uma busca mais refinada.

A Figura 14 a seguir mostra um exemplo da resposta da primeira análise para a propriedade mecânica LR. Os eixos x e y representam os hiperparâmetros utilizados, já o eixo z representa o valor médio do coeficiente de determinação calculado após as 100 rodadas de testes realizados para uma configuração de rede específica. Dessa forma, observa-se que é possível definir uma região do espaço hiperparamétrico em que a atuação do algoritmo de previsão pode ser considerada otimizada, o que auxilia em um aspecto muito importante dos experimentos realizados: o custo computacional.

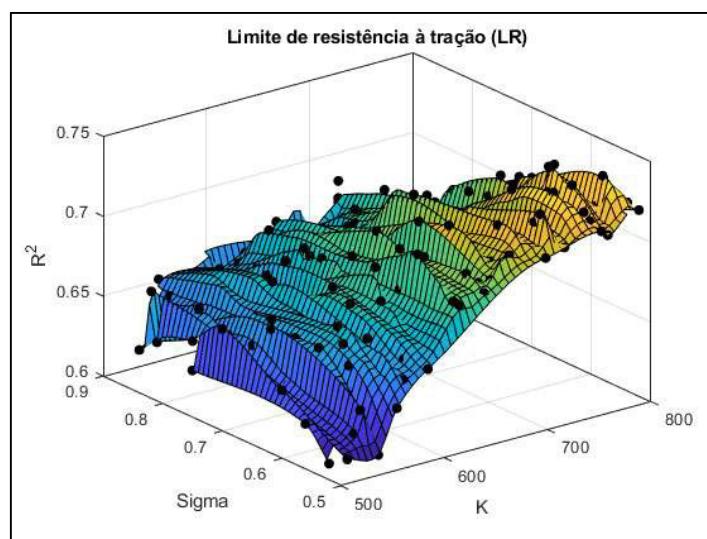
Figura 14 – Exemplo da aplicação do método *Grid Search*



Fonte: Elaborada pelo autor.

Torna-se mais vantajoso computacionalmente fazer uma varredura rápida e simples do espaço hiperparamétrico, definir uma região de interesse, e só então realizar esforços maiores para obtenção do melhor resultado. Como já se espera encontrar bons resultados nessa região, a utilização de um algoritmo como o *Random Search* é otimizada, pois faz uma busca mais robusta e com maiores chances de se encontrar as melhores respostas do algoritmo. A Figura 15 exemplifica a segunda etapa de análise da propriedade mecânica LR.

Figura 15 – Exemplo da aplicação do método *Random Search*



Fonte: Elaborada pelo autor.

4 RESULTADOS E DISCUSSÕES

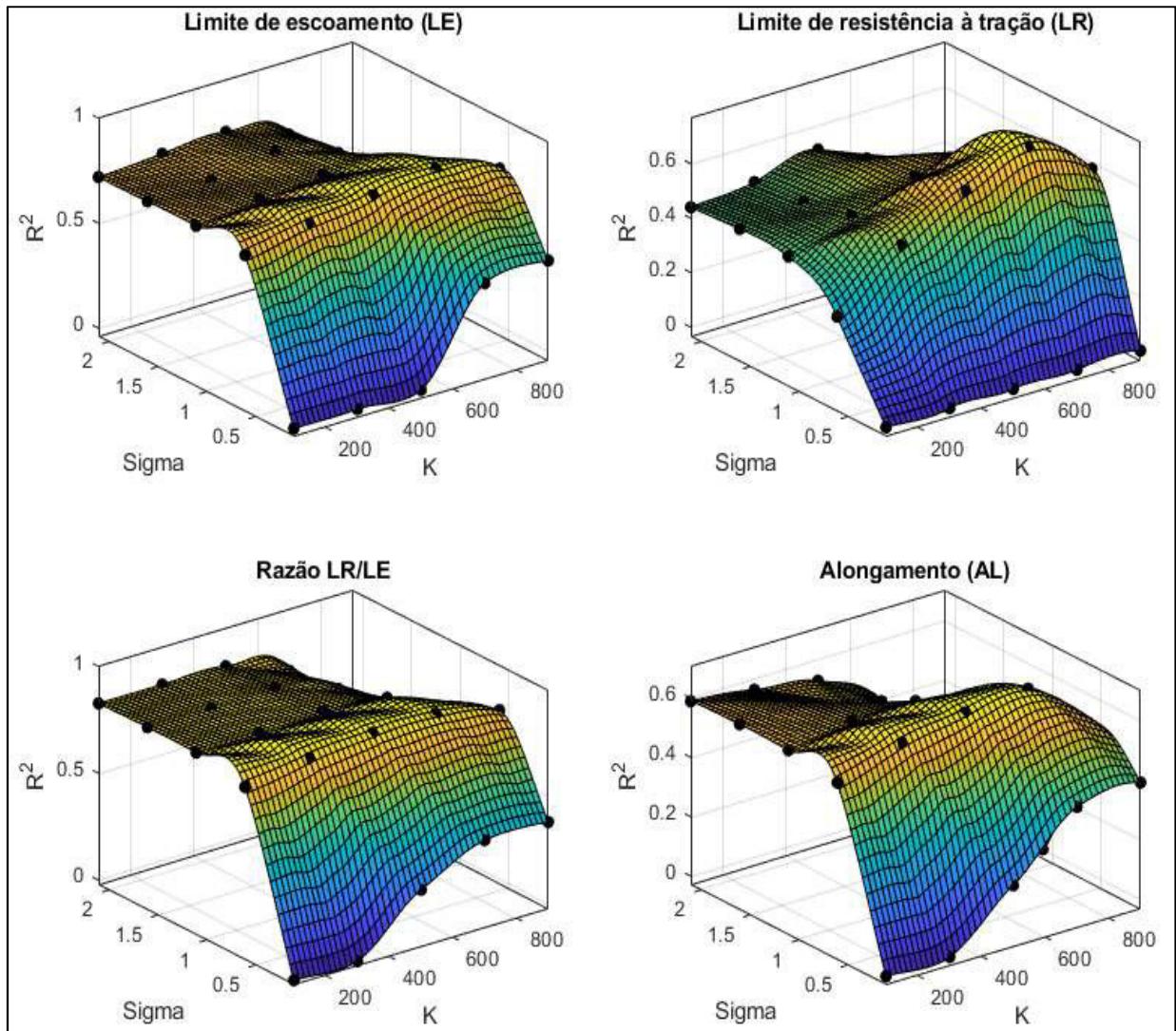
Os resultados a seguir foram obtidos utilizando-se os métodos explicados no capítulo anterior. A Tabela 2 mostra os resultados obtidos com a MLR: a propriedade mecânica com melhor previsibilidade foi a razão entre limite de escoamento e limite de resistência à tração, com um R^2 médio de 0,7633, seguido do limite de escoamento com R^2 médio de 0,5146, do alongamento com 0,5099, e do limite de resistência à tração com 0,2388. Os desvios padrões de cada propriedade não foram elevados, tendo um máximo de 0,0658 para o LR, o que demonstra a confiabilidade do método utilizado.

Tabela 2 – Resultados de previsão de propriedades mecânicas usando MLR.

Propriedades mecânicas	MLR		
	R^2 médio	R^2 máx	Desv. Padrão
<i>Limite de escoamento (Mpa)</i>	0,5146	0,6165	0,0563
<i>Limite de resistência à tração (Mpa)</i>	0,2388	0,3504	0,0658
<i>Razão entre Limite de resistência e limite de escoamento</i>	0,7633	0,8132	0,0250
<i>Alongamento (%)</i>	0,5099	0,6033	0,0400

Em relação à RBFN, a primeira etapa, como já mencionado, utilizou-se do método *Grid Search*. A busca do número de neurônios k foi realizada dentro de um intervalo de 100 a 900, com um passe de 200. A busca por σ foi realizada em um intervalo de 0,1 a 2,1, com passe de 0,5. A Figuras 16 mostra os valores obtidos para as 25 configurações de teste de cada uma das 4 propriedades mecânicas.

Figura 16 – Resultados das baterias de teste usando *Grid Search*



Fonte: Elaborada pelo autor.

Através dessas imagens é perceptível que existem regiões específicas onde os hiperparâmetros maximizam os resultados de R^2 médio. Para cada propriedade mecânica buscou-se uma região determinada para ser testada na próxima etapa pelo método do *Random Search*.

A Tabela 3 mostra as regiões de interesse encontradas pelo método *Grid Search*.

Tabela 3 – Regiões hiperparamétricas otimizadas

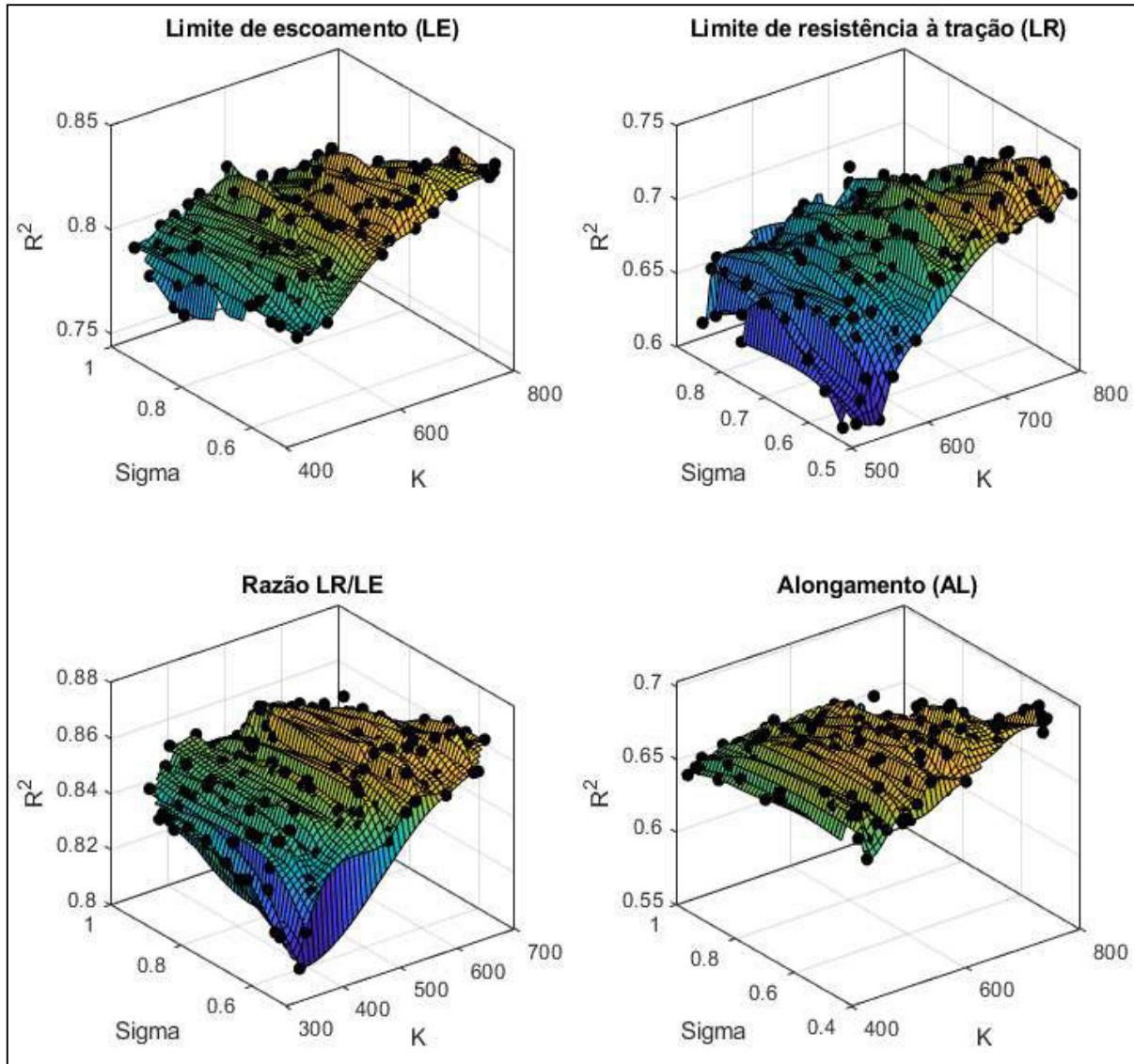
Propriedades Mecânicas	Hiperparâmetros			
	K	SIGMA (σ)		
	mín	máx	mín	máx
LE (Mpa)	400	800	0,50	1,00
LR (Mpa)	500	800	0,50	0,90
LR/LE	300	700	0,50	1,00
AL (%)	400	800	0,40	1,00

Fonte: Elaborada pelo autor.

Cada intervalo na busca de hiperparâmetro considerado foi dividido em partes para a criação de uma malha, o passe corresponde à 1/50 do intervalo total considerado. Por exemplo, para um intervalo de número de neurônios entre 400 e 800, o passe corresponde à 8, e para um sigma com variação entre 0,5 e 1,0, o passe corresponde à 0,01, criando dessa forma uma malha de busca que será analisada pelo método proposto. Ao todo foram considerados 150 configurações aleatórias que fazem parte da malha criada (número superior às 25 configurações realizadas na etapa anterior).

A Figuras 17 mostra os resultados obtidos na segunda etapa pelo método *Random Search* realizando-se a divisão proposta das regiões de interesses mostradas na Tabela 3.

Figura 17 – Resultados das baterias de teste usando *Random Search*



Fonte: Elaborada pelo autor.

É perceptível que a busca mais detalhada no espaço hiperparamétrico traz como resultado superfícies mais acentuadas, diferentemente do encontrado na Figura 16.

As Tabelas 4, 5, 6 e 7 mostram as 5 melhores configurações de rede para cada propriedade mecânica estudada. A propriedade melhor prevista, assim como na regressão multilinear, foi a razão entre limite de resistência à tração e limite de escoamento, com R^2 médio de 0,8691, seguido do limite de escoamento com 0,8437, limite de resistência à tração com 0,7403 e, alongamento com 0,7015.

Tabela 4 – Melhores resultados de previsão do LE classificados de acordo com R^2 médio

Limite de escoamento – LE				
Hiperparâmetros		Estatísticas		
K	SIGMA (σ)	R^2 médio	R^2 máx	Desvio padrão
680	0,58	0,8437	0,9080	0,0306
736	0,57	0,8429	0,9127	0,0328
784	0,53	0,8415	0,8917	0,0286
656	0,55	0,8395	0,8888	0,0285
768	0,54	0,8391	0,9027	0,0323

Fonte: Elaborada pelo autor.

Tabela 5 – Melhores resultados de previsão do LR classificados de acordo com R^2 médio

Limite de resistência à tração – LR				
Hiperparâmetros		Estatísticas		
K	SIGMA (σ)	R^2 médio	R^2 máx	Desvio padrão
758	0,59	0,7403	0,8412	0,0660
782	0,55	0,7359	0,8590	0,0665
758	0,56	0,7326	0,8619	0,0646
770	0,62	0,7318	0,8708	0,0765
740	0,56	0,7311	0,8318	0,0688

Fonte: Elaborada pelo autor.

Tabela 6 – Melhores resultados de previsão do LR/LE classificados de acordo com R^2 médio

Razão entre limite de resistência e limite de escoamento - LR/LE				
Hiperparâmetros		Estatísticas		
K	SIGMA (σ)	R^2 médio	R^2 máx	Desvio padrão
644	0,57	0,8691	0,9166	0,0254
636	0,58	0,8683	0,9340	0,0298
580	0,57	0,8672	0,9199	0,0246
604	0,65	0,8665	0,9203	0,0259
596	0,62	0,8661	0,9145	0,0262

Fonte: Elaborada pelo autor.

Tabela 7 – Melhores resultados de previsão do AL classificados de acordo com R^2 médio

Alongamento – AL				
Hiperparâmetros		Estatísticas		
K	SIGMA (σ)	R^2 médio	R^2 máx	Desvio padrão
656	0,57	0,7015	0,8085	0,0457
632	0,48	0,7005	0,7708	0,0372
600	0,51	0,6998	0,7762	0,0358
768	0,48	0,6981	0,7857	0,0363
632	0,53	0,6967	0,7755	0,0400

Fonte: Elaborada pelo autor.

A Tabela 8 a seguir compara os melhores valores encontrados com a MLR e com a RBFN. Observa-se que para todas as propriedades mecânicas obteve-se um grande aumento de previsibilidade quando a rede neural foi utilizada.

Tabela 8 – Comparativo entre os melhores resultados encontrados por cada algoritmo de previsão

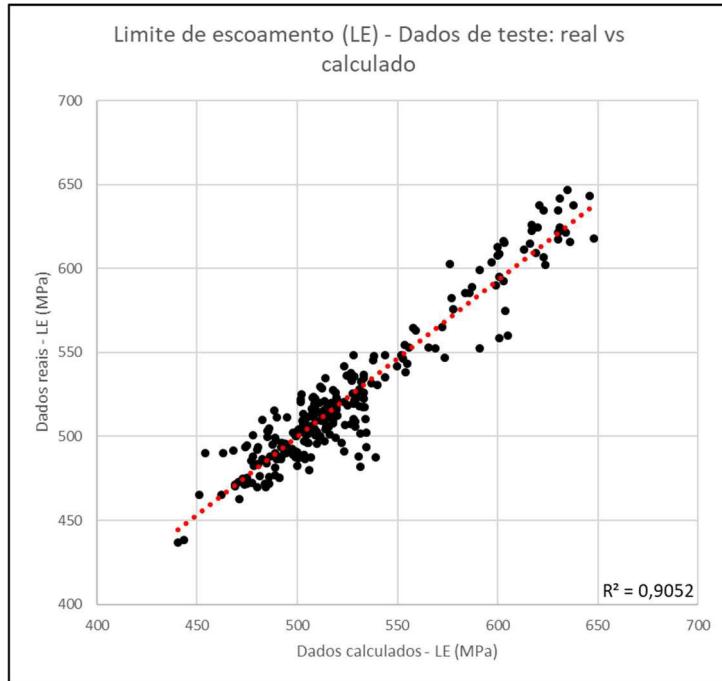
Propriedades mecânicas	R^2 médio			R^2 máx		
	MLR	RBFN	Aumento	MLR	RBFN	Aumento
<i>LE (MPa)</i>	0,5146	0,8437	64%	0,6165	0,9127	48%
<i>LR (MPa)</i>	0,2388	0,7403	210%	0,3504	0,8708	148%
<i>LR/LE</i>	0,7633	0,8691	14%	0,8132	0,9340	15%
<i>AL (%)</i>	0,5099	0,7015	38%	0,6033	0,8085	34%

Fonte: Elaborada pelo autor.

Essa situação é esperada e pode ser explicada de maneira bem simples: o formato complexo do problema, simulando um processo produtivo real, inviabiliza o uso de considerações lineares entre entrada e saída de dados, como a regressão multilinear sugere, bem como inviabiliza um estudo analítico simplista, como mencionado na introdução do trabalho. Ao invés disso, é necessário um método mais complexo como a RBFN que se utiliza de uma transformação de alta dimensionalidade para absorver características até então imperceptíveis por outros métodos, podendo reconhecer padrões de difícil compreensão e então conseguir construir uma função de previsão mais precisa.

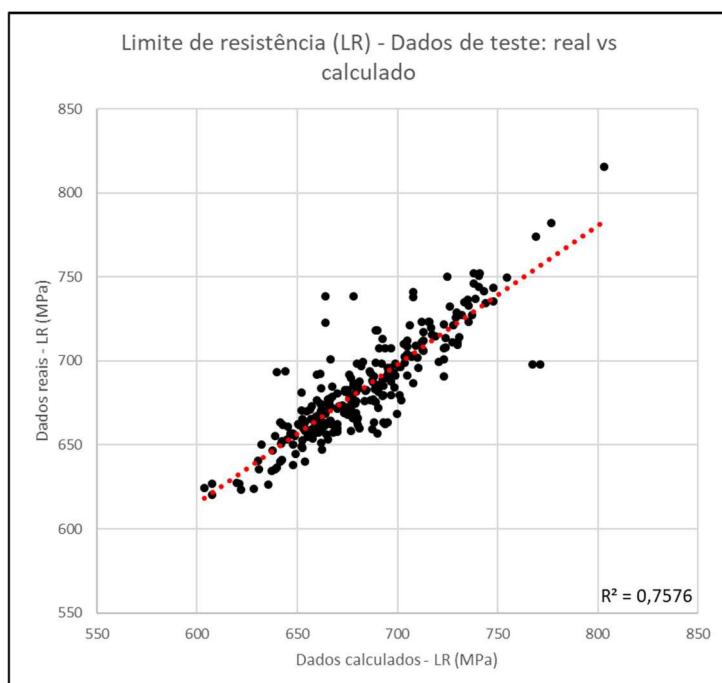
As Figuras 18, 19, 20 e 21 exemplificam por um gráfico de dispersão as respostas de uma das rodadas executadas para as melhores configurações da rede. No eixo y tem-se os valores de teste reais e no eixo x tem-se os valores de teste calculados pela rede. O R^2 de cada configuração está explícito na figura e representa apenas 1 das 100 rodadas.

Figura 18 – Exemplo de previsibilidade do LE de uma das rodadas de teste da RBFN



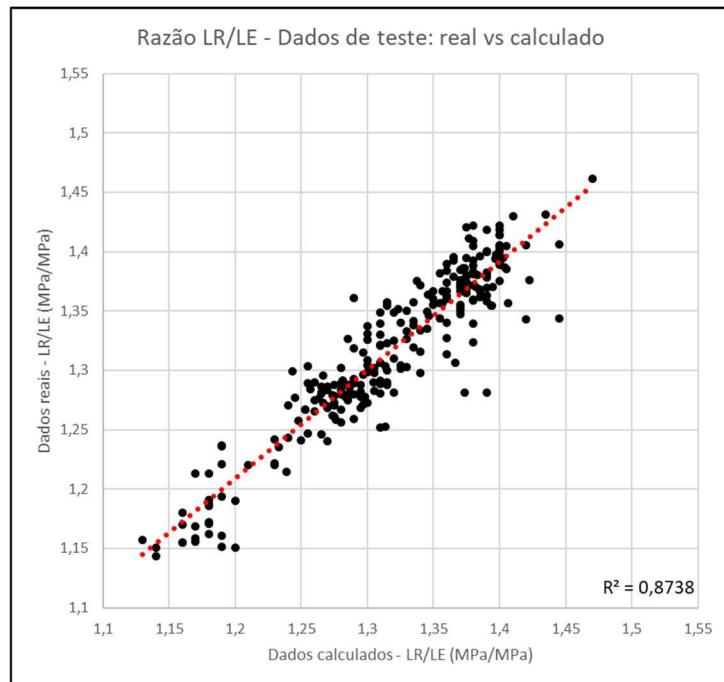
Fonte: Elaborada pelo autor.

Figura 19 – Exemplo de previsibilidade do LR de uma das rodadas de teste da RBFN



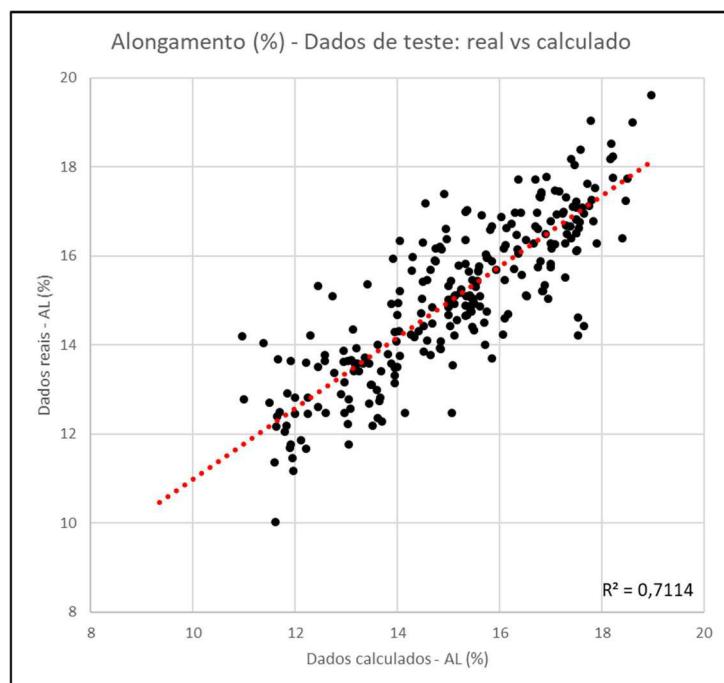
Fonte: Elaborada pelo autor.

Figura 20 – Exemplo de previsibilidade da razão LR/LE de uma das rodadas de teste da RBFN



Fonte: Elaborada pelo autor.

Figura 21 – Exemplo de previsibilidade do AL de uma das rodadas de teste da RBFN



Fonte: Elaborada pelo autor.

Nessas figuras, observa-se o que foi explicado na Seção 2.4, quanto mais próximos de uma reta estão os dados, maior é o R^2 , ou seja, maior a previsibilidade do método. De maneira geral, os valores das propriedades mecânicas reais dos dados de teste estão bem próximos dos valores das propriedades mecânicas calculadas pela rede neural.

O R^2 médio retirado dos dados de teste é importante pois dá uma maior confiabilidade estatística de qual seria o comportamento do algoritmo em uma situação real e não em uma tentativa que ao acaso conseguiu um alto valor de R^2 , como representado pelo R^2 máximo. Lembrando que essa variabilidade de resposta se deve ao fato dos dados de teste e de treino serem separados aleatoriamente para cada rodada de cada configuração de rede testada.

Vale relembrar que apenas os valores observados e estimados relacionados com dados de teste foram considerados para o cálculo do coeficiente de determinação, garantindo, dessa forma, dois aspectos importantes:

- Tentativa de reproduzir uma situação mais realística de aplicação do algoritmo onde, a priori, não são conhecidos e nem levados em conta os valores das propriedades mecânicas, garantindo que o método apenas respondeu com uma previsão calculada e baseada apenas em “experiências” do histórico de dados apresentados na etapa de treino.
- O valor do coeficiente de determinação mostrado nos resultados não sofre influência direta da etapa de treino, onde é esperado que o algoritmo apresente resultados melhores visto que está “aprendendo” com esses dados. No treino pode ocorrer o chamado *overtraining* onde, segundo Haykin (2009), o algoritmo na prática “decora” os resultados esperados, não refletindo, porém, na aplicação real/final do método.

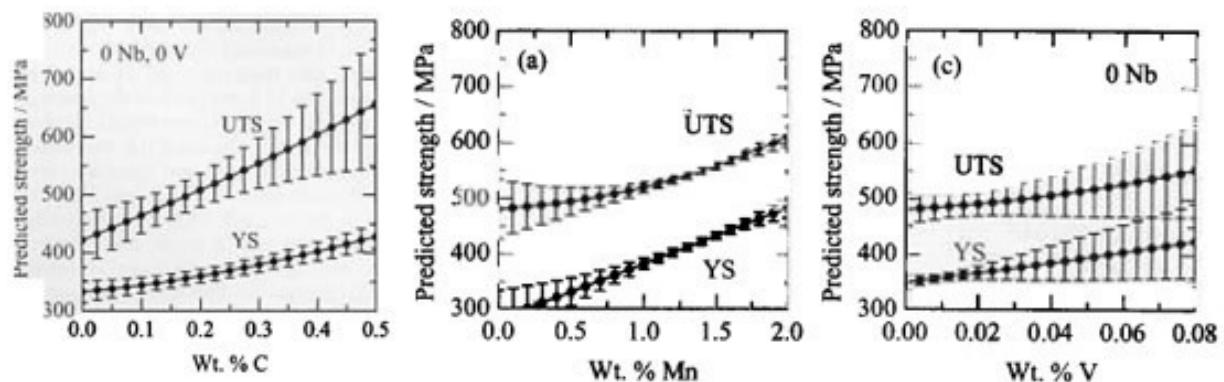
Segundo os trabalhos de Gorni e Da Silva (2009) e Jones, Watton e Brown (2005), esperava-se, oriundo dos erros intrínsecos aos métodos de medições de propriedades mecânicas, maior dificuldade na previsão do alongamento, seguido do limite de escoamento e então do limite de resistência à tração.

No presente trabalho, para a MLR, observou-se maior dificuldade na previsão do limite de resistência, estando o limite de escoamento e o alongamento em patamares parecidos de previsão. Para a RBFN, de fato, o alongamento é a propriedade mecânica que apresenta maior dificuldade de previsão, estando de acordo com as referências citadas, podendo ser explicada pelos erros embutidos na sua medição, estando seguida do limite de resistência e então do limite de escoamento.

Dessa forma, para os dois métodos, tem-se o limite de resistência como uma propriedade de mais difícil previsibilidade quando comparada ao limite de escoamento. A partir daí, buscou-se explicar esse resultado através de dois raciocínios diferentes.

O primeiro diz respeito aos estudos realizados por Singh, et al (1998). Nesse trabalho, são comparados os efeitos da adição de carbono e outros elementos de liga na previsão do limite de resistência à tração e no limite de escoamento em placas de aço microligadas e aço carbono comum. Mostra-se, através de pequenas variações na composição química que elementos como carbono, manganês, vanádio e nióbio são grandes influenciadores nas propriedades estudadas, contribuindo no crescimento das mesmas. A Figura 22 explica como se dá a variação dessas duas propriedades conforme aumenta-se a concentração de alguns elementos no aço.

Figura 22 – Variação na previsibilidade de propriedades mecânicas conforme variação de elementos químicos no aço



Fonte: Adaptado de Singh, et al (1998)

A partir do estudo de Singh, et al (1998), pode-se chegar a dois resultados importantes relacionados ao presente trabalho:

- O limite de resistência à tração é mais influenciado pela adição de elementos de liga do que o limite de escoamento;
- O limite de resistência à tração sofre maiores erros de previsão (representados pelo tamanho das barras de erro na Figura 22) quando em comparação com o limite de escoamento a partir da variação de elementos de liga;

Dessa forma, espera-se que pequenas variações na composição química do aço induzam a maiores variações no LR, introduzindo, além disso, maiores erros em sua previsão.

O segundo diz respeito aos conceitos apresentados por Hosford (2010). A princípio, sabe-se que, conforme Dos Santos (2017), variáveis como temperatura, tempo e grau de deformação plástica durante laminação à quente de um aço afetam sua recuperação e recristalização, influenciando, consequentemente, em seu limite de resistência à tração e outras propriedades mecânicas. Contudo, Hosford (2010) inclui mais uma variável, dizendo que a taxa de deformação também é parâmetro crucial. Ele diz que a tensão nominal de um material dada deformação e temperaturas fixas, pode ser representada pela Equação 4.1.

$$\sigma = C \dot{\varepsilon}^m \quad (4.1)$$

Onde σ representa a tensão nominal, C uma constante, $\dot{\varepsilon}$ uma taxa de deformação média sofrida e m uma sensibilidade à taxa de deformação.

Para a maioria dos metais, a sensibilidade à taxa de deformação, m , é baixa quando próxima à temperatura ambiente, mas aumenta com o aumento da temperatura, sendo um aumento bem rápido quando a temperatura ultrapassa a metade da temperatura de fusão do material. Considerando essa variação da sensibilidade, estima-se que em temperatura mais baixas, um aumento de 10 vezes na taxa de deformação levaria a um variação positiva da curva tensão-deformação de um material em apenas 1% ou 2%, já considerando um processo de laminação à quente, em altas temperaturas, esse mesmo aumento representaria uma variação positiva de 50% ou mais da curva tensão-deformação, significando limites de resistência finais completamente diferentes tendo como variável apenas a taxa de deformação do material.

Sendo assim, o parâmetro taxa de deformação mostra-se crucial para obtenção de valores de LR. Porém, nos dados obtidos da siderúrgica, não foram disponibilizados variáveis que pudessem estimar essa taxa, podendo esse fato ter contribuído para a menor assertividade dos métodos.

Em parte, essas duas linhas de raciocínio, em conjunto, explicam os resultados um pouco diferentes dos já encontrados e esperados por outros autores.

5 CONCLUSÃO

Por se tratar de um problema real, os resultados obtidos no trabalho com os dois algoritmos apresentaram altas taxas de acerto, conforme mostrado no capítulo anterior. De forma prática, adotando-os em um processo produtivo, seria possível estimar com boa confiabilidade as propriedades mecânicas finais de um vergalhão antes mesmo da finalização

da sua confecção, por exemplo, bastando para isso ter acesso a um histórico de produção para que os métodos propostos fossem atualizados e pudessem servir de estimadores.

Para uma siderúrgica que busca confiabilidade nas respostas de parâmetros tão importantes nos vergalhões como as propriedades mecânicas, a utilização de métodos preditivos como os apresentados mostra-se de grande auxílio nessa tarefa. Além disso, por se tratarem de métodos computacionais obtidos a partir de dados já adquiridos usualmente pela empresa, os custos inseridos na execução da tarefa são baixos.

Mostrou-se também que, para o tipo de problema em estudo, algoritmos um pouco mais complexos como a RBFN conseguem alcançar resultados bem mais precisos e corretos quando em comparação a regressões paramétricas simples como a MLR. Contudo, também se verificou que a aplicação desses métodos é de fácil implementação.

As análises realizadas mostraram que as propriedades mecânicas de mais difícil previsão foram o alongamento e o limite de resistência à tração. Em relação ao alongamento, esperava-se esse tipo de resposta pois em estudos similares foi encontrada dificuldade na previsão dessa variável, além disso, referências indicam maiores erros embutidos na aquisição desse tipo de dado durante a realização do ensaio mecânico. Já em relação à resistência à tração, entende-se que duas explicações em conjunto podem explicar o comportamento observado: os elementos de liga e os elementos residuais do aço exercem grande influência e inserem maiores variações na previsão dessa propriedade quando em comparação a outras; a taxa de deformação tem papel importante no comportamento mecânico dessa variável e não foi fornecida pela empresa.

Já as propriedades de mais fácil previsão foram o limite de escoamento e razão entre o limite de resistência e o limite de escoamento. Entende-se através disso que os principais parâmetros necessários para traçar o comportamento variável desses parâmetros foram utilizados e analisados pelos algoritmos, contribuindo para maior previsibilidade.

6 REFERÊNCIAS

- ALVARENGA JÚNIOR, Wagner José de. **Métodos de otimização hiperparamétrica: um estudo comparativo utilizando árvores de decisão e florestas aleatórias na classificação binária.** 2018.
- AMARI, Shun'ichi et al. **The handbook of brain theory and neural networks.** MIT press, 2003.
- ARAÚJO, LUIZ A. **Manual de Siderurgia.** São Paulo: Editora Arte & Ciência, 1997.
- ARAÚJO, V. T. **Caracterização e quantificação microestrutural de barras de aço produzidas via processo thermex.** Monografia (Monografia em Engenharia Metalúrgica) – Centro de Tecnologia, UFC (Universidade Federal do Ceará). Fortaleza, 2010.
- ASM METALS. **Handbook 14: Forming and Forging.** ASM International, 1993.
- ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 7480: Aço destinado a armaduras para estruturas de concreto armado–Especificação.** Rio de Janeiro, 2007.
- BENGIO, Yoshua. **Gradient-based optimization of hyperparameters.** Neural computation, v. 12, n. 8, p. 1889-1900, 2000.
- BERGSTRA, James; BENGIO, Yoshua. **Random search for hyper-parameter optimization.** Journal of Machine Learning Research, v. 13, n. Feb, p. 281-305, 2012.
- BROOMHEAD, D. S.; LOWE, D. **Multivariable functional interpolation and adaptive networks.** Complex systems, vol. 2. 1988.
- CAÇÃO, Rosário. **Testes Estatísticos–Testes Paramétricos e Não Paramétricos.** 2010.
- CALLISTER JR, William D. **Ciência e Engenharia de Materiais: Uma Introdução,** 5^a Edição. Rio de Janeiro: LTC, 2002.
- CHIAVERINI, Vicente. **Aços e ferros fundidos.** Abm, 1977.
- COSTA, Verlaine Lia; ESCORSIM, Sérgio; COSTA, Deneive Leonor. **Processo produtivo e produção de aço: a inserção do Grupo Gerdau SA no cenário mundial.** 2007.
- COVER, Thomas M. **Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition.** IEEE transactions on electronic computers, n. 3, p. 326-334, 1965.
- FENG, Hsuan-Ming; CHOU, Hsiang-Chai. **Evolutional RBFNs prediction systems generation in the applications of financial time series data.** Expert Systems with Applications, v. 38, n. 7, p. 8285-8292, 2011.

GORNI, Antonio Augusto; DA SILVA, Marcos Roberto S. **Previsão de propriedades mecânicas de bobinas laminadas a quente através de redes neurais.** Tecnologia em Metalurgia, Materiais e Mineração, v. 5, n. 4, p. 229, 2009.

HASANBEIGI, A.; PRICE, LYNN K.; MCKANE, A. T. **The State-of-the-Art Clean Technologies (SOACT) for Steelmaking Handbook.** 2010.

HAYKIN, Simon. **Neural Networks and Learning Machines.** 3 ed, p. 1-263, New York: Prentice Hall, 2009.

HOFFMANN, Rodolfo; VIEIRA, Sônia. **Análise de regressão: uma introdução à econometria.** São Paulo, 2016.

INSTITUTO AÇO BRASIL. **Processo Siderúrgico.** Disponível em: <http://www.acobrasil.org.br/site2015/processo.html>. Acesso em: 08 de out. de 2019

KAVAKLIOGLU, Kadir; KOSEOGLU, Mehmet F.; CALISKAN, Oguz. **Experimental investigation and radial basis function network modeling of direct evaporative cooling systems.** International Journal of Heat and Mass Transfer, v. 126, p. 139-150, 2018.

KOSTIN, V. A. **Mathematical formulation of carbon equivalent as a criterion for evaluation of steel weldability.** 2012.

KREBS, Claudia. **Neurociências Ilustrada.** Artmed Editora, 2013.

KRIESEL, David. **A brief introduction on neural networks.** 2007.

LERMAN, P. M. **Fitting segmented regression models by grid search.** Journal of the Royal Statistical Society: Series C (Applied Statistics), v. 29, n. 1, p. 77-84, 1980.

LIU, Jinkun. **Radial Basis Function (RBF) neural network control for mechanical systems: design, analysis and Matlab simulation.** Springer Science & Business Media, 2013.

LUO, Gang. **A review of automatic selection methods for machine learning algorithms and hyper-parameter values.** Network Modeling Analysis in Health Informatics and Bioinformatics, v. 5, n. 1, p. 18, 2016.

MARCOS, Ana González; ELÍAS, Fernando Alba. **Machine Learning en la industria: el caso de la siderurgia.** Economía industrial, n. 405, p. 55-63, 2017.

MARSLAND, Stephen. **Machine learning: an algorithmic perspective.** Chapman and Hall/CRC, 2014.

MARTINELI, Edmar. **Extração de conhecimento de redes neurais artificiais.** 1999. Tese de Doutorado. Universidade de São Paulo.

MONTGOMERY, Douglas C.; PECK, Elizabeth A.; VINING, G. Geoffrey. **Introduction to linear regression analysis.** John Wiley & Sons, 2012.

MUNNOLI, Sneha P.; BAPAT, A. U. **Clustering algorithms for radial basis function neural network.** ITSI Transactions on Electrical and Electronics Engineering (ITSI-TEEE), v. 1, p. 113-116, 2013.

OBERMEYER, Ziad; EMANUEL, Ezekiel J. **Predicting the future-big data, machine learning, and clinical medicine.** The New England journal of medicine, v. 375, n. 13, p. 1216, 2016.

PITOMBO, Cira Souza; COSTA, Aline Schindler Gomes da. **Aplicação conjunta de modelos não paramétricos e paramétricos para previsão de escolha modal.** Journal of Transport Literature, v. 9, n. 1, p. 30-34, 2015.

RAUBER, Thomas Walter. **Redes neurais artificiais.** Universidade Federal do Espírito Santo, 2005.

RIBEIRO, Priscilla Cristina Cabral; VIEIRA, Leandro da Silva. **TECNOLOGIA DA INFORMAÇÃO E COMPETITIVIDADE NA INDÚSTRIA SIDERÚRGICA BRASILEIRA: UM ESTUDO DE CASO.** 2001.

RIZZO, Ernandes Marcos Silveira. **INTRODUÇÃO AOS PROCESSOS DE REFINO PRIMÁRIO DOS AÇOS NOS FORNOS ELÉTRICOS A ARCO.** ABM, 2006.

ROBERTS, William L. **Hot rolling of steel.** CRC Press, 1983.

SCHWENKER, Friedhelm; KESTLER, Hans A.; PALM, Günther. **Three learning phases for radial-basis-function networks.** Neural Networks, v. 14, n. 4-5, p. 439-458, 2001.

SILVA, Andre Luiz da Costa E.; MEI, Paulo Roberto. **Aços e ligas especiais.** Edgard Blücher, 2008.

SILVA, Saulo Rodrigues; SCHIMIDT, Fernando. **REDUCTION OF INPUT VARIABLES IN ARTIFICIAL NEURAL NETWORKS AS FROM PRINCIPAL COMPONENTS ANALYSIS DATA IN THE MODELING OF DISSOLVED OXYGEN.** Química Nova, v. 39, n. 3, p. 273-278, 2016.

SINGH, S. B. et al. **Neural network analysis of steel plate processing.** Ironmaking and Steelmaking, v. 25, n. 5, p. 355-365, 1998.

SOLIS, Francisco J.; WETS, Roger J.-B. **Minimization by random search techniques.** Mathematics of operations research, v. 6, n. 1, p. 19-30, 1981.

SQUIRE, Larry et al. (Ed.). **Fundamental neuroscience.** Academic Press, 2012.

STEEL UNIVERSITY. **EAF User Guide.** Disponível em:
<https://content.steeluniversity.org/simulators/eaf/userguide>. Acesso em: 09 de out. de 2019

STEEL UNIVERSITY. **EAF Simulator.** Disponível em:
<https://content.steeluniversity.org/simulators/eaf/index.html>. Acesso em: 29 de nov. de 2019

UNIVERSIDADE FEDERAL DO CEARÁ. Biblioteca Universitária. **Guia de normalização de trabalhos acadêmicos da Universidade Federal do Ceará.** Fortaleza, 2013.

VON ZUBEN, IA353–Profs Fernando J.; ATTUX, Romis RF. **Redes Neurais com Funções de Ativação de Base Radial.** DCA/FEEC/Unicamp, 2001.

WITTEN, Ian H. et al. **Data Mining: Practical machine learning tools and techniques.** Morgan Kaufmann, 2016.