

UNIVERSIDADE FEDERAL DE SANTA CATARINA
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E
SISTEMAS
CURSO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
MECÂNICA

Eduardo Lacerda Pereira

**Aplicação de um modelo de aprendizado de máquina para
previsão do desgaste de fresas de topo esférico**

Florianópolis

2020

Eduardo Lacerda Pereira

**Aplicação de um modelo de aprendizado de máquina para
previsão do desgaste de fresas de topo esférico**

Trabalho de Conclusão de Curso apresentado ao Departamento de Engenharia de Produção e Sistemas da Universidade Federal de Santa Catarina como requisito parcial para obtenção do título em Engenharia, área Mecânica, habilitação Engenharia de Produção Mecânica.

Orientador: Prof. Eduardo Ferreira da Silva, PhD.

Florianópolis

2020

Ficha de identificação da obra elaborada pelo autor,
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Pereira, Eduardo Lacerda

Aplicação de um modelo de aprendizado de máquina
paraprevisão do desgaste de fresas de topo esférico /
Eduardo Lacerda Pereira ; orientador, Eduardo Ferreira da
Silva , 2020.

77 p.

Trabalho de Conclusão de Curso (graduação) -
Universidade Federal de Santa Catarina, Centro Tecnológico,
Graduação em Engenharia de Produção Mecânica, Florianópolis,
2020.

Inclui referências.

1. Engenharia de Produção Mecânica. 2. Desgaste de
Ferramentas. 3. Fresas de Topo Esférico. 4. Aprendizado de
Máquina. 5. Redes Neurais. I. , Eduardo Ferreira da Silva.
II. Universidade Federal de Santa Catarina. Graduação em
Engenharia de Produção Mecânica. III. Título.

Eduardo Lacerda Pereira

Aplicação de um modelo de aprendizado de máquina para previsão do desgaste de fresas de topo esférico

Este Trabalho de Conclusão de Curso foi julgado adequado e aprovado, em sua forma final, pelo Curso de Graduação em Engenharia de Produção Mecânica, da Universidade Federal de Santa Catarina.

Florianópolis, 12 de Agosto de 2020.



Documento assinado digitalmente

Guilherme Ernani Vieira

Data: 25/10/2020 18:01:53-0300

CPF: 888.311.759-04

Prof. Guilherme Ernani Vieira, Dr.
Coordenador do Curso

Banca Examinadora:



Documento assinado digitalmente

Eduardo Ferreira da Silva

Data: 24/10/2020 10:30:12-0300

CPF: 790.231.347-53

Prof. Eduardo Ferreira da Silva, PhD.
Orientador

Universidade Federal de Santa Catarina



Documento assinado digitalmente

Ricardo Faria Giglio

Data: 23/10/2020 08:41:09-0300

CPF: 304.078.978-36

Prof. Ricardo Faria Giglio, Dr.
Avaliador



Documento assinado digitalmente

Rolf Bertrand Schroeter

Data: 21/10/2020 17:01:51-0300

CPF: 510.631.589-15

Prof. Rolf Bertrand Schroeter, Dr.
Avaliador
Universidade Federal de Santa Catarina

AGRADECIMENTOS

A meus pais, Gabriel e Letícia, por todo o apoio e suporte dados durante meus anos de universidade. Também a meus familiares de maneira geral, assim como a meus amigos por tornarem o processo de desenvolvimento deste trabalho mais agradável.

Ao Prof. Eduardo Ferreira da Silva, por aceitar orientar um trabalho já iniciado e ajudar-me a conduzi-lo ao seu estado final. Agradeço ainda ao Prof. Rolf Schroeter pelo breve, mas produtivo, suporte dado ao trabalho.

Ao Laboratório de Mecânica de Precisão (LMP), em especial ao Prof. Walter Weingaertner e a Fernando Bordin, que me introduziram ao mundo da usinagem e abriram as portas para minha ida a Alemanha.

Ao Fraunhofer IPT, em especial ao meu chefe Willi Tontsch e ao meu então orientador Carsten Holst, por permitirem que eu iniciasse este projeto em paralelo ao meu estágio e darem o suporte necessário para este fosse realizado.

RESUMO

Este trabalho tem como objetivo utilizar um modelo de aprendizado de máquina para realizar a previsão do desgaste esperado em uma fresa de topo esférico utilizada num processo de fresamento. Ele surgiu a partir da necessidade de complementar um *software* de simulação de usinagem desenvolvido pelo Fraunhofer IPT. Além disso, os modelos de previsão de desgaste tradicionais são pouco flexíveis e costumam depender de uma grande quantidade de experimentos. A alternativa encontrada foi utilizar dados obtidos na literatura, os quais foram recriados no mencionado *software* e, após análise e tratamento, usados como entradas para diferentes redes neurais. Inicialmente foram montados modelos simplificados, a fim de verificar a viabilidade de utilizar redes neurais para resolver o problema. Uma vez verificada, o modelo foi expandido para o conjunto completo de dados, tanto de forma simplista como através de diferentes tratamentos do balanceamento do dados de entrada. Como os resultados obtidos nos modelos de regressão ficaram aquém do esperado, quando avaliados pelo seu R^2 foram testadas redes de classificação, as quais forneceram resultados com valores de acurácia na faixa de 0,9 embora forneçam informações menos detalhadas do processo.

Palavras-chave: Desgaste de Ferramentas; Fresas de Topo Esférico; Aprendizado de Máquina; Redes Neurais.

ABSTRACT

This work's objective is to use a machine learning model to predict the expected wear on a spherical end mill used in a milling process. It arose from the need to complement a machining simulation software developed at Fraunhofer IPT. Additionally, traditional wear prediction models are not very flexible and usually depend on a large number of experiments. The alternative found was to use data obtained from the literature, which were then recreated in the mentioned software and, after analysis and treatment, used as inputs for different neural networks. Initially, simplified models were built in order to verify the feasibility of using neural networks to solve the problem. Once verified, the model was expanded to the complete dataset not only in a naive setting as well as with different uses of input data balancing. As the results obtained in the regression models fell short of expectations when evaluated via their R^2 value, classification networks were tested, providing accuracy values in the range of 0,9 although less detailed information about the process can be known with them.

Keywords: Tool Wear; Ball Nose End Mills; Machine Learning; Neural Networks

LISTA DE FIGURAS

Figura 1 – Comparação entre diversos processos de fabricação . . .	15
Figura 2 – Peça com acabamento realizado por fresamento	16
Figura 3 – Fresa de topo esférico	16
Figura 4 – Esquematização do problema de pesquisa proposto . .	17
Figura 5 – Subdivisões da usinagem com ferramentas de geometria definida	21
Figura 6 – Cinemática do torneamento e fresamento	22
Figura 7 – Dimensões do cavaco no torneamento	22
Figura 8 – Fresamento concordante e discordante	24
Figura 9 – Velocidade de corte e sua influência nos mecanismos de desgaste	25
Figura 10 – Medição do desgaste de flanco VB	27
Figura 11 – Perceptron de neurônio único	33
Figura 12 – Representação gráfica das funções de ativação	35
Figura 13 – Divergência entre as perdas de treinamento e validação	37
Figura 14 – Fluxograma do trabalho realizado	40
Figura 15 – Distribuição dos materiais das peças e ferramentas antes da seleção dos artigos com os quais o trabalho será realizado	43
Figura 16 – Classe de material da peças	44
Figura 17 – Desgaste (VB) da ferramenta em função do comprimento cortado para cada um dos artigos	45
Figura 18 – Materiais das peças considerados no modelo	47
Figura 19 – Exemplo da modelagem com o caminho da ferramenta	48
Figura 20 – Distribuição dos dados em cada faixa de desgaste para o artigo 06 antes e depois da remoção dos valores nulos	51
Figura 21 – Distribuição dos valores de desgaste dentro de cada artigo	52
Figura 22 – Evolução das perdas e R^2 no treinamento inicial da rede completa	55

Figura 23 – Distribuição dos valores de desgaste dentre os dados disponíveis	56
Figura 24 – Demonstração do <i>oversampling</i> realizado antes da divisão dos dados para treinamento e validação . . .	58
Figura 25 – Distribuição dos dados de treinamento antes e após a aplicação da técnicas de <i>oversampling</i> e <i>undersampling</i>	59
Figura 26 – Evolução da perda e do R^2 durante o treinamento após a aplicação da técnicas de <i>oversampling</i> e <i>undersampling</i>	60
Figura 27 – Evolução da perda e da acurácia durante o treinamento para os modelos de classificação com 3 e com 5 classes	62
Figura 28 – Matriz de confusão para o modelo com 3 classes . .	66
Figura 29 – Matriz de confusão para o modelo com 5 classes . .	66

LISTA DE TABELAS

Tabela 1	–	Resumo dos critérios de seleção dos artigos utilizados. Em negrito são destacados os artigos que foram utilizados no trabalho.	46
Tabela 2	–	Exemplo do resultado exportado para cada interação	49
Tabela 3	–	Topologia da rede utilizada nos modelos preliminares	53
Tabela 4	–	Média das métricas de validação cruzada de cada artigo para as redes preliminares	53
Tabela 5	–	R^2 obtido ao executar-se validação cruzada nos dados balanceados	60
Tabela 6	–	Avaliação por classes dos modelos de classificação	63
Tabela 7	–	Avaliação por classes do modelo de regressão simplista	64

LISTA DE ABREVIATURAS E SIGLAS

CNC	Comando Numérico Computadorizado
DIN	Instituto Alemão para Normatização
ELU	Unidade linear exponencial
FEM	Método dos elementos finitos
IPT	Instituto de Tecnologia para Produção
ISO	Organização Internacional para Padronização
MAE	Erro absoluto médio
MSE	Erro quadrático médio
RELU	Unidade linear retificada
SELU	Unidade linear exponencial escalada
SGD	Gradiente estocástico descendente
SNN	Rede neural auto-normalizante
SSE	Soma dos erros ao quadrado
SST	Soma total dos quadrados

LISTA DE SÍMBOLOS

v_f	Velocidade de avanço
v_c	Velocidade de corte
v_e	Velocidade efetiva
a_p	Profundidade de corte
a_e	Largura de corte
VB	Desgaste de flanco
VB_B	Desgaste de flanco médio
VB_{max}	Desgaste de flanco máximo
SV_α	Deslocamento na direção do flanco
SV_γ	Deslocamento na direção da face
KB	Largura da cratera
KM	Distância do centro da cratera ao centro do gume
KT	Profundidade da cratera
f	Avanço por dente
Q_w	Taxa volumétrica de remoção de material
C_v	Tempo de vida com velocidade de corte 1 m min^{-1}
C_T	Velocidade de corte para 1 min de vida
k	Relação entre C_v e C_T
X	Vetor de <i>features</i> de entrada do modelo
x_i	Elemento do vetor de <i>features</i> de entrada do modelo
\hat{Y}	Vetor de Saída do modelo
w_i	Elemento do vetor de pesos
a	Atividade de um neurônio
$f(a)$	Função de ativação de um neurônio
λ	Hiper-parâmetro
α	Hiper-parâmetro
Y	Vetor de saídas reais do modelo
L	Função de perdas
W	Vetor de pesos
$R(\theta)$	Função de perdas com elemento de regularização
CL	Comprimento cortado

N_{caminhos}	Número de cortes paralelos realizados
n	Velocidade de rotação do eixo-árvore
delta_phi	Ângulo simulado do arco
hsp	Profundidade do cavaco simulado
bsp	Altura do cavaco simulado
lsp	Largura do cavaco simulado
R^2	Coefficiente de determinação

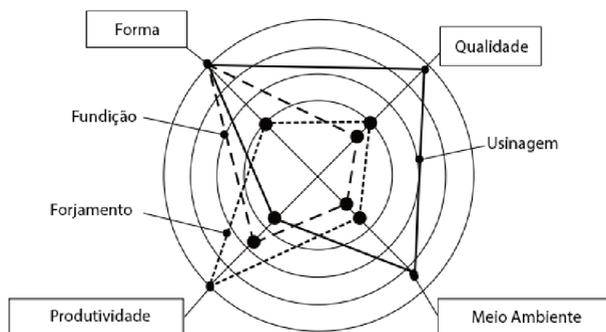
SUMÁRIO

1	INTRODUÇÃO	14
1.1	OBJETIVOS	18
1.1.1	Objetivo Geral	18
1.1.2	Objetivos Específicos	18
1.2	DELIMITAÇÃO DO TRABALHO	18
1.3	ESTRUTURA DO TEXTO	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	USINAGEM	20
2.1.1	Fresamento	23
2.1.2	Desgaste de ferramentas	24
2.1.3	Modelos de previsão do desgaste	28
2.1.3.1	Modelo de Taylor	30
2.2	APRENDIZADO DE MÁQUINA E REDES NEURAIS	31
2.2.1	A rede de neurônio único	32
2.2.2	Métodos de otimização, <i>overfitting</i> e regularização	35
3	METODOLOGIA	39
3.1	TIPO DE PESQUISA	39
3.2	ETAPAS METODOLÓGICAS	39
4	DESENVOLVIMENTO	41
4.1	OBTENÇÃO E TRATAMENTO INICIAL DOS DADOS	41
4.2	RECRIAÇÃO DIGITAL DOS ARTIGOS	46
4.3	MODELOS PRELIMINARES	51
4.4	MODELOS GERAIS	54
4.4.1	Modelo de regressão	54
4.4.1.1	Modelo simplista	54
4.4.1.2	Modelo com tratamento dos dados desbalanceados	56
4.4.2	Modelo de classificação	59
5	CONCLUSÃO	67
5.1	CONSIDERAÇÕES FINAIS	68
	REFERÊNCIAS	70

1 INTRODUÇÃO

Muitos são os processos aplicáveis à fabricação de componentes mecânicos. Dentre estes, um de grande importância é o fresamento, um processo de usinagem, que é definido pela DIN 8580 como sendo processos de fabricação por remoção de cavaco, ou seja, de material. Estes processos de fabricação se mostram relevantes por apresentarem uma enorme gama de possibilidades em termos de qualidade, produtividade e variedade de formas (TÖNSHOFF; DENKENA, 2011). Além disso, como mostrado na Figura 1, com estes processos é possível atingir-se uma excelente qualidade das peças fabricadas, tendo-se em contrapartida, uma menor produtividade em termos de unidades produzidas quando comparado a processos como fundição. Por este motivo, em mais de 80% dos casos os processos de usinagem encontram-se no fim da cadeia de produção (ILSCHNER; SINGER, 2010). Isso faz com que falhas nesta etapa sejam custosas para empresa. Somado a isto, tem-se o fato de que ferramentas desgastadas produzem superfícies cujo acabamento é pior que o projetado. Nota-se também sua elevada relevância econômica, tendo o mercado de ferramentas de usinagem movimentado durante o ano de 2019, apenas nos Estados Unidos, US\$2,4 bilhões (AMT/USCTI, 2020).

Figura 1 – Comparação entre diversos processos de fabricação



Fonte – Tradução livre de Tönshoff e Denkena (2011)

Os processos de usinagem podem ser adicionalmente divididos, de acordo com a norma DIN 8589, conforme a geometria da ferramenta utilizada e da peça fabricada, o tipo de movimento realizado pela ferramenta e a posição da superfície usinada (interna, externa) (TÖNSHOFF; DENKENA, 2011). O processo de fresamento é definido, de forma genérica, como um processo de fabricação por remoção de material, utilizando ferramentas de geometria definida, que em geral possui diversos dentes e cujo movimento de corte é circular (KLOCKE; KÖNIG, 2008). A Figura 2 mostra uma situação típica onde o processo de fresamento é utilizado. No caso de fabricação de superfícies complexas, como as pás dos rotores de compressores para motores de avião, costumam ser utilizadas fresas de topo esférico, como mostrada na Figura 3. Estas fresas, são capazes de realizar cortes também quando seu eixo não está perpendicular à peça.

A ocorrência de desgaste em ferramentas de usinagem é dependente das condições de operação às quais são submetidas. Em geral, o desgaste aumenta conforme o parâmetros de processo são intensificados (no sentido de aumentar a produtividade do processo), além de possuir causas que são de difícil modelagem por, apesar de conhecidas, atuarem simultaneamente (KLOCKE; KÖNIG, 2008). Isso faz com que a

Figura 2 – Peça com acabamento realizado por fresamento



Fonte – O Autor

Figura 3 – Fresa de topo esférico



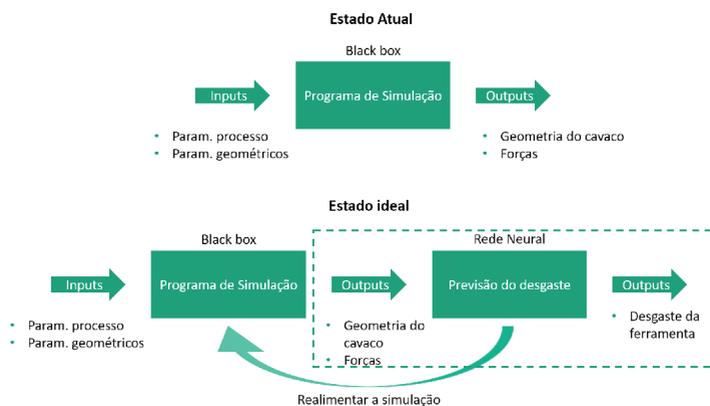
Fonte – O Autor

determinação dos parâmetros de corte para grandes produções em série costume ser feita através de extensos ensaios específicos para a peça que será produzida, enquanto peças fabricadas em menor escala dependem da experiência do operador (TÖNSHOFF; DENKENA, 2011). Cons-

truir um modelo capaz de realizar a previsão do desgaste da ferramenta significa permitir uma melhor e mais acessível determinação dos parâmetros que levam o processo aos objetivos almejados pela empresa (seja ele máxima produtividade, maior tempo de ferramenta, menor custo do processo como um todo, etc.).

Neste sentido, o Fraunhofer IPT, Instituto de Tecnologia para Produção, localizado em Aachen - Alemanha, onde o estágio curricular do autor foi realizado, possui um *software* para simulação do processo de fresamento de superfícies complexas. Este programa é capaz de fornecer informações quanto as condições geométricas dos cavacos gerados e esforços envolvidos no processo, no entanto, não é capaz de fazer previsões quanto ao desgaste que ocorre na ferramenta durante o seu uso, ou seja, a totalidade da simulação é feita considerando a geometria original da fresa utilizada. Porém, este desgaste interfere na capacidade de corte das fresas, alterando o resultado obtido no processo. Utilizando modelos de aprendizado de máquina, pretende-se construir um modelo que seja capaz de realizar essa previsão, seja para realimentar o programa de simulação, seja para realizar-se uma previsão direta a partir de medições realizadas no processo real. A Figura 4 ilustra este problema.

Figura 4 – Esquemática do problema de pesquisa proposto



Fonte – O Autor

1.1 OBJETIVOS

Nas seções a seguir estão descritos o objetivo geral e os objetivos específicos deste trabalho.

1.1.1 Objetivo Geral

Prever o desgaste de ferramentas num processo de fresamento com fresa de topo esférico.

1.1.2 Objetivos Específicos

- a) Verificar se os dados disponíveis realmente representam situações similares;
- b) Sistematizar o processo de tratamento dos dados disponíveis para aplicação numa ferramenta de aprendizado de máquina;
- c) Treinar um modelo capaz de realizar a previsão do desgaste.

1.2 DELIMITAÇÃO DO TRABALHO

Desde o início, faz-se relevante definir os limites aos quais este trabalho se aterá. Nele, não pretende-se criar um modelo genérico aplicável a todo e qualquer fresamento com fresas de topo esférico. Seus resultados são apropriados apenas em situações semelhantes às aquelas encontradas nos dados de treinamento. Além disso, os parâmetros de entrada nele utilizados devem ser de fácil obtenção, não necessitando de experimentos e/ou ensaios complexos, ou seja, mesmo empresas com poucos recursos seriam capazes de obtê-los. Esta restrição surgiu como requisito ainda na fase inicial do trabalho, realizada na Alemanha. Outra restrição que surgiu nesta fase inicial foi a de que o modelo de aprendizado de máquina utilizado seria uma rede neural, por este ter sido um trabalho que, na época, estava associado a outros trabalhos desta área em desenvolvimento no instituto.

1.3 ESTRUTURA DO TEXTO

Este trabalho está estruturado em 5 capítulos. Este é o primeiro capítulo, que traz uma introdução ao tema do trabalho e do problema de pesquisa. O Capítulo 2, intitulado Fundamentação Teórica, trata dos conceitos necessários para o entendimento do trabalho. No Capítulo 3 a metodologia do trabalho é apresentada, juntamente com uma esquete dos passos tomados no desenvolvimento. Este é aprofundado no Capítulo 4, onde está o desenvolvimento propriamente dito do trabalho e uma discussão dos resultados obtidos. Por fim, o Capítulo 5 traz as conclusões do trabalho, com algumas sugestões para trabalhos futuros e um breve comentário quanto ao processo de elaboração da monografia como um todo.

2 FUNDAMENTAÇÃO TEÓRICA

Nesta seção são apresentados conceitos-chave para o trabalho, com base em um revisão bibliográfica.

2.1 USINAGEM

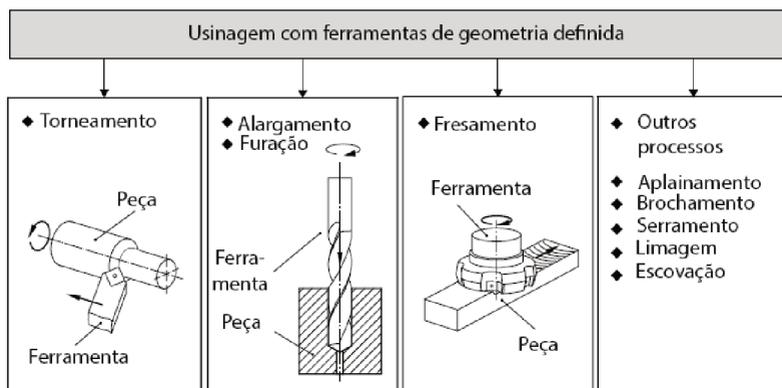
Usinagem é um dentre muitos métodos de fabricação aplicáveis a componentes mecânicos. Ao contrário de outros métodos como conformação, soldagem e manufatura aditiva, que trabalham tanto pela adição de material como pela transformação da peça, a usinagem atinge a geometria desejada da peça através da remoção de material (TÖNSHOFF; DENKENA, 2011). Apesar de este processo de fabricação poder levar a uma grande quantidade de material descartado, ele é de grande importância econômica devido às pequenas tolerâncias alcançáveis, indo de tolerâncias-padrão ISO IT¹10 até IT2, tornando-o ideal para processos de acabamento. Além disso, os processos de usinagem são de grande versatilidade, sendo empregáveis não só para produção em larga escala, como também em lotes menores ou mesmo para produtos personalizados (TÖNSHOFF; DENKENA, 2011).

Os processos de usinagem são divididos, de acordo com a norma DIN 8589, em processos onde a ferramenta de corte tem uma geometria definida, ou seja, toda a geometria da ferramenta é conhecida, e processos onde a ferramenta de corte tem uma geometria não definida, isto é, a geometria da ferramenta de corte só pode ser descrita através de métodos estatísticos, como no caso de um rebolo de retificação (TÖNSHOFF; DENKENA, 2011). A usinagem com ferramentas de geometria definida pode ser dividida ainda mais com base na cinemática do processo, como mostrado na Figura 5.

Embora cada um dos diferentes processos de usinagem tenha suas próprias particularidades cinemáticas, o processo de torneamento pode ser usado como um modelo geral e é aqui descrito de forma superficial.

¹ Sistema de tolerâncias-padrão definido pela Associação Brasileira de Normas Técnicas (1995)

Figura 5 – Subdivisões da usinagem com ferramentas de geometria definida

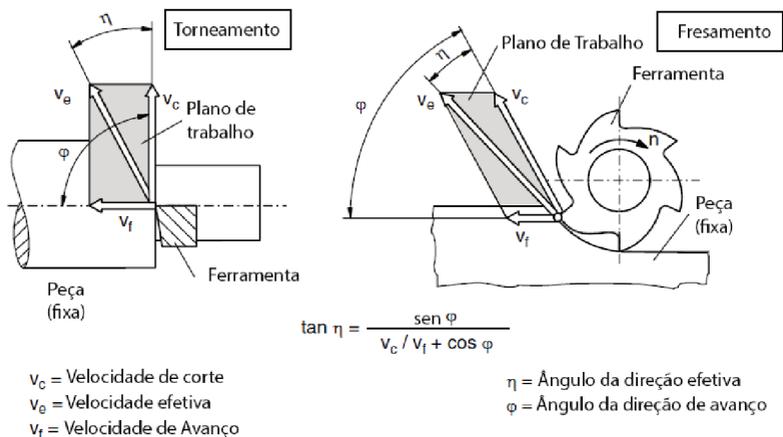


Fonte – Adaptado de Tönshoff e Denkena (2011)

Mais detalhes podem ser encontrados em Stemmer (1993), Tönshoff e Denkena (2011) e Klocke e König (2008). No torneamento, a ferramenta é pressionada contra a peça e deslocada em relação a ela num movimento chamado avanço com velocidade v_f . Simultaneamente, a peça é rotacionada em torno de seu próprio eixo com uma determinada velocidade angular. A velocidade tangencial no ponto de contato com a ferramenta se denomina velocidade de corte v_c . A soma vetorial destas duas velocidades dá a velocidade de corte efetiva v_e . Ao definir um processo de torneamento deve-se, ainda, definir a profundidade de corte a_p , que é o quanto de material será removido por avanço. As Figuras 6 e 7 ilustram estas grandezas.

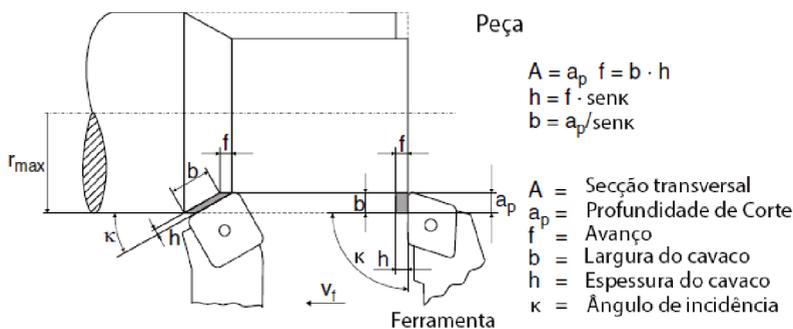
O processo de fresamento funciona de forma semelhante, mas a ferramenta assume o papel de rotacionar ao invés da peça. Da mesma forma, a velocidade de corte é definida no ponto de contato entre a ferramenta e a peça e a velocidade de avanço pela velocidade relativa dessas duas partes. Outra diferença entre estes dois processos que deve ser considerada é que dentro da fresagem cada uma dos gumes da

Figura 6 – Cinemática do torneamento e fresamento



Fonte – Tradução livre de Tönshoff e Denkena (2011)

Figura 7 – Dimensões do cavaco no torneamento



Fonte – Tradução livre de Tönshoff e Denkena (2011)

ferramenta está em contato com a peça durante um período de tempo limitado antes de perder o contato com a peça e mover-se livremente, ponto em que o gume seguinte entra em contato com a peça. Este é o chamado corte interrompido. No caso de fresamento, a_p é definido como a profundidade na peça de trabalho em que a ferramenta é definida

axialmente e tem-se, ainda, a largura de corte a_e , similar à profundidade de corte mas na direção radial. Além destas duas quantidades, o avanço por dente é utilizado como terceiro parâmetro e define a distância entre duas interações da ferramenta com a peça, influenciando a velocidade de avanço.

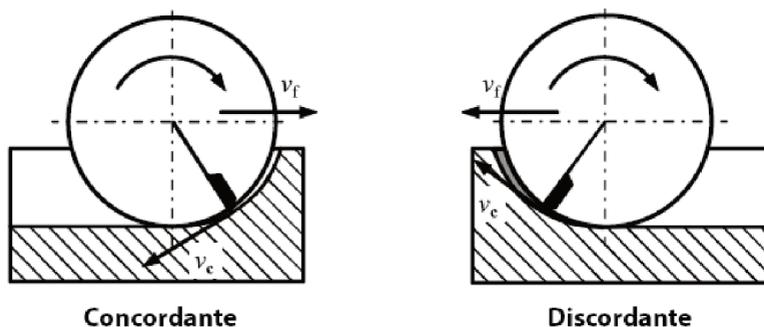
2.1.1 Fresamento

O fresamento é um processo de fabricação onde o movimento principal necessário para a usinagem da peça é feito pela ferramenta, que gira em torno de seu próprio eixo. O movimento secundário é uma translação relativa entre a ferramenta e a peça de trabalho, que pode ser feita pela ferramenta ou pela peça de trabalho ou mesmo por ambos simultaneamente, dependendo da máquina utilizada. A maioria das fresadoras está equipada com três eixos de translação independentes (X, Y e Z) e são chamadas de máquinas de 3 eixos. Por vezes, adicionam-se dois eixos de rotação à máquina, permitindo a fabricação de geometrias complexas, através da utilização de trajetórias controladas numericamente (CNC); são as chamadas máquinas de 5 eixos.

As operações de fresamento podem ser divididas em duas categorias, dependendo do avanço e do sentido de rotação da ferramenta. Considerando a peça de trabalho como fixa e todos os movimentos sendo realizados pela ferramenta, quando a componente da velocidade de corte na direção do avanço é positiva, o processo é chamado de fresamento concordante. Pelo contrário, quando esta componente é negativa, o processo é chamado de fresamento discordante. A Figura 8 torna esta distinção mais clara. A importância dessas duas classificações é que quando a fresagem é discordante a espessura do cavaco começa em zero e aumenta até que a ferramenta perca o contato com a peça. Inversamente, quando o fresamento é concordante a espessura do cavaco é máxima quando a ferramenta inicia o contato com a peça de trabalho e diminui para zero ele se encerra. O fresamento concordante é preferido na maioria dos casos porque nele o corte começa com a espessura máxima do cavaco, que é então reduzida à medida que este evolui, tendo

influência direta sobre o resultado (TÖNSHOFF; DENKENA, 2011).

Figura 8 – Fresamento concordante e discordante



Fonte – Tradução livre de Klocke e König (2008)

2.1.2 Desgaste de ferramentas

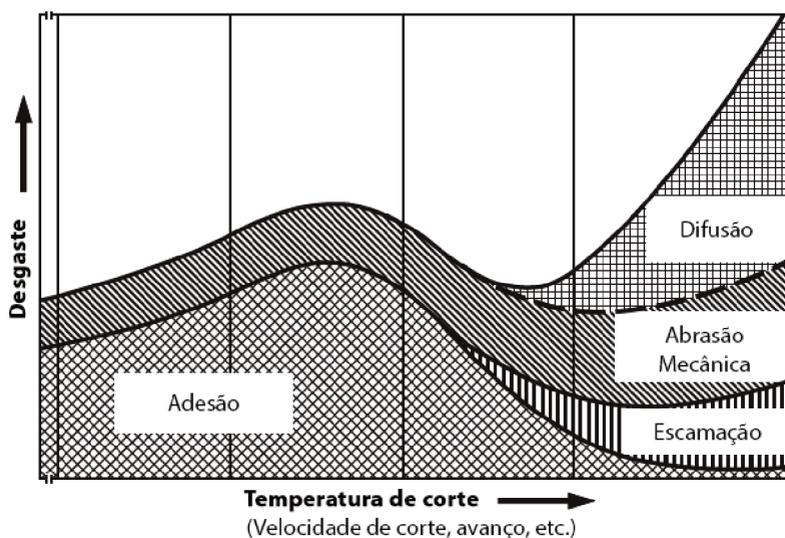
Conforme o processo de usinagem ocorre, devido a complexa combinação de esforços, ocorre desgaste da superfície da ferramenta. Com o uso, seu flanco e sua face perdem partículas até que a ferramenta atinja um estado em que deixe de ser adequada para uso. Nesta seção, será utilizada a abordagem trazida por Klocke e König (2008), complementada por outros autores onde indicado.

Mesmo que o desgaste da ferramenta não seja causado por um único mecanismo, mas por uma combinação deles (LI, 2012), é possível listar suas causas, as quais são citadas a seguir:

- Abrasão mecânica;
- Adesão;
- Solicitações mecânicas e térmicas;
- Reações triboquímicas

A Figura 9 mostra como a temperatura (e consequentemente a velocidade de corte e avanço, uma vez que estas são variáveis diretamente proporcionais) influencia qual o causa de desgaste que atuará no processo e sua magnitude. Nela, a ocorrência simultânea de diferentes causas de desgastes fica clara.

Figura 9 – Velocidade de corte e sua influência nos mecanismos de desgaste



Fonte – Tradução livre de Klocke e König (2008)

A abrasão pode ser vista como uma espécie de micro usinagem da ferramenta. Isso acontece em áreas rugosas da peça ou quando alguma partícula estranha ao processo fica presa entre o par peça/ferramenta e, devido ao movimento relativo, causam micro fissuras na superfície da ferramenta. É um desgaste cuja causa é puramente mecânica.

A adesão ocorre quando, através da formação de ligações atômicas, a peça e o material da ferramenta formam micro soldas, conhecidas como gume postiço, que são então arrancadas pelo deslizamento cons-

tante entre a ferramenta e a peça. Quando estas micro soldas são arrancadas, levam consigo partes da ferramenta, desgastando-a. O acúmulo de material (com quebra subsequente) pode acontecer tanto na cunha de corte como entre a face e a parte inferior do cavaco. O aumento na velocidade de corte leva a uma redução desses efeitos (TÖNSHOFF; DENKENA, 2011), pois aumenta a temperatura na região do processo. A propensão de uma ferramenta a sofrer desgaste por adesão só pode ser avaliada levando-se em consideração o material da ferramenta e da peça simultaneamente, já que a adesão é uma propriedade do par tribológico.

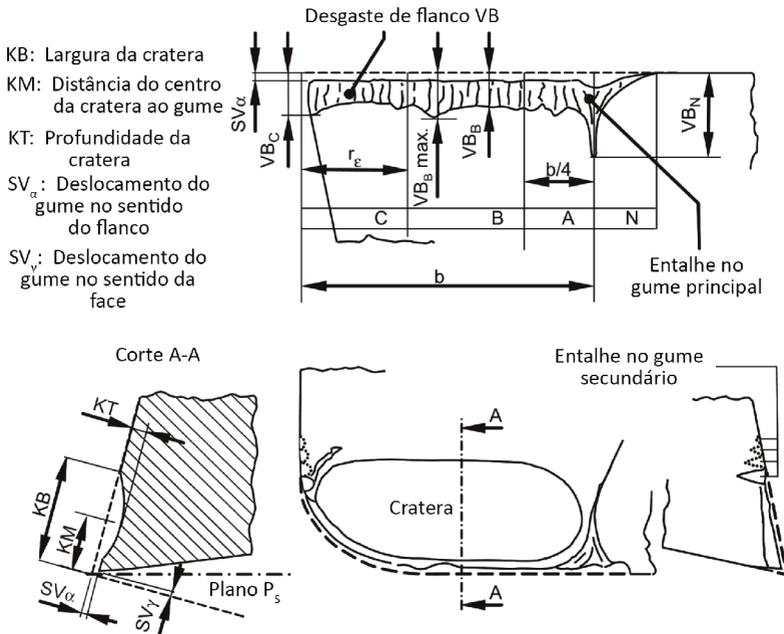
As grandes solicitações, tanto mecânicas como térmicas, são mais uma importante causa de desgaste nos processos de usinagem. Manifestam-se na ferramenta em forma de quebras, lascamentos, fissuras transversais e longitudinais e deformações plásticas. A forma exata dependerá tanto do processo ao qual a ferramenta é submetida como da própria ferramenta, em especial, quanto ao seu material.

As causas triboquímicas podem ser divididas entre difusão e oxidação. A difusão é um processo físico-químico, dependente da temperatura, onde as moléculas e/ou átomos da peça se movem para dentro da ferramenta, sendo também possível que esse processo ocorra na direção contrária. O desgaste diretamente relacionado a esse mecanismo é bastante pequeno - embora ainda mensurável - e sua maior influência é nas ferramentas com presença de determinados elementos de liga. A difusão destes pode levar a uma grande redução na resistência ao desgaste da ferramenta. A oxidação também é dependente da temperatura e pode levar a um aumento ou redução da massa da ferramenta através de reações químicas entre a ferramenta e o ambiente em que o processo ocorre (atmosfera, fluido de corte, material da peça).

O tipo específico de desgaste que será verificado na ferramenta após o processo varia muito com a intensidade dos esforços aos quais está submetida assim como com sua duração. A Figura 10 ilustra esquematicamente as medidas geralmente utilizadas para o desgaste. Uma medida muito utilizada é o VB , desgaste de flanco. Por ocorrer de forma

variada no gume, pode ser subdividida em regiões e tratada de forma média, como ocorre com VB_B , ou como sendo o máximo da região VB_{max} . Também podem ser medidos o SV_α e o SV_γ , que indicam o quanto a ferramenta deve ser deslocada na direção, respectivamente, do flanco e da face para que o processo mantenha-se constante após o desgaste da ferramenta. Por fim, a cratera ocorre na face da ferramenta, sendo descrita pelas medidas de KB , KM e KT , sendo respectivamente, a largura da cratera, a distância do centro da cratera ao gume e a profundidade da cratera.

Figura 10 – Medição do desgaste de flanco VB



Fonte – Tradução livre de Klocke e König (2008)

2.1.3 Modelos de previsão do desgaste

A escolha dos parâmetros do processo tem uma influência direta não apenas sobre a peça resultante, mas também sobre o aspecto econômico do processo. Valores maiores de a_p , a_e e f , como mostrado pela taxa volumétrica de remoção de material Q_w na Equação 1, levam a um aumento na taxa de remoção de material, que por sua vez leva a custos de processamento menores. Este aumento de produtividade, no entanto, vem com a desvantagem de aumentar as forças do processo. Isso leva a um maior desgaste da ferramenta e a um aumento nos custos de troca da ferramenta, não apenas com os custos das ferramentas em si, mas também com os custos de tempo de máquina parada. Um outro parâmetro importante para definir-se a produtividade do processo é a velocidade de corte, v_c , que tem ligação direta com a taxa de desgaste da ferramenta.

$$Q_w = a_p a_e v_f \quad (1)$$

Os parâmetros de corte recomendados pelos fabricantes de ferramentas de corte referem-se a uma vida útil da ferramenta de 15 minutos. O uso de parâmetros de corte menos intensos pode levar a uma vida mais longa da ferramenta, como aconteceu no experimento de Silva *et al.* (2017), onde o controle da temperatura da interface ferramenta-cavaco levou a um menor desgaste da ferramenta em troca de um maior tempo de corte. Uma taxa de desgaste reduzida da ferramenta significa trocas menos frequentes, com a desvantagem de causar uma menor taxa de produção e aumento dos custos salariais e de máquina. Em situações em que os salários e as máquinas são caros em comparação com as ferramentas, vale a pena aumentar a intensidade dos parâmetros de corte, resultando em uma maior taxa de produção. Entretanto, a diminuição da intensidade do processo mostra-se relevante quando a ferramenta é cara (KLOCKE; KÖNIG, 2008).

Os modelos para previsão de desgaste de ferramentas podem ser divididos em cinco categorias, de acordo com o tipo de abordagem

utilizada: (1) modelos empíricos; (2) modelos analíticos; (3) modelos numéricos; (4) modelos baseados em inteligência artificial; (5) modelos híbridos (ARRAZOLA *et al.*, 2013).

Modelos empíricos são os mais simples dentre os modelos de previsão de desgaste. A partir da realização de ensaios, ajustam curvas utilizando métodos estatísticos, sendo restritos à faixa dos experimentos realizados e não suportados pelos fundamentos físicos do processo (ARRAZOLA *et al.*, 2013).

Modelos analíticos, por outro lado, partem dos princípios físicos para prever condições como forças de corte, geometria do cavaco, tensões, deformações e temperaturas. Apesar de fornecerem informações precisas, qualquer alteração nas condições iniciais do problema exige que uma nova modelagem seja feita (ARRAZOLA *et al.*, 2013). As referências (USUI *et al.*, 1984), (HUANG; DAWSON, 2005) e (PÁLMAI, 2013) são exemplos de aplicações deste tipo de modelo. O trabalho publicado por Johansson *et al.* (2017) traz um comparativo entre diferentes modelos empíricos, aplicados num conjunto de 11 pares de ferramenta-material.

Ambos os modelos numéricos e os baseados em inteligência artificial dependem do uso de computadores para que sejam criados. No caso dos numéricos, utilizam métodos computacionais, como o método dos elementos finitos (FEM), para calcular as saídas do processo. Os modelos baseados em inteligências artificiais, por outro lado, utilizam modelos como redes neurais para encontrar estas saídas. Modelos híbridos, por fim, funcionam combinando os outros modelos existentes (ARRAZOLA *et al.*, 2013). Bencheikh *et al.* (2017), Ramírez P. *et al.* (2017), Stenberg *et al.* (2017) e Salvatore *et al.* (2013) mostram exemplos de modelos numéricos. Modelos baseados em inteligências artificiais podem ser vistos nas referências (ATTANASIO *et al.*, 2013) e (MIKOŁAJCZYK *et al.*, 2018), as quais utilizaram redes neurais artificiais para modelar o desgaste em diferentes casos.

2.1.3.1 Modelo de Taylor

O mais clássico modelo de previsão de desgaste é o modelo de Taylor, tendo sido a primeira tentativa em modelar-se a vida de uma ferramenta. Foi publicado no início do século XX e é mostrado na equação 2,

$$T_c = C_v v_c^k \quad (2)$$

sendo usada para relacionar o tempo de vida de uma ferramenta com a velocidade de corte usada pelo uso de duas constantes C_v e C_T . O parâmetro C_v representa o tempo de vida de uma ferramenta usada com uma velocidade de corte $v_c = 1$ m/min, enquanto C_T dá a velocidade de corte quando $T = 1$ min. O parâmetro k é calculado a partir de C_v e C_T , como mostrado na Equação 3 (KLOCKE; KÖNIG, 2008).

$$k = -\frac{\log C_v}{\log C_T} \quad (3)$$

A grande vantagem da equação de Taylor é que por ser representada por uma reta em um gráfico logarítmico, seus parâmetros podem ser determinados com apenas dois experimentos. É importante notar, no entanto, que estes parâmetros são constantes apenas em uma região restrita, tendo aplicabilidade limitada. Além disso, os experimentos necessários para obtenção destes parâmetros são custosos, tanto por serem destrutivos para as ferramentas ensaiadas, como por sua duração. Em um experimento realizado por Diniz *et al.* (2003), por exemplo, a usinagem de um aço AISI 52100 a 60 HRC com ferramenta de PCBN com $f = 0,08$ mm e $a_p = 0,3$ mm, foram necessários mais de 25 minutos para atingir-se o fim de vida com $v_c = 110$ m/min e mais de 14 minutos para $v_c = 175$ m/min, considerando-se $VB_b = 0,2$ mm. Neste período, tanto operador como máquina-ferramenta não estão produzindo produtos para a fábrica.

A equação de Taylor foi expandida para abranger os efeitos de outros parâmetros de corte, como profundidade de corte e avanço por

dente. Essa outra forma é mostrada na Equação 4.

$$T_c = C v_c^k f_z^{k_{fz}} a_p^{k_a} \quad (4)$$

2.2 APRENDIZADO DE MÁQUINA E REDES NEURAIIS

Atualmente, a disponibilidade de dados é maior do que nunca. A atividade das empresas gera dados secundários, relacionados ao negócio principal apenas de forma indireta. Uma grande quantidade de dispositivos é embarcada com um número cada vez maior de sensores, os *sites* são capazes de rastrear o comportamento do usuário e as empresas rastreiam clientes, fornecedores e informações de operações. Estes dados podem e devem ser usados como uma vantagem para estas empresas (MANYIKA *et al.*, 2011).

Os modelos de aprendizado de máquina utilizam dados com o objetivo de fazer previsões significativas com eles. Tipicamente, esses dados são divididos em preditores, que são os dados usados como entrada no modelo, e metas, que são seu resultado. Essas metas podem ainda ser divididas quanto ao seu tipo, podendo serem categóricas ou quantitativas, influenciando o tipo de modelo que será escolhido para fazer a previsão. Problemas cuja variável-alvo tem um tipo categórico são chamados problemas de classificação, enquanto problemas com uma variável-alvo quantitativa são chamados problemas de regressão (HASTIE *et al.*, 2009, p. 9). A presença das metas funciona como um professor para o modelo, indicando quais são seus resultados esperados e possibilitando seu ajuste. Tal tipo de modelo é chamado de modelo de aprendizado supervisionado (MACKAY, 2008).

Redes neurais são um modelo de aprendizado de máquina cuja “ideia central é extrair combinações lineares dos dados de entrada como recursos derivados e, em seguida, modelar o alvo como uma função não-linear desses recursos” (HASTIE *et al.*, 2009, p. 389). Começaram a ser desenvolvidas como uma tentativa de imitar a forma de processamento do cérebro humano, que é capaz de realizar computações paralelas

bem organizadas, em tempo real e com um baixo consumo de energia (PRIETO *et al.*, 2016).

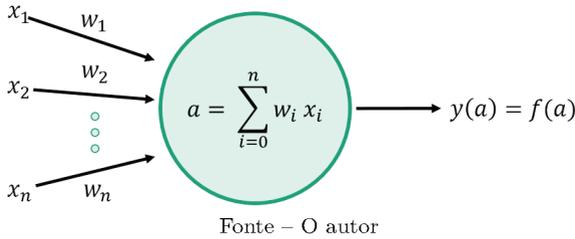
Elas não são um método novo de usar dados para fazer previsões. A história das redes neurais pode ser dividida em quatro períodos de duas décadas cada, começando na década de 1940. Durante as duas primeiras décadas (1940 e 1950) foram desenvolvidos modelos e regras de aprendizagem para neurônios individuais (que serão vistos a seguir). No período seguinte (anos 1960 e 1970), as regras de aprendizagem para redes de camada única e a aplicação de técnicas de mecânica estatística para redes recorrentes foram o estado da arte. Durante esse período, um livro que discutia as limitações de uma camada de perceptrons foi publicado por Minsky e Papert (1972), levando a uma redução considerável no desenvolvimento de redes neurais (PRIETO *et al.*, 2016). Durante as décadas de 1980 e 1990, a profundidade das redes foi aumentada. Também surgiram as aplicações dos métodos bayesianos e dos processos gaussianos. No período atual, que começou na década de 2000, o foco tem sido na melhoria dos modelos anteriores pelo uso da análise de convergência, equilíbrio estatístico, estabilidade e assim por diante (PRIETO *et al.*, 2016).

2.2.1 A rede de neurônio único

O tipo mais simples de rede neural é aquele com um único neurônio como é ilustrado na Figura 11 (MACKAY, 2008). Nestas redes um vetor de *features* de entradas X , com elementos x_i ($i = 1, \dots, j$), é dado à rede e um vetor de saída \hat{Y} é retornado como resultado. Cada uma das entradas está associada a um valor de peso w_i ($i = 1, \dots, j$). Na maioria dos casos há também um termo de viés (*bias*) w_0 , que está sempre associado a um $x_0 = 1$ fixo. A atividade a do neurônio é então calculada como na Equação 5.

$$a = \sum_{i=1}^n w_i x_i \quad (5)$$

Figura 11 – Perceptron de neurônio único



Uma vez calculada a atividade, a saída \hat{Y} é calculada como uma função da atividade, $y = f(a)$, por meio do que é conhecido como uma função de ativação. As equações 6 até 11 mostram algumas funções de ativação usuais. A mais simples é a função linear, que simplesmente produz uma combinação linear das características de entrada (HASTIE *et al.*, 2009, p. 392). Funções de limiar, como mostrado na Equação 7, são a maneira mais simples de introduzir não-linearidades na rede.

As funções sigmóides são de interesse especialmente quando se trabalha com problemas de classificação. Unidades lineares retificadas (RELU), são funções como mostradas na Equação 8, sendo capazes de introduzir não-linearidades no modelo (HINTON *et al.*, 2014) assim como de aumentar sua esparsidade, já que cada entrada menor que zero que recebem é transformada zero. As unidades lineares exponenciais escaladas (SELU) são mostradas na Equação 11. Esta classe de funções, quando usada com $\lambda \approx 1,0507$ e $\alpha \approx 1,6733$, é capaz de transformar a rede em uma rede auto-normalizante SNN, que permite arquiteturas mais profundas quando comparada com redes *feedforward* (sem realimentação) (KLAMBAUER *et al.*, 2017). Quando $\lambda \approx 1$, a Equação 11 é conhecida simplesmente como Elu ou unidade exponencial-linear e α torna-se o único parâmetro que precisa ser ajustado. Ele é normalmente tomado como 1. A Figura 12 mostra estas funções de forma gráfica.

- Linear

$$y(a) = a \tag{6}$$

- Limiar

$$y(a) = \begin{cases} -1, & \text{se } a \leq 1 \\ 1, & \text{em outros casos} \end{cases} \quad (7)$$

- Sigmoidoide

$$y(a) = \frac{1}{1 + e^{-a}} \quad (8)$$

- Relu

$$y(a) = \begin{cases} 0, & \text{se } a \leq 1 \\ a, & \text{em outros casos} \end{cases} \quad (9)$$

- Elu

$$y(a) = \begin{cases} \alpha(e^a - 1), & \text{se } a \leq 1 \\ a, & \text{em outros casos} \end{cases} \quad (10)$$

- Selu

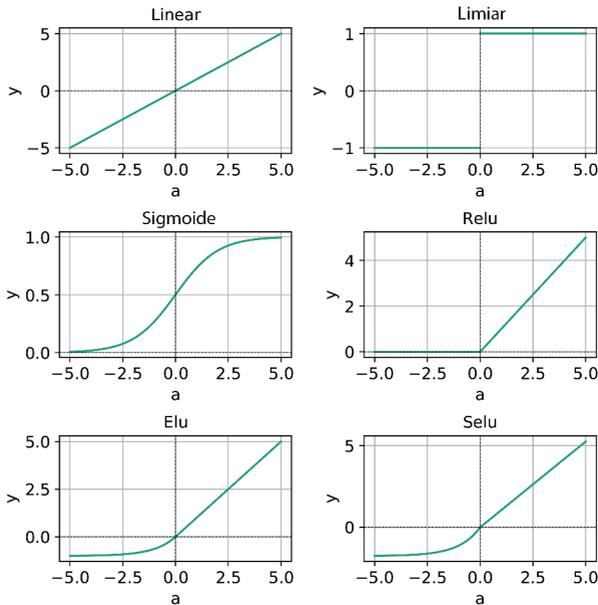
$$y(a) = \lambda \begin{cases} \alpha(e^a - 1), & \text{se } a \leq 1 \\ a, & \text{em outros casos} \end{cases} \quad (11)$$

Durante a fase de treinamento do modelo, as metas Y são conhecidas. Desta forma, é possível, comparando os valores reais com os valores fornecidos pela rede, \hat{Y} , minimizar uma função de erro L (também conhecida como função de perda) e encontrar um conjunto ótimo de pesos W . Em geral, a função de perda usada é a soma do erro quadrático médio, Equação 12, mas funções como a soma do erro absoluto médio, Equação 13, também podem ser usadas.

$$L_{\text{MSE}} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (12)$$

$$L_{\text{MAE}} = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (13)$$

Figura 12 – Representação gráfica das funções de ativação



Fonte – O Autor

2.2.2 Métodos de otimização, *overfitting* e regularização

Uma vez definida a função objetivo do modelo, também conhecida como função de perda, deve-se encontrar o conjunto de pesos W que a leva a um ponto de mínimo. É importante notar que o ponto de mínimo global desta função não necessariamente leva ao modelo ideal, pois este geralmente é um ponto que sofre de *overfitting*. A solução neste caso é adicionar um termo de regularização à função de perda (HASTIE *et al.*, 2009).

O procedimento para determinação dos pesos do modelo é comumente chamado de treinamento do modelo, e é onde ele efetivamente aprenderá as nuances do problema que permitir-lhe-ão realizar as previsões. Nesta etapa, são aplicados métodos de otimização para minimizar a função objetivo do problema. É importante notar que os algoritmos

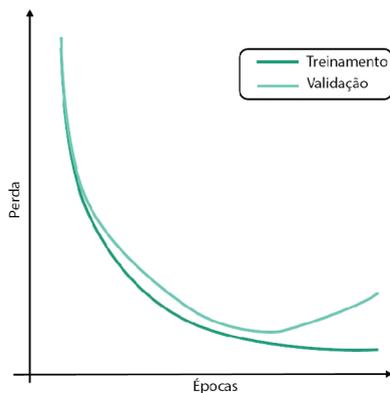
de otimização aplicados em aprendizado de máquina são diferentes dos tradicionais pois neles, como explicado por Goodfellow *et al.* (2016), busca-se otimizar uma métrica diferente da função de perdas na qual o algoritmo está rodando. Desta forma, acontece uma otimização indireta da métrica.

O algoritmo mais frequentemente utilizado no treinamento de redes neurais é o SGD (Gradiente Estocástico Descendente), cujo funcionamento é explicado em maiores detalhes por Goodfellow *et al.* (2016). Também existem variações deste algoritmo, por exemplo a adição de um termo relacionado ao *momentum* do gradiente, feita por Polyak (1964) e sua evolução, conhecida como momentum de Nesterov (SUTSKEVER *et al.*, 2013). A decisão crítica ao aplicar-se o SGD é a escolha da taxa de aprendizado a ser utilizada. Ela influencia no quão rápido o treinamento convergirá (ou se falhará, divergindo). Apesar de ser crítica, não existe um método definido para sua determinação, sendo em geral definida por tentativa e erro (GOODFELLOW *et al.*, 2016).

Um algoritmo mais recente e altamente eficiente para otimização baseada no gradiente estocástico é o *Adam*. Mais informações sobre ele podem ser em seu artigo original (KINGMA; BA, 2014).

Uma rede neural perfeitamente treinada produzirá valores de y que são iguais as metas dadas t (MACKAY, 2008). Isso, no entanto, leva ao *overfitting*. *Overfitting* acontece quando um modelo de aprendizagem de máquina é capaz de alcançar valores de perda muito baixos para os dados de treinamento, mas quando expostos a dados de teste esses valores de perda começam a aumentar como mostrado na Figura 13. Isso acontece porque ao treinar a rede, um banco com quantidade limitada de dados é usado. Estes dados possuem ruído de amostragem que não está presente em dados reais, mesmo se retirados da mesma distribuição e este ruído acaba sendo aprendido pela rede (HINTON *et al.*, 2014).

Figura 13 – Divergência entre as perdas de treinamento e validação



Fonte – O Autor

A maneira mais simples de lidar com o *overfitting* é aplicando uma interrupção precoce do processo de otimização. Como os pesos iniciais são altamente regularizados, esta técnica reduz o modelo a um modelo linear (HASTIE *et al.*, 2009, p. 398). Outra técnica utilizável é o decaimento do peso, também conhecido como regularização. Ela funciona adicionando um termo de penalidade à função de perda, que se torna a equação 14.

$$R(\theta) = L(\theta) + \lambda J(\theta) \quad (14)$$

λ é um hiper-parâmetro, significando que ele deve ser definido antes de treinar o modelo e não é aprendido. $\lambda J(\theta)$ é o termo de regularização, podendo tomar diferentes formas, amplamente discutidas por Goodfellow *et al.* (2016).

Ainda com relação à prevenção contra o *overfitting*, uma família de técnicas que pode ser utilizada é a da validação cruzada (muitas vezes referida pelo seu nome em inglês, *cross-validation*). Uma das técnicas desta família é a do *K-fold cross validation*, na qual o conjunto de dados de treinamento é dividido em K subconjuntos (ou *folds*). Destes subconjuntos, $K - 1$ serão então utilizados para fazer o treinamento e o

restante será utilizado para cálculo da métrica desejada. Este resultado é então armazenado e repete-se o processo para os subconjuntos restantes. Este processo é de importância por, conforme explicado por Hastie *et al.* (2009, p. 241), permitir que apenas um dataset seja utilizado tanto para treinamento como validação, ou seja, permite a utilização de menos dados.

3 METODOLOGIA

Neste capítulo serão apresentadas questões relativas aos métodos utilizados para elaboração deste trabalho. Primeiro é feita uma classificação quanto ao tipo de pesquisa adotado e, então, são abordados os passos tomados durante o desenvolvimento propriamente dito.

3.1 TIPO DE PESQUISA

Conforme a classificação de Gil (2014), o presente trabalho é classificado como sendo de natureza aplicada, utilizando uma ferramenta de aprendizado de máquina para a solução de um problema específico. Desta forma, possui uma abordagem qualitativa, com um propósito explicativo. Ademais, conforme Miguel (coord.) (2012), é um trabalho com uma abordagem metodológica do tipo simulação.

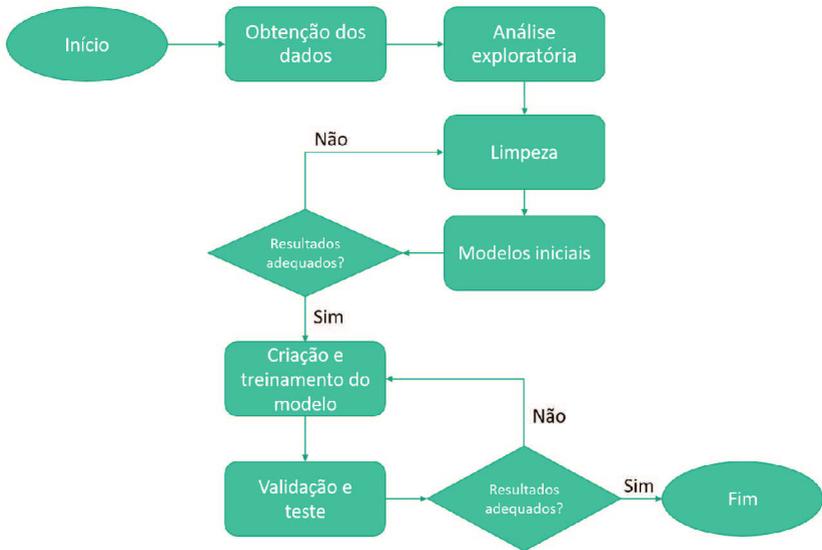
3.2 ETAPAS METODOLÓGICAS

A Figura 14 mostra as etapas utilizadas no trabalho. Seguiu-se uma abordagem bastante tradicional da ciência de dados, onde iniciou-se pela obtenção, análise exploratória e tratamento dos dados do trabalho. Então foram montados modelos preliminares, de tamanho reduzido, para verificar sua viabilidade. Uma vez verificada sua viabilidade, passou-se para o modelo completo, utilizando todos os dados disponíveis simultaneamente. Esta é uma etapa iterativa, que se encerra quando os resultados obtidos são considerados adequados.

O processo utilizado em cada uma das etapas metodológicas é apresentado de forma mais detalhada no Capítulo 4.

Para execução do trabalho, foi utilizada a linguagem Python, em sua versão 3.6. As principais bibliotecas utilizadas no tratamento dos dados e criação dos modelos foram a scikit-learn 0.22.1 (PEDREGOSA *et al.*, 2011) e keras 2.2.5 (CHOLLET *et al.*, 2015). Fez-se também, extensivo uso da biblioteca Matplotlib 3.1.1 (HUNTER, 2007) para criação dos gráficos. Além disso, os modelos foram treinados utilizando

Figura 14 – Fluxograma do trabalho realizado



Fonte – O Autor

um processador i5-7300U a 2,6 GHz e 8Gb de memória RAM. Não foi utilizada placa gráfica.

4 DESENVOLVIMENTO

Neste capítulo são detalhadas as etapas metodológicas apresentadas na Seção 3.2.

4.1 OBTENÇÃO E TRATAMENTO INICIAL DOS DADOS

O trabalho iniciou-se através de uma análise exploratória de um banco de dados de desgaste de ferramentas, contendo dados de experimentos com fresas de topo esférico. Como este banco de dados é resultado de uma revisão bibliográfica cujo objetivo inicial não era o de realizar este trabalho, o primeiro passo foi avaliá-lo, certificando-se de que dados relevantes estavam disponíveis. A fonte deste dados é apresentada na lista a seguir, na mesma ordem em que foram nomeados.

1. Sheng (2012)
2. Shafiq Tahir *et al.* (2014)
3. Pawel Twardowski *et al.* (2015)
4. Vopát *et al.* (2015)
5. Zhang *et al.* (2014)
6. Klocke *et al.* (2011)
7. Dados não publicados do Fraunhofer IPT
8. Stolorz (2017)
9. Tangjitsitcharoen *et al.* (2015)
10. Pu e Singh (2013)
11. Hood *et al.* (2013)
12. Wojciechowski e Paweł Twardowski (2012)
13. Uchikata *et al.* (2017)

14. Gopalsamy *et al.* (2010)

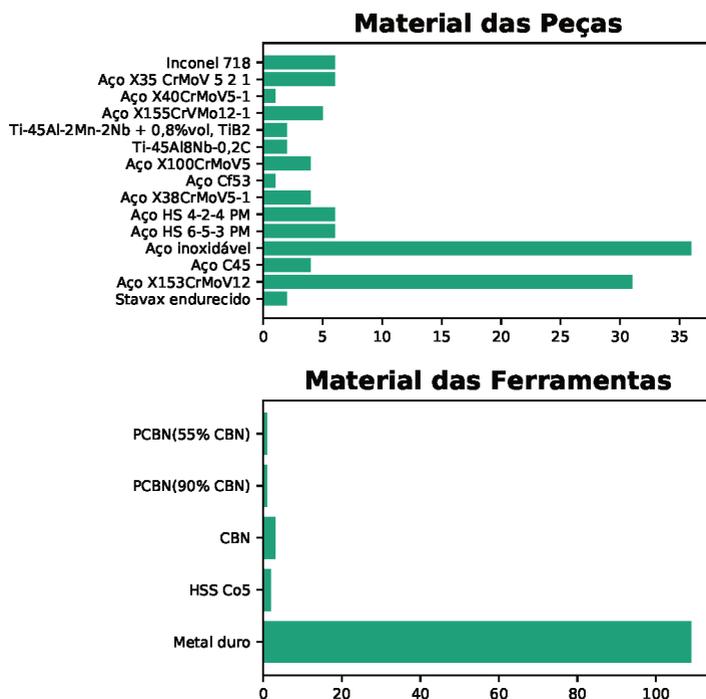
15. Sharman *et al.* (2001)

O primeiro gráfico criado, mostrado na Figura 15 tem como objetivo avaliar os tipos de materiais, tanto de peças quanto de ferramentas, que estavam disponíveis. A menos que uma grande quantidade de dados esteja disponível para cada um dos diferentes tipos de materiais, não faz sentido incluir uma grande variedade deles na rede neural. Além disso, quanto maior a variabilidade dos dados, maior é a dificuldade em criar-se um modelo capaz de realizar previsões adequadas.

Agrupando os diferentes materiais de peças em classes (de forma bastante abrangente, agrupando, por exemplo, os diferentes tipos de aços), como na Figura 16, vê-se que a maioria dos dados disponíveis é composta por diferentes tipos de aços, enquanto stavax endurecido, titânio e Inconel, foram utilizados apenas em uma quantidade reduzida de experimentos. Além disso, estes materiais possuem propriedades de usinagem muito diferentes das do aço. Por estas razões, foi decidido que apenas os aços fariam parte do modelo, fazendo com que os artigos 1, 11 e 15 fossem removidos. Da mesma forma, para os materiais das ferramentas, havia uma grande quantidade de dados sobre uma única classe de material (metal duro), enquanto os restantes não apenas eram poucos, como tinham propriedades muito diferentes (CBN e HSS). Vale lembrar ainda, que mesmo materiais pertencentes a uma mesma classe de material podem possuir características de usinabilidade bastante diferentes entre si, no entanto, escolheu-se por manter os aços de forma geral, ainda que saiba-se ser uma simplificação excessiva.

O próximo passo foi traçar a curva de desgaste em função do comprimento cortado para todos os artigos e experimentos contendo esses valores. Nesta fase, o artigo 2 foi cortado por não conter dados de comprimento de corte. Este gráfico é mostrado na Figura 17 e tem como objetivo avaliar se a forma das curvas obtidas com os dados disponíveis são compatíveis com as expectativas estabelecidas pela literatura, que estabelece haver um rápido crescimento do desgaste nos momentos

Figura 15 – Distribuição dos materiais das peças e ferramentas antes da seleção dos artigos com os quais o trabalho será realizado

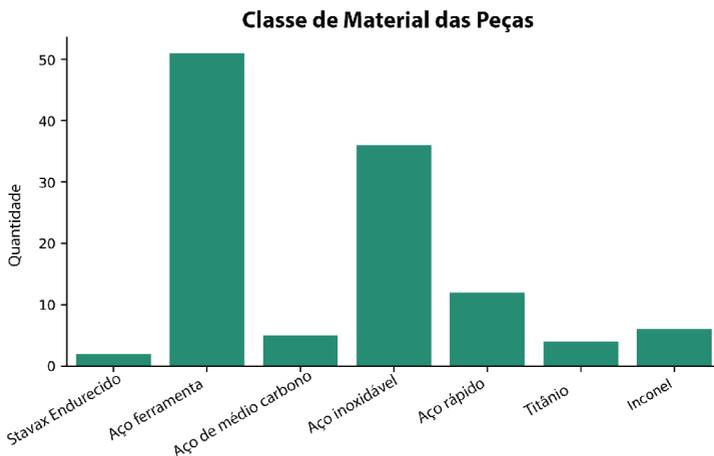


Fonte – O Autor

iniciais, uma fase de desgaste estável e uma segunda fase de desgaste elevado próximo ao fim da vida da ferramenta (JUNEJA *et al.*, 2003). Um segundo uso para este gráfico é avaliar a quantidade de dados disponíveis em cada experimento. Se um experimento possui uma quantidade insuficiente de medidas, ele não pode ser usado como *input* para o treinamento da rede neural.

De imediato, percebe-se que os artigos 5 e 13 têm uma quantidade

Figura 16 – Classe de material da peças

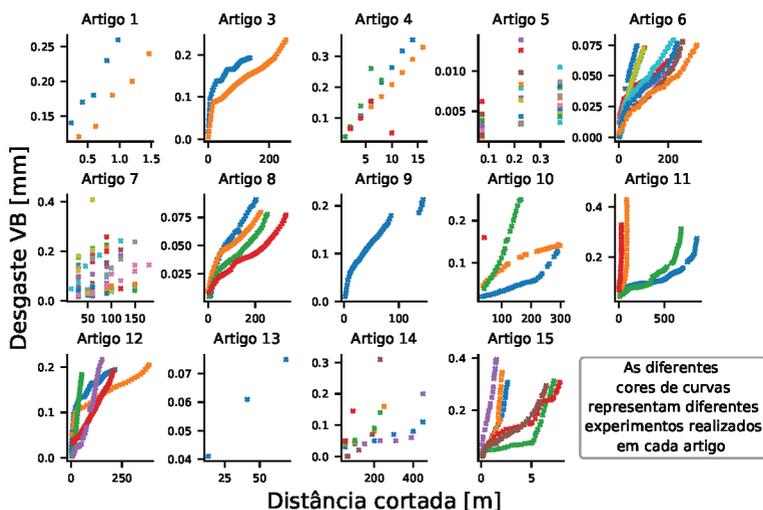


Fonte – O Autor

muito pequena de dados, com medições de desgaste realizadas em apenas três comprimentos de corte. Um segundo ponto que chama a atenção é a diferente escala presente em alguns dos experimentos. O artigo 5, por exemplo, foi realizado com valores muito pequenos de comprimento de corte e, conseqüentemente, mostra valores de desgaste muito pequenos.

De modo similar ao que acontece com o artigo 5, no artigo 1 os comprimentos de corte eram pequenos, mas apresentavam valores de desgaste mais elevados. Um terceiro artigo que chama a atenção é o número 11. Ele contém tanto experimentos com uma taxa muito alta de desgaste (quase 0,4 mm com cerca de 100 m de comprimento de corte) como experimentos onde essa taxa é muito pequena (0,2 mm de desgaste só é atingida após quase 800 m de comprimento de corte). Revisando o banco de dados, é possível identificar que os artigos 1 e 11 estavam entre os que executaram experimentos em materiais que são de difícil usinagem, Stavax Endurecido e titânio, TiAlN, respectivamente, o que significa que eles estão entre os experimentos que serão removidos do banco de dados de qualquer forma.

Figura 17 – Desgaste (VB) da ferramenta em função do comprimento cortado para cada um dos artigos



Fonte – O Autor

Ainda, adiantando o procedimentos descrito em mais detalhes na subseção 4.2, ao iniciar-se os procedimentos de simulação, mais três artigos foram removidos da base de dados. Como para recriar os experimentos todos os parâmetros de corte são necessários, os artigos 3, 9 e 12 foram removidos por não conterem dados de a_e . Aqui, o artigo 10 também teve que ser removido, mas por ser o único artigo que utilizou a ferramenta inclinada em duas direções diferentes simultaneamente.

Neste ponto, todos os dados que serão utilizados no treinamento da rede neural já estão selecionados. Os critérios utilizados para remoção ou não de determinado artigo são resumidos pela Tabela 1. Todas as ferramentas utilizadas para os experimentos restantes eram de carbetto de tungstênio, enquanto os materiais das peças e em quantos experimentos foram utilizados é mostrado na Figura 18.

Tabela 1 – Resumo dos critérios de seleção dos artigos utilizados. Em negrito são destacados os artigos que foram utilizados no trabalho.

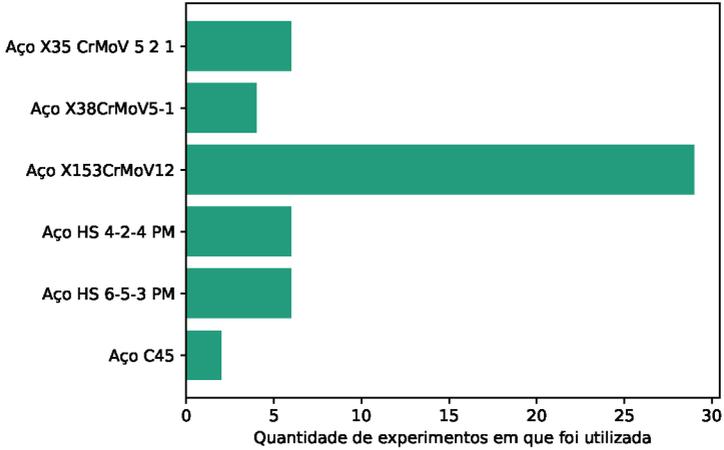
Artigo	Critério			
	Material da peça incompatível	Falta de dados	Dados Insuficientes	Ferramenta Inclinada
1	X			
2		X		
3		X		
4				
5			X	
6				
7				
8				
9				
10				X
11	X			
12				
13			X	
14				
15	X			

4.2 RECRIAÇÃO DIGITAL DOS ARTIGOS

Uma vez que os primeiros passos de limpeza foram tomados, o próximo foi recriar esses experimentos no *software* internamente desenvolvido pelo Fraunhofer IPT. Os dados de entrada para estas simulações são três arquivos, um contendo o caminho da ferramenta que deve ser simulado, um segundo contendo a modelagem da peça bruta utilizada na simulação e um terceiro contendo as configurações de cada simulação.

Para padronizar as simulações, a peça bruta foi definida como sendo um bloco com comprimento $l = 250$ mm e largura variável w , de acordo com o comprimento de corte CL e o a_e de cada experimento. O percurso da ferramenta foi definido como cortes retos paralelos ao comprimento l do bloco. Desta forma, o número necessário de cortes

Figura 18 – Materiais das peças considerados no modelo



Fonte – O Autor

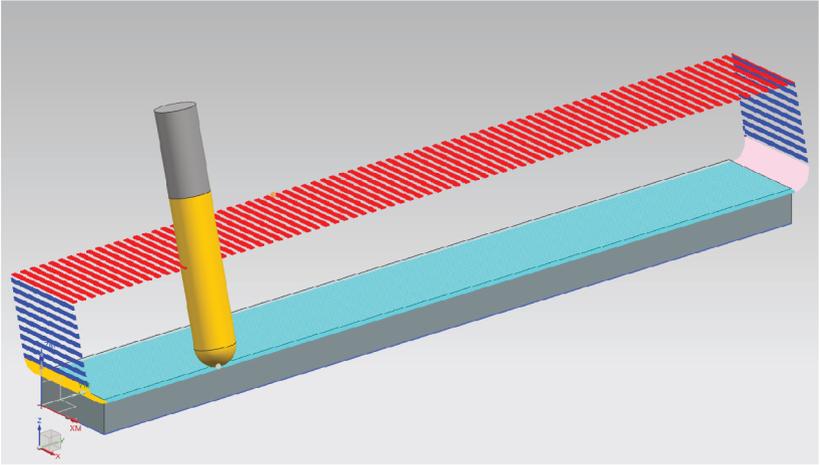
N_{caminhos} é dado pela Equação 15. Conseqüentemente, a largura necessária da peça é dada pela Equação 16. A altura h_0 do bloco base foi arbitrariamente definida como 10 mm e para fins de simulação, a peça bruta foi definida como tendo $h_1 = (10 + a_p)$ mm de altura. Isto é feito porque, ao gerar o percurso da ferramenta o modelo deve ter as dimensões finais da peça, porém a peça bruta é usada como entrada pelo *software*. Ao definir a altura da peça bruta como h_1 a profundidade de corte a_p é automaticamente definida. A Figura 19 exemplifica os modelos utilizados.

$$N_{\text{caminhos}} = \frac{CL}{l} \quad (15)$$

$$W = N_{\text{caminhos}} a_e \quad (16)$$

Depois de ter todos os modelos de peça e arquivos de trajetória ainda é necessário calcular as rotações por minuto do eixo-árvore n e as taxas de avanço v_f , que são dadas pelas equações 17 e 18 respecti-

Figura 19 – Exemplo da modelagem com o caminho da ferramenta



Fonte – O Autor

vamente. Depois de definidos todos os dados de entrada, a simulação pode ser executada.

$$n = \frac{1000 v_c}{\pi D} \quad (17)$$

$$v_f = f_z n z \quad (18)$$

A simulação executada retorna, por padrão, as dimensões do cavaco para cada uma das interações entre a ferramenta e a peça. Isto leva, no entanto, a uma enorme quantidade de dados resultantes. Como forma de contornar esta questão, foi decidido que as condições de cavaco só seriam exportadas para cada 1000 interações. Cada um desses resultados é então salvo em um arquivo .csv individual, cujos dados são exemplificados na Tabela 2. A coluna de fatia representa a posição axial da ferramenta que interagiu com a peça e conseqüentemente gerou o cavaco. Cada uma destas fatias tem uma espessura igual a 1/10 do diâmetro da ferramenta. As demais colunas representam a geometria do cavaco; δ_ϕ é o ângulo do arco no qual ocorreu a interação,

hsp é a profundidade do cavaco, bsp sua altura e lsp sua largura. Na prática, as linhas onde todos os dados são nulos representam regiões da ferramenta que não entraram em contato com a peça, de forma a não gerar nem cavaco nem sofrer desgaste.

Tabela 2 – Exemplo do resultado exportado para cada interação

fatia	delta_phi	hsp_max	hsp_avg	bsp	lsp
0	0	0	0	0	0
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	7,08	0,00772	0,00067	0,32025	0,34824
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0

Fonte – O Autor

Como a simulação não produz o desgaste causado pela interação, os dados de desgaste disponíveis no banco de dados original foram utilizados para estimá-lo. Aqui, no entanto, dois problemas surgiram.

O primeiro deles foi que o banco de dados original continha algumas medidas de desgaste em função do comprimento de corte, mas em intervalos diferentes do simulados. A solução foi usar uma regressão para determinar o desgaste nesses comprimentos de corte intermediários. Primeiro, uma regressão linear foi testada, mas esta não é confiável, pois descarta as não-linearidades presentes no processo de desgaste (Figura 17). Posteriormente foi aplicada uma regressão polinomial de terceiro grau para tentar captar este comportamento.

O segundo problema foi que os resultados da simulação têm 10 fatias para cada interação, ou seja, o desgaste em 10 posições diferentes da ferramenta, mas a base de dados contendo os valores de desgaste tem apenas uma. Nos casos em que uma única fatia interage com a peça de

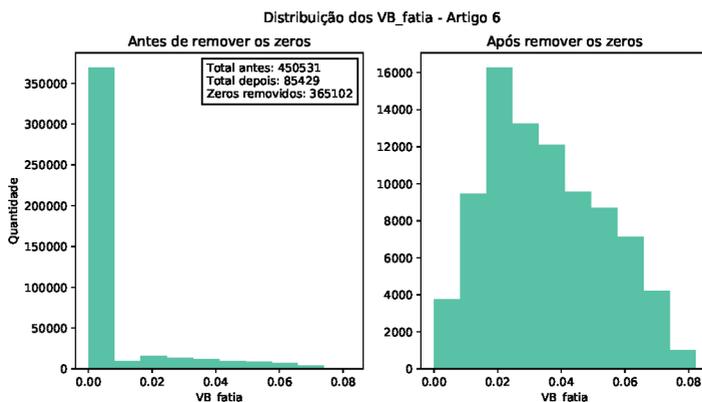
trabalho (como na Tabela 2), isso não é um problema, mas em alguns casos 2, 3 ou mesmo 4 fatias interagem com a peça simultaneamente. A solução utilizada foi dimensionar o desgaste entre as fatias, utilizando a largura máxima do cavaco de cada fatia, que, entre os resultados da simulação, tem a maior correlação com as forças de corte e, conseqüentemente, com o desgaste. A fatia com maior largura de cavaco recebe o desgaste total e as demais fatias recebem um percentual igual à proporção entre sua própria largura máxima e a largura máxima do cavaco naquela interação, conforme mostrado na Equação 19.

$$VB_{\text{fatia}} = \left(\frac{hsp_{\text{max fatia}}}{\max_{\text{interações}} hsp_{\text{max}}} \right) VB_{\text{fatia}} \quad (19)$$

Uma vez que o desgaste foi adicionado aos resultados da simulação, a velocidade de corte e a dureza do material da peça foram adicionadas a este novo banco de dados e os dados para treinamento da rede neural aparentavam estar prontos, no entanto, ao verificar-se sua distribuição ficou evidente que uma grande quantidade destes dados era na realidade nula, gerando um forte desbalanceamento, como exemplificado na Figura 20 para o artigo 06. Estes dados nulos representam momentos do processo em que nenhum corte acontece ou partes da ferramenta que não estão em contato direto com a peça de trabalho e, conseqüentemente não sofrem desgaste. Por representarem uma situação específica, que não influencia no processo de desgaste, foram removidos. Desta forma, a distribuição de desgastes para cada artigo é mostrada na Figura 21.

A base de dados, que foi utilizada para a formação da rede, contém cinco características e um alvo. As características são constituídas pelos três parâmetros geométricos do cavaco gerados na simulação com o *software* do Fraunhofer IPT, os quais ainda não foram utilizados, bem como a dureza do material usinado e o número de interações entre cada um dos gumes da ferramenta e a peça. Os outros parâmetros de corte, como profundidade e largura de corte, avanço por dente e assim por diante foram deixados de fora, já que a geometria do cavaco gerada na simulação depende deles.

Figura 20 – Distribuição dos dados em cada faixa de desgaste para o artigo 06 antes e depois da remoção dos valores nulos



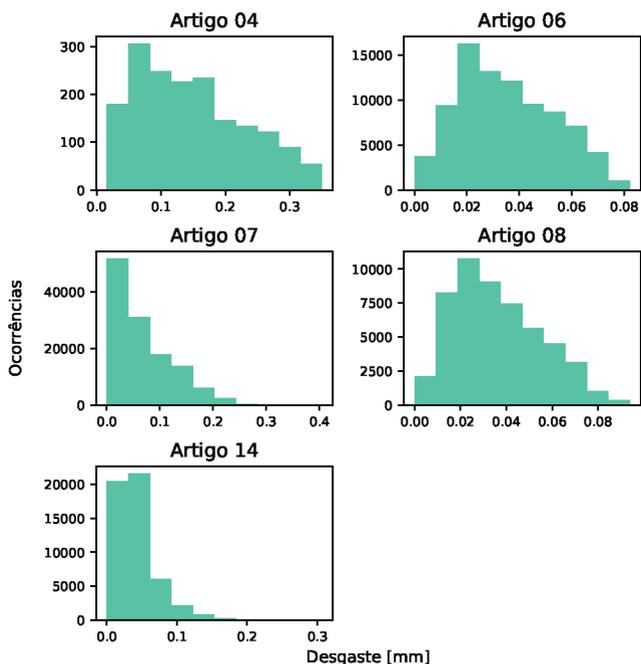
Fonte – O Autor

4.3 MODELOS PRELIMINARES

Antes de iniciarem-se os treinamentos com todo o banco de dados, foram realizados testes preliminares com versões reduzidas. Como os dados disponíveis são provenientes de diferentes artigos, usou-se justamente esta característica como critério de redução dos bancos de dados. Desta forma, foram criados cinco novos conjuntos, os quais foram usados para treinar redes neurais independentes. O objetivo destas redes é verificar a possibilidade de um modelo aprender as características de cada artigo quando isolado.

Para este treinamento, foi utilizada uma rede de arquitetura densa sem realimentação, com cinco camadas ocultas, contendo respectivamente 10/100/100/100/10 nós e cuja função de ativação foi uma Unidade linear exponencial (ELU). A camada de entrada utilizou uma função de ativação ELU e a de saída uma Unidade linear retificada (RELU). A função de perda utilizada foi o erro quadrático médio e a técnica de otimização utilizada foi a do gradiente descendente estocástico. A avaliação do desempenho de cada uma das redes treinadas foi

Figura 21 – Distribuição dos valores de desgaste dentro de cada artigo



Fonte – O Autor

feita utilizando-se validação cruzada de 5 partes, cuja saída é o valor de R^2 de cada treinamento. Por fim, a métrica final de cada rede é dada pela média dos 5 R^2 fornecidos pela validação cruzada. A topologia da rede utilizada é resumida na Tabela 3.

Neste ponto, faz-se relevante explicar a forma como o treinamento destas redes foi realizado. Os dados de treinamento são separados, aleatoriamente, em lotes (em geral referidos pelo termo em inglês, *batches*) nos quais o gradiente estocástico descendente (SGD) é aplicado. Após completado o primeiro *batch*, os parâmetros do modelo são atualizados e o SGD é aplicado ao *batch* seguinte. Este processo repete-se até que se esgotem os *batches*, definindo uma época. Esse processo repete-se até

Tabela 3 – Topologia da rede utilizada nos modelos preliminares

Parâmetro	Valor
Camadas Ocultas	5
Nós Ocultos	10/100/100/100/10
Nós de Entrada	5
Nós de Saída	1
Ativação da Camada de Entrada	elu
Ativação das Camadas Oculta	elu
Ativação da Camada de Saída	relu
Função de Perda	MSE
Otimização	SGD
Métrica	R^2

Fonte – O Autor

que o número de épocas definido para o treinamento seja atingido. No caso deste modelo, foram utilizados *batches* de 100 dados e 150 épocas.

Como mostrado na Tabela 4, todos os artigos testados apresentaram valores adequados de R^2 . Destaca-se que o valor obtido para o artigo 07, apesar de consideravelmente mais baixo que os outros, ainda é razoável. À parte da remoção dos valores nulos, já feita anteriormente, não foram utilizados outros métodos de balanceamento dos dados nesta etapa.

Tabela 4 – Média das métricas de validação cruzada de cada artigo para as redes preliminares

Artigo	MAE	R^2
04	0,0162	0,934
06	0,0050	0,838
07	0,0229	0,637
08	0,0045	0,937
14	0,0076	0,966

Fonte – O Autor

4.4 MODELOS GERAIS

Após verificar a capacidade do modelo em realizar previsões para cada um dos artigos utilizados como fonte de dados, partiu-se para a criação de uma rede neural geral, a qual seria treinada com dados provenientes de todos os artigos simultaneamente. Nas subseções a seguir serão apresentadas as duas abordagens utilizadas. Na primeira, foi utilizada uma estratégia muito similar à utilizada nos testes preliminares, buscando prever o desgaste esperado de forma numérica, no chamado modelo de regressão. Após a obtenção de resultados aquém dos desejados, partiu-se para uma segunda estratégia, onde a estimativa realizada era apenas da faixa em que este desgaste se encontraria.

4.4.1 Modelo de regressão

4.4.1.1 Modelo simplista

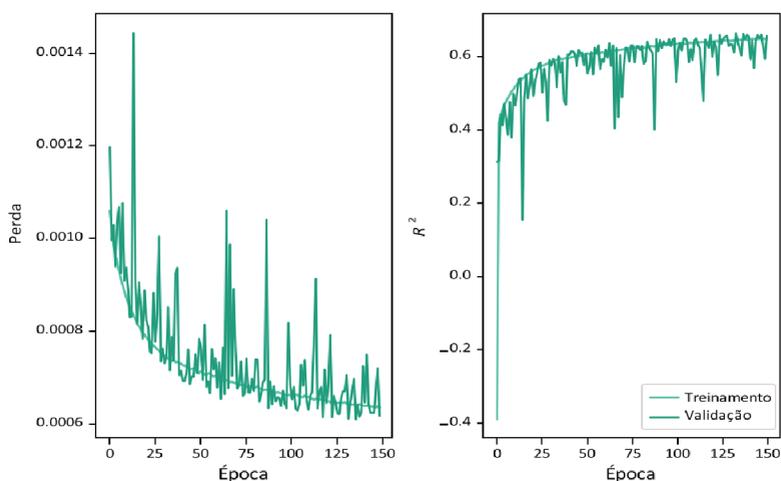
A primeira tentativa realizada na criação do modelo genérico utilizou a abordagem mais simplista possível. Os dados utilizados nos testes preliminares foram agregados, embaralhados e então divididos em conjuntos de treinamentos, validação e teste, utilizando uma proporção de 70%, 20% e 10%, respectivamente. O conjunto de treinamento utilizado no treinamento do modelo, enquanto o de validação serve para avaliar os resultados de cada treinamento. Estes dados são novos para o modelo. Por fim, os dados de teste são separados e utilizados apenas para uma avaliação final do modelo, depois da definição de todos os seus parâmetros. Estas precauções são necessárias para garantir que o modelo criado não seja específico para o *dataset*, o que configuraria um caso de *overfit*.

A rede aqui utilizada possuiu topologia exatamente igual àquela utilizada nos testes preliminares da Seção 4.3. Como resultado da validação cruzada de 5 estágios, obteve-se um R^2 médio de 0,69, o que apesar de não ser um valor desejável, também não é um valor ruim.

A Figura 22 mostra como o erro quadrático médio (que serviu como função de perda durante o treinamento) e o coeficiente de correla-

ção, R^2 , comportaram-se em um treino da rede. Após a finalização de cada época do treinamento, tanto a perda como o R^2 foram calculados para o conjunto de treinamento e o de validação. Na 150ª época tanto a variação da perda como da métrica ainda demonstravam melhoras no modelo, no entanto de forma muito lenta. Além disso, a inexistência de grande divergência entre os valores de treino e de validação indicam que com este número de épocas a rede não sofreu de *overfitting*.

Figura 22 – Evolução das perdas e R^2 no treinamento inicial da rede completa



Fonte – O Autor

Ainda com relação à Figura 22, aparecem valores negativos para R^2 , que podem parecer pouco intuitivos, no entanto, não são incorretos. De acordo com Judge (1982, p. 147), apenas em modelos lineares há garantia de que este valor será entre 0 e 1. Verificando sua definição, Equação 20, onde SSE é a soma dos erros da previsão ao quadrado e SST é a soma total dos quadrados, representando a variação dos valores reais em relação a média, percebe-se que a condição para que valores negativos ocorram é de que $SSE > SST$. No fim, a interpretação de um R^2 negativo é mesma de quando ele é nulo. Significa que o modelo

não é capaz de explicar a variabilidade da variável objetivo dadas as entradas utilizadas.

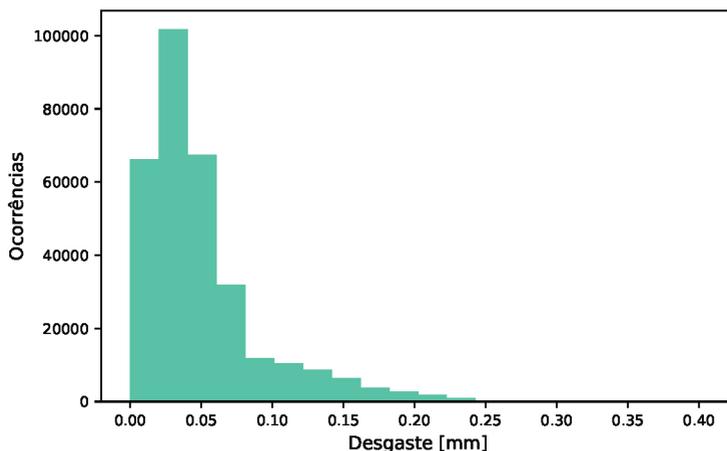
$$R^2 = 1 - \frac{SSE}{SST} \quad (20)$$

4.4.1.2 Modelo com tratamento dos dados desbalanceados

Como os resultados obtidos foram inferiores aos dos testes preliminares, relatados na seção 4.3 e acredita-se que seja possível melhorá-los, foram feitas então algumas alterações neste sentido.

Verificando a distribuição dos dados, mostrada na Figura 23, percebe-se que a população obtida ao fazer-se a união de todos os conjuntos disponíveis é extremamente desbalanceada, podendo esta ser uma das causas da piora no desempenho da rede.

Figura 23 – Distribuição dos valores de desgaste dentre os dados disponíveis



Fonte – O Autor

Existem duas abordagens básicas para a solução deste problema que, apesar de mais diretamente aplicadas em problemas de classifi-

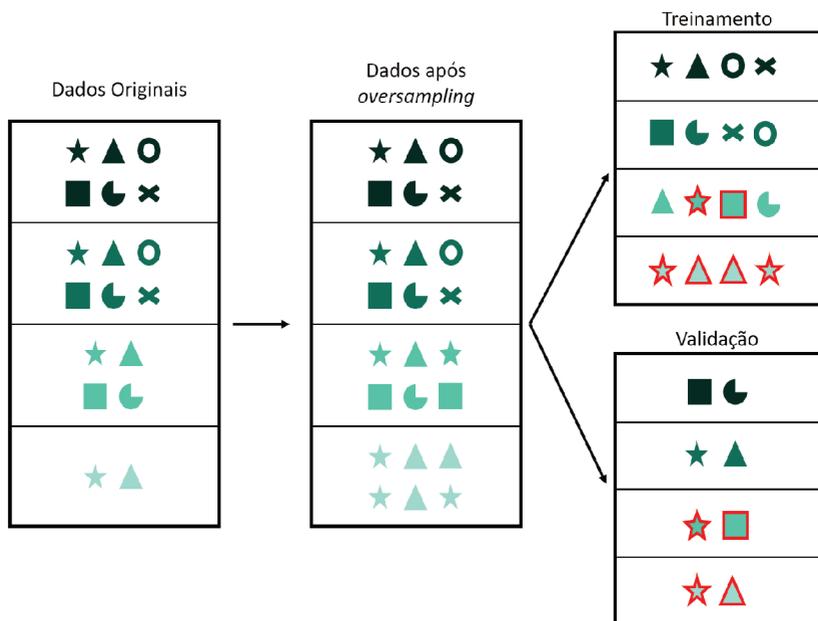
cação, podem ser adaptadas para os casos de regressão. A biblioteca “imbalanced-learn” de Lemaître *et al.* (2017) é de grande valia nestes procedimentos.

A primeira, mais simples, é chamada de *undersampling*. Ela consiste em equilibrar as diferentes categorias, removendo aleatoriamente dados daquelas mais populosas. É de escolha do usuário, de acordo com sua situação, quais categorias serão reduzidas. É possível que apenas a categoria com maior número de dados seja reduzida ou que todas exceto a menor sejam reduzidas. A desvantagem desta técnica está no descarte de uma grande quantidade de dados das categorias mais populosas.

A alternativa ao *undersampling* é o *oversampling*. Nela, a(s) categoria(s) com menos observações é aumentada artificialmente, através da realização de n seleções de observações, onde n representa o tamanho desejado da categoria. Por um lado, esta técnica permite a utilização de um conjunto de dados de treinamento maior, mas apresenta a desvantagem de acabar utilizando dados repetidos. Além disso, é de suma importância que esta seleção dos dados para aplicação do *oversampling* seja realizada após a separação dos conjuntos de treino, validação e teste a fim de evitar que dados utilizados na fase de validação e/ou teste sejam selecionados para o conjunto de treinamento. A Figura 24 exemplifica esta situação, mostrando o que acontece quando a separação dos dados é feita após o *oversampling*. Os dados que aparecem tanto no conjunto de treinamento como no de validação são destacados com um contorno vermelho. Nota-se aqui que, no desenvolvimento do projeto, este erro foi cometido, e de fato levou os resultados da rede neural treinada a serem muito melhores do que deveriam. Numa revisão do trabalho, encontrou e corrigiu-se este erro.

Desta forma, foram testadas ambas as alternativas de balanceamento dos dados de treinamento, obtendo-se as distribuições apresentadas na Figura 25. Nela, a transformação dos dados fica bastante evidente. Estes *datasets* foram então utilizados para o treinamento de duas novas redes (mantendo a topologia utilizada na Seção 4.4.1.1). Os resultados foram, no entanto, menos positivos do que o esperado.

Figura 24 – Demonstração do *oversampling* realizado antes da divisão dos dados para treinamento e validação

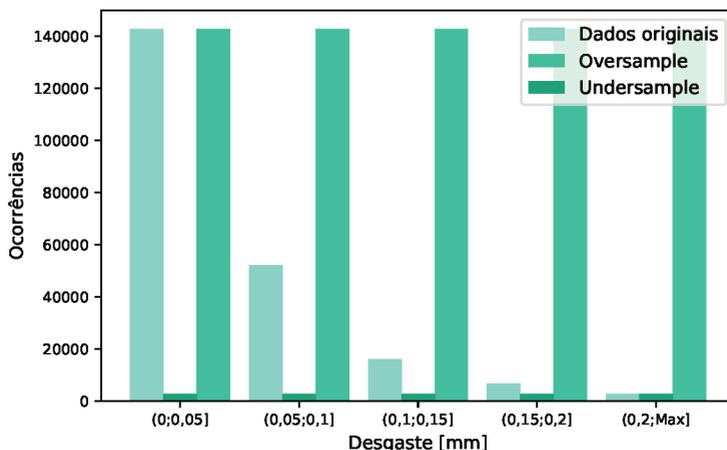


Fonte – O Autor

Como pode ser visto na Figura 26, a métrica do R^2 de validação durante o treinamento de ambas as redes foi ruim, especialmente no caso do treinamento com *undersampling*, além disso, seu crescimento após as 150 épocas rodadas já era praticamente nulo. Quanto às perdas, os resultados também foram piores que os obtidos na rede sem balanceamento dos dados, valendo destacar que a curva de perda da rede com *oversample*, ao fim do treinamento, já apresentava uma leve divergência entre o resultado com dados de treinamento e de validação, indicando uma tendência ao *overfit*.

Os resultados da validação cruzada executada com estes dados também não foram melhores, na realidade, indicaram que as redes não tiveram capacidade alguma de aprender o modelo, o que é evidenciado

Figura 25 – Distribuição dos dados de treinamento antes e após a aplicação da técnicas de *oversampling* e *undersampling*



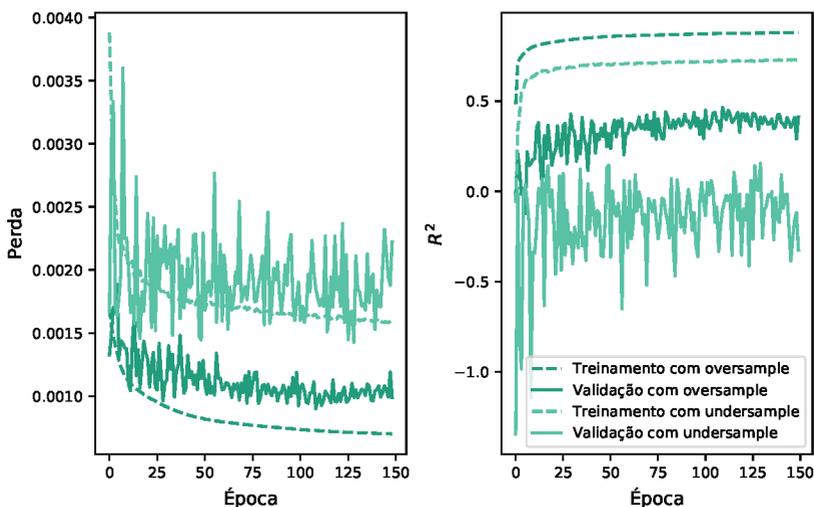
Fonte – O Autor

pelos valores negativos de R^2 , que efetivamente representam um coeficiente nulo. Estes valores de coeficiente de determinação obtidos na validação cruzada são mostrados na Tabela 5. Neste ponto, um fato curioso ocorreu, possivelmente por coincidência. As previsões realizadas com os dados de teste na rede treinada com *oversampling* resultaram num R^2 de 0,674, o que é uma leve melhora em relação ao obtido na rede simplista (e contrariando os resultados da validação cruzada). Por outro lado, por haver uma quantidade de dados muito maior neste caso (mesmo que artificiais), o tempo para treinamento da rede é bastante superior, tendo levado cerca de 30 minutos para rodar as 150 épocas, contra os cerca de 10 minutos da rede simplista.

4.4.2 Modelo de classificação

Uma vez que mesmo os melhores modelos de regressão obtidos foram incapazes de fazer uma previsão adequada do desgaste esperado,

Figura 26 – Evolução da perda e do R^2 durante o treinamento após a aplicação da técnicas de *oversampling* e *undersampling*



Fonte – O Autor

Tabela 5 – R^2 obtido ao executar-se validação cruzada nos dados balanceados

	<i>Oversample</i>	<i>Undersample</i>
1	0,360	-20,625
2	0,082	-24,474
3	-0,200	-6,886
4	-2,036	-5,323
5	-3,838	-3,856
Média	-1,126	-12,233
Desvio Padrão	1,595	8,565

Fonte – O Autor

decidiu-se por reduzir a complexidade do modelo, transformando-os de uma regressão em uma classificação. Para isso, algumas alterações foram necessárias, tantos nos dados como na própria topologia da rede.

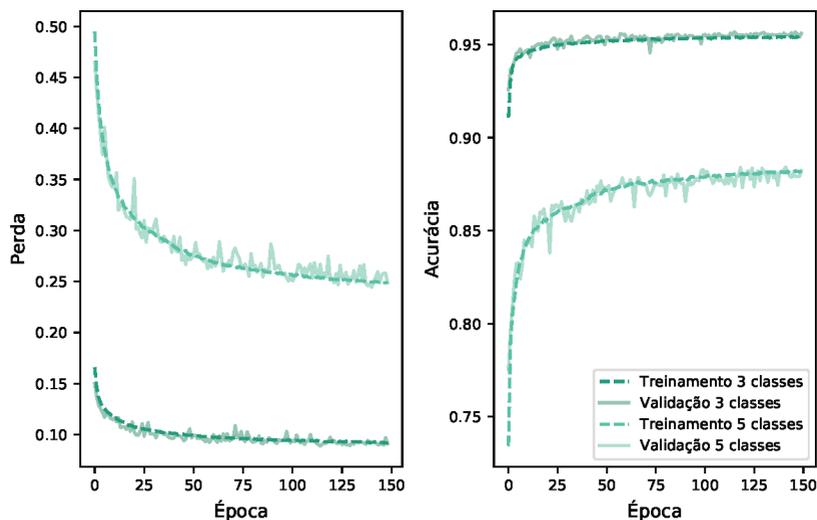
Abordando, primeiramente, a questão dos dados, é necessário iniciar-se pela classificação dos dados de desgaste. Foram criadas duas escalas de desgaste, a primeira, mais simples, usa apenas 3 classes: “baixo” (0-0,1 mm), “médio” (0,1-0,2 mm) e “alto” (0,2 mm ou mais). A segunda escala classifica os desgastes em 5 níveis: “muito baixo” (0-0,05 mm), “baixo” (0,05-0,1 mm), “médio” (0,1-0,15 mm), “atenção” (0,15-0,2 mm) e “alto” (0,2 mm ou mais). A decisão de fazer a classe “alto” abrangendo uma faixa maior de desgastes deu-se pelo fato de, no Fraunhofer IPT, o valor de 0,2 mm ser utilizado para o fim de vida da ferramenta em casos de usinagem de desbaste. Isso significa que, ao classificar-se o desgaste da ferramenta como alto, esta deverá ser substituída. Também foram descartadas as técnicas de balanceamento do *dataset*.

Quanto a questão das alterações na topologia da rede, foi feita uma alteração na camada de saída. Ao invés de apenas um nó, que no caso da regressão dava o resultado de desgaste previsto, aqui utiliza-se um número de nós igual ao número de classes do modelo. Desta forma, o resultado de cada nó indica a probabilidade de estar-se em determinada classe. Aquela classe cuja probabilidade é mais alta é, então, utilizada como a previsão. Além disso, também alterou-se a função de ativação desta camada, de *SELU* para *softmax*, uma função mais adequada a problemas de classificação. Também foram feitas alterações na função objetivo e métrica do modelo. Passou-se a utilizar a entropia cruzada categórica como função de perda e a acurácia como métrica. Por fim, o algoritmo de otimização e a taxa de aprendizado também foram alterados, passando-se do gradiente estocástico descendente para o Adam, com taxa de aprendizado de 0,001.

A evolução do treinamento para ambas as redes é mostrada na Figura 27. Nela, fica claro que o desempenho da rede com apenas três classes é melhor que o da com cinco classes. Também percebe-se que as curvas de validação acompanharam de perto as curvas de treinamento, indicando ausência de *overfitting*. Além disso, fica visível que o treinamento com 150 épocas é desnecessariamente longo para estas redes. No caso com 5 classes, cujo aprendizado levou mais tempo,

a curva de acurácia estabilizou-se por volta da septuagésima época, ou seja, o treinamento foi efetuado pelo dobro do tempo necessário, o que não necessariamente é algo positivo. O aumento da extensão do treinamento pode levar a modelos que sofrem de *overfit*.

Figura 27 – Evolução da perda e da acurácia durante o treinamento para os modelos de classificação com 3 e com 5 classes



Fonte – O Autor

Lançando-se mão dos dados de teste, é possível verificar-se mais alguns detalhes quanto ao desempenho destas redes, os quais são apresentados na Tabela 6 em termos de precisão, *recall* e suporte. A precisão indica a fração de resultados relevantes retornados pelo modelo quando comparados com o total de previsões realizadas, enquanto o *recall* é relacionado capacidade do modelo de prever a classe relevante, independente da presença de previsões falsas. Na tabela percebe-se um padrão interessante. Em ambos os modelos, com exceção da classe “alto”, houve uma redução tanto da precisão como do *recall* com a redução do suporte (número de observações de cada classe), o que era esperado. Uma possível explicação para a classe “alto” não seguir este padrão é o fato

de esta abranger uma faixa maior de desgastes. Também calculou-se a acurácia média da validação cruzada de ambos os modelos, resultando em 0,95 no caso com três classes e 0,89 no caso com 5 classes, uma diferença pequena.

Tabela 6 – Avaliação por classes dos modelos de classificação

Nível de desgaste	Precisão	Recall	F1-score	Suporte
3 Classes				
Baixo	0,97	0,98	0,98	27486
Médio	0,82	0,80	0,81	3289
Alto	0,77	0,45	0,57	413
5 Classes				
Muito Baixo	0,95	0,93	0,94	20034
Baixo	0,83	0,84	0,83	7452
Médio	0,72	0,76	0,74	2374
Atenção	0,63	0,66	0,64	915
Alto	0,64	0,72	0,68	413

Realmente, a simplificação introduzida pelo modelo de classificação aparenta ter gerado resultados melhores do que os anteriormente obtidos através da regressão. Embora os resultados aqui obtidos não possam ser comparados diretamente com os de regressão, por usarem métricas diferentes, os valores indicam uma melhora significativa. Para viabilizar esta comparação, fez-se um tratamento das previsões de teste da rede de regressão simplista (por ter sido, até aqui, a que melhor equilibrou complexidade do modelo e resultados), aplicando-se a eles uma classificação. Na prática, o que foi feito foi classificar os dados de testes, fazer uma previsão a partir dos valores numéricos e classificar esta previsão posteriormente. Então, aplicaram-se as métricas de classificação a este modelo de regressão.

Os resultados desta avaliação do modelo simplista são mostrados na Tabela 7. Calculou-se, ainda, os valores de acurácia neste caso. Para a divisão em três classes a acurácia foi de 0,94 enquanto para o de cinco classes obteve-se 0,79 como resultado. De fato, sob esta óptica o modelo

simplicista de regressão parece ser melhor do que quando avaliado como uma regressão, no entanto, seus resultados foram piores do que os do modelo de classificação em praticamente todas as métricas utilizadas

Tabela 7 – Avaliação por classes do modelo de regressão simplicista

Nível de Desgaste	Precisão	Recall	F1-Score	Suporte
3 Classes				
Baixo	0,97	0,97	0,97	27486
Médio	0,68	0,79	0,73	3289
Alto	0,94	0,31	0,46	413
5 Classes				
Muito Baixo	0,91	0,85	0,88	20034
Baixo	0,63	0,72	0,67	7452
Médio	0,54	0,72	0,61	2374
Atenção	0,52	0,38	0,44	915
Alto	0,94	0,31	0,46	413

Fonte – O Autor

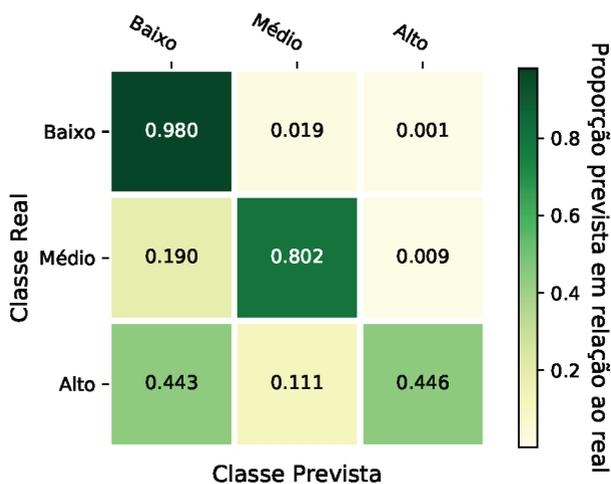
É importante ressaltar que, apesar de os resultados obtidos na rede de classificação serem, de forma geral, satisfatórios, eles não são excelentes. Dentre as métricas apresentadas para cada classe, a mais importante é o *recall* (em português, revocação), especialmente nas classes “alto”, justamente a medida mais baixa obtida. Essa medida nesta classe é importante pois representa a quantidade de ferramentas que o modelo consegue prever como tendo atingido seu fim de vida, ou seja, um valor de 1 indicaria que nenhum caso em que o desgaste da ferramenta fosse “alto” seria previsto em outra classe. Claro que, um modelo que sempre preveja a mesma classe, independente dos dados de entrada, terá um *recall* de 1 nesta classe, no entanto, sua precisão será baixíssima. Num sentido prático, é importante identificar adequadamente o fim de vida de uma ferramenta pois a utilização de ferramentas com desgaste excessivo implica em fabricação de peças que não atendem às especificações de projeto, aumentando os custos do processo devido

a refugos e retrabalhos.

Mais algumas informações sobre as redes de classificação podem ser obtidas utilizando-se suas matrizes de confusão. Na matriz de confusão é possível verificar-se, de forma absoluta ou percentual, a quantidade de previsões realizadas corretamente (elementos da sua diagonal principal), assim como em qual classe as previsões incorretas (elementos de cada linha, que não os da diagonal principal) se enquadraram. Nas Figuras 28 e 29 são apresentadas as matrizes de confusão dos dois modelos de classificação criados. Na Figura 28, por exemplo, é possível ver que cerca de 44% das observações da classe “alto” foram previstas como sendo da classe “baixo”, quase a mesma razão prevista corretamente. Isso indica uma falha do modelo com 3 classes, pois prever que uma ferramenta desgastada possui desgaste baixo é um erro mais danoso do que prever que esta mesma possui um nível de desgaste intermediário. No caso da rede com 5 classes, este erro também mostrou-se presente, mas num nível muito mais aceitável. Enquanto 73% das previsões estavam corretas, cerca de 10% estava na classe mais baixa de desgaste.

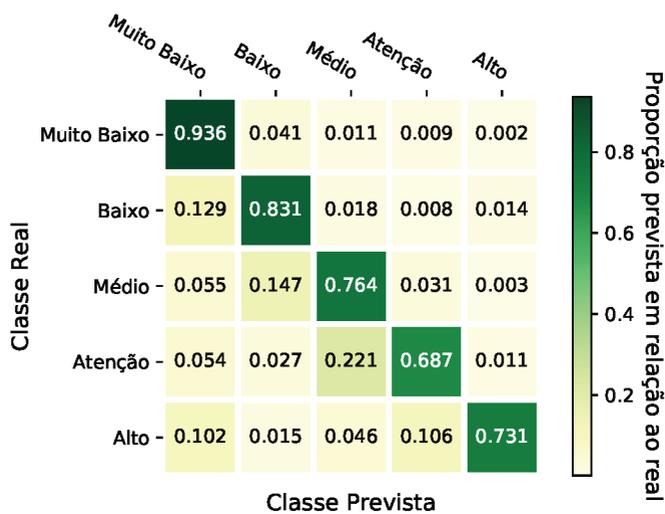
Por fim, pretendia-se realizar uma comparação entre os resultados obtidos com os fornecidos pelo modelo de Taylor, que funcionaria como um *benchmark*. Essa possibilidade foi, no entanto, descartada justamente devido a maior fraqueza de Taylor. Para que seja aplicado de forma confiável, é necessário que pelo menos três experimentos, nos quais a única variável é a velocidade de corte sejam realizados desde o ponto em que a ferramenta é nova até que seu fim de vida seja atingido. Ao analisar os dados disponíveis e utilizados para criação do modelo aqui proposto, verificou-se que apenas uma pequena parte atende ao critério de variar apenas a velocidade de corte e, destes, nenhum atendeu ao critério de ter sido realizado de forma completa.

Figura 28 – Matriz de confusão para o modelo com 3 classes



Fonte – O Autor

Figura 29 – Matriz de confusão para o modelo com 5 classes



Fonte – O Autor

5 CONCLUSÃO

Este trabalho foi construído em cima de um trabalho realizado pelo autor durante seu estágio curricular no Fraunhofer IPT, em Aachen (Alemanha). Originalmente foi criada uma rede neural cujo objetivo era fazer previsões para o desgaste de fresas de topo esférico utilizadas na usinagem de superfícies complexas. Este modelo que havia sido criado, apesar de ser capaz de obter resultados que se mostravam promissores, carecia de cuidados metodológicos tanto para sua criação como para sua avaliação. Desta forma, neste trabalho o modelo para previsão do desgaste destas ferramentas foi recriado, dando-se uma maior atenção a sua metodologia, mas utilizando o trabalho original como ponto de partida. De fato, ao fim do trabalho, obteve-se um modelo capaz de realizar esta previsão (ainda que com suas limitações), cumprindo seu objetivo geral.

A utilização de modelos de aprendizado de máquina (onde estão incluídas as redes neurais) requer uso intensivo de dados empíricos, que no caso de desgaste de ferramentas possuem custo elevado para obtenção. Desta forma, recorreu-se à literatura para sua obtenção. Estes dados foram ainda, complementados pela realização de simulações em um software desenvolvido internamente no Fraunhofer IPT, os quais forneceram dados relativos à geometria do cavaco esperada nos respectivos processos de usinagem.

Os dados coletados foram tratados, removendo inconsistências e/ou situações que não se aplicavam a pesquisa e, então, foram iniciados os testes de diferentes modelos de redes neurais. Foram utilizados, como *features* do modelo, 5 parâmetros:

- altura do cavaco
- largura do cavaco
- profundidade do cavaco
- número de interações entre cada gume da ferramenta e a peça cortada

- dureza do material usinado

Com as *features* definidas (em parte, no trabalho original) fez-se primeiramente testes com cada um dos artigos individualmente, o que forneceu boas previsões, com R^2 de 0,6 a 0,9, dependendo do artigo. O próximo passo foi de unir todos os dados e montar uma única rede neural aplicável nos diferentes casos.

Inicialmente, esta rede neural geral deveria ser muito similar à dos testes preliminares, no entanto, isto não foi possível. Tanto o modelo simplista quanto os modelos com *over-/undersampling* apresentaram resultados pouco úteis devido à sua imprecisão (o modelo simplista foi o único capaz de realizar previsões, atingindo um R^2 médio de 0,69). Desta forma, passou-se a testar redes de classificações do nível de desgaste encontrado. A rede com 5 classes de desgaste foi capaz de fazer boas previsões, com elevada acurácia na validação cruzada (0,89) e resultados adequados (ainda que não excelentes) para a precisão e *recall* da classe de desgaste denominada como “alto”, a mais relevante.

5.1 CONSIDERAÇÕES FINAIS

A escolha por trabalhar-se desde o início com redes neurais, um tipo de modelo de aprendizado de máquina relativamente complexo, pode parecer arbitrária e, de certa forma, realmente o foi. A utilização de uma rede neural era um requisito do trabalho inicialmente realizado durante o estágio curricular do autor.

Os resultados finais obtidos foram capazes de cumprir o objetivo geral do trabalho, no entanto, acredita-se ser possível aperfeiçoá-los. Uma primeira opção seria justamente a utilização de outros modelos de aprendizado de máquina, como, por exemplo, uma *Random Forest*. Poderia-se, ainda, realizar a previsão apenas do incremento de desgaste após uma interação entre a ferramenta e a peça, ao invés da previsão de desgaste absoluto após determinada quantidade de interações feita neste trabalho. Uma terceira opção seria a de realizar um trabalho mais extenso na escolha e geração das *features* a partir dos dados disponíveis,

a parte do processo de aprendizado de máquina que mais depende de intuição e experiência do usuário (DOMINGOS, 2012).

Algumas dificuldades foram encontradas durante a realização do trabalho e podem ser divididas em duas principais frentes. A primeira foi justamente a falta de experiência do autor com modelos de aprendizado de máquina, mais ainda com redes neurais. Apesar de ter estudado o assunto durante a graduação, nunca havia utilizado-o de forma prática. A segunda foi uma certa dificuldade para utilizar a biblioteca Keras, que apesar de ser uma das principais bibliotecas para redes neurais em Python, apresenta alguns *bugs* antigos, amplamente reportados no Github, mas que nunca foram corrigidos.

Os códigos utilizados no desenvolvimento deste trabalho podem ser acessados em: https://github.com/eduardolp/tcc_toolwear.

REFERÊNCIAS

AMT/USCTI. **2019 U.S. Cutting Tool Orders Close at \$2.4 Billion.** [*S.l.: s.n.*], 2020. Disponível em: <https://www.sme.org/technologies/articles/2020/february/2019-u.s.-cutting-tool-orders-close-at-2.4-billion/>. Acesso em: 19 jun. 2020.

ARRAZOLA, P. J.; ÖZEL, T.; UMBRELLO, D.; DAVIES, M.; JAWAHIR, I. S. Recent advances in modelling of metal machining processes. **CIRP Annals**, v. 62, n. 2, p. 695–718, 2013. ISSN 00078506. DOI: 10.1016/j.cirp.2013.05.006.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6158: Sistema de Tolerância e Ajustes.** [*S.l.: s.n.*], 1995.

ATTANASIO, Aldo; CERETTI, Elisabetta; GIARDINI, Claudio. Analytical Models for Tool Wear Prediction During AISI 1045 Turning Operations. **Procedia CIRP**, v. 8, p. 218–223, 2013. ISSN 22128271. DOI: 10.1016/j.procir.2013.06.092.

BENCHEIKH, I.; BILTERYST, F.; NOUARI, M.; MAKICH, H. Wear Estimation of Coated Tools Using XFEM / Level Set Function. **Procedia CIRP**, v. 58, p. 428–433, 2017. ISSN 22128271. DOI: 10.1016/j.procir.2017.03.252.

CHOLLET, François *et al.* **Keras.** [*S.l.: s.n.*], 2015. <https://keras.io>.

DINIZ, A.E; FERREIRA, J.R; FILHO, F.T. Influence of refrigeration/lubrication condition on SAE 52100 hardened steel turning at several cutting speeds. **International Journal of Machine Tools and Manufacture**, v. 43, n. 3, p. 317–326, 2003. ISSN 08906955. DOI: 10.1016/S0890-6955(02)00186-4.

DOMINGOS, Pedro. A few useful things to know about machine learning. **Communications of the ACM**, v. 55, n. 10, p. 78–87, 2012. ISSN 0001-0782. DOI: 10.1145/2347736.2347755.

GIL, Antonio Carlos. **Métodos e técnicas de pesquisa social**. 6. ed., 6ª reimpressão. São Paulo: Atlas, 2014. ISBN 978-85-224-5142-5.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. [S.l.]: MIT Press, 2016. Acesso em: 24 jun. 2020.

GOPALSAMY, Bala Murugan; MONDAL, Biswanath; GHOSH, Suktamal; ARNTZ, Kristian; KLOCKE, Fritz. Experimental investigations while hard machining of DIEVAR tool steel (50 HRC). **The International Journal of Advanced Manufacturing Technology**, v. 51, n. 9-12, p. 853–869, 2010. ISSN 0268-3768. DOI: 10.1007/s00170-010-2688-1.

HASTIE, Trevor; TIBSHIRANI, Robert; FRIEDMAN, J. H. **The elements of statistical learning: Data mining, inference, and prediction**. 2nd ed. New York NY: Springer, 2009. (Springer series in statistics). ISBN 978-0-387-84858-7.

HINTON, Geoffrey; SRIVASTAVA, Nitish; SWERSKY, Kevin. **Neural Networks for Machine Learning: Why do we need machine learning?** Toronto: [s.n.], 2014. Disponível em: https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec1.pdf. Acesso em: 5 nov. 2019.

HOOD, R.; ASPINWALL, D. K.; SAGE, C.; VOICE, W. High speed ball nose end milling of g-TiAl alloys. **Intermetallics**, v. 32, p. 284–291, 2013. ISSN 09669795. DOI: 10.1016/j.intermet.2012.09.011.

HUANG, Yong; DAWSON, Ty G. Tool crater wear depth modeling in CBN hard turning. **Wear**, v. 258, n. 9, p. 1455–1461, 2005. ISSN 00431648. DOI: 10.1016/j.wear.2004.08.010.

HUNTER, J. D. Matplotlib: A 2D graphics environment. **Computing in Science & Engineering**, IEEE COMPUTER SOC, v. 9, n. 3, p. 90–95, 2007. DOI: 10.1109/MCSE.2007.55.

ILSCHNER, B.; SINGER, Robert F. **Werkstoffwissenschaften und Fertigungstechnik: Eigenschaften, Vorgänge, Technologien**. 5., neu bearbeitete Aufl. Berlin: Springer, 2010. (Springer-Lehrbuch). ISBN 978-3-642-01733-9.

JOHANSSON, Daniel; HÄGGLUND, Sören; BUSHLYA, Volodymyr; STÅHL, Jan-Eric. Assessment of Commonly used Tool Life Models in Metal Cutting. **Procedia Manufacturing**, v. 11, p. 602–609, 2017. ISSN 23519789. DOI: 10.1016/j.promfg.2017.07.154.

JUDGE, George G. **Introduction to the theory and practice of econometrics**. New York: Wiley, 1982. (Wiley series in probability and mathematical statistics). ISBN 0-471-08277-5.

JUNEJA, B. L.; SEKHON, G. S.; SETH, Nitin. **Fundamentals of metal cutting and machine tools**. 2nd ed. New Delhi: New Age International Ltd. Publishers, 2003. ISBN 8122414672.

KINGMA, Diederik P.; BA, Jimmy. **Adam: A Method for Stochastic Optimization**. [*S.l.*: *s.n.*], 2014. Disponível em: <https://arxiv.org/pdf/1412.6980>.

KLAMBAUER, Günter; UNTERTHINER, Thomas; MAYR, Andreas; HOCHREITER, Sepp. Self-Normalizing Neural Networks, 2017. Disponível em: <http://arxiv.org/pdf/1706.02515v5>.

KLOCKE, Fritz; ARNTZ, Kristian; CABRAL, Gustavo Francisco; STOLLORZ, Martin; BUSCH, Marc. Characterization of Tool Wear in High-Speed Milling of Hardened Powder Metallurgical Steels. **Advances in Tribology**, v. 2011, p. 1–13, 2011. ISSN 1687-5915. DOI: 10.1155/2011/906481.

KLOCKE, Fritz; KÖNIG, Wilfried. **Fertigungsverfahren**. Berlin e New York: Springer, 2008. ISBN 978-3-540-23458-6.

LEMAÎTRE, Guillaume; NOGUEIRA, Fernando; ARIDAS, Christos K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. **Journal of Machine Learning Research**, v. 18, n. 17, p. 1–5, 2017. Disponível em: <http://jmlr.org/papers/v18/16-365.html>.

LI, Bin. A review of tool wear estimation using theoretical analysis and numerical simulation technologies. **International Journal of Refractory Metals and Hard Materials**, v. 35, p. 143–151, 2012. ISSN 02634368. DOI: 10.1016/j.ijrmhm.2012.05.006.

MACKAY, David J. C. **Information theory, inference and learning algorithms**. Rev. ed. Cambridge: Cambridge University Press, 2008. ISBN 9780521642989.

MANYIKA, James; CHUI, Michael; BROWN, Brad; BUGHIN, Jacques; DOBBS, Richard; ROXBURGH, Charles; HUNG BYERS, Angela. **Big data: The next frontier for innovation, competition, and productivity**. [S.l.: s.n.], 2011.

MIGUEL (COORD.), Paulo Augusto Cauchick. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2. ed. Rio de Janeiro: Elsevier, 2012. ISBN 978-8535248913.

MIKOŁAJCZYK, T.; NOWICKI, K.; BUSTILLO, A.; YU PIMENOV, D. Predicting tool life in turning operations using neural networks and image processing. **Mechanical Systems and Signal Processing**, v. 104, p. 503–513, 2018. ISSN 08883270. DOI: 10.1016/j.ymsp.2017.11.022.

MINSKY, Marvin; PAPERT, Seymour A. **Perceptrons: An introduction to computational geometry**. 2. print. with corr. Cambridge/Mass.: The MIT Press, 1972. ISBN 0262130432.

PÁLMAI, Z. Proposal for a new theoretical model of the cutting tool's flank wear. **Wear**, v. 303, n. 1-2, p. 437–445, 2013. ISSN 00431648. DOI: 10.1016/j.wear.2013.03.025.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

POLYAK, B. T. Some methods of speeding up the convergence of iteration methods. **USSR Computational Mathematics and Mathematical Physics**, v. 4, n. 5, p. 1–17, 1964. ISSN 00415553. DOI: 10.1016/0041-5553(64)90137-5.

PRIETO, Alberto; PRIETO, Beatriz; ORTIGOSA, Eva Martinez; ROS, Eduardo; PELAYO, Francisco; ORTEGA, Julio; ROJAS, Ignacio. Neural networks: An overview of early research, current frameworks and new challenges. **Neurocomputing**, v. 214, p. 242–268, 2016. ISSN 09252312. DOI: 10.1016/j.neucom.2016.06.014.

PU, Zhengwen; SINGH, Anshul. High speed ball nose end milling of hardened AISI A2 tool steel with PCBN and coated carbide tools. **Journal of Manufacturing Processes**, v. 15, n. 4, p. 467–473, 2013. ISSN 15266125. DOI: 10.1016/j.jmapro.2013.05.005.

- RAMÍREZ P., Fernando; SOLDANI, Xavier; LOYA, José; MIGUÉLEZ, Henar. A new approach for time-space wear modeling applied to machining tool wear. **Wear**, v. 390-391, p. 125–134, 2017. ISSN 00431648. DOI: 10.1016/j.wear.2017.07.015.
- SALVATORE, F.; SAAD, S.; HAMDY, H. Modeling and Simulation of Tool Wear During the Cutting Process. **Procedia CIRP**, v. 8, p. 305–310, 2013. ISSN 22128271. DOI: 10.1016/j.procir.2013.06.107.
- SHAFIQ TAHIR, Mohamad Amir; GHANI, Jaharah A.; RIZAL, Muhammad; NUAWI, Mohd Zaki; CHE HARON, Che Hassan. Prediction of Tool Wear for Ball End Nose in Milling Inconel 718 Using a Feed Forward Back Propagation Neural Network. **Australian Journal of Basic and Applied Sciences**, 8(5) Special, p. 383–390, 2014.
- SHARMAN, Adrian; DEWES, Richard C.; ASPINWALL, David K. Tool life when high speed ball nose end milling Inconel 718TM. **Journal of Materials Processing Technology**, v. 118, n. 1-3, p. 29–35, 2001. ISSN 09240136. DOI: 10.1016/S0924-0136(01)00855-X.
- SHENG, Huang. **MODEL-BASED TOOL CONDITION MONITORING FOR BALL-NOSE END MILLING**. 2012. Tese de doutorado – NATIONAL UNIVERSITY OF SINGAPORE, Singapura.
- SILVA, G. C.; MALVEIRA, B. M.; CARNEIRO, J.R.G.; BRITO, P. P.; SILVA, T. A. Wear and Thermal Analysis of WC Inserts in Turning Operations by Fuzzy Modeling. **Procedia CIRP**, v. 58, p. 523–528, 2017. ISSN 22128271. DOI: 10.1016/j.procir.2017.03.266.
- STEMMER, Caspar Erich. **Ferramentas de Corte I**. [S.l.]: Editora da UFSC, 1993.

STENBERG, Niclas; DELIĆ, Aldin; BJÖRK, Thomas. Using the SPH Method to Easier Predict Wear in Machining. **Procedia CIRP**, v. 58, p. 317–322, 2017. ISSN 22128271. DOI: 10.1016/j.procir.2017.03.234.

STOLORZ, Martin. **Prediction and compensation of machining-induced shape deviations in finish milling: Vorhersage und Kompensation von prozessinduzierten Formabweichungen beim Schlichtfräsen**. 1. Auflage. Aachen: Apprimus, 2017. Band 26/2017. (Prozesstechnologie). ISBN 978-3-86359-538-8.

SUTSKEVER, Ilya; MARTENS, James; DAHL, George; HINTON, Geoffrey. On the importance of initialization and momentum in deep learning. **Proceedings of the 30th International Conference on Machine Learning**, v. 2013/28, p. 1139–1147, 2013.

TANGJITSITCHAROEN, Somkiat; JATINANDANA, Thanathip; SENJUNTICHAJ, Angsumalin. Intelligent Prediction of Tool Wear in Ball-End Milling Process Based on Dimensionless Cutting Force Ratio. **Applied Mechanics and Materials**, v. 799-800, p. 312–318, 2015. DOI: 10.4028/www.scientific.net/AMM.799-800.312.

TÖNSHOFF, Hans Kurt; DENKENA, Berend. **Spanen: Grundlagen**. 3., erw. und neu bearb. Aufl. Berlin: Springer, 2011. (Engineering online library). ISBN 978-3-642-19771-0.

TWARDOWSKI, Pawel; LEGUTKO, Stanislaw; KROLCZYK, Grzegorz M.; HLOCH, Sergej. Investigation of wear and tool life of coated carbide and cubic boron nitride cutting tools in high speed milling. **Advances in Mechanical Engineering**, v. 7, n. 6, p. 168781401559021, 2015. ISSN 1687-8140. DOI: 10.1177/1687814015590216.

UCHIKATA, Soushi; BEAUCAMP, Anthony; TAKEUCHI, Yoshimi. 5-Axis Control Finishing for Decreased Tool Wear. **Procedia CIRP**, v. 63, p. 313–317, 2017. ISSN 22128271. DOI: 10.1016/j.procir.2017.03.320.

USUI, E.; SHIRAKASHI, T.; KITAGAWA, T. Analytical prediction of cutting tool wear. **Wear**, v. 100, n. 1-3, p. 129–151, 1984. ISSN 00431648. DOI: 10.1016/0043-1648(84)90010-3.

VOPÁT, Tomáš; PETERKA, Jozef; ŠIMNA, Vladimír; KURUC, Marcel. The Influence of Different Types of Copy Milling on the Surface Roughness and Tool Life of End Mills. **Procedia Engineering**, v. 100, p. 868–876, 2015. ISSN 18777058. DOI: 10.1016/j.proeng.2015.01.443.

WOJCIECHOWSKI, Szymon; TWARDOWSKI, Paweł. Tool Life and Process Dynamics in High Speed Ball End Milling of Hardened Steel. **Procedia CIRP**, v. 1, p. 289–294, 2012. ISSN 22128271. DOI: 10.1016/j.procir.2012.04.052.

ZHANG, Haiyan; ZHANG, Chen; ZHANG, Jilin; ZHOU, Laishui. Tool wear model based on least squares support vector machines and Kalman filter. **Production Engineering**, v. 8, n. 1-2, p. 101–109, 2014. ISSN 0944-6524. DOI: 10.1007/s11740-014-0527-1.